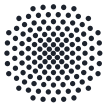
Modeling *in vitro* and *in vivo* transcription and translation with different levels of granularity

DISSERTATION

Von der Fakultät Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von
Alexander Nieß
aus Sigmaringen

Hauptberichter:	Prof. Dr.-Ing. Ralf Takors
Mitberichter:	Prof. Dr.-Ing. Dr. h.c. Matthias Reuss
Mitberichter:	Prof. Dr. Ir. Frank Delvigne
Tag der mündlichen Prüfung:	15.02.2018



University of Stuttgart
Germany

Institut für Bioverfahrenstechnik

2018

ACKNOWLEDGEMENT

This research was conducted at the Institute of Biochemical Engineering at the University of Stuttgart beginning in February 2014. During this period, many people contributed to the progress of this study whom I would like to thank:

I am indebted to Prof. Dr.-Ing. Ralf Takors for his academic assistance and support throughout this work. His patient guidance and encouragement provided a significant pillar for this research, and I am very grateful for the freedom and trust he provided for my work. Furthermore, I want to thank Prof. Dr.-Ing. Dr. h.c. Matthias Reuss and Prof. Dr. Ir. Frank Delvigne for their interest in this research and their participation in the examination committee. I also want to thank PD Dr. Martin Siemann-Herzberg and Jurek Failmezger for providing experimental data and for the scientific discussions, as well as Michael Löffler and Joana Simen for their help and discussion that led to this work. Many thanks are addressed to Silke Reu, who always helped with administration and motivation problems.

Further gratitude is granted to the students who contributed to this study: Lorenz Achtnich, Tobias Thiemt, Ferdinand Michalik, Christian Kempt, Paul Harsch, Isabell Ehrler, and Sabrina Esswein. A special thank you goes to Maike Kuschel for her great effort during her master's thesis and for all the discussions that made this work possible. The financial support for this work by the EU ("ST-FLOW" project, Grant 289326 and "Empower putida" project, Grant 635536), the Ministerium für Wissenschaft, Forschung und Kunst, Baden-Württemberg (MWK) "Nachhaltige und effiziente Biosynthesen", Grant AZ33-7533-6-195/7, BMBF (Grant FKZ031A157D and FKZ0316178A), and the University of Stuttgart is also gratefully acknowledged.

Finally, I thank all my present and past coworkers for the wonderful atmosphere at the IBVT, especially Tobias Vallon, Salaheddine Laghrami, Maria Rahnert, Gerhard Eigenstetter, Maria Hermann, Michaela Graf, Andreas Freund, Max Becker, and Lisa Junghans.

To my fiancé for her love and her support

Stuttgart, Germany *Alexander Nieß*

March 2018

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Abbreviations:

30S	Small ribosomal subunit
50S	Large ribosomal subunit
AA	Amino acid
Ala	Alanine
Arg	Arginine
ARS	Aminoacyl tRNA synthetase
Asn	Asparagine
Asp	Asparatate
ATP	Adenosine triphosphate
CAI	Codon adaptation index
CAT	Chloramphenicol acetyl transferase
CDF	Cumulative distribution function
CFPS	Cell-free protein synthesis
CTP	Cytosine triphosphate
Cys	Cysteine
DAE	Differential-algebraic equation
EF	Elongation factor
EFG	Elongation factor G
EFTs	Elongation factor thermal stable
EFTu	Elongation factor thermal unstable
FCC	Flux control coefficient
g	Gram
GFP	Green fluorescent protein
Gln	Glutamine
Glu	Glutamate
Gly	Glycine
GOI	Gene of interest
GTP	Guanine triphosphate
His	Histidine
IF	Initiation factor
Ile	Isoleucine

L	Liter
Leu	Leucine
Lys	Lysine
MCA	Metabolic control analysis
Met	Methionine
mRNA	Messenger RNA
nm	Nanometer
nt	Nucleotide
NTP	Nucleotide triphosphate
ODE	Ordinary differential equation
PFR	Plug-flow reactor
Phe	Phenylalanine
Pro	Proline
RF	Ribosome release factor
RNA	Ribonucleic acid
RNAP	RNA polymerase
rRNA	Ribosomal RNA
s	Second
Ser	Serine
SSA	Stochastic simulation algorithm
STR	Stirred-tank reactor
STY	Space-time yield
T3	Ternary complex consisting of tRNA, EFTu and GTP
TC	Transcription
Thr	Threonine
TL	Translation
t-Met f1	Formyl-methionine tRNA
tRNA	Transfer RNA
Trp	Tryptophan
TTP	Thymine triphosphate
Tyr	Tyrosine
UTP	Uracil triphosphate
Val	Valine
ZPS	Zellfreie Proteinsynthese (cell-free protein synthesis)

Symbols:

$\boldsymbol{A}$	Number of amino acid species	unitless
$\boldsymbol{C}$	Molar concentration	mol L^{-1}
$\boldsymbol{C_g}$	Codon index on gene	unitless
$\boldsymbol{D}$	Dilution Factor/rate, Diffusion coefficient	dimensionless
$\boldsymbol{d_i}$	Diffusive transport rate	mol L^{-1} s^{-1}
$\boldsymbol{d}$	Jump probability	s^{-1}
$\boldsymbol{h}$	Grid distance	nm
$\boldsymbol{k}$	First order reaction constant	s^{-1}
$\boldsymbol{k}$	Second order reaction constant	L mol^{-1} s^{-1}
$\boldsymbol{K}$	Last coding nucleotide	unitless
$\boldsymbol{K_M}$	Affinity	mol L^{-1}
$\boldsymbol{L}$	Length	nm
$\boldsymbol{N}$	Quantity	unitless
$\boldsymbol{p}$	Probability density	s^{-1}
$\boldsymbol{q}$	Queuing factor	Dimensionless
$\boldsymbol{r_{turnover}}$	Turnover of a species	s^{-1}
$\boldsymbol{R}$	Recovery	dimensionless
$\boldsymbol{T}$	Number of tRNA species	unitless
$\boldsymbol{t}$	Time	s
$\boldsymbol{V}$	Volume	L
$\boldsymbol{\dot{V}}$	Volume flow	L s^{-1}
$\boldsymbol{v}$	Reaction rate,	mol L^{-1} s^{-1}
$\boldsymbol{v_{elo}}$	Elongation rate	nt s^{-1}, amino acids s^{-1}
$\boldsymbol{x}$	RNAP position on operon	nt
$\boldsymbol{y}$	Ribosome position on mRNA	nt
$\boldsymbol{z}$	RNase position on mRNA	nt
$\boldsymbol{\nabla}$	Nabla operator	
$\boldsymbol{\alpha}$	Probability density	s^{-1}
$\boldsymbol{\mu}$	Growth rate	$hour^{-1}$
$\boldsymbol{\tau}$	Time increment	s^{-1}
$\boldsymbol{\epsilon}$	Elasticity	dimensionless

ZUSAMMENFASSUNG

Transkription und Translation sind Kernmechanismen des Lebens. Im Bereich der Biotechnologie werden diese Mechanismen weitläufig eingesetzt für die Proteinproduktion mit Anwendungsbereichen von Enzymen für Waschmittel bis zu Antikörpern für die Krebstherapie. Die Speicherung der Proteinsequenzen im Genom und der Abruf, sowie die Synthese der Zielproteine ist ein äußerst komplexes System, das auf die Interaktion vieler essenzieller Enzyme angewiesen ist.

In dieser Arbeit erfolgte die detaillierte Analyse von Transkription und Translation am einführenden Beispiel der zellfreien Proteinsynthese (ZPS). In weiteren Anwendungsfällen wird die Translations-Elongation genauer untersucht sowie der Einfluß großskaliger Bioreaktoren auf die Transkriptions-Translations-maschinerie. Im artifiziellen System der ZPS werden die Kernelemente der Proteinsynthese verwendet um Proteine zu synthetisieren ohne auf die Funktionalität der Wirtszellen angewiesen zu sein. Im Gegensatz zur *in vivo* Proteinsynthese liegt die volumetrische Produktivität jedoch um Größenordnungen niedriger.

In einem detaillierten Ansatz wurde der Effekt der Verdünnung während der Erstellung von ZPS Reaktionen analysiert. Basierend auf einem stochastischen Modell das den Transport der unterschiedlichen ternären Komplexen (Elongationsfaktor Tu (EFTu), GTP und tRNA) zu den entsprechenden Ribosomen beschreibt, konnte gezeigt werden, dass dieser Transportschritt einen der größten ratenlimitierenden Schritte der Translation *in vitro* und *in vivo* darstellt. Weiterhin konnte anhand des Modells gezeigt werden, dass die Optimierung der Ziel-Gensequenz einen sehr großen Einfluß auf die Transportraten besitzt. Eine homogene

Verteilung der Codons entlang der Gensequenz und eine homogene Nutzung der tRNA Spezies verhindert eine zusätzliche Transportlimitierung und ermöglicht höhere Translationsraten. Eine metabolische Kontrollanalyse der Transportlimitierung zeigte deutlich, dass bei optimierten Gensequenzen die reine Anzahl an Ribosomen ratenlimitierend ist, bei ineffizienter Codon Nutzung kommt es jedoch zusätzlich zu einer verringerten Translationsrate durch einen Mangel an ternären Komplexen. Diese Kontrolle zeigt sich auch bei der *in vivo* Proteinsynthese. Um die Translationsrate eines Proteins mit ineffizienter Codon Nutzung zu erhöhen ist es (bezogen auf die energetischen Kosten um die Translationsfaktoren zu synthetisieren) effizient sowohl in günstige ternäre Komplexe als auch teure Ribosomen zu investieren. Bei einer effizienten Codon Ausnutzung zeigt sich jedoch, dass es am effizientesten ist in teure Ribosomen zu investieren statt in günstige ternäre Komplexe.

Ausgehend von dieser detaillierten Betrachtung wurde der Modellumfang erweitert um die gesamte Translation zu beschreiben. Die Vergrößerung des Modellumfangs machte jedoch eine Reduzierung des Detailgrades unausweichlich. Durch die Anwendung eines dynamischen mathematischen Modells, konnten die Kernunterschiede zwischen der *in vivo* und der *in vitro* Proteinsynthese identifiziert werden. Eine Limitierung der Elongationsrate durch einen Mangel an ternären Komplexen ist hier der Hauptgrund für die geringere Translationsleistung. Durch die große Anzahl unterschiedlicher ternärer Komplexe und die starke Verdünnung bei der Erstellung von *in vitro* Reaktionen kommt es zu kumulativen Effekten. Die geringere Konzentration an Ribosomen in ZPS Reaktionen sind zusätzlich auch noch durch niedrigere ternäre Komplex Konzentrationen limitiert und die Translationsrate bricht ein. Basierend auf diesen Ergebnissen konnte ein neues experimentelles Setup etabliert werden, welches um 60 % erhöhte Translationsraten ermöglicht als das Referenzsystem. Die Erhöhung der Translationsrate durch Zugabe von EFTu und

tRNA ermöglicht jedoch nur eine begrenzte Verbesserung, da eine Verschiebung von Elongations- zu Initiationslimitierung stattfindet.

Bei der Kultivierung von Mikroorganismen im großen Maßstab auftretende Substratgradienten lösen eine Stressantwort der Mikroorganismen aus. Durch die Nutzung eines Ensemble Modells konnten die Mechanismen der stressinduzierten Transkription und Translation in einer maßstabsverkleinernden Reaktoranlage dargestellt werden. Trotz des vereinfachten Modellansatzes konnte eine qualitative Vorhersage der auftretenden Änderungen im Transkriptom der Mikroorganismen gezeigt werden. Die zeitlich sehr kurze Substratlimitierung sorgt für eine anhaltende Reaktion der Mikroorganismen welche zum großen Teil erst nachdem die Zelle die Limitierungszone wieder verlassen hat abläuft. Auf Basis dieses Modellansatzes konnten zusätzlich weitere Anpassungsmechanismen identifiziert werden. Auf Transkript Ebene kommt es bereits nach zehn Minuten zu einem stationären Zustand und einem konstanten Transkriptom für die Gesamtpopulation, das Proteom hingegen benötigt mehr als 15 Stunden um einen stationären Zustand der Population zu erreichen. Eine genauere Analyse der Population zeigt, dass es zu einer starken Inhomogenität kommt. Mehr als 65 % der Population sind mit der Stressantwort beschäftigt.

In dieser Arbeit konnte gezeigt werden, dass die Modellierung von Transkription und Translation je nach Anwendungsfall unterschiedliche Modellierungsstrategien benötigt. Die konstanten Bedingungen der ZPS erlauben einen sehr detaillierten Modellansatz, wohingegen die Beschreibung eines komplexen Bioreaktors sowie der entstehenden Stressantwort Reduktionen Vereinfachungen im Modell notwendig machen. Diese Balance zwischen Modellgröße und -komplexität sind der Schlüssel für eine erfolgreiche und effiziente Beschreibung von Transkription und Translation.

SUMMARY

Transcription and translation are key mechanisms of life. In the field of biotechnology, these mechanisms are often used to produce proteins for different applications, ranging from enzymes for detergents to antibodies for cancer treatment. Saving genetic information, and accessing and synthesizing target proteins depends on the complex interaction of many essential enzymes. In this work, transcription and translation were analyzed with an initial example of cell-free protein synthesis (CFPS). Further analysis covers the translation, elongation, and impact of large-scale bioreactors on the transcription-translation machinery. The artificial system of CFPS uses the key elements of protein biosynthesis to synthesize target proteins without relying on the functionality of the host cells. In contrast to *in vivo* protein synthesis, CFPS shows a volumetric translation rate orders of magnitude lower.

Using a detailed approach, the impact of dilution during the reaction preparation of CFPS systems was investigated. Based on a stochastic model that describes the transport of the ternary complex species (elongation factor Tu (EFTu), tRNA, and GTP) towards the suitable ribosome, this transport step was one of the key rate-limiting steps during *in vitro* and *in vivo* protein synthesis. Furthermore, optimizing the gene sequence highly influences the translation rate. A homogeneous distribution of codons along the genetic template and the homogeneous use of the different tRNA species prevents additional transport limitations and therefore allows higher elongation rates. The metabolic control analysis of the transport steps further highlights that for optimized gene sequences, the number of ribosomes limits translation rates, whereas genes with inefficient codon usage suffer from increased

elongation limitation caused by limitations of discrete ternary complex species. Such control behavior was also found under *in vivo* conditions. The most economical way to increase the translation rate of a gene with inefficient codon usage (based on the costs to synthetize the translation factors by the host cell) is to provide cheap ternary complexes and expensive ribosomes. However, for genes with efficient codon usage, it is economical to invest the energy in expensive ribosomes rather than inexpensive ternary complexes.

Continuing with this approach, the scope of the model was expanded to cover the entire translation process, which involved reducing the level of detail. By applying a dynamic mathematical model, the key differences between *in vivo* and *in vitro* protein synthesis were identified. The main reason for the lower translation rates is the limitation of ternary complexes, as the variety of tRNA species and the arising dilution during CFPS reaction preparation result in cumulative effects that lead to decreased translation rates. In addition, the lower concentration of ribosomes is combined with the ternary complex limited elongation, causing the collapse of the translation rate. Based on these results, a new experimental design with 60 % higher translation rates than the reference system was proposed. Adding EFTu and tRNA resulted in a finite increase in the translation rate due to a control switch from elongation limitation to initiation limitation.

During large-scale cultivations of microorganisms, substrate gradients are commonly caused by inefficient mixing-times, resulting in a cellular stress response of the organism. The cellular transcriptional and translational reactions were analyzed in a scale-down bioreactor by applying an ensemble model. Despite the reduced model complexity compared to the dynamic model, a qualitative prediction of a transcriptome change of the microorganism was possible. The rapid oscillating substrate limitation caused a sustained cellular reaction, which mainly elapsed after the cell left the limitation zone. Based on this modeling approach, two distinct

adaptation mechanisms were identified. On the transcript level, the bioreactor population reaches an overall steady-state after approximately ten minutes. On the other hand, the proteome of the population takes more than 15 hours to reach a steady state. A more detailed analysis of the reactor population revealed that more than 65% of the cells are undergoing the stress response caused by a short-term substrate limitation.

This work shows that modeling transcription and translation requires different strategies to describe the protein biosynthesis machinery based on the field of application. The constant conditions during CFPS reactions permit the use of a comprehensive model, whereas the description of the scale-down bioreactor requires a simplified model to describe the resulting stress response. The balance between model size and complexity is a key element for the successful and efficient description of transcription and translation.

CHAPTER 1

MOTIVATION AND OBJECTIVES

Protein biosynthesis relies on the guided polymerization of nucleotides and amino acids. Both transcription and translation (TC-TL) rely on enzymes for the correct chain elongation of mRNA and protein. Since protein synthesis is one of the key elements of reproduction, current knowledge of transcription and translation in the literature is vast. Despite the complexity of the reaction network, even small translation factors are identified and the respective interaction mechanisms are unraveled.

Although information about protein synthesis is substantial and rapidly growing, no model exists that describes all the TC-TL machinery. Even in well parameterized model organisms such as *E. coli*, the most complex model describes TC-TL based on *in vitro* protein synthesis. Most of the existing models describe the TC-TL of a single gene, whereas other models specialize in detailed mechanisms. For example the Zhang approach (Zhang et al. 2010) investigates the waiting times of ribosomes during translation elongation as a function of the coding sequence. However, TC-TL initiation and termination were neglected. There are models for different parts of protein synthesis and models for the entire TC-TL machinery, but there are no

models that describe the simultaneous protein synthesis of a complex microorganism. The existing models were developed with a given purpose in mind, such as the detailed analysis of codon usage impact or RNA polymerase (RNAP)-RNase interaction. Given the size of the biotechnological industry, a whole-cell TC-TL model to optimize protein synthesis rates is a promising approach.

Based on the following three models of protein synthesis with different levels of granularity, this study investigates the limitations and possibilities that are inherent to the description of protein biosynthesis: (i) a detailed model that describes translation elongation, (ii) a complex dynamic model that describes TC-TL for a single gene, and (iii) a simplified model that describes the cellular response to protein synthesis. With different levels of detail and numbers of mechanisms included, these models cover a broad range of modeling approaches to protein synthesis, as illustrated in figure 1.1. The following chapters provide a detailed motivation for the three approaches and present arguments for the different simulation strategies.

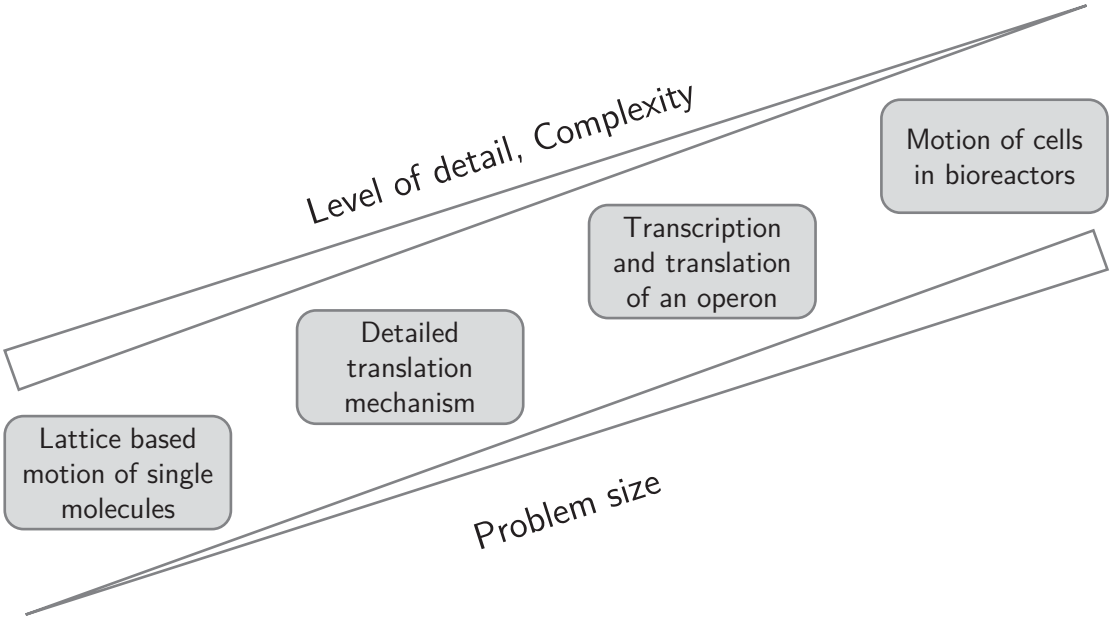


Figure 1.1: Schematic overview of this study. Different problem sizes ranging from single molecules that follow Brownian motion to an entire bioreactor population are shown. With increasing problem size, the level of detail must decline to -allow for simulation without increasing the computational effort beyond feasibility.

1.1 Diffusion-driven motion of translation factors

For achieving a high level of detail in a model, the fundamental physical effects that occur in the system must be incorporated, and the transport of molecules to the reaction site is essential for chemical reactions. The core mechanisms for the transport of molecules are convection and diffusion. Convection describes the direct transport of molecules, for example, by volumetric stream. Diffusion on the other hand is a random process underlying each molecular system, and depending on the convective velocity, diffusive transport is often neglectable.

The cytosol of prokaryotes lacks convective transport as a non-mixed system. Therefore, diffusion is the sole physical method of motion for intracellular components. While enzymatically catalyzed reactions in systems with high concentrations are often limited by the capability of the enzyme, reaction rates of diluted systems are often limited by substrate transport to the enzyme. This effect

can be explained by the fact that diffusive transport is proportional to the number of molecules. Therefore, systems with high numbers of molecules have higher diffusive transport than systems with low concentrations.

Many steps during prokaryotic translation are catalyzed by translation factors such as initiation factors (IF) and elongation factors (EF). These cofactors are essential parts of the translation machinery which enable proper functionality. Due to the motion of ribosomes in the reaction environment or along the mRNA template, these cofactors must be transported to their suitable reaction partner. Furthermore, the driving force for intracellular transport is diffusion, indicating that the diffusive transport of translation factors might be a rate-limiting step in translation. The diffusion of molecules in three-dimensional space can be simulated by either discretizing the space and allocating a concentration of molecules at each grid point, or treating each molecule as though it is wandering individually. The first approach is based on partial differential equations that describe the diffusive flux between two grid points as a function of the concentration gradient. Using a coarse grid to simulate the reaction space allows for a fast estimation of diffusive fluxes. In the second approach, each molecule is regarded as individual, which increases the number of treated objects by orders of magnitude. However, according to Dong et al. (1996), the number of tRNAs for certain species is below 1 µM (under *in vivo* conditions, and lower under *in vitro* conditions) and therefore, few molecules exist in one cell. At such a low concentration, the impact of single molecules increases. To address this aspect, a discrete simulation procedure was chosen instead.

By applying a fine grid and restricting the localization of the molecules to defined grid points, it is possible to track each molecule. Motion between two adjacent grid points can be based on probability functions, which use the diffusion coefficient and the distance between grid points to calculate the probability that motion will occur between two points. Using a stochastic simulation algorithm, numerical simulation

of single molecule motion can be performed. These algorithms are Monte-Carlo based, and during each iteration, randomly drawn variables define the type of action that occurs. In this case, the procedure defines which molecule wanders to which adjacent grid point. Since the computational effort scales with the number of objects, only systems with low numbers of molecules are calculable in an acceptable time. Using this approach, simulations used here are restricted to small compartments with a low number of molecules.

The diffusive haulage of molecules towards each other is the first step described here, followed by the collision between molecules and the subsequent reaction between them. Since the continuous model cannot describe single molecules, it is not possible to detect collision between molecules.

During this research, the use of a continuous model revealed the ternary complex, the complex responsible for amino acid transportation to the ribosome, as the main rate-limiting step during *in vitro* protein synthesis. The low concentration and distribution of tRNA in several species results in reaction kinetics that are limited by transport and not by the catalytic capability of enzymes. This part of the dissertation describes the detailed analysis of transport-limited reactions in the case of ternary complex (T3) haulage towards actively translating 70S ribosomes. The following targets are included in this chapter:

- A simulation platform must be established that can describe the three-dimensional motion of molecules based on a lattice. Furthermore, the model should describe collision-based reactions between potential reaction partners.
- With this model, the translation rate, as a function of the diffusive transport, must be evaluated for *in vivo* and *in vitro* proteisn synthesis.

1.2 Cell-free protein synthesis

The previous model regards protein synthesis as limited by the diffusive transport of ternary complexes. However, translation consists of many other mechanisms that are relevant to the protein synthesis rate. The scope of the model must therefore be increased to address all of the translation machinery, including EFTu, tRNA, and ribosomes, as well as all other translation factors such as IFs and termination factors. Given that the diffusion model with a limited number of reaction partners requires several computational hours to simulate one second of the reaction system, incorporating further molecules would render this system impossible to calculate.

Not all mechanisms of translation are indicated as limited by transport. Elongation factor G (EFG) for example, catalyzes the tRNA movement inside the ribosome and is not sequence dependent. Therefore, each individual EFG can act as a catalyst for translocation. By reducing diffusive transport towards a black-box approach, the computational demand of the model can be drastically reduced and the system can become calculable. Although reducing the complexity of mechanisms inherently reduces the investigative power of the approach, the model reduction permits the rejection of assumptions such as the constant number of actively translating ribosomes, which increases the overall detail level of the model.

In this chapter, the most simplified TC-TL apparatus in the form of cell-free protein synthesis (CFPS) is presented. CFPS systems provide a simplified platform for plasmid-based gene expression, and CFPS was the key element for deciphering the genetic code by Nirenberg & Matthaei (1961). Since then, the scope of CFPS has widely increased and it is commonly used as tool to express toxic proteins or to incorporate artificial amino acids into the polypeptide chain. In principal, CFPS reactions rely on extensively used protocols and only need the addition of a plasmid. With the addition of a plasmid, transcription starts and the synthesized mRNA allows the ribosome to produce the target protein.

The need for fast and reproducible protein expression tools increases according to the rising speed of DNA construction protocols (e.g. Gibson assembly). These tools must be accessible to be suitable for liquid handling robotic systems that automate the reaction execution. *In vivo* expression is not suitable for this case due to the high number of individual operations necessary for cloning, selection, and expression. In comparison, CFPS provides an all-in-one platform that is suitable for such cases. The expression of only the desired gene of interest (GOI) makes downstream processing easier and erases the need for cell-disrupting operations, allowing the entire process of gene expression to be automated with ease.

Current advantages in CFPS research, amongst other topics, involves the stabilization of the reaction system to achieve longer reaction durations. Key elements that increase the reaction duration are the adapted energy regeneration systems that lack byproduct synthesis. Byproducts such as acetate and lactate decrease the pH and destabilize the enzymatic machinery, while phosphate fixates Mg^{2+} ions. The concentration of the latter is crucial for optimal reaction performance, as Kim et al. (1996) demonstrated in their publication. In summary, the reaction duration was elevated from less than one hour (Pratt 1984; Kim et al. 1996) to approximately ten hours (Caschera & Noireaux 2014), yielding protein titers that linearly increase with the reaction duration.

The underlying reaction rates were poorly regarded until a recent study by Underwood et al. (2005) was published. Underwood and Swartz used a relative polysome quantification strategy to analyze the ribosome distribution over the mRNA template. They experimentally addressed the number of actively translating ribosomes and derived a mean elongation rate. These results were the first to indicate that ribosomes are an order of magnitude slower during *in vitro* protein synthesis compared to their *in vivo* counterparts. This indication of elongation-limited translation was further investigated by adding elongation factors to the reaction

setup, which revealed higher protein synthesis rates and a higher underlying elongation rate. They identified that CFPS is limited by translation elongation and not the number of actively translating ribosomes, and that the elongation rate can be increased by supplementing a mixture of all elongation factors.

At the same time, Underwood investigated the CFPS experimentally, and Arnold et al. (2005) investigated the system using a mathematical model. Their research included a vast literature search for model equations for TC-TL, and detailed mRNA degradation by exonucleases and endonucleases. The application of a sequence-oriented approach (i.e. each elongation step from a codon to the next codon is treated as an individual step) showed that transcription and mRNA degradation do not limit protein synthesis. The *in silico* addition of IFs and EFs further revealed that the average elongation rate of each ribosome limits the overall translation rate. Thus, both previously mentioned research groups independently identified translation elongation as a rate-limiting step in cell-free protein biosynthesis. Despite the identification of translation elongation as a rate-limiting step, the limiting factors in this process remain unclear. Therefore, the topic of this research is the detailed analysis of cell-free translation focusing on elongation and the regeneration of cofactors. The aims of this study are briefly outlined in the following section:

- First, published models must be evaluated and their usability for cell-free protein synthesis needs to be assessed, and the most suitable model must be implemented next. The model results can be verified by comparing them to experimentally observed protein synthesis rates.
- The functional and verified model can then be subjected to a sensitivity analysis that aims to classify the following aspects:
    1. The identification of targets to improve space-time yield
    2. The verification of predicted targets suitable for increasing the translation rate

3. The evaluation of CFPS potential by approximating *in vitro* synthesis rates and analyzing translation rates as a function of dilution
4. The application of metabolic control analysis to investigate the control of translation rates

1.3 Agent-based modeling of single cells in a stirred tank reactor-plug flow reactor population

With a validated model for the translation of a single gene, the next step is to increase the scope to include the simultaneous TC-TL of different genes. Increasing the number of genes in the dynamic model results in an enlarged system of ordinary differential equations (ODE), and the size is proportional to the number of codons regarded. CFPS systems are commonly saturated with mRNA to prohibit translation limitations, whereas *in vivo* protein synthesis is regulated by a limited number of mRNA. Consequently, the focus of this chapter is on the induction of transcription.

Biotechnological cultivations for production purposes are primarily performed in bioreactors with liquid volumes of larger than 50 m^3, and sometimes larger than 500 m^3 (Takors 2012). Production at these scales is often accompanied by decreases in product synthesis rates or unwanted byproduct formation (Lara et al. 2006). The main cause of these problems are the longer mixing times in large-scale bioreactors compared to lab-scale bioreactors. While lab-scale bioreactors achieve mixing times of less than ten seconds, large-scale bioreactors have mixing times longer than 60 seconds (Junker 2004). Due to the poor mixing conditions, inhomogeneity of the culture broth is likely to occur. Although mixing times can be decreased by changing the stirrer configuration or increasing the stirrer speed, this is strictly limited by the volumetric power input and stirrer tip speed. Increasing the stirrer speed inherently increases the shear stress on the microbial population. In conclusion, bad mixing is

an inherent drawback of large-scale bioreactors and its impact on bioprocesses must be investigated.

Recent computational fluid dynamic studies (Haringa et al. 2016) have shown that cells in large-scale bioreactors often travel between zones of different substrate concentrations, showing that substrate gradients are a problem in bioprocess upscaling. Thus, the treatment of these large bioreactors as ideally mixed systems is not possible. Large mixing times combined with process strategies that rely on feed strategies underlie the inevitable formation of substrate gradients. For example, glucose that is fed on top of the bioreactor induces a zone of high glucose concentration at the reactor head, whereas at the bottom of the reactor, a zone of glucose limitation forms. Cells that are currently at high glucose concentrations tend towards metabolic overflow, and by increasing glycolytic flux, the TCA and oxidative phosphorylation become rate limiting and mixed acid fermentation occurs. ATP regeneration by mixed acid fermentation has a significantly lower ATP per glucose yield ($Y_{ATP,S}$) than ATP regeneration by oxidative phosphorylation. In comparison to cells in high glucose concentrations, cells in low glucose concentrations have no overflow, as they tend to prepare their metabolism for substrate limitations. Both reactions include the upregulation of several genes that enable adaptation to high and low substrate concentrations. However, tracking of single cells reveals that transitions between different substrate gradient zones occur frequently (Haringa et al. 2016). For example, the protein formation induced by substrate limitations takes at least several minutes until the complete protein is available, though transitions can occur in the second time scale. Therefore, adaptation has a longer delay than the duration of these zones and repeatedly switching gene regulation on and off induces a metabolic burden to the cell.

Löffler et al. (2016) experimentally analyzed the impact of repetitive perturbation of substrate availability on an *E. coli* population in a scale-down bioreactor. They

identified a large set of genes that changed significantly compared to the cultivations without the perturbation. However, their experimental setup only allowed for the analysis of the entire population, and only average population dynamics were investigated.

This study expands on the knowledge from such scale-down experiments by providing a mathematical platform to track single cells in this reactor setup and a mechanistic-agent-based model to describe the cellular reaction towards substrate gradients. The following steps were performed in this work:

- The detailed transcription-translation model from chapter 1.1 is far too demanding in terms of computational efficiency to describe the reaction of a complete population. To overcome this limitation, a reduced modeling strategy was developed.
- Single perturbations of substrate availability were analyzed using an agent-based TC-TL model to describe the protein biosynthesis of an operon that is induced during nitrogen starvation.
- The influence of repetitive substrate perturbations on the population were simulated by coupling the agent-based model with a stochastic model to describe cellular motion in the stirred tank reactor-plug flow reactor (STR-PFR) bioreactor.
- Simulated transcript values were compared to experimental results for reaction along the PFR and for the reaction of the STR population. In addition, the influence on protein levels was analyzed by predicting the levels of synthesized polypeptides.

CHAPTER 2

THEORETICAL BACKGROUND

2.1 General modeling approaches

The following chapter presents an overview of the modelling type definitions summarized in figure 2.1 and used throughout this study. At first, models can be classified as structured and unstructured. Structured models describe systems and their inner mechanisms, whereas unstructured models regard a system as a whole. Models can then be classified as segregated and unsegregated. Unsegregated models describe a population as an average without reflecting single individuals, while segregated models describe the fate of each individual separately. Further discrimination can occur based on the underlying mathematical rules applied to modelling. Unstructured and unsegregated models benefit from a small number of equations to solve; they often rely on parameters estimated for a defined scenario, rendering them less transferable (e.g. product synthesis rate and yields). Structured and segregated models are based on detailed models of certain mechanisms and they rely on parameters that are transferable, such as enzyme affinities or catalytic rates. However, these parameters are more difficult to identify.

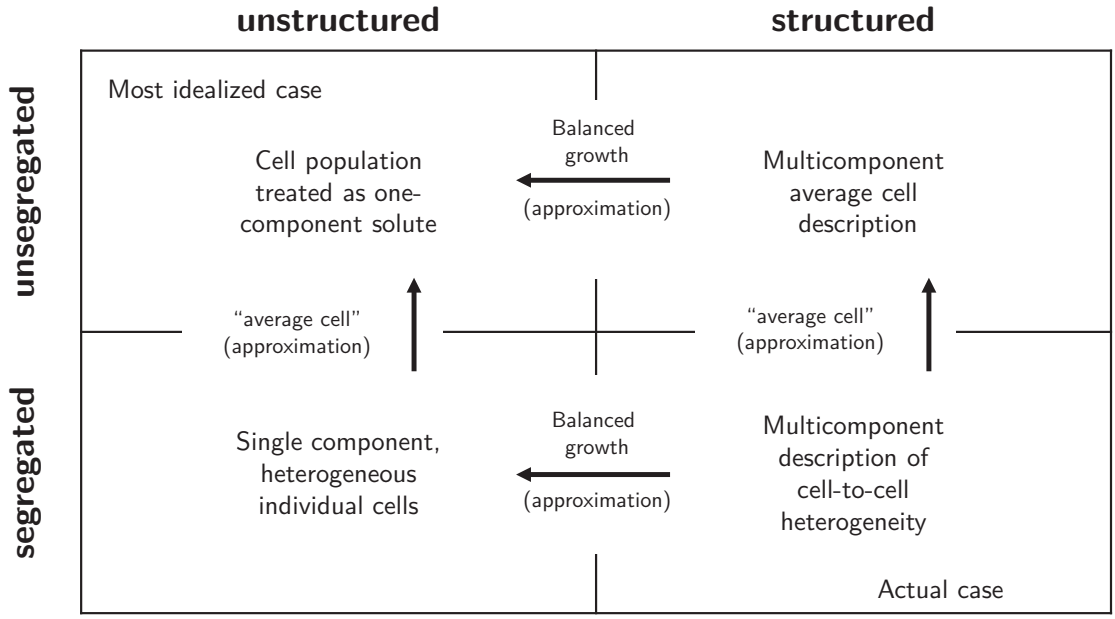


Figure 2.1: Different levels of detail for modelling the kinetic behavior of a cell population (adapted from Bailey & Ollis; (1986)).

There are two approaches to glucose consumption during cultivations. An unstructured and unsegregated model results in a simple glucose uptake rate equation, and the only necessary parameters are the maximum uptake rate and the affinity towards glucose. In comparison, a structured and segregated model describes the entire glycolysis process with reactions for each enzyme (Chassagnole et al. 2002) and therefore, a set of parameters is included for each enzyme. Changes in substrate accessibility for each cell must be regarded as well. Thus, the second model results in a drastically higher number of equations and parameters, increasing the computational effort. However, such a model can describe the reaction of cells more accurately and can be transferred to other conditions, such as deletion strains, where single reactions are knocked out.

The main disadvantage of structured and segregated models is scaling. A comprehensive model describing cellular mechanisms must be implemented and each

cell must be described individually, resulting in a quadratic scaling problem. Reducing one dimension immediately results in linear scaling with the model's level of detail, making it more manageable. Reaction systems with a small number of molecules show a significant increase in noise compared to systems with a large number of molecules. This noise is due to the underlying randomness in the reactions. According to the law of large numbers, it is possible to discriminate between these systems.

2.1.1 Law of large numbers

Despite the name, the law of large numbers is a theorem, not a physical law, which describes the result of continually repeating the same experiment. With a rising number of performed experiments, the average values of the results will tend towards the expected value if many experiments are performed. The distinction between many and few experiments is dependent on the underlying system. As the number of experimental outcomes increases, more experiments must be performed to fulfill the theorem.

Diffusion processes show the application of the law of large numbers. With a low molecule count in a reaction compartment, the local concentrations of the molecules differ, describing a heterogeneously mixed system. However, with rising molecule concentration, the local concentrations are less variable. With a large quantity of molecules, the noise of single molecule motion becomes irrelevant and the system is homogeneously mixed.

2.1.2 Discrete models

Discrete models treat a model as defined by distinct states, while model variables are only allowed to switch between states. For example, the movement of ribosomes is often treated as a discrete system. Each codon defines a state in which a ribosome

can be positioned and transitioning positions between two codons is strictly prohibited. One of the main advantages of such a modeling strategy is the reduced set of system variables. MacDonald et al. (1968) and Heinrich & Rapoport (1980), for instance, used a one-dimensional lattice to describe the kinetics of transcription and translation, respectively. Using a discrete modelling approach, both groups defined the foundations off a mechanistic description of protein biosynthesis.

Another group that relies on discrete models are stochastic simulations. These simulation types are often Monte-Carlo based, meaning that their numerical solution is based on repeated random sampling, which makes these algorithms costly to calculate. Therefore, discrete modelling is necessary to reduce the computational expenses. For example, petri nets (Hofestädt & Thelen 1998), which are often used to describe biochemical networks, provide a platform for the discrete description of systems. Based on this approach, Goss & Peccoud (1998) developed stochastic petri nets that incorporate probability distributions into the reaction network and can therefore predict noise in the underlying enzymatic system.

Discrete systems are often used in systems with few reactants where all molecules must be considered. Switching to systems with high molecule counts, the law of large numbers allows for the generalization of species or mechanisms, which results in the continuous modelling approaches that are described in the following section.

2.1.3 Continuous models

The law of large numbers indicates that noise in a given system can be neglected if the number of participants is high. Based on this assumption, model reduction can be performed by treating all molecules as one population rather than each individual molecule. Continuous models often rely on differential equations (ODE) or differential-algebraic equation (DAE) systems. ODEs describe the change of a given system variable over time allowing tracking, while DAE systems are generally more

demanding due to the algebraic equations that must be solved for each time increment. In addition, algebraic equations often introduce stability problems because consistent initial values must be provided.

Continuous models that describe TC-TL often rely on black-box approaches that treat the intracellular reaction mechanisms as irrelevant or unknown. Using the available knowledge of the system, a reduced set of kinetics such as Michaelis-Menten-like kinetics can be applied. Adapted parameters are required for this system configuration, reducing the predictability of the given model for other use cases. Two current examples of continuous black-box models for TC-TL were provided by Stögbauer et al. (2012) and Chizzolini et al. (2017). Both groups used simplified kinetics to describe protein biosynthesis based on a parameter estimation for their given experimental setup. However, the parameter estimation circumvents the transfer of their models to other setups, and estimating the underlying mechanisms that are limiting the reaction capability is difficult.

One of the major limitations of continuous models is the lack of discrimination between molecules. For example, during translation, a continuous model only describes the number of ribosomes currently located at a given codon. Direct interaction between consecutive ribosomes can therefore only be described as a probability. In comparison, discrete models link ribosomes to an mRNA, such that the interaction between the two molecules can be directly implemented into the model. In summary, continuous models enable the detailed description of a system with the drawback of high computational effort. Depending on the focus of research, detailed modeling of all system components is not always necessary. Here, hybrid models provide a balance between the level of detail and resource costs.

2.1.4 Hybrid Models

Hybrid modelling describes a mixture of discrete and continuous modelling approaches. Using continuously modelled variables combined with discrete state transitions reduces simulation expenses without describing system parts unnecessarily for the model. Inspired by hybrid models, the description of TC-TL provides a functional example of such modelling strategies. By combining the previously mentioned one-dimensional lattice discrete model and overlaying the continuous description of a molecule species (e.g. ribosomes) a detailed dynamic delineation of the protein biosynthesis is possible.

2.1.5 Deterministic versus stochastic

Based on the approach that was selected for modeling the reaction system, different simulation strategies can be chosen. A deterministic simulation approach assumes a continuous system, and a large number of molecules allows for the generalization of the system and the calculation of an average time course. In comparison, stochastic approaches do not rely on this assumption, and their fundamental calculation of change in molecules is not based on a rate, but on the probability that the reaction will be fired. Deterministic models rely on kinetic equations for all reactions (e.g. Michaelis-Menten or mass-action kinetics) and each simulation with the same parameters results in the same output. In contrast, stochastic approaches are based on reaction probabilities p for each reaction. For a given degradation reaction of molecule A,

$$A \xrightarrow{k} \emptyset$$

a mass-action-law derived kinetic rate (cf. equation 2.1) and a probability function is shown (equation 2.2). The units of degradation constants k_{deg} and the concentration C_A are equal in this case, though the units of the reaction rate and the probability function may differ.

$$v_{deg} = k_{deg}^{det}\, C_A \tag{2.1}$$

$$p_{deg} = k_{deg}^{prob}\, C_A \tag{2.2}$$

Both approaches must yield the same results for this reaction with the same initial values C_A^0. The deterministic approach, based on the assumption of a continuous system, results in an average concentration, whereas the stochastic approach results in a different time course for each simulation. However, the average concentration profile over several simulations is the same, as given by the deterministic approach. Furthermore, a stochastic approach does not neglect the noise of the system and the standard deviation of the system can be identified. Although stochastic models can identify the noise of a system, this additional information requires much more simulation effort due to the nature of the underlying algorithms and the need for several runs with different random number generator seeds.

2.2 Stochastic simulation algorithm

One of the first numerical approaches to solving the stochastic system was Gillespie's stochastic simulation algorithm (SSA; (Gillespie 1977). Although there are several more advanced algorithms that use current multicore central processors (Gibson & Bruck 2000; Gillespie 2001; Li et al. 2008; Komarov et al. 2012) or graphical computing units (Lee et al. 2010; Klingbeil et al. 2011), the SSA is still widely used due to its ordinary implementation. The fundamental mechanism of the

SSA is that a time interval is chosen such that only a single reaction is fired during the selected interval. In the example above, one molecule of species A vanishes at each time step. Choosing the fired reaction can be described based on the probability functions for each reaction. Due to the stochastic nature of these systems, the time interval in which this reaction is fired is a key element that must be chosen such that only a single reaction is fired. The resulting equations for calculating τ and choosing the reaction rely on random numbers.

Due to the nature of computers, random numbers cannot be calculated and therefore algorithms are used to calculate pseudo random numbers (PRN). One of the most efficient ways to generate such PRNs is the *MT 19937* Mersenne-Twister (Matsumoto & Nishimura 1998). This algorithm is not only efficiently implementable as a code, it provides a long period of $2^{19937} - 1$ independent PRNs before repetition occurs. Due to the simplicity of the underlying calculations, this code is parallelizable on modern computers.

2.3 Transcription and Translation

This chapter provides a short excerpt of the mechanisms of TC-TL regarded by the different models in this study.

Transcription is the first step of protein biosynthesis and it can be divided into the following three parts: initiation (describes the binding of RNAPs to DNA), elongation (describes the repetitive motion of RNAP from one nucleotide to the next and the inherent prolonging of the mRNA chain), and termination (after reaching a stop codon, the RNAP separates from the template and the resulting mRNA is released). Transcription initiation describes the association of the RNAP to the corresponding binding motive on the DNA. The RNAP then elongates from nucleotide to nucleotide from 3' to 5' on the DNA, and the mRNA is synthesized from 5' to 3'. During each elongation cycle, the corresponding nucleotide triphosphate

(NTP) is incorporated into the mRNA chain and catalyzed by the separation of pyrophosphate from the NTP. NTPs are the exclusive substrate for transcription elongation. After reaching a stop codon, where no corresponding NTP can be incorporated, the RNAP dissociates from the DNA, releasing the mRNA strand.

The cascade of reactions occurring during translation can be divided in the same way as in transcription. Due to the higher number of substrates and enzymes involved, translation is a far more complex system than transcription. Compared to transcription, translation initiation is catalyzed by three IFs. During the first step, the 30S ribosomal subunit binds the three IFs in a random order. This 30S IF complex can then bind mRNA and fMet-tRNA to form the 30S IC, which is a preinitiation complex. Hydrolysis of IF2 guanosine triphosphate (GTP) and binding of the 50S subunit is the last step during initiation; the ribosome releases all IFs and elongation can occur.

Translation elongation is a complex repetitive system which relies on a set of elongation factors. The first step during each elongation cycle is the transport of amino acids to the ribosome. Loaded tRNA (aa-tRNA) complexed with GTP-bound elongation factor Tu (EFTu) forms a ternary complex, which is the key element for transporting amino acids to the corresponding ribosome. GTP hydrolysis catalyzes this reaction step and forces the dissociation of the ternary complex. The large set of varying tRNAs for the different codons on the mRNA results in the same diversity among the T3 complexes. This step is the critical for sequence-oriented translation, where the correct amino acid is incorporated into the growing peptide chain. After transporting the amino acid to the ribosome, GTP-bound elongation factor G (EFG) forces the tRNAs inside the ribosome to move three nucleotides downstream, moving the ribosome upstream to the next codon. After each elongation step, EFG and EFTu are regenerated. No catalyst is necessary for EFG, but EFTu regeneration is catalyzed by elongation factor Ts (EFTs). The depleted tRNA is regenerated by

aminoacyl-tRNA-synthetases, which are sensitive to the corresponding tRNA species. During termination, the ribosome is pushed forwards by EFG to a stop codon where no matching tRNA is incorporated. Catalyzed by ribosome release factors, the ribosome dissociates from the mRNA and releases the remaining tRNA and the completed polypeptide chain.

2.4 Ribosome flow model

Large biomolecules such as mRNA and proteins are often described as biopolymers which consist of a defined sequence of monomers. The synthesis of such molecules is catalyzed by enzymes that are guided by a template. This lattice-based polymerization of biomolecules was initially described by Pipkin & Gibbs (1966), explaining the template-driven process of synthesizing biopolymers. In contrast to random polymerization where the polymer is expanded in a random direction, template-driven polymerization is a controlled process. Therefore, the random three-dimensional system is reduced to a guided one-dimensional system where chain elongation occurs in a single direction, controlled by template and catalyzing enzymes.

TC-TL belong to this type of polymerization process. DNA and mRNA, as the corresponding templates, provide guidance and the sequence in which monomers are added to the lengthened polymer. Gibbs et al. (Pipkin & Gibbs 1966; MacDonald et al. 1968; MacDonald & Gibbs 1969) expanded this approach with the step-wise motion of the catalyzing enzymes along the given template, and provided a theoretical analysis of translation mechanisms, though the application to cell-based translation was not performed. Moreover, Heinrich & Rapoport (1980) analyzed the steady-states of this model; identified the control of initiation, elongation, and termination on the resulting translation rate; and provided the first time-series simulations for TC-TL.

The first to successfully couple the ribosome flow model with experimentally observed translation rates was Arnold et al. (2005). The key element of their model is the sequence-oriented description of TC-TL. Sequence orientation in this context means that each individual codon of DNA and mRNA is regarded individually, and elongation describes the motion of RNAP or a ribosome to the subsequent codon. Therefore, the number of ribosomes currently translating a codon can be calculated. The impact of the sequence can also be incorporated into the elongation equations (e.g. the number of available tRNAs of the corresponding species). The ribosome flow approach coupled with many detailed reaction kinetics from the literature, as well as the estimation of missing parameters, allowed them to successfully model the CFPS synthesis of a single target gene. Based on the model of Arnold, the group of Hatzimanikatis enhanced the level of detail regarding elongation (Mehra & Hatzimanikatis 2006; Zouridis & Hatzimanikatis 2007, 2008). Each step, from binding of the corresponding tRNA to transpeptidation and exit of the free tRNA was regarded as an individual step. The resulting model consists of ten individual reactions per elongation step and provides insight into the stability of TC-TL. However, the description of experimental observations is hampered by the need for parameters. Although protein biosynthesis is one of the key elements of life, the number of application models for TC-TL is marginal.

2.5 Scale-down approach

Microbial production processes are often controlled by limited substrate addition to prevent overflow metabolism and metabolic activity beyond the technical capabilities of the reactor system, such as aeration or cooling. The design of such large production processes is regularly implemented in small lab-scale bioreactors, though differences in mixing times prohibit direct transfer from lab to production scale. A strategy for modelling large-scale heterogeneities in the culture broth with

lab bioreactors is the coupled STR-PFR system (George et al. 1993). This system permits the deflection of the main STR culture in the PFR to simulate substrate oscillations (figure 2.2 provides an overview of the scale-down approach). Löffler et al. (2016) and Simen et al. (2017) used this approach to simulate the impact of short starvations on the microbial population. Here, the STR operates in substrate-limited chemostat mode with feed and harvest, and the PFR loop introduces a substrate-limited zone where cells enter starvation. This setup allows for the investigation of short-term responses along the starvation zone to analyze the direct reaction of the cells. Furthermore, the chemostat operation provides a steady state as a reference for analysis, and allows for the analysis of the long-term impact of oscillating substrate conditions.

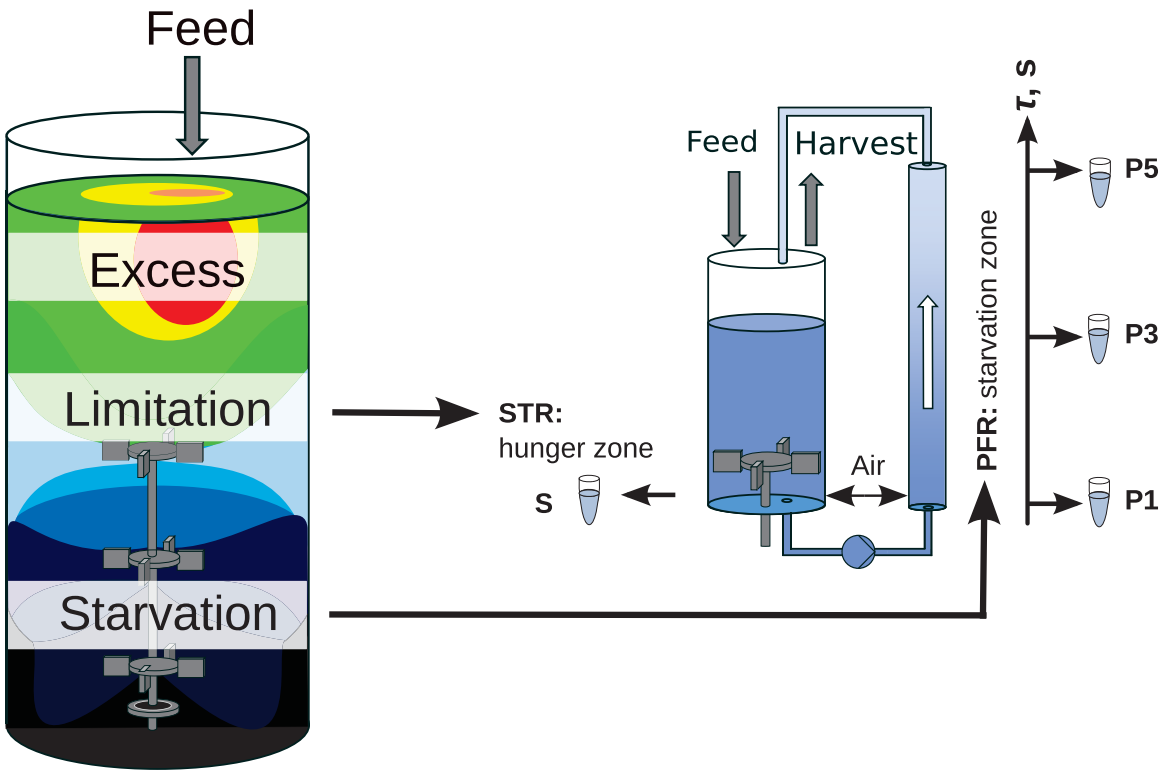


Figure 2.2: Scheme of the scale-down approach. A large-scale bioreactor and its substrate gradient is simulated by a chemostat STR (limitation zone) coupled to a PFR (starvation zone). This approach examines the influence of substrate gradients (i.e. glucose or ammonia) on the population dynamic (Nieß & Löffler et al. 2017).

CHAPTER 3

MATERIAL AND METHODS

3.1 Diffusion Model

As reported by Nieß & Failmezger et al. (2017), the ternary complex distribution and concentration are the key elements that limit *in vitro* protein synthesis. Given the challenge of further analyzing the cause of this limitation, a translation elongation model was established that solely relies on diffusive transport as a rate-limiting step in translation elongation. Due to the lack of convection in CFPS systems and living cells, diffusion is the main driving force of transport. Most diffusive problems are simulated with the reaction-diffusion equation as shown in equation 3.1. The change of concentration equals the sum of the diffusive term and the reaction term, though this approach is only suitable for systems with many reactants. In the case of translation elongation, the number of actively translating ribosomes is low and the high number of ternary complexes is divided into many species. Therefore, the rule of large numbers is not met.

$$\frac{\partial C}{\partial t} = D\ \nabla^2\ C + r \tag{3.1}$$

Another approach to describing diffusion is the lattice-based random walk, which treats each molecule individually on a three-dimensional grid. Movement is only allowed between adjacent grid points and based on the probability that a molecule travels from point *j* to *j+1.* Regarding each individual module and species makes this model a structured model, and neglecting population heterogeneity makes it an unsegregated model as shown in figure 2.1.

3.1.1 Derivation of jump probability

The probability that a molecule will jump between two grid points can be derived from the Smoluchwoski equation (equation 3.2) with $p(\vec{x}, t)$ as the probability density function of a random variable *X*.

$$\frac{\partial p(\vec{X}, t)}{\partial t} = -\nu \frac{\partial p(\vec{X}, t)}{\partial x_i} + D \frac{\partial^2 p(\vec{X}, t)}{\partial x_i^2} \tag{3.2}$$

Neglecting convection ($v = 0$) and regarding a single dimension results in the following equation:

$$\frac{\partial p(\vec{X}, t)}{\partial t} = D \frac{\partial^2 p(\vec{X}, t)}{\partial x^2} \tag{3.3}$$

In addition, discretizing the partial differential equation with central differences results in an ordinary differential equation,

$$\frac{dp(\vec{X}, t)}{dt} = D \frac{p_{i+1} - 2p_i + p_{i-1}}{\Delta x^2}, \tag{3.4}$$

where N_i describes the average number of molecules currently resting on position *j*. The mass balance of N_i is therefore:

$$\frac{dN_i}{dt} = \frac{D}{\Delta x^2}\ (N_{i+1} - 2\ N_i + N_{i-1}). \tag{3.5}$$

Based on the reaction scheme in figure 3.1, with the jump probabilities (d_i) between the adjacent grid points *i-1*, *i,* and *i+1*, a net reaction can be derived as shown in equation 3.6.

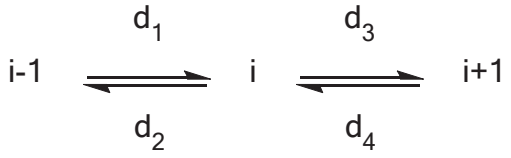


Figure 3.1: Diffusive transport reaction constants d_i for diffusive motion between point i and two adjacent points *i±1*.

$$\begin{aligned} &\frac{dN_i}{dt} = d_1 + d_4 - d_2 - d_3 \\ &\text{With } d_i = d\ N_i \\ &\frac{dN_i}{dt} = d\ (N_{i-1} - 2\ N_i + N_{i+1}) \end{aligned} \tag{3.6}$$

$$\widehat{=} \frac{D}{\Delta x^2}(N_{i-1} - 2\ N_i + N_{i+1}) \tag{3.7}$$

The specific jump probability can therefore be calculated based on the diffusion coefficient and the distance between the grid points as follows:

$$d = \frac{D}{\Delta x^2} \tag{3.8}$$

Equation 3.8 calculates the probability that molecules will jump between two grid points, and it can be used to calculate the jump probabilities for all three dimensions.

3.1.2 Random walk

The foundation of this model is a lattice-based random walk. The molecules that are considered are ternary complexes made from tRNA, EFTu, and ribosomes. The transport of ribosomes is neglected due to their much lower diffusion coefficient, and the reaction space is a three-dimensional lattice with a grid distance of h. A collision between two molecules only occurs if they touch each other, which implies that their distance is closer than the sum of their radii. A collision on the grid is defined as two molecules being on the same point. Thus, N_a was chosen to be equal to the sum of the radii of ternary complexes and ribosomes. The probability (α_i) of moving from a point towards one of its six adjacent points is defined as

$$\alpha_{travel,i} = \frac{D}{h^2} \; N_i \tag{3.9}$$

where D_i is the diffusion coefficient of molecule i and N_i is the number of molecules (i) at this position. The direction in which the molecule travels is chosen randomly with probabilities set equally (1/6 for each direction) using Gillespie's direct stochastic simulation algorithm (Gillespie 1977). Total diffusion probability α_0 is defined as the sum of all α_i

$$\alpha_0 = \sum_{i=1}^{N_{molecules}} \sum \alpha_{travel,i} \tag{3.10}$$

and $N_{molecules}$ is the number of different possible ternary complexes. The time increment is calculated as shown in equation 3.11:

$$\tau = \frac{1}{\alpha_0} ln \left(\frac{1}{r_1}\right) \qquad 3.11$$

In addition, the traveling molecule species j is selected as the smallest integer that fulfills equation 3.12,

$$\sum_{i=1}^{j} \alpha_i > r_2 \;\; \alpha_0 \qquad 3.12$$

where r_1 and r_2 are random numbers from a uniform distribution between zero and unity, calculated with a Mersenne Twister (Matsumoto & Nishimura 1998; Saito & Matsumoto 2008). After each step, the global time (t) is incremented by τ. The next necessary index is the molecule index k, which describes the discrete molecule of species j that wanders; it is chosen as a random integer between 1 and N_j (N_j is the number of molecules of species j).

The chosen molecule $N_j(k)$ then travels to one of its neighbor grid points, and the direction follows a random distribution between 1 and 6. Each movement is followed by a check-up where the new position N_{new} is scanned for possible reaction partners (in this case a ribosome with a matching anticodon). If there is no reaction partner, the algorithm moves to the next increment. If there is a possible reaction partner, the molecule update sequence is called where the ternary complex is split and the free EFTu is instantly bound to one of the free tRNAs forming a new ternary complex that is randomly relocated in the reaction space. The ribosome elongates one codon and cannot react for a timespan of t_{cat}, which correlates to the time required to refold the ribosome and prolong the peptide sequence. It is calculated as the reciprocal of the maximum specific elongation rate (24 amino acids per ribosome

per second according to Arnold et al. (2005)). During this idle time, the ribosome is not allowed to react further.

An additional phenomenon included in this model is the dissociation of ternary complexes. The probability of dissociation follows first-order kinetics as described in equation 3.13

$$\alpha_{diss,i} = N_i \; k_{diss} \tag{3.13}$$

where k_{diss} is the reaction constant for the dissociation of ternary complexes and is set to 1 s^{-1}, (Gast 1987). This expansion leads to a new equation for the total probability:

$$\alpha_0 = \sum_{i=1}^{N_{molecules}} \sum \alpha_{travel,i} + \sum_{i=1}^{N_{molecules}} \sum \alpha_{diss,i} \tag{3.14}$$

If the index j in equation 3.14 is selected as higher than the probability of travel, dissociation occurs instead of travelling. Dissociation is followed by choosing the species j and molecule $N_j(k)$. This molecule is then split and its underlying tRNA is added to the pool of free tRNA. The released EFTu binds a randomly selected free tRNA and is relocated to a random grid point.

The initial molecule distribution is set randomly for ternary complexes and ribosomes and the initial states of the ribosomes are uniformly distributed throughout the entire sequence. Translation, termination, and initiation were neglected in this model and ribosomes reaching the end of the sequence were set to the first codon. Furthermore, the entire calculation is nested in a loop with the stop criterion set to 5,000 successful elongation steps. The specific elongation rate is calculated as the slope of step number over their respective time points. The resulting slope

(elongations per second or amino acids per second) is normalized on the ribosome count, resulting in the specific elongation rate (elongations/amino acids per second per ribosome). The corresponding error is based on the deviation between ten different simulation runs with varying seeds for the random number generator.

The numbers of actively translating ribosomes (ribosomes during elongation) and EFTu are shown in table 3.1 and the number of tRNAs for the different species are shown in table 3.2. For simulation purposes, a reaction volume of 0.064 µm³ was chosen, which results in at least 80 tRNAs in the reaction volume and an even number of grid points. The diffusion coefficient for all ternary complexes was set to $D = 2.567\ 10^{-12}\ \mathrm{m^2\ sec^{-1}}$ (Fluitt et al. 2007) and diffusive transport for ribosome-mRNA complexes was neglected due to the drastically lower diffusion coefficient compared to ternary complexes.

Table 3.1: Initial number of actively translating ribosomes and EFTu during the simulation of translation in a reaction compartment with $V = 0{,}064\ \mu\mathrm{m}^3$. The number of molecules was calculated for concentrations at a given growth rate of $\mu = 1.1\ \mathrm{h}^{-1}$.

Molecule	Number	Source
Ribosomes	1,044	Rudorf & Lipowsky (2015)
EFTu	9,122	Rudorf & Lipowsky (2015)

Table 3.2: Quantity of tRNAs during the simulation of translation in a reaction compartment with $V = 0.064\ \mu m^3$. Derived from Dong et al. 1996 for growth rate of 1.1 h^{-1}.

tRNA species	Quantity	tRNA species	Quantity	tRNA species	Quantity
Ala1B	675.2	Gly3	764.6	Pro3	98.3
Ala2	122.9	His	129.1	Ser1	269.0
Arg2	916.1	Ile	729.2	Ser2	52.8
Arg3	87.1	Leu1	821.7	Ser3	208.1
Arg4	125.6	Leu2	181.9	Ser5	141.8
Arg5	94.8	Leu3	122.9	Thr1	21.6
Asn	235.1	Leu4	372.3	Thr2	102.9
Asp1	464.0	Leu5	140.7	Thr3	187.3
Cys	271.3	Lys	336.5	Thr4	192.3
Gln1	122.2	Met	158.0	Trp	159.9
Gln2	195.4	Phe	180.8	Tyr12	378.9
Glu2	929.6	Pro1	106.0	Val1	731.9

3.1.3 Costs to synthesize nucleotides

The costs of the nucleotides that are necessary for RNA synthesis are described in table 3.3. Amino acid costs were taken from Kaleta et al. (2013) where precursor costs were neglected. Taking the sequence and the costs of each amino acid into EFTu led to production costs of 1,989 ATP equivalents to synthesize the amino acids and 1,576 ATP equivalents for translation (4 ATP eq. per step). The overall synthesis cost for EFTu was therefore 3,565 ATP equivalents per molecule. Ribosomes are composed of three types of RNA (5S, 16S and 23S) which have lengths of 120, 1,542, and 2,906 nucleotides. Considering these species and their respective sequence result in costs of 42,934 ATP equivalents to synthesize the rRNA of the ribosome. The protein content of ribosomes consists of 7,459 AA with an average amino acid distribution from Spahr (1962), which leads to costs of 35,689 ATP equivalents to

synthesize the amino acids and 29,836 ATP equivalents for translation. Thus, ribosome synthesis requires 108,461 ATP equivalents per single molecule. tRNAs are an average of 76 nt in length, and an average sequence costs approximately 700 ATP equivalents per tRNA.

Table 3.3: Energy costs to produce the five different nucleotides based on the stoichiometric pathways of *E. coli*.

	ATP equivalents
UTP/TTP	6
CTP	7
ATP	11
GTP	12

3.2 Coupled transcription-translation model

The model described in this chapter is based on a former study by Arnold et al. (2005) with several adaptations, and it is published in Nieß & Failmezger et al. (2017). This model is classified as a structured and unsegregated model, as shown in figure 2.1, that describes the transcriptional and translational machinery of a CFPS system with a hybrid level of detail. While the transcriptional part was reduced to a black-box model (according to Arnold et al., CFPS is limited by translation not transcription), the translation model is a structured model that incorporates the sequential information of the gene template. Since CFPS reactions consist of cell lysates that are well mixed, no segregation was required to describe this system.

This chapter provides a detailed overview of the key characteristics of this model. Describing translation with a codon-based approach allows this model to act as platform for describing CFPS and identifying reactions that hamper the translation rate. The core is the detailed description of translation including the influence of initiation, elongation, and termination. The description of elongation has a high level

of detail to account for the gene sequence and codon usage. Figure 3.2 shows a schematic overview of the underlying mechanisms used by the dynamic model.

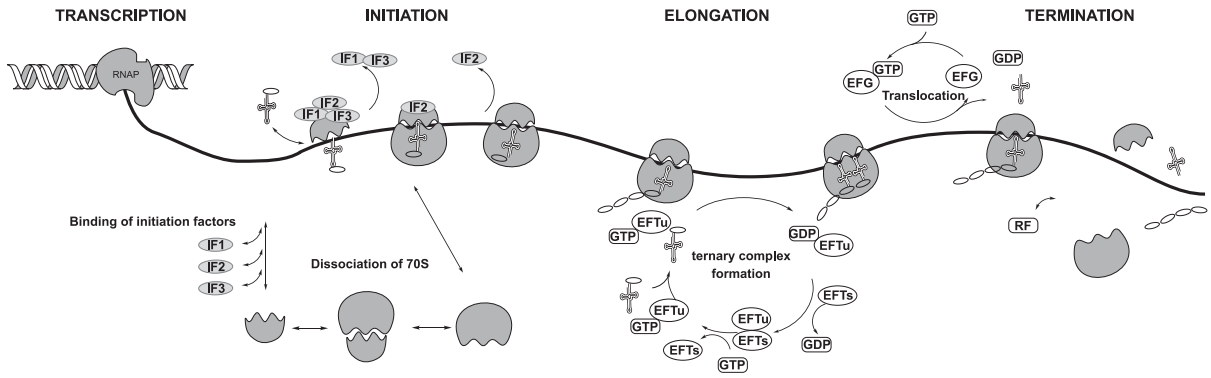


Figure 3.2: Overview of cell-free protein synthesis and reaction composition of S30 extract and added supplements. An additional schematic overview of the model's mechanisms is shown, starting with T7 RNA polymerase and subsequent sequence-oriented translation as well as the regeneration of co-factors. Abbreviations: RNAP = RNA polymerase; IF = initiation factor; EFTu = elongation factor Tu; EFTs = elongation factor Ts; EFG = elongation factor G; RF = release factor; NTP = nucleotide triphosphate; GTP = guanosine triphosphate; GDP = guanosine diphosphate (Nieß & Failmezger et al. 2017).

Compared to the previous model by Arnold et al. (2005), this model was implemented using Matlab instead of Maple V, acslX, fortran, and OptdesX, which has several advantages. First, only a single software solution is necessary compared to the previous work. Second, Matlab grants access to a large library of algorithms (e.g. solver for DAE systems, and vector and matrix operations) that are already compiled for current x64 processors. These libraries permit the use of computing clusters for more demanding calculations such as sensitivity analysis. In addition, the model was built as a framework where equations are packed into modules that can easily be exchanged or switched on and off. This versatility allows for the screening of different equations to determine their usability and switching off entire reaction cascades is possible. For example, the energy regeneration system is modularized to

easily simulate the impact of different regeneration pathways such as glycolysis or the creatine-kinase system.

The following model for protein biosynthesis is based on the comprehensive model described by Arnold et al. (2005). Using this approach, they simulated their experimental results and elucidated underlying reactions that are difficult to observe in experiments. However, the large set of equations resulted in a complex model that is challenging to solve and is not suitable for sensitivity analysis. Based on their findings that translation is the rate-limiting step in CFPS, the transcriptional part and mRNA degradation were significantly reduced to decrease the computational effort. The sequence-oriented transcription model was replaced by Michaelis-Menten-like kinetics and mRNA degradation was set to follow first-order kinetics. Based on these reductions, more than 500 differential equations could be eliminated, depending on the gene length of the target protein.

3.2.1 Transcription

Michaelis-Menten kinetics for modelling T7 polymerase activity were used to describe transcription in a black-box approach (equation 3.15), considering the plasmid (C_D) as the only substrate with the binding affinity ($K_{M,D}$) reported by Arnold et al. (2001) and a maximum activity of $0.0519\ \mu M\ min^{-1}$ (the manufacturer declaration of T7 polymerase was used). Other influences such as nucleotide triphosphates (NTPs) were neglected because their concentrations exceeded the affinity for RNA polymerase. Throughout this study, C_i describes the volumetric concentration of molecule i.

$$v_{TC} = v_{T7}^{max} \frac{C_D}{K_{M,D}+C_D} \tag{3.15}$$

The degradation of mRNA was reduced from sequence-oriented exonuclease and endonuclease reactions towards first-order kinetics (equation 3.16) with mRNA as a reactant and $k_{mRNA\,deg} = 0.0083\,min^{-1}$, as described by Hargrove & Schmidt (1989).

$$v_{mRNA\,deg} = C_{mRNA}\; k_{mRNA\,deg} \tag{3.16}$$

mRNA levels are only influenced by transcription and mRNA degradation resulting in the mass balance shown in equation 3.17,

$$\frac{dC_{mRNA}}{dt} = v_{TC} - v_{mRNA\,deg} \tag{3.17}$$

3.2.2 Translation

The key element of this modeling approach is the ribosome flow model, which describes the motion of ribosomes along an mRNA template as subsequent reactions from codon to codon. This approach calculates the number of ribosomes at each individual codon and their respective changes can be specific for the given codon. In addition, it incorporates steric hindrance between consecutive ribosomes. Translation is divided into three parts, initiation, elongation, and termination, which are treated individually in this context.

Initiation

The initiation of translation starts at ribosome dissociation from 70S to 30S and 50S subunits (association constant $K_{70S} = 5.3\,10^{7}\,\mathrm{M}^{-1}$ according to Zucker & Hershey (1986)), followed by random order binding of the three IFs (IF1, IF2 and

IF3) to the 30S subunit, forming the 30S IF complex ($C_{30S\,IF}$). The impact of bound IF on improved binding of the absent IFs was neglected, and corresponding equations and reaction rates were from Arnold et al. (2005).

The second step of initiation is the binding of the 30S IF complex towards the mRNA and fMet-tRNA in a random order. Subsequent binding of the 50S ribosomal subunit results in the release of IF1 and IF3, which is described by equation 3.18:

$$v_{TLI,70SIC} = \frac{q_j^R k_{TLI,70SIC} C_{30S\,IF}}{1 + \frac{K_{M,fMet-tRNA}}{C_{fMet-tRNA}} + \frac{K_{M,mRNA}}{C_{mRNA}} + \frac{K_{M,50S}}{C_{50S}} + \frac{K_{M,fMet-tRNA} K_{M,mRNA}}{C_{fMet-tRNA}\, C_{mRNA}}} \quad 3.18$$

Wintermeyer & Gualerzi (1983) estimated that the initiation rate constant $k_{TLI,70SIC}$ was $0.1\ s^{-1}$. Steric hindrance between ribosomes resulting in occupied ribosomal binding sites were considered by applying a queuing factor of q_j^R. This scaling factor describes the probability that a given codon, *j,* is already occupied by another ribosome, since that ribosome physically blocks $L_R = 12$ codons (Rauhut & Klug 1999). C_j^R defines the concentration of ribosomes currently translating codon *j.*

$$q_j^R = \frac{1 - \frac{\sum_{s=1}^{L_R} C_{j+s}^R}{C_{mRNA}}}{1 - \frac{\sum_{s=1}^{L_R - 1} C_{j+s}^R}{C_{mRNA}}} = \frac{C_{mRNA} - \sum_{s=1}^{L_R} C_{j+s}^R}{C_{mRNA} - \sum_{s=1}^{L_R - 1} C_{j+s}^R} = 1 - \frac{C_{j+L_R}^R}{C_{mRNA} - \sum_{s=1}^{L_R - 1} C_{j+s}^R} \quad 3.19$$

The last step during translation initiation is the release of IF2 described by the reaction rate $v_{TLI,IF2D}$, which follows first-order kinetics, according to Tomsic et al. (2000), with a reaction constant of $k_{TLI,IF2D}$ of 1.5 s^{-1}. IF2 regeneration can be

neglected due to an overhang of IF2 towards the GTP-bound species and a surplus of GTP:

$$v_{TLI,IF2D} = k_{TLI,IF2D}\ C_{70SIC} \tag{3.20}$$

Elongation

Translation elongation is the stepwise motion of ribosomes along the mRNA template. The main substrates for elongation are ribosomes as core machinery, EFG and EFTs in GTP bound form, and ternary complexes. These complexes consist of aminoacylated tRNA, EFTu, and GTP. According to Arnold et al. (2005), the net reaction rate ($v_{TLE,j+1}$.) from codon j to the subsequent codon *j+1* can be described as shown in equation 3.21. In addition, a queuing factor was introduced to consider steric interaction between ribosomes on the same mRNA strand (equation 3.19) and the maximum elongation rate k_{TLE} was set to 24 amino acids per ribosome per second (Arnold et al. 2005). The correlations of tRNAs with their respective codons (*j)* are considered as shown in the tRNA-codon table by Dong et al. (1996).

$$v_{TLE,j+1} = \frac{q_j^R k_{TLE} C_R^j}{1 + \frac{K_{M,T3_j}}{\sum C_{T3_j,i}} + \frac{K_{M,EFG\ GTP}}{C_{EFG\ GTP}}} \tag{3.21}$$

Termination

The last step during translation is the termination procedure. A ribosome located at codon *K* prior to a stop codon dissociates from the mRNA and releases the protein chain and the 70S ribosome complex. Ribosome release factors (RF) catalyze this

reaction, and according to Arnold et al. (2005), the reaction parameters are $K_{M,R_K} = 8.3\ nM$, $K_{M,GTP} = 100\ \mu M$, and $k_{TLT} = 24\ s^{-1}$.

$$v_{TLT} = \frac{k_{TLT} C_{RF}}{1 + \frac{K_{M,R_K}}{C_R^K} + \frac{K_{M,GTP}}{C_{GTP}} + \frac{K_{M,R_K} K_{M,GTP}}{C_R^K C_{GTP}}} \tag{3.22}$$

3.2.3 Regeneration of translation factors

During each elongation cycle, the GTP-bound forms of EFG and EFTu are hydrolyzed, and EFTs act as GTP donors for EFTu regeneration. Arnold et al. (2005) derived a reversible ping-pong bi-bi kinetics reaction rate for the EFTu regeneration reaction, as shown in equation 3.23. Here, the subscript t stands for the total concentration of a given molecule (equals the sum of all species). Forward and reverse reaction rate constants were set to 30 s^{-1} and 10 s^{-1} according to Ruusala et al. (1982):

$$v_{EFTu,reg} = \frac{V_f \left(C_{EFTu\ GDP}\ C_{GTP} - \frac{C_{GDP}\ C_{EFTu\ GTP}}{K_{eq,EFTu}} \right)}{D} \tag{3.23}$$

$$\begin{aligned} D = {} & K_{M,GTP} C_{EFTu\ GDP} + K_{M,EFTu\ GDP} C_{GTP} + \frac{V_f K_{M,EFTu\ GTP}}{V_r K_{eq,EFTu}} C_{EFTu\ GTP} \\ & + \frac{V_f\ K_{M,EFTu\ GDP}}{V_r\ K_{eq,EFTu}} C_{EFTu\ GTP} + C_{EFTu\ GDP} C_{GTP} \\ & + \frac{V_f K_{M,EFTu\ GTP}}{V_r K_{eq,EFTu} K_{i,EFTu\ GDP}} C_{EFTu\ GDP} C_{GTP} \\ & + \frac{V_f}{V_r K_{eq,EFTu}} C_{GDP} C_{EFTu\ GTP} \\ & + \frac{K_{M,EFTu\ GDP}}{K_{i,EFTu\ GTP}} C_{GTP} C_{EFTu\ GTP} \end{aligned} \tag{3.24}$$

$$V_f = k_{EFTs,f}\ C_{EFTs,t} \tag{3.25}$$

$$V_r = k_{EFTs,r}\, C_{EFTs,t} \tag{3.26}$$

Table 3.4: Affinity and inhibitory constants for the reversible ping-pong bi-bi kinetics of the EFTu regeneration rate equation, as described in Arnold et al. (2005).

	GTP	**EFTu GTP**	**GDP**	**EFTu GDP**
$\boldsymbol{K_M\ \mu M}$	50	1	3	2.5
$\boldsymbol{K_i\ \mu M}$	6.5	1	15	5.6

According to Pingoud et al. (1990), EFG GDP regeneration can be treated as a spontaneous reaction. The following three EFG species were considered: free, GTP, and GDP bound. The association and dissociation constants for GDP binding were set to $2.7\ 10^7\ \mathrm{M^{-1}s^{-1}}$ and $100\ \mathrm{s^{-1}}$, respectively, and the association and dissociation constants for GTP binding were set to $1.0\ 10^7\ \mathrm{M^{-1}s^{-1}}$ and $400\ \mathrm{s^{-1}}$, respectively (Pingoud et al. 1990).

$$C_{EFG,t} = C_{EFG} + C_{EFG\ GDP} + C_{EFG\ GTP} \tag{3.27}$$

Ternary complex formation describes the combination of EFTu and aa-tRNA j, forming a $T3_j$ complex, and the formation reaction is described as second-order kinetics with a first-order backward reaction, leading to following equation:

$$v_{T3,form,j} = k_{T3,form,j} C_{EFTu\ GTP} C_{tRNA,j} - k_{-T3,form} C_{T3,j} \tag{3.28}$$

tRNA recharging is catalyzed by aminoacyl-tRNA-synthetases (ARS). All ARS reactions were treated as uniform due to the lack of parameters for the ARS reaction. For the sake of simplicity, the parameters for this reaction were generalized. ARS coupled the suitable amino acid to the tRNA by forcing ATP to AMP hydrolyzation.

A reaction rate equation was derived from Arnold et al. (2005), where $K_{M,aa} = 20\ \mu M$, $K_{M,tRNA} = 0.5\ \mu M$, and $K_{M,ATP} = 100\ \mu M$ for all amino acid species. V_{ARS}^{max} was set to $357\ \text{min}^{-1}\ \text{ribosome}^{-1}$ to scale to different dilution factors (Neidhardt et al. 1977; Fersht et al. 1978):

$$V_{ARS,i,k} = \frac{V_{ARS,i,k}^{max}}{1 + \frac{K_{M,ARS,aa_j}}{C_{aa,j}} + \frac{K_{M,ARS,ATP}}{C_{ATP}} + \frac{K_{M,ARS,tRNA_j}}{C_{tRNA_j}}} \tag{3.29}$$

3.2.4 Material balances

The resulting material balances for the species regarded by this model are shown in equations 3.30 to 3.38. T indicates the overall number of different tRNA species and A indicates the overall number of different amino acid species.

$$\frac{dC_{70SIC}}{dt} = v_{TLI,70SIC} - v_{TLI,IF2D} \tag{3.30}$$

$$\frac{dC_R^1}{dt} = v_{TLI,IF2D} - v_{TLE,1} \tag{3.31}$$

$$\frac{dC_R^j}{dt} = v_{TLE,j} - v_{TLE,j-1} \quad \text{for} \quad 1 < j < K \tag{3.32}$$

$$\frac{dC_R^K}{dt} = v_{TLE,K} - v_{TLT} \tag{3.33}$$

$$\frac{dC^{Protein}}{dt} = v_{TLT} \tag{3.34}$$

$$\frac{dC_{aa,i}}{dt} = -\sum^{k} v_{ARS,j,k} \quad \text{for} \quad 1 \leq i \leq A \tag{3.35}$$

$$\frac{dC_{T3,k}}{dt} = v_{T3,form,k} - \sum_{1}^{K-1} \alpha_{j,k} v_{TLE,j} \quad \text{for} \quad 1 \leq k \leq T \quad \text{with} \quad \alpha_{j,k} \approx \frac{C_{T3,j,k}}{\sum^{i} C_{T3,j,i}} \tag{3.36}$$

$$\frac{dC_{aa_i-tRNA_k}}{dt} = v_{ARS,i,k} - v_{T3,form,k} \qquad \text{for} \qquad 1 \leq k \leq T \tag{3.37}$$

$$\frac{dC_{tRNA_k}}{dt} = \sum_{2}^{K} \alpha_{j,k} v_{TLE,j} \qquad \text{for} \qquad 1 \leq k \leq T \tag{3.38}$$

3.2.5 Initial conditions

Initial conditions for the DAE system are shown in table 3.5. Concentrations are calculated for the *in vivo* state with a growth rate of 1.1 h^{-1}.

Table 3.5: Concentrations C_i of the translation machinery for *E. coli* at a growth rate of 1.1 h^{-1}. A t as prefix for an amino acid indicates the corresponding tRNA species.

Molecule	C_i | µM	Source	Molecule	C_i | µM	Source
30S subunit	39.82	Dong et al. (1996)	t-Leu4	3.00	Dong et al. (1996)
50S subunit	39.82	Dong et al. (1996)	t-Leu5	3.65	Dong et al. (1996)
IF1	9.96	Howe and Hershey (1984)	t-Met	4.10	Dong et al. (1996)
IF2	11.95	Howe and Hershey (1984)	t-Asn	6.10	Dong et al. (1996)
IF3	7.96	Howe and Hershey (1984)	t-Pro1	2.75	Dong et al. (1996)
EFG_t	31.9	Pedersen et al. (1978)	t-Pro2	4.01	Dong et al. (1996)
$EFTu_t$	236.69	Neidhardt et al. (1977)	t-Pro3	2.55	Dong et al. (1996)
$EFTs_t$	5.0	Pedersen et al. (1978)	t-Gln1	3.17	Dong et al. (1996)
Release factor	37.6	Arnold et al. (2005)	t-Gln2	5.07	Dong et al. (1996)
t-Met f1	20.0	Arnold et al. (2005)	t-Arg2	23.77	Dong et al. (1996)
t-Ala1B	17.52	Dong et al. (1996)	t-Arg3	2.26	Dong et al. (1996)
t-Ala2	3.19	Dong et al. (1996)	t-Arg4	3.26	Dong et al. (1996)
t-Cys	7.04	Dong et al. (1996)	t-Arg5	2.46	Dong et al. (1996)
t-Asp1	12.04	Dong et al. (1996)	t-Ser1	6.98	Dong et al. (1996)
t-Glu2	24.12	Dong et al. (1996)	t-Ser2	1.37	Dong et al. (1996)
t-Phe	4.69	Dong et al. (1996)	t-Ser3	5.40	Dong et al. (1996)
t-Gly2	10.95	Dong et al. (1996)	t-Ser5	3.68	Dong et al. (1996)
t-Gly3	19.84	Dong et al. (1996)	t-Thr13	5.42	Dong et al. (1996)
t-His	3.35	Dong et al. (1996)	t-Thr2	2.67	Dong et al. (1996)
t-Ile	18.92	Dong et al. (1996)	t-Thr4	4.99	Dong et al. (1996)
t-Lys	8.73	Dong et al. (1996)	t-Val1	18.99	Dong et al. (1996)
t-Leu1	21.32	Dong et al. (1996)	t-Val2AB	6.31	Dong et al. (1996)
t-Leu2	4.72	Dong et al. (1996)	t-Trp	4.15	Dong et al. (1996)
t-Leu3	3.19	Dong et al. (1996)	t-Tyr12	9.83	Dong et al. (1996)

3.2.6 Numerical integration

The complete model consists of 410 first-order differential equations and eight algebraic equations. The resulting differential-algebraic equation (DAE) system was numerically integrated using a variable-step, variable-order solver based on the numerical differentiation formula (Shampine et al. 1999). This method was used because it is more robust with inconsistent sets of initial conditions compared to other approaches such as Dassl (Brenan et al. 1995). Compared to the previous model of Arnold et al. (2005) with more than 1,000 equations, this model has a reduced set of equations, which drastically increases numerical stability and computational speed. The DAE system is sensitive to initial conditions due to the algebraic equations, and must be balanced precisely. Therefore, a three-step algorithm was established. The first step solves the algebraic equations without regarding the differential equations. Based on these preliminary values, the second step improves the initial values and the complete DAE system is solved without active transcription. This step is crucial to balancing the initial composition of elongation factors, which cause stability issues. After the initial values are balanced, the main integration can occur including all mechanisms.

3.2.7 Dilution factor

The preparation of CFPS reactions inherently causes a dilution of the cellular components. This dilution was included in the model by incorporating a dilution factor (D) from *in-vivo* to *in-vitro* conditions, which was calculated as the ratio between the ribosome concentrations of both states. The specific loss of enzymes was factored in by calculating the recovery loss R for each reactant. According to the experimental results, the recovery for all molecules was set to one, except for ribosomes ($R^R = 0.9$) and EFTu ($R^{EFTu} = 0.8$).

$$D = \frac{C^i_{in\ vivo}\ R^i}{C^i_{in\ vitro}} \tag{3.39}$$

3.2.8 Translational control analysis

Translational control analysis identifies control steps during protein biosynthesis (Zouridis & Hatzimanikatis 2008). The impact of substrates on the resulting translation rate is described as elasticity ϵ_i and the impact of catalyzing enzymes (i.e. ribosomes) is described as the flux control coefficient (FCC). Partial differentials were assessed by deflecting the respective parameter by ± 1 %.

$$\epsilon_i = \frac{C_i}{v_{TL}} \frac{\partial v_{TL}}{\partial C_i} \tag{3.40}$$

$$FCC = \frac{C_{70S,t}}{v_{TL}} \frac{\partial v_{TL}}{\partial C_{70S,t}} \tag{3.41}$$

3.3 Ensemble Model

Describing the transcriptional and translational behavior of cells in a changing environment is a demanding task, as is the induction of transcription as a function of the substrate availability and the resulting protein biosynthesis cascade. For example, the dynamic TC-TL model is a complex system, though it only describes the protein biosynthesis of a single mRNA template. However, many different genes are transcribed and translated simultaneously under *in vitro* conditions and they influence each other, since they rely on the same pool of translation factors (e.g. RNAP, ribosomes, and ternary complexes). The dynamic and stochastic models do

not fulfill this criterion of describing the translation of multiple genes simultaneously. Furthermore, the designated model must describe cellular movement in an STR-PFR reactor system to link substrate disposability and the induction of transcription. These constraints to the model setup can only be fulfilled by a structured and segregated model as shown in figure 2.1.

These mechanisms make the model complex due to the influence of substrate disposability on protein biosynthesis and the influence of synthesized enzymes on the substrate uptake behavior. To describe the cellular reaction to the spatial substrate limitation, two models are necessary, (i) a spatial model that describes the position of each cell and (ii) a cell model that describes the reaction to starvation-induced protein synthesis.

3.3.1 Experimental setup

Using a stirred tank reactor (STR) coupled with a plug-flow reactor (PFR) as a scale-down setup allows the simulation of oscillating substrate availability to occur in large-scale cultivations (the system setup is shown in figure 2.2). According to Simen et al. (2017), the setup was used under nitrogen-limited chemostat conditions with an adjusted dilution rate of 0.2 h^{-1}. Prior to connecting the PFR, an STR chemostat cultivation with the same dilution rate was performed to provide an initial steady state (SS_0) for transcript analysis. After the initial steady state was reached, the PFR loop was added with a pump rate of 180 mL min^{-1}, forcing circulation through the PFR compartment. In the PFR loop, additional agitation is implemented to prevent oxygen starvation. Therefore, only nitrogen limitation occurs in the PFR. After passing through the PFR, the cells reenter the main STR compartment and based on this setup, long-term adaptation can be assessed in the STR and a detailed analysis of the short-term reaction along the PFR is possible. Sampling ports along the PFR allows for the withdraw of samples at residence times of 31, 70, and 110

seconds (Löffler et al. (2016). In addition, the residence times of the STR and PFR are 6.2 minutes and 125 seconds, respectively.

Due to the addition of the PFR after SS_0 and the limitation in the PFR compartment, a split of the growth rate occurs. During the PFR passage, cells are nitrogen limited and cannot grow, while cells in the STR can still grow. Based on the volumetric distribution between STR and PFR, the resulting growth rate can be calculated as shown in equation 3.42:

$$\frac{\mu_{STR}\ \tau_{STR} + \mu_{PFR}\tau_{PFR}}{\tau_{STR} + \tau_{PFR}} = D \qquad 3.42$$

In addition, samples were taken for transcriptome analysis at 25 minutes, 120 minutes, and 28 hours after installing the substrate gradient. Transcript measurements are published in Simen et al. (2017) and are available under GEO Accession GSE90743.

3.3.2 Single-cell model

TC-TL in this model is described by the motion of an RNAP i at position x_i along the DNA template, which is discretized into nucleotides. The following mechanisms describing RNAP motion are included in the model: (i) initiation at a specific time interval, (ii) defined elongation with v_{elo}^{RNAP} as the maximum elongation rate, and (iii) a minimum distance Δx between two subsequent RNAPs. These criteria lead to the following rules for RNAP movement:

- The first elongation step is treated as the initiation step and can only occur if t is in the interval of possible induction (t_{ind}).
- The minimum distance Δx between two subsequent RNAPs is fulfilled.

$$\frac{dx_i}{dt} = \begin{cases} 0 & if\ x_i = 0\ \wedge\ t \notin t_{ind} \\ 0 & if\ x_{i-1} - x_i < \Delta x\ \wedge i > 1 \\ v_{elo}^{RNAP} & otherwise \end{cases} \quad 3.43$$

The corresponding sequence part of the mRNA is synthesized during motion on the DNA. As a result, an mRNA strand i with length L_i^{mRNA} is generated according to equation 3.44:

$$L_i^{mRNA}(t) = x_i(t) \quad 3.44$$

For simplicity, the detailed mechanism of attenuation was neglected and it was assumed that transcription could only occur in the PFR compartment during nitrogen limitation and that translation was directly coupled to mRNA disposability. Each individual ribosome j and its corresponding position $y_{i,j}$ on mRNA strand i are functions of the length of the mRNA strand (L_i^{mRNA}) and the position of the next downstream ribosome y_{j-1}. Each gene on the operon structure has a different strength of the ribosome binding site and therefore, each gene results in different numbers of proteins per mRNA. This mechanism was implemented by defining the number of ribosomes $N_g^{TL,\max}$ that can translate gene g on the mRNA. In addition, the boundaries of each gene are set by the start codon (C_g^{start}, first coding nucleotide) and the stop codon (C_g^{stop}, last nucleotide of the stop codon). These mechanisms resulted in the following set of rules for ribosomal motion:

- A minimum of Δy downstream nucleotides are synthesized
- The previous ribosome is more than Δy nucleotides downstream
- The maximum number of translations for the given gene is not exceeded

$$\frac{dy_{i,j}}{dt} = \begin{cases} 0 & if\ L_i - y_{i,j} \leq \Delta y \\ 0 & if\ y_{i,j-1} - y_{i,j} < \Delta y\ \wedge j > 1 \\ 0 & N_{i,g}^{TL}(t) \geq N_g^{TL,max} \\ v_{elo}^{Ribosome} & otherwise \end{cases} \quad 3.45$$

Using the Iverson brackets that result in one if the statement inside the brackets is true and zero if the statement is false, the number of ribosomes that are active on an mRNA strand can be calculated according to equation 3.46:

$$N_{i,g}^{TL}(t) = \sum_j [y_{i,j}(t) \geq C_g^{start}] \quad 3.46$$

The number of synthesized proteins per cell from the single-cell model ($N_{g,SC}^{Protein}$) encoded by gene g can be calculated as the sum of all ribosomes acting on all mRNA strands that have passed the final nucleotide C_g^{end}.

$$N_{g,SC}^{Protein}(t) = \sum_i \sum_j [y_{i,j} > C_g^{end}] \quad 3.47$$

Furthermore, the degradation of each individual mRNA is catalyzed by RNases, starting with the initiation of degradation at the start codon of transcription (the first nucleotide of the mRNA). The elongation of degradation can be described by the position z_i on mRNA I, while the movement of RNases occurs with the elongation rate of v_{elo}^{RNase}. These assumptions lead to the following set of constraints for RNases:

- The number of active ribosomes per gene g $N_g^{TL,\max}$ is estimated as the turnover ratio of mRNAs and proteins (see below)
- Δz encodes the closest distance in nucleotides to the next ribosome downstream of z_i

$$\frac{dz_i}{dt} = \begin{cases} 0 & if\ N_{i,g}^{TL}(t) < N_g^{TL,\max} \\ 0 & if\ y_{i,N_g^{TL,\max}} - z_i \leq \Delta z \\ v_{elo}^{RNase} & otherwise \end{cases} \quad 3.48$$

The current number of mRNAs for a given gene can be calculated as the difference of transcribed mRNAs and the number of degraded mRNA. The former is derived from the lengths of the mRNA strands and the latter is derived from the RNase position on the mRNA (square brackets denote the Iverson brackets):

$$N_{g,SC}^{mRNA}(t) = \sum_i [L_i^{mRNA}(t) > C_g^{end}] - \sum_i [z_i \geq C_g^{start}] \quad 3.49$$

Each individual gene has a corresponding number of translations per mRNA $N_g^{TL,max}$, which can be calculated as the turnover ratio of mRNAs per protein for a given gene *g*. mRNA turnover $r_{turnover}^{mRNA}$ can be calculated from the transcript levels of Valgepea et al. (2013) with the degradation constant of $k_{deg}^{mRNA} = 20.79\ \mathrm{h^{-1}}$ from Chen et al. (2015). The corresponding protein turnover $r_{turnover}^{protein}$ is derived from the protein levels from Valgepea et al. (2013), neglecting active protein degradation. In addition, cytosolic dilution by microbial growth was incorporated and therefore, $k_{deg}^{protein} = \mu = 0.2\ \mathrm{h^{-1}}$. It was assumed that values of *trpB* and *trpA* were equal due to the lack of values for *trpA* in the dataset and the fact that the resulting protein complex was a tetramer consisting of two *trpA* and two *trpB* (Hyde et al. 1988). The resulting translations per mRNA are presented in table 3.6.

$$N_g^{TL.max} = \frac{r_{turnover}^{Protein}}{r_{turnover}^{mRNA}} = \frac{C_g^{Protein}\ k_{deg}^{Protein}}{C_g^{mRNA}\ k_{deg}^{mRNA}} \quad 3.50$$

Table 3.6: Calculated translations per mRNA for the *trp* operon. The value for *trpA* was extrapolated from *trpB*.

Gene	***trpE***	***trpD***	***trpC***	***trpB***	***trpA***
Translations per mRNA	4	4	5	10	10

Using the *trp* operon as an example provides the following advantages:

- the *trp* operon leads to a polycistronic mRNA (Yanofsky et al. 1981),
- the attenuation sequence in the *trpL* leader peptide allows for the coupled investigation of TC-TL, and
- the published data by Simen et al. shows that the *trp* operon is upregulated during STR-PFR cultivations.

If genes downstream of *trpL* are visible, translation initiation occurs due to the attenuation mechanisms, which was observed in this study. For simplification, the genes *trpGD* and *trpCF,* and the corresponding structural proteins were regarded as single genes (*trpG* and *trpC*, respectively) and proteins. The elongation rates of the three regarded species, RNAP, ribosomes, and RNase, were treated equally and set to the values reported by Chen et al. (2015). In addition, distances $\Delta x, \Delta y,$ and Δz between the molecules were set to 100 nucleotides (extrapolated from Bremer & Dennis (1987)).

The dataset from Simen et al. (2017) shows a delay of 30 seconds before the first elevated transcript levels are noticeable, and a corresponding induction delay time was implemented. Since no abortion sequences are reported in the literature after induction starts, it was assumed that transcription and translation would continue until the last stop codon of *trpA*.

3.3.3 Cell distribution model

The second part of the ensemble model is the process model that describes the trajectories of the cells in the STR-PFR setup and calculates residence times. Löffler et al. (2016) showed that the PFR can be treated as a plug flow as the residence time is not a distribution but a fixed value. In comparison, the ideally mixed STR shows a residence time distribution, which is constrained by the volume, the pumping rate into the PFR, and the feed and harvest streams. Regarding a single cell in the STR, the following events can occur, where cells may:

- leave the STR to enter the PFR and cycle back into STR after τ_{PFR}, the residence time in the PFR,
- be drained off by the efflux (harvest), or
- divide, setting all transcriptional and translational programs on default (no initiation of transcription or translation in the corresponding daughter cell)

Based on these events, probability functions (α_i) were defined for each event:

$$\alpha_1 = N_{STR} \frac{\dot{V}_{PFR}}{V_{STR}} \tag{3.51}$$

$$\alpha_2 = N_{STR} \frac{\dot{V}_{Feed}}{V_{STR}} \tag{3.52}$$

$$\alpha_3 = N_{STR}^0 \quad D \tag{3.53}$$

The probability $\alpha_1$is used to describe event 1, indicating the probability that a cell will leave the STR and enter the PFR; it is derived from the ratio of STR-PFR flux and STR volume. The washout of cells (event 2) was treated as equal to the dilution rate to calculate the corresponding probability α_2. The probability of cell division (α_3) is based on the set dilution rate D and the cell number N_{STR}^0 during the initial steady state prior to coupling the PFR. After entering the PFR, the constant

residence time results in a return to the STR of the cell after τ_{PFR}. Event 3, cell division, is treated as the rise of a new cell without deflected transcript and protein levels, where the mother cell receives no deflection.

The system can be solved numerically by applying Gillespie's SSA (Gillespie 1977), where each iteration is precisely the length of time required to fire a single reaction. The calculation of this time increment τ is described in equation 3.54, considering the probabilities α_i of the three events. According to equation 3.55, the corresponding reaction to be fired is chosen and the next iteration starts. The variables r_1 and r_2 are numerically generated, uniformly distributed random variables in the interval (0, 1).

$$\tau = \frac{1}{\sum \alpha_i} \ln\left(\frac{1}{r_1}\right) \tag{3.54}$$

$$\sum_{j=1}^{i-1} \alpha_j \leq r_2 \sum_{j=1}^{3} \alpha_j \leq \sum_{j=1}^{i} \alpha_j \tag{3.55}$$

Ten thousand cells were selected as a representative subpopulation to describe the overall population; this was set as the initial population number (N_{STR}^0) in the STR prior to connecting the PFR. During simulations, the position and each STR-PFR transition were monitored for all cells.

3.3.4 Coupling of single-cell and cell distribution models

Coupling the two models results in a model loop where both models influence each other. The distribution model provides the necessary induction intervals for the single-cell model which in turn influences the substrate consumption due to deflected protein levels. Since the application of a chemostat process strategy results in constant substrate conditions, the impact of enhanced metabolic capabilities on the culture broth was neglected. The initial loop pendency can be broken into a linear

pendency which results in fewer computational expenses. Since multiple consecutive STR-PFR passages described by the single-cell model can be calculated as a superposition of the individual STR-PFR passages, only one iteration of the single-cell model needs to be calculated. This results in the uncoupling of both models and enables the independent simulation of them, drastically decreasing the computational effort.

Coupling of the models is performed by logging each transition between STR and PFR, which triggers an induction. Based on these transition times t_i^{flag}, the superposition of the single-cell model can be performed as described in equations 3.56 and 3.57. By further tracking events (2) and (3), the change in the STR population distribution can be calculated. The single-cell model describes the transcriptional ($N_{g,SC}^{mRNA}(t)$) and translation patterns ($N_{g,SC}^{Protein}(t)$) for a single cell after entering the PFR, which are stored as look-up tables. In combination with the transition times from the process model, the cellular composition can be directly derived from the loop-up table. Superposition of all mRNA and protein patterns over the process duration provides a sustained transcriptional and translational pattern for each cell. Table 3.7 provides an overview of the necessary parameters for the simulation of both models.

$$N_g^{mRNA}(t) = \sum_i \begin{cases} N_{g,SC}^{mRNA}\left(t - t_i^{flag}\right) & if\ t - t_i^{flag} \leq \Delta t \\ 0 & otherwise \end{cases} \tag{3.56}$$

$$N_g^{Protein}(t) = \sum_i \begin{cases} N_{g,SC}^{Protein}\left(t - t_i^{flag}\right) & t - t_i^{flag} \leq \Delta t \\ N_{g,SC}^{Protein}(\Delta t) & otherwise \end{cases} \tag{3.57}$$

Table 3.7: Model parameters used for the simulation of both the single-cell and cell distribution models.

Parameter	Value	Unit
v_{elo}^{RNAP}	21	nt s^{-1}
$v_{elo}^{Ribosome}$	21	nt s^{-1}
v_{elo}^{RNase}	21	nt s^{-1}
Δx	100	nt
Δy	100	nt
Δz	100	nt
t_{ind}	[30 125]	s
$\dot{V}_{PFR}$	180	mL min^{-1}
$\dot{V}_{Feed}$	5	mL min^{-1}
V_{STR}	1,120	mL
D	0.2	h^{-1}
N_{STR}^{0}	10,000	cells

CHAPTER 4

RESULTS AND DISCUSSION

This chapter provides the results of this thesis, which were gathered by applying the different modeling strategies. The first chapter provides the results of the diffusion-based approach. Subsequently, the analysis of the single gene TC-TL model and the population-based approach are presented.

4.1 Diffusion limitation during translation

The lack of convection in *in vivo* and *in vitro* systems renders diffusion as the single mechanism of transport. To address the effect of diffusion as a rate-limiting step, a simulation platform that describes three-dimensional movement and collision was established. This platform can describe motion on a three-dimensional grid and the collision-driven reaction of potential reaction partners. Using this simulation platform, the effect of diffusion-driven translation elongation can be addresses and quantified.

Translation elongation, the transport of ternary complexes to the translating ribosome, was selected as an example of a diffusion-limited reaction. This step is indicated as a rate-limiting step in both *in vivo* and *in vitro* protein synthesis. The

large number of different tRNA species and the sequence-oriented incorporation of amino acids result in a high number of substrates, of which few are suitable at a specific codon.

After being depleted during an elongation step, each tRNA is released from the ternary complex and requires regeneration catalyzed by aminoacyl-tRNA-synthetases. After being regenerated, the tRNA must bind an EFTu molecule to form a new ternary complex, and the ternary complex must be transported to a ribosome that currently accepts it. Due to the low number of ribosomes in CFPS reactions and the distribution of tRNA species, the transport of ternary complexes might become a limiting step (Varenne et al. 1984; Fluitt et al. 2007; Zhang et al. 2010). In this case, the detailed transport of suitable ternary complexes to the translating ribosome was investigated. The regeneration of EFTu and tRNA was treated as instantaneous for simplicity.

4.1.1 *In vitro* validation

CFPS reactions provide an open platform for analyzing TC-TL. The single gene expression coupled with the open reaction design provides an ideal example system for model validation. Since initial concentrations of the necessary molecules are essential for this model, *in-vivo* concentrations were applied for ribosomes, tRNA, and EFTu. Ribosome concentration was reduced to 80% to accommodate the number of ribosomes that were not translating (Bremer & Dennis 1987). Shifting from *in vivo* to *in vitro* concentrations, a dilution factor was applied as described in Nieß & Failmezger et al. (2017) to adjust the concentration of all involved species.

Figure 4.1 provides an overview of the first 3,000 elongation steps of the model, and the duration of each elongation step as well as the cumulative duration are presented. The duration of each elongation step shows a wide variation. Depending on the distance that a ternary complex must travel, elongation times of up to seven

ms were identified. However, the cumulative reaction allows for easier assessment of the mean elongation rate. The initial distribution of molecules results in a non-steady state that slowly approaches steady state. After approximately 1,000 elongation steps, the slope of the total reaction time becomes constant, and to calculate the elongation rates, the first 1,000 elongation steps were depraved because the system was in a non-equilibrium state.

Underwood et al. (2005) measured an elongation rate of 1.5 ± 0.2 AA rib^{-1} s^{-1} for *in vitro* chloramphenicol acetyl transferase (CAT) synthesis. By applying a dilution factor to achieve identical ribosome concentrations, an elongation rate of 1.56 AA rib^{-1} s^{-1} was estimated. A difference between the predicted and measured elongation rates of 4 % with a *p*-value of 0.46 indicates no significant difference between the experimental results and *in silico* prediction. Furthermore, a comparison with the simulated elongation rates published in Nieß & Failmezger et al. (2017) shows that the diffusion-based model predicts slightly higher GFP elongation rates. This effect likely arises because the quantity of actively translating ribosomes in the diffusion model is fixed at 80 %, whereas the published model treats the number of actively translating ribosomes as variable. This verifies that CFPS reactions translating single mRNA species are limited by the transport of ternary complexes to ribosomes, and this effect might be further enhanced by unbalanced codon usage. In principal, the reaction diffusion model can qualitatively and quantitatively predict elongation rates for *in vitro* protein synthesis.

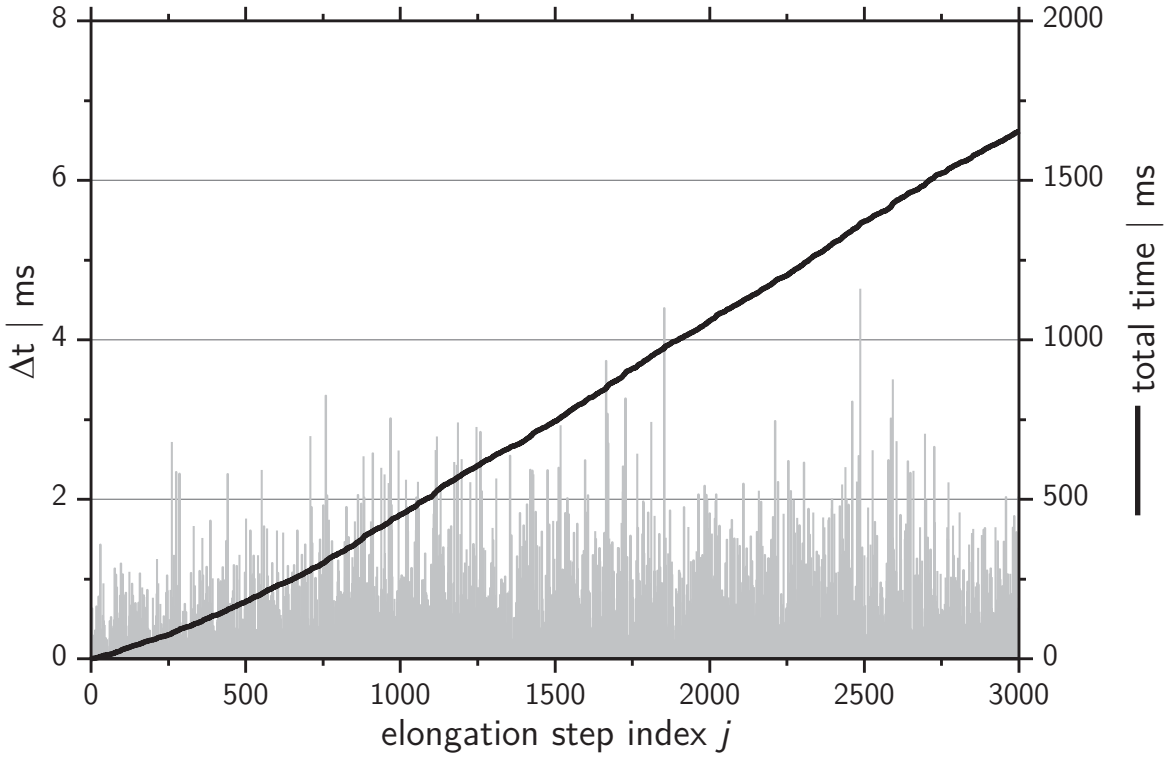


Figure 4.1. Single run of the reaction-diffusion model with GFP as a coding sequence. Evaluated time difference between two elongation steps are shown as bars and the straight line shows the overall time at which the elongation step occurred.

4.1.2 The impact of dilution

The major difference between *in vitro* and *in vivo* protein synthesis is the dilution that occurs during reaction preparation. In this chapter, the detailed influence of dilution on the transport behavior of ternary complexes was analyzed. By neglecting recovery during reaction preparation, the dilution factor becomes the single parameter influencing the elongation rate in this model. In addition, two different genes with different codon usages were analyzed, GFP as an example of heterologous gene expression, and EFTu for homologous gene expression. Both sequences have significantly different codon adaptation indexes (CAI) and gene lengths as shown in table 4.1.

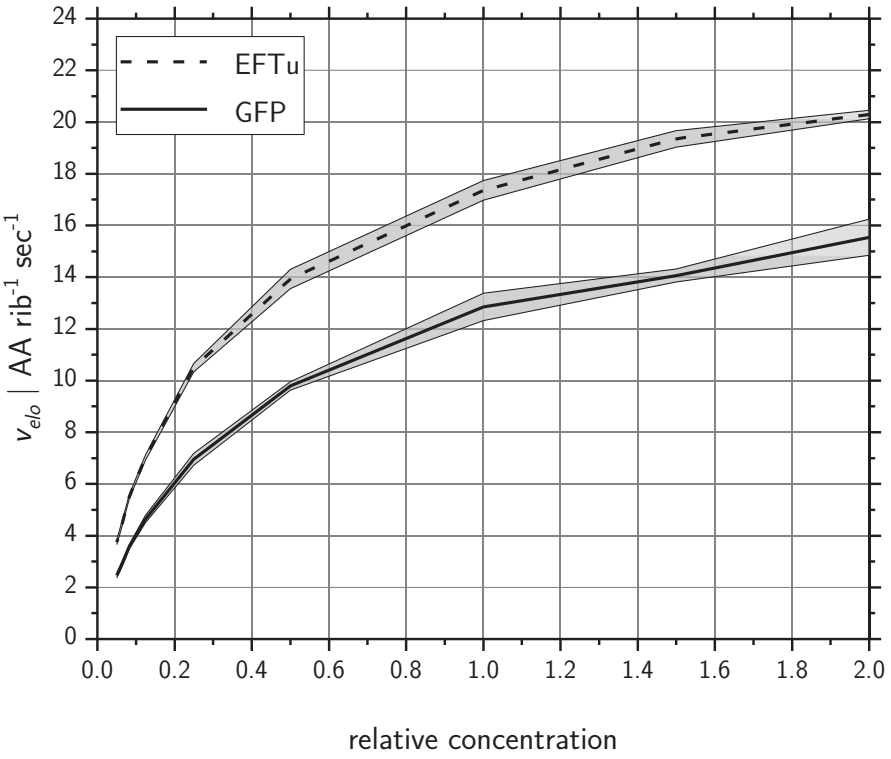


Figure 4.2: Influence of the ternary complex concentration on the average elongation rate for GFP and EFTu gene sequences. Relative concentrations are calculated based on the *in vivo* concentrations at a growth rate of 1.1 h^{-1}.

The elongation rate and relative concentration are directly coupled as shown in figure 4.2. With relative concentrations and converging *in vivo* concentrations, saturation of the elongation rate is slowly established. In this case, the catalytic reaction time becomes more relevant and limits the elongation rate. Independent from the genetic sequence, relative concentrations above unity show a decreased influence of the elongation rate and translation can no longer be regarded solely as substrate limited, as it is also limited by the catalytic efficiency of the ribosome.

Typical black-box Michaelis-Menten kinetics show the same trend. With rising substrate concentration, the impact on the reaction rate decreases until the reaction is only limited by the catalytic capability of the enzyme. Compared to diffusion, where the diffusion coefficient influences the transport rate, Michaelis-Menten kinetics rely on the affinity of the substrate to the enzyme. In principle, affinity

constants for translation elongation should be derived from diffusion-based simulations.

4.1.3 The impact of gene sequences

As shown in the previous sections, gene sequence and codon usage impact the elongation velocity, resulting in different patterns for the elongation rate. The expression of many heterologous expressed genes suffer from unfavorable codon usage compared to homogeneous expressed proteins such as EFTu, with high intracellular levels that are evolutionarily adapted to the translation conditions. In this chapter, the influence of different gene sequences with varying codon usage is quantified.

The GFP sequence of the pJOE4056.2 plasmid used by Nieß & Failmezger et al. (2017) achieves a CAI of 0.531 according to Sharp & Li (1987), indicating that this sequence is not optimized for translation in *E.coli* since a CAI of 1.0 indicates perfect codon usage and a CAI close to zero indicates unsuitable codon usage. Using the CAI optimizer platform (Puigbò et al. 2007), a new DNA sequence was generated with the same amino acid sequence resulting in a CAI of 0.723 and more homogeneous usage of tRNA species.

Table 4.1: Characteristic parameters and predicted CFPS translation rates for different gene sequences based on the reaction-diffusion model. Both *in vivo* (dilution factor = 1) and *in vitro* (dilution factor = 21.08; (Nieß & Failmezger et al. 2017) conditions were evaluated. Elongation rates were calculated as mean values of ten independent simulations.

Protein	**GFP**	**GFP***	**CAT**	**EFTu**
Length | codons	245	245	219	395
CAI | -	0.53	0.72	0.71	0.78
***in vitro* v_{elo} | AA rib^{-1} s^{-1}**	1.78 ± 0.06	2.08 ± 0.08	1.56 ± 0.09	2.98 ± 0.08
***in vivo* v_{elo} | AA rib^{-1} s^{-1}**	12.6 ± 0.4	14.5 ± 0.4	11.4 ± 0.7	17.4 ± 0.5

*Codon optimized variant of GFP

Table 4.1 shows that optimizing the GFP sequence increases the elongation rate by 19 %. GFP has a lower CAI than CAT, but a higher elongation rate, and CAT has a much lower elongation rate than EFTu despite an almost identical CAI. These comparisons show that CAI is not an exclusive parameter for comparing elongation rates. Taking gene length into account shows that the shortest gene CAT has the lowest elongation rate, followed by GFP with a slightly higher elongation rate and gene length, and then by EFTu with the highest elongation rate and the longest sequence.

These simulations indicate that tRNA usage, as shown by the CAI and gene length, influence the elongation rate. A better codon adaptation optimizes the transport rates using a higher quantity of different tRNAs. However, overuse of frequent tRNAs results in a contrary effect. During each elongation cycle, the previously used tRNA remains fixed in the ribosome and is not available for ternary complex generation. This effect reduces the number of usable ternary complexes and therefore, the elongation rate. A distribution of codons according to the available number of tRNAs can prohibit this effect and prevent consecutive identical codons (e.g. His-tags).

EFTu, as a protein with a high presence in the cytosol, is evolved for high elongation rates. Despite the suboptimal CAI, it shows high elongation rates, indicating that CAI might be a useful tool for protein synthesis optimization, but the impact of sequence and ribosome distribution on the gene must be considered as well.

For *in-vivo* conditions, the same pattern of the target sequences can be identified. CAT has the lowest elongation rate and EFTu has the highest, with the two GFP forms in between. In summary, under *in vivo* like translation conditions, the influence of CAI is reduced and the effect of gene length is increased. EFTu translation achieves elongation rates as reported for whole-cell translation (Bremer & Dennis 1987), which

was measured as the average translation rate for the entire set of genes that experienced translation.

4.1.4 Control analysis

The changing influence of substrates and enzymes on the reaction rate can be analyzed using MCA as described by Zouridis & Hatzimanikatis (2008), and Nieß & Failmezger et al. (2017). Here, the flux control coefficient (FCC) describes the influence of changing ribosome levels and the elasticity ϵ_{T3} describes the influence of ternary complexes on the resulting translation rate. Figure 4.3 illustrates that for the GFP sequence, both FCC and ϵ_{T3} have values over 0.5 and show no significant change. Consequently, translation is limited by ribosomes and ternary complexes. However, for the EFTu sequence, a significant decrease in ϵ_{T3} was identified with rising concentrations, which directly shows the impact of saturated ribosomes. Compared to the GFP sequence, the codon usage of EFTu is far more balanced and the resulting elongation rates are 50 % higher. The saturation effect emerges early compared to GFP. Due to the assumption of this model that the ratio of actively translating ribosomes is constant, the evaluation of the FCC is biased and ϵ_{T3} can be assessed with a high level of confidence.

In summary, the translation of GFP is limited by the transport of ternary complexes, even under *in vivo* conditions, whereas the translation of EFTu becomes saturated with ternary complexes with rising relative concentrations. I*n vivo* production of EFTu is almost at ideal conditions, whereas GFP production suffers from tRNA distribution. This indicates that GFP synthesis rates *in vivo* in *E. coli* show potential for further optimization.

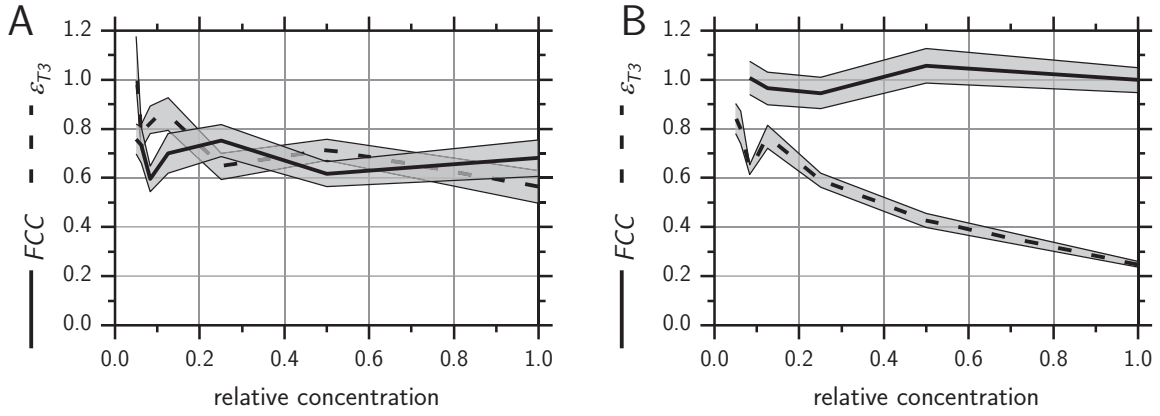


Figure 4.3: Control analysis of *FCC* and ϵ_{T3} for GFP (A) and EFTu (B) sequences at different relative concentrations.

4.1.5 Optimizing elongation rates

As shown in the previous sections, *in vivo* translation has high elongation rates due to the high level of ternary complexes, and both the number of ternary complexes and ribosomes can increase the overall translation rate. Thus, the influence of all three reaction components was analyzed by increasing them independently. The increase in components was normalized for comparison based on the ATP costs for each molecule.

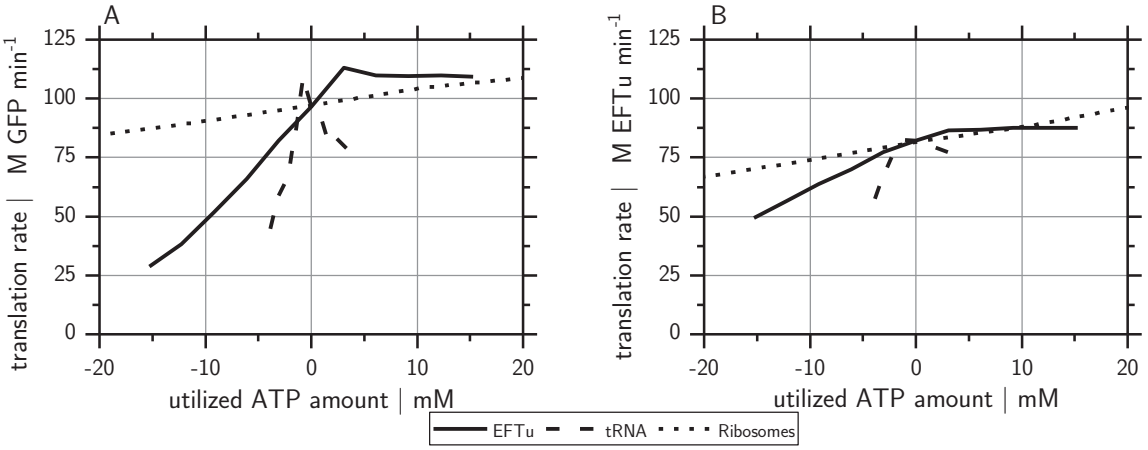


Figure 4.4: The influence of changing tRNA, EFTu, and ribosome levels on the resulting translation rate shown for GFP (A) and EFTu (B). The concentration change was normalized for the ATP costs of the individual molecules.

The impact of the three considered molecules was initially analyzed independently. For this approach, the initial concentrations were altered and the change in concentration was normalized to the costs to synthesize the molecule. According to chapter 3.1.3, costs were set to 700, 3,565 and 108,461 ATP equivalents for tRNA, EFTu and ribosomes, respectively. In addition, GFP and EFTu were analyzed as target gene sequences.

For GFP as a target sequence, three distinct patterns were identified (shown in figure 4.4): (i) the translation rate scales linear with the number of ribosomes, (ii) the translation rate scales linear with the EFTu concentration until a certain point, and then increasing the EFTu concentration has no effect, and (iii) tRNA concentration has an optimum and higher or lower concentrations, result in a lower translation rate. Effect (i) arises because only ternary complexes are moving and ribosomes are fixed. Increasing the number of ribosomes therefore reduces the average length that a ternary complex must move before colliding with a ribosome. Effect (ii) is due to the different initial conditions of EFTu and tRNA, as the initial number of tRNAs is slightly higher than the EFTu number, and increasing the EFTu number

results in more available ternary complexes. However, if more EFTu is supplied than tRNA, no further ternary complexes are generated and the translation rate cannot be increased further. The last effect (iii) is caused by two mechanisms. At lower tRNA levels, the pool of ternary complexes is decreased as well, limiting the suitable reaction partners for ribosomes. Increasing the tRNA concentration results in the accumulation of unfavorable ternary complexes. After each collision, the released EFTu binds a random tRNA and ternary complexes that are unsuitable for elongation accumulate. Although this effect is normally counteracted by ternary complex dissociation, the dissociation rate has almost no impact on the ternary complex distribution due to the low dissociation constant.

EFTu as a target sequence shows the same patterns for the molecules, though the impact of EFTu and tRNA is shallower. This is caused by the broader use of tRNAs in the EFTu sequence compared to GFP, which reduces the number of unsuitable tRNAs. Moreover, the optimal conditions for tRNA concentration are identical for both sequences despite the different codon usage. The optimal concentration is close to unaltered concentrations, indicating that the tRNA conditions *in vivo* are already optimized for maximum diffusive transport. This is also true for EFTu, where the optimal concentration is in transition from a linear to constant translation rate and is located near non-altered conditions.

Given the challenge of increasing the translation rate with limited ATP, the distribution of ATP that increases the three molecules was investigated for GFP and EFTu as target proteins. The three molecules were coupled with the total number of ATP equivalents used and distributed towards the three species indicated by the split ratio a_i, as shown in equation 4.1. In addition, the ratio for tRNAs was limited to below 0.2, since tRNAs are less expensive than other molecules and an over proportional increase in tRNAs compared to EFTu has a negative influence as shown in figure 4.4.

$$a_{tRNA} + a_{EFTu} + a_{Ribosomes} = 1 \tag{4.1}$$

$$a_{tRNA} \leq 0.2 \tag{4.2}$$

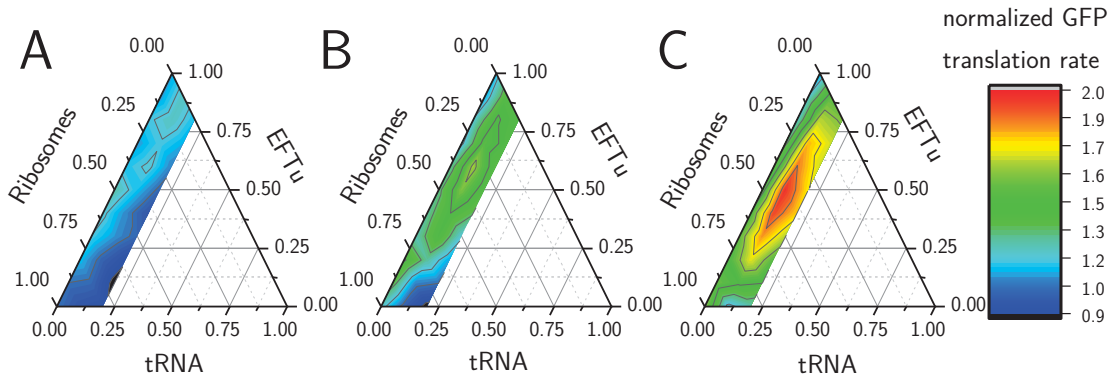


Figure 4.5: Distribution of ATP in tRNA, EFTu, and ribosomes and the resulting translation rate (normalized to unaltered conditions) for **GFP** as a target protein. The amount of ATP distributed was 1e7 (A), 5e7 (B) and 1e8 (C) ATP molecules per reaction volume (approximately 1 M, 5 M and 10 M ATP equivalents, respectively).

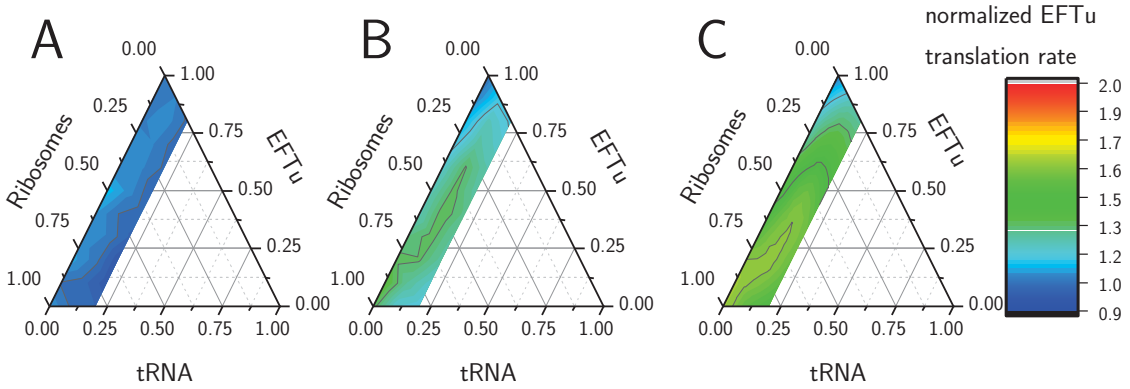


Figure 4.6: Distribution of ATP in tRNA, EFTu, and ribosomes and the resulting translation rate (normalized to unaltered conditions) for **EFTu** as a target protein. The amount of ATP distributed was 1e7 (A), 5e7 (B) and 1e8 (C) ATP molecules per reaction volume (approximately 1 M, 5 M and 10 M ATP equivalents, respectively).

Distributing a defined number of ATP equivalents to increase the concentrations of the molecules results in altered translation rates. Figures 4.5 and 4.6 show the resulting translation rates achieved by investing 1 M (A), 5 M (B), and 10 M (C) ATP equivalents, respectively. With ATP pool sizes for *E. coli* of 3.56 mM (Buckstein et al. 2008) and an ATP turnover of 311 min^{-1} (Holms et al. 1972), which relates to the ATP production of 0.9, 4.5, and 9 minutes, respectively. The results of this study show that, depending on the ratios applied, the resulting translation rate can be lower than the rate before altering the reaction composition, indicating that the interplay and balance of the three species is a critical element for the translation machinery.

Table 4.2: Peak translation rates achieved as a function of the number of ATP equivalents spent.

	GFP			**EFTu**		
ATP	**tRNA**	**EFTu**	**Ribosomes**	**tRNA**	**EFTu**	**Ribosomes**
1 M	0 %	50 %	50 %	0 %	50 %	50 %
5 M	10 %	60 %	30 %	0 %	10 %	90 %
10 M	10 %	50 %	40 %	0 %	10 %	90 %

Spending 1 M ATP equivalents on the translation machinery results in translation rates that are almost unchanged for both target sequences. With increasing ATP equivalents, the resulting translation rates significantly differ from the origin. For a broad range of distribution ratios, a constant or even decreased translation rate was observed. Only a small solution space shows increased translation rates for both GFP and EFTu as sequence. This effect increases with the number of ATP equivalents used. At 5 M, a sharp peak with significantly higher translation rates was identified, though the location of the peak differs for the two target genes. While the peak translation rate for EFTu can be found at high ribosome concentrations, the peak for GFP is at an even distribution between EFTu and ribosomes as well as a small share for tRNAs. According to the control analysis, the translation of EFTu is solely limited by ribosomes and the translation of GFP is limited by both ternary complexes and ribosomes.

Table 4.2 summarizes the peak translation rates and the corresponding distribution of ATP. These results indicate that with low ATP effort, the ATP is spent between EFTu and ribosomes to increase the translation rate. The number given to EFTu can be explained by the initial number of tRNAs, which is higher than the number of EFTu and ATP spent here and correlates with the number of EFTus needed to bridge this gap. For GFP, the shares remain relatively constant with increasing ATP, indicating limitations to the number of ribosomes and the elongation rate of each ribosome. However, as more ATP is spent, the ratio invested in ribosomes increases in the case of EFTu since the elongation rate is almost saturated compared to GFP synthesis.

Regarding protein synthesis that is not optimized such as GFP production, the inefficient codon usage must be overcome by increasing the number of ternary complexes and ribosomes. Adding both of these factors reduces the average distance that ternary complexes need to travel. On the other hand, efficiently expressed

proteins such as EFTu have a decreased average distance due to the higher number of ternary complexes suitable for a given codon. Therefore, a ternary complex hits far fewer unsuitable ribosomes before encountering a ribosome with the corresponding sequence.

Starting from a higher elongation rate, optimizing EFTu translation rates only benefits from increasing the number of ribosomes, whereas optimization of GFP translation is hampered by elongation rates which must be optimized as well. Although the costs to create further ternary complexes are orders of magnitude lower than the costs to create ribosomes, these results indicate that increasing the translation rate by increasing ternary complex concentrations is more expensive than investing in ribosomes.

This effect can also be seen in *E. coli* as published by Rudorf & Lipowsky (2015). They show that the relative increase in ribosomes is significantly higher than the increase in EFTu and tRNA with an increasing growth rate (figure 4.7). In summary, these results indicate that *in-vivo* translation is hampered by elongation velocity, but at higher growth rates, it is more economical to invest in expensive ribosomes than inexpensive ternary complexes.

The optimization of translation rates will differ depending on the GOI that will be expressed. For proteins with less optimized codon usage, translation rates are limited by elongation rates and the total number of translating ribosomes. Highly optimized proteins do not suffer from low elongation rates and translation rates are only limited by the number of ribosomes. Thus, *a priori* protein optimization can circumvent the limitation of eased elongation rates and provide an efficient tool for increasing translation rates without intervening in cellular mechanisms.

Although this model requires few parameters, the predictive quality of this approach is unique. Without any information about molecule interaction and relying

on diffusion as the sole mechanism, this approach can quantitatively describe translation elongation.

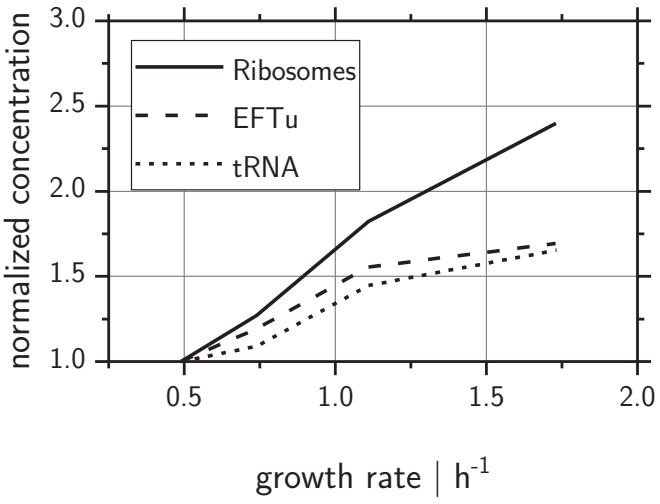


Figure 4.7: *In vivo* concentrations of ribosomes, EFTu, and tRNA as published by Rudorf & Lipowsky (2015) and normalized to the concentration of μ=0.49.

4.2 Optimizing cell-free protein synthesis

The diffusion-based analysis of *in vivo* and *in vitro* translation elongation revealed a transport bottleneck of ternary complexes. One of the main assumptions of this model was the constant number of actively translating ribosomes since transcription, translation, initiation, and termination were neglected in this approach. However, the number of actively translating ribosomes depends on these neglected mechanisms. An expanded model is necessary to assess the influence of initiation and termination as well as transcription. Since many translation factors are necessary in addition to the large computational demand of the SSA, a new modeling strategy was developed. Based on the published model by Arnold et al. (2005), the transcriptional level of detail in the model was reduced in this study. Changing the sequence-oriented

approach to transcription and mRNA degradation towards a Michaelis-Menten-like black-box description allowed for a reduction from 1145 equations to 410 equations (see chapter 3.1). In addition, changes in ribosome concentrations between lysate batches were mirrored by including a dilution factor from *in-vivo* to *in-vitro* conditions based on the measured ribosome concentration (Nieß & Failmezger et al. 2017). This allows the simulation to be adapted for varying and altered reaction conditions (e.g. supplementation of translation factors, and over-proportional loss of translation factors during lysate preparation). These modifications provide the acceleration of computational speed required for a detailed sensitivity analysis without sacrificing significance when describing translation. For their predictions, Arnold et al. (2005) used degradation mechanisms for several translation factors which induced termination of the CFPS system, and the respective degradation constants were fitted to their CFPS experimental results. However, the literature showed that CFPS reactions can be prolonged for up to ten hours without loss in STY. Therefore, translation factor degradation and energy regeneration were neglected in this system and the reaction rates were evaluated at steady-state.

4.2.1 Model validation

Using the CFPS system designed by Jurek Failmezger (Failmezger et al. 2016; Failmezger et al. 2017) resulted in a ribosome concentration of 1.7 ± 0.1 µM in the final reaction. An over-proportional loss of enzymes was only detected for ribosomes ($R^R = 0.9$) and EFTu ($R^{EFTu} = 0.8$) as described by Nieß & Failmezger et al. (2017). The evaluation of the CFPS reaction performance is summarized in table 4.3 where *in silico* predicted and *in vitro* measured translation rates are presented. For rate evaluation, only phase II with the highest translation rates was selected, since phase I is hampered by slow initial rates and phase III is hampered by disintegration of the translation machinery.

Table 4.3 shows that the model overestimates the translation rate by 63 %. GFP as a target protein has many advantages for experimental procedures, such as online measurement during CFPS reactions. However, several studies have shown that only a fraction of the synthesized GFP is fluorescent and active (Iskakova et al. 2006; Caschera & Noireaux 2014), indicating that online measurement of GFP is likely to underestimate the GFP concentration. Direct measurement techniques such as ^{14}C-Lysine incorporation circumvent the impact of protein folding. In summary, this model can predict *in vitro* translation rates with high accuracy without parameter estimation and therefore, it provides a versatile simulation platform to analyze CFPS systems and identify potential targets that can improve translation rates.

Table 4.3 Comparison of experimentally determined parameters of *in vivo* and *in vitro* translation, and simulated *in vitro* reaction parameters. *In vivo* data taken from Bremer & Dennis (1987)[a] and Underwood et al. (2005)[b].

	in vivo	***in silico* prediction**	**Measured Nieß & Failmezger et al. (2017)**
Ribosome concentration | µM	42[a]	1.7	1.7 ± 0.1
Number of actively translating ribosomes | %	80[a]	82	n.d.
Elongation rate | AA s^{-1}	18[a]	1.12	n.d.
Initiation rate | nM s^{-1}	$2.8 \cdot 10^{3}$ [b]	7.1	4.3 ± 1.07
Bulk protein synthesis rate | mg L^{-1} min^{-1}	$3.3 \cdot 10^{3}$ [b]	10.6	6.5 ± 1.6

4.2.2 Identification of rate-limiting co-factors

As described in Nieß & Failmezger et al. (2017), the identification of potential targets to improve CFPS performance was performed in two steps. An initial screening of the potential targets (all initiation factors, EFTu, EFG, EFTs, all tRNA species, RF, T7 RNA polymerase, and ribosomes) was performed to sort them into

the following categories: positive influence on the translation rate, no influence on the translation rate, and negative influence on the translation rate. Due to this screening, only EFTu and tRNA significantly enhance the translation rate, while ribosomes and T7 RNA polymerase decrease the translation rate. Increasing ribosome and T7 polymerase concentrations results in a higher number of actively translating ribosomes which leads to more tRNAs bound to ribosomes. Combined with the inefficient codon usage, a shortage of certain tRNAs arises which limits the translation elongation.

As stated previously, tRNA and EFTu are the translation factors that offer the potential to increase the CFPS translation rate. As both molecules (EFTu GTP bound and tRNA aminoacylated) form the ternary complex (T3 complex) that serves as one of the key substrates for translation elongation, their combinatory effect was further analyzed. Figure 4.8 shows that the simultaneous addition of EFTu and tRNA exceeds the increase in translation compared to adding one of them alone. These results directly indicate that *in vitro* protein synthesis is limited by the haulage of amino acids to the ribosome and strongly influenced by gene sequence and tRNA composition.

Given the potential translation rate increase by the simultaneous addition of EFTu and tRNA, a new experimental design was postulated. By doubling the number of both translation factors, *in silico* simulations predict an increase of 89 %. The concentrations were doubled due to volume constraints in the experimental reaction composition. Experimental validation of the postulated translation rate increase showed that 62±9 % of the postulated translation rate increase was achieved (Nieß & Failmezger et al. 2017).

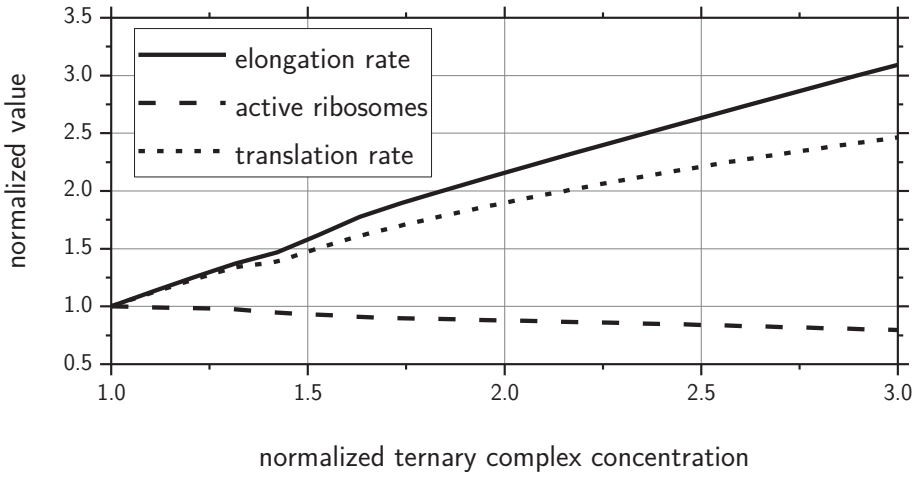


Figure 4.8: Influence of simultaneous increase of EFTu and tRNA concentration (normalized to reference values) on the resulting translation rate (dotted line), mean elongation rate (solid line) and number of active ribosomes (dashed line). All values were normalized to values at unaltered conditions.

Figure 4.8 shows that the number of actively translating ribosomes decreases, indicating that initiation of new ribosomes becomes increasingly relevant with higher concentrations of ternary complexes. To address this effect, the translational control for varying ternary complex concentrations were analyzed as described in Zouridis & Hatzimanikatis (2008). According to Nieß & Failmezger et al. (2017), the elasticities ϵ_i and the flux control coefficient *(FCC)* were evaluated. ϵ_{T3} and ϵ_{IF} describe the sensitivity of the translation rate over ternary complex concentration (C_{T3}) and initiation factor concentration, respectively. The *FCC* describes the sensitivity of the translation rate to changing ribosome concentrations. Accordingly, high values of ϵ_{IF} and *FCC* indicate the limitations of ribosomes on the mRNA and therefore, the potential limitations of translation initiation, whereas a high ϵ_{T3} indicates a limitation of translation elongation. Figure 4.9 shows that with rising ternary complex concentrations, a shift from elongation towards initiation as the limiting step of the translation rate occurs. Hence, the CFPS translation rate can be

optimized, but initiation as the next limitation results in fewer actively translating ribosomes. Therefore it is a much more promising approach to increase the concentration of all necessary translation factors, or more practically, avoiding dilution during reaction preparation.

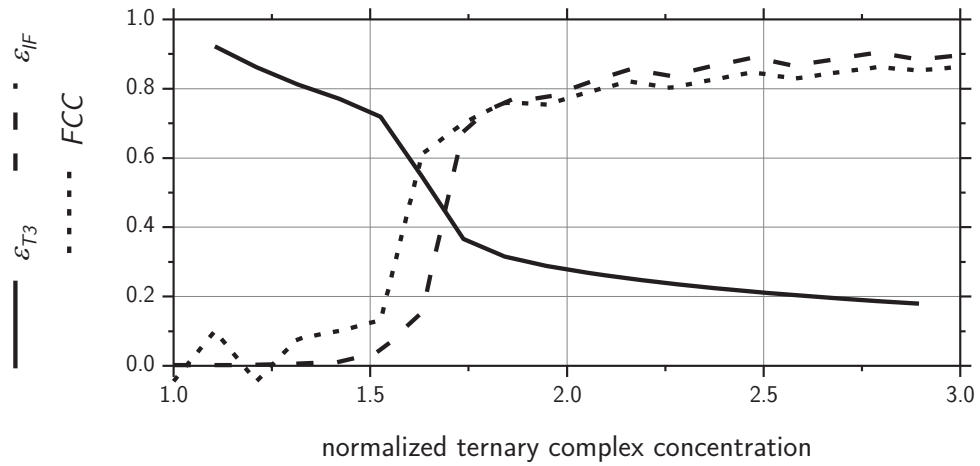


Figure 4.9: Calculated elasticities ϵ_{T3} and ϵ_{IF} and flux control coefficient (FCC) as a function of increasing normalized ternary complex concentrations. Elasticities ϵ_{T3} and ϵ_{IF} and FCC represent the sensitivity of translation rate with respect to the varying total ternary complex, initiation factor, and ribosome concentration.

4.2.3 Codon usage

Using *in silico* prediction of the translation machinery allows for detailed tracking of ribosomes and their respective location on the mRNA template. Figure 4.10 shows that more than 75 % of active ribosomes are located in the first 30 % of the coding sequence, confirming the unfavorable codon usage of the GFP sequence. The course of the cumulative distribution function (CDF) is far from an ideal linear slope, describing a homogeneous distribution.

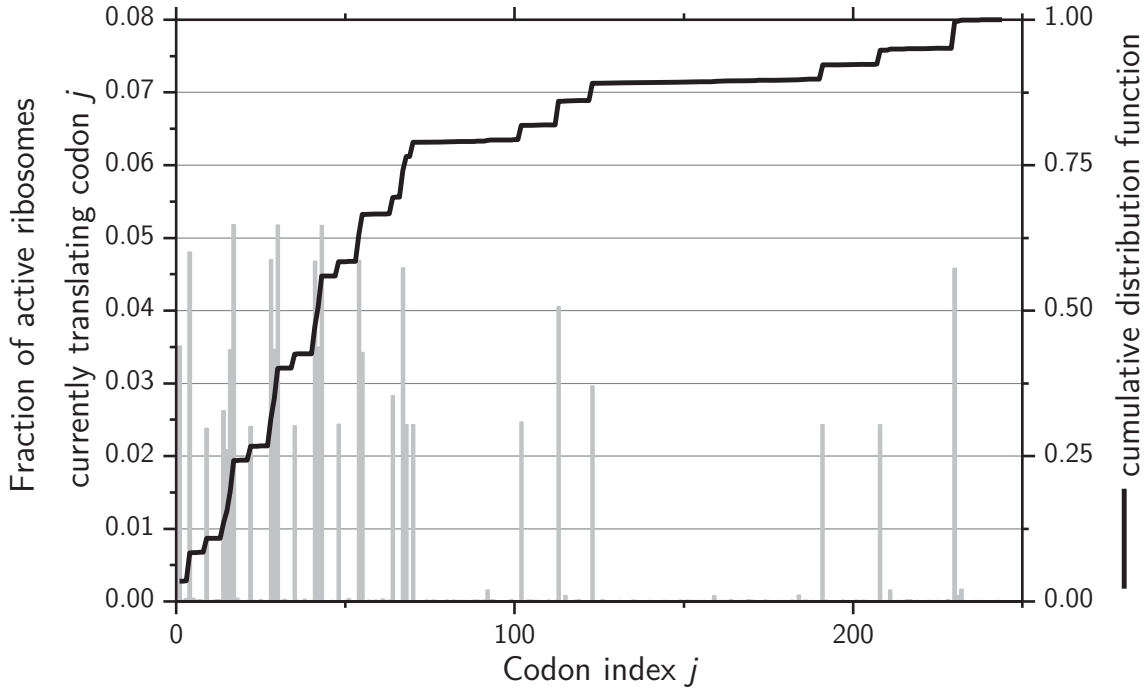


Figure 4.10: Detailed distribution of actively translating ribosomes, the respective current positions on the mRNA template, and the associated cumulative distribution function.

Using the CAI-optimized GFP variant described in chapter 4.1.3 results in a different distribution of ribosomes as shown in figure 4.11. The ribosome distribution along the mRNA is broadened and the cumulative distribution function has a lower slope at the beginning, which results in 82 % higher elongation and translation rates due to the optimized distribution of ribosomes along the mRNA. *In vivo* protein synthesis rates are highly influenced by codon usage, but since the GOI is not the only protein undergoing translation, the influence on tRNA usage is less decisive. Compared to *in vivo* translation, *in vitro* protein synthesis only translates a single gene species at once. Therefore, the effect of inefficient codon usage is stronger. In general, codon optimization provides a simple tool for optimizing *in vitro* STY without interfering with the reaction composition.

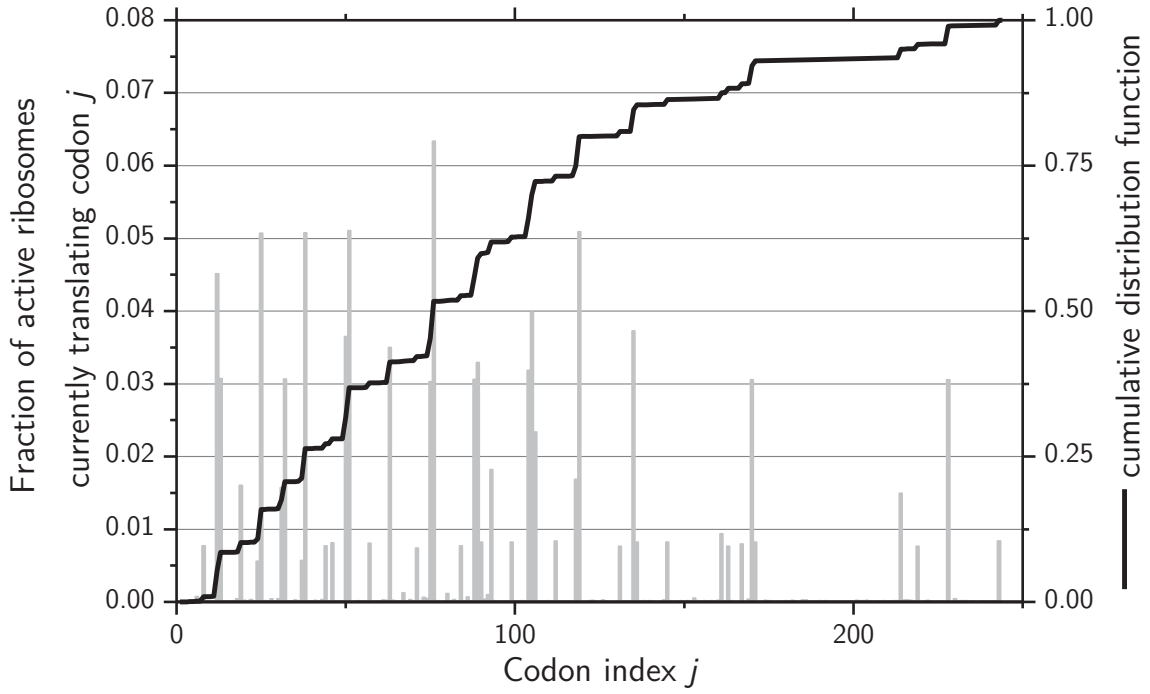


Figure 4.11: Detailed distribution of actively translating ribosomes and the respective current positions on the mRNA template for the codon-optimized sequence.

4.2.4 Estimating the potential of cell-free protein synthesis

The sequence oriented model (chapter 3.1) was designed to use *in vivo* concentrations for simulations. Modifying the dilution factor allows for the estimation of *in vitro* protein synthesis under *in vivo* conditions. The first estimation of *in-vivo* translation rates can be performed using the data published by Bremer & Dennis (1987) and Rudorf & Lipowsky (2015). Equation 4.1 results in a translation rate of approximately 145 µM s^{-1}:

$$v_{TL} = C_R^{active}\ \bar{v}_{TLE}\ N_{AA,GFP}^{-1} \qquad 4.1$$

with: $C_R^{active} = 33.85$ µM (Rudorf & Lipowsky 2015)

$\bar{v}_{TLE} = 18\ AA$ ribosome^{-1}s^{-1} (Bremer & Dennis 1987)

$$N_{AA,GFP} = 245 \text{ AA eGFP}^{-1}$$

Simulations neglecting CFPS dilution ($D = 1$) and assuming the translation of a single mRNA only showed significantly lower translation rates as the initial approximation (table 4.4, optimal conditions). Either a low number of ribosomes with a high elongation rate (table 4.4, low TC) or a high number of ribosomes with a low elongation rate (table 4.4, high TC) were found, depending on the assumed T7 concentration. Both operating points showed the same translation rate. In principle, translation rates of 3.68 g L^{-1} min^{-1} are feasible with perfectly distributed codon usage and sufficiently long sequences. In reality, such translation rates are unlikely due to the constraints imposed by the plasmid used for CFPS and the uncoupling of TC-TL. Thus, protein synthesis rates of 0.8 g L^{-1} min^{-1} are plausible for undiluted CFPS systems.

Table 4.4: Translation rates under *in vivo* conditions. Optimal conditions show values from the literature (Bremer & Dennis 1987). Simulations of *in vivo* conditions ignore dilution and investigate the impacts of transcription (TC) as indicated.

	Optimal conditions	**Simulated *in vivo* low TC**	**Simulated *in vivo* high TC**	**Simulated *in vitro***
v_{elo} **| AA rib^{-1} s^{-1}**	18	14.9	3.8	1.12
Active ribosomes | %	80	21	83	82
v_{TL} **| µM GFP min^{-1}**	145	30.3	30.8	0.38
Bulk protein synthesis rate | g L^{-1} min^{-1}	3.68	0.770	0.782	$9.65 \cdot 10^{-3}$
Dilution factor	1.0	1.0	1.0	21.08

4.3 Gene expression in microbial populations

The previous chapters regarded TC-TL for the synthesis of a single gene. In this chapter, the model scope is expanded again to cover more genes and the induction of protein synthesis.

STR-PFR systems are a common tool for modelling large-scale substrate gradients (George et al. 1993). Based on an STR chemostat cultivation coupled to an aerated PFR device, Löffler et al. (2016) modelled the impact of large-scale conditions with a lab-scale setup. This setup permits discrete analysis of the STR population and the population along the substrate gradient installed in the PFR section. Based on the steady-state conditions in the STR, small deflections of the PFR can be analyzed. By applying transcriptome analysis, the detailed reaction of *E. coli* along the PFR and in the STR was observed. They clearly indicated that oscillation substrate conditions substantially impact the transcriptional pattern of *E. coli*. The response to the substrate gradient shows elevated transcript levels after less than one minute. Further investigation of the STR population revealed that the transcript composition changes after cultivation duration and the population at the end of the cultivation has a significantly different composition.

Based on the dataset published by Simen et al. (2017) and Löffler et al. (2016) an ensemble model (Henson 2003) was established to further investigate the impacts of oscillating substrate conditions on the cellular reaction. Furthermore, the short-term and long-term reactions are analyzed.

Simen et al. (2017) showed that the *trp* operon is directly transcribed along the PFR and it was used as an exemplary operon in this study. It consists of the leader peptide *trpL* and the downstream structural genes *trpEDCBA*. Another advantage of the *trp* operon is the attenuation-based regulation (Yanofsky 2004, 2007), which indicates ongoing translation if transcripts after *trpL* are measured. In summary, the following key assumptions were made:

1. Once transcription of mRNA began, it continued until the stop signal was reached at the end of the operon, namely on the relative position 6726 nt after *trpEDCBA* (Stoltzfus et al. 1988).

2. mRNA was assumed to be immediately translated into proteins due to the attenuation system.
3. The number of translations per mRNA molecule was calculated based on the experimental findings of Valgepea et al. (2013).

4.3.1 Cellular reaction to the substrate gradient

The short-term reaction of the cell describes the reaction to the substrate gradient and the following transcriptional and translational change. Simulations of these dynamics were performed by applying the single-cell model as described in chapter 3.3 with the parameters in table 3.7 and the simulation results are shown in figure 4.12 and 4.13. For reference, the starting time for the simulation corresponded to the time a cell enters the PFR. Since the cell initially reacts to the limitation by inducing transcription, *trpL* is the first gene transcribed. As indicated by figure 4.12, transcription of *trpL* is complete and *trpE* is partly transcribed at the end of the PFR. These results directly show that the downstream genes of *trpDCBA* are not transcribed in the PFR, but in the STR. Furthermore, the appearance of *trpL* indicates that ribosomes are already bound to the mRNA and translation is ongoing.

After exiting the PFR, the degradation of mRNA is directly visible by the declining courses of *trpL* and subsequent genes. The total time required for TC-TL is approximately 8.4 minutes after the cell enters the PFR and no mRNA is visible. This also reveals that cells reentering the PFR no longer have ongoing protein synthesis. However, this model relies on the assumption of constant residence times in both reactor compartments. This assumption is valid for a PFR, but for the ideally mixed STR, a residence time distribution describes the behavior of cells more accurately. Based on this, there is a share of cells that might reenter the PFR even if transcription remains ongoing.

One key finding of this simulation is that transcription is induced in the PFR, but the main part of transcription and translation is performed when the cells enter the STR and the caused metabolic and energetic demand is delayed. This finding is supported by the findings of Löffler et al. (2016) and Simen et al. (2017). Accordingly, further investigation into oscillating substrate should focus on the PFR device and on the main population.

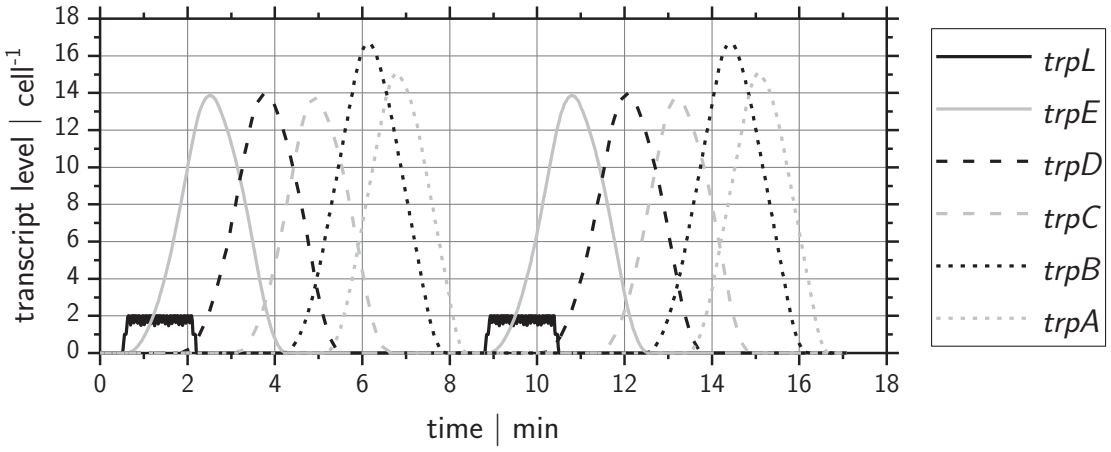


Figure 4.12: Time courses of two subsequent cell cycles comprising PFR and STR passage. mRNA profiles are shown, simulated using the single-cell model. The genes are encoded as follows: Black solid line=*trpL*, grey solid line=*trpE*, black short dashed line=*trpD*, grey short dashed line=*trpC*, black long dashed line=*trpB*, grey long dashed line=*trpA*. Grey shaded areas indicate that the cell currently passes the PFR (Nieß & Löffler et al. 2017).

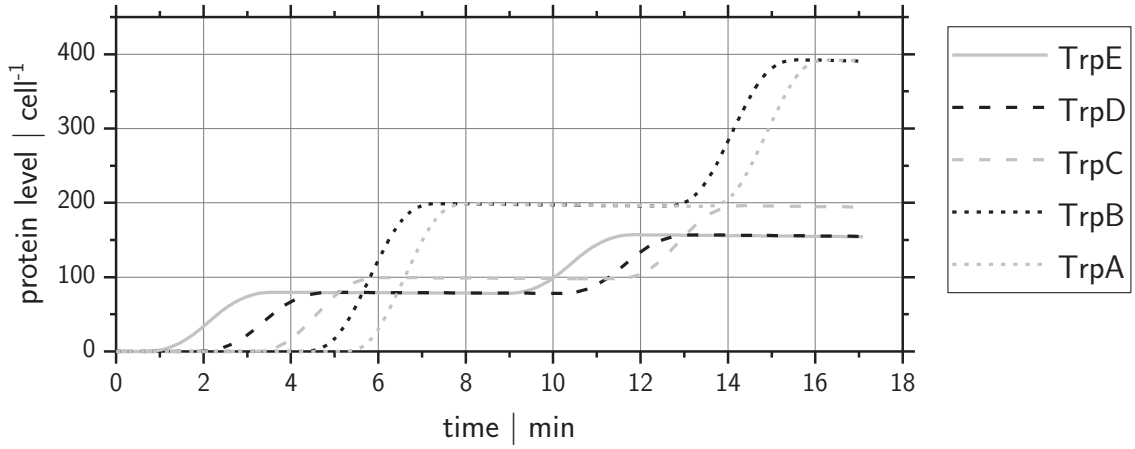


Figure 4.13: Time courses of two subsequent cell cycles comprising PFR and STR passage. protein profiles are shown, simulated using the single-cell model. The proteins are encoded as follows: grey solid line=TrpE, black long dashed line=TrpD, grey long dashed line=TrpC, black short dashed line=TrpB, grey long short line=TrpA. Grey shaded areas indicate that the cell currently passes the PFR (Nieß & Löffler et al. 2017).

Rising levels of *trpL* show that translation is already ongoing inside the PFR because the ribosomes are bound to the mRNA to facilitate the transcription of

trpEDCBA. Protein formation patterns were investigated based on the single-cell model. Along the substrate gradient, protein patterns follow the pattern of their respective mRNAs with a small delay, and most of the translation occurred in the STR, which was observed for mRNA too. In comparison to mRNA, no sharp peak in the respective protein concentration was observed due to the degradation constant of proteins, which is orders of magnitude lower than the degradation constant of mRNA. The slope of protein courses is generally narrower than that of mRNA courses. The main visible effect is that protein levels vanish with mRNA levels and therefore, proteins accumulate over time.

An STR-PFR cycle takes approximately 500 seconds (125 seconds in the PFR and 6.2 min in the STR), during which 20 mRNA copies of the *trp* operon are formed. The subsequent translation results in the formation of 80 TrpED, 100 TrpC, and 200 TrpBA copies, which correspond to translation rates of 9.6, 12, and 24 proteins per cell per minute, respectively. Furthermore, protein patterns simulated by the single-cell model rely heavily on the degradation constant for proteins, which is afflicted with errors. In the previous case, only cytosolic dilution by cell division was assumed and active protein degradation was neglected. Depending on the cell state, active degradation might or might not occur. Therefore, the impact of additional active degradation was investigated. Steady-state protein levels were estimated in the first investigation.

$$\frac{dC_{Protein}}{dt} = v_{Translation} - v_{deg} = r_{Translation} - C_{Protein}(\mu + k_{deg}^{Protein}) = 0 \qquad 4.3$$

$$C_{Protein} = \frac{v_{Translation}}{\mu + k_{deg}^{Protein}} \qquad 4.4$$

Solving equation 4.3 for the protein concentration $C_{Protein}$ results in an algebraic equation where protein levels are a function of translation rates, growth rates, and

the degradation constant. Translation and growth rates are system constants and only the degradation constant impacts the protein levels. Figure 4.14 shows the calculated protein levels as a function of the degradation constant, and measured values are illustrated for the same growth rate of 0.2 h^{-1}. These simulations indicate that degradation constants lower than 0.6 h^{-1} (corresponding to half-lives below 1.1 h) result in higher protein levels than reported. Half-lives as low as 1.1 h^{-1} are also reported in literature (Nath & Koch 1970; Lahtvee et al. 2014). Based on cycle times of 500 seconds, even low protein half-lives of 1.1 h are significantly higher and the influence of active protein degradation can be neglected.

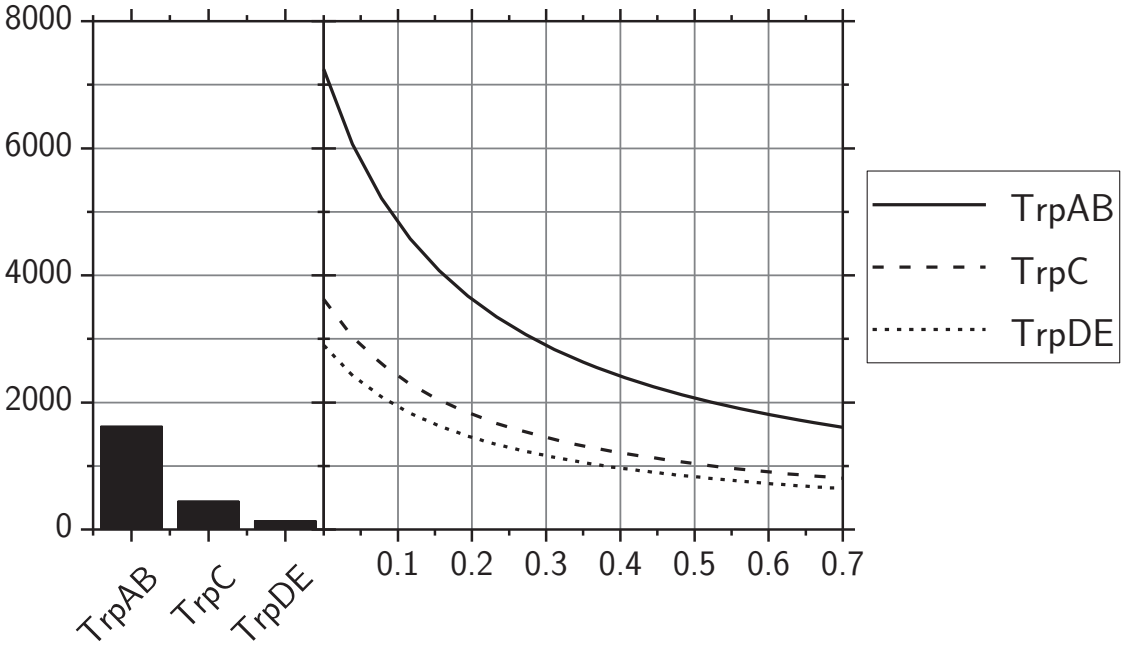


Figure 4.14: Protein levels for the new steady-state after > 15 h as a function of the degradation constant (solid lines;(Nieß & Löffler et al. 2017). Higher values show the influence of declining half-lives and thus higher degradation constants. Measured protein levels (bar graph) were taken from Valgepea et al. (2013).

4.3.2 Mechanism of long-term adaptation

The results of Löffler et al. (2016) and Simen et al. (2017) were deduced from chemostat experiments with an additional PFR to implement the substrate gradient. With and without the PFR device, the same system dilution rate of 0.2 h^{-1} was installed. The initial steady-state before connecting the PFR device allows for a detailed reference state and the chemostat operation mode allows for much longer cultivation times than fed-batch STR-PFR experiments. Using this methodology, cultivations of 28 h were performed to assess long-term adaptation. For glucose and nitrogen limitations, short-term and long-term adaptations were found.

To simulate the entire STR-PFR system, an initial 10,000 cells were used which resulted in a stable steady-state with 7526 ± 68 cells in the STR (75.0 ± 0.68 %) and 2513 ± 47 simulated cells in the PFR (25.0 ± 0.47 %). These values coincide with the volume distribution ratio in the experimental setup (74.7 vol % in the STR and 25.3 vol % in the PFR). Compared to the single-cell model where residence time distribution was neglected, this ensemble model allows for the analysis of population heterogeneities due to differing residence times in the STR. At steady-state, approximately 34% of the STR population is not induced, 48 % is undergoing a single induction, and 18 % is induced multiple times as shown in figure 4.15. Multiple inductions imply that cells reentered the PFR before their current TC-TL cascade was finished.

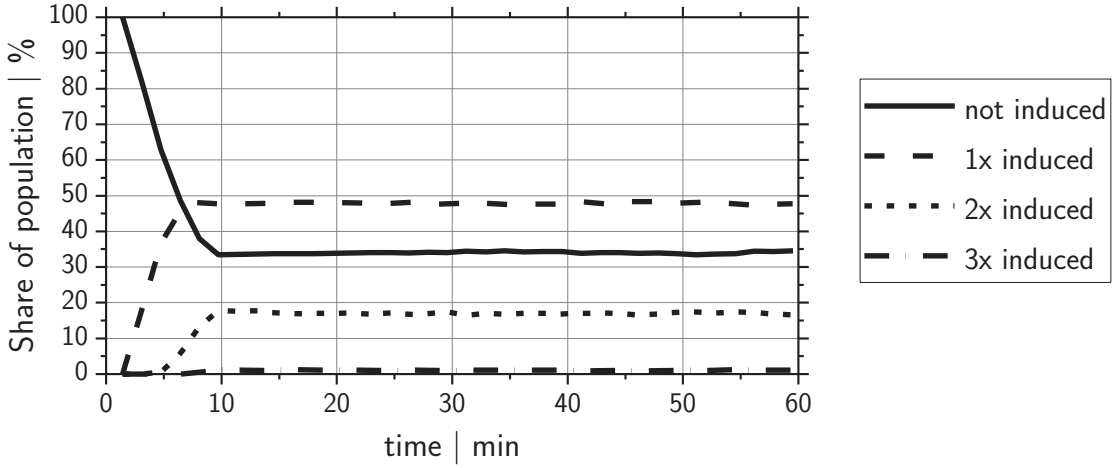


Figure 4.15: Population heterogeneity of the STR population regarding the number of inductions as discrimination criteria.

Figure 4.16 shows the predicted transcript patterns along the PFR for the entire population and simulated values were scaled to mean *trpLE* levels to allow for comparison. The ensemble model can predict the patterns of transcript reactions along the PFR, and further analysis of the STR transcriptional response is shown in Figure 4.17. Here the transcriptome of the STR population for *in silico* predicted and experimentally measured values are shown (simulated values were scaled to *trpE* for comparison). Again, the ensemble model shows the same patterns for the transcripts for the measured time points of 25 minutes, 120 minutes, and 28 hours. While traces of *trpL* are predicted, the downstream genes of *trpEDCBA* show elevated levels. Due to the marginal change over time in the transcript patterns, long-term adaptation is likely not caused by the transcript response alone.

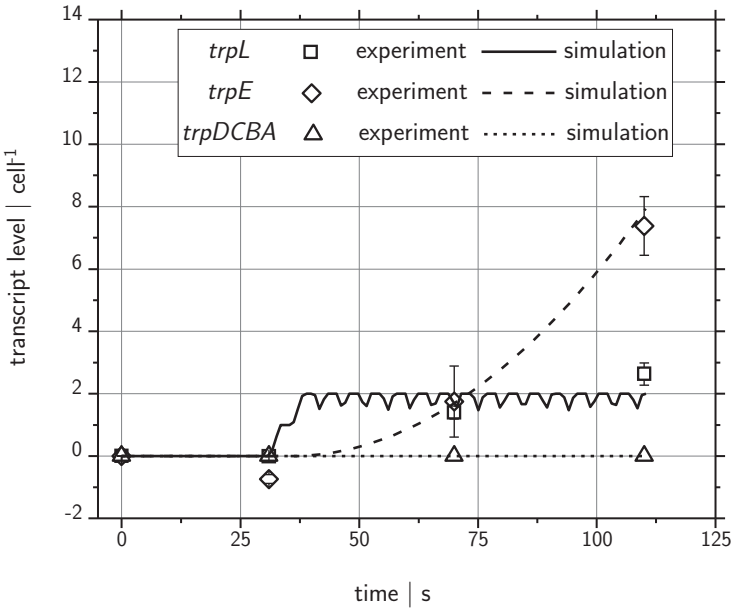


Figure 4.16: Predicted *trp* operon transcript levels for PFR residence times of 30, 71, and 110 seconds compared to measured values (scaled to mean *trpLE* level) compared to measured values (scaled to simulated *trpE* levels; (Nieß & Löffler et al. 2017).

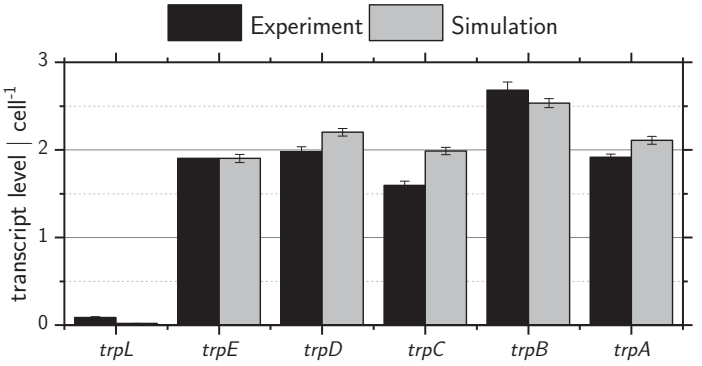


Figure 4.17: Predicted *trp* operon transcript levels for the STR population compared to measured values (scaled to simulated *trpE* levels; (Nieß & Löffler et al. 2017).

The dataset of Simen et al. (2017) shows that the *his* operon is upregulated during STR-PFR cultivations in the same manner as the *trp* operon. The *his* operon provides

the structural genes necessary for the histidine synthesis from phosphoribosyl pyrophosphate as a precursor metabolite. In addition to the *trp* operon, the *his* operon was investigated based on the same model assumptions. Figure 4.18 shows the transcriptional response along the PFR and figure 4.19 shows the transcriptional response of the STR population for the three time points. As shown for the *trp* operon, the prediction along the PFR shows conformity with the measured values. For the STR population prediction, the genes *hisDC* show slightly higher values than measured, but the pattern is still in accordance with the measured data. Downstream genes of *hisBHAFI* were neglected in this study, due to an additional promotor sequence after *hisB* that is not described by this model. As shown by the results for the *trp* and *his* operons, this model can predict PFR and STR population transcriptome changes for different genes that are induced by the substrate gradient.

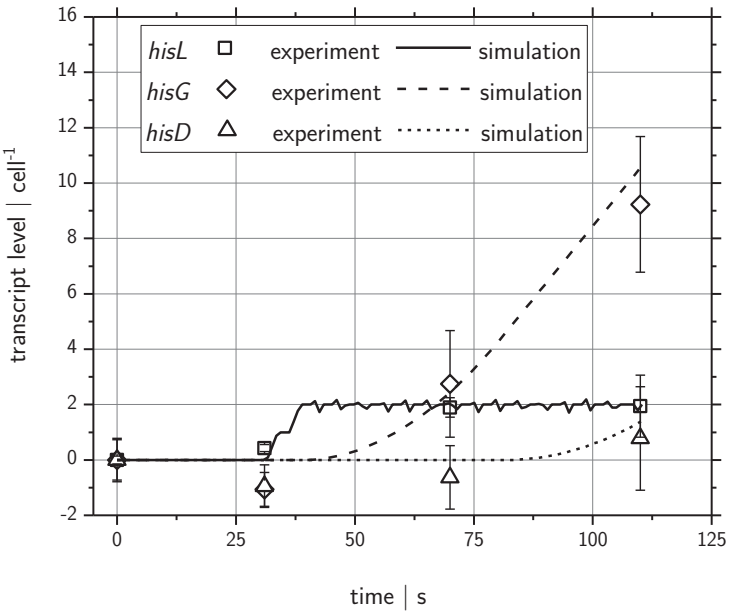


Figure 4.18: Predicted *his* operon transcript levels for PFR residence times of 30, 71, and 110 seconds compared to measured values (scaled to mean *hisLG* level)

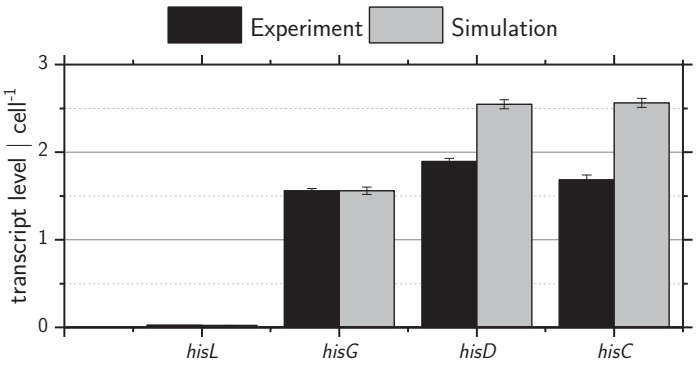


Figure 4.19: Predicted *his* operon transcript levels for the STR population compared to measured values (scaled to simulated *hisG* levels)

An analysis of protein composition was performed for the TrpA structure protein and the corresponding transcript, and additional protein degradation was neglected as described earlier. Figure 4.20 shows the patterns for both protein and mRNA. First transcript levels achieve a pseudo steady-state for the population after approximately ten minutes (slightly more than the sum of both residence times),

then protein levels slowly approach a steady-state level which takes more than 15 hours.

A comparison of the simulation results and experimentally observed transcript data revealed that the short-term response along the PFR can be predicted, as well as the impact on the STR population. The simulations and measured values indicated two distinct velocities of adaptation: short-term adaptation along the substrate gradient which results in transcriptional changes and long-term adaptation caused by the repetitive stimuli which slowly increases protein concentrations. These results indicate that long-term adaptation might be caused by the slowly rising change in protein levels due to the short-term response to oscillating substrate conditions. The simulation results indicate that the short inductions cause population heterogeneity. The population divides into cells that are induced and those that are not induced. Furthermore, the different number of inductions and PFR passages that a cell was exposed to resulted in different numbers of proteins in the cells. There are two distinct species for transcriptome, induced and not-induced, and cells fade between both states. For the proteome, no distinct pattern can be observed since cells accumulate proteins with different velocities depending on the number of PFR passages they are subject to.

Finally, this study suggests that prospective STR-PFR experiments and scale-down experiments in general can expand their focus from analysis-only transcript levels to incorporate proteome analysis. Further investigation of population heterogeneity by single-cell analysis can unravel the causes of long-term adaptation caused by oscillating substrate levels.

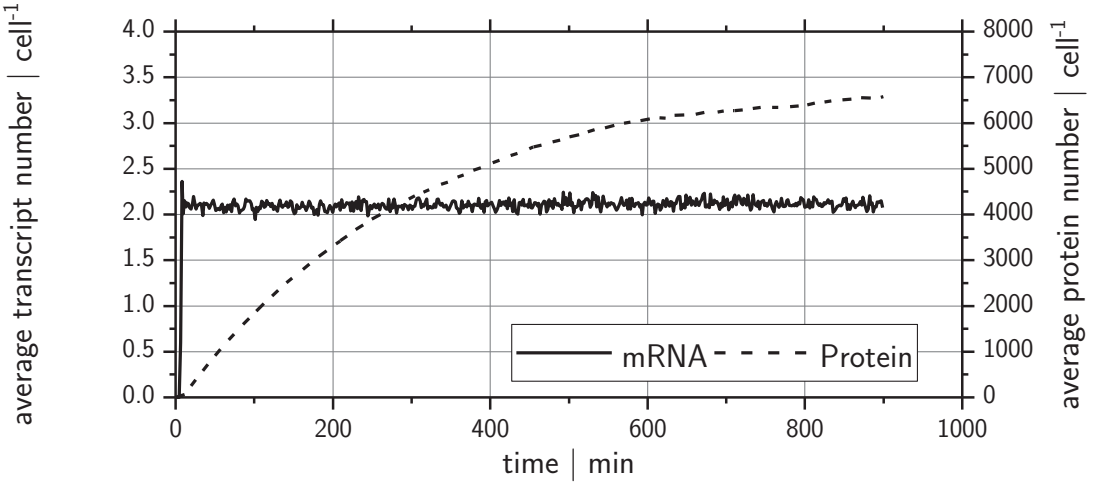


Figure 4.20: Long-term prediction of transcript and protein levels neglecting protein degradation. mRNA levels become stable levels after ~7 minutes whereas protein leveling takes more than 15 hours (Nieß & Löffler et al. 2017).

CHAPTER 5

CONCLUSION AND PERSPECTIVES

This chapter summarizes the results of the present study. Considering the objectives of this study, conclusions are provided as well as perspectives for future work in this area.

5.1 Diffusion limitation during translation

In the first part of this study, the influence of ternary complexes on the translation elongation rate was further investigated. Based on the assumption that diffusion is the limiting step in translation elongation, a simulation platform was developed to estimate the translation rate as a function of diffusive transport.

Cellular reactions are often limited by the transport of substrates towards the enzyme due to the absence of convection in most biological systems, and translation is no exception in this context. CFPS reactions provide a platform for open protein biosynthesis in lab scales, and they suffer from the same drawback if no external convection is applied. As described by Nieß & Failmezger et al. (2017), *in vitro* translation is limited by the ternary complex concentration. Elongation rates that are considerably lower than the maximum reaction velocity indicate the limitation

of substrates, not catalytic reaction time. As dilution is an inevitable operation in CFPS reactions, it can be regarded as a significant influencer of the translation rate. To address this problem, a simulation platform was established to describe translation as diffusion-collision driven.

Ternary complexes oblige to Brownian motion and can be described by a random walk algorithm. Collision between potential reaction partners was treated as a trigger for reactions. Furthermore, sequence-oriented elongation included the aspect of codon usage and tRNA distribution in the elongation rate. Necessary initiation values were taken from Nieß & Failmezger et al. (2017), and transition from *in vivo* to *in vitro* was incorporated by including a dilution factor.

Comparison with literature values (Underwood et al. 2005) for the CAT sequence shows agreement between *in silico* and *in vitro* elongation rates. Based on the hypothesis that translation is limited by the diffusion of ternary complexes even under *in-vivo* conditions, the impact of ribosomes and ternary complexes was assessed as a function of the dilution factor. This dilution occurs during cell-free reaction preparation and it is almost inevitable. For GFP as a target sequence, the flux control and elasticity of ternary complexes remain almost constant for a broad range of dilution factors indicating that the translation of GFP is limited by ribosomes and ternary complexes. Alteration of the target sequence to EFTu changes the resulting patterns. At low relative concentrations, translation elongation is limited by ribosomes and ternary complexes again. However, with rising concentrations, the impact of ternary complexes vanishes due to the saturation of ribosomes with ternary complexes under *in vivo* conditions. As a result, optimizing strategies to increase translation rates differ between varying target sequences.

The foundation of this model is the three species, ribosomes, tRNA, and EFTu, as the core machinery of translation elongation. Increasing the concentration of these three species might increase translation rates. The cost to synthesize these molecules

differs by microorganisms; tRNA as a molecule is small and cheap to synthesize and EFTu is more expensive, but ribosomes are drastically more expensive than the other molecules due to their size. For comparison, the increase in molecules in the system was normalized to the costs of ATP equivalents. Changing a single molecule concentration revealed three effects during the simulations of *in-vivo* protein biosynthesis: translation rate scales linearly with the number of ribosomes, elevating EFTu levels increases the translation rate to a certain point and then has no further impact, and tRNA shows an optimal concentration for peak translation rates and higher or lower concentrations decrease the translation rate. Moreover, the peak of the tRNA influence is almost at the same concentration that naturally occurs in microorganisms and the plateau of EFTu concentration is only slightly higher than the *in-vivo* concentration. Such behavior indicates that the cellular composition of the translation machinery is almost at an ideal composition for the most efficient transport of ternary complexes. This effect also shows the same pattern for GFP and EFTu as target sequences despite the significantly different translation rates.

The translation rate was optimized by simultaneously adding all three species. Given a defined number of ATP equivalents, the share of ATP spent between the molecules significantly changes the resulting translation rate. For the GFP sequence, the ideal ratio of molecules remains almost constant for a broad range of ATP, which corresponds to the results of the control analysis. Both the number of ribosomes and the number of ternary complexes increase the translation rate, and the distribution is only a function of the molecule costs. However, a different optimization pattern arises for the EFTu sequence. When little ATP is spent, an even distribution between EFTu and ribosomes occurs and with rising ATP levels, the ratio of ribosomes increases again coinciding with the results of the control analysis. The absolute amount spent on EFTu remains constant and is equivalent to the gap between the tRNA and EFTu of the *in vivo* concentrations.

In summary, both sequences follow the pattern of optimization that was predicted by the control analysis. GFP lacks efficient elongation and therefore, both ternary complexes and ribosomes can increase the translation rate. On the other hand, EFTu starts at higher elongation rates, the impact of increasing ternary complexes vanishes, and only ribosomes increase their translation rate further. This effect is also seen in *E. coli* with different growth rates. Increasing growth rates from 0.5 h^{-1} to 1.7 h^{-1} results in 2.5 times the ribosome concentration, but only 1.6 times the concentration of tRNA and 1.7 times the concentration of EFTu. Finally, these results show that diffusion is one of the key problems of translation elongation. The lack of convective transport in the cytosol of microorganisms is known to cause transport problems on the metabolic level. However, with ascending molecule sizes, the diffusive transport is increasingly truncated resulting in as much ATP as required for the translation machinery to increase diffusive transport. This shows that diffusive transport in the translation machinery is a problem for growing microorganisms and it increases with ascending growth rates.

This simulation framework provides a versatile platform for simulating the reaction collision mechanisms of intracellular networks. For example, the random order binding of initiation factors can be investigated with this algorithm to determine the validity of the underlying rapid equilibrium mechanism. Further optimization of the computational efficiency might allow for the incorporation of elongation and initiation. With this approach, the assumption of a constant number of actively translating ribosomes can be circumvented, though this will result in drastically increased computational effort.

5.2 Optimizing cell-free protein synthesis

The constant number of actively translating ribosomes is one of the assumptions of the diffusion model. To address this aspect, a model spanning all the translation

machinery was developed based on the published work by Arnold et al. (2005). The incorporation of sequence-oriented TC-TL allowed the model to provide unprecedented insight into the protein biosynthesis machinery. However, this level of detail requires significant computational effort. Based on their finding that transcription is not a rate-limiting step in CFPS, the sequence-oriented transcription and mRNA degradation parts of the model in this study were reduced to increase computational speed without deflecting the model outcome. Further modifications to adapt the model to *in vivo* translation allowed the model to be based on parameters available from the literature, which circumvents the need for parameter estimation. Consequently, the transition from *in vivo* to *in vitro* conditions was based on a dilution factor calculated by the ribosome ratio between the two systems. Additional implementation of recovery loss during reaction preparation increases the accuracy of the underlying co-factor concentrations.

For a comparison of model outcomes and experimentally observed translation rates, CFPS systems were used with GFP as a target protein (Failmezger et al. 2016). Fluorescence measurements were performed for the online quantification of GFP and translation rates were derived by deviating the received concentration profiles. Although fluorescence measurements are easy to setup and provide online concentration profiles for, they are only sensitive to correctly folded GFP and therefore underestimate the effective concentration of GFP (Iskakova et al. 2006; Caschera & Noireaux 2014). The comparison of *in silico* and *in vitro* translation rates indicates an over prediction of the translation rate which might arise due to the measurement. Further comparison with published results that were not relying on fluorescence measurement (Underwood et al. 2005) support the validity of this model.

GFP, as heterologous expressed protein, suffers from unbalanced codon usage as illustrated by the CAI. It achieved a CAI of 0.53 indicating inefficient codon usage, and the simulation results show that ribosomal crowding occurs along the first 30 %

of the sequence. Optimizing the sequence to a CAI of 0.72 without altering the amino acid sequence reveals an *in silico* predicted increase in the translation rate of more than 80 % and a more heterogeneous distribution of ribosomes along the mRNA. Although codon adaptation is a versatile tool for increasing translation rates in bioprocesses, the impact of codon usage drastically increases under CFPS conditions, making it a cheap tool for *a priori* optimization of *in vitro* reactions.

Codon optimization is not always possible or applicable as it depends on the gene of interest, while CFPS generally requires higher translation rates to compete with microbial protein production. The literature shows that *in-vitro* reactions are at an ideal composition (Freischmidt et al. 2010), though this study regarded final protein concentrations without investigating translation rates. Based on the reduced model, an initial screening of potential cofactors that increase translation rates was performed, and three groups of influencers were identified. The first two groups showed negative (i) or no (ii) influence on the translation rate, whereas the last group (iii) showed increased translation rates. Increasing the ribosome or the RNAP concentration resulted in lower translation rates, which can be explained by the increased number of actively translating ribosomes that bind tRNAs and result in limitations of certain tRNA species. Therefore, elongation rates were falling due to the limitation of elongation at specific codons. On the other hand, increasing EFTu and tRNA concentrations showed higher elongation rates compared to the reference, and a second sensitivity analysis was performed regarding only these two co-factors.

Considering that EFTu and tRNA build the ternary complex that is necessary for translation elongation, further analysis of their interplay was performed revealing that the simultaneous addition of both factors has a synergistic effect on the translation rate. By doubling their respective concentrations, a 117 % increase in the specific elongation rate and a 12 % decrease in the number of actively translating ribosomes was predicted, which would result in an overall increase in the translation

rate of 89 %. Using a new experimental setup with doubled EFTu and tRNA concentrations showed a 62 ± 9 % increase in the translation rate, which indicates that the hypothesis that ternary complexes are the key limiting factor of *in vitro* translation rates was confirmed. The deviations between predicted and measured translation rates might be due to uncertainties in parameters for the tRNA regeneration. Treating regeneration of each tRNA species and the corresponding mechanisms individually has great potential to increase the informative value of this model.

Simulations with elevated concentrations of ternary complexes predict that the number of actively translating ribosomes will decrease, and this effect might arise due to limitations during translation initiation. Using metabolic control analysis (Zouridis & Hatzimanikatis 2008), the impact of initiation factors, ribosomes, and ternary complexes was assessed. Under CFPS conditions, the translation rate is hampered by the number of ternary complexes indicated by an elasticity of approximately one. Increasing the ternary complex concentration results in decreasing elasticity for ternary complexes and increasing flux control of ribosomes and elasticity of initiation factors. Thus, at elevated ternary complex levels, a control switch from ternary complexes to initiation factors and ribosomes arises. Optimizing CFPS reaction systems is therefore an iterative approach that needs well-balanced increases in the essential co-factors. Having several co-factors that must be tested results in a laborious approach that can be significantly accelerated using the present model as a design platform for experiments.

Since CFPS reaction systems are commonly derived from living *E. coli* cells, estimates of translation rates can be derived from this state as well. Using the comprehensive dataset published by Bremer & Dennis (1987), a translation rate of 145 µM GFP per minute can be estimated. In addition, assuming *in-vivo*-like conditions for the model (neglecting dilution) results in a bistable system. At high

transcription rates, a low elongation rate combined with a high number of actively translating ribosomes was assessed. For lower transcription rates, a high elongation rate coupled with few active ribosomes was identified. Both states show the same translation rate of approximately 30 µM GFP per minute, which is less than 25 % of the estimated value. This deviation can be explained by the different conditions that were applied. During *in vivo* protein synthesis, dozens of genes are translated at once, whereas *in vivo* CFPS only translates a single gene at one time. Translating multiple genes simultaneously results in a far broader distribution of codons and therefore, elongation rates are less reduced as in the case of CFPS.

The PURE CFPS reaction approach (Shimizu et al. 2001) consists of separately purified translation components and allows for the individual composition of reaction preparations. The easy variation of the PURE system coupled with the predictive nature of the dynamic model promises a fast and efficient approach for the construction of *in vitro* reactions. However, the validity of this model must be proven for a broader range of variation in the underlying enzyme concentrations. Based on this approach, an *a priori* reaction optimization is possible that enables a matched reaction composition that is optimized for the given target sequence.

5.3 Gene expression in microbial populations

The last part of this study covers the aspect of TC-TL in large-scale cultivations, and primarily the cellular response towards oscillating substrate accessibility. The dynamic model provides a versatile platform for describing the translation of a single gene. However, the simultaneous TC-TL of many genes results in an ODE system that is difficult to solve. To address this application, a new modeling strategy is necessary. To describe transcriptional and translational responses in microbial populations, two models were established: a model to describe TC-TL for an exemplary operon and a model to describe cellular motion in an STR-PFR scale-

down setup. Both models combined can predict the population dynamics arising from the repetitive oscillating substrate concentrations.

In the first analysis, the transcriptional and translational response to substrate deficiency was assessed. After a short period of approximately 30 seconds, the cells started their reaction with the first RNA polymerase binding the operon structure. After 125 seconds, the cells left the substrate starvation zone and entered the limitation zone. At this point, the first two genes were transcribed, but the latter four genes were not yet under transcription. This indicates that the main part of TC-TL occurs after the starvation zone is left, and that short inductions trigger a response that takes approximately ten minutes until no further TC-TL occurs. The delayed costs of cellular reaction indicate that the STR reactor population is divided into subpopulations with or without additional maintenance.

Compared to the stiff peaks in transcript concentrations, proteins are quickly synthesized and slowly degraded due to the much lower degradation constant which leads to a steady accumulation of structural proteins. By assuming steady-state conditions, the impact of active protein degradation on the protein levels was calculated. Even degradation constants as high as the maximum values described in the literature lead to elevated protein levels in the STR population, which further supports the hypothesis that oscillating substrate conditions lead to population heterogeneity in the STR.

Based on this hypothesis, the population composition was further investigated based on a residence time distribution in contrast to the fixed assumptions that were used in the previous section. Initial simulations revealed that more than 65 % of the STR population was undergoing protein synthesis induced by substrate starvation. Over 25 % of this portion was undergoing multiple inductions, which implies that the cell reentered the PFR prior to completing the previous induction cascade. These results are supported by transcript measurements that show the same qualitative

pattern as the *in-silico* predicted values for the STR and PFR population. Further analysis of transcript progress over cultivation duration indicates almost no impact of process duration, which is supported by the short duration of a single perturbation reaction.

However, these results only include the transcript pattern. Transcriptional response takes approximately ten minutes until all transcripts are vanished, which leads to slightly higher transcript levels in the entire population. Protein levels on the other hand do not vanish, they slowly approach a new steady-state level, which is significantly higher than sole STR cultivations. This alternative steady state caused by the PFR device takes more than 15 hours to reach, compared to minutes for the transcriptional steady-state. In summary, repetitive short starvation zones lead to a bipartite reaction: a first quick transcriptional deflection that immediately restores standard conditions and an order of magnitude slower deflection of protein levels that leads to a new steady-state. Depending on the age of a single cell and the number of PFR cycles it has run through, the protein levels may vary significantly which results in further population heterogeneity.

Löffler et al. (2016) and Simen et al. (2017) mirrored the same response times. They indicated a fast response towards the gradient which describes the immediate cellular adaptation, and a slower response that was not yet allocated to cellular responses. However, the combination of *in silico* modeling and transcript data reveals that a significant deflection of the proteome composition might deflect the entire population towards an alternative steady-state. This study shows that transcriptomics can mimic the cellular reaction towards starvation, but investigation of the detailed adaptation occurring in the population must be addressed with proteomics as well.

5.4 The problem of model refinement

Three different modeling approaches were developed during the research performed in this study, each with a defined scope. Depending on the level of detail necessary to describe a system, the scope covered by a model is limited. Although it is possible to establish models that cover the entire TC-TL system as a diffusion-limited approach, they would be complicated and almost unsolvable. In contrast, the agent-based model includes fewer mechanisms that allow the scope of the model to cover the TC-TL of a whole operon and coupling to a cell distribution model. The tradeoff between the level of detail and the scope of the model is critical to the successful modeling of biochemical reactions.

A successful approach to modeling TC-TL systems uses a dynamic level of detail; mechanisms that are of less importance are treated with a lower level of detail, whereas critical reactions need a higher level of detail. By applying such a strategy, the computational effort can be reduced without sacrificing the accuracy of the model. For example, in the dynamic model, the computational effort was cut in half by reducing the entire transcriptional part to a simple black-box approach, though the model outcome is still consistent. The released computational capacity can be invested in further analysis of the model (e.g. sensitivity analysis or alternative reaction equations).

Parameter demand is another problem for large-scale models. Parameters are known for many reaction systems, but the reliance on them is questionable since most were estimated in *in vitro* systems. For example, translation elongation and ternary complex regeneration are poorly parameterized in the literature. For such a reaction system, a reduction to the lowest physical mechanisms can provide qualitative insight. In this case, the assumption of diffusion-driven reactions provided detailed insight into the reaction mechanisms without the need for individual parameters for each tRNA, which reduced the parameter demand drastically.

Although parameters for fundamental mechanisms such as diffusion are rarely described in literature, some parameters can be estimated by empirical correlations (e.g. the Stokes-Einstein correlation for diffusion coefficients).

A stable calculable model for the entire TC-TL process of a microorganism is ideal, but not realistic. Even if the complete parameter set was available, the numerical effort to solve such a model would be vast. This approach explains the diversity of models for TC-TL in the literature. Although each model was established for a given research topic and provided useful insight, the transfer of such models to other approaches is not always described or reasonable.

REFERENCES

Arnold S, Siemann M, Scharnweber K, Werner M, Baumann S, Reuss M. 2001. Kinetic modeling and simulation of *in vitro* transcription by phage T7 RNA polymerase. Biotechnol. Bioeng. 72:548–561.

Arnold S, Siemann-Herzberg M, Schmid J, Reuss M. 2005. Model-based inference of gene expression dynamics from sequence information. In: Nielsen J, editor. Biotechnology for the Future. Vol. 100. Berlin, Heidelberg: Springer Berlin Heidelberg; p. 89–179 (Advances in biochemical engineering/biotechnology).

Bailey JE & Ollis DF. 1986. Biochemical engineering fundamentals. 2. ed. New York [i.a.]: McGraw Hill. XXI, 984 S (McGraw Hill chemical engineering series). ISBN: 9780070032125.

Bremer H & Dennis PP. 1987. Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhardt FC, editor. *Escherichia coli* and *salmonella typhimurium*: Cellular and molecular biology. Vol. 2. Washington, D.C.: American Society for Microbiology; p. 1553–1569.

Brenan KE, Campbell SL, Petzold LR. 1995. Numerical solution of initial-value problems in differential-algebraic equations. Philadelphia, Pa.: Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104) (Classics in applied mathematics; vol. 14). ISBN: 978-0-89871-353-4.

Buckstein MH, He J, Rubin H. 2008. Characterization of nucleotide pools as a function of physiological state in *Escherichia coli*. J Bacteriol. 190:718–726.

Caschera F & Noireaux V. 2014. Synthesis of 2.3 mg/ml of protein with an all *Escherichia coli* cell-free transcription-translation system. Biochimie. 99:162–168.

Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, Reuss M. 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnol. Bioeng. 79:53–73.

Chen H, Shiroguchi K, Ge H, Xie XS. 2015. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. Mol Syst Biol. 11:808.

Chizzolini F, Forlin M, Yeh Martin N, Berloffa G, Cecchi D, Mansy SS. 2017. Cell-Free Translation Is More Variable than Transcription. ACS Synth. Biol. 6:638–647.

Dong H, Nilsson L, Kurland CG. 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. J. Mol. Biol. 260:649–663.

Failmezger J, Ludwig J, Nieß A, Siemann-Herzberg M. 2017. Quantifying ribosome dynamics in *Escherichia coli* using fluorescence. FEMS Microbiol. 364.

Failmezger J, Nitschel R, Sanchez-Kopper A, Kraml M, Siemann-Herzberg M. 2016. Site-specific cleavage of ribosomal RNA in *Escherichia coli*-based cell-free protein synthesis systems. PLoS ONE. 11:e0168764.

Fersht AR, Gangloff J, Dirheimer G. 1978. Reaction pathway and rate-determining step in the aminoacylation of tRNAArg catalyzed by the arginyl-tRNA synthetase from yeast. Biochemistry. 17:3740–3746.

Fluitt A, Pienaar E, Viljoen H. 2007. Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. Comp. Biol. Chem. 31:335–346.

Freischmidt A, Meysing M, Liss M, Wagner R, Kalbitzer HR, Horn G. 2010. Limiting factors of the translation machinery. J. Biotechnol. 150:44–50.

Gast F-U. 1987. Mechanistische Untersuchungen zur Fehlerkorrektur bei der ribosomalen Proteinsynthese [PhD Thesis]. Hannover: University of Hannover.

George S, Larsson G, Enfors S-O. 1993. A scale-down two-compartment reactor with controlled substrate oscillations: Metabolic response of *Saccharomyces cerevisiae*. Bioprocess Biosyst Eng. 9:249–257.

Gibson MA & Bruck J. 2000. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. J. Phys. Chem. A. 104:1876–1889.

Gillespie DT. 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81:2340–2361.

Gillespie DT. 2001. Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. 115:1716.

Goss PJE & Peccoud J. 1998. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. Proc. Natl. Acad. Sci. U.S.A. 95:6750–6755.

Hargrove JL & Schmidt FH. 1989. The role of mRNA and protein stability in gene expression. FASEB J. 3:2360–2370.

Haringa C, Noorman HJ, Mudde RF. 2016. Lagrangian modeling of hydrodynamic–kinetic interactions in (bio)chemical reactors: Practical implementation and setup guidelines. Chem. Eng. Sci.

Heinrich R & Rapoport TA. 1980. Mathematical modelling of translation of mRNA in eucaryotes; steady states, time-dependent processes and application to reticulocytest. J. Theor. Biol. 86:279–313.

Henson MA. 2003. Dynamic modeling of microbial cell populations. Curr. Opin. Biotechnol. 14:460–467.

Hofestädt R & Thelen S. 1998. Quantitative modeling of biochemical networks. In Silico Biol. 1:39–53.

Holms WH, Hamilton ID, Robertson AG. 1972. The rate of turnover of the adenosine triphosphate pool of *Escherichia coli* growing aerobically in simple defined media. Arch Mikrobiol. 83:95–109.

Howe JG & Hershey JWB. 1984. Translational initiation factor and ribosome association with the cytoskeletal framework fraction from HeLa cells. Cell. 37:85–93.

Iskakova MB, Szaflarski W, Dreyfus M, Remme J, Nierhaus KH. 2006. Troubleshooting coupled *in vitro* transcription-translation system derived from *Escherichia coli* cells: synthesis of high-yield fully active proteins. Nucleic Acids Res. 34:e135.

Junker BH. 2004. Scale-up methodologies for *Escherichia coli* and yeast fermentation processes. J. Biosci. Bioeng. 97:347–364.

Kaleta C, Schäuble S, Rinas U, Schuster S. 2013. Metabolic costs of amino acid and protein production in *Escherichia coli*. Biotechnol J. 8:1105–1114.

Kim DM, Kigawa T, Choi CY, Yokoyama S. 1996. A highly efficient cell-free protein synthesis system from *Escherichia coli*. Eur. J. Biochem. 239:881–886.

Klingbeil G, Erban R, Giles M, Maini PK. 2011. STOCHSIMGPU: parallel stochastic simulation for the Systems Biology Toolbox 2 for MATLAB. Bioinformatics. 27:1170–1171.

Komarov I, D'Souza RM, Tapia J-J. 2012. Accelerating the Gillespie τ-Leaping Method using graphics processing units. PLoS ONE. 7:e37370.

Lahtvee P-J, Seiman A, Arike L, Adamberg K, Vilu R. 2014. Protein turnover forms one of the highest maintenance costs in *Lactococcus lactis*. Microbiology. 160:1501–1512.

Lara AR, Galindo E, Ramírez OT, Palomares LA. 2006. Living With Heterogeneities in Bioreactors: Understanding the Effects of Environmental Gradients on Cells. Mol. Biotechnol. 34:355–382.

Lee A, Yau C, Giles MB, Doucet A, Holmes CC. 2010. On the Utility of Graphics Cards to Perform Massively Parallel Simulation of Advanced Monte Carlo Methods. J Comput Graph Stat. 19:769–789.

Li H, Cao Y, Petzold LR, Gillespie DT. 2008. Algorithms and software for stochastic simulation of biochemical reacting systems. Biotechnol Prog. 24:56–61.

Löffler M, Simen JD, Jager G, Schaferhoff K, Freund A, Takors R. 2016. Engineering *E. coli* for large-scale production - Strategies considering ATP expenses and transcriptional responses. Metab Eng. 38:73–85.

MacDonald CT & Gibbs JH. 1969. Concerning the kinetics of polypeptide synthesis on polyribosomes. Biopolymers. 7:707–725.

MacDonald CT, Gibbs JH, Pipkin AC. 1968. Kinetics of biopolymerization on nucleic acid templates. Biopolymers. 6:1–25.

Matsumoto M & Nishimura T. 1998. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans. Model. Comput. Simul. 8:3–30.

Mehra A & Hatzimanikatis V. 2006. An algorithmic framework for genome-wide modeling and analysis of translation networks. Biophys J. 90:1136–1146.

Nath K & Koch AL. 1970. Protein degradation in *Escherichia coli*. I. Measurement of rapidly and slowly decaying components. J Biol Chem. 245:2889–2900.

Neidhardt FC, Bloch PL, Pedersen S, Reeh S. 1977. Chemical measurement of steady-state levels of ten aminoacyl-transfer ribonucleic acid synthetases in *Escherichia coli*. J Bacteriol. 129:378–387.

Nieß A, Failmezger J, Kuschel M, Siemann-Herzberg M, Takors R. 2017. Experimentally Validated Model Enables Debottlenecking of *in vitro* Protein Synthesis and Identifies a Control Shift under *in vivo* Conditions. ACS Synth. Biol. Available from: http://pubs.acs.org/articlesonrequest/AOR-h7yNy4HKbkCkHhV4HVv3.

Nieß A, Löffler M, Simen JD, Takors R. 2017. Repetitive Short-Term Stimuli Imposed in Poor Mixing Zones Induce Long-Term Adaptation of *E. coli* Cultures in Large-Scale Bioreactors: Experimental Evidence and Mathematical Model. Front. Microbiol. 8:577.

Nirenberg MW & Matthaei JH. 1961. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc. Natl. Acad. Sci. U.S.A. 47:1588–1602.

Pedersen S, Bloch PL, Reeh S, Neidhardt FC. 1978. Patterns of protein synthesis in *E. coli*: a catalog of the amount of 140 individual proteins at different growth rates. Cell. 14:179–190.

Pingoud A, Gast F-U, Peters F. 1990. The influence of the concentrations of elongation factors and tRNAs on the dynamics and accuracy of protein biosynthesis. Biochim. Biophys. Acta. 1050:252–258.

Pipkin AC & Gibbs JH. 1966. Kinetics of synthesis and/or conformational changes of biological macromolecules. Biopolymers. 4:3–15.

Pratt. 1984. Coupled transcription-translation in prokaryotic cell-free systems. In: Hames BD, Higgins SJ, editors. Transcription and translation: A practical approach. New York: IRL Press; p. 179–210.

Puigbò P, Guzmán E, Romeu A, Garcia-Vallvé S. 2007. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. 35:W126-31.

Rauhut R & Klug G. 1999. mRNA degradation in bacteria. FEMS Microbiol. Rev. 23:353–370.

Rudorf S & Lipowsky R. 2015. Protein synthesis in *E. coli*: Dependence of codon-specific elongation on tRNA concentration and codon usage. PLoS ONE. 10:e0134994.

Ruusala T, Ehrenberg M, Kurland CG. 1982. Catalytic effects of elongation factor Ts on polypeptide synthesis. EMBO. J. 1:75–78.

Saito M & Matsumoto M. 2008. SIMD-Oriented Fast Mersenne Twister: A 128-bit Pseudorandom Number Generator. In: Keller A, Heinrich S, Niederreiter H, editors. Monte Carlo and Quasi-Monte Carlo Methods 2006. Berlin, Heidelberg: Springer Berlin Heidelberg; p. 607–622.

Shampine LF, Reichelt MW, Kierzenka JA. 1999. Solving Index-1 DAEs in MATLAB and Simulink. SIAM Rev. 41:538–552.

Sharp PM & Li W-H. 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15:1281–1295.

Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T. 2001. Cell-free translation reconstituted with purified components. Nat Biotechnol. 19:751–755.

Simen JD, Löffler M, Jäger G, Schäferhoff K, Freund A, Matthes J, Müller J, Takors R. 2017. Transcriptional response of *Escherichia coli* to ammonia and glucose fluctuations. Microb Biotechnol.

Spahr PF. 1962. Amino acid composition of ribosomes from *Escherichia Coli*. J. Mol. Biol. 4:395–406.

Stögbauer TR, Windhager L, Zimmer R, Radler JO. 2012. Experiment and mathematical modeling of gene expression dynamics in a cell-free system. Integr Biol (Camb). 4:494–501.

Stoltzfus A, Leslie JF, Milkman R. 1988. Molecular evolution of the *Escherichia coli* chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between trp and tonB. Genetics. 120:345–358.

Takors R. 2012. Scale-up of microbial processes: impacts, tools and open questions. J. Biotechnol. 160:3–9.

Tomsic J, Vitali LA, Daviter T, Savelsbergh A, Spurio R, Striebeck P, Wintermeyer W, Rodnina MV, Gualerzi CO. 2000. Late events of translation initiation in bacteria: a kinetic analysis. EMBO. J. 19:2127–2136.

Underwood KA, Swartz JR, Puglisi JD. 2005. Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. Biotechnol. Bioeng. 91:425–435.

Valgepea K, Adamberg K, Seiman A, Vilu R. 2013. *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. Mol Biosyst. 9:2344–2358.

Varenne S, Buc J, Lloubes R, Lazdunski C. 1984. Translation is a non-uniform process. J. Mol. Biol. 180:549–576.

Wintermeyer W & Gualerzi C. 1983. Effect of *Escherichia coli* initiation factors on the kinetics of N-AcPhe-tRNAPhe binding to 30S ribosomal subunits. A fluorescence stopped-flow study. Biochemistry. 22:690–694.

Yanofsky C. 2004. The different roles of tryptophan transfer RNA in regulating trp operon expression in *E. coli* versus *B. subtilis*. Trends Genet. 20:367–374.

Yanofsky C. 2007. RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. RNA. 13:1141–1154.

Zhang G, Fedyunin I, Miekley O, Valleriani A, Moura A, Ignatova Z. 2010. Global and local depletion of ternary complex limits translational elongation. Nucleic Acids Res. 38:4778–4787.

Zouridis H & Hatzimanikatis V. 2007. A Model for Protein Translation: Polysome Self-Organization Leads to Maximum Protein Synthesis Rates. Biophys. J. 92:717–730.

Zouridis H & Hatzimanikatis V. 2008. Effects of codon distributions and tRNA competition on protein translation. Biophys. J. 95:1018–1033.

Zucker FH & Hershey JWB. 1986. Binding of *Escherichia coli* protein synthesis initiation factor IF1 to 30S ribosomal subunits measured by fluorescence polarization. Biochemistry. 25:3682–3690.

AUTHOR CONTRIBUTIONS

This chapter summarizes my (Alexander Nieß) contribution to the manuscripts that were already published. The contents of the manuscripts are supplementary provided (appendix A and B).

Manuscript A

Nieß, A#., Failmezger, J.#, Kuschel, M., Siemann-Herzberg, M., & Takors, R. (2017). Experimentally validated model enables debottlenecking of *in vitro* protein synthesis and identifies a control shift under *in vivo* conditions. *ACS Synthetic Biology*.

#Both authors contributed equally to this work

Alexander Nieß (AN) designed the study, established the dynamic model and participated in the data analysis and interpretation as well as the preparation of the manuscript.

Manuscript B

Nieß, A., Löffler, M., Simen, J. D., & Takors, R. (2017). Repetitive short-term stimuli imposed in poor mixing zones induce long-term adaptation of *E. coli* cultures in large-scale bioreactors: experimental evidence and mathematical model. *Frontiers in Microbiology*, *8*.

AN designed the study and established the ensemble model. AN furthermore participated in simulation evaluation and interpretation as well as manuscript preparation.

DECLARATION OF ORIGINALITY

I declare that the submitted work has been completed by me and that I have not used any other than permitted reference sources or materials. All references and other sources used by me have been appropriately acknowledged in the work.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig angefertigt habe. Es wurden von mir nur die in der Arbeit ausdrücklich benannten Quellen und Hilfsmittel benutzt. Übernommenes Gedankengut wurde von mir als solches kenntlich gemacht.

Stuttgart, 05.03.2018

Alexander Nieß

CURRICULUM VITAE

Personal details:

Date and place of birth: 21.04.1988, Sigmaringen, Germany

Education:

02/2014-today	**PhD student** at the Institute of Biochemical Engineering, University of Stuttgart
10/2011-12/2013	**Studies in Process Engineering** University of Stuttgart, Germany Field of specialization: Biochemical and Mechanical Engineering
04/2013-12/2013	**Master thesis** at the Institute of Biochemical Engineering, University of Stuttgart 'biomass specific dynamic metabolic flux analysis'
10/2007-02/2011	**Studies in pharmaceutical Biotechnology** University of applied Sciences Biberach, Germany
03/2010-12/2010	**Internship and Bachelor thesis** Rentschler Biotechnologie, Laupheim 'Use of membrane adsorbers in downstream processing proteins from continuous bioprocesses'

Work experience:

02/2012-02/2014	**Teching Assistant** for the courses 'Numerical Methods I & II', University of Stuttgart
08/2012-03/2013	**Research Assistant** at the Institute of Biochemical Engineering University of Stuttgart
04/2011-09/2011	**Internship** Sartorius Stedim Biotech
03/2010-12/2010	**Internship and Bachelor thesis** Rentschler Biotechnologie, Laupheim 'Use of membrane adsorbers in downstream processing proteins from continuous bioprocesses'
08/2009-09/2009	**Internship** Boehringer Ingelheim

Additional Qualifications:

Languages	English (fluent, scientific writing skills)
	French (basic)
	German (native language)
Computing	MS-Office (professional knowledge)
	MatLab (professional knowledge)
	C Programming (advanced knowledge)
Personal Interests	Woodworking, electrical engineering

CHAPTER 7

APPENDICES

MANUSCRIPT A

Experimentally Validated Model Enables Debottlenecking of *in Vitro* Protein Synthesis and Identifies a Control Shift under *in Vivo* Conditions

ACS SyntheticBiology
Research Article
Cite This: ACS Synth. Biol. 2017, 6, 1913-1921
pubs.acs.org/synthbio

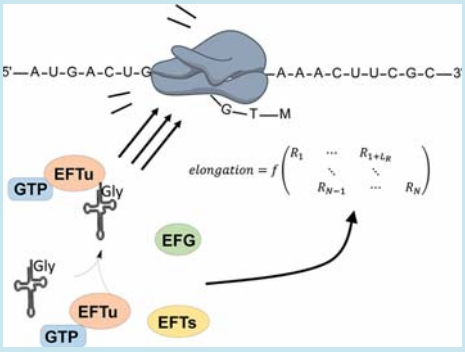
Experimentally Validated Model Enables Debottlenecking of *in Vitro* Protein Synthesis and Identifies a Control Shift under *in Vivo* Conditions

Alexander Nieß,[†,#] Jurek Failmezger,[†,#] Maike Kuschel,[†] Martin Siemann-Herzberg,[†] and Ralf Takors*[,†]

[†]Institute of Biochemical Engineering, University of Stuttgart, Stuttgart, D-70569, Germany

Supporting Information

ABSTRACT: Cell-free (*in vitro*) protein synthesis (CFPS) systems provide a versatile tool that can be used to investigate different aspects of the transcription-translation machinery by reducing cells to the basic functions of protein formation. Recent improvements in reaction stability and lysate preparation offer the potential to expand the scope of *in vitro* biosynthesis from a research tool to a multifunctional and versatile platform for protein production and synthetic biology. To date, even the best-performing CFPS systems are drastically slower than *in vivo* references. Major limitations are imposed by ribosomal activities that progress in an order of magnitude slower on the mRNA template. Owing to the complex nature of the ribosomal machinery, conventional "trial and error" experiments only provide little insight into how the desired performance could be improved. By applying a DNA-sequence-oriented mechanistic model, we analyzed the major differences between cell-free *in vitro* and *in vivo* protein synthesis. We successfully identified major limiting elements of *in vitro* translation, namely the supply of ternary complexes consisting of EFTu and tRNA. Additionally, we showed that diluted *in vitro* systems suffer from reduced ribosome numbers. On the basis of our model, we propose a new experimental design predicting 90% increased translation rates, which were well achieved in experiments. Furthermore, we identified a shifting control in the translation rate, which is characterized by availability of the ternary complex under *in vitro* conditions and the initiation of translation in a living cell. Accordingly, the model can successfully be applied to sensitivity analyses and experimental design.



KEYWORDS: *ribosomes, translation factors, translation rate, cell-free synthetic biology*

Cell-free protein synthesis (CFPS) is a powerful technique for the fast production of recombinant proteins. In contrast to traditional *in vivo* protein formation, CFPS offers advantages such as expression of genes encoding growth-inhibiting proteins, high-throughput protein production and easy access for investigating translation conditions.[1−3] Recent studies have demonstrated the potential of CFPS as a platform technique for genetic circuit engineering.[4]

Whereas early CFPS batch reactions suffer from short activity periods of less than 1 h, the implementation of sophisticated energy regeneration systems has allowed active protein formation up to 10 h,[5] reaching final product titers of several mg/mL. Nevertheless, CFPS systems achieve maximum synthesis rates of only 0.4−1 mg/mL/h,[6−11] which lags behind *in vivo* protein production rates by a factor of several hundred.[12] To compete with *in vivo* protein production, the duration of the *in vitro* reaction and the synthesis rates must both be improved, and key factors that limit the *in vitro* space-time yield must be identified.

According to current opinion, *in vivo* protein synthesis is thought to be limited by translation initiation rates,[13,14] but it is not clear whether this is also true for *in vitro* protein synthesis. Research groups have attempted to identify key steps that control *in vitro* conditions.[15] For example, Freischmidt et al.[16] investigated the impact of adding translation factors, aminoacyl-tRNA synthetases and active ribosomes, on the total amount of protein produced. Surprisingly, this did not improve the final protein concentration, and the authors concluded that stoichiometry of the translation factors was optimal in their system. However, putative impacts on translation rate (i.e., molar volumetric protein synthesis rate) were not addressed.

A detailed analysis of CFPS was performed by Underwood et al.[12] Combining polysome profile analysis with protein quantification allowed researchers to quantitatively analyze the cell-free system. Among others, Underwood et al.[12] clearly identified that a highly reduced ribosome concentration, in combination with a low specific elongation rate of around 1.5 amino acids per ribosome per second, heavily influenced the overall *in vitro* translation rate. In fact, the *in vitro* translation rate was estimated to be more than 3 orders of magnitude lower than the *in vivo* translation rate under fast growth conditions. The discrepancy between *in vitro* and *in vivo*

Received: April 7, 2017
Published: June 19, 2017

DOI: 10.1021/acssynbio.7b00117
ACS Synth. Biol. 2017, 6, 1913−1921

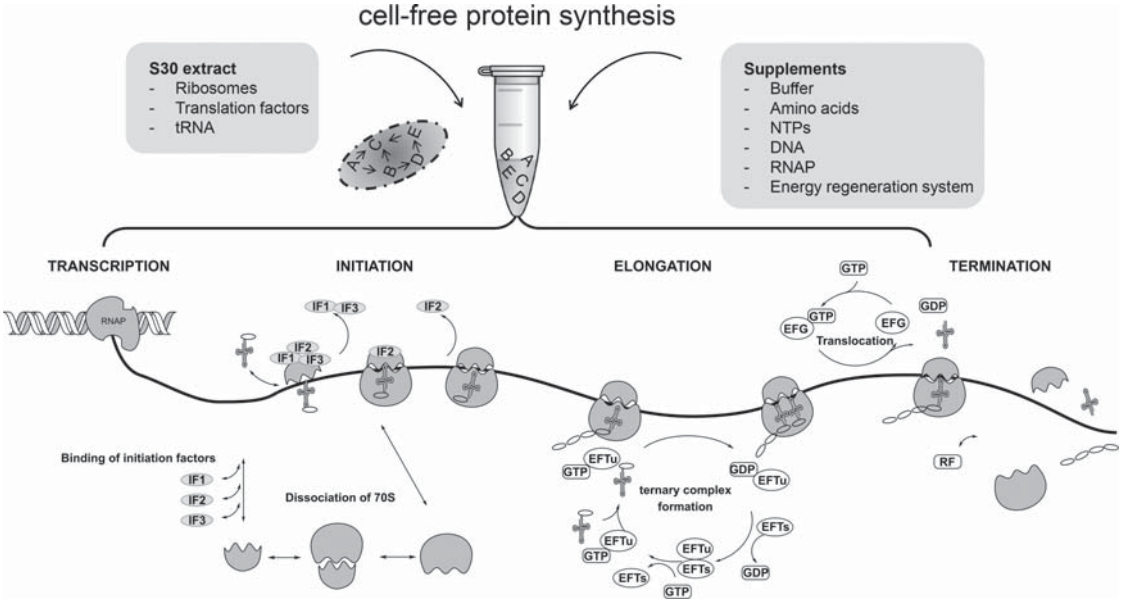


Figure 1. Overview of cell-free protein synthesis and reaction composition of S30 extract and added supplements. An additional schematic overview of the model mechanisms is shown, starting with T7 RNA polymerase and subsequent sequence oriented translation as well as regeneration of cofactors. Abbreviations: RNAP = RNA polymerase; IF = initiation factor; EFTu = elongation factor Tu; EFTs = elongation factor Ts; EFG = elongation factor G; RF = releasing factor; NTP = nucleotide triphosphate; GTP = guanosine triphosphate; GDP = guanosine diphosphate.

elongation rates could partly be overcome by supplementing the cell-free system with additional elongation factors, which increased the elongation rate by 33%. It appeared that the limitation of protein growth shifted from initiation to elongation under *in vivo* and *in vitro* conditions, indicating that translation is limited by the elongation rate of ribosomes. There are two major mechanisms involved in elongation. The first is peptide chain elongation, which is controlled by ternary complexes consisting of charged tRNA, GTP, and elongation factor Tu (EFTu), and the second is translocation, which is catalyzed by elongation factor G (EFG).

As depicted in Figure 1, the translation machinery comprises a highly complex, synergistic system that consists of a large number of components. As a result, its optimization requires sophisticated tools to properly investigate the parameter space. Arnold et al.[17] established a mathematical model that combines all steps of transcription and translation, such as mRNA synthesis and degradation, ribosomal translation, energy regeneration, and inactivation kinetics of certain proteins, to simulate a CFPS system. A DNA sequence oriented approach enabled estimation of crucial parameters such as reaction rates, which are not readily accessible experimentally. Our approach builds on a previous work by performing a thorough modeling and experimental study aimed at identifying key control elements of *in vitro* protein synthesis and defining ways in which the system can be improved.

2. RESULTS AND DISCUSSION

2.1. Model Validation. Routinely performed batch-CFPS reactions in 96-well microtiter plates always showed a three phase pattern, as depicted in Figure 2. First, the transcription−translation machinery was initiated and began producing mRNAs and proteins, and characteristic increases in translation rate were observed. Then, a pseudo steady-state was achieved, as defined by constant reaction rates. The length of phase two is dependent on the supply of substrates. Finally, phase three

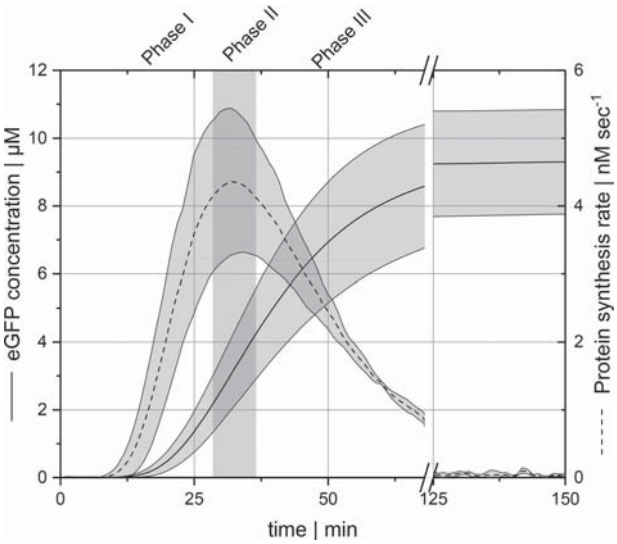


Figure 2. Time courses of protein concentration (biological triplicates) and corresponding error bands. Protein synthesis rate, derived from eGFP concentration via central differences, is also shown.

comprised a decrease in translation rate due to limiting components and decomposition of the original enzymes.

All three phases were considered for the translation analysis. An increase in GFP concentration in the observation intervals was calculated with central differences at each time point (shown in eq 1). Maximum translation rates were determined for phase two and were compared to the simulation results.

$$v_{TL}(t_i) = \frac{C_{eGFP}(t_{i+1}) - C_{eGFP}(t_{i-1})}{t_{i+1} - t_{i-1}} \quad (1)$$

With a complete set of parameters and the experimentally determined *in vitro* concentrations, it was possible to predict

DOI: 10.1021/acssynbio.7b00117
ACS Synth. Biol. 2017, 6, 1913−1921

Table 1. Comparison of Experimentally Determined Parameters of *in Vivo* and *in Vitro* Translation, and Simulated *in Vitro* Reaction Parameters

	in vivo	*in vitro* Underwood et al.	*in silico* prediction	this study
ribosome concentration (μM)	42[a]	1.6 ± 0.1	1.7	1.7 ± 0.1
amount of actively translating ribosomes (%)	80[a]	72 ± 4	82	n.d.
elongation rate (AA/sec)	18[a]	1.5 ± 0.2	1.12	n.d.
initiation rate (nM/sec)	2.8×10^3[b]	8.2 ± 0.3	7.1	4.3 ± 1.07
bulk protein synthesis rate (mg/L/min)	3.3×10^3[b]	12.3 ± 0.5	10.6	6.5 ± 1.6

[a]*In vivo* data taken from Bremer and Dennis.[22] [b]*In vivo* data taken from Underwood et al.[12]

translation rates for the given cell-free reaction system (Table 1). Our *in silico* predictions were consistent with the experimental observations of Underwood et al.,[12] although the latter studied the translation rate of chloramphenicol acetyltransferase (CAT), which has a molecular weight close to that of eGFP. On the other hand, only 61% of the predicted 90% increase in translation rate were measured experimentally, even though the simulations and experiments were focused on the same protein sequence.

Despite the fact that eGFP has the advantage of being able to utilize online monitoring via fluorescence measurement, several studies identified a drawback in that only a fraction of the synthesized eGFP is correctly folded and fluorescently active. Examples are given by Caschera et al.[18] and Iskakova,[19] who measured only 80% and 50% fluorescently active GFP, respectively. Further sophisticated efforts such as the analysis of ^{14}C-lysine incorporation could determine the degree of active eGFP. However, likewise studies were not performed as key parameters affecting eGFP folding (such as the temperature[19]) remained constant. Consequently, a stable ratio of active to nonactive eGFP was assumed for all expression rates, which most likely underestimated the measured translation rates.

Systems modeling is always a trade-off between the detailed reflection of reality and the availability of proper data sets for configuring the theoretical setup. Because likewise parameters were not available, the regeneration of tRNA is treated equally for all tRNA species disregarding putative interactions with aminoacyl tRNA synthases. Moreover, the presented model describes the transcription and translation of only one single mRNA species per time. Another characteristic feature is its hybrid structure that is given by a black box model for transcription and a detailed structured model for translation. As such, structural limitations of translation are highlighted which enables the identification of key rate controlling elements of the total processing from transcription to translation. Accordingly, the modeling approach differed from published black box models by Stögbauer et al.[20] and Chizzolini et al.[21] Notably, the model offers the opportunity for simulating *in vitro* and *in vivo* conditions, for instance by including mechanisms like mRNA secondary structure. Summarizing, we conclude that the presented model fairly represents the mechanisms of *in vitro* translation.

2.2. Model Based Identification of Limiting Factors.

To identify possible rate-limiting factors in *in vitro* translation systems, we altered several components of the translation machinery *in silico* and simulated the respective impact on translation dynamics (Figure 3). Specifically, different concentrations were tested for (i) all initiation factors (simultaneously) and selectively for (ii) EFTu, (iii) EFG, (iv) EFTs, (v) tRNA, (vi) ribosome releasing factor (RF), (vii) T7 RNA polymerase, and (viii) ribosomes. All other input parameters

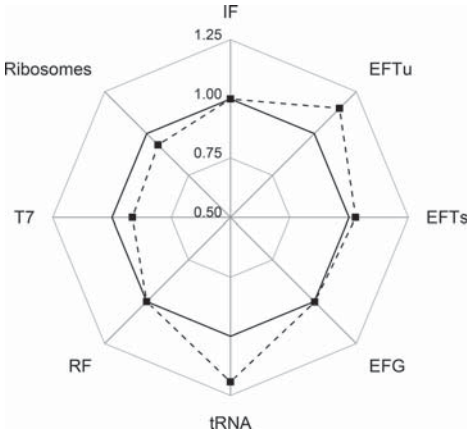


Figure 3. Sensitivity analysis of translation factors regarding translation rate. Values are derived from an *in silico* 50% increase in concentration of each translation factor. Resulting translation rate is normalized to the nonaltered rate to allow easy comparison. Factors considered are initiation factors 1−3 (IF); elongation factors G (EFG), Ts (EFTs), and Tu (EFTu); ribosome releasing factor (RF); T7 RNA polymerase; and tRNA. A value of 1 indicates no influence of this factor on the translation rate.

and reaction conditions remained constant. For comparison, the calculated translation rates were normalized to the reference condition with unchanged parameter settings.

Surprisingly, predicted translation rates persisted when the pool size of initiation factors, EFG, EFTs, and RF increased. These findings suggest that none of the factors limited translation rates under the conditions tested. In fact, increasing ribosome numbers can even decrease the translation rate, as shown in Figure 3. To be precise, raising the total number of ribosomes by 50% caused 53% increase of actively translating ribosomes which underlines that each "added" ribosome is directly incorporated in the translation machinery. This finding is consistent with previous experimental observations by Freischmidt et al.,[16] who detected a decline in translation rate with increasing ribosome density. The more the ribosomes are active, the more tRNA is bound, which decreases the amount of tRNA that can form novel ternary complexes. Ternary complex concentrations of tRNA species (according to the nomenclature of Dong et al.[24]) Ala2, Asp1, Phe, His, Lys, Asn, Pro2, Gln2, Ser3, Ser5 and Thr13 are decreased by 10−40% and their corresponding codons are required by more than 50% of the GFP sequence. Our modeling results show that the majority of ribosomes slow down translation when repetitive codons occur (e.g., His tag). Under such conditions, each ribosome is bound to a specific tRNA that is also required in the next elongation step. Accordingly, the specific tRNA becomes limiting, which decreases the elongation rate. This slowdown will also lead to ribosomal stalling.

DOI: 10.1021/acssynbio.7b00117
ACS Synth. Biol. 2017, 6, 1913−1921

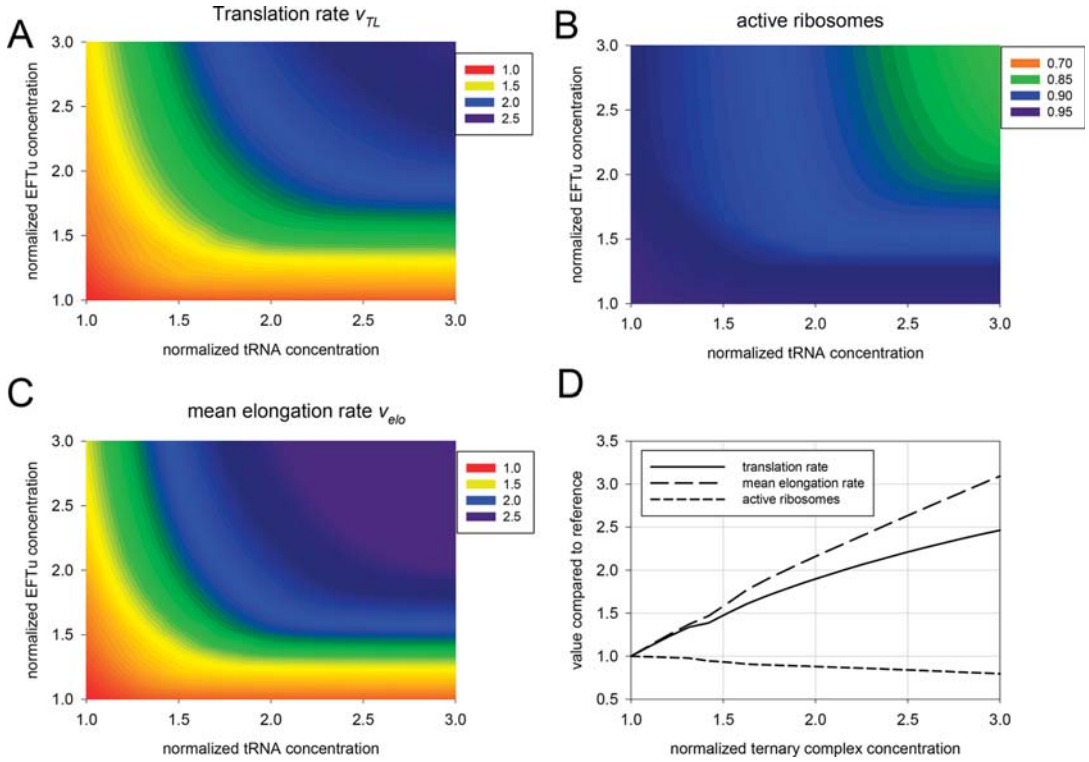


Figure 4. Influence of EFTu and tRNA concentration (normalized to reference values) and resulting normalized translation rate (A), normalized amount of active ribosomes (B) (calculated as sum of actively translating ribosomes divided by overall number of 30S/50S subunits), and normalized elongation rate (C) (average value over all codons). All values were normalized to values at unaltered conditions. Part D shows the influence of coupled EFTu and tRNA multipliers and the resulting translation rate (solid line), mean elongation rate (dotted line), and amount of active ribosomes (dashed line).

Moreover, our simulation revealed a slight decrease in translation performance due to increased RNA polymerase levels. The explanation for this follows a similar rationale to that previously described for ribosomes. An increase in mRNA concentration leads to an increase in the number of actively translating ribosomes and a lower ribosome density on the mRNA, which causes a shortage of free tRNA. In this case, the amount of freely accessible ternary complexes for the species Ala2, Ser5, and His decrease by more than 20% compared to reference levels.

Interestingly, translation rates only increased with rising levels of tRNA or EFTu. The critical role of EFTu is emphasized by the fact that EFTu is the most abundant protein in *E. coli*.[23] *In vivo* EFTu concentrations correlate with the concentration of total aminoacyl-tRNA.[24] GTP binding confers EFTu affinity to charged tRNA, the substrate for translation elongation. Accordingly, the stimulating effects of EFTu and tRNA on translation basically mirror the lack of substrate for protein formation.

As a consequence of this, there are indications that the ternary complex consisting of EFTu, tRNA, and GTP is rate-limiting in cell-free translation. To investigate the synergistic impacts of tRNA and EFTu, we analyzed the combinatorial effect of tRNA and EFTu on translation rate. As depicted in Figure 4A, the combination of tRNA and EFTu considerably increased the simulated translation rate, compared to increasing either EFTu or tRNA alone. This result supports the hypothesis that availability of the ternary complex exerts a key level of control on translation speed. Figure 4D shows that the rise in translation rate slows down as the availability of ternary complexes increases. As the translation rate is a function of the amount of actively translating ribosomes and the elongation rate, it is likely that at least one of these will limit the process. Further studies into mean elongation rate (Figure 4C) and amount of actively translating ribosomes (Figure 4B) revealed a contradictory trend for the number of actively translating ribosomes. Whereas translation rate and specific elongation rate steadily increased with tRNA and EFTu concentration, the amount of translating ribosomes decreased, and as a consequence the translation rate leveled off. These findings suggest that as specific elongation rates increase, a threshold is reached where initiation becomes limited. In other words, as the supply of tRNA and EFTu improves, the more accurately *in vitro* translation reflects the *in vivo* conditions.

2.3. Increasing Ternary Complex Concentrations in CFPS Reactions. To verify that ternary complex concentration does in fact limit *in vitro* translation rates, we doubled the EFTu and tRNA concentration by supplementing with purified components. Because a ribosome concentration of 1.45 μM was detected in the samples, 8 μM EFTu and 13 μM tRNA were added to double the respective concentrations. An equal volume of protein storage buffer was added to the reference samples. Maximum protein synthesis rates were measured under both conditions. For the given experimental setup, the model predicted an 89% increase in the translation rate. The underlying translation system shows a 117% increase in elongation rate, but the amount of actively translating ribosomes decreased by 12%. Experimental measurements

DOI: 10.1021/acssynbio.7b00117
ACS Synth. Biol. 2017, 6, 1913−1921

revealed a 62 ± 9% rise in the maximum translation rate compared to the reference (see Appendix Figure 1).

Considering the inherent drawbacks of experimental detection, we conclude that the simulation results are a good prediction of the experimental readout. Thus, our hypothesis that the ternary complex is a key rate-limiting target was confirmed.

2.4. Translational Control. Figure 4D shows that an increase in ternary complex concentration causes a proportional rise in elongation rate, but reduces the rise in translation rate. Furthermore, falling fractions of actively translating ribosomes were observed, which indicates the increasing impact of limiting initiation. Furthermore, the influence of involved reactants was investigated using metabolic control analysis. Because details of tRNA recharging via amino acylation could not be properly parametrized in the model, tRNA regeneration was finally regarded as not rate limiting.

We evaluated (i) ϵ_{T3}, (ii) ϵ_{IF}, and (iii) FCC, which represent sensitivities with respect to varying ternary complex (i), initiation factors (ii), and ribosome concentrations (iii) on the translation rate at a given C_{T3}, respectively. Variations in these impact factors were calculated by estimating the partial differentials with central differences (Figure 5). Under standard

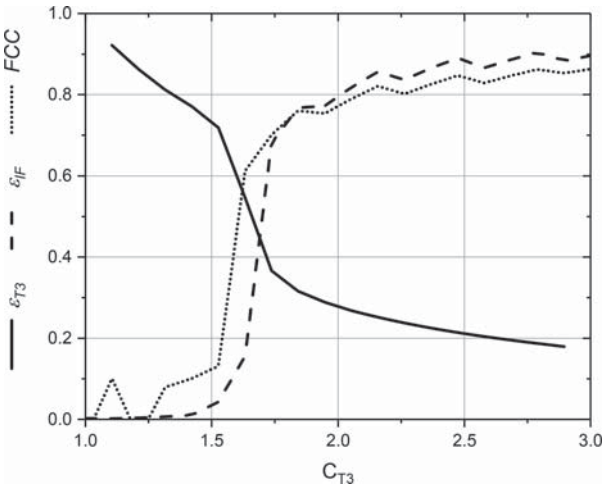


Figure 5. Calculated elasticities ϵ_{T3} and ϵ_{IF} and flux control coefficient (FCC) as a function of increasing normalized ternary complex concentrations. Elasticities ϵ_{T3} and ϵ_{IF} and *FCC* represent the sensitivity of translation rate with respect to varying total ternary complex, initiation factor, and total ribosome concentration.

in vitro conditions (normalized concentration C_{T3} = 1), the influence of ribosomes and initiation factors turned out to be negligible (FCC = ϵ_{IF} = 0). Neither changes in ribosomes nor variations in the initiation factors could improve the given translation rate. However, the translation rate could be increased significantly by increasing C_{T3}, which is indicated by an elasticity ϵ_{T3} of 1.

The stimulating impact of C_{T3} on translation rate diminished with increasing C_{T3} levels. As indicated in Figure 5, control of the translation process is taken over by the amount of actively translating ribosomes, and this number can be increased by two means: raising either the concentration of ribosomes (FCC) or the concentration of all initiation factors (ϵ_{IF}). When neither ribosomes nor initiation factors are increased, the number of actively translating ribosomes becomes reduced as the ternary complex concentration increases, as depicted in Figure 4D.

Figure 5 outlines the differences between *in vitro* and *in vivo* protein production. The first is given by the standard conditions, which may benefit from the addition of ternary complexes. The second, *in vivo* conditions, is not affected by a limiting ternary complex supply, which identifies availability of initiation factors and ribosome number as key control elements of translation rate.

Compared to the previous model[17] our hybrid approach reduced model complexity and in turn accelerated computational speed. Accordingly, the investigation of elasticities as a key issue of metabolic control analysis was enabled.

2.5. Estimation of CFPS Potential. Owing to the fact that the translation machinery is derived from viable *E. coli* cells, *in vivo* measurements of translation rates show strong potential for CFPS applications. Using the data set published by Bremer and Dennis,[22] a translation rate for eGFP of almost 150 μM per minute was estimated as theoretical upper limit for undiluted CFPS extracts (eq 2). However, the *in vivo* translation rate comprises the parallel processing of multiple different mRNA species, which implies the concomitant use of a variety of codon sequences, whereas CFPS processes only one mRNA sequence. This effect leads to one-sided use of tRNA species in CFPS, causing limited availability of distinct ternary complexes. Accordingly, the observed elongation rates in CFPS are inherently limited compared to *in vivo* conditions.

$$\nu_{TL} = C_R^{active} \bar{\nu}_{TLE} N_{AA,GFP}^{-1} \quad (2)$$

where

$$C_R^{active} = 33.85\ \mu M\ ^{25}$$

$$\bar{\nu}_{TLE} = 18\ AA\ Ribosome^{-1}\ s^{-1}\ ^{22}$$

$$N_{AA,GFP} = 245\ AA\ eGFP^{-1}$$

Neglecting CFPS dilution (D = 1), that is, assuming *in vivo*-like conditions, and assuming the translation rate of a single mRNA only, we found that the measured translation rates of Bremer and Dennis[22] could not be reached. Instead, we found either a low number of active ribosomes with a high elongation rate (Table 2, low TC), or a high number of active ribosomes with a low elongation rate (Table 2, high TC). Apparently, the "implementation" of *in vivo* concentrations had increased the active ribosome levels so extensively that the availability of ternary complexes becomes limiting and therefore the elongation rate decreases (see section 2.2, increased T7

Table 2. Translation Rates under *in Vivo* Conditions[a]

	optimal conditions	simulated *in vivo* low TC	simulated *in vivo* high TC	simulated *in vitro*
ν_{elo} AA Ribosome^{-1} s^{-1}	18	14.9	3.8	1.12
active ribosomes %	80	21	83	82
ν_{TL} μM eGFP min^{-1}	145	30.3	30.8	0.38
dilution factor	1.0	1.0	1.0	21.08

[a]Optimal conditions show values from the literature.[22] Simulations of *in vivo* conditions ignore dilution and investigate the impacts of ternary complexes (TC) as indicted.

DOI: 10.1021/acssynbio.7b00117
ACS Synth. Biol. 2017, 6, 1913−1921

concentrations). This emphasizes the close interaction between the supply of active ribosomes and ternary complexes, which creates shifting control of translation under different conditions. The finding is also supported by previous studies[17] that estimated the ratio of actively translating to total ribosomes to be below 5%, which imposed an artificial control on initiation and not on tRNA supply as it is the case in this study.

3. METHODS

3.1. Purification of EFTU. *E. coli* BL21 DE3 carrying pET24c(+)-EFTu, which was a gift from Linda Spremulli (Addgene plasmid # 31172), was grown at 37 °C in TB media (12 g/L tryptone, 24 g/L yeast extract, 0.4% glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4) containing 50 μg/mL Kanamycin, until it reached OD_{600} 0.6. Protein expression was induced by the addition of IPTG (1 mM final concentration) and performed for 12 h at 30 °C. Biomass was harvested by 20 min of centrifugation at 5000*g* at 4 °C and stored at −20 °C until purification. Cells were resuspended in lysis buffer (300 mM NaCl, 10 mM potassium phosphate buffer (pH 8.0), 1 mM DTT), and cell lysis was performed by sonication. Purification was performed using a Ni NTA gel (HIS-Select HF Nickel Affinity Gel, Sigma) according to the manufacturer's instructions. EFTu (purity > 95% according to SDS-PAGE) was buffer exchanged (10 mM Tris-HCL pH 8.0, 1 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT, 50 μM GDP, 5% (v/v) glycerol) using PD-10 columns (GE Healthcare) and then aliquoted and stored at −70 °C. The protein was quantified using a Pierce BCA Protein Assay (Thermo Scientific).

3.2. Lysate Preparation. Preparation of cell-free extract was performed as previously described.[26,27] Briefly, *E. coli* A19 was cultivated on 2 × YTPG medium in a 30 L bioreactor. Cells were harvested during the early exponential phase, and the biomass was rapidly chilled to 4 °C by pumping through a plate heat exchanger. S30 buffer (14 mM magnesium acetate, 60 mM potassium acetate, 10 mM Tris, pH 8.0, 2 mM DTT; 1 mL per gram of cell paste) was added and the biomass was lysed using a high-pressure homogenizer (EmulsiFlex-C5, Avestin, Canada) at 12 000 kPa. The extract was centrifuged twice at 30 000*g* for 30 min at 4 °C, and the cell extract was incubated at 37 °C for 80 min. After this runoff, the lysate was dialyzed using 100-times the volume of S30 buffer at 4 °C for 4 h. To clear the extract, a final centrifugation step (4000*g* for 20 min at 4 °C) was performed. The cell-free extract was flash-frozen in liquid nitrogen and stored at −70 °C.

3.3. Ribosome Quantification. Ribosomes were quantified based on 16S and 23S rRNA analysis using capillary gel electrophoresis with laser-induced fluorescence detection as previously described.[28,29]

3.4. Reaction Setup. Cell-free reactions, performed in a 250 μL scale as previously described (Failmezger et al., 2016), consisted of[27] 80 mM HEPES-KOH (pH 8.0), 1.2 mM ATP, 1 mM GTP, CTP, and UTP, 2 mM DTT, 90 mM potassium glutamate, 20 mM ammonium glutamate, 18 mM magnesium glutamate, 34 μg/mL folinic acid, 1 mM 20 amino acids, 2% PEG (8000), 80 mM creatine phosphate, 240 μg/mL creatine kinase, 3 U/μL T7 RNA Polymerase (Roche Diagnostics, Mannheim, Germany), 15 μg/mL DNA (pJOE4056.2, kindly provided by J. Altenbuchner, IIG, University of Stuttgart), 24% (v/v) S30 extract. Supplemented tRNA was derived from fast growing *E. coli* (Roche).

The production of eGFP was monitored online by fluorescence detection (excitation filter 485 nm, emission 520 nm) using a Synergy 2 plate reader (Biotek Instruments, USA) at 37 °C. Quantification of eGFP was performed using a purified eGFP standard, which was quantified by amino acid analysis.

3.5. Mechanistic Model. The mechanistic model for transcription and translation is based on the approach of Arnold et al.,[17] who were able to simulate their experimental findings of protein synthesis. However, the model was deemed too complex for parameter identification and sensitivity analysis, and was subsequently reduced. In essence, the DNA sequence oriented transcription model was replaced by a black box approach, which eliminated more than 250 differential equations, depending on the gene length of the target protein.

3.5.1. Transcription. Michaelis−Menten kinetics was used to model T7 polymerase activity, considering plasmid (C_D) as the substrate with the binding affinity ($K_{M,D}$) of Arnold et al.[30] and a maximum activity of 0.0519 μM min^{-1} (from the manufacturer). Other substrates, such as NTPs, were neglected because their respective concentrations highly exceeded the affinity toward the RNA polymerase of Arnold et al.[30]

$$\nu_{TC} = \nu_{T7}^{max} \frac{K_{M,D}}{K_{M,D} + C_D} \tag{3}$$

mRNA degradation is unavoidable because of the release of RNases during lysate preparation. Accordingly, it was considered as mass action kinetics with a degradation constant $k_{mRNA\ deg} = 0.0083\ min^{-1}$.[31]

$$\nu_{mRNA\,deg} = C_{mRNA} k_{mRNA\,deg} \tag{4}$$

Changes in available mRNA can be described as transcription (ν_{T7}) minus mRNA degradation ($\nu_{mRNA\ deg}$).

$$\frac{dC_{mRNA}}{dt} = \nu_{TC} - \nu_{mRNA\,deg} \tag{5}$$

3.5.2. Translation. The translation model is based on previously published work by Arnold et al.,[17] which describes translation using the ribosome flow approach. It allows the incorporation of DNA sequence oriented translation elongation and provides a detailed description of ribosomal movement along an mRNA template. Here, we focused on potentially achievable translation rates in *in vitro* reactions; therefore, we neglected any catalyst inactivation. For each codon *j* in the DNA-sequence, we balanced the concentration of ribosomes (C_j^R) that were active on this codon.

3.5.2.1. Initiation. Translation initiation (TLI) is treated as a multistep process. First, ribosomal 70S units dissociate into 30S and 50S subunits (association constant $K_{70S} = 5.3 \times 10^7\ M^{-1}$, according to Zucker and Hershey[32]). Subsequently, the 30S subunit binds to three initiation factors (IF1, IF2, and IF3) in a random order. The binding of IF is assumed to reach "instantaneous" equilibrium, leading to formation of the 30S IF complex. Equilibrium constants and conservation equations were taken from Arnold et al.[17]

In the second step, the 30S IF complex binds to mRNA and fMet-tRNA, again in a random order. The resulting complex then links to the 50S subunit and triggers the release of IF1 and IF3. A net reaction rate equation was derived from Arnold et al.[17]

$$\nu_{TLI,70SIC} = \frac{q_j^R k_{TLI,70SIC} C_{30SIF}}{1 + \frac{K_{M,fMet-tRNA}}{C_{fMet-tRNA}} + \frac{K_{M,mRNA}}{C_{mRNA}} + \frac{K_{M,50S}}{C_{50S}} + \frac{K_{M,fMet-tRNA} K_{M,mRNA}}{C_{fMet-tRNA} C_{mRNA}}} \tag{6}$$

DOI: 10.1021/acssynbio.7b00117
ACS Synth. Biol. 2017, 6, 1913−1921

The initiation rate constant $k_{TLI,70SIC}$ was set to 0.1 s^{-1}.[33] Any influence of ribosomal hindrance was incorporated by applying a queuing factor q_j^R, which reduces the elongation rate with increasing ribosome density. For an elongation step from codon j to $j + 1$, the queuing factor corresponds to the density of ribosomes on the L_R codons further downstream (L_R = 12 codons, the number of codons occupied by a single ribosome according to Rauhut et al.[34] Ribosomes on these codons are close enough to the elongating ribosome to reduce the elongation rate.

$$q_j^R = \frac{1 - \frac{\sum_{s=1}^{L_R} C_{j+s}^R}{C_{mRNA}}}{1 - \frac{\sum_{s=1}^{L_R-1} C_{j+s}^R}{C_{mRNA}}} = \frac{C_{mRNA} - \sum_{s=1}^{L_R} C_{j+s}^R}{C_{mRNA} - \sum_{s=1}^{L_R-1} C_{j+s}^R}$$

$$= 1 - \frac{C_{j+L_R}^R}{C_{mRNA} - \sum_{s=1}^{L_R-1} C_{j+s}^R} \quad (7)$$

After formation of the 70S IC by the reaction $v_{TLI\ 70SIC}$, IF2 is released by GTP hydrolysis of the IF2 GTP complex with reaction rate $v_{TLI,IF2D}$, following first-order kinetics with $k_{TLI,IF2D}$ = 1.5 s^{-1}.[35] This reaction is the last step of initiation and ends with the approach of one ribosome at the first codon. According to Arnold et al.,[17] the released IF2 is assumed to be bound to GTP, therefore it is not necessary to consider the regeneration of IF2.

$$v_{TLI,IF2D} = k_{TLI,IF2D} C_{70SIC} \quad (8)$$

3.5.2.2. Elongation. Translation elongation (TLE) is regarded as DNA sequence oriented with individual reaction rates per elongation step. Accordingly, the elongation factors G (EFG, GTP coupled form) and Tu (EFTu, GTP coupled form) and the codon-dependent tRNA were considered. EFTu and tRNA were used as a pair, combined in the ternary complex (T3). The elongation rate from codon j to codon $j + 1$ was derived from Arnold et al.[17] and is defined as $v_{TLE,j+1}$. Calculation of the queuing factor used equation 5 and k_{TLE} was set to 24 amino acids per ribosome per second.[17] The individual tRNA/codon match for codon j followed the tRNA-codon table from Dong et al.[24]

$$v_{TLE,j+1} = \frac{q_j^R k_{TLE} C_R^j}{1 + \frac{K_{M,T3_j}}{\sum C_{T3_{j,i}}} + \frac{K_{M,EFGGTP}}{C_{EFGGTP}}} \quad (9)$$

The recharging of tRNAs was treated as a rapid equilibrium binding of substrates, and the lack of available parameters (for example only kcat values for 10 amino acids are described in literature) made it necessary that kinetic constants were assumed to be the same for all tRNA species. We also assumed that all aminoacyl tRNA synthetases are equally functional in the system. A detailed description of the mechanism can be found in the Supporting Information section 1.1.

3.5.2.3. Termination. Translation termination (TLT) describes the process by which the ribosome (located at the codon prior to the stop codon K) dissociates from the mRNA and releases the 70S ribosomes and the completed peptide chain. This reaction is catalyzed by releasing factors (RF). According to Arnold et al.,[17] reaction parameters were set to K_{M,R_K} = 8.3 nM, $K_{M,GTP}$ = 100 μM and k_{TLT} = 24 s^{-1}.

$$v_{TLT} = \frac{k_{TLT} C_{RF}}{1 + \frac{K_{M,R_K}}{C_R^K} + \frac{K_{M,GTP}}{C_{GTP}} + \frac{K_{M,R_K} K_{M,GTP}}{C_R^K C_{GTP}}} \quad (10)$$

3.6. Numerical Integration. The complete model comprises more than 400 first order differential equations and eight algebraic equations. The resulting differential–algebraic equation (DAE) system was numerically integrated using a variable-step, variable-order solver based on the numerical differentiation formula.[36] This method was chosen because it proved to be more robust toward inconsistent sets of initial conditions compared to that of other approaches such as in the work of Dassl.[37]

3.7. Dilution Factor. As a consequence of sample processing, concentrations of the *in vivo* systems were diluted by a factor D compared to that of *in vitro* conditions. The latter was derived from experimental studies that quantified ribosome concentrations and even considered likewise bias due to lysate preparation.[28] Recovery for all molecules was set to 1, except for ribosomes (R^R = 0.9) and EFTu (R^{EFTu} = 0.8).

$$D = \frac{C_{invivo}^i R^i}{C_{invitro}^i} \quad (11)$$

R^i = recovery of molecule i during lysate preparation

3.8. Translational Control Analysis. The influence of different reaction substrates i on the resulting translation rate can be described in terms of elasticity.[38] The influence of the active enzyme (in this case the total amount of ribosomes) on the translation rate can be described by the flux control coefficient. Partial differentials were assessed by deflecting the respective parameter by ±1%.

$$\epsilon_i = \frac{C_i}{v_{TL}} \frac{\partial v_{TL}}{\partial C_i} \quad (12)$$

$$FCC = \frac{C_{70S,t}}{v_{TL}} \frac{\partial v_{TL}}{\partial C_{70S,t}} \quad (13)$$

4. CONCLUSION

In summary, the reduced model presented here was able to accurately predict translation rates in CFPS systems, and showed that fine-tuning of parameters such as tRNA composition may further improve predictability. The results show that (i) we successfully identified a model that describes cell-free protein synthesis, (ii) the model is able to accurately predict translation rates, and (iii) the model can be used for sensitivity analysis, for examplel, to identify rate-controlling steps.

In comparison to black box models (e.g., refs 20 and 21), we show that using a detailed mechanistic model facilitates the analysis of the complex translational network (e.g., the amount of free ternary complexes for all species can be unraveled). Using this systems biology-based methodology, we found that concentrations of EFTu and tRNA are limiting *in vitro* translation rates. This bottleneck can be partially overcome by supplementing the reaction system with EFTu and tRNA. However, this also shifts translation control to initiation, thereby increasing the cost of cell-free protein synthesis. In essence, we have identified a shift of translation control that is induced by EFTu and tRNA, making it necessary to iteratively optimize similar CFPS. Nevertheless, *in vitro* translation rates still lag behind *in vivo* activities. An equilibrated increase in the

DOI: 10.1021/acssynbio.7b00117
ACS Synth. Biol. 2017, 6, 1913–1921

main components (as shown above) will help to improve performance. However, it is important to consider that *in vitro* applications are inherently hampered by one-sided translation of distinct mRNA, which complicates the one-by-one comparison with *in vivo* data. In summary, our model provides guidance with respect to bridging the gap between *in vitro* and *in vivo* protein synthesis.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.7b00117.

Additional information for the mechanistic model; results of the optimized experimental design (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: (+49) 711 685 64535. Fax: (+49) 711 685 65164. E-mail: takors@ibvt.uni-stuttgart.de.

ORCID

Alexander Nieß: 0000-0001-8078-1449

Author Contributions

#A.N. and J.F. contributed equally to this work. A.N. and M.K. performed the modeling; J.F. performed the experiments; A.N., J.F., M.S.-H., and R.T. wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Eike Krauter and Michael Kraml for their assistance with the cell-free reactions. We further gratefully acknowledge the funding of this work by the Bundesministerium für Bildung und Forschung (BMBF; Grant FKZ031A157D) and the European Union ("ST-Flow" project, Grant 289326 in the framework 7 program KBBE.2011.3.6-03).

■ REFERENCES

(1) Chen, H., Xu, Z., Xu, N., and Cen, P. (2005) Efficient production of a soluble fusion protein containing human beta-defensin-2 in *E. coli* cell-free system. *J. Biotechnol. 115*, 307−315.

(2) Catherine, C., Lee, K.-H., Oh, S.-J., and Kim, D.-M. (2013) Cell-free platforms for flexible expression and screening of enzymes. *Biotechnol. Adv. 31*, 797−803.

(3) Katzen, F., Chang, G., and Kudlicki, W. (2005) The past, present and future of cell-free protein synthesis. *Trends Biotechnol. 23*, 150−156.

(4) 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 Synth. Biol. 5*, 344−355.

(5) Caschera, F., and Noireaux, V. (2015) A cost-effective polyphosphate-based metabolism fuels an all E. coli cell-free expression system. *Metab. Eng. 27*, 29−37.

(6) Kim, D. M., and Swartz, J. R. (2001) Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnol. Bioeng. 74*, 309−316.

(7) Jewett, M. C., and Swartz, J. R. (2004) Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng. 86*, 19−26.

(8) Calhoun, K. A., and Swartz, J. R. (2005) An economical method for cell-free protein synthesis using glucose and nucleoside monophosphates. *Biotechnol. Prog. 21*, 1146−1153.

(9) Kim, T.-W., Oh, I.-S., Keum, J.-W., Kwon, Y.-C., Byun, J.-Y., Lee, K.-H., Choi, C.-Y., and Kim, D.-M. (2007) Prolonged cell-free protein synthesis using dual energy sources: Combined use of creatine phosphate and glucose for the efficient supply of ATP and retarded accumulation of phosphate. *Biotechnol. Bioeng. 97*, 1510−1515.

(10) Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J., and Swartz, J. R. (2008) An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol. 4*, 220.

(11) Kim, H.-C., Kim, T.-W., and Kim, D.-M. (2011) Prolonged production of proteins in a cell-free protein synthesis system using polymeric carbohydrates as an energy source. *Process Biochem. 46*, 1366−1369.

(12) Underwood, K. A., Swartz, J. R., and Puglisi, J. D. (2005) Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. *Biotechnol. Bioeng. 91*, 425−435.

(13) Laursen, B. S., Sorensen, H. P., Mortensen, K. K., and Sperling-Petersen, H. U. (2005) Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev. 69*, 101−123.

(14) Jacques, N., and Dreyfus, M. (1990) Translation initiation in Escherichia coli. Old and new questions. *Mol. Microbiol. 4*, 1063−1067.

(15) Pedersen, A., Hellberg, K., Enberg, J., and Karlsson, B. G. (2011) Rational improvement of cell-free protein synthesis. *New Biotechnol. 28*, 218−224.

(16) Freischmidt, A., Meysing, M., Liss, M., Wagner, R., Kalbitzer, H. R., and Horn, G. (2010) Limiting factors of the translation machinery. *J. Biotechnol. 150*, 44−50.

(17) Arnold, S., Siemann-Herzberg, M., Schmid, J., and Reuss, M. (2005) Model-based inference of gene expression dynamics from sequence information, in *Biotechnology for the Future* (Nielsen, J., Ed.) pp 89−179, Springer Berlin Heidelberg, Berlin, Heidelberg.

(18) Caschera, F., and Noireaux, V. (2014) Synthesis of 2.3 mg/mL of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie 99*, 162−168.

(19) Iskakova, M. B., Szaflarski, W., Dreyfus, M., Remme, J., and Nierhaus, K. H. (2006) Troubleshooting coupled in vitro transcription-translation system derived from Escherichia coli cells: synthesis of high-yield fully active proteins. *Nucleic Acids Res. 34*, e135.

(20) Stögbauer, T. R., Windhager, L., Zimmer, R., and Radler, J. O. (2012) Experiment and mathematical modeling of gene expression dynamics in a cell-free system. *Integrative biology: quantitative biosciences from nano to macro 4*, 494−501.

(21) Chizzolini, F., Forlin, M., Yeh Martin, N., Berloffa, G., Cecchi, D., and Mansy, S. S. (2017) Cell-Free Translation Is More Variable than Transcription. *ACS Synth. Biol. 6*, 638−647.

(22) Bremer, H., and Dennis, P. P. (1987) Modulation of chemical composition and other parameters of the cell by growth rate, in *Escherichia coli and salmonella typhimurium. Cellular and molecular biology* (Neidhardt, F. C., Ed.) pp 1553−1569, American Society for Microbiology, Washington, D.C.

(23) Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol. 25*, 117−124.

(24) Dong, H., Nilsson, L., and Kurland, C. G. (1996) Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J. Mol. Biol. 260*, 649−663.

(25) Rudorf, S., and Lipowsky, R. (2015) Protein synthesis in *E. coli*: Dependence of codon-specific elongation on tRNA concentration and codon usage. *PLoS One 10*, e0134994.

(26) Liu, D. V., Zawada, J. F., and Swartz, J. R. (2005) Streamlining Escherichia coli S30 extract preparation for economical cell-free protein synthesis. *Biotechnol. Prog. 21*, 460−465.

(27) Pratt (1984) Coupled transcription-translation in prokaryotic cell-free systems, in *Transcription and translation: A practical approach*. (Hames, B. D., and Higgins, S. J., Eds.) pp 179−210, IRL Press, New York.

(28) Failmezger, J., Nitschel, R., Sanchez-Kopper, A., Kraml, M., and Siemann-Herzberg, M. (2016) Site-specific cleavage of ribosomal RNA in *Escherichia coli*-based cell-free protein synthesis systems. *PLoS One 11*, e0168764.

(29) Failmezger, J., Ludwig, J., Nieß, A., and Siemann-Herzberg, M. (2017) Quantifying ribosome dynamics in Escherichia coli using

fluorescence. *FEMS Microbiol. Ecol. 364*, fnx055 DOI: 10.1093/femsle/fnx055.
(30) Arnold, S., Siemann, M., Scharnweber, K., Werner, M., Baumann, S., and Reuss, M. (2001) Kinetic modeling and simulation of in vitro transcription by phage T7 RNA polymerase. *Biotechnol. Bioeng. 72*, 548−561.
(31) Hargrove, J. L., and Schmidt, F. H. (1989) The role of mRNA and protein stability in gene expression. *FASEB J. 3*, 2360−2370.
(32) Zucker, F. H., and Hershey, J. W. B. (1986) Binding of Escherichia coli protein synthesis initiation factor IF1 to 30S ribosomal subunits measured by fluorescence polarization. *Biochemistry 25*, 3682−3690.
(33) Wintermeyer, W., and Gualerzi, C. (1983) Effect of Escherichia coli initiation factors on the kinetics of N-AcPhe-tRNAPhe binding to 30S ribosomal subunits. A fluorescence stopped-flow study. *Biochemistry 22*, 690−694.
(34) Rauhut, R., and Klug, G. (1999) mRNA degradation in bacteria. *FEMS Microbiol Rev. 23*, 353−370.
(35) Tomsic, J., Vitali, L. A., Daviter, T., Savelsbergh, A., Spurio, R., Striebeck, P., Wintermeyer, W., Rodnina, M. V., and Gualerzi, C. O. (2000) Late events of translation initiation in bacteria: a kinetic analysis. *EMBO. J. 19*, 2127−2136.
(36) Shampine, L. F., Reichelt, M. W., and Kierzenka, J. A. (1999) Solving Index-1 DAEs in MATLAB and Simulink. *SIAM Rev. 41*, 538−552.
(37) Brenan, K. E., Campbell, S. L., and Petzold, L. R. (1995) in *Numerical solution of initial-value problems in differential-algebraic equations*, Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), Philadelphia, Pa.
(38) Zouridis, H., and Hatzimanikatis, V. (2008) Effects of codon distributions and tRNA competition on protein translation. *Biophys. J. 95*, 1018−1033.

DOI: 10.1021/acssynbio.7b00117
ACS Synth. Biol. 2017, 6, 1913−1921

MANUSCRIPT B

Repetitive Short-Term Stimuli Imposed in Poor Mixing Zones Induce Long-Term Adaptation of *E. coli* Cultures in Large-Scale Bioreactors: Experimental Evidence and Mathematical Model

frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 June 2017
doi: 10.3389/fmicb.2017.01195

Repetitive Short-Term Stimuli Imposed in Poor Mixing Zones Induce Long-Term Adaptation of *E. coli* Cultures in Large-Scale Bioreactors: Experimental Evidence and Mathematical Model

*Alexander Nieß, Michael Löffler, Joana D. Simen and Ralf Takors**

Institute of Biochemical Engineering, University of Stuttgart, Stuttgart, Germany

Rapidly changing concentrations of substrates frequently occur during large-scale microbial cultivations. These changing conditions, caused by large mixing times, result in a heterogeneous population distribution. Here, we present a powerful and efficient modeling approach to predict the influence of varying substrate levels on the transcriptional and translational response of the cell. This approach consists of two parts, a single-cell model to describe transcription and translation for an exemplary operon (*trp* operon) and a second part to characterize cell distribution during the experimental setup. Combination of both models enables prediction of transcriptional patterns for the whole population. In summary, the resulting model is not only able to anticipate the experimentally observed short-term and long-term transcriptional response, it further allows envision of altered protein levels. Our model shows that locally induced stress responses propagate throughout the bioreactor, resulting in temporal, and spatial population heterogeneity. Stress induced transcriptional response leads to a new population steady-state shortly after imposing fluctuating substrate conditions. In contrast, the protein levels take more than 10 h to achieve steady-state conditions.

Keywords: scale-down, hybrid modeling, population heterogeneity, adaptation times, *Escherichia coli*

OPEN ACCESS

Edited by:
Jan-Ulrich Kreft,
University of Birmingham,
United Kingdom

Reviewed by:
Peter Neubauer,
Technische Universität Berlin,
Germany
Frank Delvigne,
University of Liège, Belgium

****Correspondence:***
Ralf Takors
takors@ibvt.uni-stuttgart.de

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *29 April 2017*
Accepted: *12 June 2017*
Published: *28 June 2017*

Citation:
Nieß A, Löffler M, Simen JD and Takors R (2017) Repetitive Short-Term Stimuli Imposed in Poor Mixing Zones Induce Long-Term Adaptation of E. coli Cultures in Large-Scale Bioreactors: Experimental Evidence and Mathematical Model. Front. Microbiol. 8:1195. doi: 10.3389/fmicb.2017.01195

INTRODUCTION

Large-scale industrial bioprocesses make use of reactors ranging from 100 to 800 m^3 reaction volume. For aerobic processes, stirred tank reactors are still preferred, albeit alternative setups such as airlift reactors may be attractive if reactor sizes exceed the volume of about 500 m^3. All reactors have in common that gradients of substrates, dissolved gases and pH occur, which are the consequence of poor mixing conditions (Nienow et al., 1997). Cells are circulating in these reactors, thereby frequently passing through zones of different substrate availability. Accordingly, cellular interactions are repeatedly triggered (Oldiges and Takors, 2005; Lara et al., 2006; Neubauer and Junne, 2010; Takors, 2012). Noteworthy, related regulatory responses are not limited to changes of metabolism but also comprise transcriptional and translational programs (Löffler et al., 2016, 2017; Simen et al., 2017).

Often, microbial processes are controlled by limited substrate feeding to avoid non-wanted overflow metabolism and to prevent too high metabolic activity that may exceed the technical capacities of aeration and cooling. Industrial examples are the implementation of glucose or ammonia limitations (Neubauer et al., 1995). Recently, Chubukov et al. (2014) outlined that proper nitrogen (or phosphate) limitation may even increase biomass specific substrate uptake during production phases when cell growth is strongly limited. Michalowski et al. (2017) further succeeded to engineer the *E. coli* HGT host for likewise conditions.

Löffler et al. (2016) and Simen et al. (2017) studied the scenario of frequently occurring glucose or ammonia limitations by using a conventional STR-PFR (stirred tank reactor—plug flow reactor) setup as described by George et al. (1993). Unlike previous investigations, these studies installed steady-state growth conditions before large-scale gradients were repeatedly imposed on the cells by connecting the PFR to the STR. As such, a distinct reference steady-state was created that enabled quantitative and highly accurate analysis of the metabolic and transcriptional responses of the cells on the installed glucose or ammonia gradients.

These data sets are the experimental basis for the modeling approaches presented in this study. By exploiting the metabolic and transcriptional time series it will be investigated whether and how similar dynamics can be modeled to predict short- and long-term regulatory responses of *E. coli*. Related data-driven models can serve as the core for ensemble modeling (Henson, 2003) to predict large scale cellular performance *in silico* and *ab initio*.

For the sake of simplicity, transcriptional dynamics of the tryptophan operon were chosen as an illustrative example. It has been shown by Simen et al. (2017) that the repetitive exposure to nitrogen starvation induced the frequent transcription of the *trp* operon. Considering the well-known attenuation control (Yanofsky, 2004, 2007), the expression of downstream genes *trpEDCBA*, consequently, indicates not only the ongoing transcriptional response on environmental triggers but also the start of protein translation. Accordingly, modeling *trp* expression dynamics needs to fulfill several challenges: (i) Short-term transcript dynamics observed in the PFR must be predicted, (ii) long-term transcript responses of the whole population should be mirrored, and (iii) the different time-scales of transcriptional and translational dynamics have to be reflected, too. This study will outline that every constraint is properly met by a simple mechanistic model.

MATERIALS AND METHODS

Experimental Setup

Oscillating substrate availability was simulated in a stirred-tank-reactor (STR) plug-flow-reactor (PFR) scale-down approach. **Figure 1** shows the schematic setup of the system. As Simen et al. (2017), the STR system was operated as nitrogen limited chemostat cultivation with a dilution rate of 0.2 h^{-1} (5 mL min^{-1}). The well-mixed bioreactor was simulated by using the steady-state condition in STR without the PFR (SS_0). After characterization of SS_0 the PFR was connected and a fraction

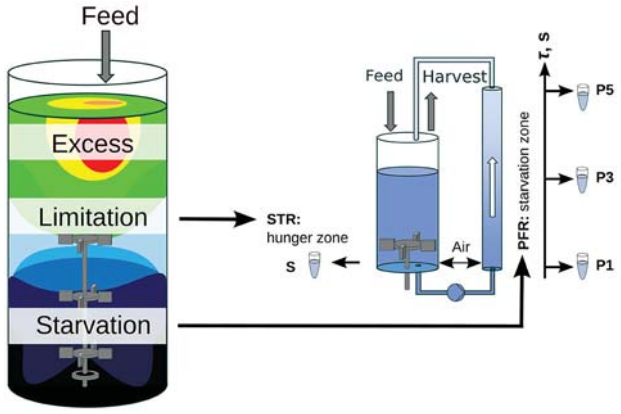


FIGURE 1 | Scheme of the scale-down approach. A large-scale bioreactor and its substrate gradient is simulated by a chemostat STR (limitation zone) coupled to a PFR (starvation zone). This approach allows the examination of the influence of substrate gradients (i.e., glucose or ammonia) on the population dynamic.

of cells cycled through the PFR loop. No additional feed was added into the PFR. Therefore, the cells shift from nitrogen limitation to starvation along the PFR. The experimental design allows the observation of the transcriptional responses along the PFR (short-term) and over the process time in the STR (long-term). The system characteristics and cultivation conditions were published in Löffler et al. (2016). Residence times τ of STR and PFR were estimated to be 6.2 min and 125 s, respectively. Samples for transcriptome analysis were taken at sampling ports P1, P3, and P5 in the PFR with corresponding residence times of 31, 70, and 110 s. Volumes in STR and PFR were 1.12 and 0.38 L, respectively. Biosuspension was continuously pumped through the PFR (180 mL min^{-1}).

Compared to SS_0 with a growth rate of 0.2 h^{-1} in the STR, the STR-PFR setup splits growth rates individually in the two compartments. Whereas, the total growth rate of the STR-PFR system is still 0.2 h^{-1}, no growth can be expected inside PFR when the nutrient is completely consumed. Accordingly, PFR can be subdivided into a first part considering still growing cells and a subsequent part characterized by no growth. Because the total system runs with the dilution rate of 0.2 h^{-1}, STR growth rates must be increased accordingly to compensate missing cell growth in the PFR. Calculation of growth rates in the STR can be performed based on the residence time distributed average growth rate that is set equal to the dilution rate. For the sake of simplicity and because nutrient starvation occurred very rapid in PFR, no growth was assumed to be existent in the total PFR compartment. Therefore, the average growth rate can be split into two different growth rates for each compartment (Equation 1).

$$\frac{\mu_{STR}\,\tau_{STR} + \mu_{PFR}\tau_{PFR}}{\tau_{STR} + \tau_{PFR}} = D \qquad (1)$$

Samples for transcriptome analysis were taken at 25 and 120 min as well as at 28 h after installing the substrate gradient.

Transcript measurements are published in Simen et al. (2017) and are available under GEO Accession GSE90743.

Single-Cell Model

For the agent based transcription-translation model, DNA and mRNA templates are discretized in nucleotides, defining a 1D lattice. Movement of RNA polymerases (RNAPs) is treated according to Equation (2). Here, x describes the relative position of nucleotides inside the operon starting with the first mRNA encoding nucleotide of the operon sequence. RNAP movement is based on the elongation rate v_{elo}^{RNAP} and the distance Δx between two subsequent polymerases. The following criteria were considered for RNAP motion:

- The first elongation step is treated as the initiation step and can only occur if t is in the interval of possible induction (t_{ind}).
- The minimum distance Δx between two subsequent RNAPs is fulfilled.

$$\frac{dx_i}{dt} = \begin{cases} 0 & if\ x_i = 0\ \wedge\ t \notin t_{ind} \\ 0 & if\ x_{i-1} - x_i < \Delta x\ \wedge i > 1 \\ v_{elo}^{RNAP} & otherwise \end{cases} \quad (2)$$

For each passed nucleotide on the DNA sequence, the respective nucleotide in the mRNA sequence is transcribed. The resulting mRNA strand i with length L_i^{mRNA} can be directly derived from x_i.

$$L_i^{mRNA}(t) = x_i(t) \quad (3)$$

For simplification, we neglected the modeling of the attenuation process considering terminator/antiterminator interactions and assumed ongoing translation only during nitrogen starvation instead. Position $y_{i,j}$ of a ribosome j on mRNA strand i is a function of L_i^{mRNA} and the position of the previous ribosome y_{j-1}. The number of ribosomes that translate a gene g ($N_g^{TL,\max}$) can vary and depends on the gene itself. Ribosomal motion on a gene g starts at C_g^{start} (first coding nucleotide) and stops at C_g^{end} (third nucleotide of the terminating codon). The necessary criteria for translation are stated as follows:

- At least Δy nucleotides downstream are already synthesized.
- The previous ribosome is more than Δy nucleotides further downstream.
- The maximum number of translations for the given gene is not exceeded.

$$\frac{dy_{i,j}}{dt} = \begin{cases} 0 & if\ L_i - y_{i,j} \le \Delta y \\ 0 & if\ y_{i,j-1} - y_{i,j} < \Delta y\ \wedge j > 1 \\ 0 & N_{i,g}^{TL}(t) \ge N_g^{TL,\max} \\ v_{elo}^{Ribosome} & otherwise \end{cases} \quad (4)$$

The number of ribosomes acting on a single mRNA i is calculated following the Iverson brackets (Equation 5). These brackets return 1 if the term inside is true and 0 if the term is false.

$$N_{i,g}^{TL}(t) = \sum_j \left[y_{i,j}(t) \ge C_g^{start} \right] \quad (5)$$

The amount of synthesized proteins per cell from the single-cell model ($N_{g,\,SC}^{Protein}$) encoded by gene g can be calculated as the sum of all ribosomes acting on all mRNA strands that have passed the final nucleotide C_g^{end}.

$$N_{g,\,SC}^{Protein}(t) = \sum_i \sum_j \left[y_{i,j} > C_g^{end} \right] \quad (6)$$

Each mRNA strand is expected to be degraded by RNases. Initiation of mRNA breakdown begins at the start codon of transcription. Movement along mRNA is encoded by position z_i on strand i and depends on the degradation elongation rate v_{elo}^{RNAse}. The following constraints define the motion of RNAses:

- The number of active ribosomes per gene g $N_g^{TL,\max}$ is estimated as the turnover ratio of mRNAs and proteins (see below)
- Δz encodes the closest nucleotide distance to the next ribosome downstream of z_i

$$\frac{dz_i}{dt} = \begin{cases} 0 & if\ N_{i,g}^{TL}(t) < N_g^{TL,max} \\ 0 & if\ y_{i,\,N_g^{TL,max}} - z_i \le \Delta z \\ v_{elo}^{RNase} & otherwise \end{cases} \quad (7)$$

Accordingly, the current amount of mRNA per gene is calculated as the difference of already synthesized mRNAs and the amount of degraded mRNAs. The first is modeled from the number of complete mRNA strands synthesized. The second mirrors the amount of RNases that have passed the first codon.

$$N_{g,\,SC}^{mRNA}(t) = \sum_i \left[L_i^{mRNA}(t) > C_g^{end} \right] - \sum_i \left[z_i \ge C_g^{start} \right] \quad (8)$$

$N_g^{TL,\max}$ is calculated as the turnover ratio of mRNAs per protein for a given gene g. Protein turnover $r_{turnover}^{Protein}$ was calculated based on protein levels at a growth rate of 0.2 h^{-1} ($k_{\deg}^{Protein} = \mu$) (Valgepea et al., 2013). Active protein degradation was neglected and only growth based dilution was considered. mRNA turnover $r_{turnover}^{mRNA}$ was calculated based on the levels measured by Valgepea et al. (2013) with average half-lives of 2 min ($k_{\deg}^{mRNA} = 20.79\ h^{-1}$) (Chen et al., 2015). However, no mRNA measurements of *trpA* were given in this data set. We thus assumed the translations per mRNA for *trpA* to be the same as for *trpB*, due to the fact that the resulting protein complex is a tetramer consisting of two *trpA* and two *trpB* (Hyde et al., 1988). *TrpL*, the leader peptide, was neglected in this calculation. **Table 1** shows the resulting translations per mRNA.

$$N_g^{TL.\,\max} = \frac{r_{turnover}^{Protein}}{r_{turnover}^{mRNA}} = \frac{c_g^{Protein}\ k_{\deg}^{Protein}}{c_g^{mRNA}\ k_{\deg}^{mRNA}} \quad (9)$$

We used the *trp* operon as an example for several reasons: (i) The *trp* operon leads to a polycistronic mRNA (Yanofsky

TABLE 1 | Calculated translations per mRNA for the *trp* operon.

Gene	*trpE*	*trpD*	*trpC*	*trpB*	*trpA*
Translations per mRNA	4	4	5	10	10

The value for trpA was extrapolated from trpB.

et al., 1981), (ii) the attenuation sequence in the *trpL* leader peptide allows the coupled investigation of transcription and translation (see **Figure 2**), and (iii) the published data by Simen et al. shows that the *trp* operon is upregulated during STR-PFR cultivations. Accordingly, translation must have happened if transcripts of genes downstream of *trpL* are measured, as it is the case in the data sets used for this study. Simplification was made by treating the structural genes *trpGD* and *trpCF* and their corresponding proteins as single genes (*trpG* and *trpC*, respectively) and proteins.

All three actively moving species (RNAP, ribosomes and RNase) are treated as equally fast and their elongation rate was taken from the RNAP elongation rate reported by Chen et al. (2015) and set to 21 nucleotides s^{-1} (see **Table 2**). Minimum distances Δx, Δy, and Δz were set to 100 nucleotides each [which is larger than (Bremer and Dennis, 1987) estimated for a growth rate of 0.5 h^{-1}].

Each PFR passage induced transcription, however, with a delay of 30 s based on experimental observations. Once induction has started and RNAP has passed the attenuation sequence, transcription was assumed to continue until the terminator sequence after *trpA* was reached (see **Figure 2**).

Cell Distribution Model

The ensemble cell model needs to be embedded in a process model for describing the flow wiring and residence times of the cells in the compartments. The PFR is considered as a plug flow reactor showing almost equally distributed residence times for all cells. The STR is assumed to be ideally mixed, thus, having a residence time distribution constrained by the reaction volume and the throughput. Dilution and growth rate additionally influence the population.

For population balancing, the following events were considered to track the fate of each individual cell:

Cells may

1. leave the STR for entering the PFR and cycle back into STR after τ_{PFR}, the residence time in the PFR
2. be drained off by the efflux (harvest)
3. divide, setting all transcriptional and translational programs on default (no initiation of transcription or translation in the corresponding daughter cell)

The following probability functions α_i were defined

$$\alpha_1 = N_{STR} \frac{\dot{V}_{PFR}}{V_{STR}} \tag{10}$$

$$\alpha_2 = N_{STR} \frac{\dot{V}_{Feed}}{V_{STR}} \tag{11}$$

$$\alpha_3 = N^0_{STR} D \tag{12}$$

For modeling event (1), the rate $\alpha_1$is used, indicating that a cell leaves the STR and enters the PFR again. Washout of cells (event 2) was treated equally with the dilution rate as flux value (α_2). The probability for cell division (α_3) is based on the set dilution rate D and the cell number N^0_{STR} during SS_0. Return of cells from the PFR compartment was fixed to occur after τ_{PFR} passed. Cells that are washed out by event (2) are deleted from the system and newly born cells from event (3) are treated as default daughter cells without any transcriptional deflection.

The reaction system was numerically solved by applying Gillespie's stochastic simulation algorithm (Gillespie, 1977). Time increment τ was solved based on the sum of the three reaction events considering the probabilities as indicated in Equation (13). The chosen reaction i is calculated, based on Equation (14). r_1 and r_2 are uniformly distributed random variables in the interval (0, 1).

$$\tau = \frac{1}{\sum \alpha_i} \ln\left(\frac{1}{r_1}\right) \tag{13}$$

$$\sum_{j=1}^{i-1} \alpha_j \leq r_2 \sum_{j=1}^{3} \alpha_j \leq \sum_{j=1}^{i} \alpha_j \tag{14}$$

Simulations were performed using 10,000 cells, assuming uniform distribution in the STR (N^0_{STR}) before it is connected to the PFR. Simulations tracked cell numbers in the STR and the PFR as well as each transition of a cell from STR to PFR.

Coupling of Single-Cell and Cell Distribution Model

To minimize computational efforts, the impact of single-cell metabolic activities on the local environment was ignored. In essence, cells were considered to travel through a "frozen" bioreactor background that triggers transcriptional and translational responses as reflected in the single-cell model. For balancing the population distribution properly, the history of every cell was tracked. As the trigger "PFR" is of outstanding importance, the entrance of each cell into the PFR was logged. The resulting in a set of time flags (t_i^{flag}) for each cell that was stored for the total simulation period, which allows detailed tracking of the cells motion in the STR-PFR setup. Additionally, the events (2) and (3) were tracked for each cell allowing the calculation of the population distribution in the STR and the PFR at each time step of simulation.

The simulation approach allowed the independent solution of the single-cell and cell distribution model. Simulations of the single-cell model resulted in distinct mRNA ($N^{mRNA}_{g,\,SC}(t)$) and protein ($N^{Protein}_{g,\,SC}(t)$) patterns for every cell entering and leaving the PFR and this constant sequence can be stored as look-up table. In the distribution model, each flag indicates start of induction, whose sequence is stored in the look-up table. Duration of an induction phase is defined from entering the PFR at t^{flag} until the last mRNA is degraded at $t^{flag}+\Delta t$. Superposition of all transcriptional and translational patterns over the cells lifetime results in a continuous description of transcriptional and translational patterns in the STR-PFR system.

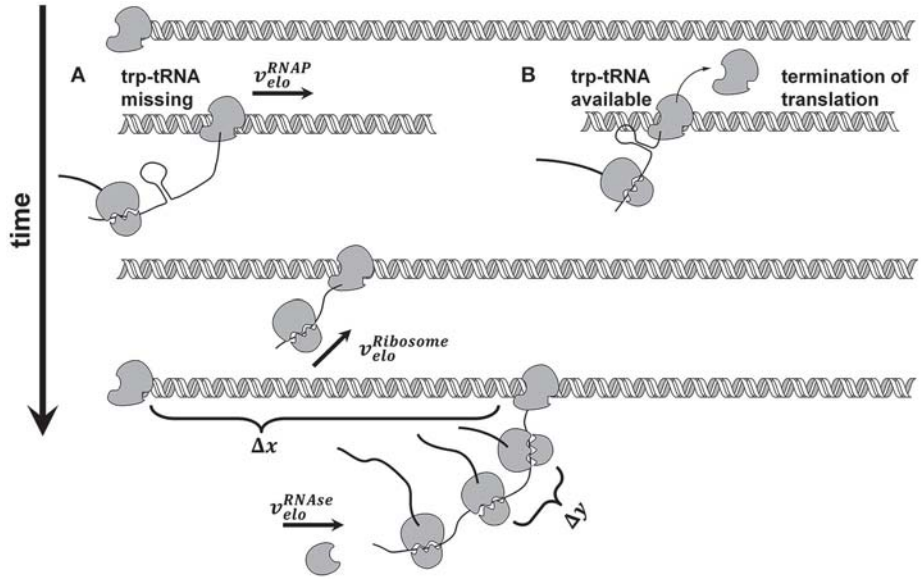


FIGURE 2 | Model of coupled transcription and translation. RNA polymerase ( ) binds to the *trp* promotor and starts transcribing with constant elongation rate. After transcription of each genes ribosome binding site, translation takes place and ribosomes ( ) elongate with constant elongation rate that is set equal to RNAP rate. If trp-tRNA is missing, translation continues **(A)**, if trp-tRNA is available, a terminator sequence forms and translation stops **(B)**. Each gene has its own number of translations before degradation takes place. Degradation by RNases ( ) starts at the 5′ end and continues from gene to gene if the mentioned number of translations already took place.

Cellular growth by event (3) is treated as generation of a new default cell without any additional mRNA and protein content without altering the mother cell.

$$N_g^{mRNA}(t) = \sum_i \begin{cases} N_{g,SC}^{mRNA}\left(t - t_i^{flag}\right) & if\ t - t_i^{flag} \leq \Delta t \\ 0 & otherwise \end{cases} \quad (15)$$

$$N_g^{Protein}(t) = \sum_i \begin{cases} N_{g,SC}^{Protein}\left(t - t_i^{flag}\right) & t - t_i^{flag} \leq \Delta t \\ N_{g,SC}^{Protein}(\Delta t) & otherwise \end{cases} \quad (16)$$

RESULTS

Key assumptions

Löffler et al. (2016) and Simen et al. (2017) showed that the repeated oscillation of the substrate availability of *E. coli*, simulated with a STR-PFR system, induce repeated on/off switching of several hundred genes. Among them, the frequent activation of the tryptophan operon could be observed (Simen et al., 2017). The mathematical model comprising the (Equations 2–16) was used to describe not only short- and long-term transcriptional dynamics but also to estimate the impact on protein formation by linking the transcription with the translation machinery. The following key assumptions were made: (i) Once transcription of mRNA has started, it continued until the stop signal was achieved at the end of the operon, namely on the relative position 6726 nt after *trpEDCBA* (Stoltzfus et al., 1988), (ii) mRNA was assumed to be immediately translated into proteins. The number of active ribosomes per gene (encoding mRNA) was calculated based on the experimental findings of Valgepea et al. (2013).

TABLE 2 | Model parameters used for simulation of both single-cell and cell distribution model.

Parameter	Value	Unit
ν_{elo}^{RNAP}	21	Nucleotides per second
$\nu_{elo}^{Ribosome}$	21	Nucleotides per second
ν_{elo}^{RNase}	21	Nucleotides per second
Δx	100	Nucleotides
Δy	100	Nucleotides
Δz	100	Nucleotides
t_{ind}	[30 125]	Seconds
$\dot{V}_{PFR}$	180	mL min^{-1}
$\dot{V}_{Feed}$	5	mL min^{-1}
V_{STR}	1,120	mL
D	0.2	h^{-1}
N_{STR}^0	10,000	cells

Modeling Short-Term Transcriptional Dynamics

The simulation of transcriptional dynamics during a single PFR passage was achieved using the single-cell model with the parameters listed in **Table 1**. **Figure 3A** depicts mRNA courses of two subsequent PFR-STR passages. At $t = 0$, the PFR entering cell is induced and initiates transcription after the experimentally observed delay time of 30 s. Then, transcription

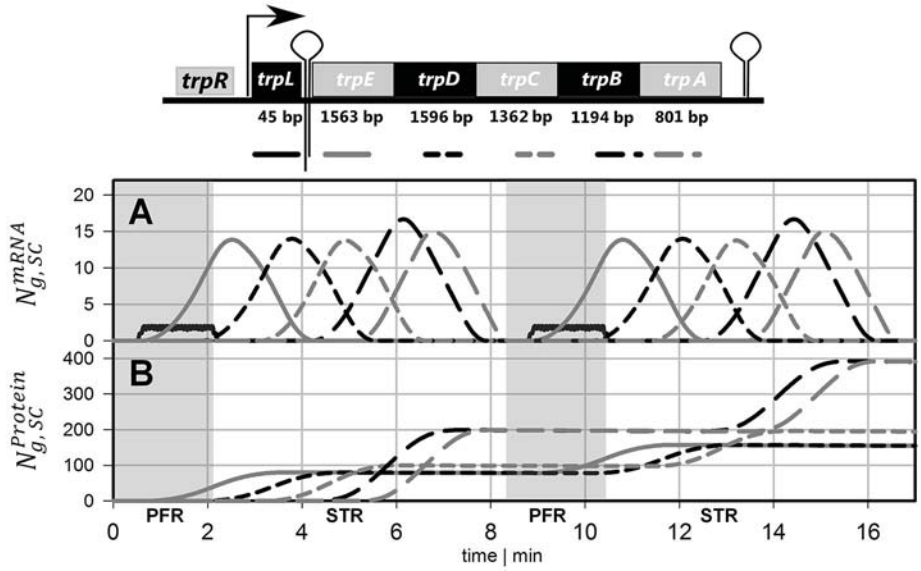


FIGURE 3 | Time courses of two subsequent cell cycles comprising PFR and STR passage. **(A)** mRNA and **(B)** protein profiles are shown, both simulated using the single-cell model. The genes and their gene products are encoded as follows: Black solid line = *trpL*, gray solid line = *trpE*, black short dashed line = *trpD*, gray short dashed line = *trpC*, black long dashed line = *trpB*, gray long dashed line = *trpA*. Gray shaded areas indicate that the cell currently passes the PFR.

of the *trp* operon starts with *trpL*. As shown, *trpL* is fully and *trpE* partially transcribed before the cell leaves the PFR. Accordingly, the remaining genes downstream of the operon were transcribed after the cell reenters the STR. Shortly after initiation, degradation of *trpL* mRNA has started, as indicated by the constant mRNA levels. After leaving the PFR, the cell stops further RNAP initiation and RNases immediately degrade the remaining transcripts. Noteworthy, all gene transcripts were fully degraded (except for a small residual of *trpA*) before the cells again reentered the PFR.

Modeling Protein Formation

Based on the single-cell model, translation of mRNA was simulated as depicted in **Figure 3B**. It was assumed that protein formation started as soon as the ribosomal binding site was transcribed. Because *trpL* encodes the leader peptide, translation modeling was simply focused on *trpEDCBA*. First, TrpE proteins were produced while the cells passed the PFR compartment. Downstream proteins were translated after the cells reentered the STR. Consequently, the majority of translation happened in STR. Protein formation is delayed and multiplexed compared to mRNA production. Accordingly, dynamics of protein courses are less steep than those of transcript levels. The latter are characterized by fast transcription and fast mRNA degradation that finally lead to sharp peaks of transcript contents. Protein degradation is slower by orders of magnitude. Consequently, only moderate pool dynamics and even protein accumulation was observed after PFR-STR transits.

Each PFR-STR cycle lasted for about 500 s. During this period, cells managed to produce 20 mRNA copies of the complete *trp* operon. Subsequent translation enabled the formation of 80 TrpED, 100 TrpC, and 200 TrpBA copies (considering the ribosomal stoichiometry of **Table 1**) with corresponding translation rates of 9.6, 12 and 24 proteins per cell per minute.

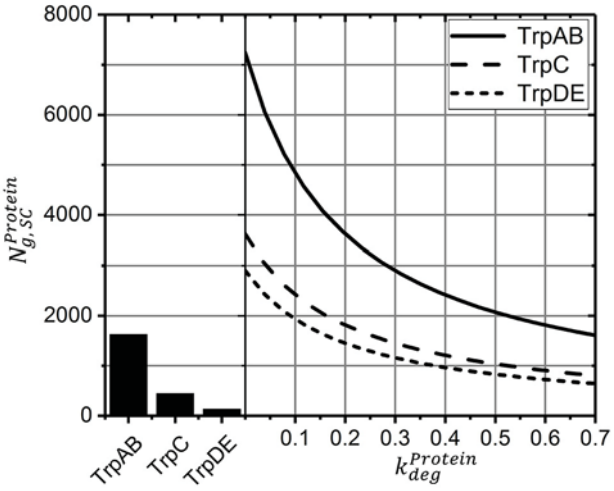


FIGURE 4 | Protein levels for the new steady-state after >15 h as a function of the degradation constant (solid lines). Higher values show the influence of declining half-lives and thus higher degradation constants. Measured protein levels (bar graph) were taken from Valgepea et al. (2013).

As outlined above, protein degradation is known to be much slower than mRNA decay which enabled the simplified simulation of protein dynamics shown in **Figure 3B**. However, the scenario may change if steady-states are analyzed. As outlined in equation 17, steady-state protein levels will be dependent on

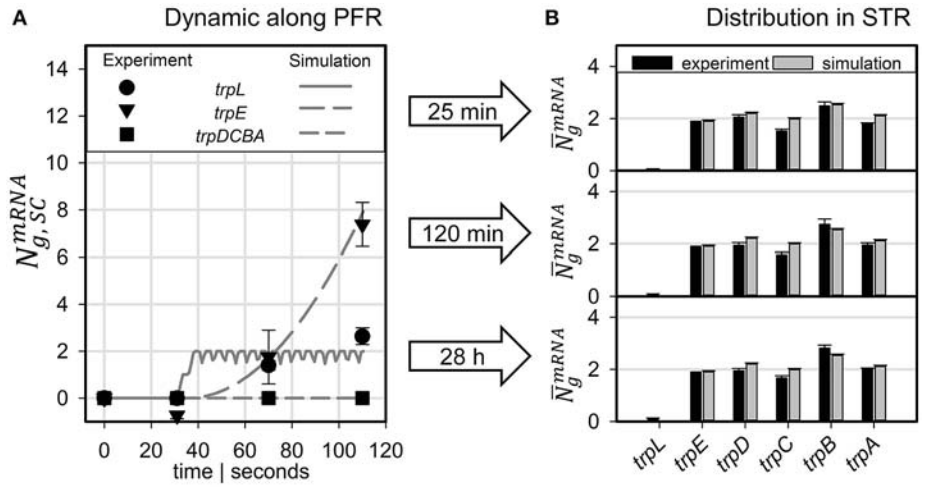


FIGURE 5 | Predicted transcript levels for **(A)** PFR residence times of 30, 71, and 110 s compared to measured values (scaled to mean *trpLE* level) and **(B)** for the STR population compared to measured values (scaled to simulated *trpE* levels).

the degradation constant.

$$\frac{dc_{Protein}}{dt} = r_{Translation} - r_{\mathrm{deg}} = r_{Translation} - c_{Protein}\, k_{\mathrm{deg}}^{Protein} = 0 \quad (17)$$

$$c_{Protein} = \frac{r_{Translation}}{k_{\mathrm{deg}}^{Protein}} \quad (18)$$

Because the individual degradation constants for the *trp* gene products are unknown, simulation studies were performed and summarized in **Figure 4**. In essence, results for $k_{\mathrm{deg}}^{Protein} = 0$ indicate protein loss simply based on cell drain under continuous operating conditions whereas results with $k_{\mathrm{deg}}^{Protein} > 0$ consider additional protein degradation with the given rates. For comparison, experimental results are indicated, too. As shown, when $k_{\mathrm{deg}}^{Protein}$ exceeds 0.6 h^{-1} (which corresponds to half-lives lower than 1.1 h) simulated protein levels are smaller than those reported for the given growth rate of 0.2 h^{-1}. Accordingly, the simplifying assumption to neglect protein degradation for simulating STR-PFR dynamics is validated as half-lives of 1.1 h fairly exceed cycling times of about 500 s (about 0.12 h).

Simulating Long-Term Adaptation

As indicated in Löffler et al. (2016) and Simen et al. (2017), the STR-PFR experiments were performed as a continuous cultivation. First, glucose- or ammonia limited steady-states were installed cultivating the cells in STR only. Then, the PFR was connected while retaining the total system dilution rate of 0.2 h^{-1}. As such, not only short-term transcript dynamics could be elucidated by sampling the PFR but also long-term adaptation of the whole population by studying transcript patterns in STR during the adaptation period of 28 h after PFR connection.

For simulation studies, the cell and the process model were linked predicting a stable distribution of 7526 ± 68 tracked cells in the STR (75.0 ± 0.68%) and 2513 ± 47 simulated cells in the PFR (25.0 ± 0.47%). Accordingly, the simulated cell population matched well with the volumetric setup comprising 74.7 vol% in the STR and 25.3 vol% in the PFR.

Neglecting the residence time distribution in the STR indicates that cells in the STR are always induced as shown in **Figure 3**. Therefore, population heterogeneity is not observable. Including residence time distribution for a perfectly mixed reactor reveals the existence of several subpopulations. Whilst 34% of the cells are currently not induced, 48% of the cells are currently induced once and 18% of the population are induced multiple times. Multiple inductions in this context indicate that the cell reenters the PFR while still being induced from a previous PFR passage, resulting in multiple transcription events (time dependency is shown in Supplementary Material).

Figure 5A compares measured and simulated transcript dynamics of the *trp* operon while passaging through the PFR. Notably, measured transcript dynamics were very similar so that measurements taken after 25, 120 min and 28 h were cumulated and indicated by a common variance. According to the modeling constraints, mRNA production started after 30 s which is in good agreement with experimental observations for *trpL* and for *trpE*. Synthesis of further downstream genes *trpDCBA* was neither predicted nor measured.

The long-term adaptation of the population was simulated for the exemplary time points of 25 and 120 min as well as for 28 h (see **Figure 5B**). Again, experiments and simulation results show a high agreement for all conditions. This also holds true for the short *trpL* mRNA which was hardly detected in the PFR, confirming the simulation.

To compare the dynamics of transcript and protein adaptation toward new steady-states, both species were simulated. For transcript studies, the average *trpA* transcription was considered. Protein formation of TrpA disregarded putative degradation and simply considered continuous cell drain under steady-state

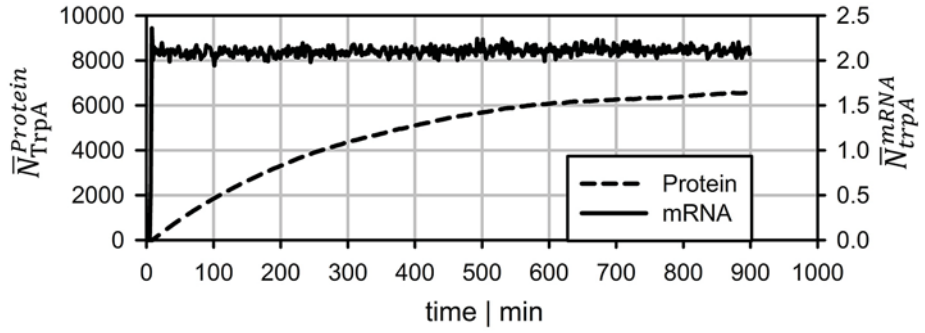


FIGURE 6 | Long-term prediction of transcript and protein levels neglecting protein degradation. mRNA levels reach stable levels after ∼7 min whereas protein leveling takes more than 15 h.

conditions. **Figure 6** clearly outlines the different speeds. Whereas, transcript levels converge to a new steady-state within 10 min (slightly more than a PFR-STR cycle), proteins need about 15 h.

DISCUSSION

The transcriptional dynamics observed in the STR-PFR experiments of Löffler et al. (2016) and in particular Simen et al. (2017) were modeled using a combined cell and process model. By focusing on details of transcription and translation in the cell model, a set of 10,000 individual cells was created and tracked during their repeated passages through the STR-PFR system. Inherently, the modeling approach mirrors a mechanistic understanding linking external stimuli with cellular transcriptional responses thereby excluding putative stochastic events (Elowitz et al., 2002; Avery, 2006). Accordingly, the modeling approach showed fundamental characteristics of an ensemble model, as outlined by (Henson, 2003). Here, we used the *trp* operon as an example because its polycistronic mRNA consisting of five structural genes and a leader peptide was repeatedly transcribed envisaging ammonia limitation (Simen et al., 2017) and, most importantly, its induction was followed by attenuation which directly linked transcription and translation of the gene products. Only by using the approach of ensemble modeling, individual cell fates could be tracked which finally explain the occurrence of population heterogeneity.

Comparing the experimental observations of transcript dynamics with modeling predictions, high agreement between simulations and experimental data can be observed (**Figure 5**). The qualification holds true not only for the prediction of short-term transcript dynamics in the PFR but also for the long-term adaptations in the STR, visualized by analyzing samples up to 28 h after initial connection of the PFR with the STR. Notably, the high precision of transcript prediction was achieved without any parameter regression. Only literature documented parameters were chosen to fix the setting of the ensemble model. Again, this finding is qualified as a confirmation of the basic approach.

Protein formation was assumed to start immediately after mRNA transcription. Unlike mRNA degradation, no distinct decay kinetics for the *trp* genes were known. Simulation studies of **Figure 4** revealed that realistic protein half-lives should be about 1.1 h, which is in the range of experimental observations for other proteins (Nath and Koch, 1970; Lahtvee et al., 2014). Accordingly, impacts of protein degradation on short-term kinetics can be ruled out. However, the long-term adaptations indicated in **Figure 6** are likely to be affected. The additional consideration of decay kinetics will likewise reduce steady-state levels.

One of the key findings of the STR-PFR studies of Löffler et al. (2016) and Simen et al. (2017) was the observation that PFR induced regulatory responses are propagated into STR finally causing the adaptation of the whole population. Exactly this phenomenon could be modeled as well. **Figures 3**, **5** document that only *trpL* and *trpE* are fully transcribed in PFR whereas the transcription of the rest of the operon continued in the STR. Subsequently, most of the stress induced cellular burden occurred after a time-delay in the well-mixed STR compartment and not immediately in the PFR, the origin of the trigger. As a consequence, the population in the STR is very heterogeneous, consisting of cells in different transcription and translation states. Some cells should be still propagating the PFR induced stress response, whereas others may have completed the same. Moreover, given that the STR and the PFR compartments do not physically exist in large-scale bioreactors, cells are expected to co-exist next to each other while circulating around. Similar studies have already be performed investigating the lifelines of fluctuating cells (Haringa et al., 2016; Kuschel et al., 2017).

CONCLUSION

The ensemble model used in this study succeeded to predict experimental observations of long- and short-term transcriptional dynamics with high precision and without parameter adjustments. As such, the approach demonstrated its fundamental suitability for predicting large-scale population heterogeneity as a consequence of local stress triggers. Accordingly, likewise modeling approaches open the door for an *in silico* scale-up design, simulating large-scale performance of the cells *ab initio*.

This study illustrates that locally induced stress responses are propagated into different regions of the bioreactor thereby creating temporal and spatial inhomogeneity of the population. Notably, cellular reactions do happen on different time

scales: Whereas transcriptional responses require <10 min, translational changes may continue for more than 10 h to reach new steady-states. Additionally, metabolic responses may occur which are likely to precede the transcriptional reaction. The hierarchical sequence of regulatory responses is overlaid with dynamics of mass transfer, mixing and process control which make it necessary to track individual cell responses properly for predicting large-scale performance of the total culture.

AUTHOR CONTRIBUTIONS

AN performed the modeling, designed the study, and prepared the manuscript. ML and JS performed the experiments and prepared the manuscript. RT designed the study and prepared the manuscript.

FUNDING

The authors gratefully acknowledge the funding by the Bundesministerium für Bildung und Forschung (BMBF, Grant FKZ0316178A).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01195/full#supplementary-material

REFERENCES

Avery, S. V. (2006). Microbial cell individuality and the underlying sources of heterogeneity. *Nat. Rev. Microbiol.* 4, 577–587. doi: 10.1038/nrmicro1460

Bremer, H., and Dennis, P. P. (1987). "Modulation of chemical composition and other parameters of the cell by growth rate," in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology, Vol. 2*, ed F. C. Neidhardt (Washington, DC: American Society for Microbiology), 1553–1569.

Chen, H., Shiroguchi, K., Ge, H., and Xie, X. S. (2015). Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol. Syst. Biol.* 11:808. doi: 10.15252/msb.20159000

Chubukov, V., Gerosa, L., Kochanowski, K., and Sauer, U. (2014). Coordination of microbial metabolism. *Nat. Rev. Microbiol.* 12, 327–340. doi: 10.1038/nrmicro3238

Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183–1186. doi: 10.1126/science.1070919

George, S., Larsson, G., and Enfors, S.-O. (1993). A scale-down two-compartment reactor with controlled substrate oscillations: metabolic response of *Saccharomyces cerevisiae*. *Bioprocess Eng.* 9, 249–257. doi: 10.1007/BF01061530

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361. doi: 10.1021/j100540a008

Haringa, C., Tang, W., Deshmukh, A. T., Xia, J., Reuss, M., Heijnen, J. J., et al. (2016). Euler-Lagrange computational fluid dynamics for (bio)reactor scale-down: an analysis of organism life-lines. *Eng. Life Sci.* 16, 652–663. doi: 10.1002/elsc.201600061

Henson, M. A. (2003). Dynamic modeling of microbial cell populations. *Curr. Opin. Biotechnol.* 14, 460–467. doi: 10.1016/S0958-1669(03)00104-6

Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W., and Davies, D. R. (1988). Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*. *J. Biol. Chem.* 263, 17857–17871.

Kuschel, M., Siebler, F., and Takors, R. (2017). Lagrangian trajectories to predict the formation of population heterogeneity in large-scale bioreactors. *Bioengineering* 4:27. doi: 10.3390/bioengineering4020027

Lahtvee, P.-J., Seiman, A., Arike, L., Adamberg, K., and Vilu, R. (2014). Protein turnover forms one of the highest maintenance costs in *Lactococcus lactis*. *Microbiology* 160, 1501–1512. doi: 10.1099/mic.0.078089-0

Lara, A. R., Galindo, E., Ramírez, O. T., and Palomares, L. A. (2006). Living with heterogeneities in bioreactors: understanding the effects of environmental gradients on cells. *Mol. Biotechnol.* 34, 355–382. doi: 10.1385/MB:34:3:355

Löffler, M., Simen, J. D., Jager, G., Schaferhoff, K., Freund, A., and Takors, R. (2016). Engineering, *E. coli* for large-scale production - Strategies considering ATP expenses and transcriptional responses. *Metab. Eng.* 38, 73–85. doi: 10.1016/j.ymben.2016.06.008

Löffler, M., Simen, J. D., Müller, J., Jager, G., Laghrami, S., Schaferhoff, K., et al. (2017). Switching between nitrogen and glucose limitation: unraveling transcriptional dynamics in *Escherichia coli*. *J. Biotechnol.* doi: 10.1016/j.jbiotec.2017.04.011. [Epub ahead of print].

Michalowski, A., Siemann-Herzberg, M., and Takors, R. (2017). *Escherichia coli* HGT: engineered for high glucose throughput even under slowly growing or resting conditions. *Metab. Eng.* 40, 93–103. doi: 10.1016/j.ymben.2017.01.005

Nath, K., and Koch, A. L. (1970). Protein degradation in *Escherichia coli*. I. Measurement of rapidly and slowly decaying components. *J. Biol. Chem.* 245, 2889–2900.

Neubauer, P., and Junne, S. (2010). Scale-down simulators for metabolic analysis of large-scale bioprocesses. *Curr. Opin. Biotechnol.* 21, 114–121. doi: 10.1016/j.copbio.2010.02.001

Neubauer, P., Haggstrom, L., and Enfors, S. O. (1995). Influence of substrate oscillations on acetate formation and growth yield in *Escherichia coli* glucose limited fed-batch cultivations. *Biotechnol. Bioeng.* 47, 139–146. doi: 10.1002/bit.260470204

Nienow, A. W., Edwards, M. F., and Harnby, N. (eds.). (1997). *Mixing in the Process Industries, 2nd Edn*. Oxford; Boston, MA: Butterworth-Heinemann.

Oldiges, M., and Takors, R. (2005). Applying metabolic profiling techniques for stimulus-response experiments: chances and pitfalls. *Adv. Biochem. Eng. Biotechnol.* 92, 173–196. doi: 10.1007/b98913

Simen, J. D., Löffler, M., Jäger, G., Schäferhoff, K., Freund, A., Müller, J., et al. (2017). Transcriptional response of *Escherichia coli* to ammonia and glucose fluctuations. *Microb. Biotechnol.* doi: 10.1111/1751-7915.12713. [Epub ahead of print].

Stoltzfus, A., Leslie, J. F., and Milkman, R. (1988). Molecular evolution of the *Escherichia coli* chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between trp and tonB. *Genetics* 120, 345–358.

Takors, R. (2012). Scale-up of microbial processes: impacts, tools and open questions. *J. Biotechnol.* 160, 3–9. doi: 10.1016/j.jbiotec.2011.12.010

Valgepea, K., Adamberg, K., Seiman, A., and Vilu, R. (2013). *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Mol. Biosyst.* 9, 2344–2358. doi: 10.1039/c3mb70119k

Yanofsky, C. (2004). The different roles of tryptophan transfer RNA in regulating trp operon expression in *E. coli* versus B. subtilis. *Trends Genet.* 20, 367–374. doi: 10.1016/j.tig.2004.06.007

Yanofsky, C. (2007). RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. *RNA* 13, 1141–1154. doi: 10.1261/rna.620507

Yanofsky, C., Platt, T., Crawford, I. P., Nichols, B. P., Christie, G. E., Horowitz, H., et al. (1981). The complete nucleotide sequence of the tryptophan operon of *Escherichia coli*. *Nucleic Acids Res.* 9, 6647–6668. doi: 10.1093/nar/9.24.6647

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.