MATHEMATICAL MODELING OF CELL-FREE PROTEIN SYNTHESIS SYSTEMS

Mohamed Akbar Ahmed Aqeel

University of Sri Jayewardenepura, Sri Lanka

BSc. Honors Degree

June 2022

MATHEMATICAL MODELING OF CELL-FREE PROTEIN SYNTHESIS SYSTEMS

A dissertation submitted to The Department of Chemistry of the University of Sri Jayewardenepura in partial fulfillment of the requirement for the Bachelor of Science honors degree in Chemistry
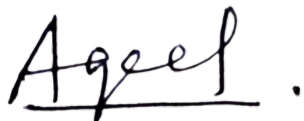
By

M.A.A.Aqeel

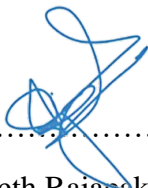University of Sri Jayewardenepura

June 2022

DECLARATION

The work described in this dissertation was carried out by me in the Department of Chemistry, University of Sri Jayewardenepura, Sri Lanka under the guidance of Dr. Suneth Rajapaksha and has not been submitted elsewhere.

10/07/2022

………………………………………   ……………………………………….

M.A.A.Aqeel   Date of submission

……………………………………….

Dr. Suneth Rajapaksha

Project Supervisor

……………………………………….

Prof. Champa Jayaweera

Head/Department of Chemistry

University of Sri Jayewardenepura

ACKNOWLEDGEMENT

First, I would like to thank my supervisor, Dr. Suneth Rajapaksha, Senior Lecturer, Department of Chemistry, University of Sri Jayewardenepura for his positive encouragement and knowledge. I am thankful to my mother for listening to every piece of scientific information I gathered for this Research.

LIST OF ABBREVIATIONS

GE   Gene expression

CFPS   Cell-free protein synthesis

NTP   Nucleotide triphosphate

AA   Amino acids

rDNA   Recombinant DNA

ODE   Ordinary differential equation

ODE-MM   Ordinary differential equation-based mathematical model

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Cell-free protein synthesis (CFPS) is a platform to synthesize proteins in a cell-free format. DNA molecules are artificially added into these open systems to produce the desired product. In this work, an ordinary differential equation-based mathematical model (ODE-MM) was developed to address the behavior of such CFPS systems. The model consists of a set of differential equations that describes the basic mechanics of cell-free gene expression (GE) and the effects of DNA inputs (i.e., in the form of initial doses), resource pools, and RNA degradation on product yield. For simulations, the required model parameters and initial concentrations of state variables are obtained from the available literature, and a typical model based on the literature values was developed (Model-L).

The model showed possible protein production yield could be only ~23 pM. But simulating the increment on resource pools showed that the maximum concentration of target protein obtainable reaches up to ~ 0.4 nM. Increasing the DNA input showed that the protein product obtainable proportionally increases up to a certain limit. For the model, the critical limit is ~2 nM. Simulating the inhibition of nuclease activity in the model showed that the maximum concentration of product yield achievable is ~1 nM. Increasing the concentrations of ribosomes in the model, increased the product yield with a gradient of ~0.022.

CHAPTER 1

INTRODUCTION

1.1 Cell-free protein synthesis (CFPS) systems

In bacteria, nutritional and environmental conditions of the surroundings are relayed through small molecules. The rate of GE can be increased or decreased accordingly to the type of signal presented, and it can also modulate the concentrations of background machinery involved in GE. Therefore, utilizing microbial cultures by genetic engineering and DNA manipulation techniques to produce certain compounds at considerable yields might give confounding results.[7,9]

Extracting the background machinery required for GE, such as ribosomes, RNA polymerases, energy-producing molecules, and other co-factors from the controlling elements of the bacterial cells (their inherent genes and cell membranes) and assembling them in a reaction vessel[21,25,32] to conduct GE independently is a promising alternative.[13,16,26] This can be done by growing microbial cultures in a suitable medium and harvesting them at an appropriate time. During the harvest, certain lytic compounds are added into the medium to destroy the cell wall and cell membranes of the bacterial cells. This in turn makes the cytoplasmic contents seep out into the surrounding environment. After this lysis process, the lysate is centrifuged, and the clear supernatant is extracted.[6] The supernatant/extract obtained will contain the necessary components to conduct GE independently and can be programmed with DNA to express certain target proteins. These types of expression systems are called cell-free protein synthesis (CFPS) systems.[35,45] The method to obtain cell-free extract is outlined in Figure 1.1.
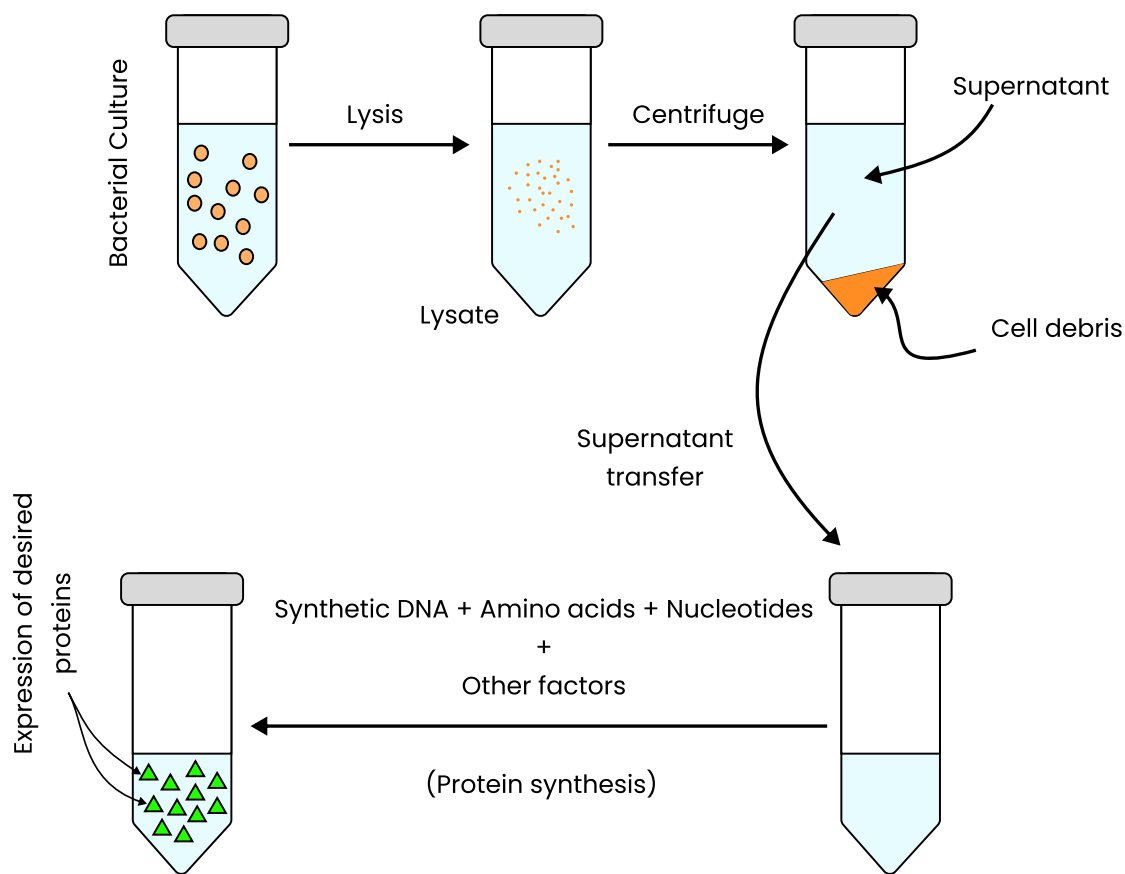
Figure 1.1: Preparation of cell-free extract for CFPS experiments. Note the addition of DNA and resources to commence gene expression.

The amount of background machinery available to conduct GE in the extract is determined by the type of bacteria, centrifuging conditions (sedimentation rate), culture medium, and the specific growth stage of the microbial culture at the time of harvesting.[3,4,6] By carefully choosing these parameters, certain cytoplasmic environments can be mimicked inside the extract. For example, an extract can be rendered to produce more proteins than any other compound. In addition to this feature, chemical compounds that are not familiar with the norms of biology can also be added to these systems to improve certain features such as

addition of non-canonical amino acid incorporation into proteins that can possibly improve the stability.[17]

1.2. Programming the extract with DNA

The capability to join different DNA segments from different sources allowed scientists to build recombinant DNA (rDNA) molecules.[6,45] The introduction of these rDNA molecules into a cell-free extract to amplify the synthesis of certain proteins is popular among pharmaceutical industries.[1,5] Because the bottom-up synthesis of certain therapeutic biomolecules (e.g., protein hormones) from scratch with considerable yield and desired stereochemistry may not be possible with the available energy landscape of chemical reactions or sometimes expensive with available synthetic capabilities. Therefore, utilizing standard rDNA techniques to assemble the set of genes involved in the biosynthesis of the desired product into a cluster and then introducing it into cell-free systems for expression (which will eventually give a polycistronic RNA during transcription) is a profitable alternative.[12] The concentration of desired protein harvested per batch depends on the concentration of DNA, the amount of background machinery, and the resources available to the CFPS systems.

1.3 Advantages of CFPS systems

The open nature of CFPS allows researchers to directly control, monitor, automate, and maintain accurate concentrations of molecular resources for the biochemical reactions proceeding in the extract.[34,48] Prototyping of novel metabolic circuits, decoding the complex biochemical interactions of gene networks for treatment, and biomanufacturing complex and synthetically "impossible" therapeutic biomolecules are all made possible by these systems

since they can allow direct addition of chemical components to the CFPS systems that are unorthodox or even toxic within cell-dependent norms.[13,23,27,47] Mixing the extract with another extract of a different source is also possible and it can provide unique cytoplasmic environments for the CFPS systems that never existed in natural cells.[33] Manufacturing therapeutics that are in demand among small populations or individuals (for example, niche drugs for rare genetic, metabolic, and allergy conditions) can be synthesized by these platforms because it doesn't require any expensive infrastructure from pharmaceutical companies.[28,48] Energy efficiency compared to cell-based expression systems is high due to the resources provided to CFPS will be consumed only for the synthesis of the target product. In contrast, cell-based systems partition their energy reservoir for homeostasis, division, and cellular activities, and not only for protein synthesis.[6,12]

1.4 Limitations of CFPS systems

The full control to manipulate and predict the outcome of GE from these systems is limited. Because the extract will always contain components that can affect the information flow from DNA to target proteins.[48] A particular example is the RNA digesting enzymes called ribonucleases (RNases).[29,30] These enzymes can degrade the RNA molecules synthesized in CFPS, which are required to synthesize target proteins. The synthesized target proteins can also be susceptible to degradation if the extract contains another family of enzymes called proteases.[41] These molecules are part and parcel of these types of CFPS extracts and cannot be separated.[48] Keeping the costs low is a high priority when operating these systems and it is more economical to execute DNA programs in a small number of large batches than in many small batches. But the volumes of CFPS extracts that can be obtained from microbial cultures

are only in the order of a few microliters.[2,19] This makes the process somewhat expensive to scale up and exists as a niche method in biotechnology. Another limitation is that the CFPS systems cannot perpetuate the process of GE without a continual supply of energy-rich molecules since they lack the mechanisms to regenerate energy-rich molecules.[46,47]

1.5 Modeling biosystems and pathways

To address the behavior of biochemical systems, each biochemical interaction of the system must be decomposed and analyzed.[8] This can be done by representing each reaction of the system as an equation (rather than interaction diagrams). The set of equations to describe the system can be called as a model. By developing models, the different interpretations available to describe the behavior of the system can be reduced (i.e., reduction in ambiguity). One of the advantages of constructing a model is that certain changes to the system can be simulated and their effects can be predicted. Therefore, will act as a cornerstone to decode certain metabolic and genetic diseases associated with biochemical systems.

1.6 Significance of modeling CFPS

Even though this branch of biotechnology is growing, the detailed quantitative models to address the behavior of CFPS systems are still missing.[2] The product yield obtained from CFPS within the batch reaction time is interconnected with the concentrations of DNA initially added and the concentrations of molecular resources available in the extract.[1,45] Developing mathematical models that are constrained by these parameters is essential to test the potential

and their limitations. Further, it helps to focus on a particular problem while other aspects and problems are abstracted away.[8]

1.7 Objectives of the study

Main objective

- To develop a mathematical model that grasps the basic mechanics of gene expression in CFPS systems.

Specific objectives

- To understand the model behavior towards the concentration of initial DNA inputs and product yield.
- To understand the model behavior towards the resource constraints confronted by CFPS systems.
- To check the influence of RNA degradation on the protein yield attainable and the response towards initial DNA inputs.

CHAPTER 2
MODEL DEVELOPMENT

The GE consists of two stages.[10] The first stage, transcription, involves transferring of information encoded in DNA to a complementary RNA chain. Transcription is carried out by a protein complex called RNA polymerase (RNAP), which binds to the specific nucleotide sequences present in the gene. After binding it diffuses one-dimensionally along the DNA, concomitantly forming RNA strand from nucleotide precursors.[4,10,12] The second step of gene expression is translation, in which the RNA molecule binds a protein called a ribosome, which reads the nucleotide sequence and produces a corresponding target protein. Translation, like transcription, diffuses along the RNA chain concomitantly forming the desired proteins from amino acid building blocks.[10,11,12]

The difficulty in modeling genetic systems is that the molecules involved in the biochemical reactions of GE are often present in discrete amounts.[8] The background machinery that impacts GE is often present in small quantities: the number of ribosomes and RNAPs is regularly in the hundreds or less. Moreover, genes themselves are almost always present at very low numbers. When the molecule count is low, ensemble-based kinetics such as rate laws cannot be used to address the behavior.[15] However, in the CFPS systems, the open nature allows the users to maintain the necessary molecular components to be in ensemble amounts.[8]

Mass action-based formalisms are preferred to address GE in CFPS because the biochemical interactions can be simply represented in terms of effective rate constants and concentrations. In this work, an ordinary differential equation-based mathematical model (ODE-MM) was built to encompass the critical biochemical events associated with CFPS

systems in terms of effective rate constants and concentration of molecular components.[2,8] The

model was built on two assumptions:

Assumption 1 – The model is homogeneous, and small in volume. Therefore, reactions can be

described independent of position (i.e., the reason for the modifier - ordinary).

Assumption 2 – There are ensemble number of species present, therefore, the molecular

abundance can be described by a concentration that varies continuously.

The developed model consists of a set of differential equations (i.e., rate equations).

Solving these equations gives the time-varying changes in species concentrations. The

behavior of CFPS under various conditions can be simulated by adjusting the effective

parameters and concentrations of the model rate equations. The simulations are carried out via

the software R (version 4.1.3) and the differential equations are solved using the backward

differentiation formula (BDF) available in the R package deSolve (version 1.31).[5,14] The

typical information flow of canonical GE applies to CFPS systems since they mimic the

cytoplasmic environment of bacterial cells. The typical events of GE in a cell-free system are

given in Figure 2.1 and the reaction scheme is given in Scheme 2.1.
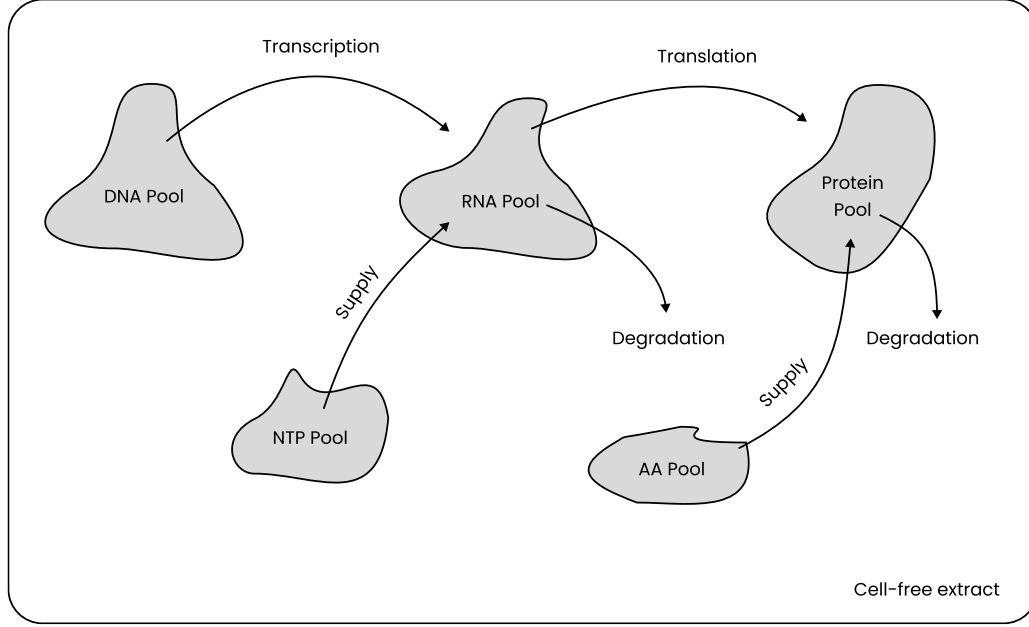
Figure 2.1: Gene expression in CFPS. The DNA, RNA, and AA pools refers to the initial DNA

nucleotide triphosphate, and amino acids added to the extract.
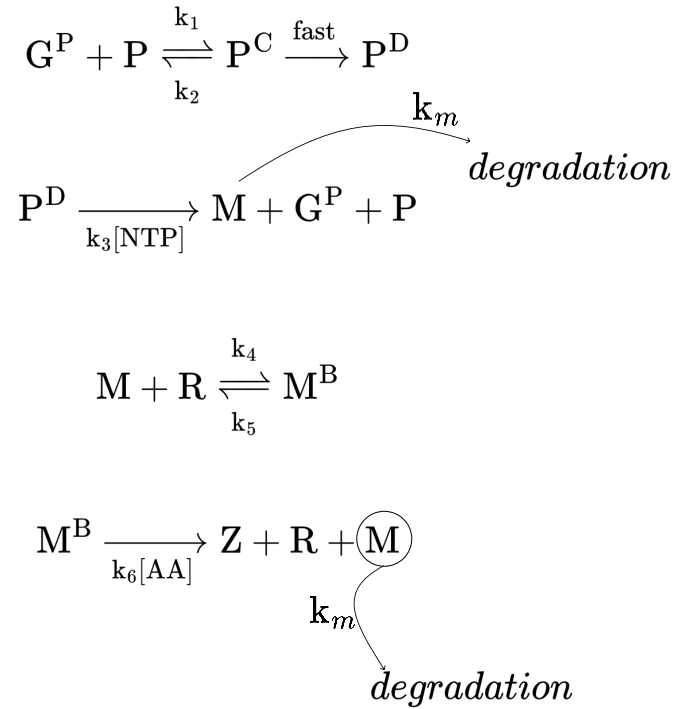
Scheme 2.1: The reaction scheme for gene expression in CFPS. Where G^p = initial [DNA], P

= [RNA polymerase/RNAP], NTP = [nucleotide triphosphate], AA = [amino acids], P^C =

[transiently bound promoter], P^D = [isomerized promoter], R = [ribosome], M = [RNA chain], M^B = [RNA-ribosome complex], and Z = [target protein] respectively.

2.1 Modeling of Transcription

The model begins with the transient binding of RNA polymerase (RNAP) towards the promoter region, forming a promoter complex.[10,12] The reaction for RNAP and promoter binding is assumed to be a second-order reaction given as,

$$G^P + P \underset{k_2}{\overset{k_1}{\rightleftharpoons}} P^C \tag{2.1}$$

Where G^P, P, P^C, k_1, and k_2 specify the [promoter], [RNAP], [promoter complex], RNAP association rate constant, and RNAP dissociation constant respectively. The promoter complex should undergo several isomerization steps to initiate transcription.[4,11,36] But the number of isomerization steps and the factors involved are not fully characterized in the literature, therefore, here, it is assumed that RNAP binding is first order and immediately transitions the promoter complex to the required isomer. Therefore, equation 2.1 can be rewritten as,

$$G^P + P \underset{k_2}{\overset{k_1}{\rightleftharpoons}} P^C \overset{fast}{\longrightarrow} P^D \tag{2.2}$$
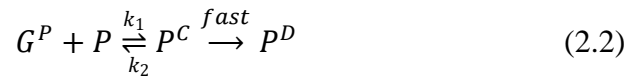
Where P^D specifies the required isomer for transcription initiation. The above reaction is very rapid on the timescale of GE reactions; therefore, the concentration behavior of P^C is neglected during formulations. After this conversion, the RNAP complex will escape from the promoter region and translocate downstream of the gene, concomitantly synthesizing the RNA. For this model, it is assumed that the synthesized RNA chain will be released to the CFPS extract only

when the terminating signals are confronted at the very end of the gene sequences.[6,12] The reaction for the RNA release, promoter escape and RNAP release can be given as,

$$P^D \xrightarrow{k_3 \cdot [NTP]} M + G^P + P \qquad (2.3)$$
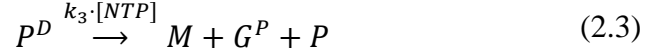
Where M and k_3 specify the [RNA chain] and second-order rate constant for RNA production. It should be noted that the rate constant in equation 2.3 requires nucleotide triphosphate (NTP) available in the CFPS extract to proceed. The consumption of NTP can be given as,

$$NTP \longrightarrow \varphi \,(\text{consumption}) \qquad (2.4)$$
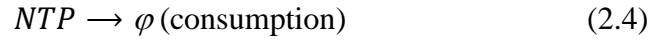
The degradation of RNA chain in the presence of RNase can be given as,

$$M \xrightarrow{k_m} NT \qquad (2.5)$$

Where $[NT]$ and k_m specify the nucleotides and degradation rate constant respectively.[38,40,41,42] In this formulation (equation 2.5), the RNA after degradation cannot replenish the NTP pool unless it is phosphorylated with certain enzymes and, here, it is assumed the CFPS system is incapable of regenerating NTP from NT.[6,12]

2.2 Modeling of Translation

The RNA chains that are released to the cell-free extract at the end of transcription will be recognized by ribosomes and can bind transiently to form a ribosome-RNA complex. This transient interaction is assumed to be second-order in both species and given as,

$$M + R \underset{k_5}{\overset{k_4}{\rightleftharpoons}} M^B \qquad (2.6)$$

Where R, M^B, k_4, and k_5 specify the [ribosome], [ribosome-RNA complex], forward rate constant for complex formation, and complex dissociation respectively. Like transcription, the bound complex can tread the RNA chain concomitantly synthesizing the desired protein product by adding complementary amino acids to the RNA chain. After reaching the terminating sequences of the RNA chain, the complex will dissociate to give the protein product.[36,37] This reaction is dependent on the [amino acids] (AA) available in the extract and can be given as,

$$M^B \xrightarrow{k_6 \cdot [AA]} M + R + Z \tag{2.7}$$

Where Z and k_6 specify the target protein product and second-order rate constant for protein production respectively. The degradation of target proteins is assumed to be negligible and not formulated explicitly.[41] The reaction for consumption can be given as,

$$AA \longrightarrow \varphi \, (\text{consumption}) \tag{2.8}$$
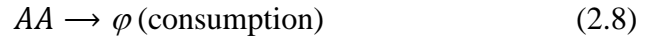
The differential equations describing the complete model are given in Table 2.1.

Table 2.1 : The differential equations for the complete model	
Molecular Component[1]	**Corresponding differential equations**
The promoter, $[G^P]$	$\dfrac{d[G^P]}{dt} = -k_1[G^P][P] + (k_2 + k_3[NTP])[P^D]$
RNAP, $[P]$	$\dfrac{d[P]}{dt} = -k_1[G^P][P] + (k_2 + k_3[NTP])[P^P]$

Promoter Complex[2], $[P^D]$	$$\frac{d[P^D]}{dt} = k_1[G^P][P] - (k_2 + k_3[NTP])[P^D]$$
RNA, $[M]$	$$\frac{d[M]}{dt} = k_3[P^D] - k_4[M][R] + (k_5 + k_6[AA])[M^B] - k_m[M]$$
Ribosome, $[R]$	$$\frac{d[R]}{dt} = -k_4[M][R] + (k_5 + k_6[AA])[M^B]$$
RNA-ribosome complex, $[M^B]$	$$\frac{d[M^B]}{dt} = k_4[M][R] - (k_5 + k_6[AA])[M^B]$$
Target protein, $[Z]$	$$\frac{d[Z]}{dt} = k_6[AA][M^B]$$
NTP pool, $[NTP]$	$$\frac{d[NTP]}{dt} = -k_3[NTP][P^D]$$
AA pool, $[AA]$	$$\frac{d[AA]}{dt} = -k_6[AA][M^B]$$

[1] The molecular component NT is not modeled explicitly since it cannot contribute to the NTP pool without any expenditure of energy-rich molecules.

[2] The molecular component P^C is not formulated since the isomerization is rapid (equation 2.2), therefore, $P^C \cong 0$ always.

CHAPTER 3

RESULTS AND DISCUSSION

The differential equations of the model were solved using the model parameters and initial condition given in Table 3.1. The model with these initial settings has been given the name Model L and is used as a reference to compare the changes that are introduced into the model.

Table 3.1: Model parameters and initial conditions for CFPS system	
Model parameters	
(Obtained from the lacZ operon kinetic data available in the literature)[3]	

Description	Value
RNAP association constant, k_1	$6 \times 10^9 M^{-1}min^{-1}$
RNAP dissociation constant, k_2	$600\ min^{-1}$
The rate constant for RNA synthesis[a], k_3	$1 \times 10^{10}M^{-1}min^{-1}$
Ribosome association constant, k_4	$6 \times 10^9 M^{-1}min^{-1}$
Ribosome dissociation constant, k_5	$135\ min^{-1}$
The rate constant for protein production[b], k_6	$1 \times 10^{10}M^{-1}min^{-1}$
The rate constant for RNA degradation, k_m	$18\ min^{-1}$

Batch time/reaction time, t_b	30 min

Initial conditions[c]

Description	Initial concentration (nM)
Promoter concentration, $[G^P]$	1.00
RNAP concentration, $[P]$	1.00
Promoter complex concentration, $[P^D]$	1.00
Ribosome concentration, $[R]$	1.00
RNA concentration, $[M]$	0.00
RNA-ribosome concentration, $[M^B]$	0.00
Target protein concentration, $[Z]$	0.00
NTP concentration, $[NTP]$	1.00
AA concentration, $[AA]$	1.00

[a,b] The values are not available in the literature, therefore, the maximum theoretical value an enzyme can achieve is used instead.

[c] The initial value can be maintained at these concentrations without any complications therefore, arbitrarily chosen.

3.1 Model-L

From the simulation of Model-L shows that both promoter complex, G^P and $RNAP$, P reaches a steady state after ~600ms. Only a fraction of promoters and RNAPs (~1%) are involved in the formation of promoter complex, P^D and the rest remain freely in the CFPS extract. The fractional promoter occupancy is given in Figure 3.1 and the concentration behavior of RNAP is given in Figure 3.2. at equilibrium, the relationship between promoter, RNAP, and bound-promoter can be given as,

$$\frac{k_1}{k_2} = \frac{[P^D]}{[G^P][P]}$$

The model shows, that the initial DNA input of 1nM gives only 10 pM of promoter complex. This bottlenecking will be reflected in the RNA chain synthesis because it requires the RNAP-bound promoters as a reactant (Equation 2.3) to proceed with GE. The concentration behavior of the bound-promoter is shown in Figure 3.3. To increase (i.e., to relieve from bottleneck) the fraction of bound-promoter, P^D two factors in the Model-L can be changed. First, can be done by changing the ratio of rate constants (i.e., equilibrium association constant , K_a) as in the above equation or can be done by simply increasing the initial DNA input (Section 3.2). Changing the value of K_a can be done by DNA manipulation techniques, and the effect of such changes are not modeled in this research since the manipulation one rate constant can change the other value congruently.
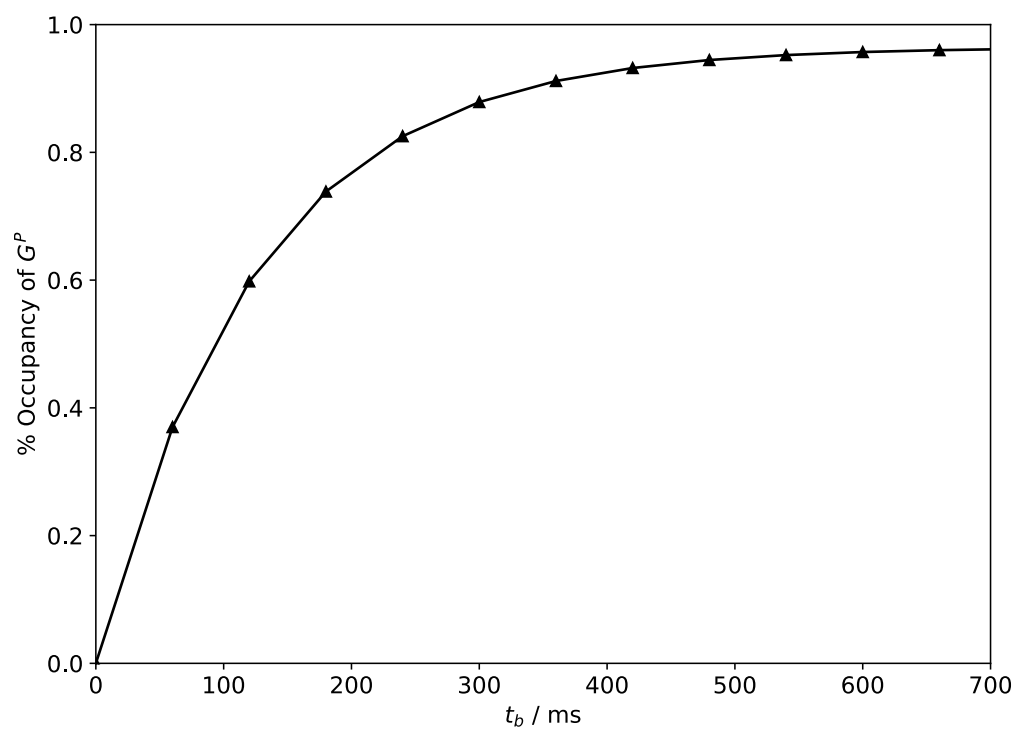
Figure 3.1: Percentage occupancy of promoters.
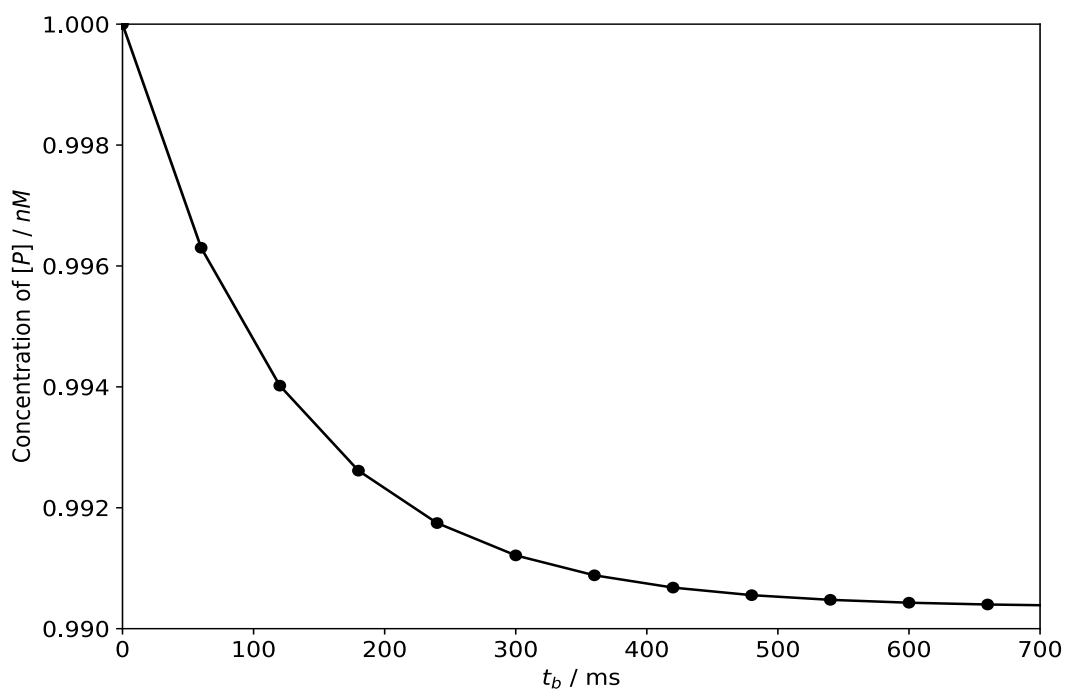
Figure 3.2: Concentration behavior of RNAP.
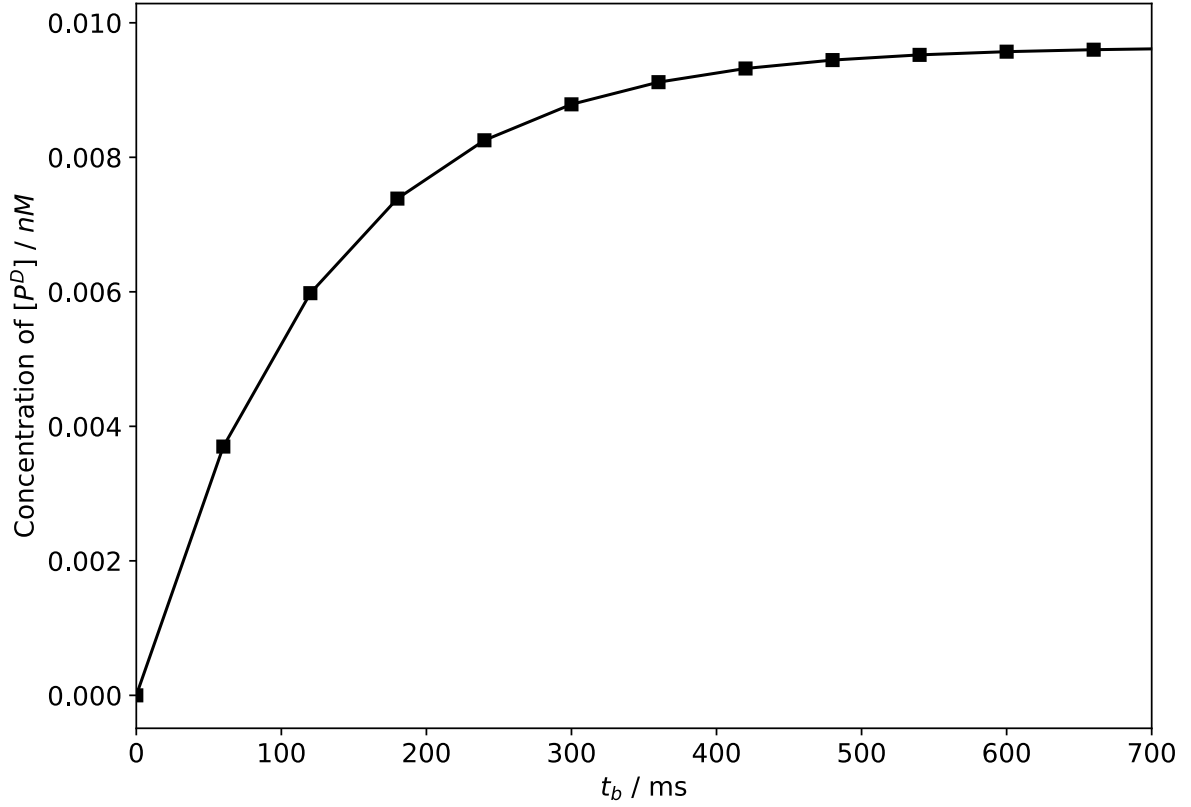
Figure 3.3: Concentration behavior of RNAP-promoter complex.

In addition to this bottleneck effect, RNA degradation will also play a crucial role in target product yield. In the presence of RNA degradation, the CFPS extracts cannot maintain RNA levels at a steady state to keep the information flow from DNA to target proteins. This behavior is evident in Model-L, where the concentration of RNA chain goes through a maximum of 5 pM at ~15s after initializing the batch and decays exponentially. This behavior is shown in Figure 3.4.
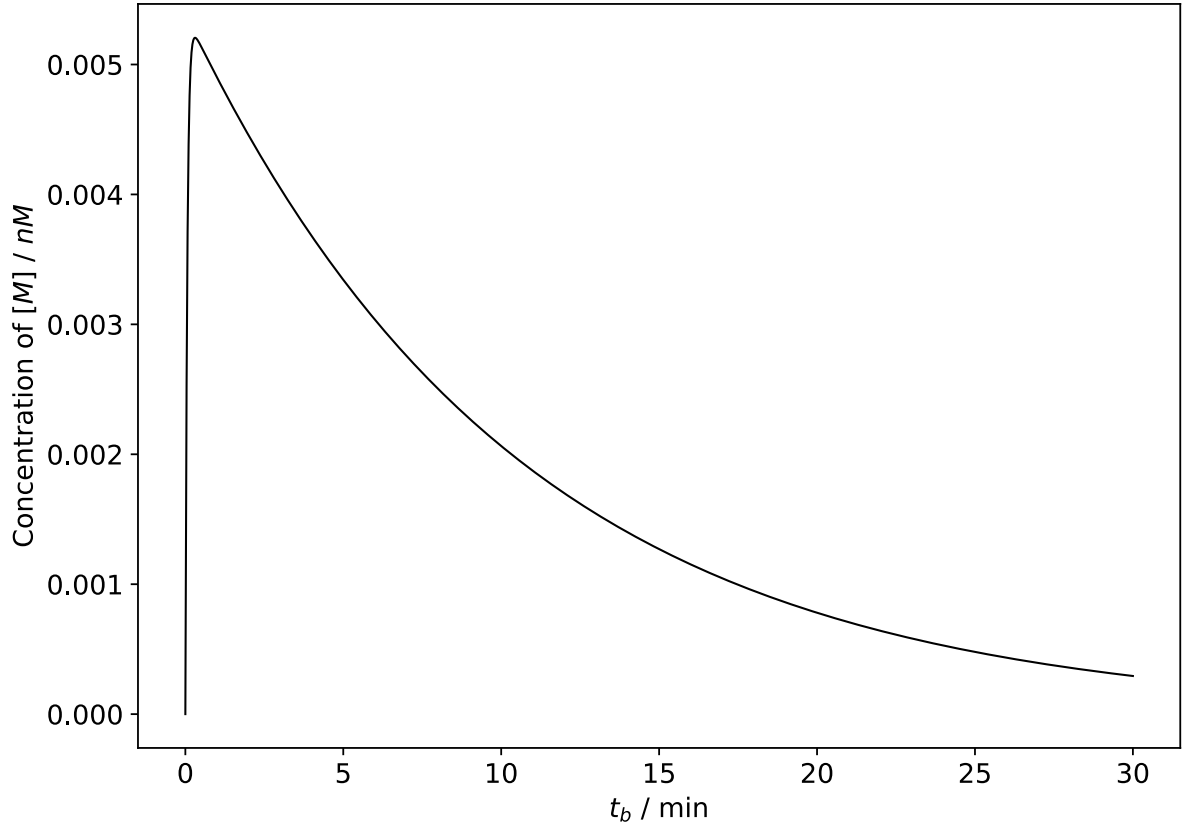
Figure 3.4: Concentration behavior of RNA.

This degradation behavior should be controlled inside the CFPS systems since it will cause indirect depletion of the energy-rich NTP pool available to the system. This depletion can be given as,

$$NTP \rightarrow RNA \xrightarrow{k_m} NT$$

Therefore, the NTPs added to the CFPS extract will be channeled into this path and the system will retain no RNA molecules eventually (reason for the exponential decay behavior). The concentration profile of NTP depletion is given in Figure 3.5.
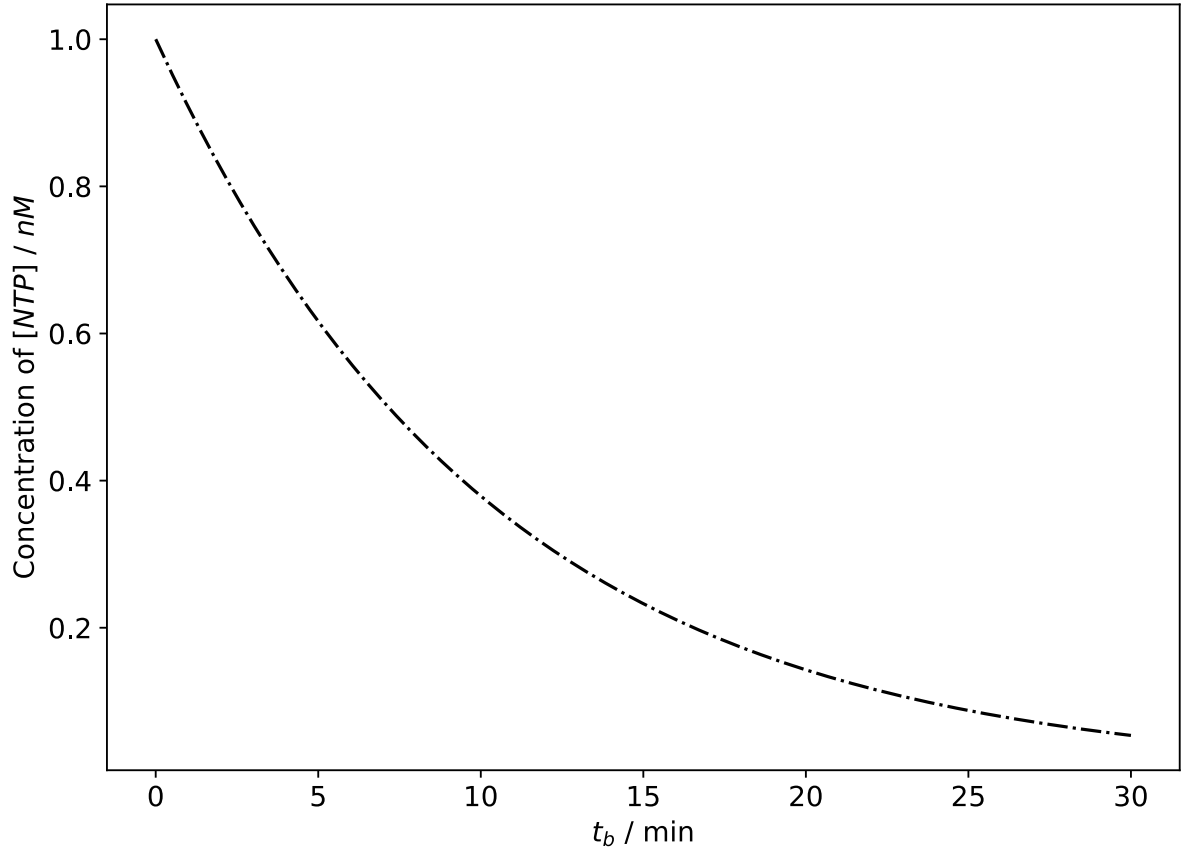
Figure 3.5: Concentration behavior of NTP.

In bacterial cells, this resource depletion is not usually observable since they have the necessary enzymes and methods to replenish the NTP pools.[39] In contrast to NTP pools, the AA pools show only a 2% decrease throughout the batch time of Model-L. The reason for this behavior can be attributed to the low production of RNA in the extract and the negligible protein degradation rate previously assumed (Equation 2.7 & 2.8). The concentration behavior of AA pool is given Figure 3.6.
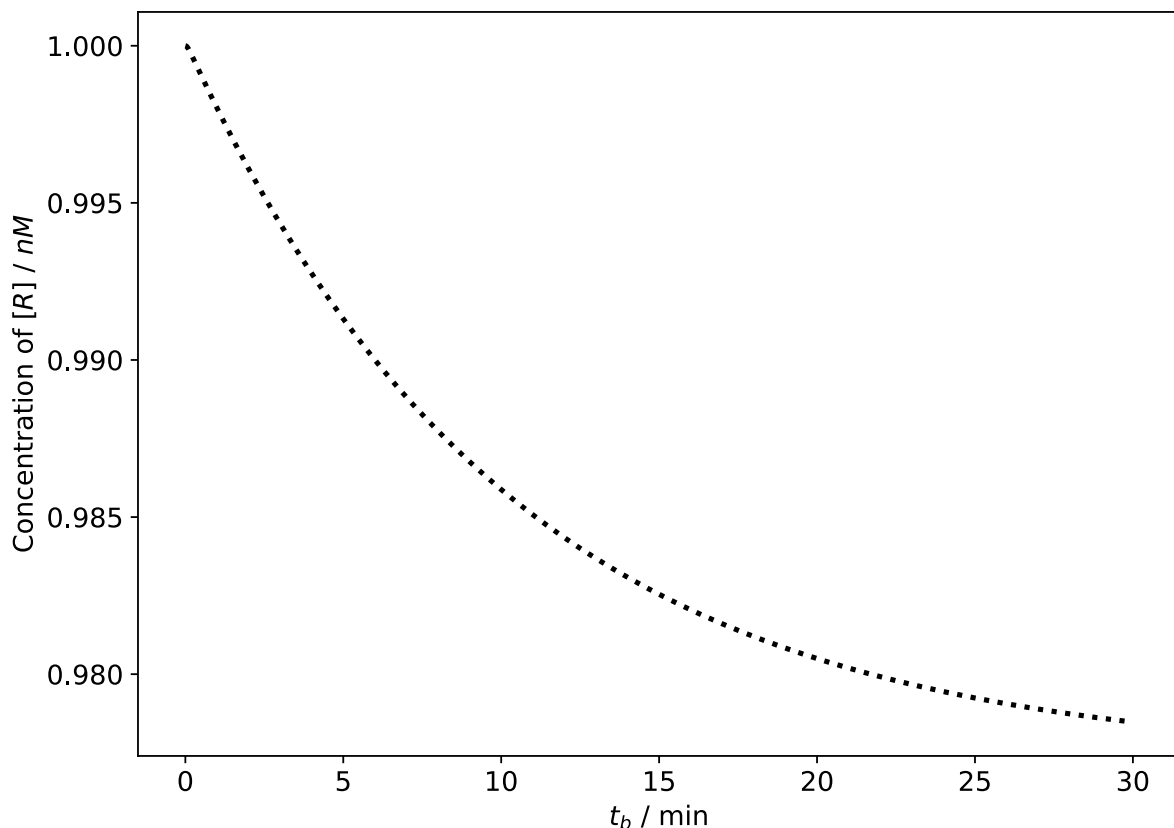
Figure 3.6: Concentration behavior of AA pool

Further, the NTP pool depletion 'discourages' the ribosomes to initiate or engage itself in protein synthesis. The ribosomes engaged in protein synthesis (which is 0.03% of the total ribosomes) are recovered completely after approximately 30min of batch initialization. The ribosome engagement and recovery are given in Figure 3.7. The concentration of ribosome goes through a minimum of 30pM at ~15s after the batch initialization. So far, the concentrations of background machinery that are used to initialize the systems (i.e., ribosomes, RNAP, and DNA) are almost remains the same in the CFPS extract (RNAP ~99%, DNA ~99%, and ribosomes ~100%) and indicates that the Model-L conditions are not a suitable setting to obtain target proteins at high yield. Further, it should be noted that the concentration behavior

of ribosome is inversely proportional to the RNA molecules available in the extract (Equation 2.6).
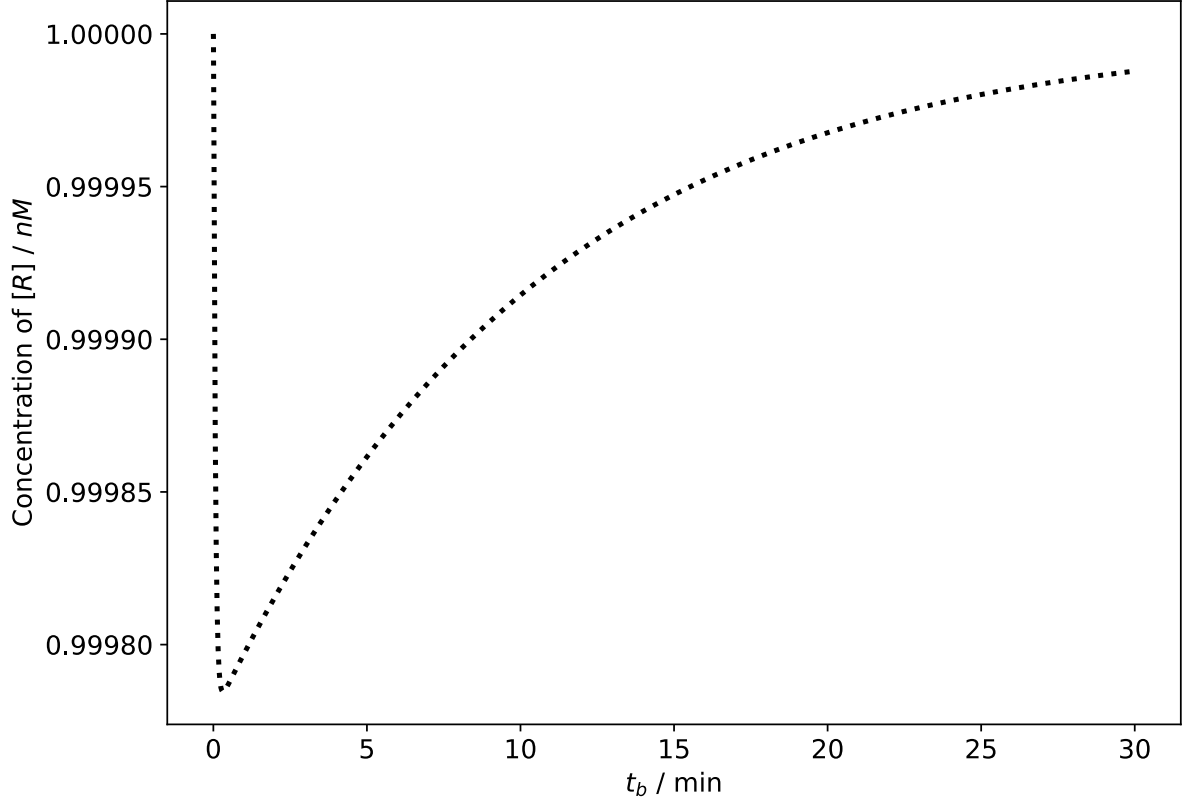
Figure 3.7: Recovery behavior of ribosomes

The propensity of Model-L towards translation can be elucidated by the ribosome-RNA complex formation (M^B formation) and is given in Figure 3.8. The concentration behavior shows that extremely small concentration of M^B is produced. The ribosome-bound RNA complex goes through a maximum of ~0.23 pM at 15 s of batch initialization and decays exponentially.
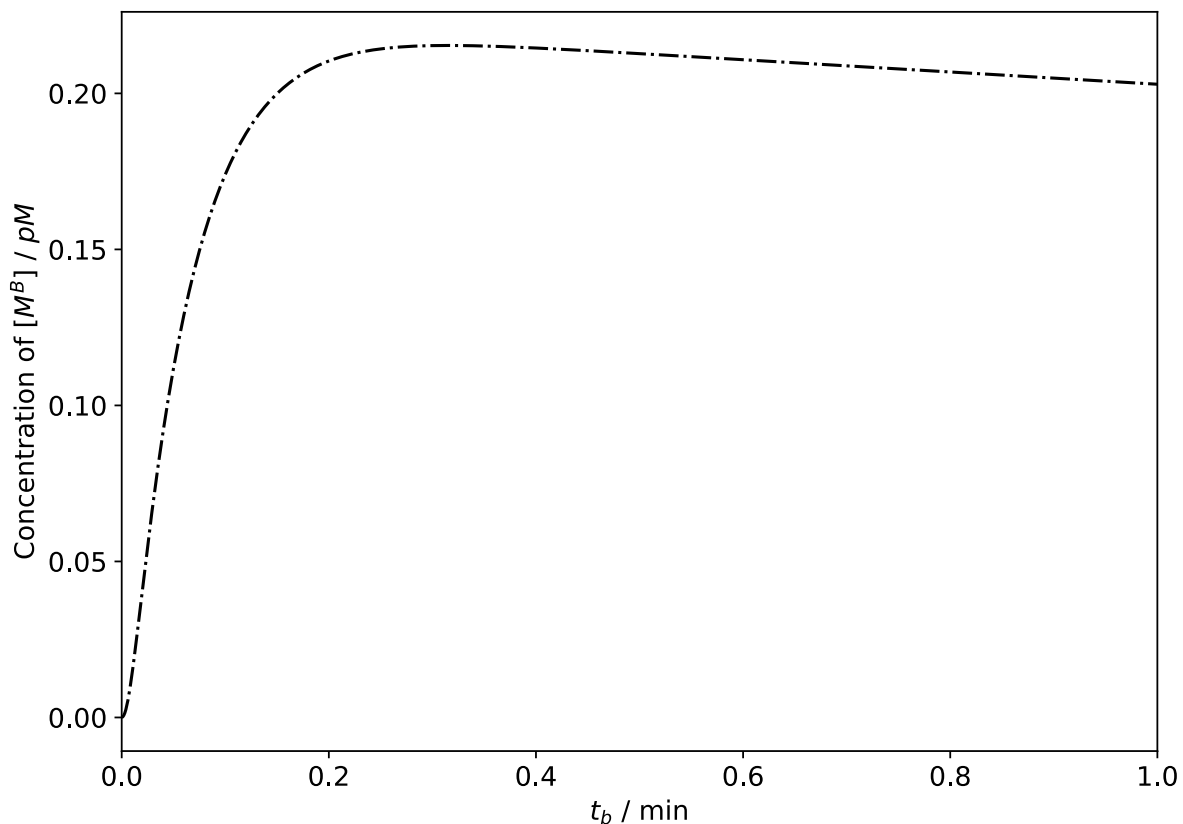
Figure 3.8: Concentration behavior of ribosome-RNA complex

The concentration behavior of target protein production (Z) is shown Figure 3.9. This hyperbolic saturation curve observed is due to the lack of protein degradation in the Model-L (assumed in Section 2.2) and resource constraints available to the CFPS extract. From the results, it is evident that the maximum concentration of target protein attainable by Model-L is ~23 pM (even after 30 min of batch reaction time). This may seem to be very discouraging at first but the open nature of CFPS systems allows for manipulation of the biochemical reactions of GE by adding necessary compounds required in CFPS. Therefore, the number of parameters and initial conditions that can be changed individually or in combination is high in these kinds of CFPS systems. But in this research, few conditions and parameters are changed relative to

the model settings of Model-L (Table 2.1). The complete R code to replicate the behavior of Model-L is given in APPENDIX 1.
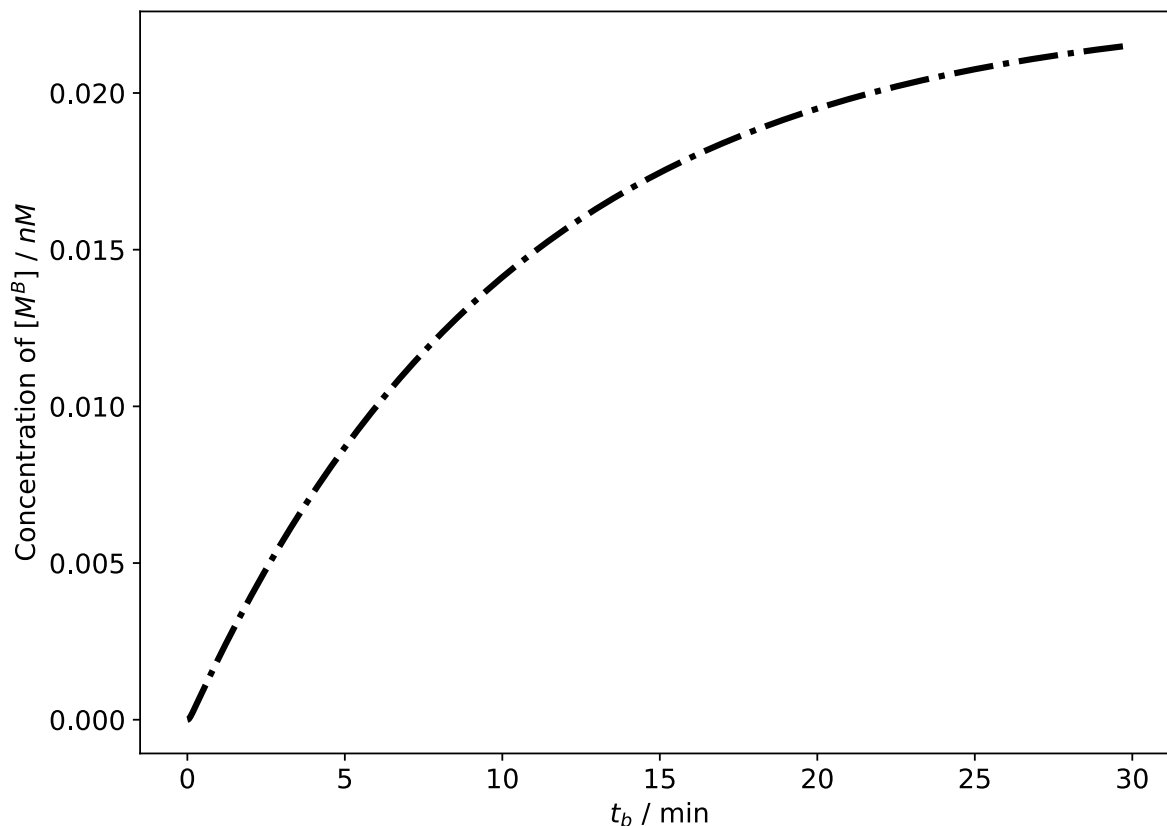
Figure 3.9: Concentration behavior of target protein

3.2 Effect of DNA concentration on target protein yield

From the simulation results of Model-L, it is evident that RNAP binding on to the promoter reaches equilibrium immediately (Figure 3.1). Therefore, increasing the initial DNA input to the CFPS system can increase the promoter concentration according to the Le Chatelier principle (Equation 2.1). This can concomitantly improve other reactions that are directly or indirectly associated with promoter complex formation (Equations 2.3, 2.6, and 2.7).

Experimental observations show that, by increasing the DNA input, the concentration of target protein in the cell-free extract can be increased and after a critical DNA input, the concentration of target protein reaches a plateau and no longer increases to further addition of DNA.[18,20] Testing this behavior on Model-L has shown the same behavior as in experimental observations. The simulation result in response to the initial DNA input is given in Figure 3.10.
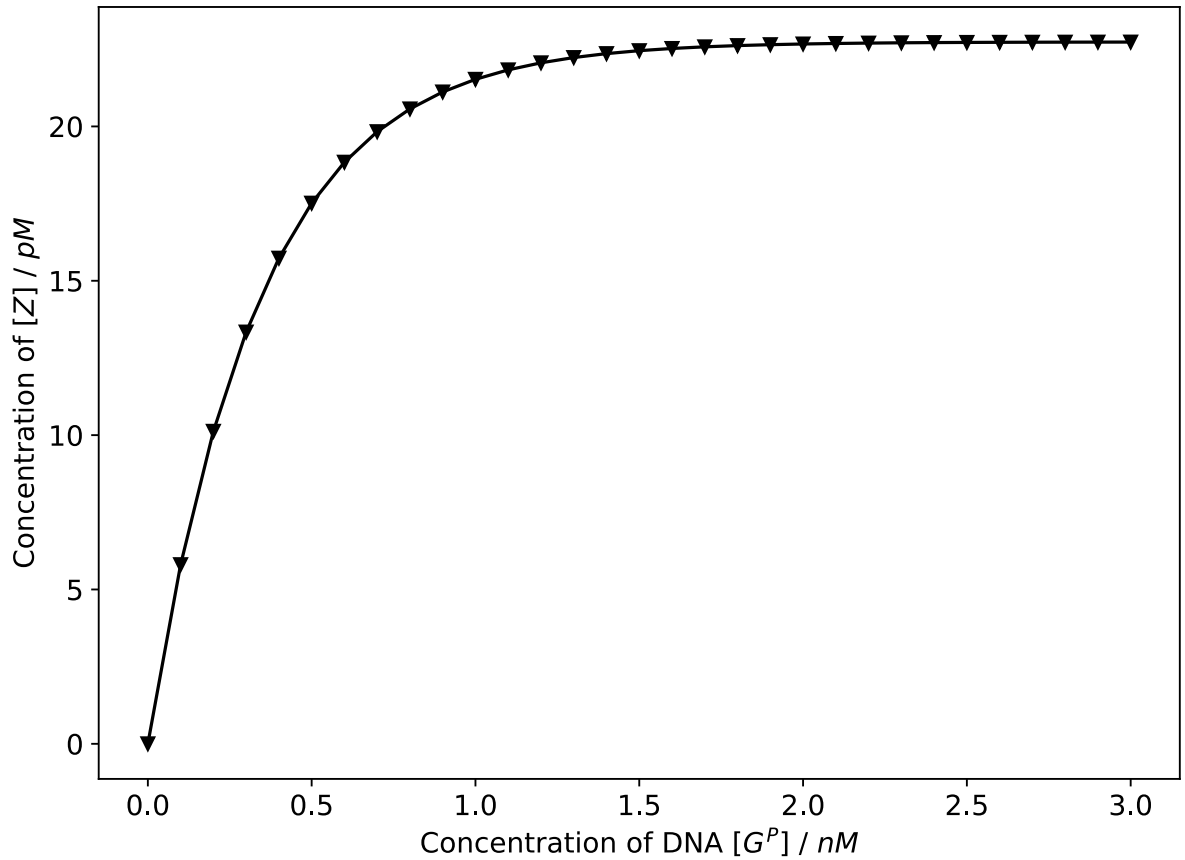
Figure 3.10: The concentration behavior of Model-L towards various DNA inputs.

The maximum concentration of target protein achievable in Model-L is ~25 pM and this value is reached at critical DNA input of ~2 nM. Increasing the DNA input may not be always suitable if the cost associated with the gene manipulation technique is high. The possible alternative is to increase the batch time/reaction time t_b. The product yield of Model-

L in the presence of different combinations of batch time and initial DNA input is given in Figure 3.11.
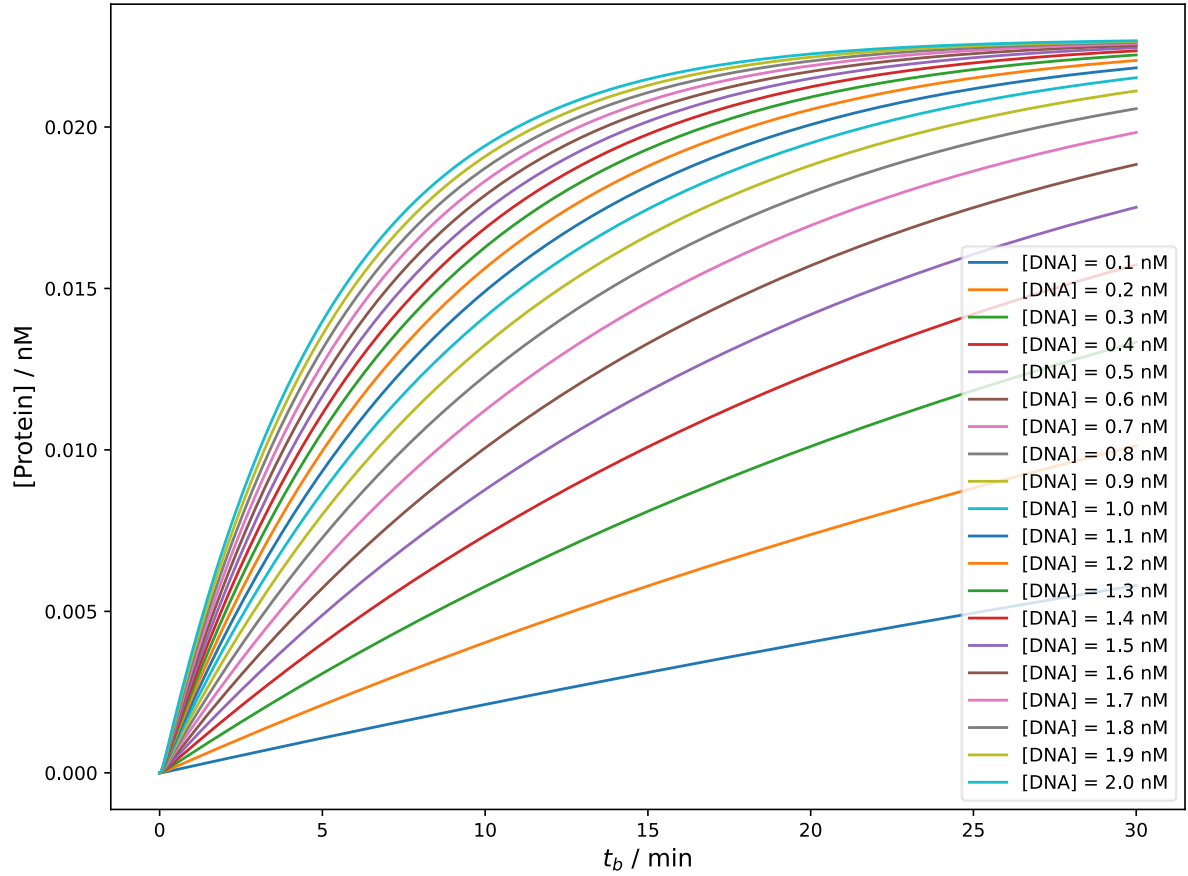
Figure 3.11: Target protein yield with different combinations of reaction time, t_b and DNA inputs.

From these results it is evident that increasing the initial DNA input also reduces the reaction time to achieve a fixed concentration of protein product (bend of the curve increasing with high initial DNA inputs). Therefore, fixing the reaction time and increasing DNA input results in an increase the product yield. For example, the Model-L (initial [DNA] = 1 nM), for a fixed batch time of 15 min give the target protein concentration of ~16 pM, while a CFPS

system with DNA input of 2nM can give ~21 pM of target protein (31.25% increase). This percent increase will become more prominent for smaller batch times (i.e., improved synthesis time) when DNA input is increased.

The main goal of CFPS is to produce the desired protein concentration at high yield within the short amount of synthesis time.[1,22,24] This behavior is not possible sometimes due to the theoretical limits imposed on the CFPS systems such as protein solubility (Section 3.2). Therefore, at some instances, the researchers get to specify the theoretical protein concentration attainable for different Model-L settings, and Figure 3.11 will show the time to achieve that specific concentration. For example, if the maximum soluble protein concentration for Model-L is 10 pM at 0.5 nM DNA input, then the time required to achieve this value is ~13 min. But increasing the DNA input to 1 nM (assuming no change in protein solubility due to increasing DNA input), the time required to achieve this value reduces to ~5 min (reduction in synthesis time is ~61.5%). This information is useful for scientists and synthetic biologists to come up with an economically feasible CFPS systems. The R code for dose-response behavior is given in APPENDIX 1.

3.3 Effect of resource pools on CFPS

According to this research, it is assumed that the required resources for CFPS systems are only NTPs and AAs. To test these effects, the resource pool values of Model-L were changed congruently (i.e., the change d[AA] = d[NTP]). The simulation results show that, in the presence of large resource pools, the system would respond better to the initial DNA inputs.

The Model-L results for the response towards DNA inputs is given in Table 3.2 and the behavior is given pictorially in Figure 3.12.

Table 3.2: Simulated data of DNA inputs in various resource pool concentrations

[Resource Pool]/nM →	2	4	6	8	10
[DNA]/nM ↓	[Z]/nM	[Z]/nM	[Z]/nM	[Z]/nM	[Z]/nM
0.1	5.80 E-3	2.14 E-2	4.47 E-2	7.39 E-2	1.08 E-1
0.2	1.01E-2	3.74 E-2	7.82 E-2	1.30 E-1	1.90 E-1
0.3	1.33 E-2	4.93 E-2	1.03 E-1	1.71 E-1	2.51 E-1
0.4	1.57 E-2	5.82 E-2	1.22 E-1	2.03 E-1	2.97 E-1
0.5	1.75 E-2	6.49 E-2	1.36 E-1	2.26 E-1	3.32 E-1
0.6	1.88 E-2	6.98 E-2	1.46 E-1	2.44 E-1	3.58 E-1
0.7	1.98 E-2	7.35 E-2	1.54 E-1	2.57 E-1	3.78 E-1
0.8	2.06 E-2	7.63 E-2	1.60 E-1	2.67 E-1	3.92 E-1

0.9	2.11 E-2	7.83 E-2	1.64 E-1	2.74 E-1	4.03 E-1
1.0	2.15 E-2	7.99 E-2	1.68 E-1	2.80 E-1	4.11 E-1
1.1	2.18 E-2	8.10 E-2	1.70 E-1	2.84 E-1	4.17 E-1
1.2	2.21 E-2	8.19 E-2	1.72 E-1	2.87 E-1	4.22 E-1
1.3	2.22 E-2	8.25 E-2	1.73 E-1	2.89 E-1	4.25 E-1
1.4	2.24 E-2	8.30 E-2	1.74 E-1	2.91 E-1	4.28 E-1
1.5	2.25 E-2	8.33 E-2	1.75 E-1	2.92 E-1	4.30 E-1
1.6	2.25 E-2	8.36 E-2	1.76 E-1	2.93 E-1	4.31 E-1
1.7	2.26 E-2	8.38 E-2	1.76 E-1	2.94 E-1	4.32 E-1
1.8	2.26 E-2	8.39 E-2	1.76 E-1	2.94 E-1	4.33 E-1
1.9	2.27 E-2	8.41 E-2	1.76 E-1	2.94 E-1	4.34 E-1
2.0	2.27 E-2	8.41 E-2	1.77 E-1	2.95 E-1	4.34 E-1
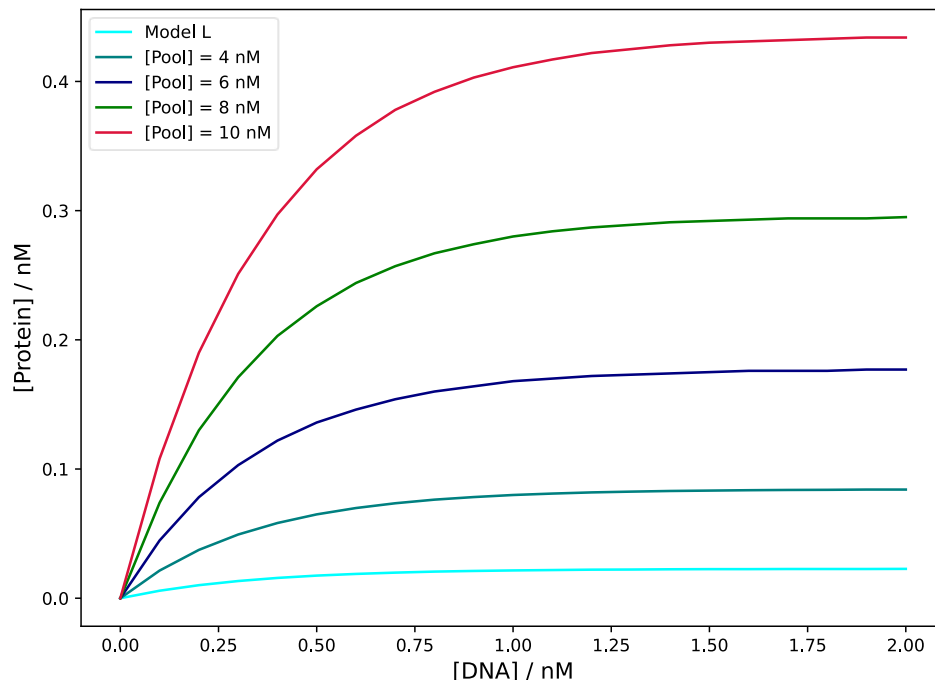
Figure 3.12: Model-L behavior towards DNA inputs under various resource concentrations. Note that [NTP + AA] are shown in the legend.

The maximum concentration of target protein achievable by increasing the size of resource pools for Model-L ranges from ~22 pM – 0.4nM when the pool size is increased from 2nM – 10nM (it should be noted [NTP] + [AA] are given as total resource pool). This phenomenal increase may convince researchers to manipulate the size of resource pools more than any other parameters available for manipulation. But it should be noted that large resource pools inherit large ionic strength into the CFPS extract and can alter the solubility of the target protein. The complete R code to reproduce these results are given in APPENDIX 1.

The effect of resource pools on the RNA and protein concentration of Model-L is shown in Table 3.3. The behavior of Model-L RNA towards resource pool is shown in Figure 3.13.

Table 3.3: Simulated data of Model-L, RNA and Protein concentrations		
[Pool]/nM	[Z]/nM	[M]/pM
0.2	2.32 E-4	2.90 E-2
0.4	9.19 E-4	5.80 E-2
0.6	2.05 E-3	8.72 E-2
0.8	3.62 E-3	1.16 E-1
1.0	5.60 E-3	1.46 E-1
1.2	8.00 E-3	1.75 E-1
1.4	1.08 E-2	2.05 E-1
1.6	1.40 E-2	2.34 E-1
1.8	1.76 E-2	2.64 E-1
2.0	2.15 E-2	2.94 E-1
2.2	2.58 E-2	3.23 E-1
2.4	3.05 E-2	3.53 E-1
2.6	3.55 E-2	3.83 E-1
2.8	4.09 E-2	4.13 E-1

3.0	4.66 E-2	4.43 E-1
3.2	5.26 E-2	4.74 E-1
3.4	5.90 E-2	5.04 E-1
3.6	6.56 E-2	5.35 E-1
3.8	7.26 E-2	5.65 E-1
4.0	7.99 E-2	5.96 E-1

From the simulations, it is evident that the RNA concentration of the CFPS system increases linearly with the gradient of ~0.00029. This indicates that the Model-L suffers from severe RNA degradation (i.e., 25% increase in [pool] yields only ~0.03% in [RNA], Equation 2.5). Further, the simulation results for target protein (Z) concentration in the presence of varying resource pool sizes show that for very small pool concentrations, there are no quantifiable production of proteins (from ~ 0 – 0.1 nM increase in pool concentration show no protein production from the CFPS systems). After this threshold the protein concentration increases linearly with the gradient of ~0.058. This buffering behavior is shown in Figure 3.14.
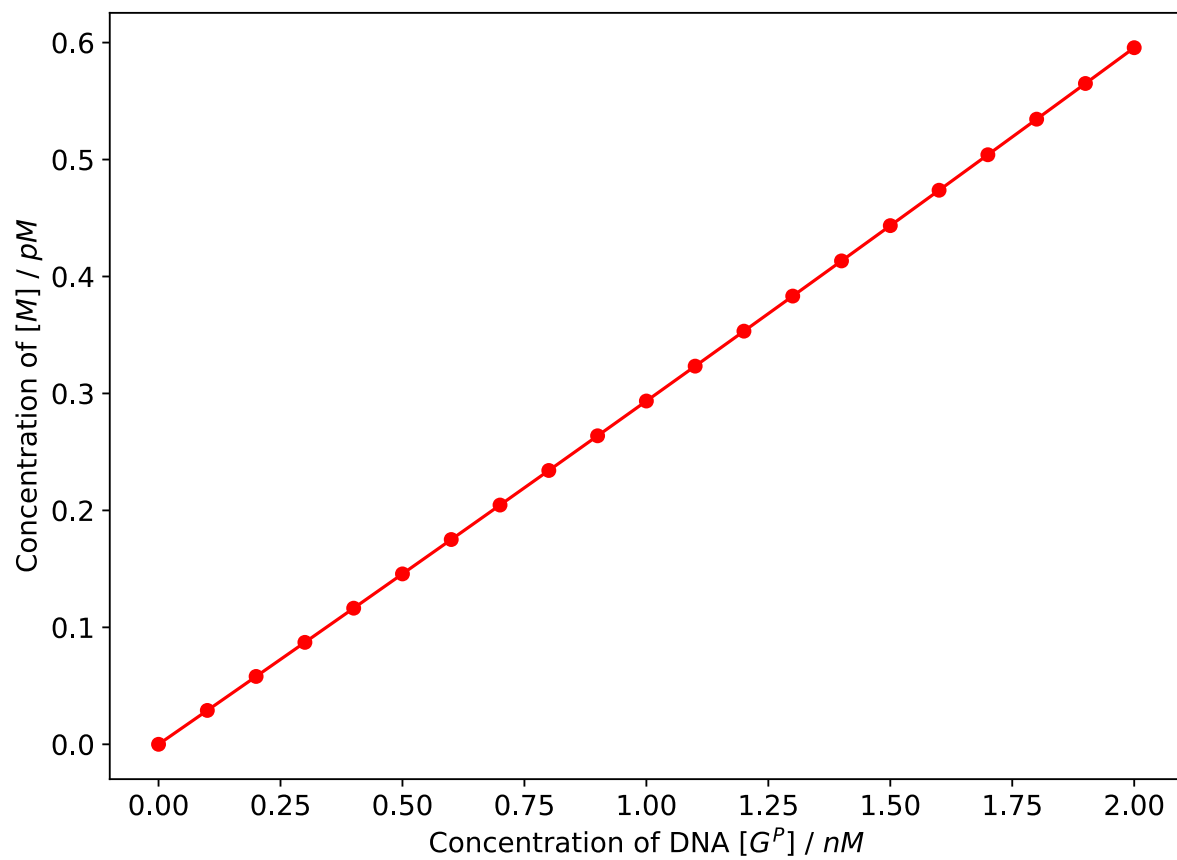
Figure 3.13: Model-L RNA concentration in various DNA inputs
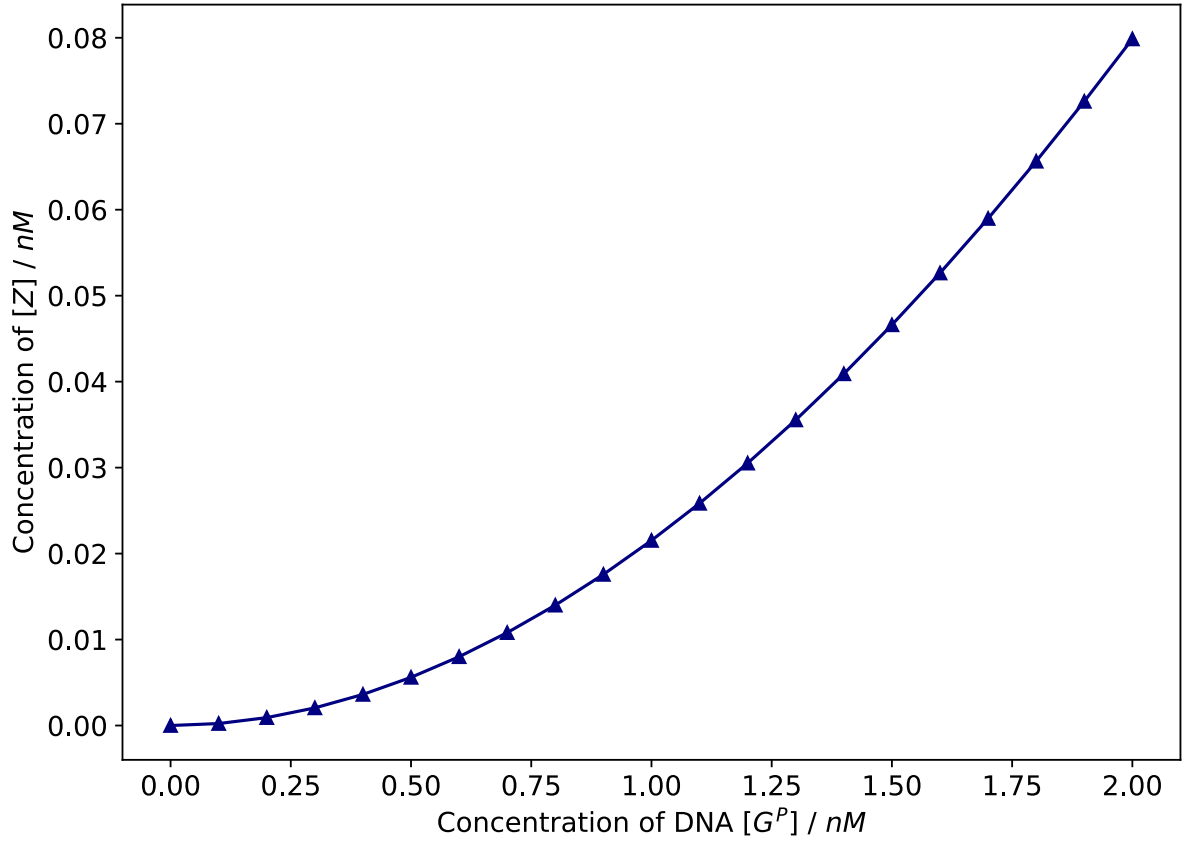
Figure 3.14: Model-L target protein concentration at various DNA inputs.

3.4 Effect of RNA chain degradation

The indirect effect of RNA degradation on resource depletion is discussed in Section 3.1 (Figure 3.5). Here, the magnitude of their effects is analyzed through simulation. The results show that for Model-L, if the RNase activity is removed (i.e., by setting the k_m value to be zero in Equation 2.5), the maximum achievable product yield reaches 1nM. Recall that, Model-L previously yielded only ~22 - 23 pM of target protein when $k_m = 18$ min^{-1} (but now, it is ~4500% increase). The effects of various k_m values on protein concentration are simulated and given in Table 3.4.

k_m (min^{-1})	[Z]/nM
Table 3.4: Simulated data of Model-L final protein concentrations on various degradation rate constant	
0.00	1.00
1.00	3.25 E-1
2.00	1.78 E-1
3.00	1.23 E-1
4.00	9.34 E-2
5.00	7.54 E-2
6.00	6.32 E-2
7.00	5.44 E-2
8.00	4.78 E-2
9.00	4.26 E-2
10.00	3.84 E-2
11.00	3.50 E-2
12.00	3.21 E-2

13.00	2.97 E-2
14.00	2.76 E-2
15.00	2.58 E-2
16.00	2.42 E-2
17.00	2.28 E-2
18.00	2.15 E-2
19.00	2.04 E-2
20.00	1.94 E-2

If the activity increases, the ability to produce target proteins diminishes in an exponential fashion.

Effect of RNA chain also follows this same behavior but more "victimized" since they are the substrates for RNases (as the k_m value approaches zero the RNA concentration increases ~200 times from initial settings). The simulated data in the presence of various RNA degradation rate is given in Table 3.5.

Table 3.5: Simulated data of Model-L RNA for various RNA degradation rates

k_m (min^{-1})	[Z]/nM

0.0	9.08 E-1
0.1	1.55 E-1
0.2	5.03 E-2
0.3	2.64 E-2
0.4	1.76 E-2
0.5	1.32 E-2
0.6	1.05 E-2
0.7	8.79 E-2
0.8	7.53 E-3
0.9	6.58 E-3
1.0	5.85 E-3
1.1	5.26 E-3
1.2	4.78 E-3
1.3	4.39 E-3
1.4	4.05 E-3
1.5	3.76 E-3

1.6	3.51 E-3
1.7	3.29 E-3
1.8	3.09 E-3
1.9	2.92 E-3
2.0	2.77 E-3

The response of Model-L towards DNA inputs in the presence of varying RNA degradation rates, was simulated and their results given in Table 3.5. The effect of RNA degradation on both species are shown in Figure 3.15. The complete R code to reproduce the behavior on both RNA and target protein is given in APPENDIX 1.
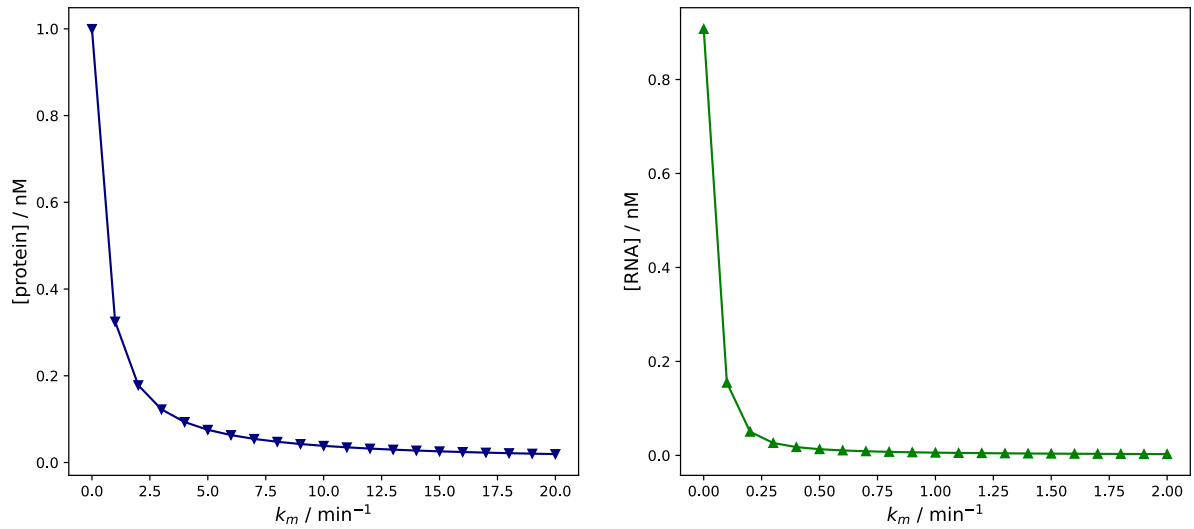
Figure 3.15: Effect of RNA degradation on Model-L, RNA and protein pools.

The simulations show that the response toward DNA inputs is "desensitized" in the presence of increased RNase activities (Table 3.6). These results are shown in Figure 3.16.

k_m/(min^{-1}) →	0	9	18	27	36
[DNA]/nM ↓	[Z]/nM	[Z]/nM	[Z]/nM	[Z]/nM	[Z]/nM
0.1	8.04 E-1	1.16 E-2	5.80 E-3	3.87 E-3	2.91 E-3
0.2	9.50 E-1	2.01 E-2	1.01 E-2	6.76 E-3	5.08 E-3
0.3	9.84 E-1	2.65 E-2	1.33 E-2	8.92 E-3	6.69 E-3
0.4	9.94 E-1	3.12 E-2	1.57 E-2	1.05 E-2	7.90 E-3
0.5	9.97 E-1	3.47 E-2	1.75 E-2	1.17 E-2	8.80 E-3
0.6	9.99 E-1	3.73 E-2	1.88 E-2	1.26 E-2	9.47 E-3
0.7	9.99 E-1	3.93 E-2	1.98 E-2	1.33 E-2	9.96 E-3
0.8	9.99 E-1	4.07 E-2	2.06 E-2	1.38 E-2	1.03 E-2
0.9	1.00	4.18 E-2	2.11 E-2	1.41 E-2	1.06 E-2

Table 3.6: Simulation data of DNA inputs at various RNA degradation rates

1.0	1.00	4.26 E-2	2.15 E-2	1.44 E-2	1.08 E-2
1.1	1.00	4.32 E-2	2.18 E-2	1.46 E-2	1.10 E-2
1.2	1.00	4.37 E-2	2.21 E-2	1.48 E-2	1.11 E-2
1.3	1.00	4.40 E-2	2.22 E-2	1.49 E-2	1.12 E-2
1.4	1.00	4.42 E-2	2.24 E-2	1.50 E-2	1.12 E-2
1.5	1.00	4.44 E-2	2.25 E-2	1.50 E-2	1.13 E-2
1.6	1.00	4.46 E-2	2.25 E-2	1.51 E-2	1.13 E-2
1.7	1.00	4.47 E-2	2.26 E-2	1.51 E-2	1.14 E-2
1.8	1.00	4.47 E-2	2.26 E-2	1.51 E-2	1.14 E-2
1.9	1.00	4.8 E-2	2.26 E-2	1.51 E-2	1.14 E-2
2.0	1.00	4.9 E-2	2.27 E-2	1.52 E-2	1.14 E-2
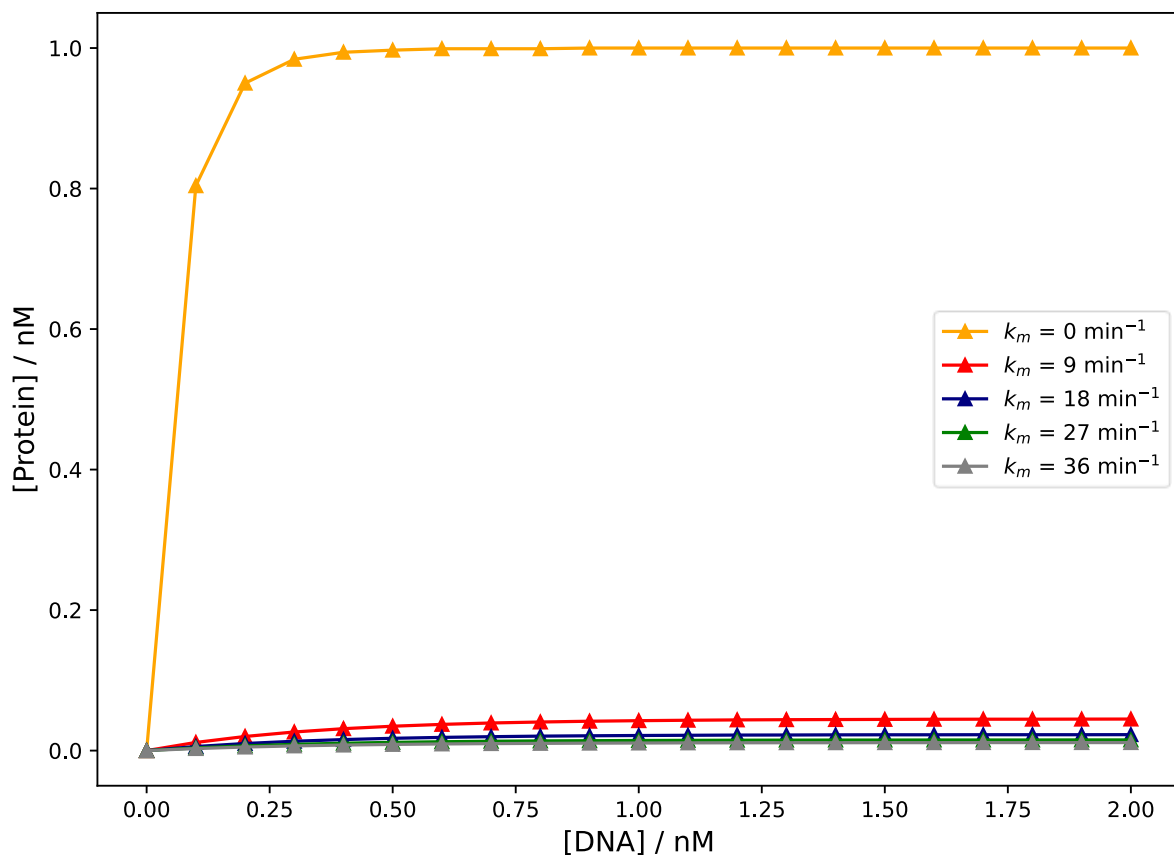
Figure 3.16: Model-L behavior toward DNA inputs in the presence of RNA degradation.

3.5 Effect of RNAP & ribosome concentrations

Increasing the RNAP concentration of Model-L shows no effect on target protein production (data not shown but can be reproduced from the codes given in APPENDIX 1). This can be correlated to the fact that the model contains very little RNA throughout the reaction process (maximum concentration of only 5 pM is achievable) and increasing the RNAP will have very little effect on the product yield. However, increasing the ribosome concentration increased the protein production yield linearly with the gradient of ~0.022. The simulation data for this are shown in Table 3.7 and the data is shown Figure 3.17.

Table 3.7: Simulated data of Model-L ribosome concentration versus product yield	
[R]/nM	[Z]/nM
0.1	2.17 E-3
0.2	4.34 E-3
0.3	6.50 E-3
0.4	8.66 E-3
0.5	1.08 E-2
0.6	1.30 E-2
0.7	1.51 E-2
0.8	1.73 E-2
0.9	1.94 E-2
1.0	2.15 E-2
1.1	2.37 E-2
1.2	2.58 E-2
1.3	2.79 E-2
1.4	3.00 E-2

1.5	3.21 E-2
1.6	3.42 E-2
1.7	3.63 E-2
1.8	3.84 E-2
1.9	4.05 E-2
2.0	4.26 E-2
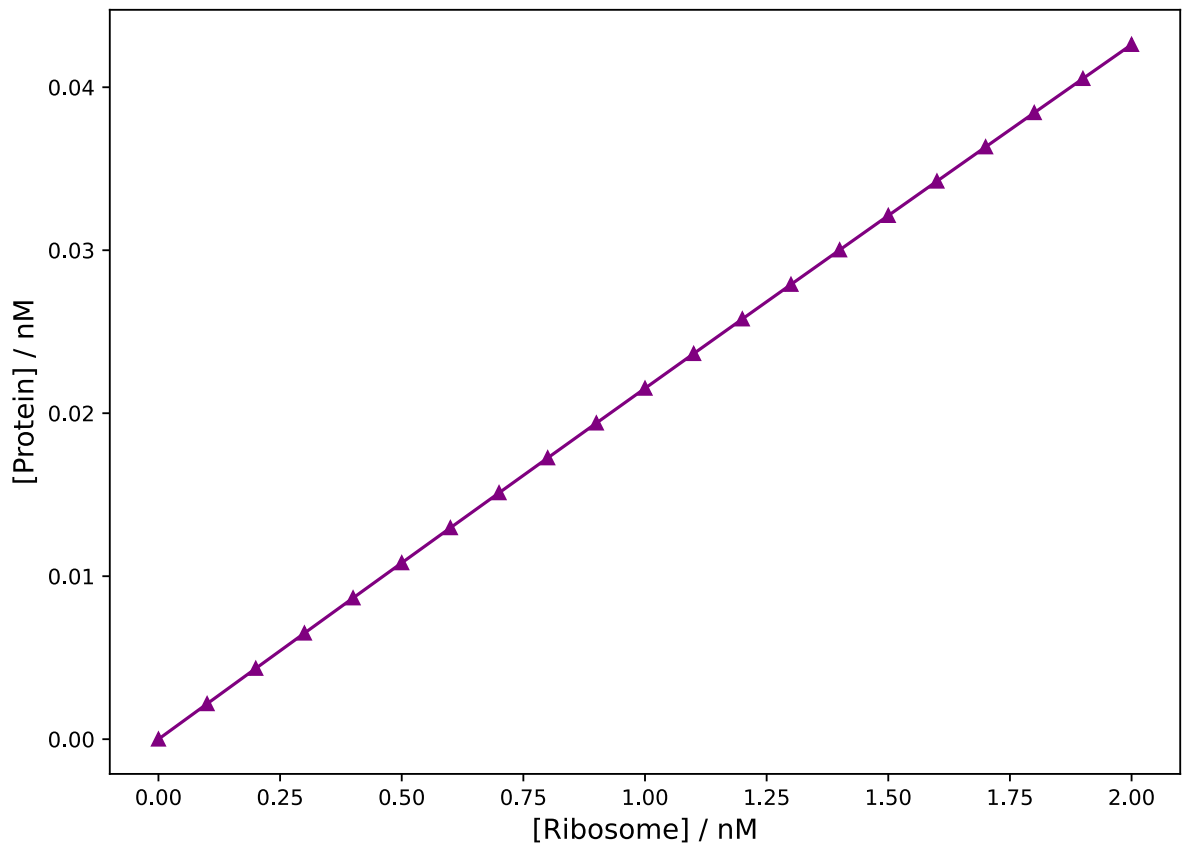
Figure 3.17: Model-L behavior towards the changes in ribosome concentration.

CHAPTER 4
CONCLUSION

From the simulation results, it is evident that CFPS systems can be severely affected by the degradative enzymes that are inevitably transferred into the CFPS extract during preparation. The resources also must be supplied constantly to continual production of the desired product at desired concentrations. The NTP pools are expensive molecules and potential scaling up might be a huge risk in investment cost. The developed model can act as a model framework for certain communities that are interested in metabolic circuit engineering, prototyping biochemical reactions, and biomanufacturing certain pharmaceuticals since it is necessary to develop models based on constrained biochemical parameters. Even though changes in certain parameter can be simulated the practicality of that physically achieving that change is questionable.

The results obtained from simple models (such as Model-L) by simulation may not capture the limitations invoked by chemistry and physics such as the protein solubility discussed previously. Therefore, the model should incorporate parameters such as pH, isoelectric pH, reaction temperature, batch volume, the type of microbial culture used for extract preparation and so on to scale up these systems.

CHAPTER 5

FUTURE WORKS

The possibilities to extend CFPS to a large scale is still a discouraging step for investors and pharmaceutical industries to produce drugs or therapeutic biological structures due to its small volumes and requirement for expensive resources. However, its simplicity, and open nature allows some freedom to mitigate certain undesired behaviors. For example, addition of RNase inhibitors can improve the CFPS extract performance (which is proven by simulation in this work) is promising. Further engineering the extract with certain metabolic reactions to produce the required resources from cheap simple molecules is also another approach to overcome the hurdles imposed by these CFPS methods.

REFERENCES

(1) Burrington, L. R.; Watts, K. R.; Oza, J. P. Characterizing and Improving Reaction Times for E. Coli-Based Cell-Free Protein Synthesis. *ACS Synth. Biol.* **2021**, *10* (8), 1821–1829. https://doi.org/10.1021/acssynbio.1c00195.

(2) Marshall, R.; Noireaux, V. Quantitative Modeling of Transcription and Translation of an All-E. Coli Cell-Free System. *Sci. Rep.* **2019**, *9* (1), 11980. https://doi.org/10.1038/s41598-019-48468-8.

(3) Kierzek, A. M.; Zaim, J.; Zielenkiewicz, P. The Effect of Transcription and Translation Initiation Frequencies on the Stochastic Fluctuations in Prokaryotic Gene Expression. *J. Biol. Chem.* **2001**, *276* (11), 8165–8172. https://doi.org/10.1074/jbc.M006264200.

(4) Darzacq, X.; Shav-Tal, Y.; de Turris, V.; Brody, Y.; Shenoy, S. M.; Phair, R. D.; Singer, R. H. In Vivo Dynamics of RNA Polymerase II Transcription. *Nat. Struct. Mol. Biol.* **2007**, *14* (9), 796–806. https://doi.org/10.1038/nsmb1280.

(5) Soetaert, K.; Petzoldt, T.; Setzer, R. W. Solving Differential Equations in R: Package deSolve. Journal of Statistical Software 2010, 33 (9), 1 - 25. DOI: 10.18637/jss.v033.i09 (acccessed 2022/06/26).

(6) Brown, T. A.; ProQuest. Gene cloning and DNA analysis : an introduction; Wiley Blackwell, 2016.

(7) Ferrer-Miralles, N.; Domingo-Espín, J.; Corchero, J. L.; Vázquez, E.; Villaverde, A. Microbial factories for recombinant pharmaceuticals. Microb Cell Fact 2009, 8, 17. DOI: 10.1186/1475-2859-8-17.

(8) Ingalls, B. P.; ProQuest. Mathematical modeling in systems biology : an introduction; MIT Press, 2013.

(9) Sanchez-Garcia, L.; Martín, L.; Mangues, R.; Ferrer-Miralles, N.; Vázquez, E.; Villaverde, A. Recombinant pharmaceuticals from microbial cells: a 2015 update. Microb Cell Fact 2016, 15, 33. DOI: 10.1186/s12934-016-0437-3.

(10) Raven, P.; Johnson, G.; Singer, S.; Mason, K.; Losos, J. Biology, 11th ed.; McGraw-Hill Education: New York, NY, 2017.

(11) Tymoczko, J. L.; Berg, J. M. Biochemistry: A Short Course, 3rd ed.; W.H. Freeman: New York, NY, 2015.

(12) Garrett, R. H.; Grisham, C. M. Biochemistry, 6th ed.; Cengage Learning: Boston, MA, 2017.

(13) Podolsky, K. A.; Devaraj, N. K. Synthesis of Lipid Membranes for Artificial Cells. Nat. Rev. Chem. 2021, 5 (10), 676–694. https://doi.org/10.1038/s41570-021-00303-3.

(14) R Core Team. R: A Language and Environment for Statistical Computing; 2021.

(15) Stamatakis, M.; Mantzaris, N. V. Comparison of Deterministic and Stochastic Models of the Lac Operon Genetic Network. *Biophys. J.* **2009**, *96* (3), 887–906. https://doi.org/10.1016/j.bpj.2008.10.028.

(16) Georgi, V.; Georgi, L.; Blechert, M.; Bergmeister, M.; Zwanzig, M.; Wüstenhagen, D. A.; Bier, F. F.; Jung, E.; Kubick, S. On-Chip Automation of Cell-Free Protein Synthesis: New Opportunities Due to a Novel Reaction Mode. *Lab Chip* **2016**, *16* (2), 269–281. https://doi.org/10.1039/c5lc00700c.

(17) Cui, Z.; Johnston, W. A.; Alexandrov, K. Cell-Free Approach for Non-Canonical Amino Acids Incorporation into Polypeptides. *Front. Bioeng. Biotechnol.* **2020**, *8*, 1031. https://doi.org/10.3389/fbioe.2020.01031.

(18) Müller, J.; Siemann-Herzberg, M.; Takors, R. Modeling Cell-Free Protein Synthesis Systems-Approaches and Applications. *Front. Bioeng. Biotechnol.* **2020**, *8*, 584178. https://doi.org/10.3389/fbioe.2020.584178.

(19) Mavelli, F.; Marangoni, R.; Stano, P. A Simple Protein Synthesis Model for the PURE System Operation. *Bull. Math. Biol.* **2015**, *77* (6), 1185–1212. https://doi.org/10.1007/s11538-015-0082-8.

(20) Stögbauer, T.; Windhager, L.; Zimmer, R.; Rädler, J. O. Experiment and Mathematical Modeling of Gene Expression Dynamics in a Cell-Free System. *Integr. Biol. (Camb.)* **2012**, *4* (5), 494–501. https://doi.org/10.1039/c2ib00102k.

(21) Rustad, M.; Eastlund, A.; Jardine, P.; Noireaux, V. Cell-Free TXTL Synthesis of Infectious Bacteriophage T4 in a Single Test Tube Reaction. *Synth. Biol.* **2018**, *3* (1). https://doi.org/10.1093/synbio/ysy002.

(22) Vilkhovoy, M.; Adhikari, A.; Vadhin, S.; Varner, J. D. The Evolution of Cell Free Biomanufacturing. *Processes (Basel)* **2020**, *8* (6), 675. https://doi.org/10.3390/pr8060675.

(23) Kopniczky, M. B.; Canavan, C.; McClymont, D. W.; Crone, M. A.; Suckling, L.; Goetzmann, B.; Siciliano, V.; MacDonald, J. T.; Jensen, K.; Freemont, P. S. Cell-Free Protein Synthesis as a Prototyping Platform for Mammalian Synthetic Biology. *ACS Synth. Biol.* **2020**, *9* (1), 144–156. https://doi.org/10.1021/acssynbio.9b00437.

(24) Li, J.; Gu, L.; Aach, J.; Church, G. M. Improved Cell-Free RNA and Protein Synthesis System. *PLoS One* **2014**, *9* (9), e106232. https://doi.org/10.1371/journal.pone.0106232.

(25) Ayoubi-Joshaghani, M. H.; Dianat-Moghadam, H.; Seidi, K.; Jahanban-Esfahalan, A.; Zare, P.; Jahanban-Esfahlan, R. Cell-Free Protein Synthesis: The Transition from Batch Reactions to Minimal Cells and Microfluidic Devices. *Biotechnol. Bioeng.* **2020**, *117* (4), 1204–1229. https://doi.org/10.1002/bit.27248.

(26) Tran, K.; Gurramkonda, C.; Cooper, M. A.; Pilli, M.; Taris, J. E.; Selock, N.; Han, T.-C.; Tolosa, M.; Zuber, A.; Peñalber-Johnstone, C.; Dinkins, C.; Pezeshk, N.; Kostov, Y.; Frey, D. D.; Tolosa, L.; Wood, D. W.; Rao, G. Cell-Free Production of a Therapeutic Protein: Expression, Purification, and Characterization of Recombinant Streptokinase Using a CHO Lysate. *Biotechnol. Bioeng.* **2018**, *115* (1), 92–102. https://doi.org/10.1002/bit.26439.

(27) Koch, M.; Faulon, J.-L.; Borkowski, O. Models for Cell-Free Synthetic Biology: Make Prototyping Easier, Better, and Faster. *Front. Bioeng. Biotechnol.* **2018**, *6*, 182. https://doi.org/10.3389/fbioe.2018.00182.

(28) Doerr, A.; de Reus, E.; van Nies, P.; van der Haar, M.; Wei, K.; Kattan, J.; Wahl, A.; Danelon, C. Modelling Cell-Free RNA and Protein Synthesis with Minimal Systems. *Phys. Biol.* **2019**, *16* (2), 025001. https://doi.org/10.1088/1478-3975/aaf33d.

(29) Provost, P.; Dishart, D.; Doucet, J.; Frendewey, D.; Samuelsson, B.; Rådmark, O. Ribonuclease Activity and RNA Binding of Recombinant Human Dicer. *EMBO J.* **2002**, *21* (21), 5864–5874. https://doi.org/10.1093/emboj/cdf578.

(30) Kamm, R. C.; Smith, A. G. Ribonuclease Activity in Human Plasma. *Clin. Biochem.* **1972**, *5* (4), 198–200. https://doi.org/10.1016/s0009-9120(72)80033-x.

(31) Dondapati, S. K.; Stech, M.; Zemella, A.; Kubick, S. Cell-Free Protein Synthesis: A Promising Option for Future Drug Development. *BioDrugs* **2020**, *34* (3), 327–348. https://doi.org/10.1007/s40259-020-00417-y.

(32) Kattan, J.; Doerr, A.; Dogterom, M.; Danelon, C. Shaping Liposomes by Cell-Free Expressed Bacterial Microtubules. *ACS Synth. Biol.* **2021**, *10* (10), 2447–2455. https://doi.org/10.1021/acssynbio.1c00278.

(33) Stech, M.; Quast, R. B.; Sachse, R.; Schulze, C.; Wüstenhagen, D. A.; Kubick, S. A Continuous-Exchange Cell-Free Protein Synthesis System Based on Extracts from Cultured Insect Cells. *PLoS One* **2014**, *9* (5), e96635. https://doi.org/10.1371/journal.pone.0096635.

(34) Sawasaki, T.; Ogasawara, T.; Morishita, R.; Endo, Y. A Cell-Free Protein Synthesis System for High-Throughput Proteomics. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (23), 14652–14657. https://doi.org/10.1073/pnas.232580399.

(35) Terada, T.; Kusano, S.; Matsuda, T.; Shirouzu, M.; Yokoyama, S. Cell-Free Protein Production for Structural Biology. In *Springer Protocols Handbooks*; Springer Japan: Tokyo, 2016; pp 83–102.

(36) Cramer, P. Organization and Regulation of Gene Transcription. *Nature* **2019**, *573* (7772), 45–54. https://doi.org/10.1038/s41586-019-1517-4.

(37) Schuller, A. P.; Green, R. Roadblocks and Resolutions in Eukaryotic Translation. *Nat. Rev. Mol. Cell Biol.* **2018**, *19* (8), 526–541. https://doi.org/10.1038/s41580-018-0011-4.

(38) Zhao, B. S.; Roundtree, I. A.; He, C. Post-Transcriptional Gene Regulation by MRNA Modifications. *Nat. Rev. Mol. Cell Biol.* **2017**, *18* (1), 31–42. https://doi.org/10.1038/nrm.2016.132.

(39) Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J. Post-Translational Modification: Nature's Escape from Genetic Imprisonment and the Basis for Dynamic Information Encoding: Information Encoding by Post-Translational Modification. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2012**, *4* (6), 565–583. https://doi.org/10.1002/wsbm.1185.

(40) Yang, E.; van Nimwegen, E.; Zavolan, M.; Rajewsky, N.; Schroeder, M.; Magnasco, M.; Darnell, J. E., Jr. Decay Rates of Human MRNAs: Correlation with Functional Characteristics and Sequence Attributes. *Genome Res.* **2003**, *13* (8), 1863–1872. https://doi.org/10.1101/gr.1272403.

(41) Shin, J.; Noireaux, V. Study of Messenger RNA Inactivation and Protein Degradation in an Escherichia Coli Cell-Free Expression System. *J. Biol. Eng.* **2010**, *4* (1), 9. https://doi.org/10.1186/1754-1611-4-9.

(42) Ross, J.; Kobs, G. H4 Histone Messenger RNA Decay in Cell-Free Extracts Initiates at or near the 3′ Terminus and Proceeds 3′ to 5′. *J. Mol. Biol.* **1986**, *188* (4), 579–593. https://doi.org/10.1016/s0022-2836(86)80008-0.

(43) Gregorio, N. E.; Levine, M. Z.; Oza, J. P. A User's Guide to Cell-Free Protein Synthesis. *Methods Protoc.* **2019**, *2* (1), 24. https://doi.org/10.3390/mps2010024.

(44) Garenne, D.; Haines, M. C.; Romantseva, E. F.; Freemont, P.; Strychalski, E. A.; Noireaux, V. Cell-Free Gene Expression. *Nat Rev Methods Primers* **2021**, *1* (1), 1–18. https://doi.org/10.1038/s43586-021-00046-x.

(45) Voloshin, A. M.; Swartz, J. R. Efficient and Scalable Method for Scaling up Cell Free Protein Synthesis in Batch Mode. *Biotechnol. Bioeng.* **2005**, *91* (4), 516–521. https://doi.org/10.1002/bit.20528.

(46) Tinafar, A.; Jaenes, K.; Pardee, K. Synthetic Biology Goes Cell-Free. *BMC Biol.* **2019**, *17* (1), 64. https://doi.org/10.1186/s12915-019-0685-x.

(47) Thoring, L.; Dondapati, S. K.; Stech, M.; Wüstenhagen, D. A.; Kubick, S. High-Yield Production of "Difficult-to-Express" Proteins in a Continuous Exchange Cell-Free System Based on CHO Cell Lysates. *Sci. Rep.* **2017**, *7* (1), 11710. https://doi.org/10.1038/s41598-017-12188-8.

(48) Brookwell, A.; Oza, J. P.; Caschera, F. Biotechnology Applications of Cell-Free Expression Systems. *Life (Basel)* **2021**, *11* (12), 1367. https://doi.org/10.3390/life11121367.

APPENDICES

APPENDIX 1

Figure A.1: R code to reproduce the behavior of Model-L

```
                              initial_settings.R

library(deSolve)

state.variables <- c(gp = 1E-9, p0 = 1E-9, xs = 0, m0 = 0,
                     r0 = 1E-9, mr = 0, z0 = 0,
                     AA = 1E-9, NT = 1E-9)


parm.values <- c(k1 = 6E9, k2 = 600, k3 = 1E10,
                 k4 = 6E9, k5 = 135, k6 = 1E10,
                 k7 = 18)

con.profile <- function(t, state.variables, parm.values){
  with(as.list(c(state.variables, parm.values)),{

    dgp <- -k1*gp*p0 + k2*xs + k3*NT*xs
    dp0 <- -k1*gp*p0 + k2*xs + k3*NT*xs
    dxs <-  k1*gp*p0 - k2*xs - k3*NT*xs
    dm0 <-  k3*NT*xs - k4*m0*r0 + k5*mr + k6*AA*mr - k7*m0
    dr0 <- -k4*m0*r0 + k5*mr + k6*AA*mr
    dmr <-  k4*m0*r0 - k5*mr - k6*AA*mr
    dz0 <-  k6*AA*mr
    dAA <- -k6*AA*mr
    dNT <- -k3*NT*xs

    return(list(c(dgp, dp0, dxs, dm0, dr0, dmr, dz0,
                  dAA, dNT)))
  })
}

out <- ode(y = state.variables,
           parms = parm.values,
           method = "bdf",
           times = seq(0, 30, 0.001),
           func = con.profile)
```

Figure A.2: R code to reproduce DNA inputs in the presence of RNA degradation

```
                              Z0_vs_G0_km_change.R

library(deSolve)

km.range <- c(0, 9, 18, 27, 36)
g0.range <- seq(0, 2E-9, 1E-10)

df <- data.frame(km = numeric(),
                 g0 = numeric(),
                 z0 = numeric())

n <- 1

for (i in km.range) {
  for (j in g0.range) {

    init.values <- vector(mode = 'numeric')
    init.values <- append(init.values, c(i, j))

    state.variables <- c(gp = j, p0 = 1E-9, xs = 0, m0 = 0,
                         r0 = 1E-9, mr = 0, AA = 1E-9, NT = 1E-9,
                         z0 = 0)


    parm.values <- c(k1 = 6E9, k2 = 600, k3 = 1E10,
                     k4 = 6E9, k5 = 135, k6 = 1E10,
                     k7 = i)

    con.profile <- function(t, state.variables, parm.values){
      with(as.list(c(state.variables, parm.values)),{

        dgp <- -k1*gp*p0 + k2*xs + k3*NT*xs
        dp0 <- -k1*gp*p0 + k2*xs + k3*NT*xs
        dxs <-  k1*gp*p0 - k2*xs - k3*NT*xs
        dm0 <-  k3*NT*xs - k4*m0*r0 + k5*mr + k6*AA*mr - k7*m0
        dr0 <- -k4*m0*r0 + k5*mr + k6*AA*mr
        dmr <-  k4*m0*r0 - k5*mr - k6*AA*mr
        dAA <- -k6*AA*mr
        dNT <- -k3*NT*xs
        dz0 <-  k6*AA*mr

        return(list(c(dgp, dp0, dxs, dm0, dr0, dmr, dAA,
                      dNT, dz0)))
      })
    }

    out <- ode(y = state.variables,
               parms = parm.values,
               method = "bdf",
               times = seq(0, 30, 0.001),
               func = con.profile)

    zx.value <- unname(out[nrow(out), ncol(out)])

    init.values <- append(init.values, zx.value)

    df[n,] <- init.values

    n <- n + 1
  }

}
```

Figure A.3: R code to reproduce DNA input in the presence of varying resource pools

```
                                Z0_vs_G0_pool_change.R

library(deSolve)

pool.range <- c(1E-9, 2E-9, 3E-9, 4E-9, 5E-9)
g0.range <- seq(0, 2E-9, 1E-10)

df <- data.frame(pool = numeric(),
                 g0 = numeric(),
                 z0 = numeric())

n <- 1

for (i in pool.range) {
  for (j in g0.range) {

    init.values <- vector(mode = 'numeric')
    init.values <- append(init.values, c(i, j))

    state.variables <- c(gp = j, p0 = 1E-9, xs = 0, m0 = 0,
                         r0 = 1E-9, mr = 0, AA = i, NT = i,
                         z0 = 0)


    parm.values <- c(k1 = 6E9, k2 = 600, k3 = 1E10,
                     k4 = 6E9, k5 = 135, k6 = 1E10,
                     k7 = 18)

    con.profile <- function(t, state.variables, parm.values){
      with(as.list(c(state.variables, parm.values)),{

        dgp <- -k1*gp*p0 + k2*xs + k3*NT*xs
        dp0 <- -k1*gp*p0 + k2*xs + k3*NT*xs
        dxs <-  k1*gp*p0 - k2*xs - k3*NT*xs
        dm0 <-  k3*NT*xs - k4*m0*r0 + k5*mr + k6*AA*mr - k7*m0
        dr0 <- -k4*m0*r0 + k5*mr + k6*AA*mr
        dmr <-  k4*m0*r0 - k5*mr - k6*AA*mr
        dAA <- -k6*AA*mr
        dNT <- -k3*NT*xs
        dz0 <-  k6*AA*mr

        return(list(c(dgp, dp0, dxs, dm0, dr0, dmr, dAA,
                      dNT, dz0)))
      })
    }

    out <- ode(y = state.variables,
               parms = parm.values,
               method = "bdf",
               times = seq(0, 30, 0.001),
               func = con.profile)

    zx.value <- unname(out[nrow(out), ncol(out)])

    init.values <- append(init.values, zx.value)

    df[n,] <- init.values

    n <- n + 1
  }

}
```