Impact of Chemical Dynamics of Commercial PURE Systems on Malachite Green Aptamer Fluorescence

Zoila Jurado[1,*] and Richard M. Murray[1]

[1]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA
[*]Corresponding author: zjuradoq@caltech.edu

March 16, 2024

Abstract

The malachite green aptamer (MGapt) is known for its utility in RNA measurement *in vivo* and lysate-based cell-free protein systems. However, MGapt fluorescence dynamics do not accurately reflect mRNA concentration. Our study finds that MGapt fluorescence varies in commercial PURE systems, leading to inaccurate RNA calculations. MGapt fluorescence is not solely linked to waste products or buffering agent interactions. MGapt dynamics are also independent of DNA and RNA concentrations, suggesting an influence beyond general system dynamics. Specific to the commercial system, we posit that MGapt fluorescence is significantly affected by the system's chemical properties, governed notably by the presence of dithiothreitol (DTT). We propose a model that, on average, accurately predicts MGapt measurement within a 10 % margin, leveraging DTT concentration as a critical factor. This model sheds light on the complex dynamics of MGapt in cell-free systems and underscores the importance of considering environmental factors in RNA measurements using aptamers.

Introduction

Cell-free protein synthesis (CFPS) systems can be broadly categorized into two types. The first and most commonly used CFPS system is based on cell lysate. The cell lysate-based system uses cellular machinery harvested from the cell [1]. The second category contains all the necessary transcription and translation proteins for *E. coli*, each cultured and purified individually and combined at known concentrations, known as PURE — Protein synthesis Using purified Recombinant Elements [2]. A variant of PURE is OnePot PURE, where all 36 proteins are co-cultured and purified together [3].

Using PURE facilitates the modeling of the CPFS system, allowing for the development of a complete model and techniques to seamlessly integrate it into the "design-build-test" pipeline for genetic circuit construction or synthetic cell assembly. Models in recent years have modeled the translation of peptides for PURE using chemical reaction networks [4, 5], and our previous work has added to these models by expanding the user peptide and the addition of transcription [6]. Validation of all models requires accurate transcription and translation monitoring. While translation can be measured using fluorescent proteins, transcription is more challenging, leading to the reliance

on RNA aptamers, such as malachite green aptamer. Initially used as a means to control gene expression in *S. cerevisiae* [7], MGapt has been used as a means to study RNA production, RNA dynamics, and investigating trade-offs between transcription and translation in cell-free protein systems with protein expression of deGFP [8, 9, 10].

Studies using MGapt to measure RNA production in CFPS have been used in lysates [9, 11]. Our use of MGapt in PURExpress reveals surprising dynamics, which may explain why the MGapt measurements appear absent from PURE CFPS. Generally, the exceptional specificity of aptamers allows the discrimination between closely related isoforms or different conformational states of the same target molecule [12, 13]. However, MGapt has been known to bind to other triphenylmethane dyes such as crystal violet (CV), tetramethylrhodamine (TMR), and Pyronin Y (PY) [14, 15]. Furthermore, it has been found that free MGapt reduces the amount of RNA folded in the correct binding conformation, and metal ions, though not required for high-affinity bind [16], stabilize the complexes with non-native ligands. In contrast, the complex with the original selection target is stable at low salt and without divalent metal ions [17]. The destabilization of MGapt through the use of organic solvent, the addition of Mg^{2+}, DTT, and other ions has been described in the past [18, 19, 20, 21, 15, 22].

In this paper, we demonstrate how the chemical composition of commercial PURE may destabilize MGapt, leading to different aptamer states corresponding to different fluorescence levels. We will first describe the observed MGapt expression in commercially available PURE and inconsistencies between what we could predict and what was observed, such as saturation time and dynamics of MGapt when no transcription is involved. We then discuss the potential effects of DTT in the system and propose a model that uses DTT as a driving force. Finally, we provide experimental validation of the measured MGapt model of the PURE system.

Results and Discussion

Expression and fluorescence of MGapt in different PURE systems. We initially evaluated the consistency of MGapt dynamics across two distinct PURE systems to ascertain whether this phenomenon is inherent to PURE. In commercial PURE systems, the protocol recommends incubating the reaction for 2-4 hours [23, 24], during which protein production saturates. RNA production also would be expected to saturate around the 2-hour mark to be consistent with the reaction lifespan. However, MGapt fluorescence does not saturate at 2 hours. In comparison in PURExpress and OnePot PURE, both expressing MGapt from plasmid DNA with construct pT7-MGapt-tT7 (light blue) and pT7-MGapt-UTR1-deGFP-tT7 (dark blue), MGapt fluorescence dynamics were striking, shown in Figure 1.

Experimental results indicate that the two PURE systems had different MGapt dynamics, irrespective of the total MGapt production. While each system exhibited a monotonic upward trend of MGapt fluorescence, the PURExpress reaction (Figure 1a) showed that MGapt fluorescence saturated at around six hours. On the other hand, the OnePot PURE reaction (Figure 1b) showed that MGapt fluorescence saturated at around two hours. The MGapt saturation could indicate that MGapt production continues after two hours in PURExpress, effectively out-living OnePot PURE's RNA production. It didn't appear to be logical for the transcription process to persist beyond the translation. Furthermore, if MGapt fluorescence is considered an accurate reflection of RNA production, it would also suggest that the RNA production rate changes around four hours. The significant disparity in MGapt production dynamics suggested MGapt fluorescence might increase due to chemical differences between PURExpress and OnePot PURE.

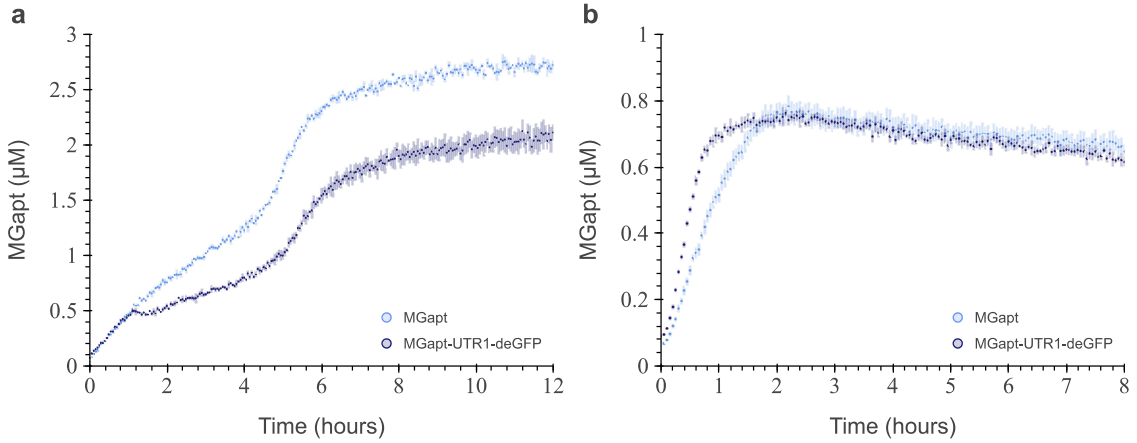

Figure 1: MGapt concentration in PURExpress and OnePot CFPS system. The MGapt concentration over time for transcriptions of pT7-MGapt-tT7 (light blue) and pT7-MGapt-UTR1-deGFP-tT7 (dark blue) at 5 nM, in (**a**) PURExpress and (**b**) OnePot, made in lab. Experimental data consisted of three replicates (circles and respective error bars).

Effects of buffering conditions MGapt fluorescence. Upon investigating the difference, we first the two systems chemical compositions [2, 25]; outlined in Table S1. We discovered numerous salts in one system but not the other, which we concluded should not significantly impact MGapt fluorescence. The one major difference in the two systems was the reducing agent used, tris(2-carboxyethyl)phosphine (TCEP) versus dithiothreitol (DTT). PURExpress uses 1 mM of DTT [26] while OnePot PURE uses 1 mM of TCEP [3]. DTT and TCEP both reduce disulfide bonds, but TCEP has the advantages of being significantly more stable in the absence of a metal chelator and less inhibitive in labeling with maleimide [27]. Additionally, compared to DTT, TCEP is more effective, non-volatile, and does not readily oxidize above pH 7.5 [28]. It is important to note that the PURE reaction occurs above pH 7.5.

MGapt fluorescent was measured under four different chemical conditions to measure the effect of the reducing agent on MGapt fluorescence. Three different reducing agents were tested: TCEP, glutathione (GSH), and DTT), and the waste product pyrophosphate (PPi). The experiment consisted of three replicates of 10 µL reactions with the respective additives (TCEP, GSH, DTT, and PPi) at experimentally relevant concentrations, and purified RNA of MGapt-UTR1-deGFP at 0.51 µM. MGapt fluorescence was read in BioTek plate reader (610/650) for six hours at 37 °C and calibrated to µM using calibration curve Figure S3. Figure 2 summarizes the average MGapt concentration measured over the six-hour window, full time course measurements are depicted in Figure S1.

Only, TCEP and PPi samples showed concentration effects on the total MGapt fluorescence. Higher concentrations of TCEP and PPi resulted in higher MGapt fluorescence, except at 4 mM PPi, where MGapt fluorescence decreased. The results from Figure 2 did not provide conclusive evidence that the reducing agent, DTT, directly affected MGapt fluorescence nor that other reducing agents (TCEP or GSH) would be a better alternative.

The fluorescence of MGapt is affected by the chemical properties of the PURE system. While Figure 2 implies that the fluctuating concentration of PPi might lead to increased

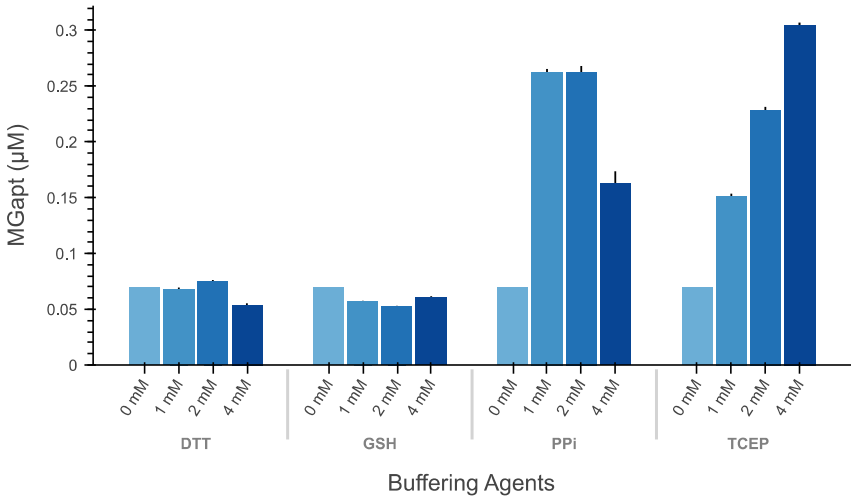

Figure 2: Measured MGapt concentration under different buffer conditions. Average measured MGapt concentration of purified RNA coding for MGapt-UTR1-deGFP in different reducing agents and waste chemicals over six hours. Each section of the bar plot is titled with the respective buffer used, and the chemical concentrations are labeled in different shades of blue. The negative control was subtracted from the test conditions, and standard deviations of triplicates are represented by black bars.

MGapt fluorescence, it's important to note that waste production would be directly correlated with system expression. Therefore, we wouldn't anticipate a greater or quicker production of PPi in the PURExpress compared to the OnePot PURE. To avoid any possible interference of transcription products with MGapt, purified RNA of MGapt-UTR1-deGFP was added removing the largest PPi production source, RNA strand elongation. Additionally, to isolate the effect of the PURE reaction on MGapt fluorescence, the translation reactions were minimized by using RNA containing only the MGapt. Upon first observation, it is observed that the RNA only containing MGapt measures lower than the RNA of MGapt-UTR1-deGFP. Additionally, can see that in both cases, with 0.49 µM of MGapt-deGFP-RNA (crosses) or 0.53 µM of MGapt-RNA (circles), the measured MGapt concentration is not constant, Figure 3a. we had anticipated an initial reduction in the fluorescence of MGapt within the first 30 min, followed by relatively stable readings afterward, similar to the MGapt measurements depicted in Figure S1.

However, when we normalize each result by its respective maximum, Figure 3b, the concentration of MGapt seems to increase gradually over time, despite the absence of transcription. Interestingly, the environmental effect appeared to be similar for both the TL-null and TL-only conditions, indicating that the additional ribosome binding site and deGFP sequences could not be solely responsible for affecting MGapt fluorescence.

To further demonstrate that the measured MGapt concentration over time was not dependent on the production of RNA or protein, additional PURE reactions were performed at different concentrations of DNA, Figure 3c. The experiment consisted of three replicates of 10 µL with DNA plasmid pT7-MGapt-UTR1-deGFP-tT7 at six different DNA concentrations dispensed using an Echo 525 Acoustic Liquid Handler and the final DNA concentration was calculated on the rounded amount of DNA that was dispensed. To account for total RNA production, the measured MGapt

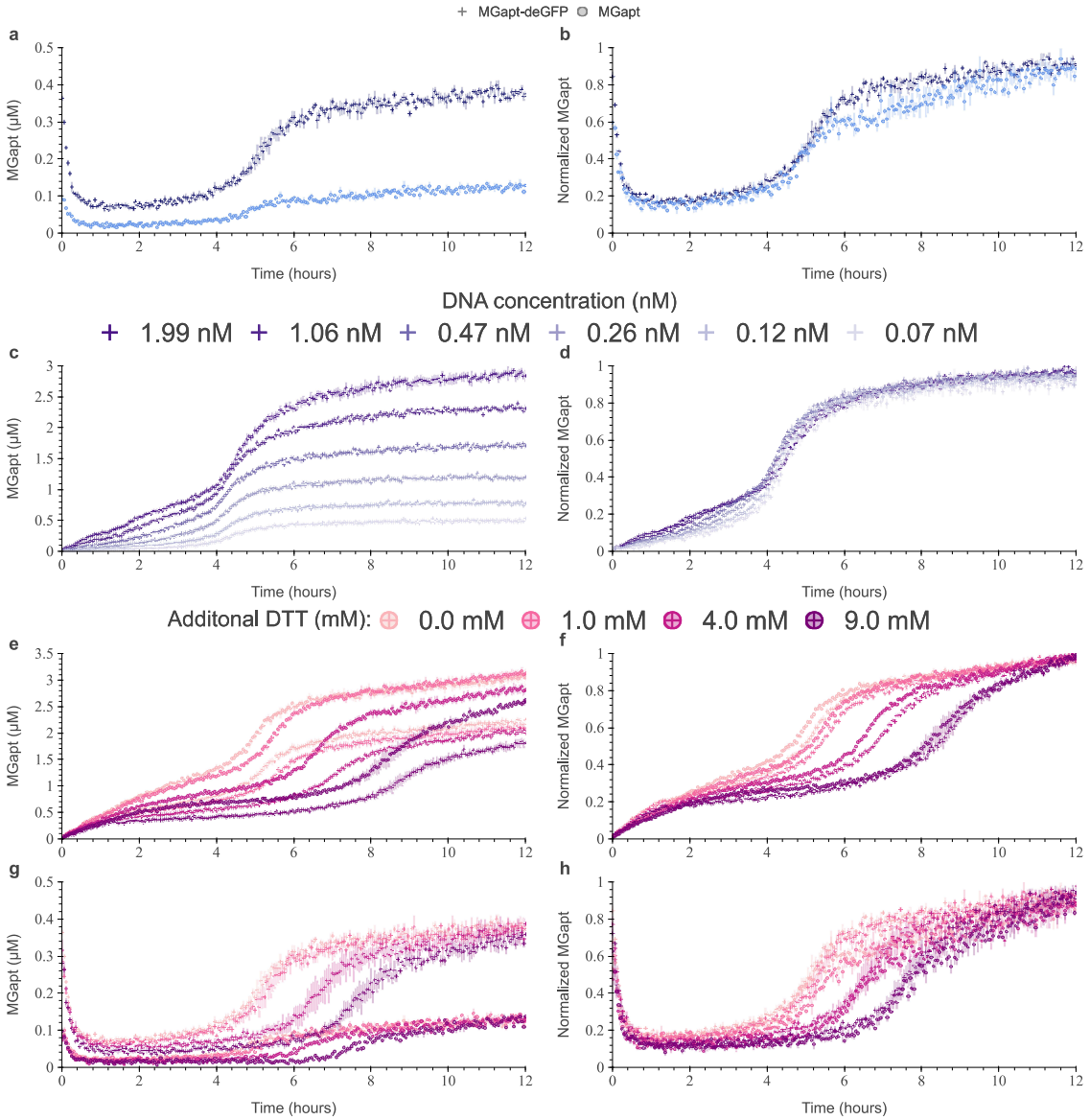

Figure 3: Measured MGapt concentration dynamics in PURExpress. PURExpress reactions of 10 µL with three technical replicates containing 10 µM of Malachite Green dye and 8 units of RNAse inhibitor. (**a**) Starting from purified RNA of MGapt at 0.53 µM (light blue crosses) and MGapt-UTR1-deGFP at 0.49 µM (dark blue circles) and (**b**) after data is normalized. (**c**) Expressing plasmid pT7-MGapt-UTR1-deGFP-tT7 at various DNA concentrations and (**d**) after data is normalized. (**e**) Expressing of plasmid pT7-MGapt-UTR1-deGFP-tT7 at 4.93 nM and pT7-MGapt-tT7 at 5.03 nM with increasing DTT concentration and (**f**) after data is normalized. (**g**) Starting from purified RNA of MGapt-UTR1-deGFP at 0.49 µM and MGapt at 0.53 µM with increasing DTT concentration and (**h**) after data is normalized.

concentrations were normalized, Figure 3d. The initial RNA production rate showed variation based on DNA concentrations, but all normalized MGapt concentration measurements converged. Notably, all concavity transitions occurred around the 4-hour mark regardless of RNA production rate, indicating that MGapt dynamics are unaffected by DNA concentration. The increase in MGapt concentration over time further supports that MGapt fluorescence is affected by the chemical environment of the commercial PURE system, independent of transcription and translation.

The observed inflection point around 4 hours, in Figures 3a-d, suggests the presence of system repression by specific chemical substrates that degrade over time. As illustrated in the model for PURE translation [4], there are 483 reactions with a nonzero rate, and 278 reactions occur without the presence of DNA: tRNA charging, NTP degradation, and energy recycling. These reactions also occur in OnePot PURE, indicating that this MGapt response wouldn't be exclusive to one system. This leads back to the distinctions between OnePot PURE and commercial PURE, particularly their selection of reducing agents—DTT versus TCEP. While the reducing agent may not directly impact MGapt fluorescence, it could potentially influence other chemical reactions that affect MGapt fluorescence. To test the effects of DTT on measured MGapt concentration, DTT was added to PURExpress expressing DNA plasmids of pT7-MGapt-UTR1-deGFP-tT7 and pT7-MGapt-tT7 at 5 nM. Additional DTT to reach a final added DTT concentration of 1 mM, 4 mM,and 9 mM, Figure 3e. Results showed that increasing DTT concentrations shifted the infection point of measured MGapt to the right. Following data normalization, Figure 3f, we observed that the measured MGapt concentrations for both plasmids overlap, indicating Mgapt effects are primarily due to the addition of DTT.

Finally, the previous experiment was repeated with purified RNA to underscore the concept that DTT has a direct impact and can effectively represent other auxiliary reactions and chemicals affecting MGapt. In this iteration, purified RNA of MGapt-UTR1-deGFP and MGapt at approximately 0.5 µM concentration was combined with varying concentrations of DTT, see Figure 3g. Consistent with previous findings, as the concentration of DTT increased, the addition of DTT seems to inhibit MGapt fluoresces. This effect is more evident in the normalized data, Figure 3h, data collapse based on added DTT concentration. Based on these experimental results, we conclude that DTT acts as a direct or indirect suppressor of MGapt fluorescence, which degrades over time due to its volatile nature. This signifies that using MGapt measurements as an indicator for RNA is inaccurate as it does not account for DTT's effect.

MGapt transitions different aptamer states with different fluorescence levels. We assert that MGapt in the commercial PURE reaction goes through different equilibrium and intermediates states, as MGapt has been previously reported [22, 15, 19]. The concentration of DTT plays a crucial role in meditating, promoting, and/or facilitating the instability of MGapt and the conversion between the two chemical states, as demonstrated in Figure 3e-h. Using DTT as a proxy for the full chemical reactions responsible for altering MGapt states and thus fluorescence in the commercial PURE reaction, we propose the model depicting MGapt inhibition by DTT, as illustrated in Figure 4.

To model the detailed step-by-step inhibition and conversion of MG_{dye} bounded to MGapt ($MGapt_{bound}$) to a less fluorescent $MGapt_{altered}$, we built a chemical reaction network (CRN) in a CRN compiler tool called BioCRNpyler [30]. We utilize the functionalities provided by BioCRNpyler model the reactions shown in Figure 4:

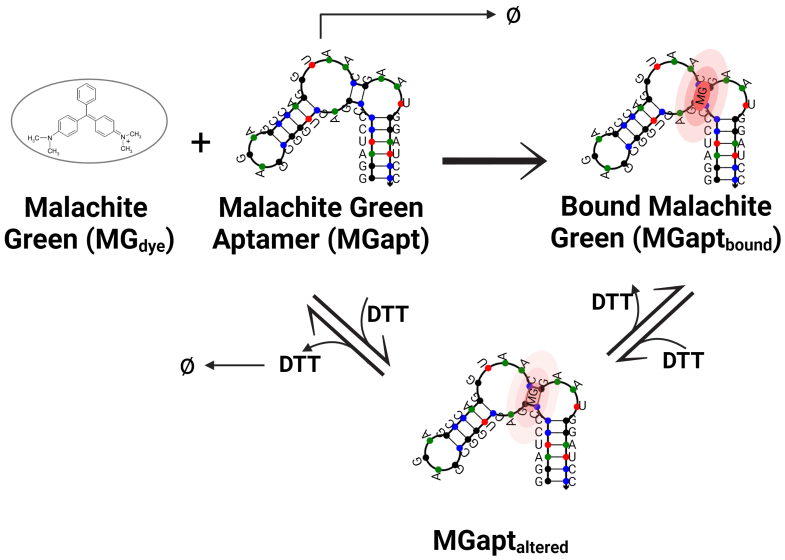

Figure 4: Inhibition of malachite green aptamer by DTT. Illustration of proposed interaction of DTT on bound fluorescent malachite green aptamer (MGapt). Adapted from chemical structure Stead *et al.* [19] and malachite green aptamer NuPack predicted secondary structure [29]. Created with BioRender.com

$$\mathrm{MGapt} + \mathrm{MG_{dye}} \xrightarrow{\mathrm{k_{MGapt}}} \mathrm{MGapt_{bound}} \tag{1}$$

$$\mathrm{MGapt_{bound}} + \mathrm{DTT} \underset{\mathrm{ku_{DTT}}}{\overset{\mathrm{kb_{DTT}}}{\rightleftharpoons}} \mathrm{MGapt_{altered}} \tag{2}$$

$$\mathrm{MGapt_{altered}} \underset{\mathrm{kb_{mgapt}}}{\overset{\mathrm{ku_{mgapt}}}{\rightleftharpoons}} \mathrm{mgapt} + \mathrm{DTT} + \mathrm{MG_{dye}} \tag{3}$$

$$\mathrm{mgapt} \xrightarrow{\mathrm{k_{RNAdeg}}} \emptyset \tag{4}$$

$$\mathrm{DTT} \xrightarrow{\mathrm{k_{DTTdeg}}} \emptyset \tag{5}$$

Beginning with transcribed and folded malachite green aptamer state (MGapt), malachite green oxalate ($\mathrm{MG_{dye}}$) binds to mgapt to form fully fluorescent malachite green aptamer ($\mathrm{MGapt_{bound}}$). Concentrations of DTT induce instability in $\mathrm{MGapt_{bound}}$ through a reversible process, leading to an intermediate state with reduced fluorescence ($\mathrm{MGapt_{altered}}$). The $\mathrm{MGapt_{altered}}$ state can be further destabilized, causing dissociation into its three components: $\mathrm{MG_{dye}}$, mgapt, and DTT. Finally, unbound states such as mgapt and DTT are susceptible to degradation. The introduction of the altered state, $\mathrm{MGapt_{altered}}$, which exhibits lower fluorescence compared to $\mathrm{MGapt_{bound}}$, implies that the total MGapt fluorescence is a linear combination of both states, $\mathrm{MGapt_{measured}} = \mathrm{MGapt_{bound}} + \mathrm{c_1 MGapt_{altered}}$. The $\mathrm{MGapt_{measured}}$ represents the measured fluorescence detected by the BioTek plate reader or any other fluorescent reader.

Analysis and identification of parameters in the proposed model. To analyze the model of the effects of DTT on MGapt measured, we measured the MGapt fluorescence in a 10 µL reaction done in triplicate using PURExpress. The reaction contained 0.22 µM of RNA of MGapt-UTR1-deGFP. To determine the coefficients of the linear combination, we utilized the lowest concentration measurement to indicate when 100 % of the aptamer was in the $\mathrm{MGapt_{altered}}$ state. The coefficient c_1 was then determined based on the percentage of $\mathrm{MGapt_{measured}}$ and total added RNA ($\mathrm{MGapt_{total}}$), resulting in the final linear system:

$$\mathrm{MGapt_{measured}} = \mathrm{MGapt_{bound}} + 0.15\,\mathrm{MGapt_{altered}} \tag{6}$$

$$\mathrm{MGapt_{total}} = \mathrm{MGapt_{bound}} + \mathrm{MGapt_{altered}} \tag{7}$$

To parameterize the CRN model constructed using BioCRNpyler, we initialized the parameters by hand-tuning the reactions for bounded conditions by steps. Due to the time required to prepare the reactions before measuring on BioTek, it was necessary to determine the initial conditions at the time of reading. Employing the system of equations, we computed the initial conditions based on the $\mathrm{MGapt_{measured,t=0}}$ reading. Our calculations revealed that approximately 45 % of the system was in the $\mathrm{MGapt_{bound}}$ state, while another 55 % was in the $\mathrm{MGapt_{altered}}$ state. The specific values of the initial conditions utilized are outlined in Table S2.

For accurate predictions of the measured MGapt concentration in PURE from the BioTek readings, we conducted model training using experimental data from the measured MGapt concentration of 0.22 µM RNA with the MGapt-UTR1-deGFP construct. Training this model using parameter identification was challenging due to the combined fluorescence of two distinct states in the measured data. With the ability to train only on one species, the effective concentration of $\mathrm{MGapt_{bound}}$ was calculated using the set of linear equations. Ultimately, in addressing the inherent noise in the experimental data, we employ Bayesian inference to derive a distribution of potential parameter values based on the experimental data. To accomplish both tasks of evaluating identifiability and determining posterior parameter distributions, we utilize a biological data analysis pipeline [31] implemented with the Python package Bioscrape [32].

Parameters for identification were determined by conducting a local sensitivity analysis of all species in the model across all parameters and time. The sensitivity analysis heatmap for DTT's interaction with RNA of MGapt-UTR1-deGFP model is depicted in Figure 5a. We selected parameters with the highest sensitivity regarding MGapt output fluorescence. Among the seven reaction rates in equations (1)-(5), we identify $\mathrm{k_{MGapt}}$, $\mathrm{kb_{DTT}}$, $\mathrm{ku_{mgapt}}$, $\mathrm{kb_{mgapt}}$, and $\mathrm{k_{DTTdeg}}$ as the most sensitive parameters.

Subsequently, we utilize Bayesian inference tools in Bioscrape to ascertain the posterior parameter distributions for these five parameters. The corner plot in Figure 5b illustrates the posterior parameter distributions and their covariance, providing a sampling distribution for predicting output using the fitted model. Model simulations using parameter values drawn from the posterior, alongside experimental data, are presented in Figure 5c. The final parameter values utilized are detailed in Table S4.

Validation and assessment of model capturing DTT's impact on MGapt fluorescence. To validate the model of the effects of DTT on MGapt measured, we conducted simultaneous tests using multiple RNA concentrations. The RNA of MGapt-UTR1-deGFP at final concentrations of 0.41 µM, 0.86 µM, 1.26 µM and 1.67 µM were read for 12 hours at 37 °C in a BioTek reader at 610/650 (ex/em). The initial concentrations of MGapt and $\mathrm{MGapt_{altered}}$ at $t = 0$ were determined as previously outlined and are listed in Table S3. Finally, utilizing equation (6), the $\mathrm{MGapt_{measured}}$

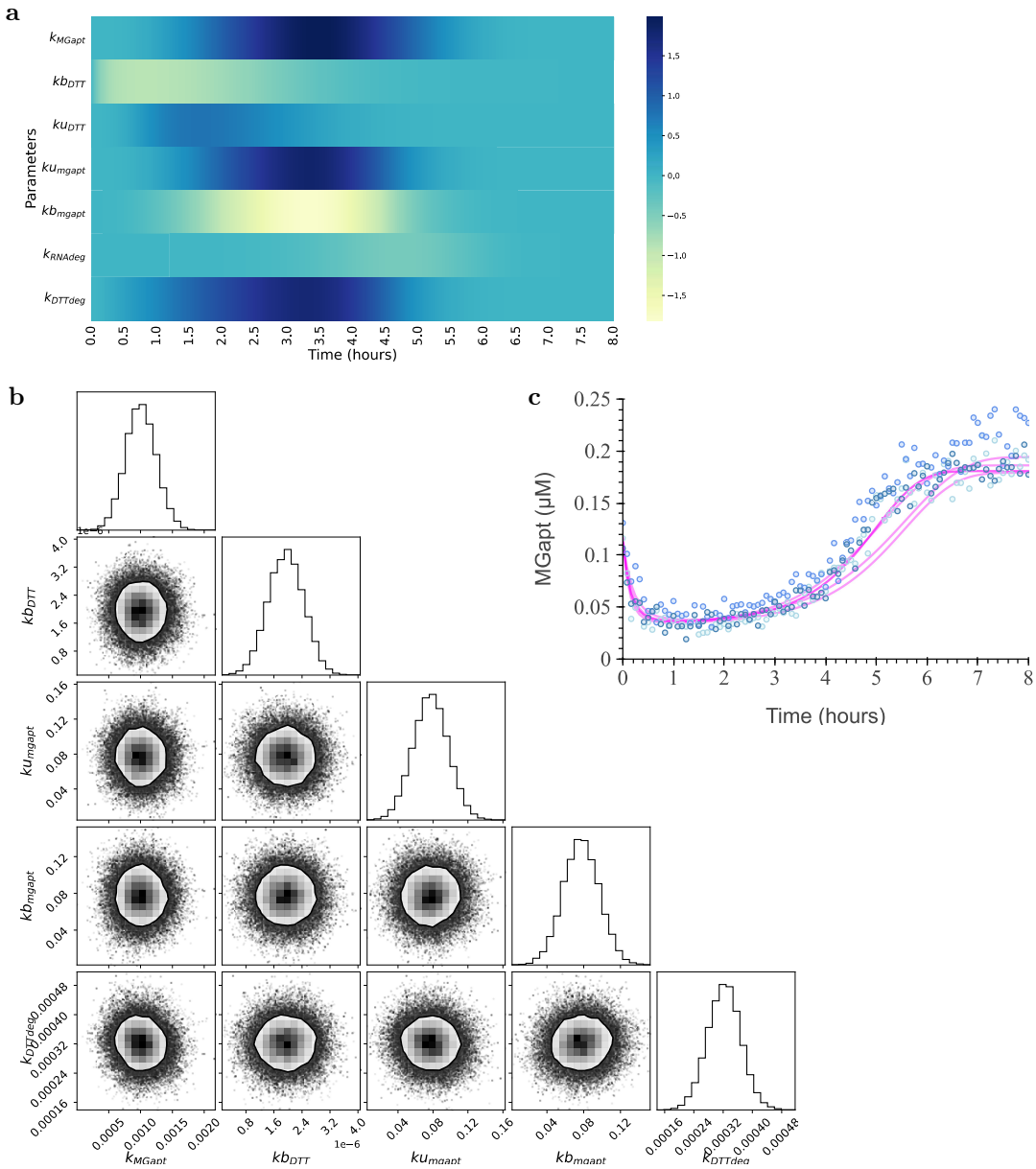

Figure 5: Modeling, analysis, and parameter inferencing of DTT effects on MGapt model. (**a**) The sensitivity of the MGapt to the CRN model parameters for the initial four hours. (**b**) The posterior distributions of parameters obtained after running Bayesian inference on k_{MGapt}, kb_{DTT}, ku_{mgapt}, kb_{mgapt}, and k_{DTTdeg}. The corner plot depicts the covariance of the two parameters, with the contour showing the 75 % probability region for the parameter values. (**c**) With parameter values for k_{MGapt}, kb_{DTT}, ku_{mgapt}, kb_{mgapt}, and k_{DTTdeg} sampled from the posterior distributions, the model simulations (magenta lines) are shown alongside the experimental data for three biological replicates (scattered blue points).

was computed and superimposed onto the experimental results using their respective colors shown in Figure 6.

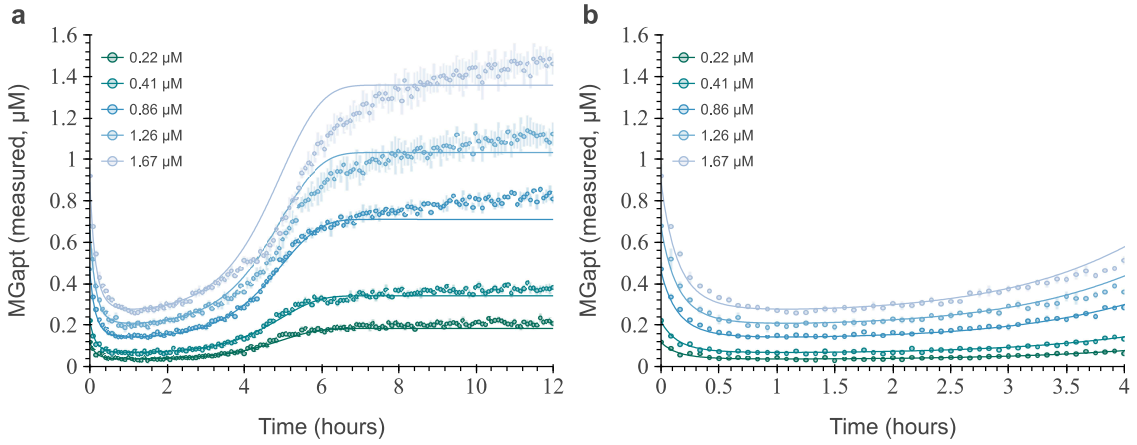

Figure 6: Modeled MGapt$_{\text{measured}}$ at different initial RNA concentrations The modeled MGapt$_{\text{measured}}$ of BioTek at different RNA concentrations (solid line) overlaying with experimental data, three replicates (circles and error bars, of respective colors). (**a**) Full measured time of 12 hours. (**b**) Zoomed, reaction relevant time of 4 hours.

As seen in Figure 6a, the predicted MGapt$_{\text{measured}}$ concentration remains reasonably accurate throughout the measured time for concentrations below 1 µM. The increased error above below 1 µM of RNA is most likely due to changes in DTT degradation or increased stability of the MGapt$_{\text{altered}}$ state. Regardless, the model does recapitulate the increase of the measured MGapt concentration and its overall dynamics. Furthermore, as indicated in the manuals of commercial PURE systems, the reaction is generally monitored for a 2 h- 4 h rather than 12 hour. Upon closer examination of the pertinent 4 h reaction window, highlighted in Figure 6b, it becomes evident that the model accurately predicts the measured MGapt$_{\text{measured}}$. Moreover, the model consistently predicts the measured MGapt within an average of 10 % margin of error (refer to Figure S4) within the initial 4 hour period for RNA concentration is greater than 1 µM, longer the recommended reading time. The accuracy and the capturing of the dynamics of measured MGapt fluorescence demonstrate that DTT can serve as a representative model for MGapt fluorescence, commonly employed in measuring RNA production. Moreover, the simulation can be used in reverse to calculate the total RNA in a system by calculating the total RNA ratio between measured and total to calculate back the amount of RNA, as illustrated in Figure S5.

Conclusion

Our experiments reveal that the chemical makeup of commercial PURE systems influences measured MGapt concentration. Specifically, we observe a correlation between the chemical properties of DTT, given that DTT is absent in OnePot PURE but present in other commercial PURE systems. Though DTT and TCEP are effective reducing agents commonly used in biochemical research; DTT is more sensitive to oxidation, less stable, and more violative. We hypothesize that the concentration

of DTT suppresses the fluorescence of MGapt by reversibly converting it into alternative forms with reduced fluorescence, which are indistinguishable and challenging for readers to account for. The degradation of DTT over time, driven by its chemical properties, enables MGapt to return to its fully fluorescent state, dynamics observed using purified RNA. We demonstrated that we could lengthen the suppression time by increasing the initial concentration of DTT in the expression of either transcription, translation expression, or both.

Moreover, given the inability to reduce the concentration of DTT in most commercial PURE systems, we introduce a model to predict the measured MGapt concentration by accounting for DTT's impact on the state of MGapt. We identified a distribution of possible parameters in the model with the experimental data at one concentration of RNA for MGapt-UTR1-deGFP. We validated our model by accurately predicting the measured MGapt for the same RNA construct at five different concentrations within the relevant 2 hour period.

This model is essential in comprehending and accurately quantifying transcription within cell-free expression systems, specifically commercial PURE systems. Employing our methodology enables the construction of mathematical models based on the measured MGapt concentration and can be utilized to retroactively calculate the total quantity of RNA used in the PURE reaction. More importantly, the realization that MGapt can transition between various fluorescence states in cell-free expression systems suggests potential inaccuracies in our understanding and assessment of RNA production. While MGapt is widely and extensively utilized as an aptamer, this phenomenon may not be exclusive. Exploring other aptamers and monitoring fluorescence changes over time could prove beneficial in identifying additional chemical reactions or environmental factors that might influence the measured RNA concentration.

Materials and Methods

Computational Modeling and Simulations. In this paper, we build a deterministic model of two fluorescently different MGapt states driven by the concentration of DTT. This model is based on MGapt different states effects by organic solvents found Stead *et al.*, Zhou *et al.*, and Da Costa *et al.* [22, 15, 19]. We use the chemical reaction network (CRN) formalism to create the detailed mechanistic model using a CRN compiler called BioCRNpyler [30]. The computational model was developed to take a total RNA concentration of MGapt-UTR1-deGFP and predict the measured MGapt concentration using a BioTek plate reader. Our CRN model consists of five distinct species and seven total reactions.

The BioCRNpyler tool generates models in the Systems Biology Markup Language (SBML)[33], a standard format for biological modeling. These SBML files can be simulated using any compatible SBML simulator. In our case, we utilize the Bioscrape Python package [32] for simulation. Bioscrape converts the Chemical Reaction Network (CRN) model to ordinary differential equations (ODEs) and employs Python's `odeint` to solve them based on specified initial conditions. Each reaction rate in the CRN is expressed using mass-action propensity [34] for this conversion. We opt for Bioscrape because it supports sensitivity analysis and Bayesian inference tools alongside model simulations. Local sensitivity analysis is conducted for each SBML model to determine the sensitivity of measured species to all parameters over time. Subsequently, we identify the most sensitive parameters using experimental data. Parameter identification is achieved through a Bayesian inference algorithm implemented in Bioscrape, utilizing the emcee Python package [35]. By incorporating experimental data, we obtain probability distributions for each identified parameter through Bayesian inference. Model simulations, using parameter values sampled from these posterior probability distributions,

are then compared against experimental data to assess model prediction quality. These posterior probability distributions also quantify the uncertainty in the data, highlighting a significant advantage of Bayesian inference methods.

Summarizing the computational analysis for the PURE transcription model, we identified five out of seven parameters using the experimental data. The final trained model predicts the BioTeK measurement of MGapt concentration over time for the RNA sequence of MGapt-UTR1-deGFP. All calculations and plotting were performed using the standard stack of Python packages – NumPy [36], SciPy [37], Pandas [38], Matplotlib [39], Bokeh [40], and Seaborn [41]. The simulation time for one RNA concentration is less than a tenth of a second on a personal computer running AMD Ryzen 7 4700U 2.0 GHz with 16GB of RAM.

DNA constructs. The original DNA plasmids were procured from the Arbor Biosciences myTXTL Toolbox 2.0 plasmid collection [42], pTXTL-T7p14-mGapt and pTXTL-T7p14-deGFP. The deGFP protein is on a T7 promoter with a UTR1 ribosome binding site (RBS). Primers from Integrated DNA Technologies were employed to clone the MGapt between the promoter and the RBS site. Detailed information regarding the forward and reverse primers utilized can be found in the Supplementary Information Table S5.

Subsequently, both plasmids were initially transformed into JM109 cells and cultivated overnight on plates containing carbenicillin resistance at 37 °C. Colonies from each plate were selected and cultured in 4.5 mL LB medium with 4.5 µL of carbenicillin (100 mg/µL) overnight. Glycerol stocks of each plasmid were prepared by mixing 500 µL of liquid culture with an equal volume of 50 % glycerol, while the remaining culture was miniprep using the Qiagen Miniprep Kit. Before running DNA in the cell-free reaction, all plasmids underwent an additional PCR purification step utilizing a QiaQuick column (Qiagen) to eliminate excess salt. The purified plasmids were then eluted and stored in nuclease-free water at 4 °C for short-term storage and at −20 °C for long-term storage.

RNA constructs. Purified RNA was made using the original plasmid of pTXTL-T7p14-mGapt (pT7-MGapt-tT7) and the modified pTXTL-T7p14-deGFP (pT7-MGapt-UTR1-deGFP-tT7) plasmid. Purified plasmids were first amplified and linearized using Q5 Hot Start High-Fidelity DNA Polymerase (NEB); forward and reverse primers are listed in Table S5. Eight reactions of 125 µL, for a total volume of 1 mL were run in a thermocycler. The thermocycler was run for 30 cycles with an elongation time of 15 sec or 45 sec, respectively, and annealing temperature of 65 °C. The completed PCR reactions were then combined into three microcentrifuge tubes of 333 µL to which 33 µL sodium acetate (3 M) and 1 mL Ethanol (100 %) was added. All three tubes were placed at −80 °C for 20 min and then centrifuged at 16 000 xg for 30 min at 4 °C to remove all traces of ethanol. The precipitated DNA was then resuspended in 50 µL of nuclease-free water, combined, and stored at 4 °C.

Using the HiScribe T7 High Yield RNA Synthesis Kit (NEB) 1 µg of the amplified and linearized DNA construct was transcribed in an 80 µL reaction volume and incubated overnight at 37 °C. Before the RNA purification procedure, the transcription reaction underwent DNAse I treatment for 15 min at 37 °C to remove the linear DNA template. Ethanol (100 %) was then added to achieve a final concentration of 35 %, to precipitating the RNA. Next, using PureLink RNA Mini Kit (Invitrogen), RNA was isolated following the protocol for 'Purification of RNA from liquid samples' starting from Step 2-Bind RNA. The yield and quality of the RNA samples were analyzed using the NanoDrop2000c measuring at UV absorbance at 260 nm. The purified RNA was flash frozen in liquid nitrogen and stored at −80 °C.

PURE reactions and fluorescence measurements. The PURE reactions were prepared according to the NEB PURExpress (E6800) protocol, adapted for a 10 µL reaction volume, and incubated in a 384-well plate (Nunc) at 37 °C. A concentration of 5 nM DNA was utilized, unless

otherwise specified, along with 8 units of RNAse inhibitor (NEB), and 10 µM malachite green oxalate added to each reaction. The RNA concentrations varied and were specified in the corresponding experimental setup description. DNA and RNA were added using an Echo 525 Acoustic Liquid Handler and final concentrations were recalculated based on dispensed volumes. The remaining components were made into a master mix with 5 % excess, added to a 384-well plate using a repeater pipette.

Fluorescence readings were obtained using a BioTek Synergy H1 plate reader (BioTek) at 3-minute intervals for 12 hours at 37 °C. Excitation/emission wavelengths were set at 610/650 nm for MGapt with a gain of 150. All samples were analyzed using the same plate reader. For RNA calibrations, relative fluorescence units (RFUs) were converted to µM of RNA using purchased MGapt from IDT. Dilutions of MGapt (rArCrUrGrGrArUrCrCrCrGrArCrUrGrGrCrGrArGrArGrCrCrArGrGrUrArArCrGrArArUrGrGrArUrCrCrArArU) in PBS were run, and multiple reads per concentration were averaged to make a calibration plot, Figure S3.

Acknowledgments

We thank Zachary Martinez for their encouragement during the start of this project. We sincerely appreciate Manisha Kapasiawala, Dr. Yan Zhang, and Dr. John Marken for their invaluable insights and thought-provoking discussions on the experiments and modeling framework throughout the project. We would also like to acknowledge the contributions of Dr. Ayush Pandey, who provided assistance and feedback on the proposed model's modeling, analysis, and inference. Their involvement has been indispensable to its completion.

Research supported by the National Science Foundation award number 2152267. The computations presented here were conducted in the Resnick High Performance Computing Center, a facility supported by the Resnick Sustainability Institute at the California Institute of Technology, Pasadena, CA, USA.

References

[1] Zachary Z. Sun, Clarmyra A. Hayes, Jonghyeon Shin, Filippo Caschera, Richard M. Murray, and Vincent Noireaux. Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *JoVE*, page e50762, 2013.

[2] Yoshihiro Shimizu, Akio Inoue, Yukihide Tomari, Tsutomu Suzuki, Takashi Yokogawa, Kazuya Nishikawa, and Takuya Ueda. Cell-free translation reconstituted with purified components. *Nature Biotechnology*, 19(8):751–755, 2001.

[3] Barbora Lavickova and Sebastian J. Maerkl. A simple, robust, and low-cost method to produce the PURE cell-free system. *ACS Synthetic Biology*, 8(2):455–462, 2019.

[4] Tomoaki Matsuura, Naoki Tanimura, Kazufumi Hosoda, Tetsuya Yomo, and Yoshihiro Shimizu. Reaction dynamics analysis of a reconstituted *Escherichia coli* protein translation system by computational modeling. *Proceedings of the National Academy of Sciences*, 114(8):E1336–E1344, 2017.

[5] Tomoaki Matsuura, Kazufumi Hosoda, and Yoshihiro Shimizu. Robustness of a reconstituted *Escherichia coli* protein translation system analyzed by computational modeling. *ACS Synthetic Biology*, 7(8):1964–1972, 2018.

[6] Zoila Jurado, Ayush Pandey, and Richard M. Murray. A chemical reaction network model of PURE. *bioRxiv*, 2023.

[7] Dilara Grate and Charles Wilson. Inducible regulation of the *S. cerevisiae* cell cycle mediated by an RNA aptamer–ligand complex. *Bioorganic & Medicinal Chemistry*, 9(10):2565–2570, 2001.

[8] Elisa Franco, Eike Friedrichs, Jongmin Kim, Ralf Jungmann, Richard Murray, Erik Winfree, and Friedrich C. Simmel. Timing molecular motion and production with a synthetic transcriptional clock. *Proceedings of the National Academy of Sciences*, 108(40):E784–E793, 2011.

[9] Ryan Marshall and Vincent Noireaux. Quantitative modeling of transcription and translation of an all-*E. coli* cell-free system. *Scientific reports*, 9(1):11980, 2019.

[10] Manisha Kapasiawala and Richard M. Murray. Probing metabolism in an *E. coli*-based cell-free system reveals a trade-off between transcription and translation. *bioRxiv*, pages 2023–03, 2023.

[11] Dan Siegal-Gaskins, Zoltán A. Tuza, Jongmin Kim, Vincent Noireaux, and Richard M. Murray. Gene circuit performance characterization and resource usage in a cell-free "breadboard". *ACS Synthetic Biology*, 3(6):416–425, 2014.

[12] Richard Conrad and Andrew D. Ellington. Detecting immobilized protein kinase C isozymes with RNA aptamers. *Analytical Biochemistry*, 242(2):261–265, 1996.

[13] Daniela Proske, Martin Hofliger, Richard M. Soll, Annette G. Beck-Sickinger, and Michael Famulok. A Y2 receptor mimetic aptamer directed against neuropeptide Y. *Journal of Biological Chemistry*, 277(13):11416–11422, 2002.

[14] Jeremy R. Babendure, Stephen R. Adams, and Roger Y. Tsien. Aptamers switch on fluorescence of triphenylmethane dyes. *Journal of the American Chemical Society*, 125(48):14716–14717, 2003.

[15] Jason B. Da Costa, Aurelia I. Andreiev, and Dieckmann Thorsten. Thermodynamics and kinetics of adaptive binding in the malachite green RNA aptamer. *Biochemistry*, 52(38):6575–6583, 2013.

[16] James M. Carothers, Jonathan A. Goler, Yuvraaj Kapoor, Lesley Lara, and Jay D. Keasling. Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity. *Nucleic Acids Research*, 38(8):2736–2747, 2010.

[17] Jason B. Da Costa and Thorsten Dieckmann. Entropy and Mg^{2+} control ligand affinity and specificity in the malachite green binding RNA aptamer. *Molecular BioSystems*, 7(7):2156–2163, 2011.

[18] Steven M. Plakas, Daniel R. Doerge, and Sherri B. Turnipseed. Disposition and metabolism of malachite green and other therapeutic dyes in fish. *Xenobiotics in Fish*, page 149, 1999.

[19] Sara L. Stead, Helen Ashwin, Brian H. Johnston, Anne Dallas, Sergei A. Kazakov, Jonathan A. Tarbin, Matthew Sharman, Jack Kay, and Brendan J. Keely. An RNA-aptamer-based assay for the detection and analysis of malachite green and leucomalachite green residues in fish tissue. *Analytical Chemistry*, 82(7):2652–2660, 2010.

[20] Joshua E. Sokoloski, Sarah E. Dombrowski, and Philip C. Bevilacqua. Thermodynamics of ligand binding to a heterogeneous RNA population in the malachite green aptamer. *Biochemistry*, 51(1):565–572, 2012.

[21] Tao Pan, Da-Wei Liu, Sui-Zhou Ren, Jun Guo, and Guo-Ping Sun. Study on decolorization of triphenylmethane dyes by DTT. *Huan Jing ke Xue= Huanjing Kexue*, 33(3):866–870, 2012.

[22] Yubin Zhou, Hong Chi, Yuanyuan Wu, Robert S. Marks, and Terry W. J. Steele. Organic additives stabilize RNA aptamer binding of malachite green. *Talanta*, 160:172–182, 2016.

[23] New England BioLabs. *PURExpress®* In Vitro *Protein Synthesis*. New England BioLabs, Ipswich, MA, 2020.

[24] PUREfrex. *PUREfrex® 1.0 Cell-free Protein Synthesis Kit*. PUREfrex, Kashiwa, Chiba, Japan, 2024.

[25] Laura Grasemann, Barbora Lavickova, M. Carolina Elizondo-Cantú, and Sebastian J. Maerkl. OnePot PURE cell-free system. *JoVE*, (172):e62625, 2021.

[26] Yasuaki Kazuta, Tomoaki Matsuura, Norikazu Ichihashi, and Tetsuya Yomo. Synthesis of milligram quantities of proteins using a reconstituted *in vitro* protein synthesis system. *Journal of Bioscience and Bioengineering*, 118(5):554–557, 2014.

[27] Elise Burmeister Getz, Ming Xiao, Tania Chakrabarty, Roger Cooke, and Paul R. Selvin. A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Analytical Biochemistry*, 273(1):73–80, 1999.

[28] Joan Christine Han and Grace Yang Han. A procedure for quantitative determination of tris (2-carboxyethyl) phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Analytical Biochemistry*, 220(1):5–10, 1994.

[29] Joseph N. Zadeh, Conrad D. Steenberg, Justin S. Bois, Brian R. Wolfe, Marshall B. Pierce, Asif R. Khan, Robert M. Dirks, and Niles A. Pierce. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32(1):170–173, 2011.

[30] William Poole, Ayush Pandey, Andrey Shur, Zoltán A. Tuza, and Richard M. Murray. BioCRNpyler: Compiling chemical reaction networks from biomolecular parts in diverse contexts. *PLOS Computational Biology*, 18(4):e1009987, 2022.

[31] Ayush Pandey, Makena L. Rodriguez, William Poole, and Richard M. Murray. Characterization of integrase and excisionase activity in a cell-free protein expression system using a modeling and analysis pipeline. *ACS Synthetic Biology*, 12(2):511–523, 2023.

[32] Ayush Pandey, William Poole, Anandh Swaminathan, Victoria Hsiao, and Richard M. Murray. Fast and flexible simulation and parameter estimation for synthetic biology using bioscrape. *Journal of Open Source Software*, 8(83):5057, 2023.

[33] Michael Hucka, Andrew Finney, Herbert M. Sauro, Hamid Bolouri, John C. Doyle, Hiroaki Kitano, Adam P. Arkin, Benjamin J. Bornstein, Dennis Bray, Athel Cornish-Bowden, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[34] Domitilla Del Vecchio and Richard M. Murray. *Biomolecular Feedback Systems*. Princeton University Press Princeton, NJ, 2015.

[35] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: the MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306, 2013.

[36] Charles R. Harris, K. Jarrod Millman, Stéfan J. Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, et al. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020.

[37] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020.

[38] Wes McKinney et al. pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, 14(9):1–9, 2011.

[39] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[40] Bokeh Development Team. Bokeh: Python visualization library, 2023.

[41] Michael L. Waskom. Seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.

[42] Arbor Biosciences. *myTXTL T7 Expression Kit*. Michigan, United States of America, 2019.