Understanding the lifetime and rate of protein production in cell-free reactions while maximizing energy use

Thesis by

Ankita Roychoudhury

In Partial Fulfillment of the Requirements for

the degree of

Control & Dynamical Systems Minor

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2021

(Defended June 1, 2021)

ACKNOWLEDGEMENTS

Thank you to Prof. Richard Murray for guidance throughout the project and this wonderful opportunity; William Poole and Ayush Pandey for teaching me important modelling tools and supporting me with insightful discussion; Manisha Kapasiawala, Zoila Jurado, Miki Yun, Andrey Shur, Rory Williams, and David Garcia for help learning lab techniques and experiment suggestions; Prof. Justin Bois for teaching me how to think through various types of problems; and Liana Merk for emotional support.

ABSTRACT

Liposomes, or vesicles, offer promising applications in fields including biofuel synthesis, drug delivery, and toxin removal. Programmable liposomes can be used for optimal protein synthesis and to prototype genetic technologies. However, one of the major challenges is the short lifetime of protein production. Here, we add metabolites and molecules to cell-free reactions at different times to interrogate their importance. Through this testing, we find that ATP only slightly enhances protein synthesis, and ADP can help a reaction reach steady state faster. We also note that the excess of particular molecules, such as NAD and 3PGA, can halt protein production. With this data, we developed a more accurate chemical reaction network-based model for cell-free reactions. We also begin to study an unexplained discrepancy in protein production between bulk and vesicle dynamics. To quantify protein synthesis, we use *E. coli* extract and energy buffer, often called cell-free transcription and translation (TXTL), with a chosen DNA template both within vesicles (encapsulated) and without (bulk). We have also been able to uncover fundamental properties of transcription/translation systems. We supplement this data with computational models utilizing chemical reaction networks. We established a vesicle setup with membrane pores and supplemental energy buffer on the outside which increased the efficiency of protein synthesis. By using chemical reaction network models, we have highlighted differences and similarities between models and experiments. With this setup, vesicles can be used for more complicated applications, such as drug delivery or genetic construct testing.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND/OR TABLES

Number

NOMENCLATURE

TXTL. A mixture of E. Coli extract and energy buffer that can produce protein given a DNA template.

Bulk. TXTL mixtures in a plate reader, with no compartmentalization.

Encapsulated. Mixtures placed inside a lipid bilayer, forming a vesicle.

Vesicle, Liposome, Synthetic Cell. A lipid bilayer with a custom solution inside and outside.

Alpha-hemolysin. A passive membrane protein with a 3 kDA pore.

GFP. Green Fluorescent Protein. A protein that omits green wavelengths.

C h a p t e r 1

INTRODUCTION

Synthetic Cells

Bottom-up construction of synthetic cells holds the unique promise to allow for the design of controlled environments for uses varying from biofuel synthesis to targeted drug or DNA delivery. When building synthetic cells, there are various subsystems that should be considered. These include, but are not necessarily limited to, spatial organization, metabolic subsystems, sensing and signaling, regulation and computation, and actuation.[2] There is a growing interest in the development and application of these genetically-programmed synthetic cells, which do not replicate or divide. To uncover how the components of a cell can work together, it is important to study and gain a comprehensive understanding of their building blocks.[1] These synthetic cells are developed by creating liposomes in the lab, within which we can place a specific set of reagents designed by the scientist. These can be used for biophysical studies. For example, Litschel et al. were able to encapsulate patterning proteins into vesicles to study different patterns formed over time.[1] Actin monomers have also been placed within liposomes such that their polymerization and shape formation can be studied.[1]

By developing a programmable vesicle chassis that we understand completely, it will be easier to repurpose biochemical circuitry to do a variety of diverse tasks.[2] It is then possible to build machines that can self-destruct after performing their function. Constructing synthetic cells allows us to program their destruction and to control whether the organisms enter our biosphere.[2] Synthetic cells can be customized to contain the bare-minimum components. They can thus be used as environments in which more complex engineered systems can be implemented with less off-target interactions and undesired noise. In this process, there is ample opportunity to uncover what is truly known about cellular processes. By supplementing experiments with simulations, there is an additional level of confidence in our knowledge. By implementing reactions and solving them as chemical reaction networks in our models, it is easier to test hypotheses for underlying mechanisms, such as enzymatic mechanisms or transcription.

To form the vesicles, an oil-emulsion technique with phosphatidylcholine (POPC) lipids can be used.[1,4] For small scale transcription and translation, TXTL is used. TXTL is an extract composed of parts and proteins from *Escherichia coli* that are necessary for transcription and translation. A schematic is shown in Figure 1. Given a deGFP DNA template with TXTL within a vesicle, we are able to visualize protein production. A typical cell-free protein synthesis reaction contains a DNA template, *E. coli* extract, and an energy buffer solution with amino acids, NTP's, and energy regeneration molecules.[1] Moving forward, we refer to the energy buffer and *E. coli* extract mixture as TXTL.

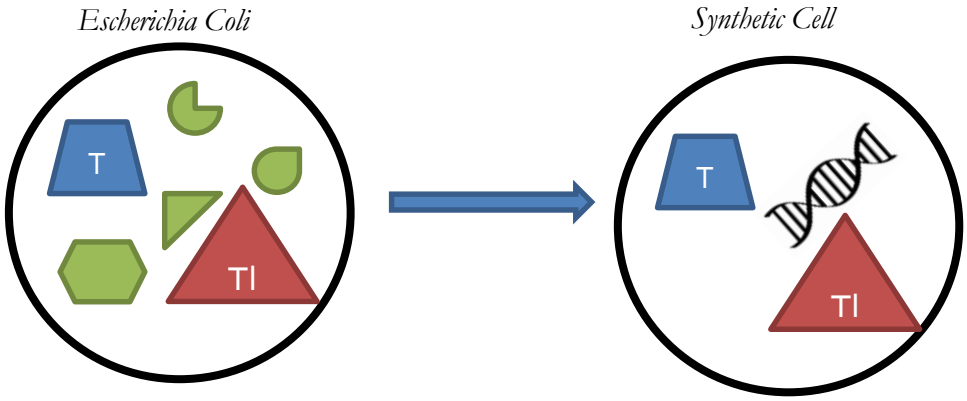


Figure 1. A depiction of a synthetic cell. The transcriptional (orange) and translational (blue) machinery from *E. coli* is extracted and placed into a liposome with the desired DNA template. The undesired machinery, including genomic DNA and debris, are represented by the green shapes.

Limitations

Current synthetic cell technologies can be used in both batch (bulk) and within vesicles (encapsulated). They are used for optimal biofuel production, since controlling the molecules are in specific compartments can result in higher efficiency and yield by reducing off-target interactions.[5]

However, energy for transcription and translation is limited. There is opportunity to address this issue and find new ways to optimize energy use. Various metabolic pathways have been shown to replenish energy sources for a longer time period, although it is difficult to investigate and quantify the toxic side effects these metabolites could have on protein production.[6] For example, Opgenworth et al. developed a molecular rheostat with isobutanol and glucose that was able to regenerate ATP.[7] Systems like these have promise, although there are around 15 enzymes necessary for this pathway, introducing opportunity for unwanted enzymatic reactions.

There are also studies that integrate proteins observed naturally, such as ATP synthase and proton pumps, that are capable of regenerating ATP with minimal side effects.[8] However, ATP synthase is a large, complex membrane protein. Spontaneously integrating this into a vesicle from purified protein or genetic code *in vitro* is a unique challenge on its own.

For our study, we will investigate encapsulated protein synthesis with TXTL since currently the most common mixture used for cell-free reactions. By perturbing this system, we hope to get protein synthesis to last longer than 10-12 hours, its current lifetime.[1] There is a need to find a complementary mechanism for energy lifetime extension that is simple and mitigates unnecessary crowding or toxic effects.

Further, different dynamics are often observed in bulk and encapsulated experiments. For example, GFP signal is significantly weaker in vesicles compared to in bulk when given the same percentage of molecules.[1] It remains quantitatively unexplained how and why protein production rates and steady states differ.[1] Investigating these differences can help develop a more universal platform for cell-free systems to significantly reduce variability.

Research Approach

In this study, we chose to study what components affect protein production in bulk and vesicle dynamics. This investigation involves the metabolic subsystem and spatial organization of liposomes.[2] More specifically, we studied how to maximize protein production and to uncover the differences between bulk and encapsulated dynamics.[2,9]

We are able to extend the lifetime of processes within cells derived from liposomes by exploiting compartments and membrane channels. To do so, we use ATP, or adenosine triphosphate, which is an organic compound that acts as the energy source for many different processes in cells. It has been repeatedly shown that ATP is the limiting factor in bulk and encapsulated protein synthesis, which we also investigated.[1] We also take advantage of compartments to optimize energy use within a vesicle by offering opportunity for more energy to enter and for molecules that may cause toxic side effects to leave the vesicle.

Previous work has shown that vesicles with passive membrane channels (α-hemolysin) surrounded by energy buffer have extended the steady state values of deGFP production.[10] When integrated into the membrane, it allows components from energy buffer, such as NTPs, to enter the cell to extend protein production.[10] Further, these pores also give toxic molecules a chance to escape the vesicle. For example, ATP use can result in higher free phosphate concentration, a small molecule that can have toxic effects due to off-target phosphorylation. In a vesicle with pores, there is more opportunity for free phosphate to leave the cell, as indicated in Figure 2. However, adding additional energy buffer in bulk experiments does not always guarantee greater protein production. By adding reagents, such as energy buffer or extract, in the beginning or middle (spike) of protein production in bulk and vesicle, we were be able to collect sufficient data to develop models in attempt to explain any observed differences.

We used TXTL in vesicles so we could control the creation of desired protein from DNA templates. We hypothesize that models and experiments will reveal the different dynamics in bulk vs. encapsulated reactions, and allow us to design ways to maximize energy use. An efficient, longer-lasting method to provide energy required for internal reactions will allow us to carry out more complex, sustainable experiments. We are able to broaden the range of possible research in synthetic cells if we can measure responses, production, etc. for longer time periods.[1]

With this study, we are better able to understand what molecules dictate the lifetime of protein production and how encapsulation affects cellular processes. By discerning what components are crucial for energy regeneration, we can understand how metabolism truly works in cells. We can understand if energy is the limiting factor for many existing experiments. Longer lifetimes will allow for more synthesis of bio-compatible materials, accurate environmental monitoring and remediation, self-assembly of complex multi-cellular machines, etc.[11]

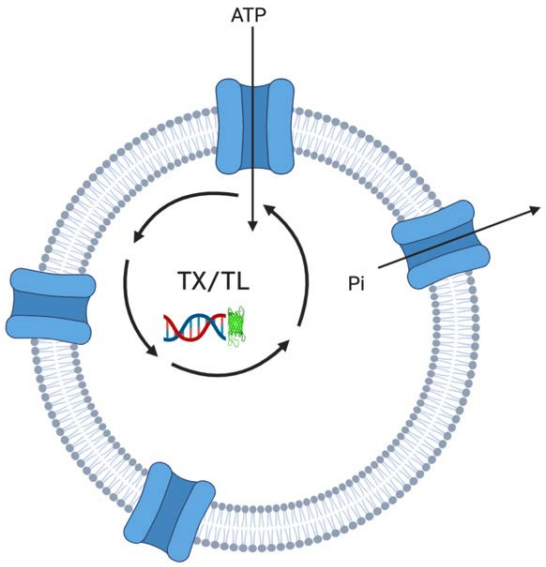


Figure 2. Schematic of α-hemolysin membrane pores on a vesicle with DNA and TXTL inside. Ideally, we would see some outward diffusion of toxic molecules, such as free phosphate (P_i), and inward diffusion of desired molecules, like ATP.

We model our data via chemical reaction networks (CRNs). Software, such as BioCRNpyler, allow us to choose enzymatic mechanisms, as well as rate constants and initial conditions.[12] We can model bulk experiments simply by including our assumptions about degradation, leak, what processes use ATP, etc.

Previously published studies have attempted to develop thorough models for cell-free reaction systems in MATLAB.[13] This pre-existing model, which we often refer to as the original model throughout this study, considers the binding of DNA and RNA with RNA polymerases, ribosomes, ribonucleases, and transcription factors.[13] More detailed reactions, species, and parameters can be found in the Appendix. The model includes transcription, translation, RNA degradation, ATP regeneration, and DNA degradation. Our experiments aim to build on this model.

Results Summary

In this thesis, we have quantified how different reagents may affect protein production in bulk and encapsulated settings. We note that there is ATP leak without DNA in TXTL mixtures, and that adding ATP only slightly increase protein production. We are not working in the regime of DNA saturation, so adding DNA can still cause greater levels of protein at steady state. Adding too much fuel sequesters all the energy and neglects translation. Spiking in reagents also does not function to revive protein production.

In an encapsulated setting, adding alpha-hemolysin and energy buffer on the outside is able to give more protein production. We also find that GFP expression in vesicles is lower than in bulk experiments, adding DNA increases steady state protein production, and adding ADP causes encapsulated cell-free reactions to reach steady state faster.

Chapter 2

RESULTS

There is ATP Use without DNA in TXTL Mixtures

In a pre-established MATLAB model, it was believed that there would be minimal ATP use in a mixture with buffer and extract without DNA.[13] This is because DNA is an important reagent that utilizes the transcription and translational machinery available in extract, thereby causing ATP to be used.

We chose to test this assumption by setting up an experiment to test GFP production and ATP use with different combinations of buffer, extract, and DNA. From this data, we can understand how ATP is depleted without DNA and we can see if the lack of ATP and regeneration pathways can allow for some protein production. To make sure these setups were correct, we tested GFP fluorescence, shown in Figure 3. We note that only the positive control well has fluorescence, as expected. The other wells do not have all the necessary components to guarantee protein production. This also indicates that energy buffer is critical for protein production.

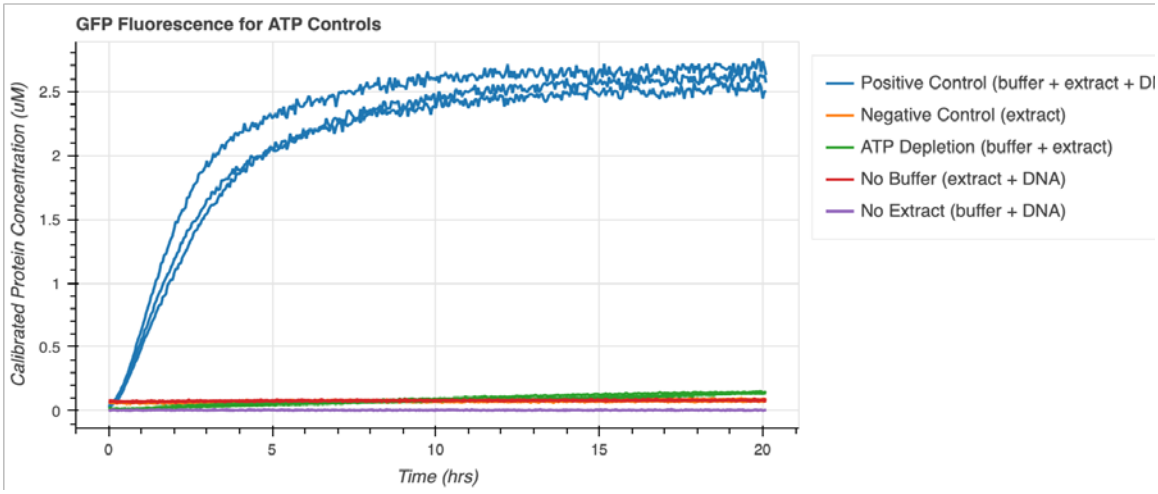


Figure 3. GFP Fluorescence Data with deGFP DNA. Only the positive control well reaches high levels of protein production, confirming all components are critical.

After confirming the GFP controls, we took ATP time course data with a firefly luminescence assay for the same samples that would allow us to understand the inner energy dynamics of a TXTL mixture (Fig. 4). A Hamilton Automated Liquid Handling Robot was used to minimize pipetting errors. The reaction with no extract starts off with the highest amount of ATP because it has the highest ratio of ATP from the buffer. The positive control and ATP depletion wells have slightly different starting ATP concentrations, likely due to noise in sample collection or presence of some extra ATP in the extract. We do see a fairly significant decrease of ATP over time in this sample, likely due to natural degradation within the buffer and off-target reactions with the DNA. The ATP depletion and the positive control

wells both have downward trends around the same magnitude. This counterintuitively indicates that DNA does not have a large impact on the ATP levels. Given DNA, we would expect there to be higher metabolic activity, resulting in more ATP use. This may indicate that there are many off-target ATP use pathways with only buffer and extract. Mitigating these could allow for more efficient protein production.

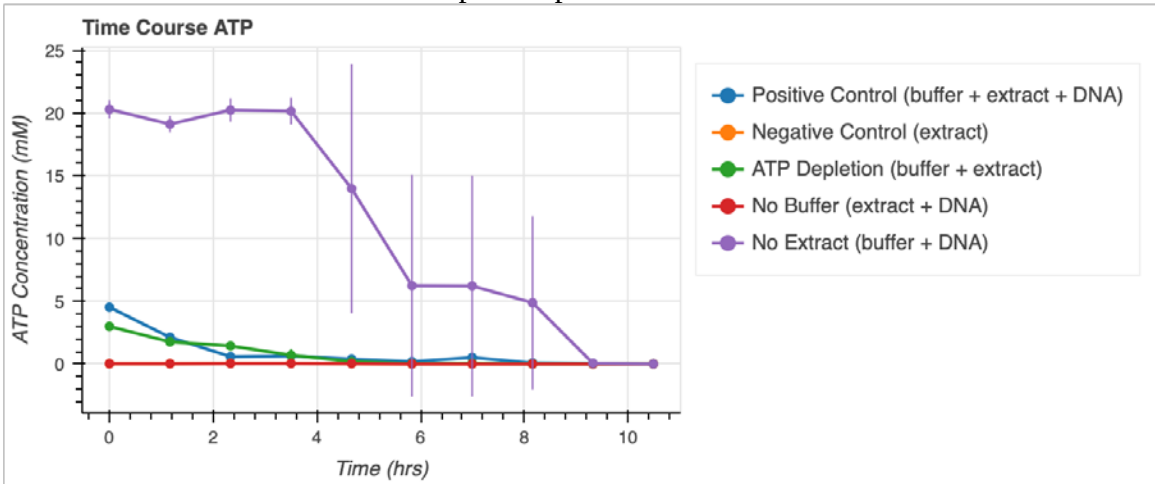


Figure 4. ATP Controls Data. In this setup, we see that all curves have a downward trend. Notably, the buffer + extract curve consumes ATP by hour 6.

These data confirm that there are likely off-target pathways or processes that use up energy. Next, we adjusted an existing model to accommodate for this phenomenon. The model shown in Figure 5 is a mixture of extract and buffer without DNA, mimicking the ATP depletion well in the experiments. We experimentally observed that ATP is used up by hour six, so we make an adjustment in the original model. The preexisting model had NTP use by hour 15, however, we have adjusted the rate of NTP degradation (from 1.77e-5 to 4.77e-4 /s) to account for our experiments. Note that we assume most NTP's are ATP's in the model.

Model Adjustment based off GFP control

Based off the GFP control shown in Figure 3, we next aimed to develop a more accurate model in BioCRNpyler. We noted that by tuning the initial concentrations of RNAases's (0 to 4.5e-3 mM), RNA polymerases (7.73e-4 to 1.33e-3 mM) , and ribosomes (2.73e-2 to 7.73e-4 mM). we were able to get the correct steady state value of protein. Then, by lowering the rate of binding in transcriptional processes (4.48 to 0.48 /s), the time to steady state approach in the simulation also closely matched the experimental data. Figure 5c,d show the data with adjusted parameters. We call this the original model moving forward for simplicity, since we will be performing initial condition perturbations on it and would like to note the differences.

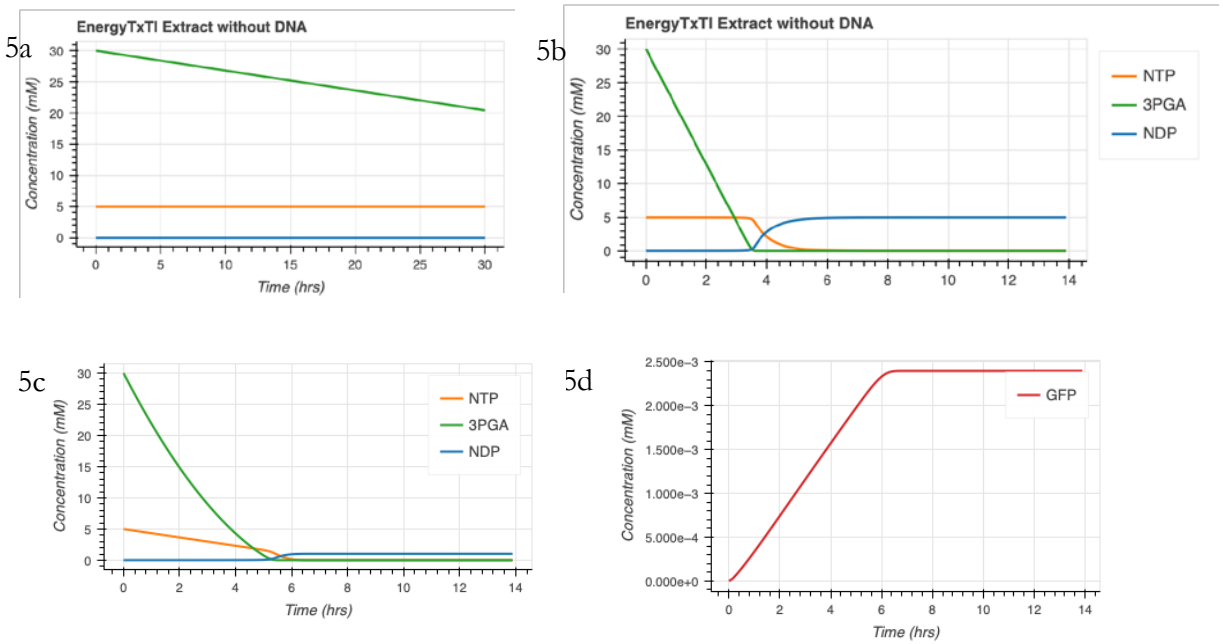


Figure 5. Simulation of an extract + buffer mixture, with no DNA. 5a) The simulations for the original model (1.77e-5 /s). 5b) The simulations for the adjusted models with new parameter values (rate of NTP degradation = 4.77e-4 /s). In this model, we see ATP use by hour 5, a closer match to experimental data. 5c) Simulations with adjusted parameters, called the original model moving forward. Initial conditions of RNAase's, RNA polymerases, and ribosomes are adjusted. Rate of transcriptional binding processes is lowered. 5d) The GFP steady state much closer matches the experimental data.

Adding ATP only slightly increases protein production

Next, we studied whether simply adding ATP to the reaction setup would affect the steady state level of protein production. To our cell-free reaction setup, which has extract, buffer, and deGFP DNA, we added 5 mM ATP, 10 mM ATP, and 25 mM ATP (Fig. 6a). After studying the deGFP endpoint values and the slopes at the beginning (Fig. 6b), we concluded that adding ATP in the beginning results in slightly more protein production. The endpoint values and the slopes are both slightly greater than the positive control. This indicates that adding ATP in the beginning may minorly affect the initial rate of protein production. Note that there is a large amount of variability in these samples, so it can be difficult to discern whether there is a substantial difference, or if we are observing the effects of noise.

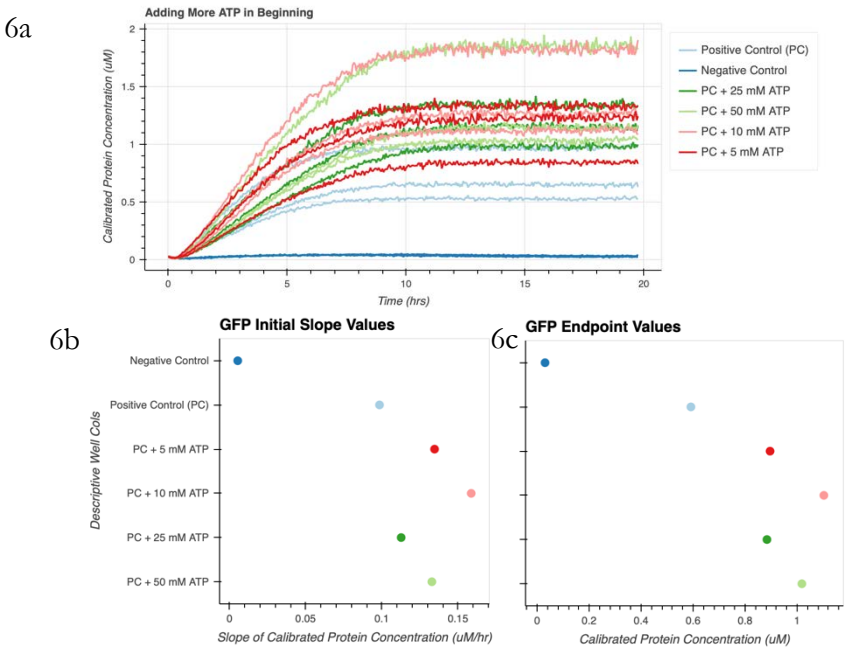


Figure 6. Adding ATP in the beginning. 6a) GFP time course data. n=3 for each sample well. 6b) GFP values of the initial slope (from hour 1 to hour 4) and 6c) endpoint values.

Next, we take a look at the simulation for this experiment (Fig. 7). We did this by changing the initial concentration from 5 mM ATP to 5.5 mM. We see that the system with more ATP actually has slightly less GFP at the end point (Fig. 7b). This is likely because 3PGA is used quicker, resulting in less time for protein production.

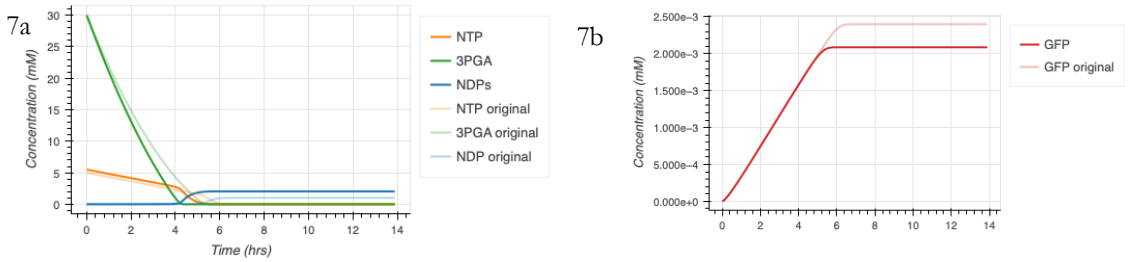


Figure 7. Simulations with more NTP (5 to 5.5 mM) in the beginning. 7a) NTP is used at a similar rate and 7b) results in slightly lower GFP production.

One of the reasons adding ATP may not result in substantially increased protein production is accumulation of toxic free phosphate, which is not included in the model. We hypothesized that there is more ATP use when ATP is added, by other off-target processes. So, we took ATP endpoint data for the same setup. We see that the positive control well had the most leftover ATP. The 10 mM ATP had much less ATP. This could indicate that addition of a certain amount of ATP starts allowing other off-target processes to begin. We also studied the free phosphate endpoint values in attempt to see if adding ATP indicates there is more P_i, which could have toxic effects that resulted in no increase of steady state protein production (Fig. 9). We saw that up to 25 mM ATP, the P_i endpoint values were

comparable to that of the positive control. However, adding 50 mM ATP resulting in an excess of leftover P_i.

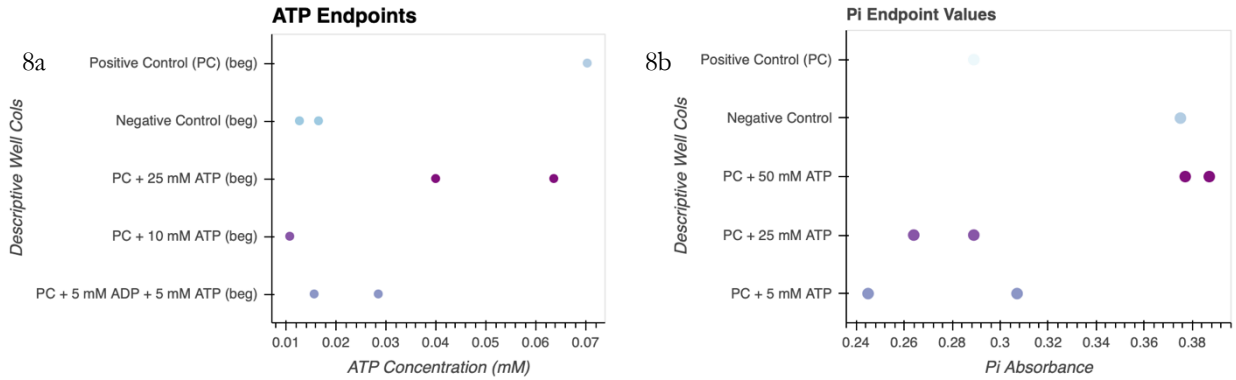
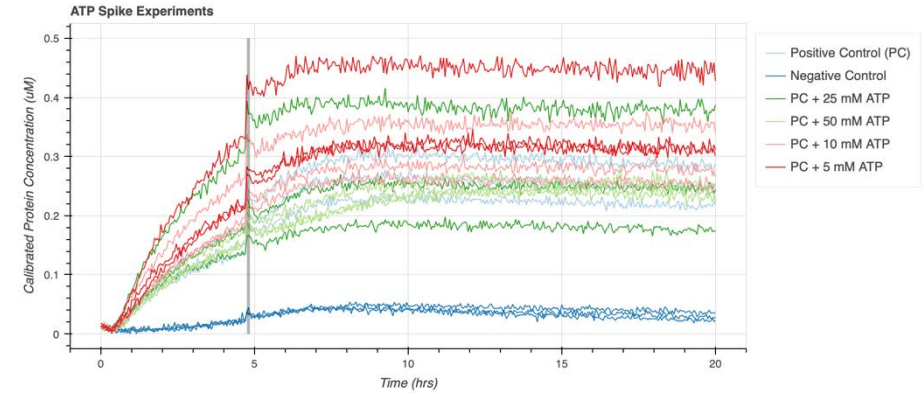


Figure 8. ATP and P_i End point for ATP beginning experiments. 8a) We see that there is more leftover ATP in the positive control. 8b) Free phosphate endpoint values are all below the positive control except for the 50 mM ATP well.

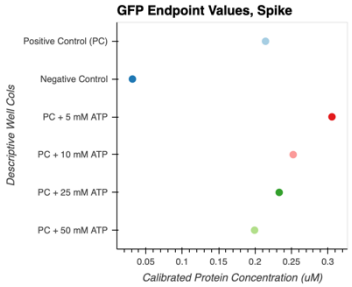
Spiking in ATP around hour five can increase steady state GFP production

Experimentally, spiking in ATP in the middle of protein production has little effect on the steady state deGFP levels. This may indicate there are auxiliary processes that consume ATP that we have failed to include in our model. From Figure 9, we see that adding 5 mM ATP slightly increases deGFP levels. However, there is not a large difference between the stady state values for the positive control and ATP spikes (Fig. 9b). This indicates that spiking ATP is not able to recover protein production. We also took ATP endpoint values for a similar spike experiment setup to study how much leftover ATP there was (Fig. 9c). Interestingly, we again saw that there were similar amounts of ATP leftover in the positive control well vs the ATP spike wells (except for one of the 25 mM samples). This aligns with our hypothesis that other non-protein production pathways may be consuming ATP when added in excess.

9a



9b



9c

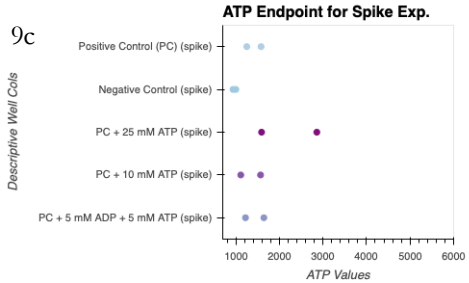


Figure 9. 9a) ATP time course, 9b) endpoint, and 9c) GFP endpoint values for a spike experiment. Adding ATP between hour four and five still results in similar amounts of steady state production values.

Adding too much NAD and 3-PGA completely kills protein production

Next, we tried to add more NAD and 3-PGA in the beginning of a TXTL reaction. We chose NAD (nicotinamide adenine dinucleotide) because it is an important cofactor involved in metabolism. 3-PGA is the fuel source in these reactions and functions to regenerate ATP. However, we note that adding these molecules in concentrations that are too high functions to completely kill protein production (Fig. 10). We see that the protein steady state values and initial rates when adding NAD and 3-PGA are close to the negative control sample. This could occur if NAD triggered other off-target pathways in the mixture. Further, adding too much 3-PGA may cause most of the energy to be devoted towards breaking down these fuel compounds as opposed to protein production. Perhaps adding more intermediate values of 3PGA and NAD would result in more optimal protein production.

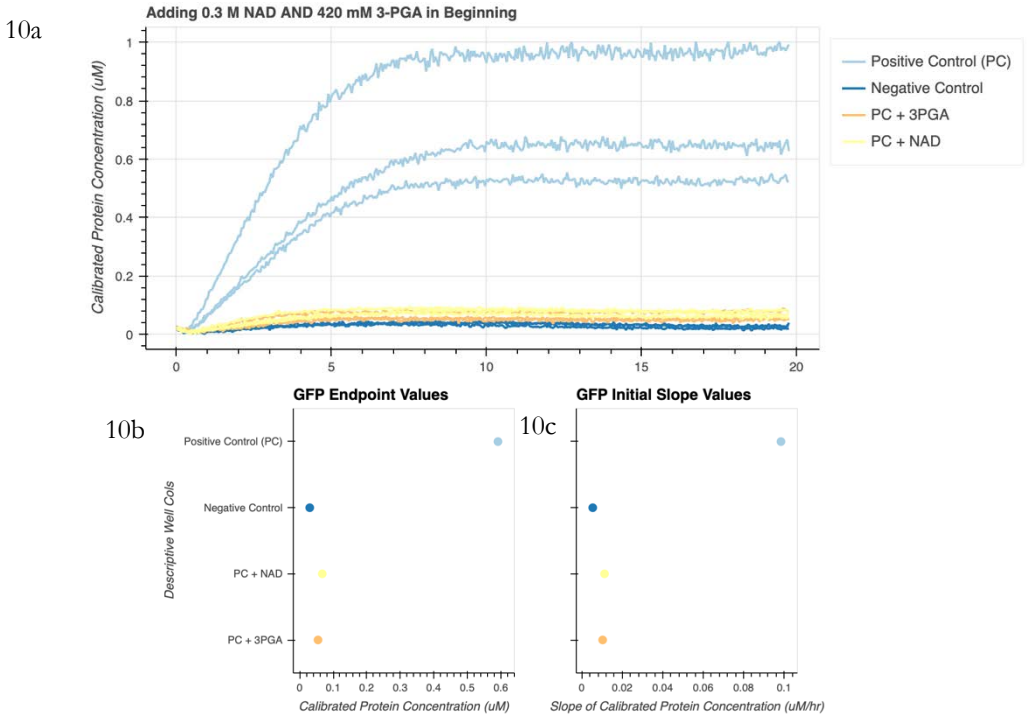


Figure 10. High initial concentrations of NAD and 3-PGA halt protein production. 10a) GFP Time-course curves. 10b) Average endpoint values. PC + NAD and PC + 3PGA are close to the negative control. 10c) Initial slopes (from hour one to four) of the samples. The sample wells are close to the negative control again, indicating protein production is minimal.

Now, we can take a look at the simulations for this system. Currently, NAD is not included in our models. Its direct involvement in cell-free reactions is unknown. However, 3PGA is modelled as an energy regeneration mechanism. However, when we add an excess of it, changing its initial condition from 30 to 100 mM, we see the increased protein production in the original model (Fig. 11b).

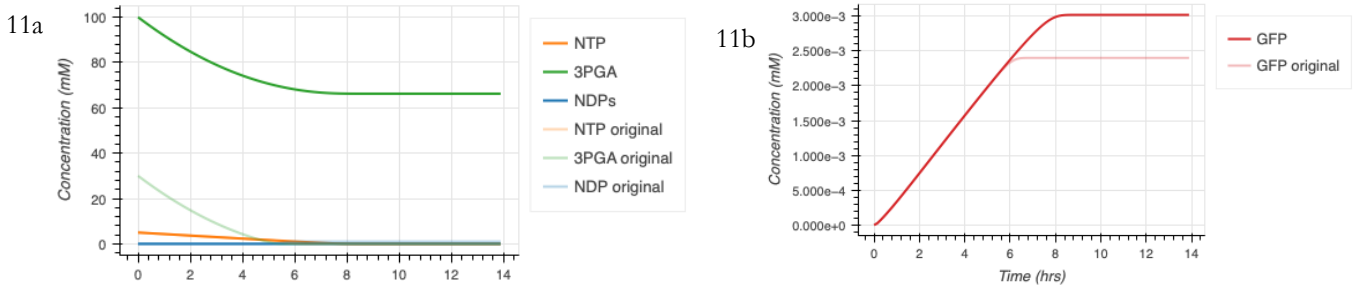


Figure 11. 3PGA simulations show increase in steady state protein level. 11a) Comparing the original and adjusted models shows that there is a lot of leftover 3PGA and 11b) higher GFP steady state levels.

Thus, it was necessary to adjust our models to account for the experimental data. Our hypothesis is that presence of too much 3PGA results in off-target pathways being activated. To model this, a reaction was added that broke down 3PGA into another molecule (2PGA) with the use of energy at a rate of 5e-6 /s. We note that when initial conditions are normal, around 30 mM for 3PGA, the GFP protein production is similar to the original model (Fig. 12b). However, when 3PGA initial condition is changed to 100 mM, protein production is severely inhibited, as desired (12d).

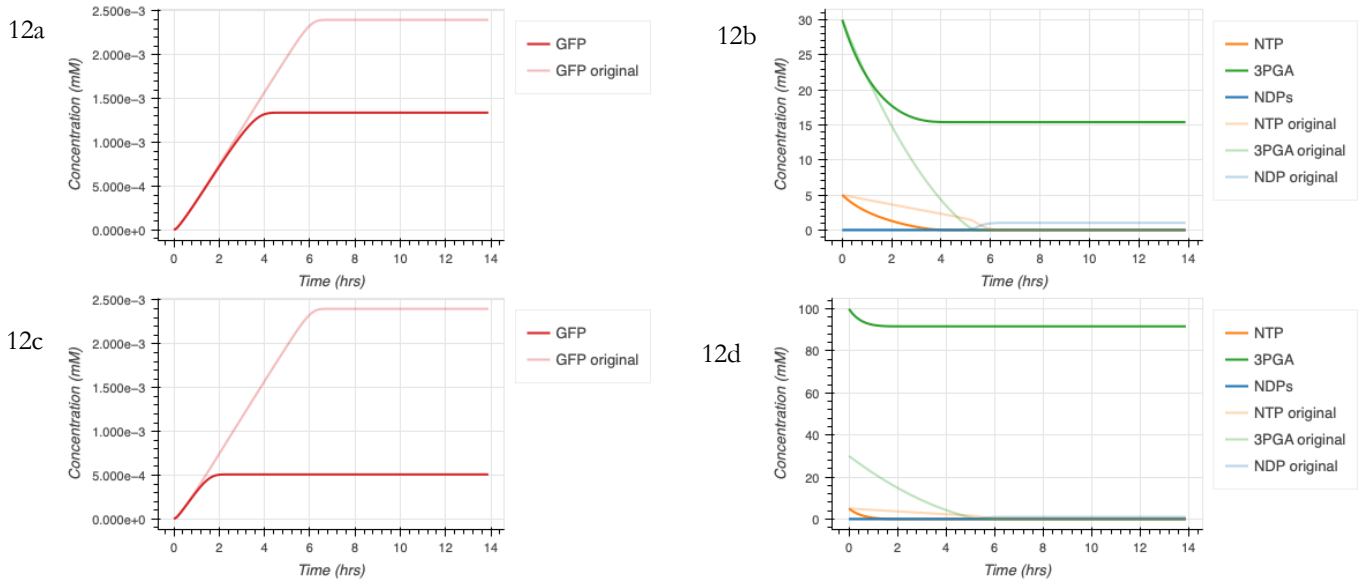


Figure 12. Adjusted model with 3PGA off-target breakdown better captures the dynamics of the experimental data. 12a) A reaction of 3PGA + NTP → 2PGA with a rate of 5e-6 /s was added to the original model. The 3-PGA is used a bit quicker. 3PGA's initial condition remains at 30 mM. 12b) GFP levels are similar when the extra 3PGA breakdown step is added. 12c) The simulation when 3PGA initial condition is set to 100 mM for the model with the added 3PGA breakdown step is limited by the concentration of NTP. 12d) We see significantly lower GFP steady state values when the 3PGA initial condition is increased.

By adding a 3PGA breakdown reaction, we are better able to capture the experimental dynamics. There are likely off-target reactions of 3-PGA and all the energy is dedicated to its breakdown, resulting in substantially lower steady state protein production.

Next, we see that the data from the spike experiments for NAD and 3PGA support our hypotheses. Spiking in high concentrations of NAD and 3PGA after most of the protein production has occurred does not greatly affect steady state GFP levels (Fig. 13). This may indicate that the ATP was used towards transcription and translation in the beginning, and so adding potentially harmful molecules later during this process does not have as large of an effect because the desired pathways are already triggered. The steady state GFP levels of NAD and 3PGA are a bit lower and higher than the positive control respectively. This is likely due to random noise in the different wells. In conclusion, only high initial concentrations of NAD and 3PGA are detrimental to protein production.

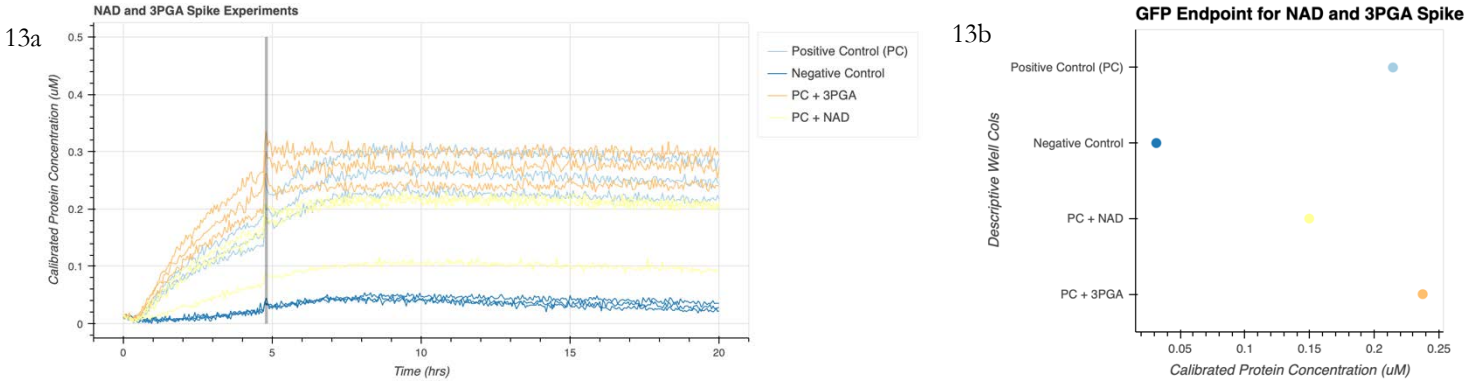


Figure 13. NAD and 3PGA spike experiments do not halt protein production. 13a) GFP time course of the different samples shows NAD and 3PGA do not greatly affect steady state GFP levels. 13b) Endpoint values of NAD and 3PGA are near the positive control.

Extract very slightly increases protein production in bulk TXTL reactions

Next, we tried adding a higher initial concentration of Rosetta extract to bulk TXTL reactions. We chose the Rosetta strain because it is optimized for protein production purposes. Extract contains RNA polymerases and ribosomes, which are the main factors that would affect protein production. Experimentally, we observe a very minor increase in protein production (Fig. 14).

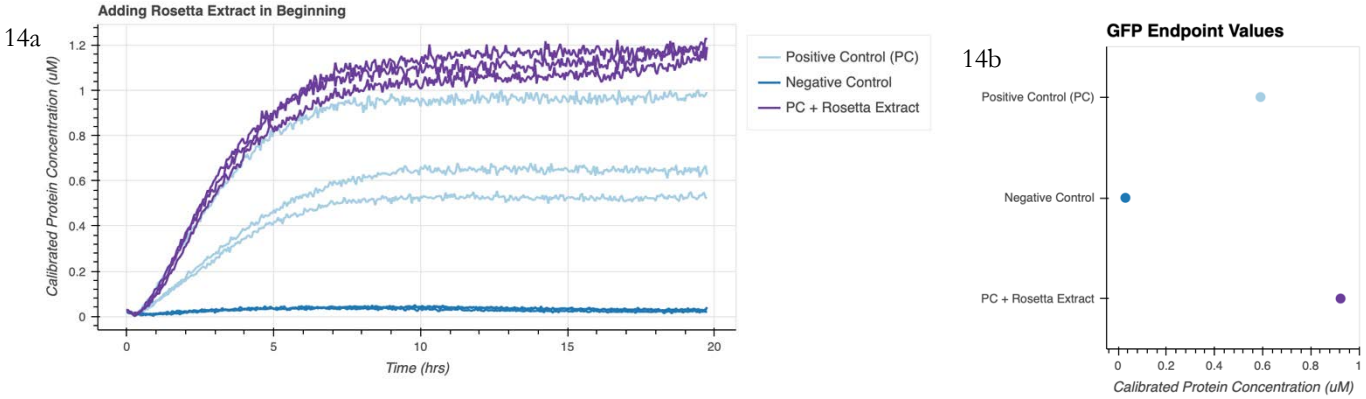
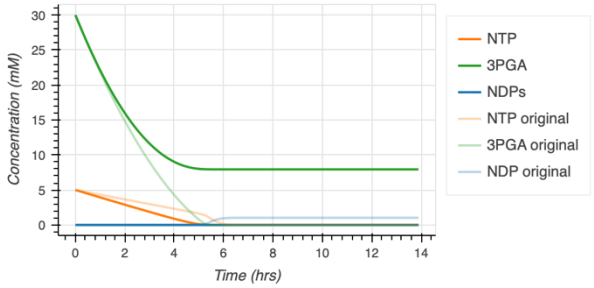


Figure 14. Adding more extract in the beginning slightly increases protein production. 14a) GFP time course trace shows a minor increase in the PC + Rosetta Extract well. 14b) Average endpoint values confirms the small increase in the sample with additional Rosetta Extract.

Now, we study the simulations for this setup. We increase the initial concentrations of RNAase's to 0.005mM, RNA polymerases to 0.0014 mM, and ribosomes to 0.00078 mM. We note that NTP and fuel sources are used quicker, and we see an increase in GFP production, as seen experimentally.

15a



15b

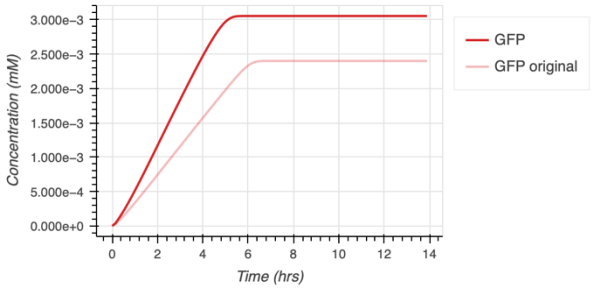
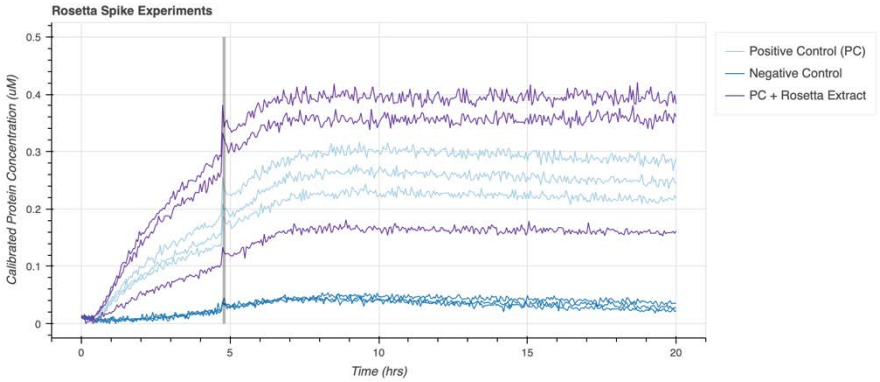


Figure 15. Simulations of increased RNA polymerase and ribosome initial conditions cause a slight increase in steady state protein production. 15a) Fuel's and NTP's are used a bit quicker in the adjusted model. 15b) A very slight increase in GFP production is seen with the new initial conditions.

Now, we can take a look at the dynamics when Rosetta Extract is spiked into the system between hour four and five. We see that adding Rosetta between hour four and five does not drastically change protein production. The slight increase viewed in the steady state protein values (Fig. 16b) is likely due to random noise intrinsic to these small TXTL reactions. Thus, we note that adding RNA polymerases and ribosomes has a very slight impact on protein production.

16a



16b

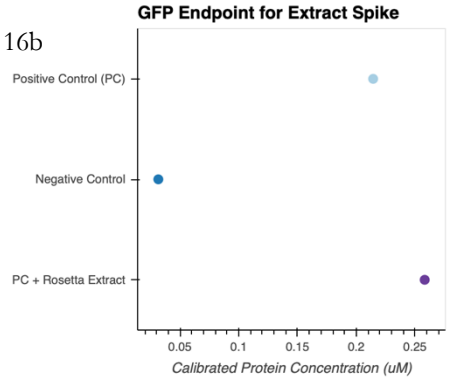
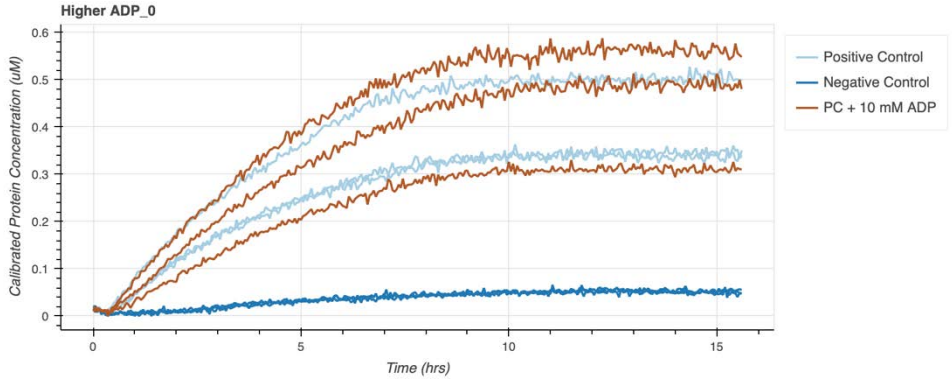


Figure 16. Rosetta spike experiments have slight effects on protein production. 16a) Adding extract (composed of RNA polymerases and ribosomes) has a minimal effect on GFP levels. 16b) The GFP endpoint level is slightly higher, likely due to noise.

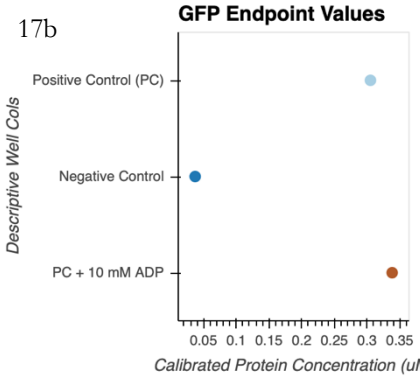
Adding Energy Buffer and ADP slightly affects protein production

Now, we added about 1.5 uL of 10 mM ADP and energy buffer separately to the TXTL mixtures. Interestingly, their addition slightly mitigated protein production (Fig. 17, 18). ADP's main role in protein production is reacting with 3PGA to regenerate ATP. However, ADP is also produced from ATP use, which occurs immediately, so adding extra ADP in the beginning functions to generate extra ATP in the beginning.

17a



17b



17c

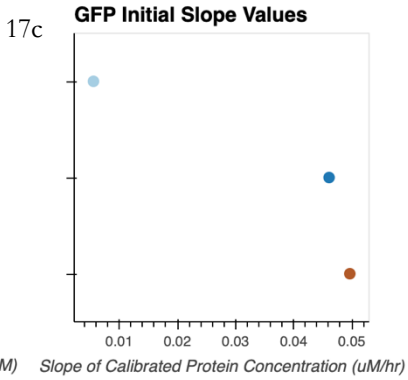
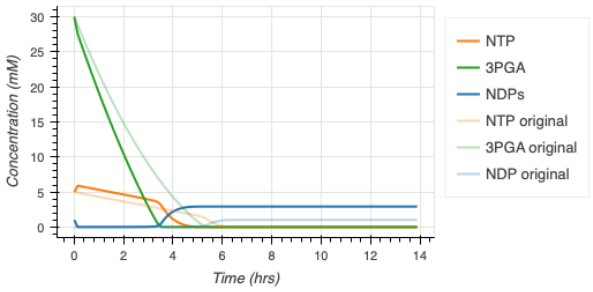


Figure 17. Higher initial concentration of ADP does not substantially affect protein steady state. 17a) GFP time course of controls and addition of 1.46 uL 10 mM ADP. Both 17b) the GFP steady state levels and 17c) initial rates are not largely affected by the addition of ADP.

If we take a look at the simulations when more ADP is added in the beginning (2 mM as opposed to 0 mM), we see a slight decrease in GFP production, similar to the additional ATP simulations. (Fig. 18b). The NDP supplied in the beginning is used immediately, as is some 3PGA, resulting in a small increase in the amount of NTP (Fig. 18a). This contributes to the quicker approach to steady state.

18a



18b

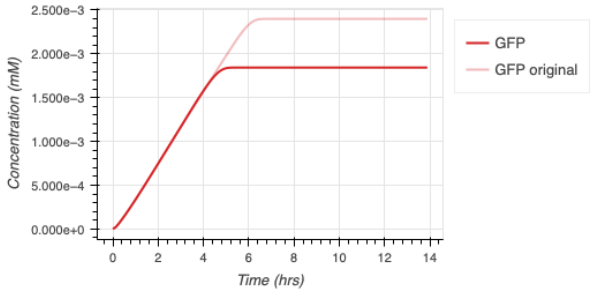


Figure 18. ADP simulations 18a) use NDP and 3PGA immediately, generating more NTP in the beginning. This results in 18b) a lower steady state value.

Then, we note that spiking in ADP between hour four and five doesn't largely affect protein production either (Fig. 19). This is likely because 3PGA is mostly used up or degraded by now, so it cannot to regenerate much ATP with the addition of ADP.

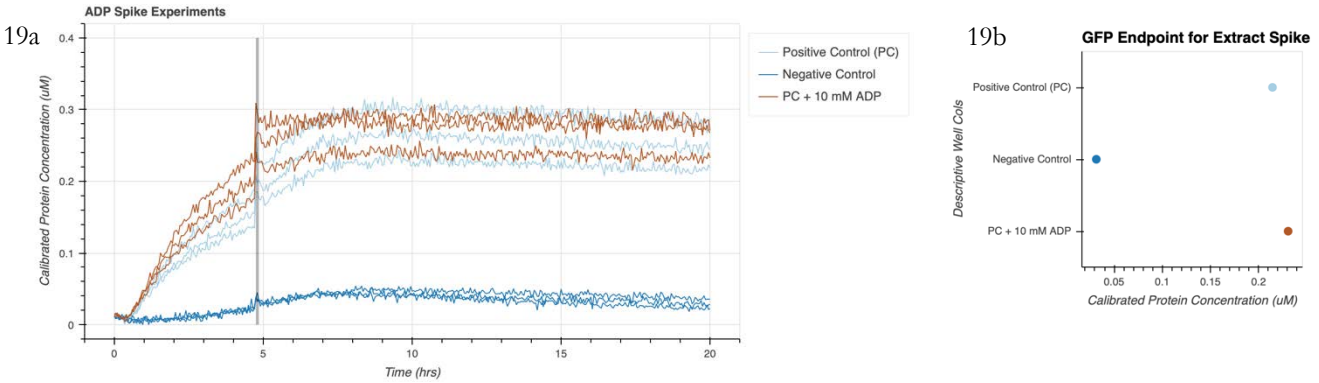


Figure 19. ADP spike experiment 22a) GFP time course and 19b) steady state values are very similar to the positive contro19

Then, we noted that adding additional energy buffer lessened protein production (Fig. 20). Both the steady state values and initial rates in the sample with energy buffer wells were lower than the positive control values. Energy buffer is composed of amino acids, salts, and an energy solution, which contains ATP and the fuel regeneration pathways. The reason for this phenomena is likely due to the addition of magnesium and potassium salts. These can negatively affect protein production when added in un-optimal concentrations and can also affect fluorescent readout. Because of the complexity of the components in energy buffer, it cannot easily be modeled.

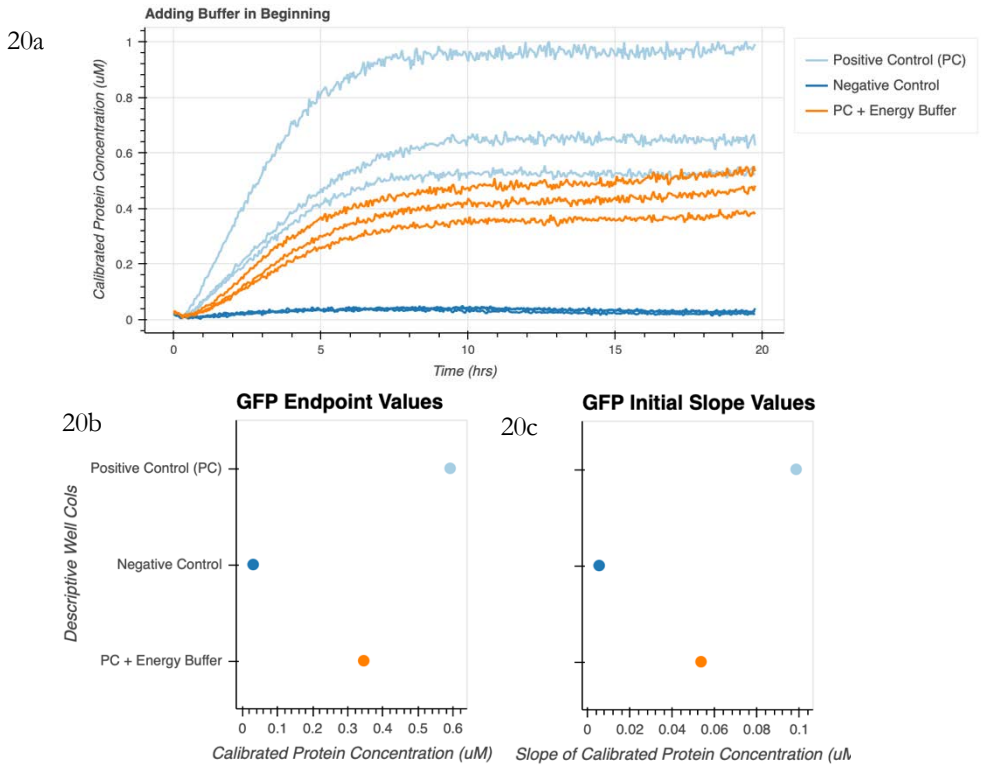


Figure 20. Experimental data when adding buffer in the beginning slightly mitigates protein production. 20a) GPF time course with controls and additional energy buffer. The 20b) steady state protein levels and 20c) initial slopes of the energy buffer well are lower than the positive control.

Next, we can take a look at the spike data for the energy buffer. Similarly, we see a minor decrease in protein production steady state levels (Fig. 21). This is likely due to the increased concentration of magnesium and potassium salt, which can introduce toxicity if added in levels too high, or even affect fluorescence readout.

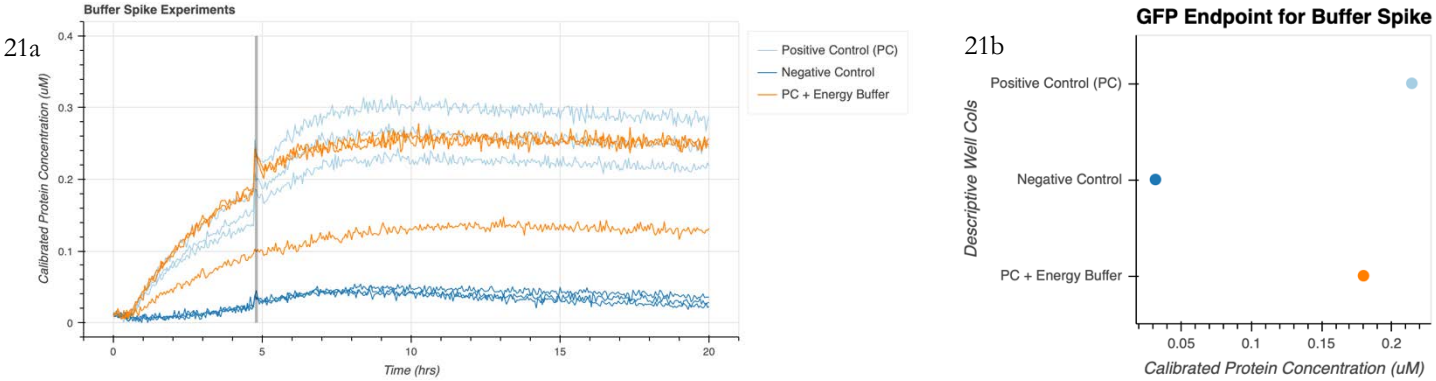


Figure 21. 21a) GFP traces and 21b) steady state protein values of the sample with additional energy buffer tends to be slightly lower than the positive control

Adding DNA increases protein production.

Next, we add in more DNA and see that it does increase steady state production levels (Fig. 22c). Thus, we note that our setup is not in the regime of DNA saturation. We can remedy this by creating more concentrated DNA solutions.

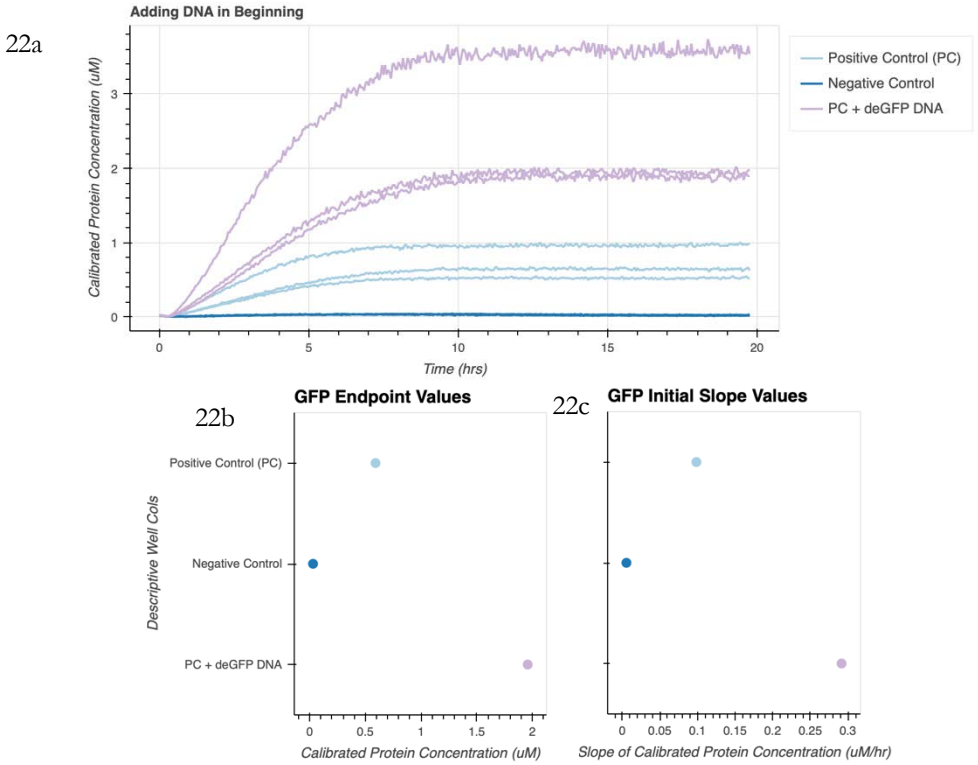


Figure 22. Adding more DNA in the beginning increases steady state production. 22a) Time course GFP data with more DNA and controls. 22b) Protein steady state values and 22c) initial slopes show that adding DNA increases GFP production and rate.

The model assumes we are the regime of DNA saturation. Thus, adding more DNA decreases steady state values a little (Fig. 23). However, it is interesting to note that the dynamics of the curve match more closely to the experimental data when we increase the DNA concentration (from 1 mM to 2 mM). NTP's are also used quicker. An almost saturated regime is likely a more accurate representation of the dynamics within an experiment.

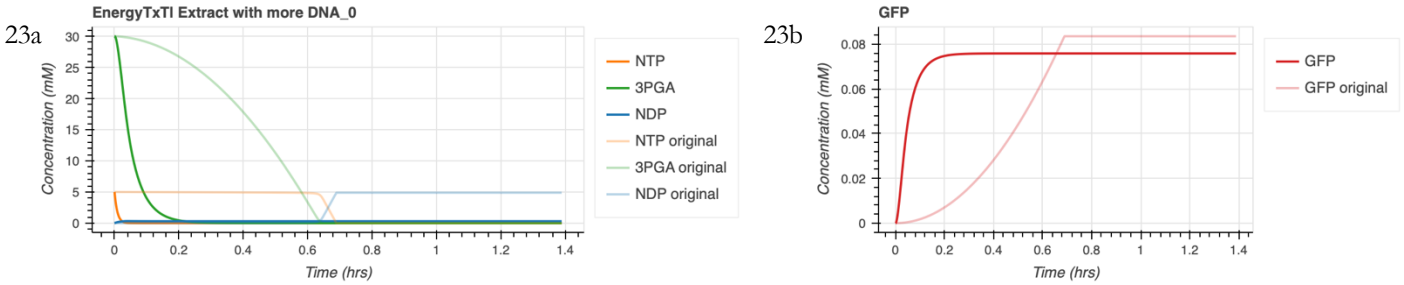


Figure 23. Simulations with higher initial concentration DNA. The original simulation has 1 mM initial DNA while the adjusted simulation has 2 mM. 23a) NTP and fuel are used quicker with more DNA. This results in a curve 23b) that more matches our experimental data.

Next, we study the experimental data when DNA is spiked in between hours four and five. Note that there is not substantially higher GFP production (Fig. 24). This may indicate that the TXTL mixture is not capable of taking and producing protein from a new DNA template any longer. Thus, adding DNA in the beginning increases protein production, however, adding it around hour five does not result in more steady state GFP.

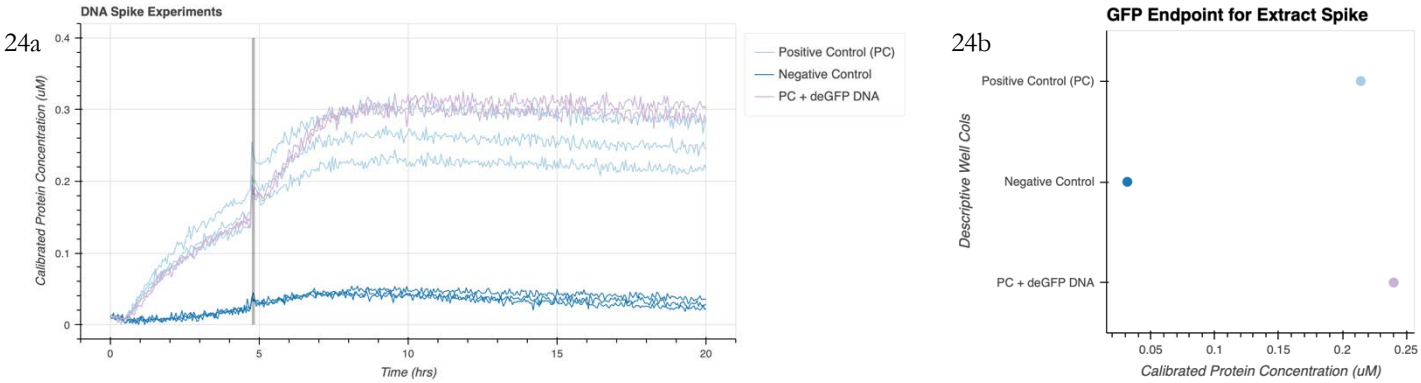


Figure 24. Data for DNA Spike Experiments shows that adding DNA in the middle does not have a significant effect on 24a) the GFP time course data or 24b) steady state protein production.

A TXTL mixture cannot take new protein template between two and four hours

Next, we wanted to add in DNA at different time points to study when TXTL mixtures die out. We spiked in DNA at hours two, four, and six. Interestingly, we saw that, when added at two hours, we are able to get ~50% of the optimal protein production (Fig. 25). However, when spiking in DNA at four hours, we get almost no additional protein production. This indicates the extract and buffer mixtures dies out between two and four hours and thus is no longer able to make protein. This is likely due to lack of fuel or energy. Without DNA, ATP is depleted quickly, by hour six. Adding DNA later on in the process may not leave enough energy for transcription and translation to continue.

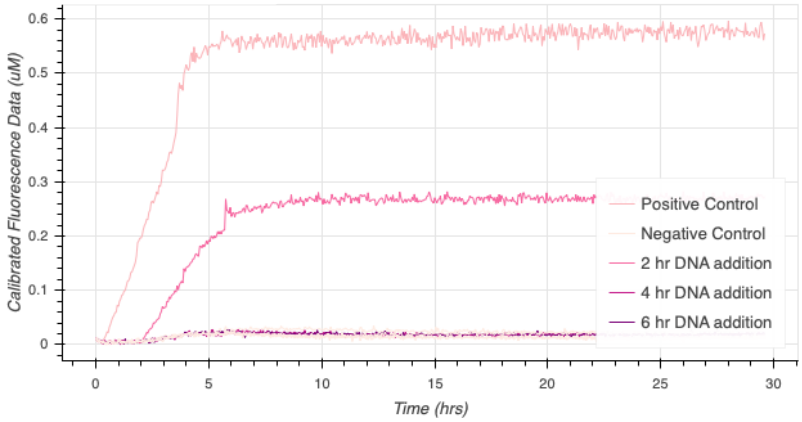


Figure 25. Adding in DNA at different times shows that extract cannot make protein from a new DNA template provided between two and four hours.

Vesicle positive control data has low GFP fluorescence

As a positive control for our vesicle experiments, we took time course data for a vesicle with a TXTL mixture inside and vesicle outer solution on the outside. The mCherry channel represents Rhod-PPE, which marks the lipid membrane. The GFP channel represents the fluorescence from the deGFP DNA. Vesicles were made via an oil-emulsion method. Inside the vesicles, we place extract, buffer, and deGFP DNA.

We observe faint GFP expression for the first two hours (Fig. 26). However, at around 12 hours, the GFP signal is not visible. This could be due to long exposure and photobleaching issues. However, we also note that the vesicle seems slightly deformed. This could be due to the difference in salt concentration inside and outside the vesicle or due to photobleaching. When time between images was increased, we did not see as significant of deformations as seen below, indicating that photobleaching may have been the issue.

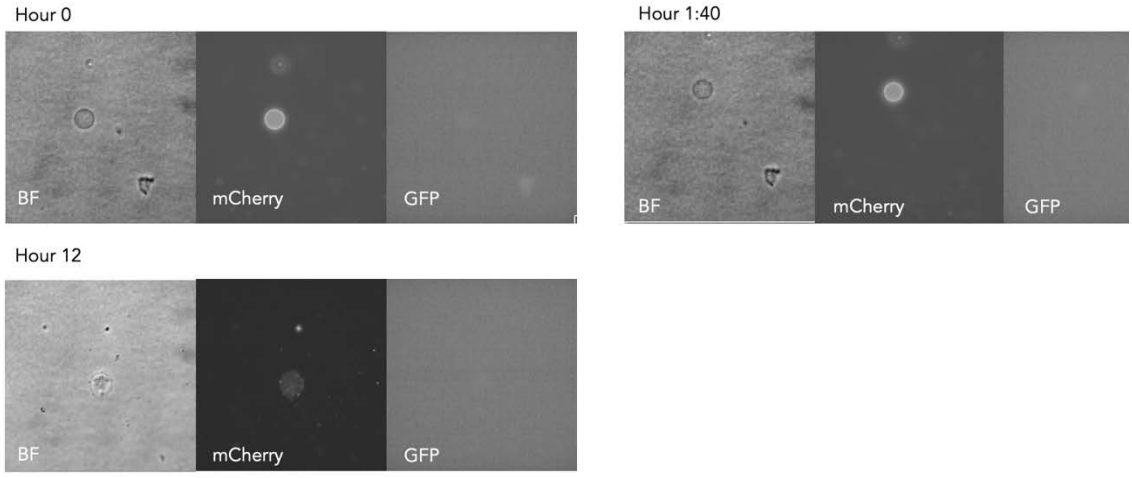


Figure 26. Time-course vesicle control data has faint values of GFP expression.

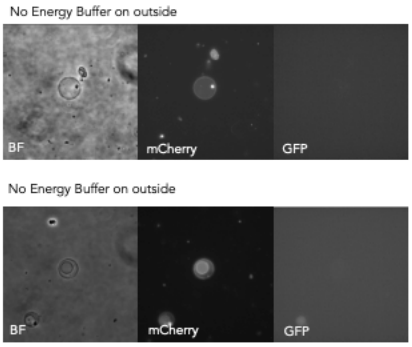
The same setup in bulk has strong GFP expression by hour six, a property not observed in vesicles. This discrepancy could be explained by uneven partitioning of molecules or because the reaction volume is significantly smaller within a vesicle.

When alpha-hemolysin is integrated into vesicles, additional energy buffer can be supplied on the outside to increase protein production.

Next, we made vesicles that had alpha-hemolysin, a small membrane protein with a 3 kDA pore. This would allow us to try and supply reactions within the vesicle with extra material from the outside in attempt to test if protein production could be elongated in an encapsulated environment. A TXTL reaction with deGFP was placed inside the vesicles. Energy buffer and vesicle outer solution (200 mM Glucose, 100 mM HEPES), were placed on the outside. About 1 uL of 100 uM alpha-hemolysin was added for a solution with 30 uL vesicles diluted in outer solution and 20 uL of either outer solution (negative control) or energy buffer.

Interestingly, we observed that the vesicles without energy buffer on the outside had low deGFP expression (Fig. 27a) while the vesicles with additional energy buffer on the outside had higher deGFP expression (Fig. 27b). This indicates that alpha hemolysin is integrating and likely allowing the addition of energy buffer inside vesicles, resulting in the higher steady state fluorescence of GFP.

27a)



27b)

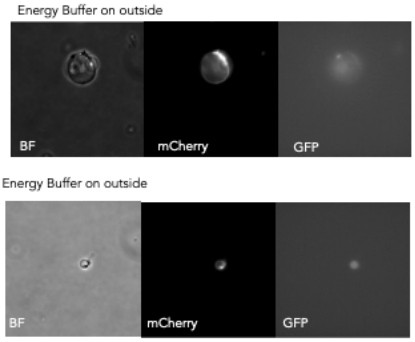


Figure 27. Testing the effect of adding energy buffer to the outside of vesicles with alpha-hemolysin. 27a) Without energy buffer, there is minimal deGFP levels after one day. 27b) With buffer, we see stronger deGFP signals, indicating that energy buffer is possibly reviving and replenishing important small molecules within the vesicle.

Thus, we have established a successful to perform spike experiments within vesicles as well. It takes alpha-hemolysin about two hours to integrate into the membrane, after which the reagent on the outside can slowly begin to seep in.

Adding a higher initial concentration of DNA in vesicles resulted in significantly more protein production.

Next, we added a higher concentration of DNA within vesicles (Fig. 28). We noted that the steady state GFP levels were significantly higher than the positive control. Similar to the bulk experiments, this indicates that we are not in the regime of DNA saturation within the positive control vesicles.

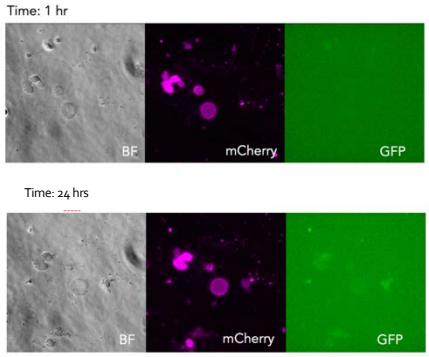


Figure 28. Vesicles with more DNA have an increase in GFP over time, more than the positive control.

Adding more ADP in vesicles results in a faster approach to GFP steady state

When we add higher initial volumes of 10 mM ADP within vesicles (Fig. 29), we get high GFP expression within the first hour. This is not observed in the positive control vesicles. This is likely due to the fact that ADP reacts with 3PGA to generate ATP. This increase in ATP production causes GFP production to reach steady state faster.

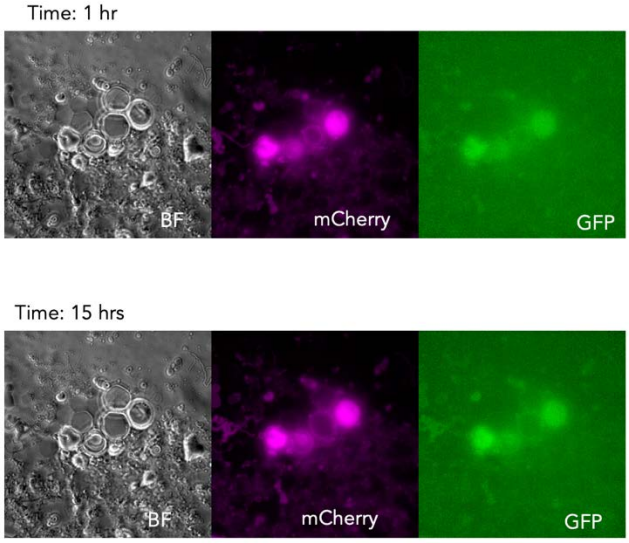


Figure 29. Vesicles with more ADP reach a higher level of GFP quicker than the positive control.

In an encapsulated setting, GFP expression was high by hour one. In bulk with additional ADP in the beginning, GFP expression reaches steady state around five hours (as opposed to six without) according to the model.

Chapter 3

DISCUSSION

In this project, we aimed to investigate the dynamics of bulk and encapsulated experiments. We performed experiments with higher concentrations of various reagents initially, and experiments where we spiked in concentrations of reagents in the middle of a reaction. We used chemical reaction network-based simulations to test the true depth of our pre-existing knowledge. By updating the models to match experiments, we learnt that we were missing integral degradation and leak reactions. We also noted how the model was successful in the face of various perturbations. The similar experimental setup outside and within vesicles allowed us to compare their nonlinear dynamics. This has remained an open question in the synthetic cell community, which we began to tackle[10].

Previously, researchers have shown that protein production can be lengthened in bulk. This has been done with various ATP regeneration mechanisms. For example, phospoenol pyruvate and pyruvate kinase can be used to regenerate ATP from ADP. However, it has been shown that collection of free phosphate functions to inhibit protein production due to random phosphorylation resulting in toxic effects[14]. We have been able to support these claims with our work showing that an increased level of ATP does not necessarily correspond to higher levels of protein production, but can correlate with higher levels of free phosphate. It has also been shown that ATP is often the limiting factor of cell-free reactions[14]. While our data does not support this, it does not refute it either. Adding ATP does not result in more protein, however this could occur because ATP is sequestered by off-target processes. We have shown that ATP is used over time in a mixture of buffer and extract, without DNA. This indicates there must be off-target processes which consume ATP. In order to properly test this hypothesis, it will be necessary for future researchers to attempt to identify these undesired pathways and cut off their energy supply. This can be done by inhibiting enzymes or by using a system such as PURE, that contains *only* the components integral for transcription and translation instead of general *E. coli* extract [15]. Although more expensive, tests comparing ATP use in TXTL vs PURE will reveal information regarding off-target pathways.

Our exploration with spike experiments in the TXTL setup has given new insights. We were able to probe our system to see if protein production could be regenerated. We observed different dynamics amongst the beginning and spike experiments for the same reagent. For example, additional extract in the beginning of a reaction had a substantial positive effect, whereas adding extract at hour four had almost no effect. Similarly, adding additional DNA at the beginning had a larger effect on steady state protein production than adding more DNA around hour five.

We also note that adding an excess of NAD and 3PGA halts protein production. NAD is proven to be integral in the regulation of transcription[16]. NAD can bind to proteins that function to change chromatin structures. A large perturbation in NAD levels is guaranteed to affect the efficiency of transcription [16]. This could be a large reason an excess of NAD causes

shut-down of protein production. Further, 3PGA is the source of fuel in these reactions. There is a reaction that uses energy to transfer a phosphate group from 3PGA into 2PGA.[16] In the presence of excess 3PGA, it is possible this process is preferred, thereby removing any available 3PGA. In this case, 3PGA would not be able to regenerate ATP, depleting the energy resource significantly. We were able to include a 3PGA breakdown reaction in our model that resulted in simulations which closely modeled the experimental data.

Further, we have generally worked in a regime close to DNA saturation. However, adding more DNA resulted in more protein production. It would be interesting to repeat these experiments with a more concentrated stock of deGFP DNA, to study if similar dynamics are seen in the saturation regime.

To continue, it remains unexplained why adding energy buffer resulted in a slight decrease of protein production. This could be due to suboptimal salt concentrations, since magnesium and potassium are part of the buffer. The addition of buffer also changes the pH, which can have negative effects on protein production. Buffer also contains 3PGA, which we have shown halts protein production in excess. Further studies into each of these possibilities could identify why the addition of energy buffer minorly inhibits protein production.

Additionally, we note that adding a DNA template between two and four hours did not result in any protein production. This is likely due to the lack of ATP. We notice that a mixture of buffer and extract completely runs out of ATP by hour six. However, between hours two and four, there is less than half the concentration of ATP than the initial time point. Thus, it is possible there is not enough ATP to trigger the transcription and translation of a new DNA template between two and four hours.

Next, we aimed to investigate the dynamics of experiments in vesicles. Experiments with alpha-hemolysin in vesicles with energy buffer have been done before.[10] We were similarly able to show that addition of energy buffer increases our steady state production of GFP within vesicles. With this setup, we can perform the spike experiments similar to those carried out in the bulk settings. The desired reagent, such as ATP, energy buffer, or 3PGA, can be placed outside the vesicle. Alpha-hemolysin is also added to the outside of the vesicle mixture. After alpha-hemolysin is integrated into the membrane, the desired reagent on the outside can slowly spike into the vesicle. This can help uncover the differences in nonlinear dynamics between bulk and encapsulated experiments. In the future, it will be important to model vesicle and α-hemolysin by including parameters which code for the concentration of α-hemolysin and how long it may take for transport to occur.

Vesicles with membrane pores give molecules with toxic effects a chance to leave the compartment, as well as allowing beneficial molecules to enter. However, the opposite effects can also be seen (healthy molecules leave, toxic molecules enter). In attempt to mitigate this, future research can be focused on directed transport of various molecules. We chose to use alpha-hemolysin, a passive membrane channel, so we could probe and learn from the system. Yet, in order to maximize the efficiency of encapsulated protein production, directed export of toxic molecules and directed import of beneficial ones will be valuable.

In a vesicle setting, we were able to show that the addition of ADP in the beginning of a reaction can trigger a faster approach to protein steady state. This can be valuable in situations where speed is desired. For example, if the timescale of a particular signaling

protein should reach steady state quicker than another one, perhaps for communication purposes, then additional ADP can be added in the beginning to ensure this control.

In addition, modelling TXTL reactions offer new insight into the dynamics.[12] However, previous models failed to consider the leak and degradation reactions of ATP. They also need to include possible off-target pathways, such as the breakdown of 3PGA. By including toxic effects, we have been able to develop models that more accurately represent our data.

With the information discussed, future research can be directed towards further optimizing and greatly affecting protein production in vesicles. Given the success of that research *in vitro*, the suite of possible experiments carried out in synthetic cell can be greatly expanded.

Chapter 4

MATERIALS AND METHODS

Cell-Free Reactions

The DNA used for all the experiments is pR(pLambda):UTR1:deGFP:T500 on ColE1 (pBEST/IA_v1-1) backbone. This template expresses deGFP on a constitutive promoter and is often used to characterize strength of a new extract batch.

The energy buffer is derived from TXTL Toolbox 2.0. A recipe to make 1 buffer is of 2.24 uL of 400 mM Mg-glutamate, 5.99 uL of 3000 mM K-glutamate, 22.33 uL of 6 mM amino acids, 6.41 uL of 14 mM energy solution, and 0.32 uL water. Salt calibrations were performed for Mg and K, see SI.

The energy solution is made up of 50 mM HEPES pH 8, 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg/mL tRNA, 0.26 mM coenzyme A, 0.33 mM NAD, 0.75 mM cAMP, 0.068 mM folinic acid, 1 mM spermidine, 30 mM 3-PGA, 1 mM DTT, 2% PEG8000, either 10−15 mM maltose or 20−40 mM maltodextrin[13]. We optimized Mg^{2+} and K^{+} concentrations for our extract.

Extract is derived from *E. coli* cytoplasmic extract. We create a Rosetta strain, which is optimized for protein production. The protocol follows the TXTL Toolbox 2.0[13]. *E. coli* glycerol stocks are grown in 2xYT with phosphate and chloramphenicol. After incubation and growth, pellets are weighted and vortexed. Then, the solution is centrifuged and the lysate is collected and aerated. Dialysis is performed at 4 ºC for 45 minutes and the final solutions are flash frozen with liquid nitrogen and stored at -80 ºC. This can be used for at least a year.

A cell-free reaction has deGFP DNA at a concentration of 1 nM, typically 1.17 uL of 8.99 nM stock. 4.375 uL of buffer, 3.5 uL of extract, and 1.46 uL of water to bring the total reaction volume to 10.5 uL. For the experiments where reagents were added in the beginning of spiked in, the volume of water (1.46 uL) was replaced with the particular reagent.

Encapsulation

We encapsulate TXTL and GFP DNA in vesicles to test the dynamics of protein synthesis in vesicles. We use chloroform, POPC lipids, and Rhodamine (Rhod-PPE) for encapsulation. POPC is a type of lipid used to make up the membrane and Rhodamine is a red fluorescent molecule that binds to lipid membranes, allowing us to confirm the presence of vesicles. We utilize an oil-emulsion method. We let a mixture of 100 uL 50 mg/mL cholesterol in chloroform, 200 uL 25 mg/mL POPC, and 10 uL 1 mg/mL Rhod-PPE sit in a 55 degC bead heater until the chloroform has evaporated. Then, 1 mL of mineral oil is added and left at 55 degC for 15 minutes, then left to cool on the bench for 5 minutes.

The following are then incubated on ice for each sample: 1 tube of 150 uL lipid/oil on top of 225 uL outer solution (100 mM HEPES, 200 mM Glucose), 1-2 tubes of 225 uL outer solution (for washes), 20 uL inner solution (which is often 2x the volumes of the bulk

reaction), 500 uL of lipid/oil, and an empty tube to store the final vesicle solution is. Then, the 20 uL of inner solution is added to 500 uL of lipid/oil and vortexed for 2 minutes. This cortexed mixture is placed in the tube with 150 uL lipid/oil and 225 uL outer solution, which is then centrifuged at 4 ºC, 14,000 g for 10 minutes. The oil layer is removed and the pellet is transferred to a tube with 225 uL outer solution. This mixture is then mixed properly and centrifuged again at 4 ºC, 9,000 g for 5 minutes (a wash step). This wash step can be repeated up to two more times, depending on the size of the pellet. If the pellet is too large, there are likely substances other than vesicles, so more washes are performed. At last, the pellet is isolated and vesicles are prepared.

Fluorescence Microscopy of Bulk Experiments

To quantify the amount of protein produced in bulk experiments, we use fluorescence microscopy. deGFP DNA is placed into each reaction, and the output of GFP represents the amount of protein production. BioTek H1MF plate readers were used with excitation/emission of 585/615 nm and gain settings of 61 and 100, the optimal settings for deGFP detection. This allowed us to quantify protein production in real time, typically over a 24 hour period.

Confocal Microscopy of Vesicles

To visualize and quantify protein synthesis within vesicles, we used an Olympus IX81 Confocal Microscope. We are able to capture brightfield, GFP, and mCherry images. The brightfield is for the general environment, GFP to visualize protein production of deGFP, and mCherry detects the Rhod-PPE, which fluoresces and attaches to the vesicle membrane. Exposures were set at 100 ms. For time course data, we typically took images every 30 minutes for 12 hours. Micro-Manager and ImageJ software were used for image analysis.

Vesicles were placed on an agar pad for imaging. Agar pad makeup was 0.0520 g of low melt agarose in 3 mL vesicle outer solution (made up of 100 mM HEPES and 200 mM Glucose). This mixture was then microwaved for 30 seconds and 1 mL placed onto 22 mm x 22 mm glass cover slips after a 5 minute cool down period. Then, another cover slip is placed on top and left to sit for 45-60 mins. After the pads were ready, 15-20 uL of the vesicle solution was pipetted evenly onto the pad, it was left to dry for 15-20 minutes, and then placed on an imaging disc and imaged.

Firefly Luminescence ATP Assay

An ATP Detection Assay Kit – Luminescence was ordered from Cayman Chemical (Item No. 700410). It was stored in -20 ºC. Calibrations were performed as directed for the assay. A slope of 0.6178, intercept of 10.283, and an average blank value of 11.5 was used to transfer luminescence to ATP concentration. To make 0.25 mL of ATP assay reaction mixture, 0.05 mL of 5x Detection Assay Buffer, 0.2 mL nuclease free water, and 4.35 uL of DTT were mixed together. 0.25 mL of this was mixed with 1.25 uL of D-Luciferin (3 mg in 100 uL) and 0.25 uL Luciferase to make the final reaction mixture. 30 uL of this mixture was added to 3 uL of sample. This was incubated for 15 minutes, then endpoint luminescence data with gain 135 was collected in a H1MF BioTek plate reader.

Alpha-hemolysin

Pure Alpha-hemolysin was ordered from Sigma (catalog number: H9395). It is a lyophilized powder with sodium citrate buffer. This was stored in 4 ºC fridge. 1 uL of 100 uM alpha-hemolysin was used for 50 uL vesicles.

Automation

An Echo 525 Acoustic Liquid Handler was used to pipette all the reagents of cell-free reactions except for extract, which is too viscous for some settings. It can pipette with 25 nL resolution.

A Hamilton STARlet Liquid Handling Robot was used to collect ATP time course data. This robot was used to incubate ATP assay in 4 ºC, transfer 30 uL of the ATP assay to the appropriate wells, incubate for 15 minutes, and place the plate in a Biotek H1MF plate reader for luminescence data.

BioCRNpyler as a Model Generator and Systems Biology Markup Language (SBML) for Software Communication

We used BioCRNpyler (version 1.0.2) to model biochemical reactions in the form of Chemical Reaction Networks (CRNs). By providing high level specifications of our reaction pathways, such as molecule 1 + enzyme 1 → molecule 2, the software is able to develop sophisticated models which were then simulated via Bioscrape. Parameter values were toggled, included decreasing and increasing rate constants and initial conditions. Details about parameter values can be found on the corresponding GitHub for this research[10].

The EnergyTxTl mechanism was used. See SI for species, chemical reactions, and parameter values. This mechanism accounts for the dynamics in typical cell-free reactions with and without DNA.

Bioscrape for Model Output and Simulation

Bioscrape (version 1.0.2.2), or Bio-circuit Stochastic Single-cell Reaction Analysis and Parameter Estimation, is a Python package that can model and simulate the CRNs for our proposed mechanisms. We chose Bioscrape because it has options for different functions common in synthetic biology and useful for our project, such as protein synthesis delay, cell growth, cell divisions, and stochastic simulations. It is able to outputs results in a Panda's data-frame, which can be easily manipulated for plotting.

This software was used to simulate the reactions generated from EnergyTxTl in BioCRNpyler.

BIBLIOGRAPHY

1. Vincent Noireaux and Allen Liu. The New Age of Cell-Free Biology. *Annual Reviews*. 22:51-77. (Feb 2020).
    a. This review was integral to understanding the background in the field and the relevant progress needed moving forward.
2. Richard Murray. Genetically Programmed Artificial Cells and Multi-Cellular Machines. *Vannevar Bush Faculty Fellows Program*. (2017).
    a. This proposal helped explain the background and intuition behind pursuing cell free protein synthesis and synthetic cell problems.
3. Arbor Biosciences. The myTXTL system is a comprehensive solution for protein engineering and synthetic biology applications. *Arbor Biosciences*. (2014).
    a. This article gave a coherent review of how to use and apply the TXTL system.
4. Pautot, Sophie, et al. Production of Unilamellar Vesicles Using an Inverted Emulsion. *Langmuir*. 19: 2870-2879. (2003).
    a. This article was used to understand the technique of oil-emulsion to make vesicles.
5. Takasuka, T. et al. Cell-free translation of biofuel enzymes. *Methods Mol Biol*. 1118:71-95. (2014).
    a. We used this paper to compare and understand the importance of biofuel production and how CFPS was beneficial.
6. Garamella, J. et al. Chapter Nine – TXTL-based approach to synthetic cells. *Methods in Enzymology*. 617(217-239). (2019).
    a. We learnt about the benefits of encapsulating TXTL in synthetic cells.
7. Opgenworth, P. A molecular rheostat maintains ATP levels to drive a synthetic biochemistry system. *Nature Chemical Biology*. (2017).
    a. In this article, we learnt about a potential energy regeneration mechanisms that has promise for cell free protein production.
8. Altamura, E. et al. Light-driven ATP production promotes mRNA biosynthesis inside hybrid multicompartment artificial protocells. *Biorxiv* (2020).
    a. This article proved that ATP synthase could be integrated into vesicle membranes designed in lab, which could be used for ATP regeneration.

9. Rustem Ismagilov and Michael Maharbiz. Can we build synthetic, multicellular systems by controlling developmental signaling in space and time? (2007). Doi: 10.1016/j.cbpa.2007.10.003
    a. This article helped us understand the importance of multicellular systems and compartmentalization.
10. Garamella, J. et al. The All *E. coli* TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS Publications*. (Jan 2016).
    a. In this paper, we learnt about the updated TXTL reaction setup. We also recreated the experiment that alpha-hemolysin can be used to replenish energy buffer within vesicles.
11. Anderson, M. Energizing Eukaryotic Cell-Free Protein Synthesis With Glucose Metabolism. *FEBS Letters*, 589(15): 1723-1727. (2015).
    a. Here, we read about a potential method for ATP regeneration utilizing glucose.
12. Poole, William, et al. BioCRNpyler: Compiling Chemical Reaction Networks from Biomolecular Parts in Diverse Contexts. *Biorxiv*. (2020).
    a. This paper gave us insight about the background behind chemical reaction network modeling.
13. Singhal, V. et al. A MATLAB toolbox for modeling genetic circuits in cell-free systems. Vol 6. (2021).
    a. This paper gave us a base model to develop on top of.
14. Lee, Kyung-Ho and Kim, Dong-Myung. Recent advances in development of cell-free protein synthesis systems for fast and efficient production of recombinant proteins. *FEMS Microbiology Letters*. 365(17). (July 2018).
    a. This article gave us more insight into the developments of CFPS to optimize speed.
15. Wei, Eric and Endy, Drew. Experimental tests of functional molecular regeneration via a standard framework for coordinating synthetic cell building. *bioRxiv*. (Mar 2021). Doi: https://doi.org/10.1101/2021.03.03.433818.
    a. In this article, we learnt about different ways to approach the question of energy regeneration.
16. Ghosh, S. et al. NAD: A master regulator of transcription. *Biochimica et Biophysics Acta (BBA)*. 1799(681-693). (2010).
    a. Here, we learnt the importance of NAD, and how it has integral roles in transcription.
17. Jedrzejas, M. et al. Structure and mechanism of action of a novel phosphoglycerate mutase from *Bacillus stearothermophilus*. *The EMBO Journal*. 19(7): 1419-1430. (Apr 2000).
    a. In this article, we learnt about a potential mechanisms for a 3PGA breakdown reaction.

18. Adamala, K. Engineering genetic circuit interactions within and between synthetic minimal cells. *Nature Chemistry*. (2016)
    a. From this article, we learnt about a use case for interactions between artificial and real cells.
19. Takahashi, M. Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription-Translation (TX-TL) Systems. *ACS Synthetic Biology*. (2014)
    a. This article helped us understand the importance of TXTL and the current speed limitations.
20. Verchère, A. *In vitro* Investigation of the MexAB Efflux Pump from *Pseudomonas aeruginosa*. *JoVE Journal Biology*. (84), e50894, doi:10.3791/50894 (2014).
    a. In this article, we learnt about another use case for synthetic cells.
21. Ahmad Kalil and James Collins. Synthetic biology: applications come of age. *Nature Reviews.* 11, 367-379. (May 2010).
    a. We used this article to spearhead discussion about the overall concept of synthetic biology, and where synthetic cells may fit in.
22. Open collaboration supporting the science and engineering of building synthetic cells. https://www.buildacell.org/
    a. We used this website to understand the motivation behind building synthetic cells.
23. Caschera, Filippo and Noireaux, Vincent. Synthesis of 2.3 mg/ml of protein with an all *Eschericia coli* cell-free transcription-translation system. *Biochimie*. 1-7. (2013).
    a. This article highlighted the current drawbacks and limitations for CFPS with TXTL.
24. Garamella, J. et al. Chapter Nine – TXTL-based approach to synthetic cells. *Methods in Enzymology*. 617(217-239). (2019)
    a. This article gave more detailed reaction composition and procedures that allowed us to calculate concentrations for our TXTL mixtures.
25. Domitilla Del Vechhio and Richard Murray. Biomolecular Feedback Systems. (2014).
    a. This book helped to introduce important ideas of feedback, nonlinearities, and perturbations.

A p p e n d i x A

SUPPLEMENTARY INFORMATION

The base EnergyTXTL species and reactions from BioCRNpyler.

Species(N = 14) = {
metabolite[amino_acids] (@ 30.0),
  found_key=(mech=initial concentration, partid=None, name=amino_acids).
  search_key=(mech=initial concentration, partid=, name=amino_acids).
metabolite[Fuel_3PGA] (@ 30.0),
  found_key=(mech=initial concentration, partid=None, name=Fuel_3PGA).
  search_key=(mech=initial concentration, partid=, name=Fuel_3PGA).
metabolite[NTPs] (@ 5.0),
  found_key=(mech=initial concentration, partid=None, name=NTPs).
  search_key=(mech=initial concentration, partid=, name=NTPs).
protein[Ribo] (@ 0.0273),
  found_key=(mech=initial concentration, partid=None, name=protein_Ribo).
  search_key=(mech=initial concentration, partid=, name=protein_Ribo).
protein[RNAP] (@ 0.00933),
  found_key=(mech=initial concentration, partid=None, name=RNAP).
  search_key=(mech=initial concentration, partid=, name=RNAP).
dna[GFP] (@ 1e-06),
  found_key=(mech=initial concentration, partid=None, name=GFP).
  search_key=(mech=initial concentration, partid=, name=GFP).
complex[protein[Ribo]:rna[GFP]] (@ 0), complex[protein[RNAase]:rna[GFP]] (@ 0), complex[dna[GFP]:protein[RNAP]] (@ 0), complex[complex[protein[Ribo]:rna[GFP]]:protein[RNAase]] (@ 0), protein[RNAase] (@ 0), metabolite[NDPs] (@ 0), protein[GFP] (@ 0), rna[GFP] (@ 0),
}

Reactions (12) = [
0. dna[GFP]+protein[RNAP] <--> complex[dna[GFP]:protein[RNAP]]
 Kf=k_forward * dna_GFP * protein_RNAP
 Kr=k_reverse * complex_dna_GFP_protein_RNAP_
  k_forward=4.48
  found_key=(mech=energy_transcription_mm, partid=None, name=kb).
  search_key=(mech=energy_transcription_mm, partid=P, name=kb).
  k_reverse=2.48889e-06
  found_key=(mech=energy_transcription_mm, partid=None, name=ku).
  search_key=(mech=energy_transcription_mm, partid=P, name=ku).

1. metabolite[NTPs]+complex[dna[GFP]:protein[RNAP]] --> metabolite[NTPs]+dna[GFP]+protein[RNAP]+rna[GFP]
 Kf=k_forward * metabolite_NTPs * complex_dna_GFP_protein_RNAP_
  k_forward=0.010833333333333334

2. metabolite[NTPs]+complex[dna[GFP]:protein[RNAP]] --> complex[dna[GFP]:protein[RNAP]]

Kf=k_forward * metabolite_NTPs * complex_dna_GFP_protein_RNAP_
k_forward=3.25
found_key=(mech=energy_transcription_mm, partid=None, name=ktx).
search_key=(mech=energy_transcription_mm, partid=P, name=ktx).

3. rna[GFP]+protein[Ribo] <--> complex[protein[Ribo]:rna[GFP]]
Kf=k_forward * rna_GFP * protein_Ribo
Kr=k_reverse * complex_protein_Ribo_rna_GFP_
k_forward=0.819
found_key=(mech=energy_translation_mm, partid=None, name=kb).
search_key=(mech=energy_translation_mm, partid=rbs, name=kb).
k_reverse=0.002853659
found_key=(mech=energy_translation_mm, partid=None, name=ku).
search_key=(mech=energy_translation_mm, partid=rbs, name=ku).

4. 4metabolite[NTPs]+metabolite[amino_acids]+complex[protein[Ribo]:rna[GFP]] --> 4metabolite[NTPs]+metabolite[amino_acids]+rna[GFP]+protein[Ribo]+protein[GFP]
Kf=k_forward * metabolite_NTPs^4 * metabolite_amino_acids * complex_protein_Ribo_rna_GFP_
k_forward=0.192

5. 4metabolite[NTPs]+metabolite[amino_acids]+complex[protein[Ribo]:rna[GFP]] --> complex[protein[Ribo]:rna[GFP]]+4metabolite[NDPs]
Kf=k_forward * metabolite_NTPs^4 * metabolite_amino_acids * complex_protein_Ribo_rna_GFP_
k_forward=19.2
found_key=(mech=energy_translation_mm, partid=None, name=ktl).
search_key=(mech=energy_translation_mm, partid=rbs, name=ktl).

6. metabolite[Fuel_3PGA]+metabolite[NDPs] --> metabolite[NTPs]
Kf=k_forward * metabolite_Fuel_3PGA * metabolite_NDPs
k_forward=0.02
found_key=(mech=one_step_pathway, partid=NTPs_production, name=k).
search_key=(mech=one_step_pathway, partid=NTPs_production, name=k).

7. metabolite[NTPs] --> metabolite[NDPs]
Kf=k_forward * metabolite_NTPs
k_forward=1.77e-05
found_key=(mech=one_step_pathway, partid=NTPs_degredation, name=k).
search_key=(mech=one_step_pathway, partid=NTPs_degredation, name=k).

8. rna[GFP]+protein[RNAase] <--> complex[protein[RNAase]:rna[GFP]]
Kf=k_forward * rna_GFP * protein_RNAase
Kr=k_reverse * complex_protein_RNAase_rna_GFP_
k_forward=1.0
found_key=(mech=rna_degredation_mm, partid=None, name=kb).
search_key=(mech=rna_degredation_mm, partid=rna_GFP, name=kb).
k_reverse=1.26582e-06
found_key=(mech=rna_degredation_mm, partid=None, name=ku).
search_key=(mech=rna_degredation_mm, partid=rna_GFP, name=ku).

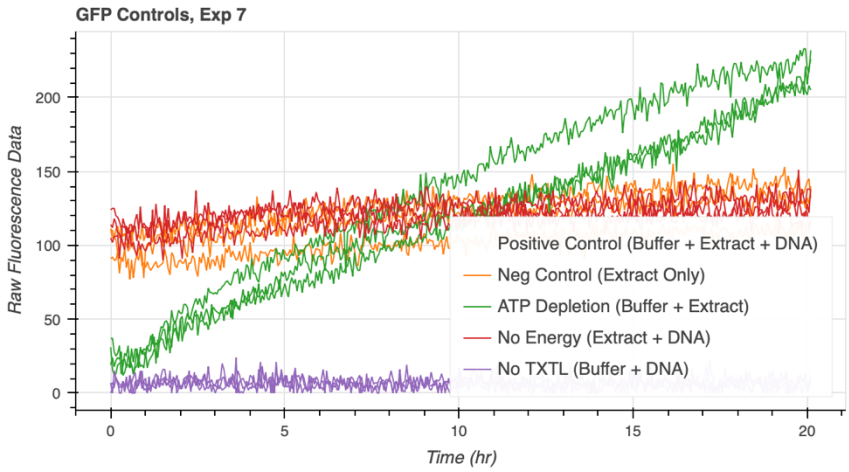
9. complex[protein[RNAase]:rna[GFP]] --> protein[RNAase]
Kf=k_forward * complex_protein_RNAase_rna_GFP_
k_forward=0.000555556
found_key=(mech=rna_degredation_mm, partid=None, name=kdeg).
search_key=(mech=rna_degredation_mm, partid=rna_GFP, name=kdeg).

10. complex[protein[Ribo]:rna[GFP]]+protein[RNAase] <--> complex[complex[protein[Ribo]:rna[GFP]]:protein[RNAase]]
 Kf=k_forward * complex_protein_Ribo_rna_GFP_ * protein_RNAase
 Kr=k_reverse * complex_complex_protein_Ribo_rna_GFP__protein_RNAase_
  k_forward=1.0
  found_key=(mech=rna_degredation_mm, partid=None, name=kb).
  search_key=(mech=rna_degredation_mm, partid=complex_protein_Ribo_rna_GFP_, name=kb).
  k_reverse=1.26582e-06
  found_key=(mech=rna_degredation_mm, partid=None, name=ku).
  search_key=(mech=rna_degredation_mm, partid=complex_protein_Ribo_rna_GFP_, name=ku).

11. complex[complex[protein[Ribo]:rna[GFP]]:protein[RNAase]] --> protein[Ribo]+protein[RNAase]
 Kf=k_forward * complex_complex_protein_Ribo_rna_GFP__protein_RNAase_
  k_forward=0.000555556
  found_key=(mech=rna_degredation_mm, partid=None, name=kdeg).
  search_key=(mech=rna_degredation_mm, partid=complex_protein_Ribo_rna_GFP_, name=kdeg).

GFP Fluorescence in a buffer + extract well steadily increases.

We observe this phenomena in our experiments. It may be due to increase in pH, which affects fluorescence. We also see that there is a non-zero fluorescence for the extract only well, which has a slight yellow color.



Magnesium and Potassium salt calibration for Rosetta Extract