CHEME7770 Final Report
Tina Ye (ty369) and Xiaojing Ma (xm89)
May 21, 2020

COVID-19 Protein Expression Analysis in Cell-Free System

Abstract

The SARS-Cov-2 virus outbreak that started in December, 2019 caused a global pandemic, leading a spectrum of COVID-19 disease from light cold-like symptoms to fatal infection. It raised an urge to understand the virus and develop effective therapy. A number of non-structural proteins has been identified with proposed sequences and functionalities. (1) In this project, a model is proposed to analyze the expression of nonstructural proteins, nsp1, nsp10, and RdRp. A sequence specific modeling is used to predict the optimal productivity of these COVID19 proteins under steady state and the modeling also reveals that the RNAP elongation rate has a high impact on the maximum optimal productivity. Besides, with parameters tuning, an Excel solver is used to perform analysis on nsp1 varies in time. Lastly, the inhibition of the ssrA-tag GFP's expression by nsp1 is discussed.

Introduction

Coronavirus disease 2019 (COVID-19) is a current pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which more or less affects everyone, including health issues, social interactions, and economics, in the USA and around the world. The outbreak started in Wuhan, China, and then has spread worldwide. As of 18 May 2020, there were more than four million confirmed cases and three hundred and sixteen thousand deaths worldwide. (2) It is a respiratory disease that can transmit through direct contact with patients or contaminated objects. Since now there is no medicine that has demonstrated efficacy in the treatment of SARS-CoV-2 virus. Even though several potential vaccines are under development, the availability for widespread use may take at least one year while vaccines are approved.

Due to the severe situation of COVID-19 globally, Olin Therapeutics in Cornell University are asked to use a commercial *Escherichia coli* (*E. coli)* TX-TL cell-free protein synthesis (CFPS) system to develop a synthetic gene expression system for COVID-19 genes. Our team aims to better understand certain proteins that play important roles in virus translation and transmission.

The cell free protein synthesis (CFPS) system is an important tool that enable genetic work without the use of living cells, which allows one to directly control translation and transcription. Cell-free system is used because of its versatility and broad application, such as producing antibodies, and therapeutic proteins, which enables better studies in areas such as gene circuit and high throughput protein. (3) CFPS system is also relatively economically friendly compares to cell cultures. The TX-TL process can be achieved by cell extracts from *E. coli*. (4) The commercial *E. coli* TX-TL CFPS system, in particular, based on PANOxSP system, is considered to develop a synthetic gene expression system for COVID-19 genes in this project.

A group of non-structural proteins (NSP) studied in this project contribute to viral replication and transcription. Specifically, RNA-dependent RNA polymerase (RdRp), nsp1, and nsp10 are investigated in this study. RdRp, also known as NSP12, plays an essential role in the replication and transcription cycle of COVID-19 virus. Thus, it is considered as a primary target for antiviral drugs. Nsp1 of coronavirus (CoVs) also has gotten people's attention as a potential virulence factor and a target for CoV vaccine development due to its role in the downregulation of host innate immune response to CoV infection. (5) The expression of NSP1 promotes degradation of expressed RNA transcripts and host endogenous mRNAs, which inhibit host protein synthesis. In addition, NSP10 is another protein with concerns as a cofactor, which can activate multiple replicative enzymes, such as NSP14 and NSP16 in coronavirus.

CHEME7770 Final Report
Tina Ye (ty369) and Xiaojing Ma (xm89)
May 21, 2020

Methodology

FBA Sequence Specific Modeling Construction

The coding template is kindly provided by Professor Jeffrey Varner research lab. Some of the key parameters such as T7 polymerase's transcription saturation coefficient, mRNA degradation rate constant, and σ_{70} concentration, dissociation constant, and hill coefficient, are taken from Michael Vilkhovoy et al's paper. [6] The protein synthesis is assumed mainly based on the PANOxSP system with slight modification. [6, 7] Details in parameters usage can refer to the actual code published on GitHub. (https://github.com/ty369/ChemE7770_Final_Project_Report). Amino acid sequence of the protein nsp1 and nsp10 are obtained from Yang Zhang research lab and sequence of RdRp is obtained from UniPort database. Full sequences and nucleotides composition are shown in Appendix A.

Fundamental reactions that are involved in the protein production are shown in Table 2 in Appendix B, with the first six reactions occurring within the boundary we defined and the last nine are fluxes that go in or out of the boundary. [8] Internal RNA polymerase, tRNA, and rRNA production is also considered, detailed reactions shown in Table 3, Appendix B. [8] Beside protein synthesis reactions and flux exchanges across boundary, essential metabolic pathways are considered which include: glycolysis, pentose phosphate pathway, entner-doudoroff pathway, TCA cycle, oxidative phosphorylation, overflow metabolism, amino acid formation and degradation, pyrimidine synthesis, and purine synthesis. Collecting all the reactions involved in the boundary and across the boundary, a matrix, S, with rows of metabolites and columns of reaction and flux are created. In this case, we have 146 rows and 265 columns. Since the number of columns is greater than the rank of S (146), there are more than one set of solutions to the matrix. Thus, bounds and restrictions are set for parameters to obtain a desired optimal result on the objective function that we are interested in. The objective function here is the maximum optimal productivity of the non-structural proteins.

Time-dependent Nsp1 and mRNA Expression Construction

The method to develop expression of nsp1 protein in unsteady state is based on literature. [9] After analyzing the results of western blot (WB) assay, which is widely used in molecular biology to detect specific proteins in a sample of tissue extract, of nsp1 in low and high expression, the relationship of NSP1 and time can be found.

To express the mRNA level varies in time, we collect information about the numbers of *E. coli* cells involved in mRNA/protein calculations from literature. [10] In order to find approximate numbers of mRNA, we start with the characteristic protein concentration and use the formulas below.

$$\frac{N_{protein}}{V} \; R = \frac{N_{protein}}{t} \; r = \frac{v}{d} \; N_{mRNA} = \frac{R}{r}$$

N_protein, V, R, t, r, v, d, N_mRNA stands for number of protein, volume, rate of protein production per cell, doubling time, rate of translation, distance between ribosomes, and number of mRNA respectively. In *E. coli* cells, the values of N_protein, V, t, v, d are $3*10^6$ proteins, $1um^3$, 1 hour, 10 to 50 bases/s, 30-300 bases respectively. After getting the numbers of mRNA, we calculate the percentage of reduction of mRNA by NSP1 based on values in literature. [11] Then the relationship of mRNA with NSP1 is found.

The Expression and Inhibition of GFP-ssrA Protein in CFPS

Method described in Garamella et al's paper is applied here to find the expression of GFP-ssrA protein in the cell- free system. [12] After analyzing points in figures, the relationship of GFP-

ssrA and time is found. The same influence of NSP1 on both native mRNA and GFP-ssrA's mRNA is assumed. After applying a percentage change (76%) of GFP-ssrA expression by NSP1, how NSP1 acts on the expression of GFP-ssrA is found.

Details calculations and actual data fitting is shown in Excel in the GitHub link mentioned earlier in this report.

Results and Discussion

Objective Defined

In this project, we have three main goals: First, investigate different constraints' effect on optimal productivity of RdRp, NSP1, and NSP10 in the cell free system. In this goal, flux balance analysis with sequence specific modeling is used to generate result of protein synthesis in cell free system. This goal helps us have a better understanding on the three non-structural protein and serve as a potential evaluation to generate massive amount of the non-structural protein efficiently for research study. After having some insight about factors that affect non-structural protein production under steady state, we aim to study how nsp1 protein are express and influence other non-COVID19 protein in a time dependent manner. Thus, our second goal is to propose modeling to reveal the expression of COVID19 nonstructural protein NSP1 with time dependency, and how it influences mRNA expressions; This goal aims to give insight about how long can the virus be active and its concentration in cells after infection. Furthermore, we would like to see how the accumulation of nsp1 in the environment affects other non-COVID19 protein, and we have out third goal: literature based modeling on the effects of NSP1 on the expression of GFP-ssrA protein.

FBA Sequence-Specific Modeling under Steady State

The modified model provides the maximum optimal productivity of the protein and it is based on promoter T7 for the simulation. First, we would like to see how the carbon numbers affect the optimal production of proteins where nsp1 has 872 carbon, RdRp has 704 carbons and nsp10 has 636 carbons. We run the simulation of the three under three different case scenarios and the data is shown in Fig.1. The plot shows maximum optimal productivity for the three proteins under three case scenarios including one: amino acid uptake supplied with de novo synthesis; two: amino acid supplied with no synthesis, and three: amino acid synthesis only. From the plot, it shows that nsp10 has the largest optimal productivity among the three proteins while it contains the least number of carbon. This indicates that optimal productivity is inversely proportional to carbon number of the protein. Also, the productivity stays the same

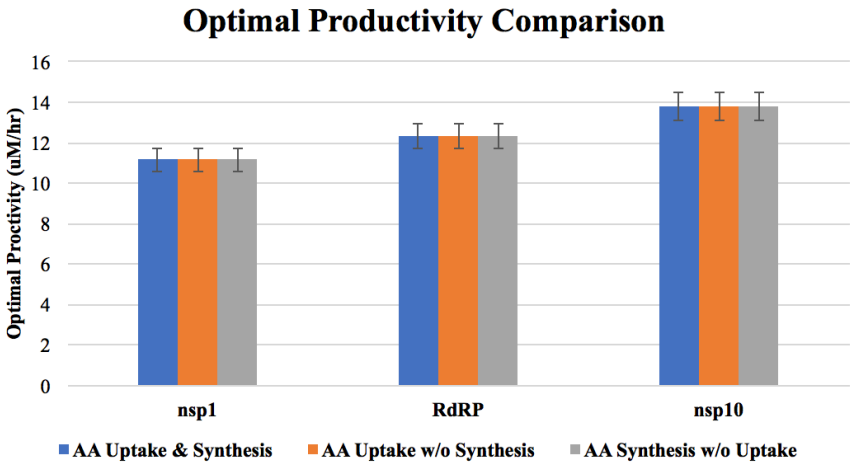


Figure. 1 Maximum optimal productivity comparison among three coronaviral proteins.

across different cases for each protein reveals that the change of amino supply or synthesis does not really affect the optimal productivity.

Next, from this FBA model, we want to see how different parameters affect our protein's productivity. The data is generated based on case 1 where both amino acid uptake and synthesis are present. Guided by the Vilkovoy et al, we test different values of the RNA polymerase (RNAP) elongation rate, ribosome elongation rate, RNAP's and ribosome's concentrations. We found that the elongation rate plays an important role in affecting productivity. Fig. 2 and Fig. 7 (App. C) shows optimal productivity versus the RNAP and ribosome elongation rate respectively. The result indicates that the generated maximum optimal productivity increases as the elongation rate increases. Similar plots for changing the concentration of RNAP and ribosome are also listed in Appendix C. Based on the slope of the generated trend lines, changing the concentration has relatively less influence on the productivity outcome, which matches the conclusion stated in Vilkovoy et al's paper from a much larger scale of global sensitivity analysis. [6] However, comparing the concentrations of ribosome and RNAP, RNAP level seems to have a larger effect on productivity. This may be caused by a difference of units for ribosome concentration are considered here compared to the original model. Another finding in the plots are the changing the elongation rate or the RNAP/ribosome concentration seems to have a greater effect on the productivity of nsp10 compares to the other two protein. This may imply that protein size has dominate effects compare to the altering the RNAP or ribosome condition. Also note that in this study, we consider mainly the expression of RdRp and for future study, effects of RdRp on the amplification of other non-structural proteins will be considered.

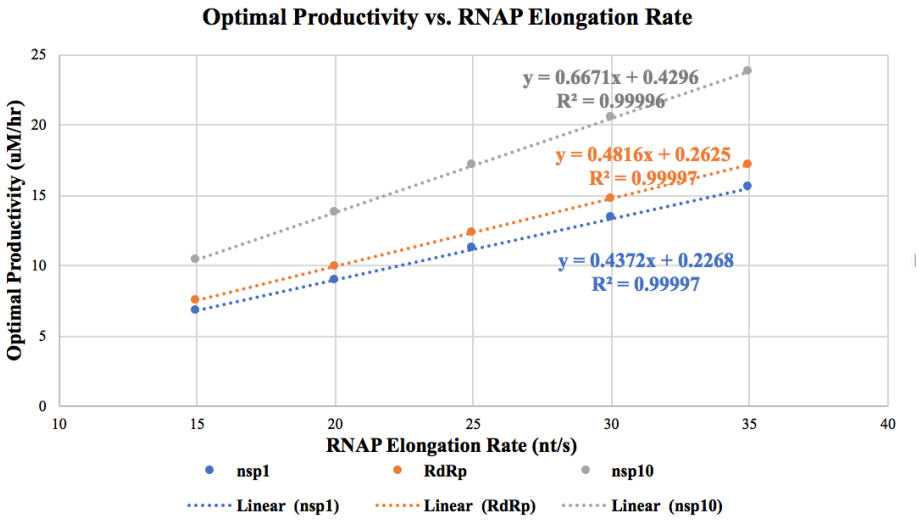


Figure. 2 Maximum optimal productivities of the three proteins under different RNAP elongation rate.

Nsp1 Expression with Time Dependency

As a precondition, COVID-19 is assumed to behave similar to other viruses, so the results from literature [9] analyzed by WB assays can be used for project analysis. After infecting normal cells, the levels of NSP1 expression have been measured at different times shown in Fig. 3. According to the figure, NSP1 with low expression has started to express after 9 hours post infection (hpi), and reached its high expression at 12 hpi, remaining high until 18 hpi. Nsp1 with high expression has started to express at 6 hpi, and remained high until 18 hpi. Thus, both high level and low level expressions of NSP1 increase at late times, telling NSP1 is a late viral protein.

We then are interested in knowing how the expression of mRNA is compared with the expression of nsp1 since nsp1 is known to inhibit transcription of mRNA. The condition of *E. coli* CFPS systems are considered due to the easy access to the information for *E. Coli*. After reading the values directly from the literature's figure, [10] numbers of *E. coli* cells per 100ml sample at different time point have been organized in Table 4 (App. D). In order to find approximate numbers of mRNA in the *E. coli* cell, the calculation started with the characteristic protein concentration has approached below.

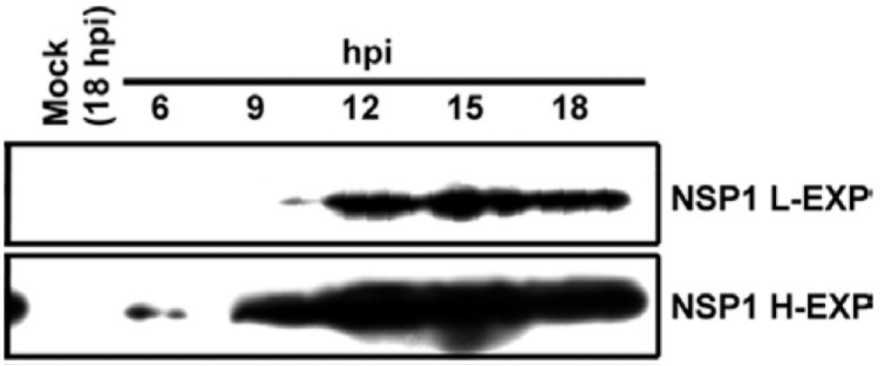


Figure. 3 The results of WB assay of NSP1 in low and high expression.

$$\frac{N_{protein}}{V} = 3\ X\ 10^6 \frac{proteins}{um^3}$$

$$R = \frac{N_{protein}}{t} = \frac{3\ X\ 10^6\ proteins}{3000\ s} = 10^3\ proteins/s$$

$$r = \frac{v}{d} = 0.1\sim1\ \text{protein/mRNA/s}$$

$$N_{mRNA} = \frac{R}{r} = \frac{10^3\ proteins/s}{0.1\sim1\ \text{protein/mRNA/}s} = 10^3\sim10^4\ mRNA/cell$$

For convenience, 1000 mRNA/cell has been chosen and we multiply it with the numbers of *E. coli* cells to get possible numbers of mRNA shown in Table 4. As mentioned in the introduction before, NSP1 expression inhibits host protein synthesis by degrading expression of mRNAs. From literature, [11] it has compared relative mRNA levels with and without NSP1. The results show that mRNA levels can reach 300 without NSP1, but decrease to 70 with NSP1. The percentage change, about 76%, is assumed to be the same as the project's. Thus, the numbers of mRNA with NSP1 can be calculated and recorded in Table 4. To understand better about how NSP1 influences the expression of mRNA, mRNA concentration versus time is plotted and shown in Fig. 4. At early time, from 0 to 2 hours, there is no big difference between numbers of mRNA with or without NSP1. Starting from 2.5 hours, NSP1 promotes degradation of mRNA faster. This result matches the discussion before, that NSP1 is a late viral protein and we expect that it may have inhibition affects in later stage.

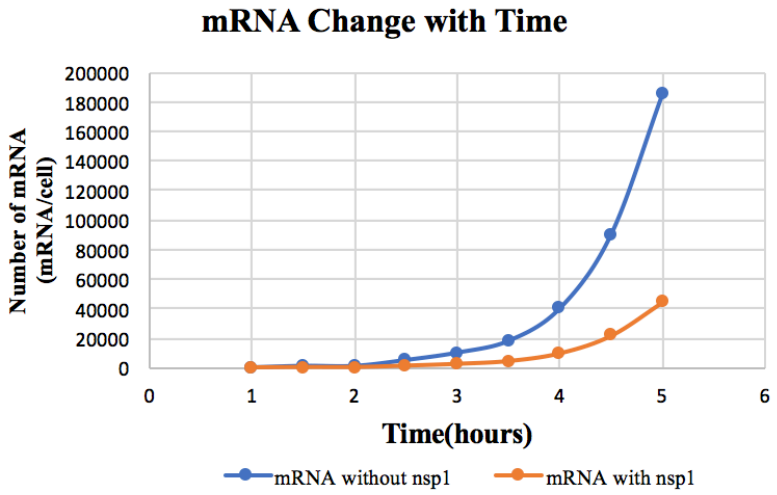


Figure. 4 mRNA with or without NSP1 change with time.

CHEME7770 Final Report
Tina Ye (ty369) and Xiaojing Ma (xm89)
May 21, 2020
Inhibition of GFP-ssrA Protein by nsp1

We assume the GFP-ssrA expression in *E. coli* as the pervious nsp1 and mRNA expression is analyzed in *E. coli* system. In *E. coli* cells, the ATP- dependent ClpXP protease contributes to degradation of ssrA- tagged proteins. After collecting some data from literature [12] in Excel, we are able to construct a plot to illustrate concentration of GFP-ssrA at different time steps, shown in Fig. 5. The plot shows that the higher level of clpXP causes more rapid degradation of GFP-ssrA. For convenience, GFP-ssrA with 0 nm clpXp is used for analysis of the inhibition mention below.

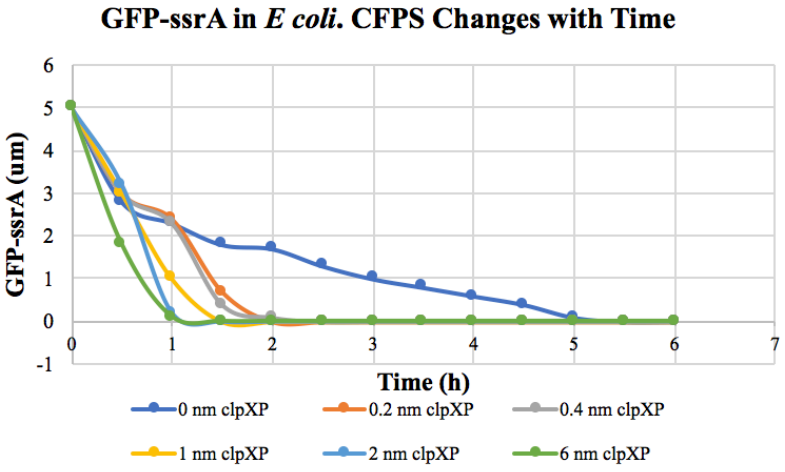


Figure. 5 GFP-ssrA in E coli. CFPS change with time.

In order to see how NSP1 acts on GFP-ssrA expression, we assume NSP1 has the same influence on both mRNA and GFP-ssrA, which means they have the same percentage change of 76%. After adding GFP-ssrA and NSP1 in a cell free reaction, NSP1 causes 76% reduction of GFP-ssrA at each time. As shown in Fig. 6, NSP1 promotes degradation of GFP-ssrA. The calculated result is a little bit different from what we expected. As mentioned before, NSP1 is a late viral protein, thus it should influence the expression of GFP-ssrA in the late time. However, from Fig. 6, the expression of GFP-ssrA reduces fast at the beginning. This unexpected result happens maybe because we assume the same percentage reduction at all time points, while the reduction may change at different times.

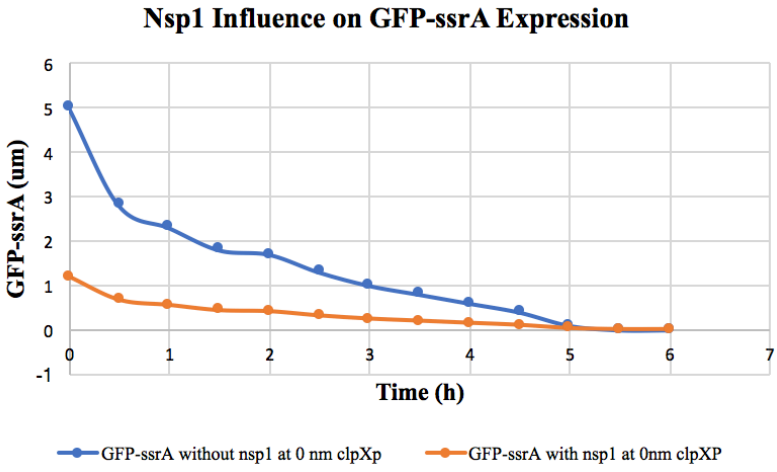


Figure. 6 Nsp1 influence on GFP-ssrA expression.

Conclusions

As a conclusion, some key findings are noticed. First and foremost, from the Sequence-Specific analysis, the maximum optimal productivity of a cell-free reaction is inversely proportional to carbon number of the protein. The higher number of carbon a protein contain, the less maximum optimal productivity will be generated relatively based on the FBA model. Also, the translation elongation rate is the most important factor controlling productivity, while RNAP

and ribosome abundance have a relatively modest effect. From the study of expression of nsp1 with time, it shows that nsp1 increases at late times post infection. Thus, NSP1 is a late viral protein. Some controversies occurs when comparing the results of nsp1's inhibition effect on the GFP-ssrA expression, which shows that the expression of GFP-ssrA was inhibited by nsp1 started at early stage. This may be caused by imprecise calculations when we give assumption that same percentage reduction of mRNA and GFP-ssrA at all time points. Also, nsp1 promotes degradation of GFP-ssrA in cell free reaction. At last, the amount of clpXP influences the expression of GFP-ssrA in *E. coli* CFPS, the higher the clpXp, the quicker the GFP-ssrA degrades.

Acknowledgements

This project would not have been possible without the support of course advisors. Many thanks to our adviser, Prof. Jeffrey Varner, who answers our questions, gives our ideas about the project, and helps us figure out coding in Julia. Also thanks to our other advisor, Prof. Matthew Paszek, who offered support, without his course knowledge we cannot have any ideas about our projects. We also acknowledge the sequence specific model template provided by Varner lab.

References

1. "SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) Sequences." National Center for Biotechnology Information, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/.
2. "Coronavirus (COVID-19)." Google News, Google, news.google.com/covid19/map?hl=en-US&gl=US&ceid=US:en.
3. Swartz J. Developing cell-free biology for industrial applications. *J Ind Microbiol Biotechnol* 2006;33(7):476e85.
4. Katzen F, Chang G, Kudlicki W. The past, present and future of cell-free protein synthesis. Trends *Biotechnol* 2005;23(3):150e6.
5. Krishna N, Sydney I. R,Kumari G. L, Shinji M, Coronavirus nonstructural protein 1: Common and distinct functions in the regulation of host and viral gene expression. *Virus Research, Elsevier B.V.*, 16 Apr. 2015
6. Vilkhovoy, M., Horvath, N., Shih, C.-H., Wayman, J. A., Calhoun, K., Swartz, J., and Varner, J. D. (2017) Sequence Specific Modeling of *E. coli* Cell-Free Protein Synthesis. *ACS Synthetic Biology*.
7. Jewett, M. C., and Swartz, J. R. (2004) Mimicking theEscherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and Bioengineering* 86, 19–26.
8. Allen, T. E., and Palsson, B. Ø. (2003) Sequence-Based Analysis of Metabolic Demands for Protein Synthesis in Prokaryotes. *Journal of Theoretical Biology* 220, 1–18.
9. Martinez-Alvarez, L., Pina-Vazquez, C., Zarco, W., and Padilla-Noriega, L. (2013) The shift from low to high non-structural protein 1 expression in rotavirus-infected MA-104 cells. *Memórias do Instituto Oswaldo Cruz* 108, 421–428.

10. Buzatu, D. A., Moskal, T. J., Williams, A. J., Cooper, W. M., Mattes, W. B., and Wilkes, J. G. (2014) An Integrated Flow Cytometry-Based System for Real-Time, High Sensitivity Bacterial Detection and Identification. *PLoS ONE 9*.
11. Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., Kubo, H., and Makino, S. (2006) Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proceedings of the National Academy of Sciences* 103, 12885–12890.
12. Garamella, J., Marshall, R., Rustad, M., and Noireaux, V. (2016) The All *E. coli* TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS Synthetic Biology* 5, 344–355.

Appendix. A Non-structural Protein's Amino Acid Sequence.

nsp1 QHD43415_1(L=180aa) (https://zhanglab.ccmb.med.umich.edu/COVID-19/)
MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQHLKDGTCGLVEVEK
GVLPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGETLGVLVPHVGEIPVA
YRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQENWNTKHSSGVTRELM
RELNGG

nsp10 QHD43415_10 (L=139aa) (https://zhanglab.ccmb.med.umich.edu/COVID-19/)
AGNATEVPANSTVLSFCAFAVDAAKAYKDYLASGGQPITNCVKMLCTHTGTGQAITVT
PEANMDQESFGGASCCLYCRCHIDHPNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVC
TVCGMWKGYGCSCDQLREPMLQ

RdRp (L=141aa) (https://www.uniprot.org/uniprot/V5YMF8)
WDYPKCDRAMPNMLRIMASLILARKHSTCCNLSHRFYRLANECAQVLSEMVMCGGSL
YVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIADKYVRNLQHKLYQNL
YRNRD VDHEFVSEFY AYLRKHFSMMI

Nucleotides	Nsp1	nsp10	RdRp
A	126	94	105
T	94	78	91
G	189	130	122
C	131	115	105
Total bp	540	423	417

Table 1. Nucleotides ATGC content breakdown.

CHEME7770 Final Report
Tina Ye (ty369) and Xiaojing Ma (xm89)
May 21, 2020
Appendix. B Fundamental Reaction Involved in Protein Synthesis.

Simplified, fundamental reaction set for protein production

Transcription initiation:	$G + \text{RNAP} \xrightarrow{v_1} G^*$
Transcription:	$G^* + n\text{NTP} \xrightarrow{v_2} \text{mRNA} + G + \text{RNAP} + 2n\text{P}_i$
mRNA decay:	$\text{mRNA} \xrightarrow{v_3} n\text{NMP}$
Translation initiation:	$\text{mRNA} + \text{rib} \xrightarrow{v_4} \text{rib}^*$
Translation:	$\text{rib}^* + a\text{AAtRNA} + 2a\text{GTP} \xrightarrow{v_5} a\text{tRNA} + 2a\text{GDP} + 2a\text{P}_i + \text{rib} + \text{mRNA} + \text{protein}$
tRNA charging:	$\text{AA} + \text{tRNA} + \text{ATP} \xrightarrow{v_6} \text{AMP} + 2\text{P}_i + \text{AAtRNA}$
Exchange fluxes:	$\text{AA}_{ext} \xrightarrow{b_1} \text{AA}$
	$\text{NTP}_{ext} \xrightarrow{b_2} \text{NTP}$
	$\text{protein} \xrightarrow{b_3} \text{protein}_{ext}$
	$\text{NMP} \xrightarrow{b_4} \text{NMP}_{ext}$
	$\text{ATP}_{ext} \xrightarrow{b_5} \text{ATP}$
	$\text{AMP} \xrightarrow{b_6} \text{AMP}_{ext}$
	$\text{GTP}_{ext} \xrightarrow{b_7} \text{GTP}$
	$\text{GDP} \xrightarrow{b_8} \text{GDP}_{ext}$
	$\text{P}_i \xrightarrow{b_9} \text{P}_{i\,ext}$

Table 2. Fundamental Reaction Set for Protein Production. [8]

Reactions added to the fundamental system when including the internal production of RNAP, tRNA, and rRNA

RNAP:	$G_P + \text{RNAP} \xrightarrow{v_{1P}} G_P^*$
	$G_P^* + n_P\text{NTP} \xrightarrow{v_{2P}} \text{mRNA}_P + G_P + \text{RNAP} + 2n_P\text{P}_i$
	$\text{mRNA}_P \xrightarrow{v_{3P}} n_P\text{NMP}$
	$\text{mRNA}_P + \text{rib} \xrightarrow{v_{4P}} \text{rib}_P^*$
	$\text{rib}_P^* + a_P\text{AAtRNA} + 2a_P\text{GTP} \xrightarrow{v_{5P}} a_P\text{tRNA} + 2a_P\text{GDP} + 2a_P\text{P}_i + \text{rib} + \text{mRNA}_P + \text{RNAP}$
tRNA:	$G_t + \text{RNAP} \xrightarrow{v_{1t}} G_t^*$
	$G_t^* + n_t\text{NTP} \xrightarrow{v_{2t}} \text{tRNA} + G_t + \text{RNAP} + 2n_t\text{P}_i$
rRNA:	$G_r + \text{RNAP} \xrightarrow{v_{1r}} G_r^*$
	$G_r^* + n_r\text{NTP} \xrightarrow{v_{2r}} \text{rib} + G_r + \text{RNAP} + 2n_r\text{P}_i$

Table 3. Reactions for Internal Production of RNAP, tRNA, and rRNA. [8]

Appendix. C Maximum Optimal Productivity Variation Plots.

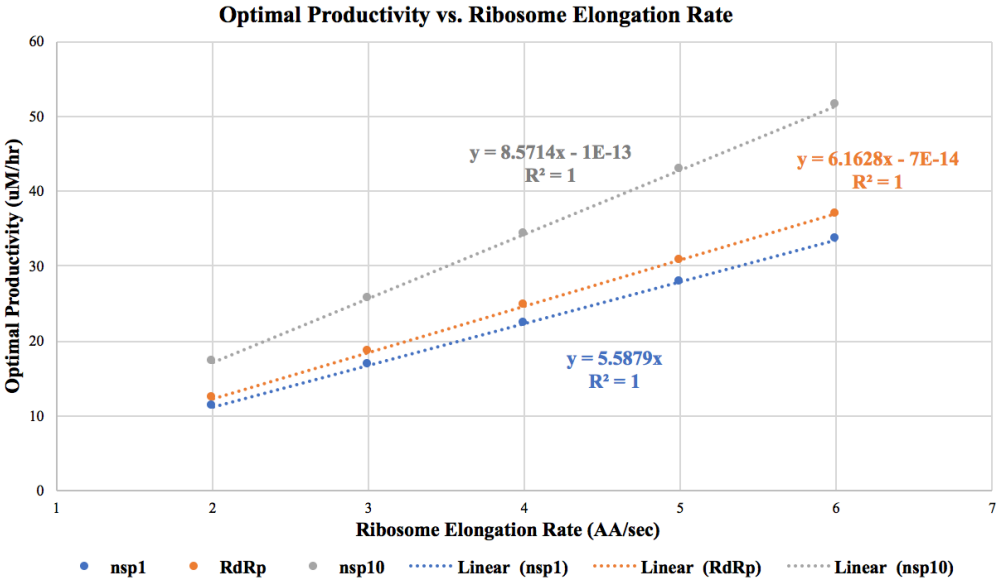


Figure. 7 Maximum Optimal Productivity varies with Ribosome Elongation.

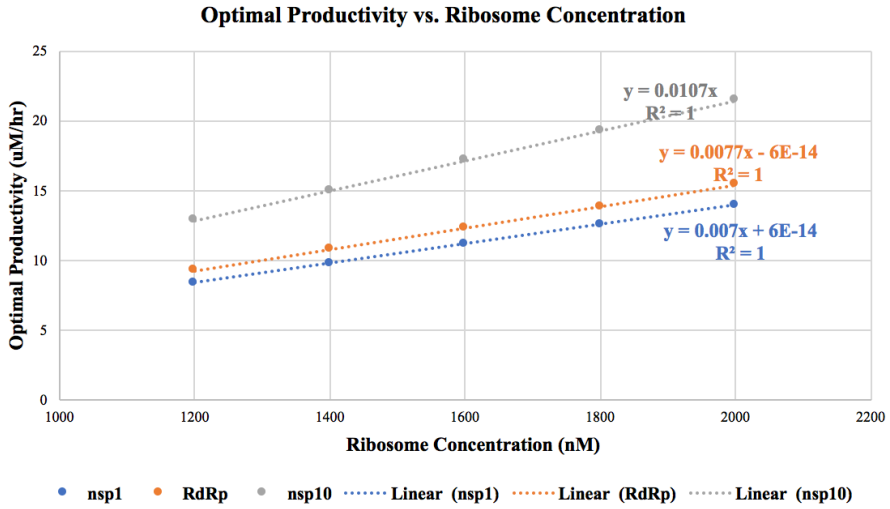


Figure. 8 Maximum Optimal Productivity varies with Ribosome Concentration.

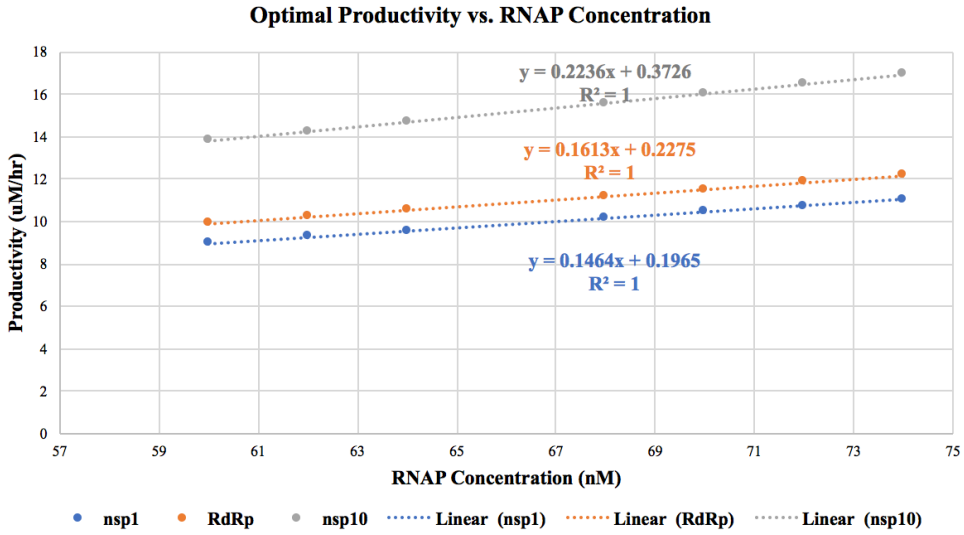


Figure. 9 Maximum Optimal Productivity varies with RNAP Concentration.

Appendix. D

Number of mRNA Changes with Time			
Time(hours)	**E. coli cells per 100 ml sample**	**# of mRNA**	**# of mRNA with nsp1**
1	0	0	0
1.5	1	1000	240
2	1	1000	240
2.5	5	5000	1200
3	10	10000	2400
3.5	18	18000	4320
4	40	40000	9600
4.5	90	90000	21600
5	185	185000	44400

Table. 4 Calculations for Number of mRNA Change with Time.

Appendix. E Presentation Slides

Presentation slide is attached below and can also be accessed through Github links:
https://github.com/ty369/ChemE7770_Final_Project_Report

COVID-19 Protein Expression Analysis in Cell-Free System

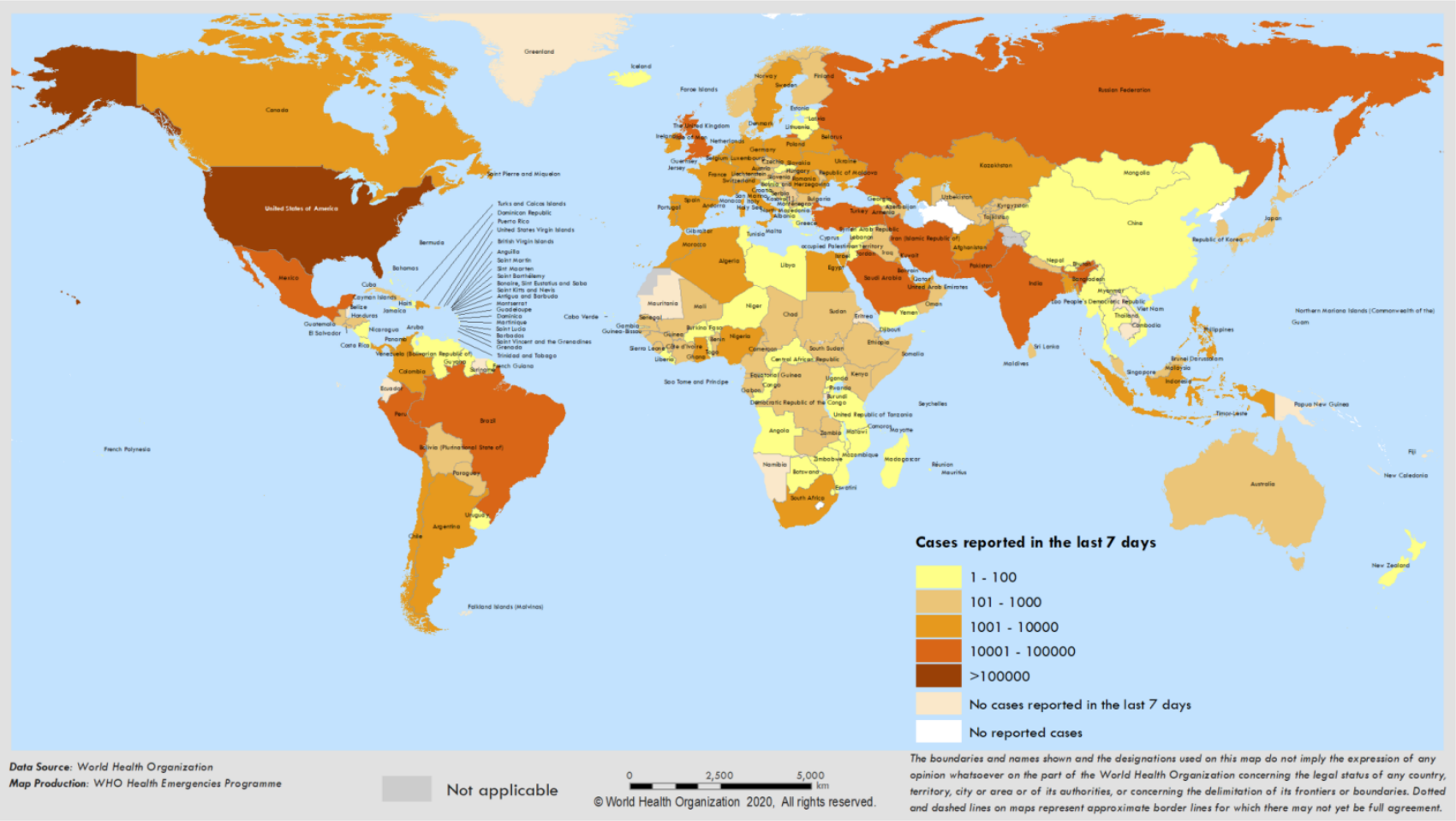
ChemE 7770 Final Project
Tina Ye (ty369) & Xiaojing Ma (xm89)
May 14th 2020

Outline

1. Motivation and Background Information.
2. Rationale of the Simulation Approach.
3. Quantitative and Qualitative Analysis.
4. Key Findings and Conclusion.

COVID-19 is A Current Pandemic

Figure 1. Number of confirmed COVID-19 cases reported in the last seven days by country, territory or area, 6 May to 12 May**



- 4.34M confirmed cases; 297K death by May 13th.
- Transmission: contact with contaminated objects or between persons. (CDC)
- Affects our normal life and the economy.

Project Tasks

1. Investigate different constraints' effect on optimal productivity of RdRp, nsp1, and nsp10 in the cell free system.

2. Propose modeling to reveal the expression of COVID19 nonstructural protein nsp1 varies with time.

3. Review literature's modeling on the effects of nsp1 on the expression of GFP-ssrA protein.

SARS-CoV-2 Proteins for this Study

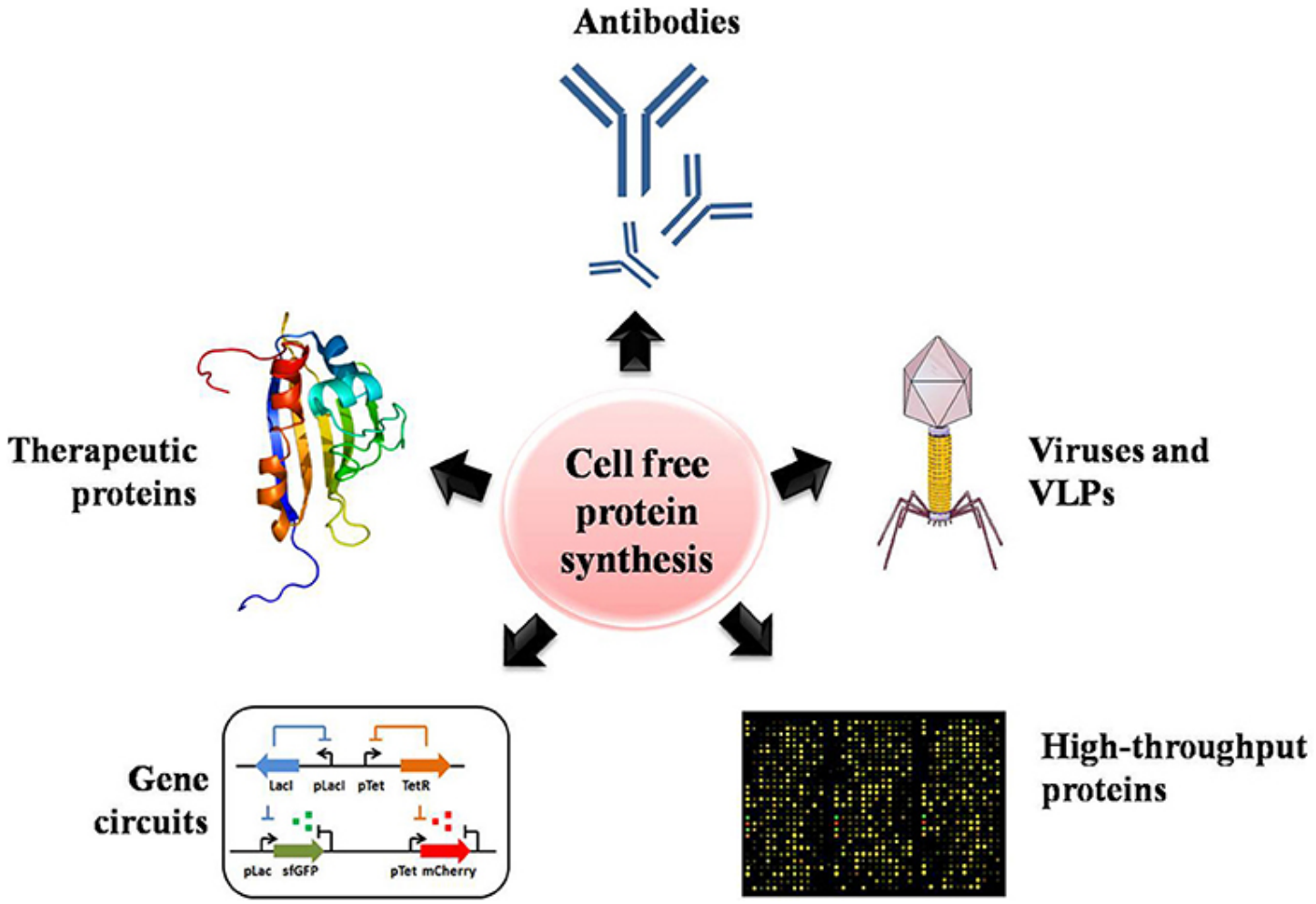
- **RdRp** (the RNA-dependent RNA polymerase, also named nsp12):
  - central component of coronaviral replication/transcription machinery
  - primary target for the antiviral drug, remdesivir.

- **nsp1** (non-structural protein 1): inhibit host gene expression
  - degradation of expressed RNA transcripts and host endogenous mRNAs

- **nsp10** (non-structural protein 10): a critical cofactor for activation of multiple replicative enzymes
  - binds and stimulates both the nsp14 and nsp16 activities.

Yan Gao et al, 2020; Wataru Kamitani et al, 2006; Bouvet M et al, 2014

Cell Free Protein Synthesis as Model Condition



- The commercial *E. coli* TX-TL cell-free protein synthesis (CFPS)

- Advantage:
  - high efficiency
  - flexibility
  - low cost

1. FBA Sequence-Specific Modeling

Major Reactions Involved for Protein Synthesis

Simplified, fundamental reaction set for protein production

Transcription initiation:	$G + \text{RNAP} \xrightarrow{v_1} G^*$
Transcription:	$G^* + n\text{NTP} \xrightarrow{v_2} \text{mRNA} + G + \text{RNAP} + 2n\text{P}_i$
mRNA decay:	$\text{mRNA} \xrightarrow{v_3} n\text{NMP}$
Translation initiation:	$\text{mRNA} + \text{rib} \xrightarrow{v_4} \text{rib}^*$
Translation:	$\text{rib}^* + a\text{AAtRNA} + 2a\text{GTP} \xrightarrow{v_5} a\text{tRNA} + 2a\text{GDP} + 2a\text{P}_i + \text{rib} + \text{mRNA} + \text{protein}$
tRNA charging:	$\text{AA} + \text{tRNA} + \text{ATP} \xrightarrow{v_6} \text{AMP} + 2\text{P}_i + \text{AAtRNA}$
Exchange fluxes:	$\text{AA}_{ext} \xrightarrow{b_1} \text{AA}$
	$\text{NTP}_{ext} \xrightarrow{b_2} \text{NTP}$
	$\text{protein} \xrightarrow{b_3} \text{protein}_{ext}$
	$\text{NMP} \xrightarrow{b_4} \text{NMP}_{ext}$
	$\text{ATP}_{ext} \xrightarrow{b_5} \text{ATP}$
	$\text{AMP} \xrightarrow{b_6} \text{AMP}_{ext}$
	$\text{GTP}_{ext} \xrightarrow{b_7} \text{GTP}$
	$\text{GDP} \xrightarrow{b_8} \text{GDP}_{ext}$
	$\text{P}_i \xrightarrow{b_9} \text{P}_{i\,ext}$

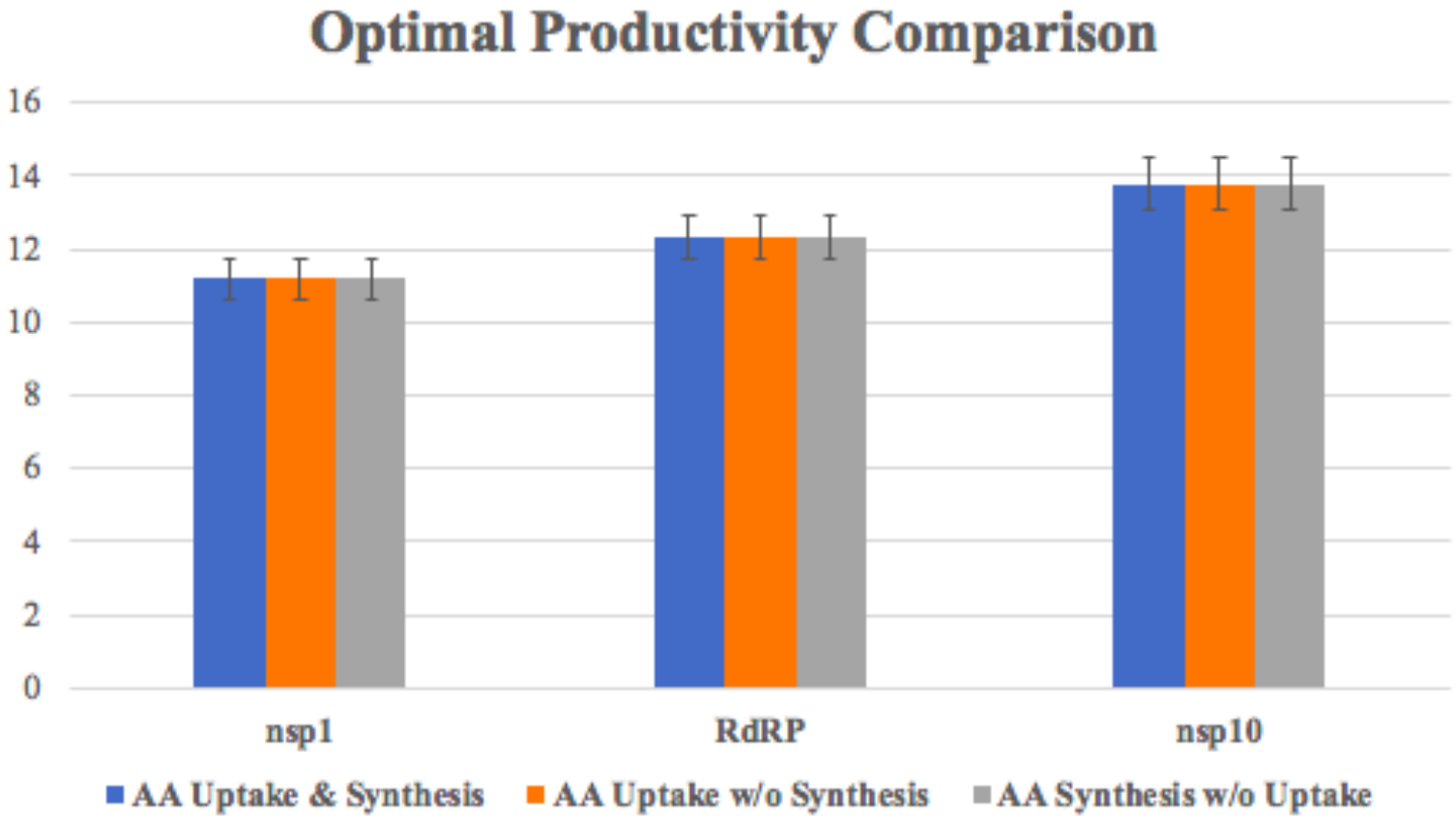
Eqn for internal production of RNAP, tRNA, and rRNA showed in Appendix.

- G is the protein of interest: RdRp, nsp1, or nsp10
- Assume commercial *E. coli* myTXTL extract
- Assume Steady State Condition in Cell-Free System
- Adopted coding template from Varner Lab

T. Allen et al, 2003, M. Vilkhovoy, 2018

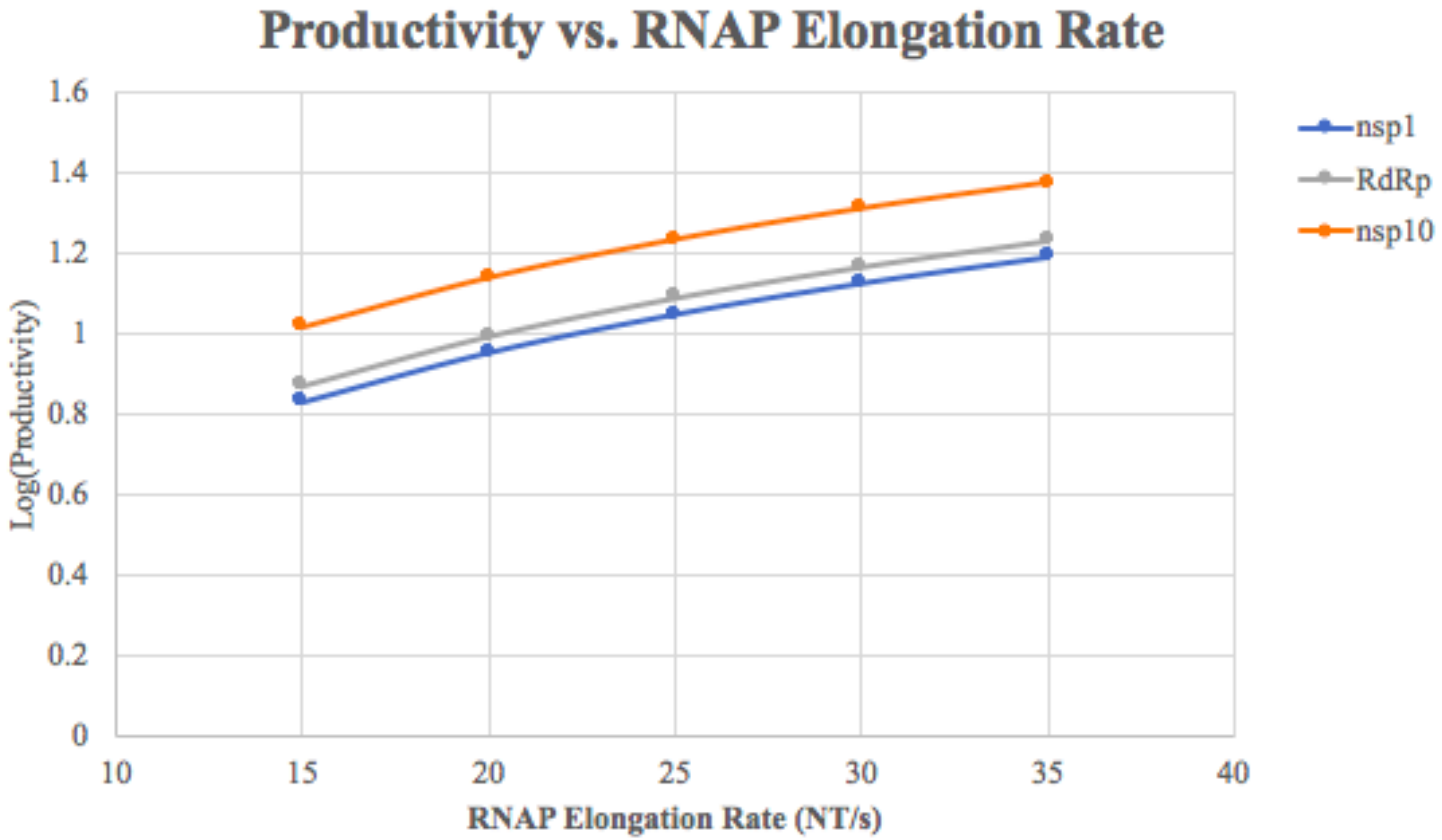
1. FBA Sequence-Specific Modeling

Simulation Result



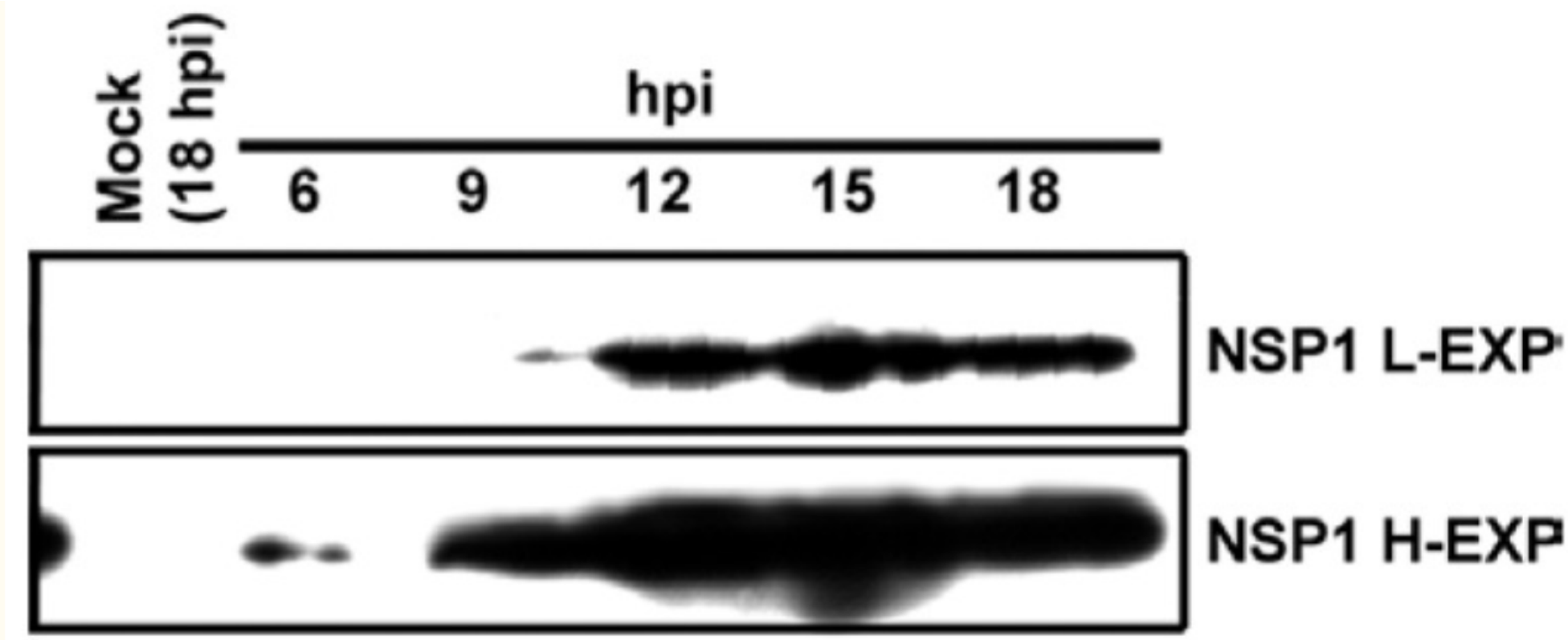
- Optimal productivity inversely proportional to carbon number of the protein.
  - nsp1:180 aa, 872 C
  - RdRp: 141 aa, 704 C
  - nsp10: 139 aa, 636 C
- AA uptake w/ or w/o de novo synthesis seems to not affect productivity.

1. FBA Sequence-Specific Modeling



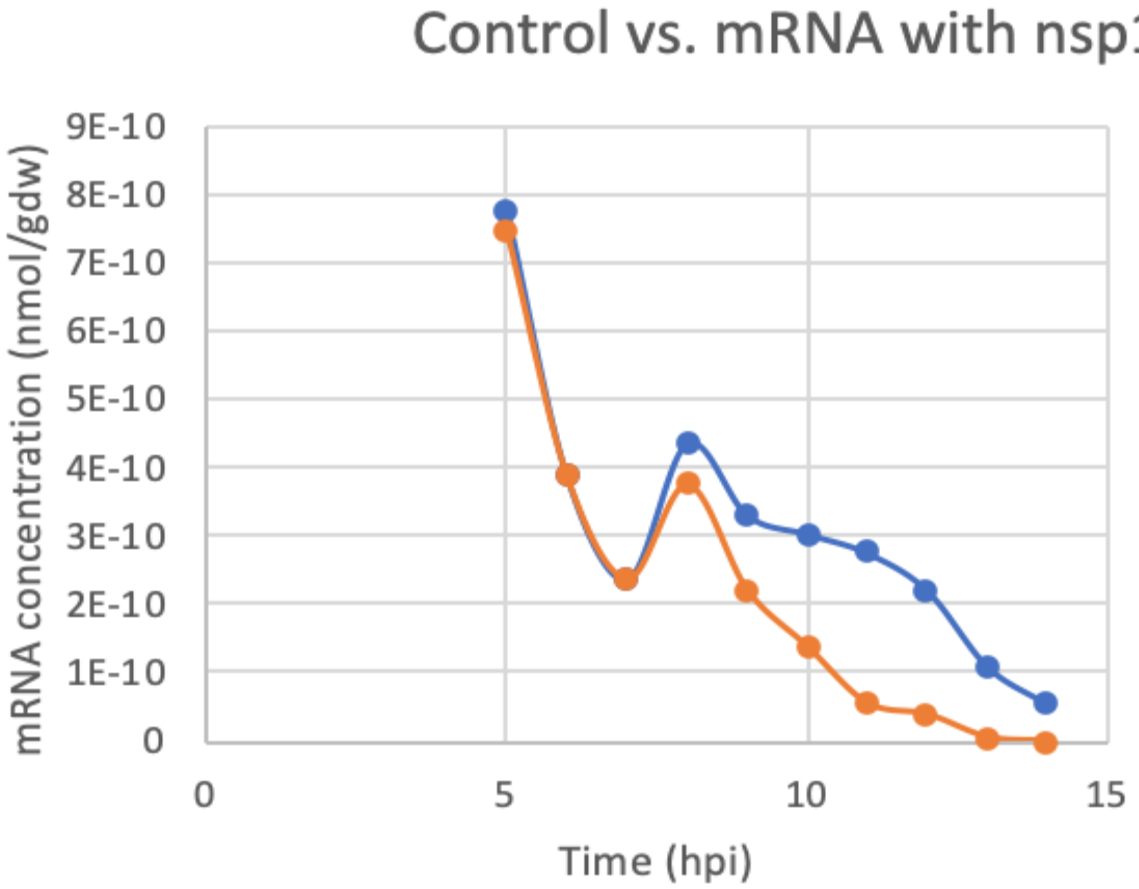
- Maximum optimal productivity increases proportionally to translation elongation rate.
- RNAP elongation rate has greater influence on productivity than for RNAP concentration.

2. The expression of nsp1 varies with time



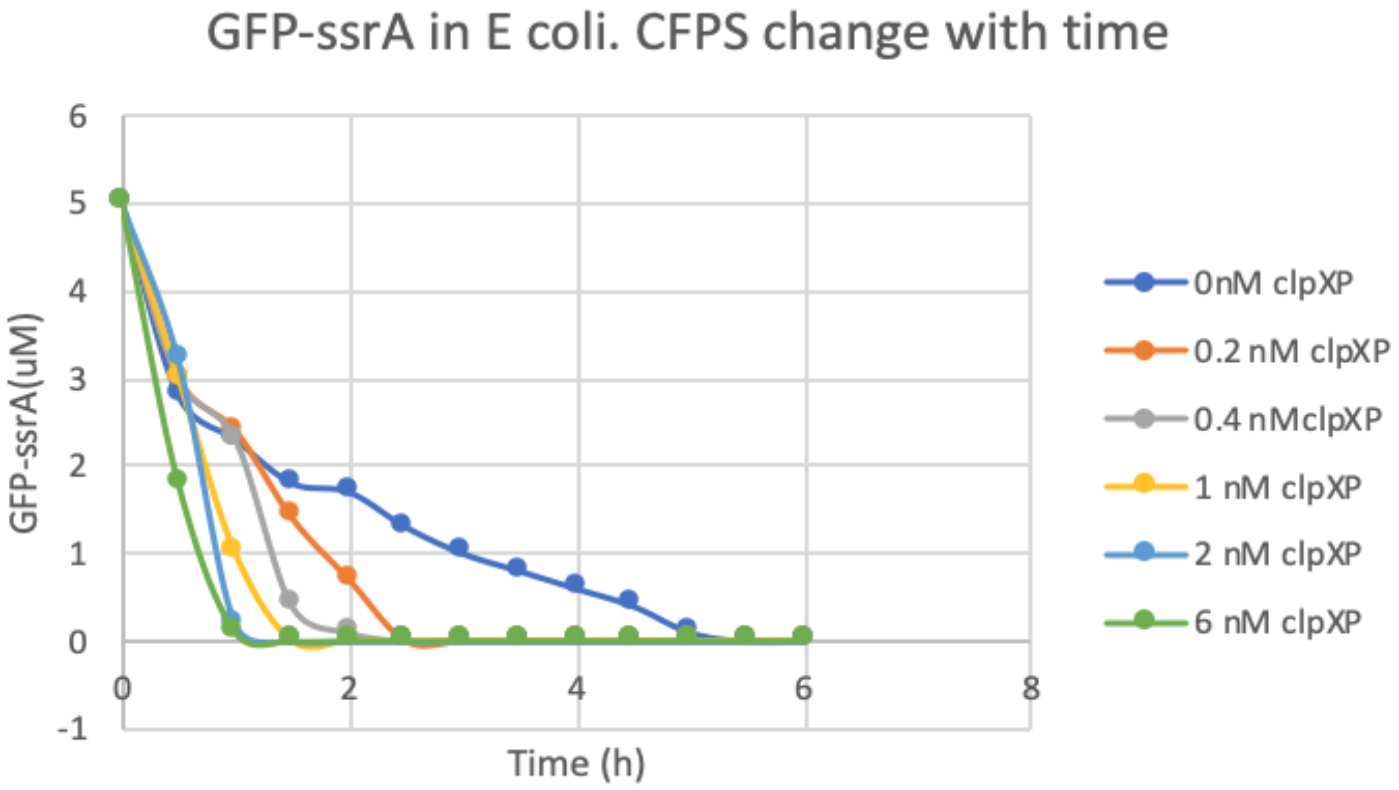
Laura Martínez-Álvarez et al, 2013

2. The expression of mRNA with nsp1



nsp1-CD and nsp1 employed the same mechanism to inhibit translation

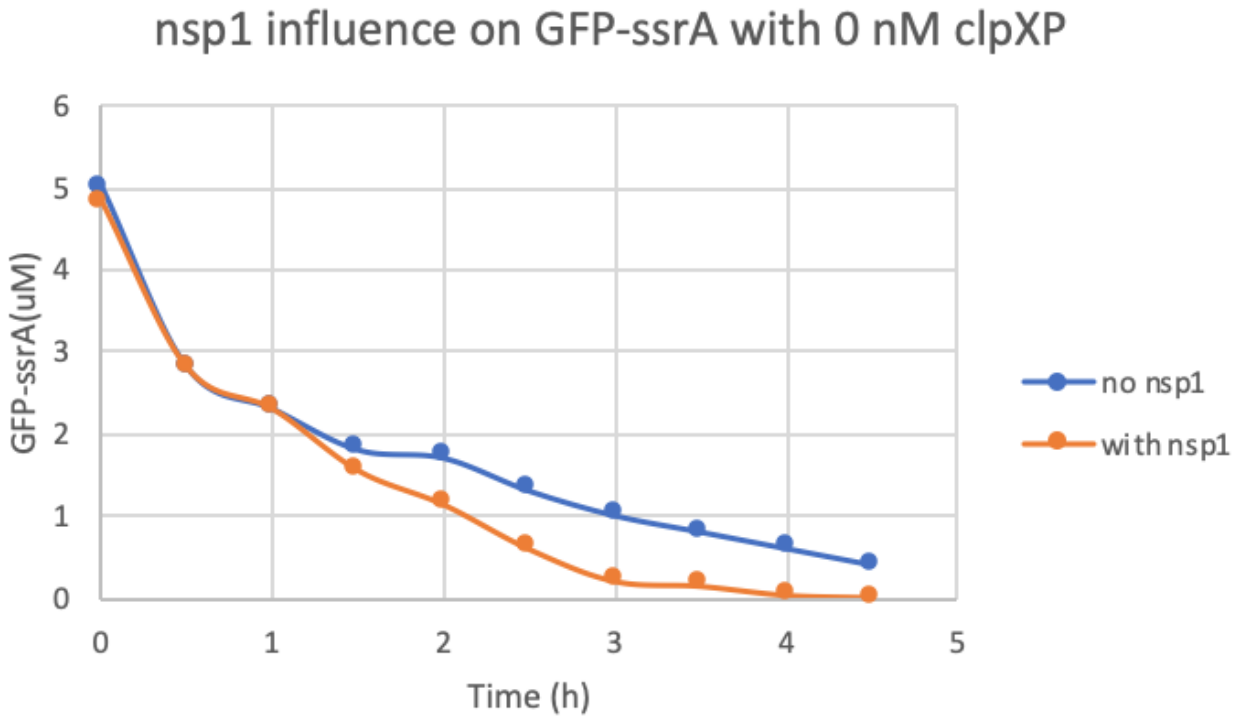
3. The expression of GFP-ssrA protein in CFPS



In E. coli, the ATP-dependent ClpXP protease contributes to degradation of ssrA-tagged proteins.

3. nsp1 Inhibits Production of GFP-ssrA

Time(hpi)	% change	GFP-ssrA(um) with 0nm clpXP	with nsp1
0	0.03571	5	4.82142857
0.5	0	2.8	2.8
1	0	2.3	2.3
1.5	0.13924	1.8	1.54936709
2	0.33333	1.7	1.13333333
2.5	0.54545	1.3	0.59090909
3	0.8	1	0.2
3.5	0.825	0.8	0.14
4	0.95	0.6	0.03
4.5	1	0.4	0



Conclude with Key Findings

- Optimal productivity inversely proportional to carbon number of the protein.
- Translation elongation rate greatly affects productivity compares to RNAP and ribosome abundance.
- NSP1 is a late viral protein.
- The amount of clpXP influence the expression of GFP-ssrA in *E coli*. CFPS.
- The nsp1 promotes degradation of GFP-ssrA in cell free reaction.

Thank you for your attention!

Reference

Bouvet, Mickaël, et al. “Coronavirus Nsp10, a Critical Co-Factor for Activation of Multiple Replicative Enzymes.” *Journal of Biological Chemistry*, 12 Sept. 2014, www.jbc.org/content/289/37/25783.full.

CohenApr, Jon, et al. “COVID-19 Vaccine Protects Monkeys from New Coronavirus, Chinese Biotech Reports.” *Science*, 6 May 2020, www.sciencemag.org/news/2020/04/covid-19-vaccine-protects-monkeys-new-coronavirus-chinese-biotech-reports.

“Disease Background of COVID-19.” *European Centre for Disease Prevention and Control*, 8 May 2020, www.ecdc.europa.eu/en/2019-ncov-background-disease.

Farrell, Christopher M., et al. “Cytoplasmic Degradation of SsrA-Tagged Proteins.” *Wiley Online Library*, John Wiley & Sons, Ltd, 17 Aug. 2005, onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2958.2005.04798.x.

Gao, Yan, et al. “Structure of the RNA-Dependent RNA Polymerase from COVID-19 Virus.” *Science*, American Association for the Advancement of Science, 10 Apr. 2020, science.sciencemag.org/content/early/2020/04/09/science.abb7498.

Ghose, Tia. “How Are People Being Infected with COVID-19?” *LiveScience*, Purch, 7 Apr. 2020, www.livescience.com/how-covid-19-spreads-transmission-routes.html.

Martínez-Álvarez, Laura, et al. “The Shift from Low to High Non-Structural Protein 1 Expression in Rotavirus-Infected MA-104 Cells.” *Memorias Do Instituto Oswaldo Cruz*, Instituto Oswaldo Cruz, Ministério Da Saúde, June 2013, www.ncbi.nlm.nih.gov/pmc/articles/PMC3970611/.

PMC, Europe. *Europe PMC*, europepmc.org/article/MED/23035226.

“UNITED STATES DEPARTMENT OF LABOR.” *Safety and Health Topics | COVID-19 - Background | Occupational Safety and Health Administration*, www.osha.gov/SLTC/covid-19/background.html.

Appendix 1. Equation for Internal Production of RNAP, tRNA, and rRNA

Reactions added to the fundamental system when including the internal production of RNAP, tRNA, and rRNA

RNAP:	$G_P + \text{RNAP} \xrightarrow{v_{1P}} G_P^*$
	$G_P^* + n_P\text{NTP} \xrightarrow{v_{2P}} \text{mRNA}_P + G_P + \text{RNAP} + 2n_P\text{P}_i$
	$\text{mRNA}_P \xrightarrow{v_{3P}} n_P\text{NMP}$
	$\text{mRNA}_P + \text{rib} \xrightarrow{v_{4P}} \text{rib}_P^*$
	$\text{rib}_P^* + a_P\text{AAtRNA} + 2a_P\text{GTP} \xrightarrow{v_{5P}} a_P\text{tRNA} + 2a_P\text{GDP} + 2a_P\text{P}_i + \text{rib} + \text{mRNA}_P + \text{RNAP}$
tRNA:	$G_t + \text{RNAP} \xrightarrow{v_{1t}} G_t^*$
	$G_t^* + n_t\text{NTP} \xrightarrow{v_{2t}} \text{tRNA} + G_t + \text{RNAP} + 2n_t\text{P}_i$
rRNA:	$G_r + \text{RNAP} \xrightarrow{v_{1r}} G_r^*$
	$G_r^* + n_r\text{NTP} \xrightarrow{v_{2r}} \text{rib} + G_r + \text{RNAP} + 2n_r\text{P}_i$

Appendix 2. calculations for task 2

Assumption: 100 mRNA/cell

Table 1				
Time(hpi)	mRNA (% of total)	mRNA (mRNA/cell)	control (% of total)	control (mRNA/cell)
5	13.5	13.5	14	14
6	7	7	7	7
7	4.3	4.3	4.3	4.3
8	6.8	6.8	7.9	7.9
9	4	4	6	6
10	2.5	2.5	5.5	5.5
11	1	1	5	5
12	0.7	0.7	4	4
13	0.1	0.1	2	2
14	0	0	1	1

Appendix 3. calculations for task 2

Calculations					
mRNA with nsp1(mol)	mRNA with nsp1 (nmol)	mRNA with nsp1 (nmol/gdw)	control (mol)	control (nmol)	control (nmol/gdw)
2.24178E-23	2.24178E-14	7.4726E-10	2.3248E-23	2.32481E-14	7.74936E-10
1.1624E-23	1.1624E-14	3.87468E-10	1.1624E-23	1.1624E-14	3.87468E-10
7.14048E-24	7.14048E-15	2.38016E-10	7.1405E-24	7.14048E-15	2.38016E-10
1.12919E-23	1.12919E-14	3.76398E-10	1.3119E-23	1.31186E-14	4.37286E-10
6.64231E-24	6.64231E-15	2.2141E-10	9.9635E-24	9.96347E-15	3.32116E-10
4.15144E-24	4.15144E-15	1.38381E-10	9.1332E-24	9.13318E-15	3.04439E-10
1.66058E-24	1.66058E-15	5.53526E-11	8.3029E-24	8.30289E-15	2.76763E-10
1.1624E-24	1.1624E-15	3.87468E-11	6.6423E-24	6.64231E-15	2.2141E-10
1.66058E-25	1.66058E-16	5.53526E-12	3.3212E-24	3.32116E-15	1.10705E-10
0	0	0	1.6606E-24	1.66058E-15	5.53526E-11

Appendix 4 calculations for task 3

Time(h)	GFP-ssrA(um) with 0nm clpXP	0.2nm clpXP	0.4nm clpXP	1nm clpXP	2nm clpXP	6nm clpXP
0	5	5	5	5	5	5
0.5	2.8	3	3	3	3.2	1.8
1	2.3	2.4	2.3	1	0.2	0.1
1.5	1.8	1.4	0.4	0	0	0
2	1.7	0.7	0.1	0	0	0
2.5	1.3	0	0	0	0	0
3	1	0	0	0	0	0
3.5	0.8	0	0	0	0	0
4	0.6	0	0	0	0	0
4.5	0.4	0	0	0	0	0
5	0.1	0	0	0	0	0
5.5	0	0	0	0	0	0
6	0	0	0	0	0	0

Appendix 5. Gene Sequence for FBA

nsp1 QHD43415_1(L=180) (https://zhanglab.ccmb.med.umich.edu/COVID-19/)

MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQHLKDGTCGLVEVEKGVLPQLEQPYVFIKRSDARTAPHGHV
MVELVAELEGIQYGRSGETLGVLVPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQENWNTKHSSG
VTRELMRELNGG

Nsp10 QHD43415_10 (L=139) (https://zhanglab.ccmb.med.umich.edu/COVID-19/)

AGNATEVPANSTVLSFCAFAVDAAKAYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEANMDQESFGGASCCLYCRCHIDHP
NPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGMWKGYGCSCDQLREPMLQ

RdRp (L=141) (https://www.uniprot.org/uniprot/V5YMF8)

WDYPKCDRAMPNMLRIMASLILARKHSTCCNLSHRFYRLANECAQVLSEMVMCGGSLYVKPGGTSSGDATTAYANSVFNICQAV
TANVNALLSTDGNKIADKYVRNLQHK LYQNLYRNRD VDHEFVSEFY AYLRKHFSMM I