Experiment and quantitative modeling of cell-free gene expression dynamics

Tobias Roland Stögbauer

Dissertation

an der Fakultät für Physik

der Ludwig–Maximilians–Universität

München

vorgelegt von

Tobias Roland Stögbauer

aus Bad Griesbach

München, den 05.10.2012

Erstgutachter: Prof. Dr. Joachim O. Rädler

Zweitgutachter: Prof. Dr. Hermann Gaub

Tag der mündlichen Prüfung: 05.12.2012

Contents

Zusammenfassung

Genexpression, d. h. die zelluläre Synthese von Proteinen, setzt sich aus den Teilschritten Transkription (mRNA Synthese auf Grundlage der DNS), Translation (Proteinsynthese auf Grundlage der mRNA) und Proteinfaltung zusammen. Aufgrund der großen Zahl an Wechselwirkungen einzelner Komponenten ist dieser Prozess *in vivo* sehr komplex, was eine quantitative mathematische Modellierung extrem aufwendig macht. Indem auf vereinfachte *in vitro* Modellsysteme zurückgegriffen wird, können Teilaspekte der zellulären Genexpression im Detail studiert werden. Mit "zellfreier Genexpression" wird die biochemische Synthese von Proteinen *in vitro* bezeichnet. Die dabei eingesetzten zellfreien Systeme bestehen aus den Hauptbestandteilen der zellulären Transkriptions- und Translationsmaschinerie wie Polymerasen, Ribosomen, Aminosäuren und Nukleotiden, gelöst in Puffer. Diese Systeme werden entweder aus Zellextrakt gewonnen oder aus aufgereinigten Einzelkomponenten komplett zusammengesetzt.

Im dieser Arbeit wurde Genexpression in einem zellfreien System genutzt, um ein quantitatives Modellsystem der Genexpressionskinetik *in vitro* zu entwickeln. Zu diesem Zweck wurden Techniken erarbeitet, die reproduzierbare quantitative Fluoreszenzmessungen der zellfrei produzierten mRNA- und Proteinmenge erlauben. Hier wurde das grün fluoreszierende Protein GFP exprimiert, da es sowohl in prokaryotischen als auch in eukaryotischen Wirten produziert werden kann und durch seine Fluoreszenz nachweisbar ist. Für den Nachweis der mRNA kamen spezifische molekulare Fluoreszenzmarker (*molecular beacons*) zum Einsatz. Basierend auf Messreihen der Kinetik von mRNA und GFP Synthese wurde ein mathematisches Modell gewöhnlicher Differentialgleichungen entwickelt. Das Modell berücksichtigt die Sättigung der zellfreien Genexpressionsmaschinerie bei hohen DNS Konzentrationen sowie Verzehr und unspezifischen Abbau von Ressourcen im Laufe der Synthesereaktion. Dadurch stimmen Daten und Modell bei simultaner Datenanpassung mit einem einzigen Satz von Parametern über fünf Größenordnungen der DNS Konzentration und über den gesamten Verlauf der Kinetik sehr gut überein. Dank seiner

Vorhersagekraft ist das hier aufgestellte Modellsystem ein nützliches Werkzeug für Studien zur Transkriptions- und Translationsmechanik, zu Regulationsnetzwerken und Modifikationen einzelner Komponenten.

Im nächsten Schritt wurde untersucht, in wie weit der vorgestellte Ansatz auf ein System anwendbar ist, das einer natürlichen Zelle stärker ähnelt, nämlich mRNA bzw. DNS transfizierten Zellen. Dabei zeigte sich, dass das Modell die Kinetik DNS transfizierter Zellen nur dann zufriedenstellend wiedergeben kann, wenn die Modellparameter sehr stark von Literaturwerten abweichen. Andererseits beschreibt das einfache Ratenmodell die Kinetik mRNA transfizierter Zellen exakt. Das könnte ein Hinweis darauf sein, dass der Transportprozess der synthetisierten mRNA vom Zellkern in das Zytoplasma einen bisher unbekannten Einfluss auf den Verlauf der Genexpression hat. Mit Hilfe dieser Modellierung konnten die Degradationsraten von mRNA und GFP sehr exakt ermittelt werden. Des Weiteren wurde ein stochastisches Modell adaptiert, mit welchem die Aufnahme der mRNA durch eine Zelle wiedergegeben werden konnte, was eine Abschätzung der mittleren Anzahl an mRNA Molekülen pro transfizierter Zelle erlaubt.

Zellfreie Systeme eignen sich bestens, um die Auswirkung von modifizierten Komponenten auf die Expressionskinetik zu analysieren. Beispielsweise ist bekannt, dass D_2O das Zellwachstum hemmt, seine genaue Wirkung auf die Proteinsynthese ist aber noch nicht vollständig analysiert. D_2O ist besonders dadurch interessant, dass es als kleine globale Störung aller Teilschritte der Proteinsynthese betrachtet werden kann. Es zeigte sich, dass Transkription, Translation und Proteinfaltung von D_2O beeinflusst werden, allerdings in unterschiedlichem Maße. Vergleichsmessungen mit *E. coli* zeigten, dass das künstliche System stärker auf D_2O reagiert, als die wesentlich komplexeren lebenden Organismen. Diese Erkenntnisse könnten ferner dabei helfen, den Ertrag von Genexpression in D_2O zu optimieren, was für die zellfreie Herstellung deuterierter Proteine für die Kernspinresonanzspektroskopie bedeutsam ist.

Ein wichtiges Ziel der synthetischen Biologie ist die Konstruktion einer künstlichen minimalen Zelle. Eine Herangehensweise ist, zellfreies Extrakt in Lipidvesikeln einzuschließen. Besonders bedeutsam ist hierbei, Transmembranproteine zu exprimieren und korrekt gefaltet in die künstliche Membran einzubauen. Dies ermöglicht z. B. den Nährstoffaustausch mit der Umgebung, oder erlaubt, definierte Bindungsstellen zu schaffen. Im Rahmen dieser Arbeit wurde das Transmembranprotein Ephrin-B2 zellfrei exprimiert und *in situ* in Vesikel eingebunden. Diese Methode kann ferner dazu genutzt werden, auf einem mikrostrukturierten Substrat definierte Bindungsstellen für Zellen zu schaffen.

Summary

Genexpression that is the cellular synthesis of proteins is comprised of the sub-steps transcription (mRNA synthesis based on the DNA master), translation (protein synthesis based on the mRNA) and protein folding. Owing to the large number of interactions between individual components this process is very complex *in vivo* and therefore mathematical modeling is extremely laborious. By means of simplified *in vitro* model systems individual aspects of cellular gene expression can be studied in detail. Cell-free gene expression denotes the biochemical synthesis of proteins *in vitro*. Cell-free systems are comprised of the predominant components of the cellular transcription and translation machinery like polymerases, ribosomes, amino acids and nucleotides dissolved in buffer. These systems are either cellular extracts or are reconstituted from purified components.

In this thesis gene expression in a cell-free system was used to develop a quantitative model system of the gene expression kinetics *in vitro*. To this end techniques were developed that enable reproducible quantitative fluorescence measurements of mRNA and protein synthesized in a cell-free system. Here, the green fluorescent protein GFP was used since it can be expressed both in prokaryotic and eukaryotic hosts and it can be detected by its fluorescence. In order to measure mRNA levels specific molecular fluorescence markers (*molecular beacons*) were used. Based on measurements of the kinetics of mRNA and GFP synthesis a mathematical model of ordinary differential equations was developed. The model accounts for the saturation of the cell-free gene expression machinery at high DNA concentrations as well as for the unspecific degradation of resources in the course of the synthesis reaction. Thereby the model can concurrently fit the kinetics of mRNA and GFP expression with one consistent set of parameters over five orders of magnitude of DNA concentration. Due to its predictive power the model described here can be a valuable tool for studies of transcription and translation mechanics, gene regulation networks and modifications of individual components.

In the next step it was reassessed how well the modeling ansatz described here can be

applied to a system that resembles a living cell more closely, that is to say mRNA and pDNA transfected cells respectively. It was found that the model can satisfactorily fit the kinetics of pDNA transfected cell only if model parameters far away from literature values are used. On the other hand the simple model fits the kinetics of mRNA transfected cells very well. This could indicate that the transport process of synthesized mRNA into the cellular nucleus has an influence on the course of the protein synthesis that is so far not exactly understood. By means of this model the degradation rates of mRNA and GFP could be determined with great accuracy. Furthermore, a stochastic mRNA uptake model was adapted which satisfactorily fits the transfection efficiency of mRNA vectors and allows for an estimation of the number of mRNA molecules per transfected cell.

Cell-free systems are well suited for studying modifications of the gene expression machinery. For example, D_2O impedes cellular growth and proliferation but its effect on protein synthesis has not been ascertained in detail. D_2O is furthermore interesting to study because it can be viewed as a small global disturbance of all partial steps of gene expression. It was found that all transcription, translation and protein folding are affected by D_2O albeit to varying grades. Comparison measurements with *E. coli* showed that the artificial system reacts more strongly to D_2O than the far more complex living organism. These insights could further help to optimize the yield of protein expressed in D_2O which is important for the production of deuterated proteins for nuclear magnetic resonance spectroscopy.

A major goal of synthetic biology is the development of an artificial minimal cell. One approach to this is to encapsulate cell-free extract in lipid vesicles. In this regard it is important to express transmembrane proteins and insert them into the artificial membrane in their correctly folded state. This would for example allow for nutrient uptake or allow to create defined binding sites. In this thesis the transmembrane protein ephrin-B2 was expressed in a cell-free system and incorporated into vesicle membranes *in situ*. One future application of this setup is the construction of defined cellular binding sites on a microstructured substrate.

Chapter 1

Introduction

The 20th century brought an enormous increase of knowledge in all fields of science and technology. This swift progress stems from an interplay between improved measurement and analysis technologies and successive new scientific insights. Medicine and biology are two fields that have especially profited from said progress. Over the course of only a few decades the composition and functioning of living organisms were uncovered down to the intra-cellular level. With the uncovering of more and more functions growing interest arose about how all these functions are linked together. In other words one wants to know all interactions of the components of an organism and how the totality of the individual interactions gives rise to the behavior of the complete organism. This interdisciplinary field of study has been termed "systems biology". Two of the pioneers in this field were the later Nobel Prize laureates Alan Lloyd Hodgkin and Andrew Fielding Huxley who published a much-noticed mathematical model of neurons in 1952 [1]. However, systems biology as a prominent field really took off not until the turn of the 21st century. At this time large amounts of data on genomics and proteomics became available and the increased computing power and broad adoption of the Internet allowed to exploit these data. However, the goal to fully understand the regulation network of a single cell – let alone a whole organism – is complicated by the immense complexity of the interactions. In this connection the complementary engineering based approach of "synthetic biology" may help to further our understanding. Synthetic biology just like systems biology is a recent interdisciplinary field of study. It pursues the creation of artificial biological functions not found in nature. Synthetic biology aims to reverse engineer natural systems, to build new systems that do not already exist and to equip existing organisms with new functionalities. Artificial analogies of organisms with their reduced complexity can also be used as model

systems to test and refine systems biology models. This approach is often highlighted with Richard Feynman's famous quote "What I cannot create, I do not understand".

A prominent goal of synthetic biology efforts is to create a minimal cell, that is a construct which mimics natural cellular behavior up to self replication. To this end model systems for the individual components that make up a living cell like molecular motors, cellular membranes, encapsulation techniques and dynamic pattern formation are being investigated [2, 3]. A further cellular component which is intensely studied is gene regulation. Elowitz and Leibler constructed a synthetic transcription regulation network termed the "repressilator" that exhibits mutual negative regulation. Inserted into *E. coli* the system oscillated with periods longer than the cell cycle, which means the cellular state was transmitted to the next generation [4]. The authors also observed noisy behavior which can be attributed to stochasticity of the system's components. In this pioneering work the synthetic system was still incorporated into a natural host. The next logical step is to create a completely artificial framework and this can be achieved by using so called cell-free systems.

Cell-free systems denote mixtures of resources and enzymes that can be used to synthesize protein from a DNA template *in vitro*. Although these systems were originally developed for (compared to cell cultures) quick and easy to handle protein synthesis they have become a prominent tool in synthetic biology applications. Noireaux et al. developed a genetic circuit exhibiting transcriptional activation and repression cascades in a cell-free system [5]. The protein produced in one stage of their circuit serves as activator or repressor of the next stage.

In order to bring this approach closer to the goal of creating a minimal cell it is necessary to enclose the cell-free system with an artificial barrier that mimics the cellular membrane. Vesicles made of artificial phospholipid bilayers are a promising model system since phospholipids are the main ingredient of cellular membranes (the other is membrane proteins). Preparation techniques exploit the self-organization property of the amphiphilic molecules into bilayer vesicles in aqueous solutions [6]. However, these techniques are usually time consuming and may not be compatible with even medium level ion concentrations. This is problematic since it is desirable to capture also the early phase of gene expression which requires a fast loading mechanism of cell-free system solution into the vesicles. Furthermore cell-free extracts contain various species of ions. Noireaux and Libchaber developed a new vesicle formation technique that fulfills these requirements. Droplets of cell-free extract in oil form an emulsion with phospholipids forming a monolayer at the interface. These

droplets are then spun down into buffer and lipids at the oil-buffer interface interact with the monolayer around the droplets to form a bilayer. The authors then observe synthesis of green fluorescent protein (GFP) inside the vesicles. They brought their concept even closer to the goal of a minimal cell by expressing a second protein, namely the transmembrane protein α-hemolysin in their vesicles. They showed that α-hemolysin successfully incorporated into the vesicle membrane and formed pores through which nutrients could pass into the vesicle. This way the life-time of the gene expression reaction in the vesicles was increased up to four days [7, 8]. Membrane proteins are challenging to synthesize as they tend to aggregate in aqueous solutions and therefore require the application of detergents. However, as mentioned above cell-free synthesized membrane proteins can incorporate into artificial lipid membranes *in situ*. Apart from α-hemolysin this has also been demonstrated for a number of other membrane proteins [9–12]. Furthermore, Dittrich et al. developed a microfluidics setup which creates an emulsion of micro-droplets of cell-free system in oil [13]. These droplets are very stable, monodisperse and can be produced with a high throughput of up to 30 droplets per second.

A different approach is to fabricate micro compartments directly in a microfluidics chip [14–17]. In combination with a sophisticated system of channels and valves such microarrays can be used for example for high throughput proteomics studies [18, 19].

Cell-free systems are usually a mixture of cell lysate and some added components and therefore their composition is not entirely known. However, recently a cell-free system was presented that is comprised solely of purified components [20, 21]. Such a reconstituted system guarantees that no interference due to unknown and unwanted components will distort the outcome of an experiment.

The aim of this thesis was to establish a cell-free system as a model system for studying gene expression *in vitro*. In order to do so experimental and theoretical tools were developed that permit quantitative conclusions from experimental data. The results of the *in vitro* model were compared to gene expression kinetics in living cells. Subsequently the established knowledge of the particular cell-free system was used to examine how exactly an agent (here deuterated water) modifies the individual kinetic rates. Lastly an extension of the experimental procedure is presented with which a membrane protein could be shown to be synthesized and incorporated into artificial lipid membranes.

This thesis is arranged as follows: in chapter 2 basic theoretical concepts are presented. A short description of the individual steps and necessary components of the gene expression mechanism is given. This is followed by an overview over the history, composition and

different experimental setups of cell-free gene expression systems. Subsequently GFP which was used as fluorescent probe is briefly introduced. Furthermore the differences in physical properties of deuterated water in comparison to normal water are presented.

In chapter 3 the experimental protocols for sample preparation, measurement procedure and data analysis are explained in detail. Fluorescent microscopy and fluorescence correlation spectroscopy are briefly presented. Special focus is given to the description of the developed calibration standard for quantitative GFP fluorescence measurements.

In chapter 4 a mathematic model of cell-free gene expression kinetics based on differential equations for individual kinetics rates is presented. A Michaelis-Menten like ansatz was incorporated into the model to account for different template DNA concentrations. Furthermore, the model assumes finite pools of resources for transcription and translation to account for the observed cessation of *in vitro* protein synthesis after a fixed time. This way cell-free gene expression kinetics could be fitted from the early to the late phase and this over five orders of magnitude of DNA concentration. It was found that the translation step is the bottle neck in GFP synthesis and that already terminated cell-free expression could be restarted by introducing fresh ribosomes into the system. These findings could help to improve yield strategies in cell-free protein synthesis. This properly characterized cell-free system in combination with the predictive model presented here is a valuable tool for studies on transcription and translation kinetics or on *in vitro* gene regulation networks.

In chapter 5 the kinetic rate based modeling approach is applied to a system that is more closely alike a living cell: DNA and mRNA transfected cells respectively. The kinetics of mRNA but not pDNA transfected cells could be rigorously modeled with this basic approach. Distributions for the mRNA and GFP degradation rates as well as number of mRNA molecules taken up by a cell were derived. Assuming that mRNA is taken up and released in statistically independent quanta (the lipoplexes used for transfection), the average number of mRNAs per lipoplex could be estimated. Interestingly, kinetics of DNA transfected cells could not be fitted adequately, probably hinting at some cooperativity effect *in vivo* that is so far not understood. This work demonstrates that mRNA transfection in combination with computer modeling can be a powerful tool for time-resolved gene expression studies for example of siRNA knockdown.

In chapter 6 a cell-free system is used to study how the influence of deuterated water on gene expression also observed *in vivo* affects the individual rates: transcription, translation and maturation. It was found that D_2O affects individual polymerase differently. With a prokaryotic DNA polymerase an inhibitory effect was found whereas with a viral RNA

polymerase D$_2$O was observed to amplify the synthesis efficiency. Interestingly, the D$_2$O effect was less pronounced *in vivo* although gene expression in a living organism is much more complex. The results presented could help to improve yield strategies for the cell-free synthesis of deuterated proteins which is necessary for nuclear magnetic resonance studies of protein structures.

Chapter 7 covers the cell-free synthesis of a transmembrane protein (ephrin-B2) and its *in situ* incorporation into a model membrane. Furthermore, it is shown how microstructured substrates can be used to induce alignment and deformation of coexisting lipid phases in substrate adhering artificial membranes. With membrane proteins that exhibit a preference for one lipid phase such a system could be used to create a substrate with defined cellular binding sites for automated microscopy studies.

Finally chapter 8 gives an outlook to possible future applications of this thesis' findings as well as new questions that have arisen and need to be answered.

Chapter 2

Basic concepts

2.1 Cell-free gene expression

2.1.1 The course of gene expression

Living cells come in many different shapes and sizes depending on the respective organism and the cell's individual task. However, even the most simple cell is made up of a highly complex composition of ingredients like DNA, RNAs, energy sources, ions, proteins or organelles (Fig. 2.1 A). This composition has to be constantly monitored and adapted by feedback mechanisms so that the cell stays functional. This is especially true for the composition of the proteins. In order to synthesize new proteins or replenish degraded ones the cell accesses its "long term memory", namely its DNA. The demand for the production of a certain type of protein produces a signal which activates the gene encoding this protein. Depending on the specific regulatory motif of the gene there are different ways how this activation step works (for a review on transcription regulation see [22, 23]). The next steps are what Francis Crick, one of the co-discoverers of the double helix structure of the DNA, called the "central dogma of molecular biology" [24]: DNA is transcribed into RNA which in turn is translated into a protein. This is a one-way process, information is not communicated from proteins to DNA (Fig. 2.1 B).

Transcription The process of transcribing the genetic information carried by the DNA into RNA is basically the same for prokaryotic and eukaryotic organisms.[1] Activation of

---

[1]Prokaryotes are bacteria and archaea. They have no cellular nucleus, their DNA swims in the cytoplasm and takes the form of a ring. Eukaryotes (some unicellular organisms and almost all known multi-cellular organisms) have a nucleus which houses their DNA. Their DNA takes the form of strongly condensed
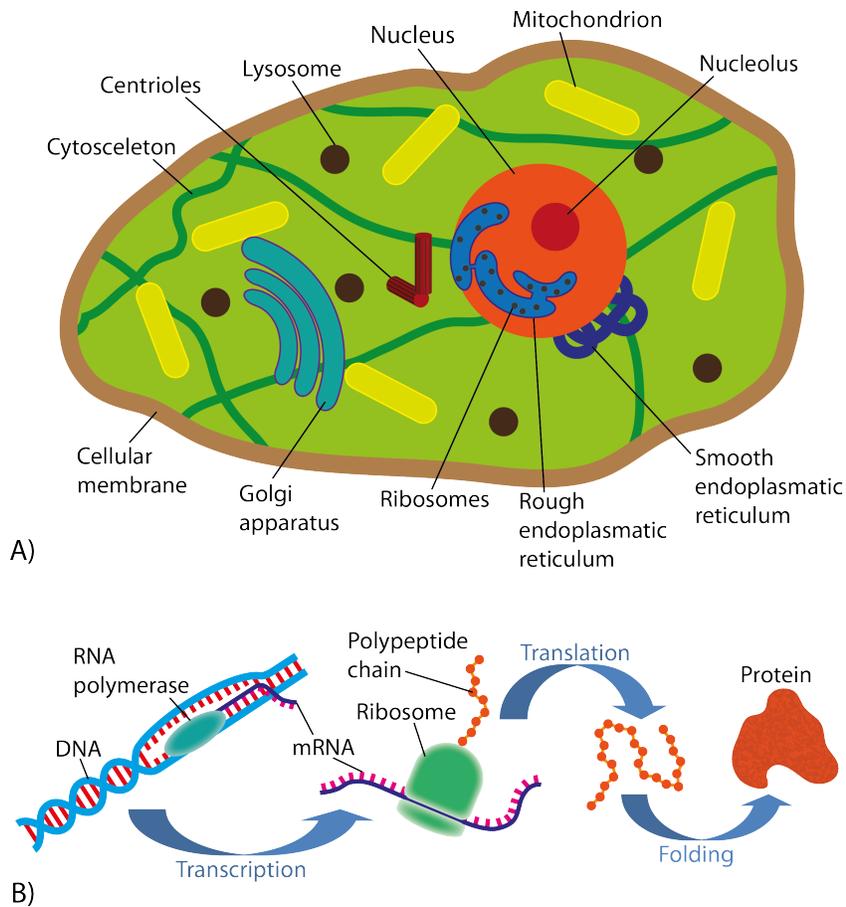
Figure 2.1: A) Some of the major components that make up an eukaryotic cell. B) The "central dogma of molecular biology": DNA is transcribed into mRNA by RNA polymerase and mRNA is translated into a polypeptide chain by a ribosome. The polypeptide then folds into a functional ribosome. This picture is greatly simplified, only the core components are shown. In nature a great variety of additional components is necessary in order to synthesize the multitude of different proteins that are contained in a cell.

a gene means that the enzyme RNA-polymerase can now bind to a specific binding site at the start of the gene, the promoter, and begin reading it. In doing so the polymerase separates the DNA double helix by breaking the hydrogen bonds between the strands and adds complementary RNA nucleotides to the coding strand. Then the sugar-phosphate backbone of the RNA is formed, the DNA-RNA hydrogen bonds are broken and the RNA separates from the DNA. At the end of the gene, the terminator, the polymerase is released from the DNA. RNA differs from DNA in the fact that it is single stranded, contains ribose instead of deoxyribose in its backbone and the nucleotide thymine is replaced by uracil. A major difference between eukaryotes and prokaryotes is that in eukaryotes transcription takes place in the nucleus (which prokaryotes lack) and the RNA is then transported into the cytoplasm for further use. RNA polymerases achieve typical transcription speeds of 20-40 nucleotides per second [25].

Translation Possible products of transcription are different types of RNA, the most prominent of which are messenger RNA (mRNA) that encodes a protein, ribosomal RNA (rRNA) which is a part of the ribosomes and transfer RNA (tRNA) which acts as carrier for single amino acids. Translation is a much more complex process than transcription since a sequence of four different nucleotides now has to be used to assemble a sequence of up to 22 different amino acids of which the protein is composed of. In this connection is the genetic code divided into groups of three nucleotides, the codons, which encode an amino acid each (some amino acids are encoded by more than one codon). Some codons (UAA, UAG, UGA) do not encode an amino acid but are the stop signal of the translation process. The translation process is catalyzed by ribosomes, complex cellular machines composed of proteins and rRNAs. Translation works as follows: ribosomes consist of two separate subunits which unite on the mRNA. The now functional ribosome provides a binding site at which tRNAs recognize their specific codon on the mRNA and bind to it. This brings the amino acids that the tRNAs carry on their back into close contact. Subsequently a peptide bond is created between the carboxy-terminus of the already synthesized polypeptide chain and the amino group of the newly arrived amino acid. In this manner the polypeptide chain grows out of the ribosome until the stop codon on the mRNA is reached and the translation process terminates.

Although the ribosomes in prokaryotes and eukaryotes have the same function, the structure and composition of them is different. Phrased in Svedberg, the unit of sedimenta-

---

superstructures, the chromosomes.

tion in a centrifuge, prokaryotes and eukaryotes have $70\,S$ and $80\,S$ ribosomes respectively. Eukaryotic ribosomes translate about two amino acids per second, prokaryotic ribosomes even about 20 amino acids per second [25].

Protein folding   The product of the translation step is a polypeptide chain which is not yet a functional protein. Firstly, this chain has to fold into the correct three-dimensional configuration of the target protein. Due to its length and flexibility one can expect the polypeptide chain to be able to fold into a great number of configurations. This number however is severely limited by noncovalent bonds (hydrogen bonds, ionic bonds and van der Waals attractions) between different parts of the chain. The conformation that is ultimately adopted is the one with the maximum strength of these interactions as well as with optimal shielding of nonpolar amino acid side chains from the hydrous environment (hydrophobic effect) and therefore minimal free energy.[2]

Although the conformation of each protein species is unique there are two smaller scale patterns that can be found frequently – the so called α helix and β sheet (Fig. 2.2). Their ubiquity stems from the character of their emergence which is hydrogen bonds between N–H and C=O groups of the polypeptide backbone that do not involve particular amino acid side chains.[3]

The crowded interior of living cells makes proper folding of proteins difficult therefore special helper proteins called "molecular chaperons" assist the amino acid sequence in adopting the correct conformation [25].

Posttranslational modification   Some classes of proteins require additional modifications apart from correct folding in order to become fully functional [25, 29]. There are many different types of modifications. Some examples are oxidation, cofactor binding, disulfide bond formation, glycosylphosphatidylinositol (GPI) anchor formation, or addition of a polysaccharide group (glycosylation).

With respect to synthesizing a specific protein in a host that does not naturally express that protein (e. g. through plasmid DNA transfection) or in a cell-free system as described later on one has to make sure that either the target protein does not require any

---

[2]The postulate that a protein's native structure is defined by its amino acid sequence is known as "thermodynamic hypothesis" or "Anfinsen's dogma" after Nobel Laureate Christian Anfinsen [26, 27].

[3]In this context biologists defined four levels of organization of the protein structure: the amino acid sequence is called "primary structure", α helices and β sheets comprise the "secondary structure", the whole three-dimensional conformation is the "tertiary structure" and a complex comprised of more than one protein is termed "quarternary structure".
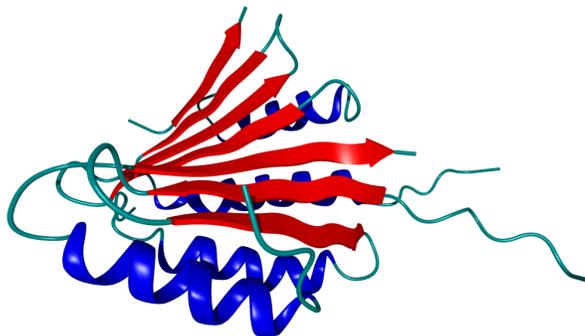
Figure 2.2: As an example for the crystal structure of a protein the ribbon diagram of SRP9/14 is shown. α helices are depicted in blue and β sheets are depicted in red (taken from [28]). The ribbon diagram representation has been established as a powerful tool for the visualization of the often times complex 3D configuration of proteins.

modifications specific to its natural host or provide such modifications artificially.

2.1.2 History of cell-free systems

Pioneering work on the mechanism of gene expression that would eventually lead to commercially available easy to use cell-free protein synthesis systems started in the 1950s when it was found that encapsulation in a cell is not necessary for the gene expression reaction [30]. Littlefield et al. showed RNA translation in a cell-free system of rat liver in 1955 [31] and other groups subsequently demonstrated protein synthesis in *E. coli* extracts [32–34].[4] However, protein that was synthesized in these early papers stemmed from native mRNA which was still included in the cell-free extract. Nirenberg and Matthaei were the first to achieve translation of an exogenous RNA in an *E. coli* cell-free system in 1961 [36]. This was further improved by removing endogenous DNA and RNA from the cell-free extract by nuclease treatment and by using the more stable DNA instead of RNA as the carrier of the genetic information [37, 38]. Besides *E. coli* extract cell-free systems based on rabbit reticulocyte lysate and wheat germ extract were also developed [39, 40]. These three systems are the basis for most commercially available *in vitro* protein synthesis kits nowadays. These kits do not contain endogenous polymerases but are loaded with highly promoter

---

[4]With these early publications one has to bear in mind that many cellular functions were still unknown. For example, only in 1959 had it been firmly established that ribosomes are a fundamental part of protein synthesis [35].

specific and fast working polymerases like the one of the T7 phage. This also means that a standardized promoter sequence on the DNA template can be used, in this case the T7 promoter.

Protein synthesis systems made up from cell extracts may exhibit several disadvantages like short reaction lifetime due to heavy waste of energy resources and sub optimum protein yield due to the presence of leftover nucleases and proteases. It was reasoned that a system reconstituted solely from purified components could circumvent these problems. Indeed, in 2001 Shimizu et al. presented such a system which they termed the PURE system ("Protein synthesis using recombinant elements") [20, 21]. A reconstituted system has the additional advantage that the target protein does not require a polyhistidine-tag for purification. Instead all the system's components are prepared with such a protein tag and the target protein can easily be purified by ultrafiltration and subsequent affinity chromatography [20].

2.1.3   Components of cell-free systems

A minimal cell-free system consists of about 100 individual components which can be broadly classified into four categories (see Tab. 2.1). Nucleoside triphosphates (NTPs) provide the necessary energy for the protein synthesis reaction and are also used as raw material for mRNA production. Translation requires the aid of initiation, elongation and release factors (IF, EF, RF) as well as ribosome recycling factor (RRF). Aminoacylation is the process of covalently binding an amino acid to the appropriate tRNA. An efficient energy regeneration system prevents premature cessation of protein synthesis (see also sec. 2.1.4).

The much lower complexity of transcription (solely NTPs and RNA polymerase) compared to the translation system explains why it is beneficial to add the additional transcription step to a cell-free system and use stable DNA instead of much more easily degradable mRNA as template.

Cell-free systems differ from cells not only in their reduced complexity of composition but the components are also much more diluted. Polymerase and ribosome concentrations are about one order of magnitude lower and the overall protein concentration is about two orders of magnitude lower (the interior of a cell must not be thought of as a fluid with widely distributed solid bodies but a tightly packed conglomeration of solid bodies with almost no space in between, see Fig. 2.3).

Many proteins require the help of chaperons or specific post-translational modifications

Cell-free system		*E. coli*
Transcription system:	DNA $(3\,nM)$	$5\,nM$
	RNA Polymerases $(100\,nM)$	$500 - 800\,nM$
	NTP mix $(1 - 2\,mM)$	$1.3 - 7.0\,mM$
Translation system:	Ribosomes $(2.4\,\mu M)$	$30\,\mu M$
	20 amino acids $(300\,\mu M)$	$1.5\,mM$
	Transl. factors IF, EF, RF, RRF	
	tRNA mix	
	Aminoacylation comp.	
Energy sources:	NTP mix $(1 - 2\,mM)$	$1.3 - 7.0\,mM$
	Regeneration system	
Other components:	Buffer	
	(Chaperons)	
	(Glycosylation comp.)	
	(Phosphorylation comp.)	

Table 2.1: Components of a minimal cell-free system; concentrations of some key elements in a cell-free system and in *E. coli* are give for comparison [21, 47].

like glycosylation or phosphorylation which are not provided by cell-free systems [41]. It has been found that the addition of chaperons helps *in vitro* synthesized proteins to fold correctly in some cases but fails in others [42–44]. Different strategies are being pursued with respect to cell-free glycosylation or phosphorylation, for example a cell-free system based on insect cell extract has been shown to provide core protein glycosylation enzymes [45]. Another suggested approach is to use non-natural amino acids that carry the desired modification [46] (for a review on this matter see [41]).

2.1.4   Energy supply in cell-free systems

Energy in the form of ATP is used up by the aminoacylation reaction and in the form of GTP energy is required to fuel the ribosomes [30]. However, it is not sufficient to simply provide an ample supply of these energy resources in a cell-free system. NTPs are quickly hydrolyzed by NTPases present in the system as well as by non beneficial reactions competing with gene expression [30, 49]. For the protein synthesis reaction to run over an extended period of time it is necessary to somehow maintain proper NTP levels. Two different approaches to this challenge have been developed: continuous-flow/continuous-exchange cell-free systems and the incorporation of an energy regeneration motif.

In continuous-flow systems the reaction reservoir containing the transcription/trans-
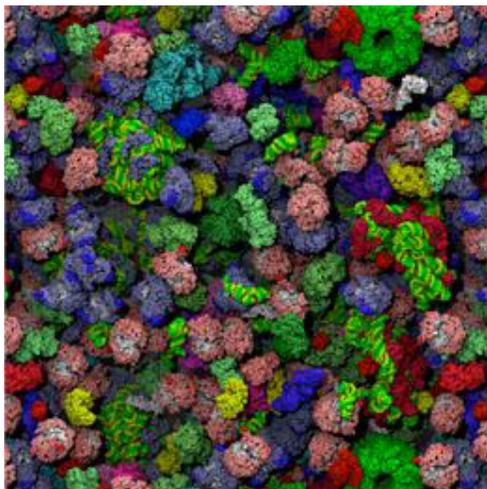
Figure 2.3: The interior of cells is very crowded with almost no free space in between individual components. Here a snapshot of a simulation of protein diffusion inside a cell is shown. Due to the crowded environment diffusion coefficients can be ten times lower than in solution (anomalous diffusion) (reprinted by permission from Macmillan Publishers Ltd: Nature News [48], copyright (2009)).

lation machinery (DNA template, polymerases, ribosomes etc.) is continuously fed with resources (NTPS, amino acids etc.). Conversely reaction products (inorganic phosphate, NMPs, synthesized protein) are continuously removed through an ultrafiltration membrane which holds back the components of the reaction machinery due to their higher molecular weight. This way *in vitro* protein synthesis can easily go on for $40\,h$ instead of only two to three hours like it does in batch format [50]. Continuous-exchange formats use a similar but somewhat simpler approach where the feeding solution is not pumped actively but diffuses into the reaction volume through a dialysis membrane [41] (see Fig. 2.4 A). Both approaches suffer from two drawbacks: the setup is more complex than a simple batch format cell-free system and they cannot readily be used in micro-volume, high-throughput format, an important field of application for cell-free protein synthesis systems.

An "energy regeneration system" means that ADP and GDP that accumulate in the protein synthesis reaction are converted back to ATP and GTP. There are a number of possibilities to achieve this which differ in the efficiency of conversion and in the cost of the required components – an important factor in the success of individual commercial protein synthesis systems. For example, phosphoenol pyruvate and pyruvate kinase or creatine phosphate and creatine kinase can be included in a cell-free system to set up an energy regeneration motif, see Fig. 2.4 B and [51, 52].
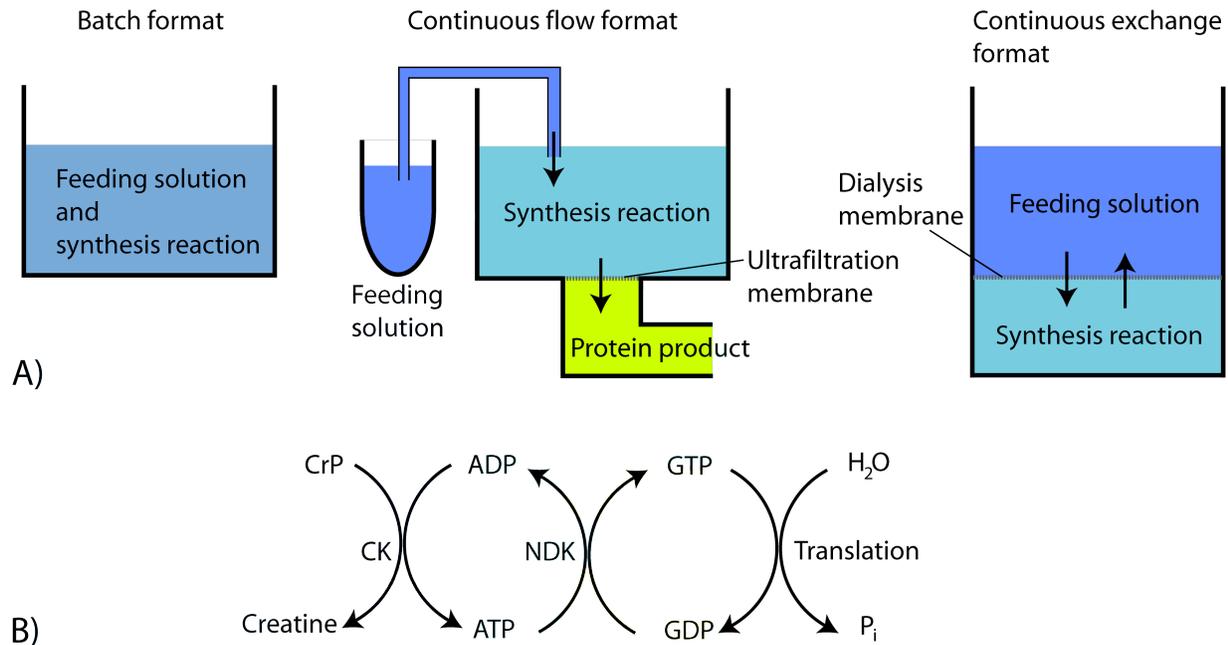
Figure 2.4: A) Different modes of operation have been developed for cell-free systems. The batch format is the easiest to set up and operate. However, depletion of resources and/or accumulation of waste will lead to the cessation of protein synthesis after about $3\,h$. The continuous flow format was developed in order to counter this problem. Buffer containing the necessary resources is continuously fed into the reaction volume while protein product and reaction waste is pumped down through an ultrafiltration membrane. The complexity of this setup however makes it impractical for regular laboratory use. Therefore a third format was developed. In continuous exchange cell-free systems the feeding buffer and the reaction volume are separated by a dialysis membrane. Reaction resources as well as by-products can diffuse through this membrane [41]. B) A typical energy regeneration scheme incorporated in cell-free systems. Creatine phosphate (CrP) and ADP are converted to creatine and ATP by creatine kinase (CK). Conversely ATP is used by nucleoside diphosphate kinase (NDK) to convert GDP to GTP. GTP in turn is used up in the translation reaction. In this scheme inorganic phosphate (P_i) accumulates as a byproduct and can eventually contribute to the cessation of cell-free protein synthesis [52].

2.2   Exogenous gene expression in living cells

The expression of endogenous genes in a cell is under the control of a sophisticated regulation network where a specific signal is needed to activate genes. One example of such an activation mechanism are certain signaling molecules that bind to the DNA and cause a conformational change which makes the promoter of the gene accessible. Another example are signaling molecules that bind to a repressor that blocks a promoter. This causes the repressor to detach from the DNA and activates the gene [25]. A cell usually carries only one or two copies of a gene, thus when a gene is being transcribed only some tens of mRNAs are produced before the gene is switched off again. These short intervals of mRNA transcription are termed "bursts". Regulatory proteins are only present in an order of several hundred [53]. These small copy numbers cause endogenous gene expression in cells to be noisy (that is showing strong fluctuations over time [54, 55]). From this also follows that most of the noise arises from transcription whereas translation contributes only weakly due to the higher number of molecules involved.

Cells do not only express endogenous genes but they can also be made to express exogenous ones. Viruses have developed strategies to overcome the barrier that the cellular membrane poses and to insert their genetic code into the cell. This code is then processed by the cellular gene expression machinery but with the difference that the virus DNA has no off-switch and the expression does not happen in bursts but in one continued rise until eventual cell death.

The stage of development of artificial gene ferries is nowadays sufficiently high to use them as reliant vectors for gene delivery to eukaryotic cells. A widespread class of these artificial gene vectors are cationic lipids or polymers that encapsulate the DNA. The delivery process of such vectors to an eukaryotic cell consists of cellular uptake via the endosomal pathway, escape from the endosome, transport to and entry of the nucleus and finally unpacking of the enclosed DNA (compare Fig. 5.1 in chapter 5). If such exogenous genes do not need a signaling molecule to be activated, their gene expression profile does not show the typical burst behavior of endogenous genes but a continuous rise of the target protein level until a steady state between protein production and degradation via proteases is reached [56].

2.3   Modeling gene expression with rate equations

There are various approaches to modeling gene expression behavior which differ in the level of detail they feature [57]. Binary models only confer "on" and "off" states to individual genes. The computational cost of such models is comparably low but due to the strong simplification such models fail to reproduce the more detailed features of gene expression.

One step more complex are deterministic rate equation models. In these models the dynamics of the concentration of key components like mRNAs or proteins are examined. Each component is assumed to be synthesized and degraded with a certain rate. Mathematically this can be described using systems of ordinary differential equations (ODEs) of the form

$$\frac{d}{dt}B = \sigma A - \delta B \tag{2.1}$$

where A and B are concentrations of mRNAs or proteins and B is the product of A; σ is the synthesis rate of B and δ its degradation rate. Such models can be solved analytically if they are not too complex and otherwise numerically using well established computational tools. Although these models are per se deterministic they can also be adopted to reflect stochastic effects by introducing noise terms that account for varying levels of components of the gene expression apparatus.

The highest level of detail is conveyed by the class of stochastic kinetics models. This approach accounts for stochastic synthesis and decay of individual components. It is well suited to analyze reaction fluctuations due to small copy numbers of participants. However, the drawback is a high computational cost and this class of models does not yield an analytical solution which could further the understanding of the system [55, 57].

Cell-free gene expression comprises of large systems, that is intrinsic stochasticity does not play a dominant role. In fact the protein yield is predetermined by DNA template amount or experimental timing (compare experimental results in chapter 4). In this study it was desired to derive a model that is readily comprehensible and easily adaptable to individual cell-free systems. In addition the model was to have comparably low computational cost in order to quickly run it multiple times to test modifications. For these reasons the ODE based approach was chosen.

2.4   Green fluorescent protein – GFP

2.4.1   History of GFP

In 1955 D. Davenport and J. Nicol published their studies on bioluminescence of hydrome-dusae [58]. One species they studied was the jellyfish *Aequorea victoria*. It took another seven years before in 1962 Osamu Shimomura and coworkers reported the extraction and purification of a protein from *A. victoria* which fluoresced when induced with Ca^{++} [59]. This protein which they termed "Aequorin" fluoresced in a bluish light ($\lambda_{max} = 470 \, nm$) when purified whereas *A. victoria* itself showed greenish fluorescence ($\lambda_{max} = 509 \, nm$, see Fig. 2.5 A). Shimomura et al. speculated that this discrepancy might be due to an energy transfer from Aequorin to another fluorescent protein present in *A. victoria*, namely GFP which they had also been able to purify. This assumption was verified in the early 1970s [60, 61]. GFP differed from Aequorin and any other fluorescent protein known to that date in the fact that it does not require continuous uptake of chemical energy like Ca^{++}, solely excitation by UV or blue light. This would later prove to be one of the major factors in the usefulness of GFP in bioanalytical research as target cells do not have to be injected with potentially disturbing chemicals.

The structure of the GFP chromophore was first described by O. Shimomura in 1979 [62] and later in 1993 Cody et al. proposed some slight changes in the model [63]. The crystal structure of wild type GFP was first described by Yang et al. [64] in 1996.

However, it can be said that the true start of the success story of GFP was in the early 1990s. Prasher et al. cloned its cDNA and first described its nucleotid sequence in 1992 [65]. Based on this success, Chalfie et al. in 1994 showed that it is possible to transfer the GFP gene into new host organisms like *E. coli* or the roundworm *C. elegans* and have them produce the functional i.e. fluorescent protein [66]. Another important step in the growing usefulness of GFP was the generation of mutants with improved fluorescence properties and different wavelengths like blue, cyan and yellow (Fig. 2.5 B) which was first reported by the group of R. Tsien [67]. Nowadays GFP and its derivatives are widely used as marker genes in a wide variety of bioanalytical applications. For their contributions to this field Osamu Shimomura, Martin Chalfie and Roger Tsien were awarded the 2008 Nobel Prize in Chemistry [68].
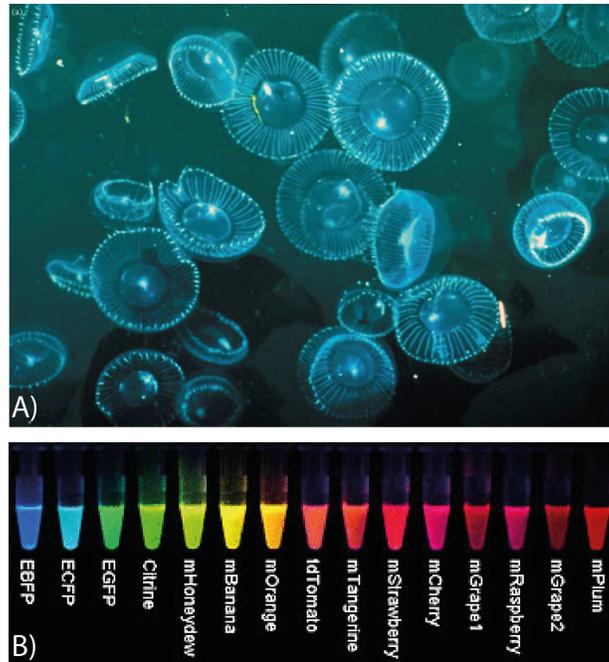
Figure 2.5: A) GFP was first isolated from the jellyfish *A. victoria*. The bluish green color of *A. victoria* as it is shown here stems from GFP that receives its excitation energy from another fluorescent protein termed Aequorin via Förster resonance energy transfer (FRET). Aequorin fluorescence in turn can be induced by Ca^{++}. This mechanism explains why the jellyfish does not need an excitation light source in order to exhibit GFP fluorescence (reprinted by permission from John Wiley & Sons, Inc.: Journal of Microscopy [69], copyright (2005)). B) Mutants of GFP were created that not only exhibit improved fluorescence properties but also have their excitation/emission maximum at different wavelength than wild type GFP. This library of mutants now spans the complete visible spectrum from blue to red light (taken from [70]).
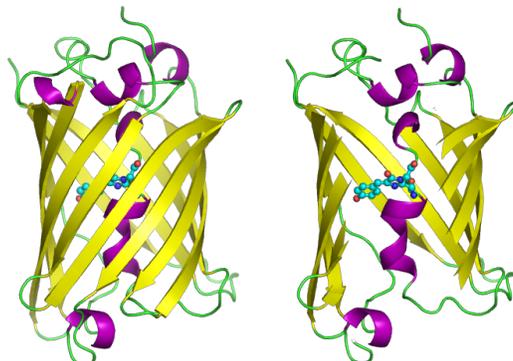
Figure 2.6: GFP is an eleven stranded *beta*-barrel protein. The top and bottom of the barrel are sealed by short α-helices. The chromophore is located in the center of the barrel were it is protected from environmental influences (taken from [73]).

2.4.2 Fluorescence mechanism

GFP has a β-barrel structure consisting of 11 strands with short α-helices which seal the top and bottom of the barrel and also hold the chromophore in place in the middle of the barrel. Thus in correctly folded GFP the chromophore is completely protected from environmental influences like quenching by outside oxygen (Fig. 2.6) [64]. The chromophore is composed of the amino acid sequence serine-tyrosine-glycine on positions 65-67. The formation of the chromophore in a freshly translated protein works as follows: through dehydration Gly-67 cyclizes with the carbonyl group of Ser-65 to form imidazolin-5-one. Then, in the presence of molecular oxygen dehydrogenation occurs and creates a conjugated π-electron resonance system comprising of the imidazolin ring and the tyrosine phenyl ring (Fig. 2.7) [71, 72].[5] As one can see the formation of the GFP fluorophore (the so called GFP maturation) does not depend on any *A. victoria* specific enzymes but it is fully autocatalytic. This property makes the application of GFP as such a versatile reporter gene possible in the first place.

2.5   Heavy water

In 1931 Harold Urey substantiated the isotope deuterium (for this discovery he was awarded the 1934 Nobel Prize in Chemistry) [74]. Two years later Gilbert Lewis reported about the purification of deuterium oxide (also called heavy water), that is a water sample with

---

[5]An interactive Java tutorial depicting the formation of the GFP fluorophore can be found at [72].
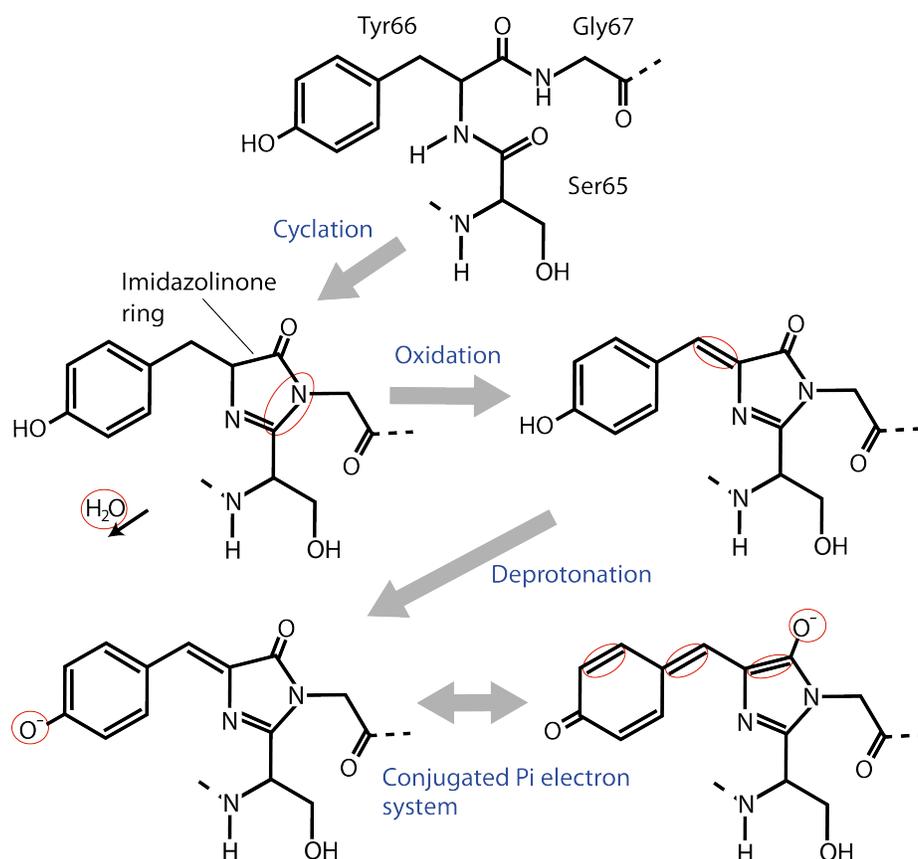
Figure 2.7: GFP maturation as described in Ref. [71, 72]: the first step is a torsional rearrangement of the amino acid sequence which brings the carbonyl group of Ser-65 into close proximity of the amino group of Gly-67 (here the result of this rearrangement is shown). Subsequently the system dehydrates and an imidazolin-5-one ring forms (cyclation). Then in the presence of molecular oxygen the $\alpha - \beta$ carbon bond of Tyr-66 oxidates and thus the conjugated π-electron system is expanded to include Tyr-66. The protonated form of GFP is responsible for the GFP excitation peak at $395\,nm$ whereas the deprotonated form is excited at $470\,nm$ (eGFP at $488\,nm$).

deuterium (^{2}H) instead of protium (^{1}H) atoms [75].[6] Deuterium oxide can also be found in nature: a small percentage of natural water molecules (about 0.01%) contains deuterium instead of protium. Heavy water features slight differences in its physical properties compared to normal water: it is about 11% denser, its viscosity is about 23% higher, the energy required for dissociation is higher etc. An overview over the physical properties of H_2O and D_2O is given in table 2.2.

Heavy water can be produced for example by electrolysis utilizing the kinetic isotope effect: since deuterium is twice as heavy as protium the vibrational zero-point energy of its bonds is much lower and the activation energy of reactions that break these bonds is accordingly higher [76]. Electrolysis therefore breaks light water into hydrogen and oxygen but leaves deuterium oxide largely intact. Commercially available heavy water can have purities as high as 99.9%.

The fields of application of deuterium oxide are quite diverse. For example it is used in nuclear power plants of the heavy water reactor type. Heavy water absorbs neutrons to a lesser extent than light water therefore it is possible to use non-enriched uranium in these reactors. Deuterium oxide is further used in nuclear magnetic resonance (NMR) spectroscopy because it has a distinct magnetic moment and can be clearly separated from the background water signal. Especially NMR studies of proteins with more than $20\,kDa$ stringently require $80-90\%$ deuteration [77]. Another field of application was the Sudbury Neutrino Observatory (in operation 1999-2006) which was filled with $1000\,t$ of heavy water. In neutrino detectors using light water only electron scattering of neutrinos is detectable. In heavy water neutrinos can also interact with the deuterium nucleus, fissioning it. This weak force interaction can happen via a Z boson (for electron neutrinos) or via a W boson (for all neutrino flavors). In this manner was it possible to experimentally verify neutrino oscillations and solve the solar neutrino problem [78].

Moreover heavy water affects cellular proliferation if it has sufficiently accumulated in the organism. It is assumed that the main target of the inhibitory effect of heavy water is the mitotic spindle formation process. Eukaryotic cells are unable to divide without the mitotic spindle. A more detailed description of the effects of heavy water on living organisms can be found in section 6.1.

---

[6]To be precise the term heavy water could also be applied to water containing the oxygen isotopes ^{17}O and ^{18}O.

	Hydrogen oxide	Deuterium oxide
Molar mass (g/mol)	18.015	20.029
Max. Density (g/cm^3)	0.999975	1.10589
pH and pD resp. (at $25°C$)	7.00	7.43
Viscosity (at $20°C$, $mPa\,s$)	1.0016	1.2467
Freezing point $(°C)$	0.00	3.82
Boiling point $(°C)$	100.00	101.4
Melting heat (J/mol)	6.01	6.34
Evaporation heat (J/mol)	44.01	45.47

Table 2.2: Comparison of the physical properties of H_2O and D_2O.

Chapter 3

Materials and Methods

3.1 Fluorescence microscopy

The spontaneous emission of light when an excited electron state goes into a state of lower energy is called luminescence. This phenomenon is subdivided into fluorescence which comprises spin allowed transitions (that is $\Delta S = 0$) between energy states and phosphorescence which comprises spin forbidden transitions ($\Delta S \neq 0$). Typical life times of excited states are up to a few micro seconds for fluorescence and from milliseconds to hours for phosphorescence (Fig. 3.1 A). The emitted photon of a fluorophore usually has lower energy than the one absorbed because the fluorophore relaxes to the lowest vibrational level of the excited state before it relaxes to its electronic ground state via photon emission. This effect is called Stokes-shift after the Irish physicist George G. Stokes. Fluorescence microscopy makes use of this shift as it makes it possible to separate excitation and emission light in one optical path through the use of suitable filters (Fig. 3.1 B). Non-cytotoxic fluorophores like GFP and its derivatives can be bound to virtually any component of a cell in order to monitor it *in vivo*. Because fluorophores emit their own light the resolution of fluorescence microscopes is not limited by Abbe's law[1].

High excitation intensities or sufficient excitation/emission cycles lead to the photochemical destruction of fluorophores. This process is called photobleaching. Until recently mercury-vapor lamps were the prevalent light source for fluorescence microscopes. These lamps emit light in the whole visible spectrum as well as at UV wavelengths. A filter is used to extract the desired excitation wavelength. However, such filters are not 100%

---

[1] According to Abbe's law the minimum distance of to points that can still be identified as distinct is $d = \frac{\lambda}{2n \sin \alpha}$ where λ is the wavelength, n the refractive index and α half the angular aperture.
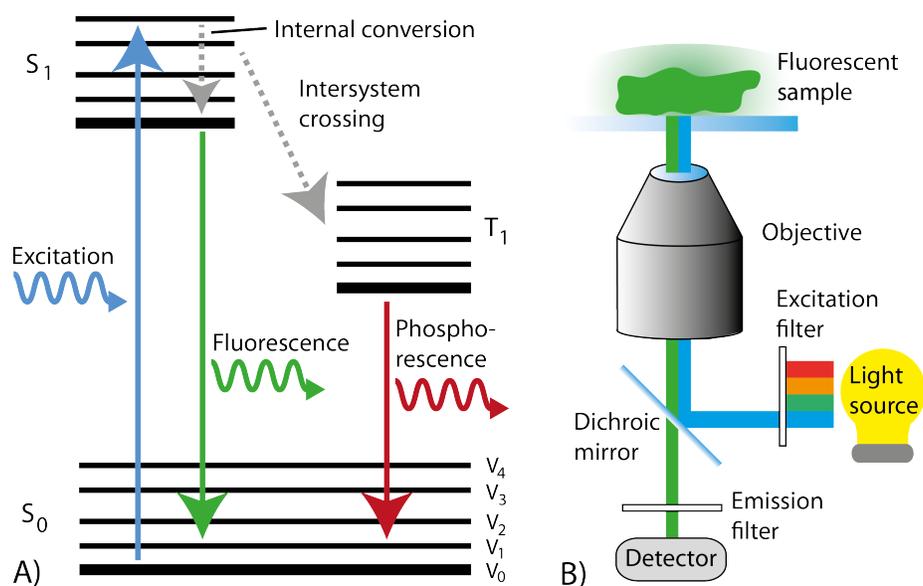
Figure 3.1: A) The possible transitions of electrons between different energy states in a fluorophore (Jablonski diagram). Relaxation by emission of a photon competes with radiation-free processes. In a proper fluorophore photon emission is the dominant process. Transitions that involve spin conversions have significantly longer lifetimes than spin allowed transitions. B) Schematic setup of a fluorescence microscope. Excitation and emission light are centered around different wavelengths due to the Stokes shift. Thus their beam paths can be separated using a dichroic mirror.
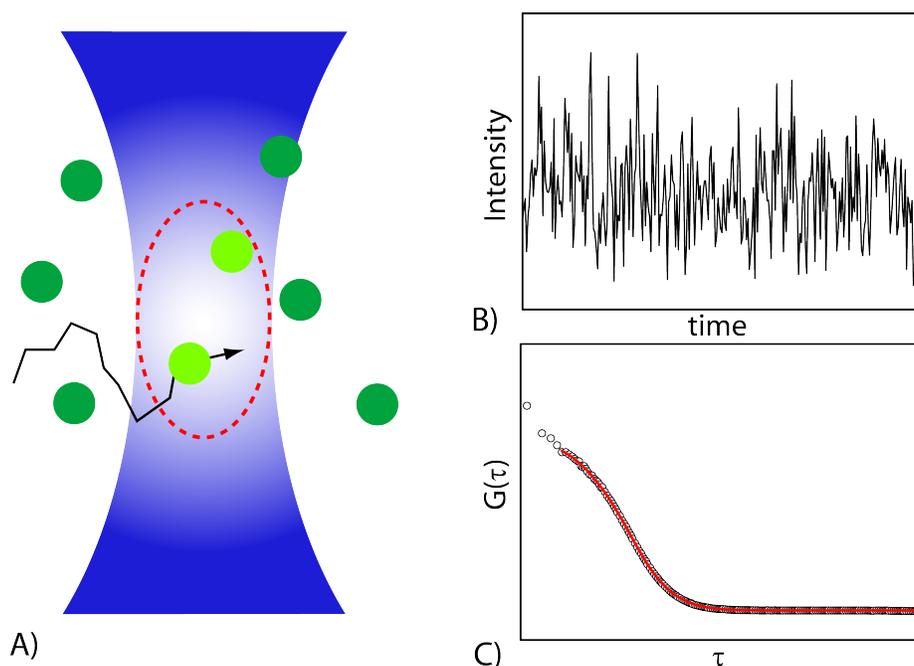
Figure 3.2: A) Technically FCS is usually realized on a confocal fluorescence microscope. Confocal microscopes are equipped with a pinhole in front of the detector which only light from the confocal plane can pass as well as from slightly above and below due to the diffraction limit. This yields a confocal volume as depicted by the red oval. The detector has to feature fast single photon sensitivity therefore photon multiplier tubes or avalanche photo diodes are used. B) Fluorescent molecules diffuse in and out of the confocal volume which leads to fluctuations in the recorded fluorescence intensity. C) Autocorrelation data of the data in B) and best fit according to equation 3.3.

impermeable for the wavelengths they should block and a small percentage is transmitted. This way photobleaching is stronger than it would be with ideal filters, especially because of the high energy UV light transmitted. Over the last years high performance LEDs have been developed and are now a preferred light source for fluorescence microscopy despite the disadvantage that a different LED is needed for each excitation wavelength. In addition to avoiding the harmful UV light the illumination profile of LEDs is preadjusted and is in general much more uniform than that of mercury-vapor lamps.

3.2 Fluorescence Correlation Spectroscopy – FCS

FCS is a fluorescence fluctuation analysis method with single molecule sensitivity. It was developed in the 1970s by Douglas Magde, Elliot Elson and W. W. Webb [79–81]. Fluores-

cent particles diffuse in and out of an excitation volume which leads to temporal fluctuations in the recorded fluorescence intensity (Fig. 3.2). In FCS the agreement between the signal at time point t and after a time lag $t + \tau$ is analyzed via temporal autocorrelation:

$$G(\tau) = \frac{\langle \delta F(t) \delta F(t + \tau) \rangle}{\langle F(t) \rangle^2} \tag{3.1}$$

Here the pointed brackets denote averaging over time and $\delta F(t) = F(t) - \langle F(t) \rangle$ is the deviation from the mean fluorescence signal. FCS is technically realized with a confocal microscope setup[2]. Thus the excitation volume is usually of the order of one femtoliter and typical sample concentrations are in the nanomolar regime. Assuming proper adjustment of the optical pathway the excitation volume can be well approximated by a 3D Gaussian:

$$I(x, y, z) = I_0 \exp \left(-\frac{2(x^2 + y^2)}{\omega_{xy}^2} - \frac{2z^2}{\omega_z^2} \right) \tag{3.2}$$

where I_0 denotes the peak intensity and ω_{xy}, ω_z are the Gaussian's radii in the respective directions. With this a model for the observed intensity fluctuations can be created. In the simple case that the fluorescent particles are driven by Brownian motion and can diffuse unrestricted in three dimensions the analytical solution is [82]:

$$\boxed{G(\tau) = \frac{1}{N} \cdot \frac{1}{1 + \frac{\tau}{\tau_D}} \cdot \frac{1}{\sqrt{1 + \frac{\tau}{S^2 \tau_D}}}} \tag{3.3}$$

Here N is the mean number of particles in the excitation volume, $S = \omega_z / \omega_{xy}$ is called structure parameter and τ_D is the average diffusion time through the focal volume in the xy-plane. τ_D is related to the diffusion coefficient of the particle D via

$$\tau_D = \frac{\omega_{xy}^2}{4D} \tag{3.4}$$

Employing the Stokes-Einstein relation FCS also returns the hydrodynamic radius R of the diffusing particles:

$$D = \frac{k_B T}{6 \pi \eta R} \tag{3.5}$$

---

[2]A confocal setup is basically a fluorescence microscope with a pinhole in front of the detector. Only emission light from the focal plane can pass this pinhole which means that the sample can be dissected along the optical axis and a 3D image can be created. The detector is usually a photomultiplier or an avalanche photo diode for very high sensitivity.

where k_B is the Boltzmann constant, T the temperature of the sample and η the viscosity of the medium.

Thus FCS is a highly sensitive analysis method for quantitative studies of diffusion coefficients, concentrations, chemical reaction kinetics or hydrodynamic radii. Great advantages of FCS are single cell sensitivity and the ability to observe fluorescently tagged particles *in situ* in living cells.

3.3 Considerations for quantitative measurements

An important requirement of this thesis was to measure fluorescence quantitatively – even for very low concentrations of fluorophores. Therefore considerable effort was put into identifying and eliminating sources of errors. For example in the spectrometer it was found that it is necessary to fix the cuvette with a chock in the sample holder otherwise small misalignments produce considerably different results in repeated measurements. Furthermore, different batches of cell-free kit were found to produce divergent GFP levels. This may be due to intrinsic variations in composition of the batches or due to unequal handling in the laboratory (like small inaccuracies in aliquoting the cell-free kit or different storage times). Associated experiments should therefore be conducted using one batch of cell-free kit if possible. In addition is it advisable to create a stock of cell-free kit components for parallelized measurements instead of pipetting components for each replica individually in order to minimize the effect of pipetting inaccuracy.

Experimental timing is a further aspect that needs to be heeded. First, crucial components of the cell-free system begin to degrade after thawing and measurement results diverge if the timing is not consistent. Second, at sample volumes of only some micro liters evaporation is a problem and samples need to be quickly sealed during preparation in order to avoid noticeable volume loss and therewith change of component concentrations. In this context one needs to keep in mind that polydimethylsiloxane (PDMS) which is often used to create structures for microfluidics is not completely air tight. Therefore micro liter volume cell-free expression in PDMS casts is prone to evaporation during measurements, especially when the experiment is not conducted at room temperature but at $37°C$.

It was also found that adjusting the sample temperature to exactly $37°C$ is not trivial on the fluorescence microscope and FCS setups. Good accuracy is necessary because the protein synthesis reaction is temperature sensitive. On the spectrometer and on the plate reader respectively the sample is put in a enclosed temperature controlled box where the

temperature control works well. On microscopes the equivalent to this ideal would be to enclose the whole microscope setup in a thermo box. When only the sample is heated like with the ibidi Heating System (ibidi, Germany) one has to keep heat loss in mind. This means that samples should not be put solely on thin bottoms like cover glass slips but instead on, for example, six well plates or petri dishes which such heating systems are made for and which are thicker and show less heat loss. Attention should be paid to the fact that in such heating chambers the actual temperature can diverge from the predefined one depending on room temperature, air conditioning or chamber type. A digital thermometer should be used to check the proper setting. Immersion objectives should be avoided where possible as it was found that objective heaters may not be sufficient to counteract heat loss through an immersion objective that is in direct contact with the sample.

The need for good quantitative accuracy also sets a limit as to how far samples could be downsized. Downsizing is beneficial first because commercial cell-free kits are expensive and second because it allows for massive parallelization of measurements in micro well format. However it was found that at sample volumes below about ten micro liters measured fluorescence levels varied considerably between individual measurements. Note that these fluctuations can not be explained by low copy numbers of cell-free system components: the minimum sample volume tested was $0.2\,nl$. This volume contains about

- $1 - 10^4$ DNA molecules (depending on the concentration)

- 10^3 RNA-polymerases

- 10^4 ribosomes

- 10^6 amino acids

Unequal filling due to pipetting inaccuracy, surface effects due to imperfect passivation of surfaces (cell-free kit components carry his-tags and may bind to surfaces) as well as evaporation of samples during measurements may all contribute to the observed discrepancies.

It is further necessary to take the comparable instability of mRNA into account. RNAses are ubiquitous and very sturdy[3] therefore certain precautions like using RNAse-free pipette tips, cleaning the working space with RNAse removal agent and treating sample holders in a plasma oven are inevitable in order to avoid sample contamination.

---

[3]In fact RNAse A can be isolated by boiling cellular extract until all other enzymes are denaturated.
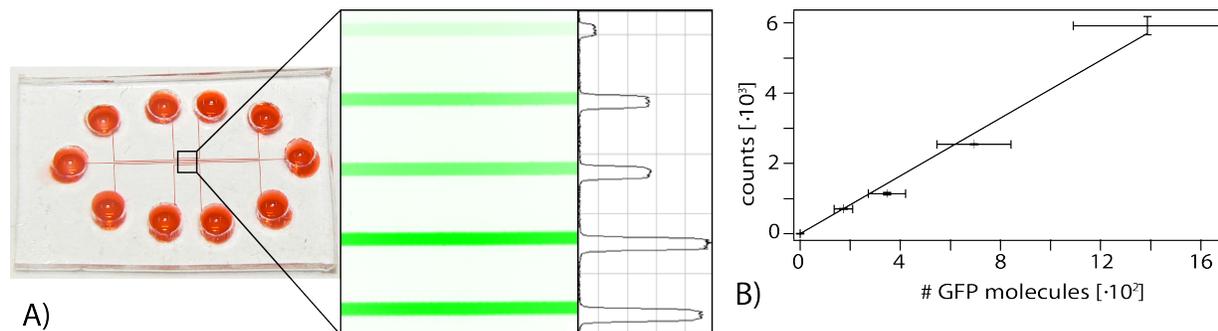
Figure 3.3: A) Micro channels for calibration. In the photograph on the left the channels are filled with red food coloring for better visibility. The center shows a fluorescence image of the five channels filled with GFP calibration standard (from the top channels two and three and four and five respectively contain the same GFP concentration). On the right a histogram following a vertical line through the five channels is shown. B) Calibration curve for translating arbitrary fluorescence units into numbers of GFP molecules. Similar calibration curves were established for spectrometer and plate reader.

3.4 Calibration standard for GFP

GFP synthesis was measured fluorescently on a number of instruments (spectrometer, plate reader, fluorescence microscope) depending on the requirements of the respective measurement. These instruments yield arbitrary fluorescence units as readout. In order to translate this readout into a meaningful quantity (that is number of GFP molecules or GFP concentration) it is therefore necessary to establish a calibration curve. This can be done by measuring the fluorescence of a GFP standard with known concentration.

GFP standard was expressed in plasmid transformed *E. coli*. The plasmid contained exactly the same GFP sequence as the one used for experiments thus ensuring complete analogy of the resultant GFP. A his-tag was attached to the end of the GFP sequence in order to purify it using high performance liquid bioaffinity chromatography. GFP stock in phosphate buffered saline (PBS) pH 7.4 contained the following additives: Tween 20 (0.001%) to prevent binding to surfaces and sodium azide 0.002% as a preservative. The GFP concentration of the stock solution was measured photometrically[4] on a Nanodrop 1000 spectrophotometer (Thermo Scientific, Germany) with $\epsilon_{GFP}(488\,nm) = 55000\,M^{-1}cm^{-1}$. The concentration was also measured by fluorescence correlation spectroscopy for verification.

---

[4]Absorbance and concentration of a solution are connected by the Lambert-Beer law: $A = \epsilon \cdot c \cdot d$ with $\epsilon(\lambda)$ the molar extinction coefficient of the solvent at the respective wavelength and d the optical path length.
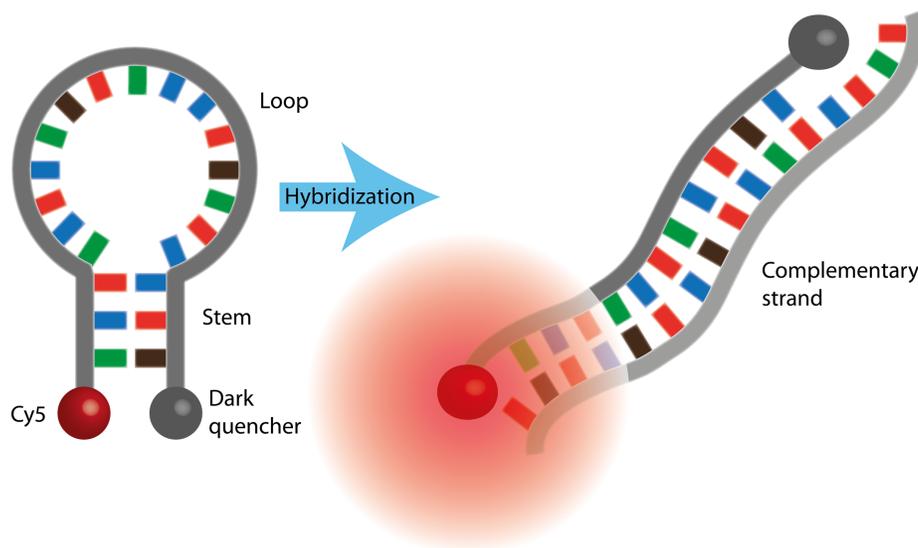
Figure 3.4: Working principle of molecular beacons: In the absence of the target sequence the beacons fold back on themselves. In that case the ends are in close contact and the fluorophore is efficiently quenched. However, binding to the target is energetically favorable and therefore in its presence the beacons open up and fluorescence can be detected.

Calibration curves for spectrometer, plate reader and fluorescence microscope were established using the same instrument settings and sample volumes as in the measurements. For calibration of the microscope GFP calibration solution was filled into micro channels cast in PDMS, see Fig. 3.3A. PDMS casts were produced using a silicon master created with standard photolitography methods. PDMS chips were glued to glass cover slips via oxygen-plasma hydrophilisation. The channels were passivated by treating them with PBS containing $3\,\mu M$ Pluronic F108 for $20\,h$ at $4°C$. They were subsequently washed with PBS three times and dried under a nitrogen flow. Knowledge of the channel height (width $20\,\mu m$, height $8\,\mu m$, length $10\,mm$) allows for calculation of GFP molecules in a defined region of interest.

3.5   Molecular beacons for mRNA labeling

If mRNA levels are to be measured fluorescently it is necessary to apply a fluorescent label. As described in section 4.2 labeling with unspecific RNA dye was not successful because of a strong background signal (likely due to the dye binding to tRNA). Therefore a dye was needed which specifically binds to mRNA and which is also not toxic for the cell-free transcription apparatus. Molecular beacons fulfill these requirements [83]. These beacons
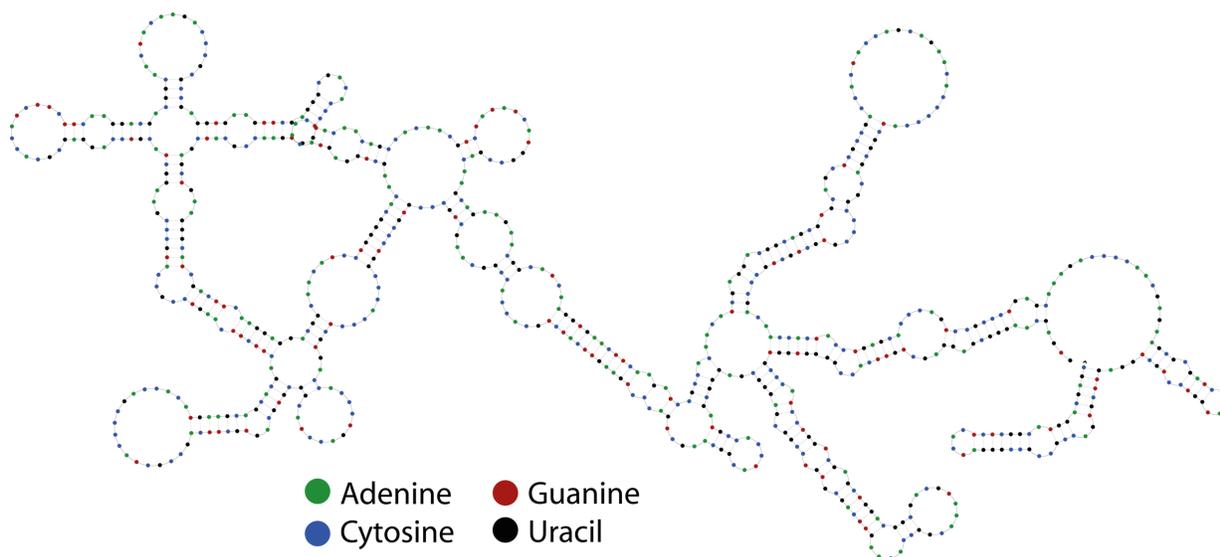
Figure 3.5: Minimum free energy structure of the cell-free transcribed mRNA at $37°C$ (calculated with the NUPACK software).

are oligonucleotides with a stem-loop structure (see Fig. 3.4). The stem forms because two complementary sequences at the ends of the beacons anneal at temperatures lower than their melting temperature. The loop sequence is complementary to a part of the target mRNA and hybridizes with it when it comes into close proximity. Since the loop sequence is longer than the stem the hybrid is more stable than the hairpin structure and it is therefore the prevalent configuration. This separates the two ends of the beacon and therewith the fluorescent donor-quencher pair attached to said ends[5]. From this point on fluorescence of the donor can be detected.

The minimum free energy structure of the cell-free transcribed mRNA at $37°C$ was modeled using the software package NUPACK (`www.nupack.org`), see Fig. 3.5. A readily accessible stem-loop sequence was chosen as binding site for the beacons. Molecular beacons with the sequence $5'-TCGCCGTGGGGGTGTTCTGCTGGTGTGGTCGGCG-3'$ were ordered from biomers.net (Germany). Beacons carried the fluorescent dye Cy5 on the $5'$ end and the dark quencher BBQ-650 on the $3'$ end. Since each mRNA is labeled with exactly one Cy5 molecule there is a one to one relation between the Cy5 and mRNA concen-

---

[5]Fluorophores in very close proximity (less than $10\,nm$) whose respective emission/excitation spectra overlap exhibit non-radiative energy transfer from the "donor" to the "acceptor" due to dipole-dipole interactions [84]. This mechanism is termed "Förster Resonance Energy Transfer" (FRET) after the German physicist Theodor Förster. FRET is a highly sensitive technique for measuring changes in length as the transfer efficiency E scales with the inverse of the sixth power of the distance: $E = \frac{1}{1+(r/R_0)^6}$. R_0 is the so called Förster radius, it depends on the specific donor/acceptor pair.
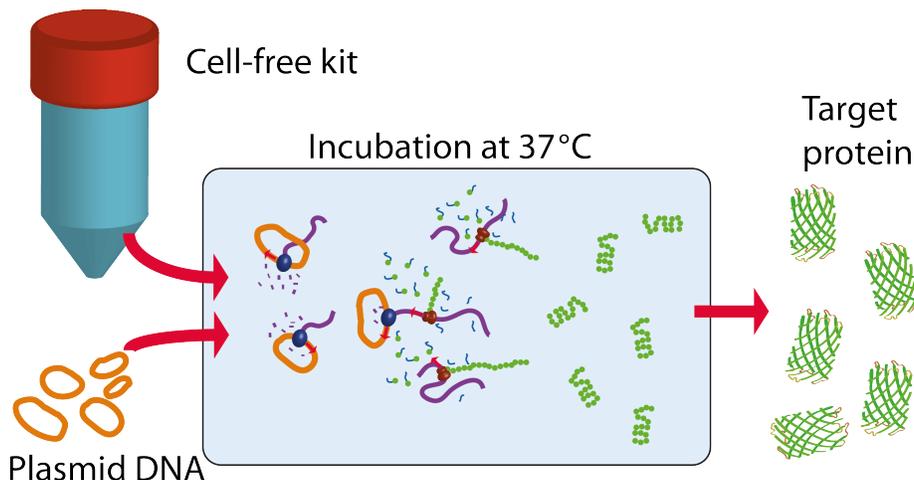
Figure 3.6: Principle of cell-free protein synthesis in a combined transcription/translation system.

tration in the sample. A Cy5 calibration curve using a Cy5 dilution series was established similar to the one for GFP described above. Molecular beacons were added to samples in about tenfold the expected mRNA concentration to ensure saturation of binding partners.

RNA mimics of GFP   It should be noted that recently (after RNA measurements reported here were done) a new promising variant of RNA labels was reported [85]. In GFP the fluorophore sequence ($Ser^{65} - Tyr^{66} - Gly^{67}$) is located in the center of the molecule, shielded from outside influences (compare Sec. 2.4.2). Paige et al. describe a GFP analog based on a RNA aptamer that binds a fluorophore in a similar way. They used derivatives of the GFP fluorophore for their work because these do not show cytotoxicity and fluoresces only when bound to the protective aptamer (free GFP fluorophores dissipate their energy non-radiatively). The authors managed to produce a palette of RNA-fluorophore complexes which cover the visible spectrum from $300\,nm$ to $500\,nm$. They demonstrated that these aptamers are transcribed in living cells and can be used to depict RNA dynamics within them after a brief incubation with the fluorophore. As the authors note this system might be of great merit for RNA-RNA, RNA-protein interaction studies as well as simultaneous observation of various RNAs.

3.6 Cell-free protein synthesis

Plasmid DNA for cell-free GFP synthesis The *gfpmut1* mutant of the GFP gene taken from the pEGFP-N1 vector (BD Biosciences Clontech, Germany) was cloned into the pET-23b vector (Novagen, USA). pET-23b carries a T7 promoter sequence and is therefore suitable for cell-free gene expression. gfpmut1 exhibits higher photostability and fluorescence compared to wild-type GFP. It also has its excitation maximum shifted to $488\,nm$ which better matches the standard fluorescein microscope filters. pET-23b plasmid DNA was purified in deionized water. The concentration was measured with a NanoDrop spectrophotometer (Thermo Scientific, USA) by absorption at $280\,nm$. Aliquots were stored at $-20°C$.

Sample preparation When preparing samples for cell-free gene expression great emphasis was put on working RNAse free. The working space on the laboratory bench was treated with RNAse-ExitPlus solution (AppliChem, Germany) and solely RNAse free pipette tips (Biozyme Diagnostik, Germany) were used. Sample holders were subjected to treatment in a plasma oven immediately prior to measurements.

The reconstituted cell-free system PURExpress (New England Biolabs, Germany) was used as described by the manufacturer. In short, PURExpress is provided in to solutions "A" and "B" which can be stored in aliquots at $-80°C$ for at least six month. PURExpress contains T7 RNA polymerase in about $100\,nM$ concentration and ribosomes in about $2.4\,\mu M$ concentration. For an experiment components "A" and "B" were mixed, DNA was added at the desired concentration and the total sample volume was adjusted to $25\,\mu l$ using RNAse free water. After filling the sample into the sample holder it was covered with mineral oil in order to prevent evaporation. A digital temperature control unit was used to maintain a constant $37°C$ during measurements.

Data acquisition Kinetics of cell-free GFP synthesis were recorded fluorescently. Initially measurements were performed on a Fluorolog-3 spectrometer (Horiba Jobin Yvon, Germany) due to the unparalleled sensitivity of spectrometers compared to other instruments like fluorescence microscopes or microplate readers. The employed sample holder was a fused silica ultra-micro cuvette (Hellma, Germany). Instrument parameters were as follows: excitation wavelength $488\,nm$, recorded emission spectrum $500 - 600\,nm$, width of spectrometer slits $8\,nm$, integration time $0.1\,s$. Data points were recorded every 10 to 20 seconds.

After fine tuning the measurement protocol sufficient sensitivity was also achieved with a Fluostar Optima plate reader (Optima, Germany). Using the plate reader is advantageous because here multiple measurements can be conducted in parallel. Here the employed sample holder was a 96 well plate with V-bottoms (Greiner Bio-one, Germany). Instrument parameters were: excitation/emission filter set $485/520\,nm$, gain 2000, ten flashes per well, bottom optics.

Data analysis   Data were analyzed using the OriginPro 8.5G software. GFP emission spectra recorded on the spectrometer were integrated and plotted vs. measurement time. Plate reader data were background corrected by shifting time traces such that the second data point of each curve has zero fluorescence intensity. The correction was performed this way because the first data point of each curve showed instrument dependent erratic behavior. Mean values were created out of associated data sets. Statistical outliers were omitted from averaging.

3.7   Real-time PCR

Polymerase chain reaction (PCR) is an *in vitro* technique for the amplification of DNA sequences [86, 87]. It was invented in 1983 by Kary Mullis who received the Nobel Prize in Chemistry for his invention in 1993. PCR is nowadays one of the most important tools of molecular biology. Examples for its application are screening for hereditary diseases, determination of genetic fingerprints or functional gene analysis. The working principle of PCR is based on the enzymatic amplification of a DNA sequence using a DNA polymerase. The polymerase needs to be thermostable in order to endure the high temperature steps in PCR (see below). For example the Taq polymerase originally isolated from the thermophilic bacterium *Thermus aquaticus* fulfills this requirement. A typical PCR run goes as follows: DNA template, polymerase, deoxynucleoside triphosphates (dNTPs) oligonucleotide sequences complementary to the $3'$ end of the sense and anti-sense strand of the DNA (primer) and buffer are mixed and placed in a thermocycler. The thermocycler executes 20-30 cycles of a temperature steps. Each cycle consists of the following steps:

- Denaturation: At about $95°C$ the hydrogen bonds between the DNA double strands are melted. This step takes about 30 seconds.

- Annealing: At $55 - 65°C$ (about $5 - 10°C$ below the melting point of the primers)

      primers hybridize with the complementary DNA sequence. This step takes about 30 seconds.

- Elongation: At its optimum working temperature (about $72°C$) the polymerase, starting at the $3'$ end of the primer, synthesizes a new DNA strand complementary to the template. Taq polymerase adds about 500 bases per 30 seconds.

The newly synthesized DNA strands do not have a defined stop, they end when the polymerase unbinds from the template at the end of the elongation step. However, the at the $3'$ end overly long strands serve as templates in the next cycle. Primers bind at their $3'$ region and since the strands already have the correct $5'$ end the strands synthesized now have the correct length. As the new strands themselves serve as template in consecutive cycles strands with the correct (short) sequence accumulate exponentially whereas overly long strands do so only linearly (since only the original DNA serves as template for them).

    Real-time PCR is a variation of the standard PCR work flow where the amount of product is measured in each cycle. To this end a fluorescent dye like Sybr Green which fluoresces after binding to double stranded DNA is added to the sample and the thermocycler needs to be equipped with a fluorescence reader.

mRNA purification and cDNA production   A cell-free transcription/translation reaction with pET 23b plasmid and PURExpress cell-free kit was conducted following the manufacturer's instructions. However, samples were diluted with D_2O in the desired concentration as described in the text. After $4\,h$ of cell-free reaction at $37°C$ the cell-free transcribed mRNA was purified using the "RNA cleanup" protocol of the RNeasy Mini Kit (Qiagen, USA). Subsequently complementary DNA (cDNA) was reverse transcribed out of the mRNA with the SuperScript II Reverse Transcriptase kit (Invitrogen, USA).

PCR sample preparation   Samples consisted of $2\,\mu l$ cDNA, $2\,\mu l$ each of forward (5'-cgc cac cat ggt gag caa gg-3') and reverse (5'-ggt tgt cgg gca gca gca cg-3') primer, $10\,mM$ dNTP mix, $2\,\mu l$ of $25\times$ concentrated Sybr Green, $2\,U$ Taq DNA polymerase (New England Biolabs, Germany) and $5\,\mu l$ of $10\times$ reaction buffer.

Data acquisition   Real-time PCR was carried out on a C1000 thermal cycler equipped with the CFX96 Real-time Detection System (Bio-Rad, Germany). The PCR protocol was as follows: initiation at 95°C for 5min. This was followed by 30 steps of denaturation

($94°C$, $30\,s$), annealing ($65°C$, $30\,s$) and extension ($72°C$, $30\,s$). Final extension was done at $72°C$ for 5min.

Data analysis   In PCR the amount of DNA and therewith the Sybr Green fluorescence grows exponentially with the cycle number. Therefore the sooner the observed fluorescence crosses a manually defined threshold the higher is the amount of original template in the respective sample. The cycle number at which each PCR curve crosses the threshold is called the C_t value. C_t values were plotted against the D_2O concentration in the respective sample and corresponding C_t values were averaged. Outliers were omitted from the statistics.

3.8   Fabrication of lipid vesicles and supported lipid bilayers

All lipids (Avanti Polar Lipids, USA) were stored in chloroform at $-20°C$. Lipid solutions of the desired composition were mixed under the fume hood using glass vials and syringes. Subsequently the chloroform was evaporated under a nitrogen stream followed by over night storage of the vials in a vacuum oven at room temperature. Lipid compositions for giant unilamellar vesicles (GUVs) were not dried out but kept in chloroform.

Supported lipid bilayers (SLBs) were prepared as follows: SLBs consisted either of pure 1-Stearoyl-2-Oleoyl-Sn-Glycero-3-Phosphatidylcholine (SOPC) or of a mixture of $90\,mol\%$ SOPC and $10\,mol\%$ N-((6-(biotinoyl)amino)hexanoyl)-1,2-dihexadecanoyl-sn-glycero -3-phosphoethanolamine (biotin-X DHPE). Dried lipids were resuspended in isopropanol at $2\,mg/ml$ and spin coated on the substrate at $2000\,rpm$ for $2\,s$, followed by $3000\,rpm$ for $30\,s$. The coated substrate was subsequently dried out in vacuum oven for at least $15\,min$ and then assembled in a microscope chamber, filled with deionized water (DI) and left over night. In the water the lipids arranged to form a bilayer on top of the substrate. Excess lipids were washed away by thorough rinsing of the chamber with DI. The DI was then exchanged for phosphate buffered saline (PBS). This was carried out so that the bilayer was never exposed to air as the air-water intersection would immediately destroy the bilayer. In the case of SLBs with biotin-X DHPE at this point $40\,\mu g/ml$ streptavidin was added to the chamber and left to incubate for $30\,min$. Subsequently the chamber was rinsed with DI again and then the DI was replaced by glucose solution.

GUVs were prepared with the electroformation method (compare [6]): Lipids dissolved

in chloroform ($20\,mol\%$ cholesterol, $40\,mol\%$ brain sphingomyelin, $40\,mol\%$ dioleoyl-1,2-sn-phophatidylcholine (DOPC) and $0.1\,mol\%$ Texas Red DHPE (1,2-dihexadecanoyl-sn-glycero-3-phosphoethanolamine, triethylammonium salt at $2\,mg/ml$) were put on indium tin oxide (ITO) plates and dried out in a vacuum oven. The ITO plates were then assembled face to face in a watertight Teflon chamber and the space in between was filled up with sucrose solution. Subsequently an electric field ($10\,Hz$, $2\,V$) was applied for two hours. During this time the chamber was heated to $40°C$, that is above the phase transition temperature of the ternary lipid mixture. GUVs were then harvested with a Pasteur pipette and added to a microscopy chamber filled with glucose solution to a final ratio of about 1:2 to 1:6.

Small unilamellar vesicles (SUVs) were prepared via extrusion using a mini-extruder (Avanti Polar Lipids): dried lipids were resuspended in carbonate buffer ($1-2\,mg/ml$) and pressed through micropores of the desired size ($200\,nm$ or $1\,\mu m$). A suspension of vesicles with a very monodisperse diameter formed. Mini-extruder and syringes were washed with ethanol, DI and buffer prior to use. After use the cleaning procedure was repeated in the opposite order. For expression of the transmembrane protein ephrin-B2-GFP $8\,\mu l$ SUV solution, $17.5\,\mu l$ PURExpress cell-free kit and $2\,\mu l$ of ephrin:GFP DNA ($131\,ng/\mu l$) were mixed and heated to $37°C$ for three hours.

3.9   Microstructuring of glass substrates

Microstructures on glass substrates were created using standard photolitography techniques: glass substrates were covered with photo resist, put under a chrome master with the desired structure and illuminated with UV light. The chrome master exhibited $2\,\mu m$ wide stripes in $4\,\mu m$ intervals. Subsequently the photoresist was developed and reactive ion etching (RIE) was used to create the micropattern. Photoresist remains were cleaned off with piranha solution (3:1 H_2SO_4/H_2O_2). Patterns of alternating smooth and rough stripes were fabricated using borosilicate slides (etching depth of the rough stripes about $30\,nm$). Groove patterns were fabricated using fused silica slides (groove depth about $1\,\mu m$).

Chapter 4

Modeling gene expression dynamics in a cell-free system

Content of this chapter is the subject of publication [P2] (reproduced by permission of The Royal Society of Chemistry).

4.1 Quantitative measurements with a cell-free system

As described in section 2.1 since their first introduction cell-free systems have been developed into a versatile tool capable of synthesizing a wide variety of proteins. Even the notoriously difficult to handle membrane proteins can be synthesized *in vitro* either with the use of detergents or by letting them embed into an artificial membrane *in situ* [9–11]. Another field of application for cell-free systems are microfluidic devices [16, 17]. Cell-free expression in combination with microfluidics allows for massive parallelization of measurements and can greatly quicken protein interaction studies. This is of interest for example for identifying and testing new drugs as most drugs target proteins [14]. In the same manner such devices could be used for the detection of toxins for example in food or drinking water [15]. Furthermore, with this approach one can identify the individual components of cellular signaling networks and quantify their binding affinities [18, 19].

Beyond purely qualitative work and depending on the actual question being investigated it is necessary to describe the dynamics of protein synthesis quantitatively. Deterministic rate equation based models are a standard approach for modeling this kind of biochemical

reactions. For example Karzbrun et al. presented a rate equation model that describes the initial rising phase of gene expression in a cytoplasmic *E. coli* extract [88]. Their model uses ten free parameters for protein synthesis and degradation rates as well as concentrations of components. However, this model does not incorporate the late phase of cell-free protein synthesis when protein levels saturate. Such a saturation can arise either from a steady state between protein synthesis and degradation or from expiration of the synthesis reaction due to a lack of resources in the absence of protein degradation. For applications like high throughput screening where the yield of protein is of interest it is necessary to provide a model which covers also the late plateau phase.

In the following an ODE model is presented which consistently fits the entirety of the GFP synthesis kinetics in a reconstituted cell-free system. A set of transcription and translation measurements was used to calibrate the rate constants in the model to the cell-free system used. Knowledge of the rates enables the model to predictively describe protein synthesis dynamics.

4.2   Experimental results

Cell-free GFP synthesis was carried out in $25\,\mu l$ format in a spectrometer cuvette and the fluorescence signal was recorded as a function of time. This measurement format was chosen because it offers the necessary sensitivity and dynamic range for a quantitative evaluation of the data (for more details compare Sec. 3.3). The reconstituted cell-free system that was used here has the advantage that its composition is well controlled as only necessary components are included. This is especially important with respect to its lack of RNAses and proteases (compare sections 2.1.2 and 2.1.3).

GFP expression as a function of DNA   First the amount of synthesized GFP as a function of template DNA was measured. The cell-free system was run with DNA concentrations between $0.3\,pM$ and $30\,nM$, that is five orders of magnitude, and the kinetics of GFP synthesis was recorded as shown if Fig. 4.1. The GFP levels rise linearly in the first $3\,h$ and reach a plateau afterwards. The individual slopes as well as GFP yields are inherent to the respective template DNA concentration up to a concentration of about $1\,nM$. At this point the protein synthesis machinery is saturated and additional template is not processed as shown in Fig. 4.2. A maximum GFP synthesis rate of about $200\,nM/h$ and a maximum GFP yield of about $300\,nM$ was found.
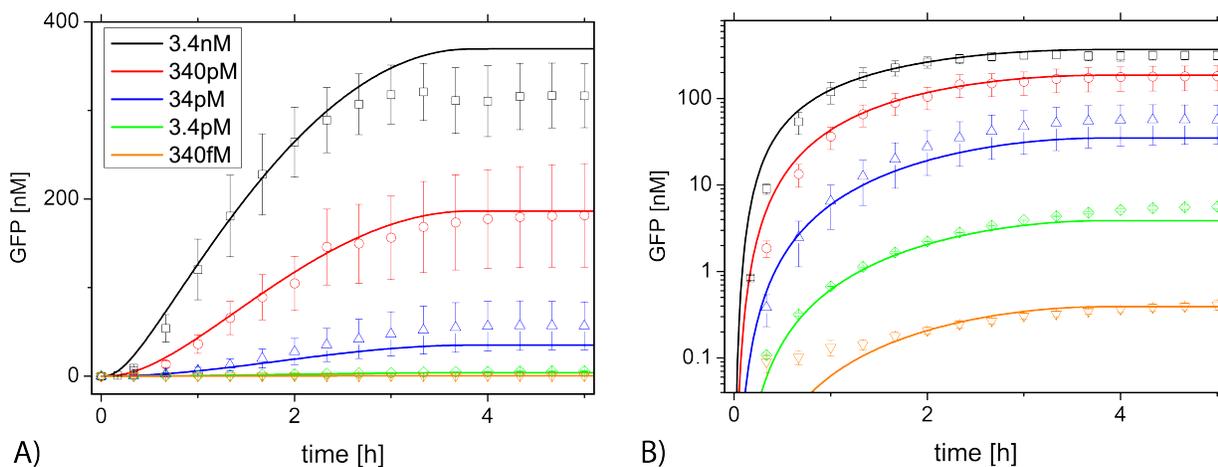
Figure 4.1: Cell-free synthesized GFP as a function of different plasmid DNA concentrations. A) shows the data in a linear graph and B) in semilogarithmic representation. With higher DNA concentrations GFP levels grow faster. Note that GFP levels start to saturate after about three hours in all cases, however at different plateau levels. The lines depict the best concurrent fit of the model to the data (see text for details).

The data in Fig. 4.1 show that the final GFP yield is a function of template DNA in the linear regime. However, the observed plateaus can not simply be explained by an expiration of the GFP synthesis reaction due to the exhaustion of some resource like NTPs or amino acids. Under this assumption the reaction should expire after the same amount of GFP had been synthesized regardless of DNA concentration. Less template should simply lead to a slower rise to this universal GFP level.

The different plateaus can also not be explained as a steady state between GFP synthesis and degradation like in the *in vivo* case described in chapter 5. As advertised by the manufacturer the PURE reconstituted cell-free system should be essentially free of RNAse and protease. Indeed in the measurements of GFP maturation described further below no GFP degradation was observed.

Expiration of cell-free GFP synthesis   Expiration at different GFP plateau levels can be explained by the hypothesis that some key component of the cell-free system like polymerases, ribosomes or NTPs degrades during the first $3\,h$ after initiation. The PURE system is stored as aliquots of two solutions "A" and "B" at $-80^\circ C$. As such it is stable for at least six month. Initiation of a measurement means that the two solutions are mixed, DNA is added and the temperature is set to $37^\circ C$. In order to check this hypothesis a timing measurement was carried out: aliquots of cell-free system were activated by mixing
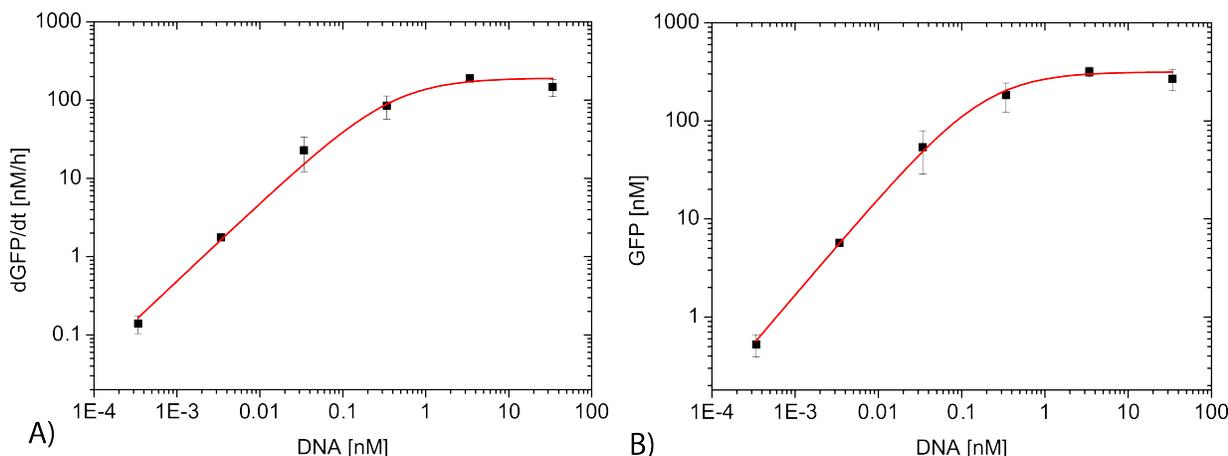
Figure 4.2: A) GFP synthesis rate and B) maximum yield as a function of template DNA. Higher concentrations of DNA template lead to faster synthesis as well as higher yield of GFP. However, at about $1\,nM$ DNA the cell-free machinery begins to saturate in a Michaelis-Menten like manner. The red curves depict a Michaelis-Menten fit to the data shown to guide the eye.

and heating. A fixed amount of DNA was added at different time points and the resulting kinetics of GFP synthesis were recorded (Fig. 4.3). With increasing delay time continuously lower yields of GFP are obtained. In agreement with the measurements depicted in Fig. 4.1 the cell-free system becomes inactive after a delay time of about $3\,h$.

In the following it was aimed to identify the exact component that is degraded. It has been shown that in cytoplasmic extracts NTP hydrolysis, accumulation of inorganic phosphate and degradation of enzymes can play a role in the expiration of protein synthesis [51, 89, 90]. The longer life time of the transcription step compared to the complete protein synthesis reaction shown further below in Fig. 4.5 implies that the sought after component is part of the translation step (GFP maturation does not involve components of the cell-free system). NTP hydrolysis was ruled out as the cause for the expiration of GFP synthesis as addition of fresh NTPs in $4\,mM$ concentration (about two times the concentration in the PURE system) in the plateau phase more than $3\,h$ after initiation did not restore protein production. However addition of $2.6\,\mu M$ fresh ribosomes ($2.4\,\mu M$ ribosomes are included in the PURE system) did indeed restore GFP synthesis. In this case the GFP level reached 180% of the usual yield (see Fig. 4.4).

Transcription rate   In the next step the transcription rate was measured individually. This required some sort of label for the mRNA. Karzbrun et al. [88] used radioactive
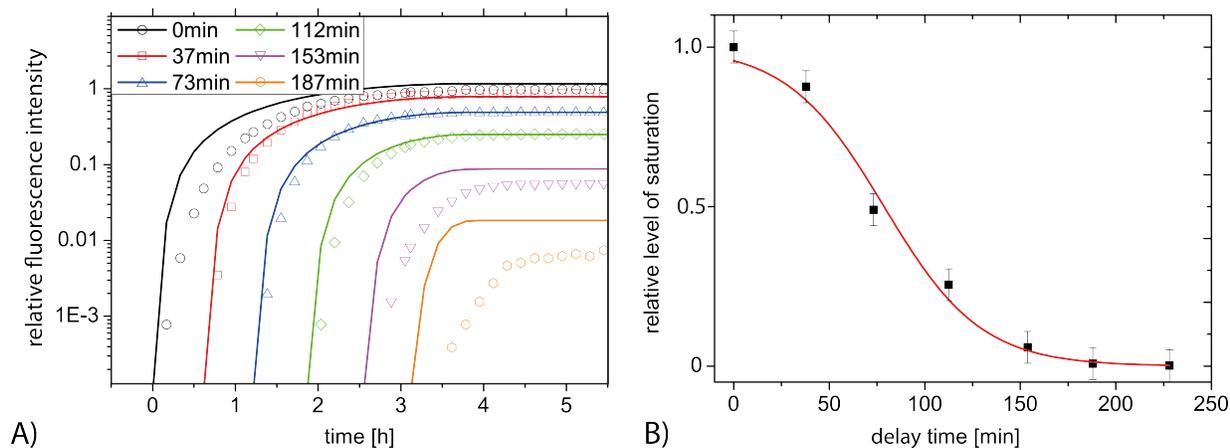
Figure 4.3: Unspecific degradation of the cell-free system leads to GFP plateaus as a function of DNA concentration (compare text). This effect was quantified as follows: a fixed amount of DNA was added to the cell-free system with different delay times. With growing delay time the final yield of GFP decreased (A). After about three hours delay time cell-free expression had ceased. The curves in (A) show the best concurrent fit of the model to the data (see text for details). The red curve in (B) was added to guide the eye.
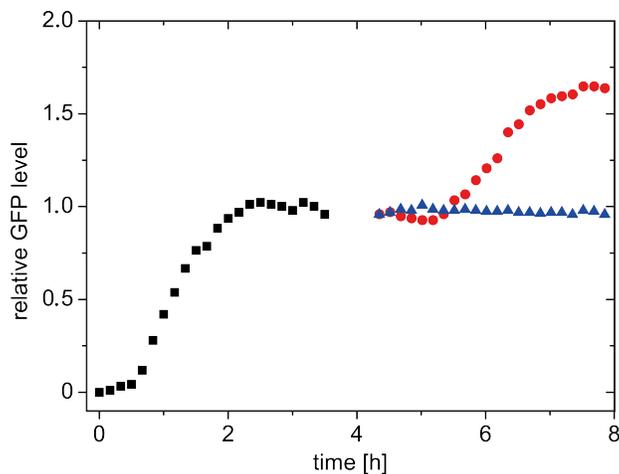
Figure 4.4: Measurements were undertaken to identify the exact limiting component of the cell-free system that causes the observed cessation of the GFP synthesis reaction after three hours: from the data shown in Fig. 4.5 it is known that cell-free transcription is not affected by this degradation of performance after three hours. Therefore a component of the translation step must be affected. After more than three hours of GFP expression (black squares) fresh resources were added to the sample. NTPs are a typical target of unspecific hydrolysis in cell-free systems. However, fresh NTPs did not restart GFP synthesis (blue triangles). On the other hand, fresh ribosomes did restart GFP production such that a final yield of 180% of GFP was achieved (red dots).

nucleotides for mRNA labeling in cytoplasmic *E. coli* extract. Osada et al. [91] applied
"Quant-iT RiboGreen" fluorescent RNA dye for RNA labeling in a "transcription only"
solution. However, radioactive labeling was not available and it was found that the Ri-
boGreen dye produced a high background noise in the cell-free transcription/translation
system which was attributed to fluorescent markers binding to tRNAs. This prevented
the non ambiguous detection of mRNA synthesis. Therefore another approach was chosen:
specific labeling of mRNA using molecular beacons (compare Sec. 3.5 for details). In the
presence of a complimentary strand - in this case the mRNA - molecular beacons uncurl
from the closed hairpin state and bind to the mRNA. This results in a fluorescence signal
of the FRET fluorophores. Cell-free mRNA synthesis measured this way is depicted in Fig.
4.5 A. The transcription life time depends on DNA concentration unlike translation which
universally expires after $3\,h$. At $6.8\,nM$ DNA transcription expires after about $6\,h$ whereas
at $1.7\,nM$ and lower concentrations it continues for over $8\,h$. Note that transcription there-
fore outlives cell-free translation. It is also interesting to note that mRNA levels in the
cell-free system exceed GFP levels (compare Fig. 4.5 A and 4.5 B). From this follows that
not all mRNAs are successfully translated into GFP. No significant mRNA degradation
was observed in the PURE system.

GFP maturation rate   GFP has to fold into its correct three-dimensional conformation
before its fluorophore becomes active (compare Sec. 2.4.2). This time limiting maturation
step was measured by inactivating the ribosomes in the first $3\,h$ of protein synthesis. An
increase of measured GFP fluorescence after this time point is due to maturation of already
synthesized protein (Fig. 4.6). This relaxation of GFP levels was fitted with an exponential.
The obtained GFP maturation time is $t_{mat} = 5.0 \pm 0.7\,min$ and the maturation rate is
therefore $0.2\,min^{-1}$. This value is in good agreement with a maturation time of $6.5\,min$
which was reported for GFP maturation in *E. coli* [92].

Looking at late times after ribosome inactivation no decrease of the GFP abundance
was observed meaning that protease activity was insignificant in the reconstituted cell-free
system.

Comparison with other cell-free systems   It is of interest to know if insights gained
from the PURE system can suitably be carried over to other cell-free systems as well.
Therefore GFP expression kinetics were measured in an eukaryotic extract (based on rabbit
blood) and in an *E. coli* extract (Fig. 4.7). Here GFP synthesis expired somewhat sooner
than in the reconstituted system, that is after about $2\,h$ compared to $3\,h$ and the final GFP
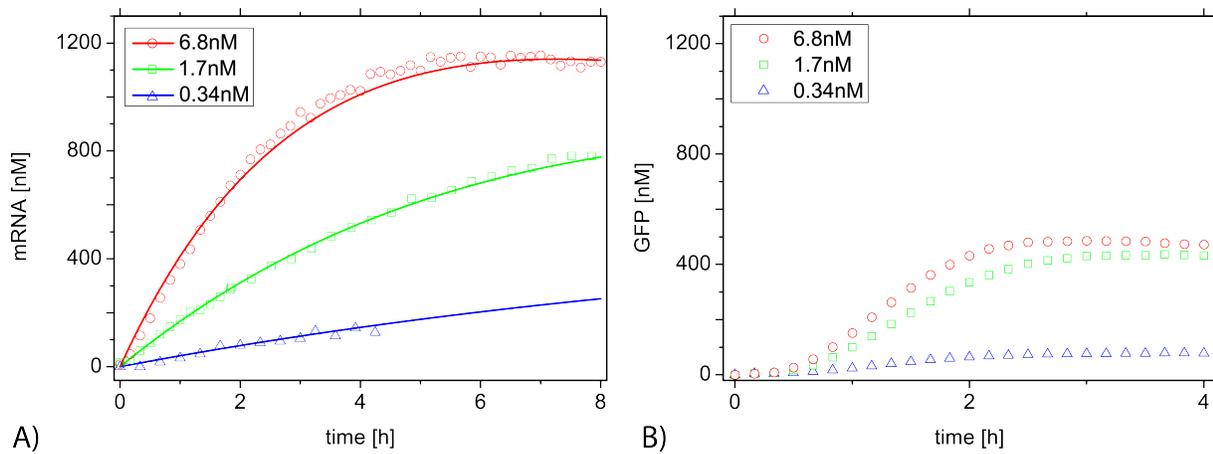
Figure 4.5: Transcription rate in the cell-free system measured individually: molecular beacons complementary to the mRNA were added to the sample. These beacons carry the fluorophore Cy5 which only fluoresces in the bound state because of quenching in the unbound state. A) depicts the kinetics of mRNA synthesis as a function of DNA concentration. In B) the corresponding curves of GFP synthesis are shown for comparison. As can be seen cell-free transcription lasts longer and reaches higher values than translation. Note also that in A) the time point at which transcription saturates depends on the DNA concentration: at higher concentrations mRNA levels saturate earlier. The lines in A) depict the best concurrent fit to the data according to the model (see text for details).
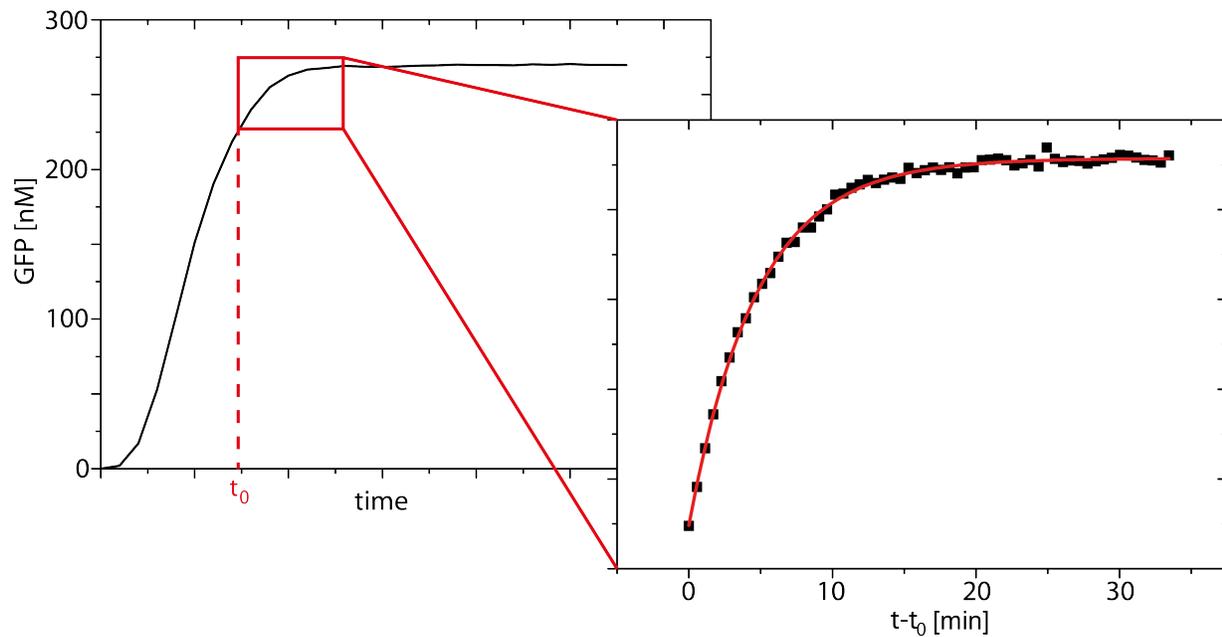
Figure 4.6: The GFP maturation rate measured individually: during the rising phase of cell-free GFP expression the antibiotic chloramphenicol was added to the sample. This inactivates the ribosomes and residual increase of GFP fluorescence after this time point is due to GFP maturation. This increase in fluorescence can be fitted with an exponential and the maturation rate be extracted (a cartoon of the measurement principle is shown on the left, actual data on the right).
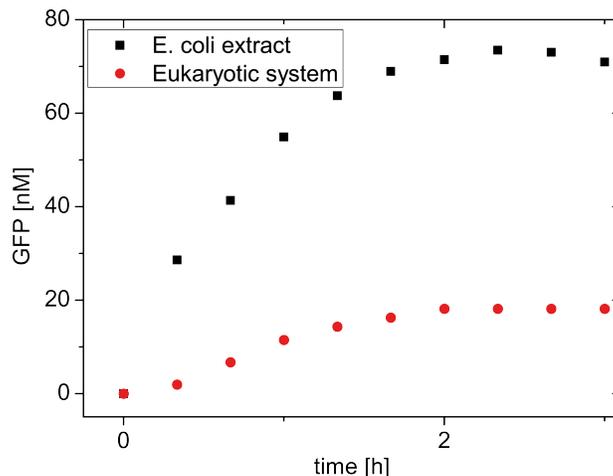
Figure 4.7: In order to verify the insights gained from studying the PURE cell-free system an *E. coli* based extract as well as an eukaryotic extract were quickly analyzed ($3.4\,nM$ of template DNA was used, compare with Fig. 4.1 A). It was found that overall the GFP synthesis kinetics of these extracts exhibit a similar shape compared to the PURE system albeit the GFP yield is somewhat lower and the reaction terminates already after two hours.

abundance was approximately one order of magnitude lower. Noticeably, GFP synthesis was not restarted by introducing fresh ribosomes in the plateau phase as it was the case with the PURE system.

4.3 Concurrent fit with a rate equation model

The objective in modeling was to develop a predictive model of cell-free transcription/translation. To this end a concurrent and quantitative fit to the calibration measurements shown in Figs. 4.1, 4.3 and 4.5 was created. A standard ODE based approach containing rate equations for transcription, translation and maturation was used. This scheme was complemented with finite resources for the transcription and translation step respectively (see Fig. 4.8). When developing the model the main focus was to keep the number of free parameters as small as possible while still concurrently fitting:

- early and late phase traces of GFP synthesis dependent on DNA concentration (Fig. 4.1)

- saturation of protein yield at about $1\,nM$ DNA concentration (Fig. 4.2)

- expiration of translation after $3\,h$ independent of DNA addition (Fig. 4.3)
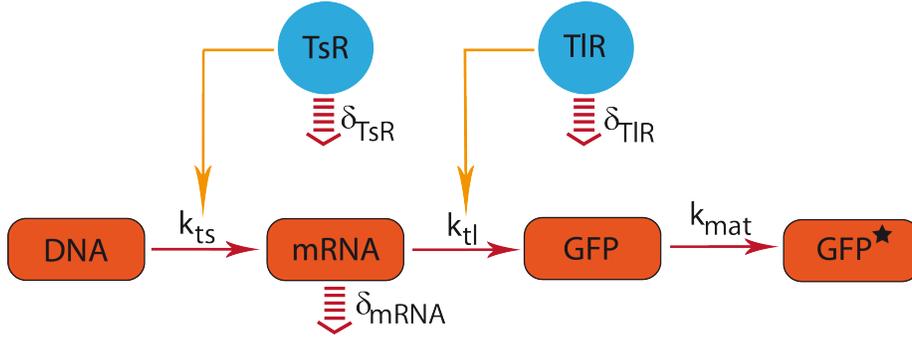
Figure 4.8: Rate model of cell-free GFP synthesis. Both transcription and translation consume energy and raw materials out of the finite pools of resources in the samples. The resources also degrade independently of the protein synthesis reaction. Degradation also limits the mRNA life time whereas GFP is stable over the course of one measurement.

These considerations led to the following model:

$$\frac{d}{dt}mRNA = \frac{k_{ts} \cdot TsR \cdot DNA}{K_s + DNA} - \delta_{mRNA} \cdot mRNA \tag{4.1}$$

$$\frac{d}{dt}GFP = \frac{k_{tl} \cdot TlR \cdot mRNA}{K_l + mRNA} - k_{mat} \cdot GFP \tag{4.2}$$

$$\frac{d}{dt}GFP^* = k_{mat} \cdot GFP \tag{4.3}$$

$$\frac{d}{dt}TsR = -\frac{kc_s \cdot TsR \cdot DNA}{K_s + DNA} \tag{4.4}$$

$$\frac{d}{dt}TlR = -\frac{\delta_{TlR} \cdot TlR}{K_{TlR} + TlR} \tag{4.5}$$

For transcription and translation Michaelis-Menten like equations were applied in order to account for the observed saturation effect. During model development mass-action kinetics as well as Hill functions were tested for transcription and translation. However, it was found that it is not possible to properly fit the GFP traces over all five orders of magnitude of DNA concentration with mass-action kinetics. That is only either low or high concentrations could be fitted satisfactorily but not both simultaneously. On the other hand Hill equations for transcription and translation produced a good fit to the entirety of the data. However, the optimized fits returned Hill coefficients very close to one which means that the Hill equations were basically reduced to the Michaelis-Menten like equations (4.1) and (4.2) shown above.

The prefactors TsR and TlR account for the observed expiration of GFP synthesis after $3\,h$. They stand for the resource pools of molecules of the transcription and trans-

Parameter	Model result	Previous studies	Ref.
k_{ts}	$18.2\,nM/min = 2.2\,NTP/s$	$250\,NTP/s$ (in vivo)	[93]
k_{tl}	$16.1\,nM/min = 0.03\,aa/s$	$1.8\,aa/s$	[20]
kc_s	$1.1 \cdot 10^{-2}$	n/a	
δ_{mRNA}	$7.8 \cdot 10^{-4}/min$	$8.3 \cdot 10^{-2}/min$	[88]
δ_{TlR}	$4.5 \cdot 10^{-3}/min$	n/a	
K_s	$8.5\,nM$	$0.8 - 4.2\,nM$	[94]
K_l	$65.8\,nM$	$14 - 30\,nM$	[95]
K_{TlR}	$6 \cdot 10^{-5}$	n/a	

Table 4.1: Optimized fit parameters and literature values given for comparison.

lation step respectively like polymerases, NTPs, ribosomes, tRNAs and so forth. This simplification is necessary since the exact concentration of the individual molecules is not known. This also means that meaningful initial values for TsR and TlR could not be deduced. They were therefore set to one and scaling parameters were installed. During model development two manners of consumption of TsR and TlR were studied: first, degradation independent of transcription and translation respectively and second, consumption by transcription/translation. Thus the preliminary rate equations for TsR and TlR were:

$$\frac{d}{dt}TsR = -\frac{kc_s \cdot TsR \cdot DNA}{K_s + DNA} - \frac{\delta_{TsR} \cdot TsR}{K_{TsR} + TsR} \tag{4.6}$$

$$\frac{d}{dt}TlR = -\frac{kc_l \cdot TlR \cdot mRNA}{K_l + mRNA} - \frac{\delta_{TlR} \cdot TlR}{K_{TlR} + TlR} \tag{4.7}$$

In either equation the first term represents consumption of the resources and the second term resource degradation. The parameters kc_s, δ_{TsR}, kc_l and δ_{TlR} were optimized simultaneously with the other fit parameters of the model. Thus the optimal value of δ_{TsR} was found to be close to zero after multiple optimization steps. Hence degradation of TsR independent of transcription was omitted from the final model. Analogously it was found that kc_l had its optimum close to zero and consumption of TlR was consequently omitted. Thus the rate equations for TsR and TlR were implemented into the final model in the form of equations (4.4) and (4.5). The time course of TsR and TlR exhaustion as implemented in the model is shown in Fig. 4.9.

By means of the model (eqs.(4.1)-(4.5)) a concurrent fit to the data of transcription and translation was created. First a numerical solution of the rate equations was fitted to the data using a downhill simplex algorithm followed by least squares minimization. The eight
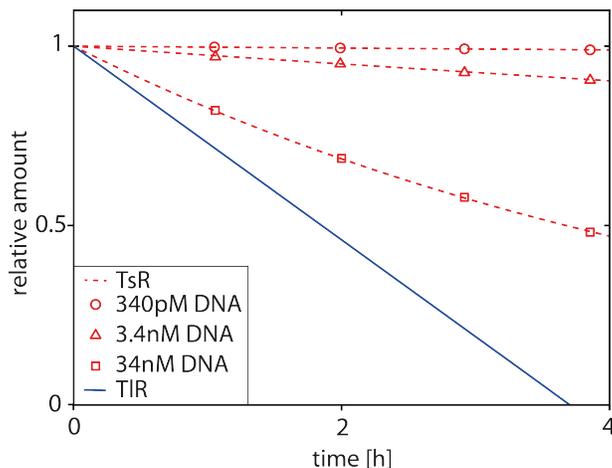
Figure 4.9: DNA dependent consumption of transcription resources TsR and DNA independent degradation of translation resources TlR as predicted by the model. Because the exact composition of the resources as well as their amount is not known the starting amount was set to one and scaling parameters were introduced.

free fit parameters of the model were weighted equally. Subsequently a Bayesian ensemble of the parameters was created using a Markov chain Monte Carlo method. The variance of each parameter value was found to be small indicating that a stable fit was reached. The optimum fit is represented by the solid lines in Figures 4.1, 4.3A, 4.5A (see also Tab. 4.1 for the optimum fit parameter values).

The transcription and translation rates are lower than their *in vivo* values. This observation is in agreement with data from Karzbrun et al. [88] who reported a similar drop for an *E. coli* based cell-free extract. Here the transcription rate of T7 polymerase is $2.2\,NTP/s$ while the *in vivo* value is $250\,NTP/s$. The translation rate of $0.03\,aa/s$ obtained from the fit is also lower than it has been originally estimated for the PURE system. The constants kc_s, δ_{TlR} and K_{TlR} which are associated with the notion of finite resources TsR and TlR have no counterpart in previous literature. δ_{mRNA} is two orders of magnitude lower than reported for a cytoplasmic *E. coli* cell-free extract. This is reasonable as the reconstituted system is essentially free of RNAses and proteases. Consequently no proteolytic degradation of GFP was observed. The obtained values for the Michaelis constants K_s and K_l are comparable to values found in previous literature.

4.4 Discussion

Cell-free systems were originally developed as a convenient means to synthesize proteins that are hard to produce in cells like toxic proteins. Nevertheless due to their comparative simplicity and ease of modification they also exhibit great potential as *in vitro* model platforms for transcription/translation studies. This however necessitates knowledge of kinetic parameters, expected protein yield, necessary template DNA concentration and reaction speed of the particular cell-free system which is generally not provided. Here a procedure of how to specify gene expression dynamics of a cell-free system through a combination of a set of calibration experiments and a mathematical model is established. The presented model concurrently fits the entirety of the data with one set of eight parameters. This was achieved by expanding the classical rate equation scheme as it is known from work *in vivo* (chapter 5) in two key points: first, a Michaelis-Menten like ansatz to account for saturation at high DNA concentrations and second, finite resource pools for transcription and translation. In this manner the model presented here adequately captures both the early rising phase and the late plateau phase of cell-free protein synthesis. It does so consistently over five orders of magnitude of DNA concentration with one set of fit parameters. After calibrating the model to the respective cell-free system used in an experiment the model predictively describes time courses of mRNA and protein synthesis as a function of template concentration and experimental timing. The fact that here a more general model (Hill formalism for protein synthesis and two different decay processes for transcription and translation resources) was reduced to a simpler form solely through parameter optimization of simulated values to experimental data is quite remarkable. In principle a more general formalism should result in a better fit because of the trade-off between model simplicity and fit quality. Here the experimental data seem to be well suited for discriminating the models and support the simplification. Note however that in the case of significant RNAse or protease activity in a particular cell-free system the more general expressions (4.6) and (4.7) for TsR and TlR time courses have to be tested.

   In many applications of cell-free systems a protein yield as high as possible is desired. It is therefore of great interest to know the exact cause of cessation of protein synthesis in a specific system. Through careful analysis of the data the translation step was identified as the bottle neck in the PURE system. It was found that in the reconstituted system accumulation of inorganic phosphates or NTP hydrolysis are not problematic as protein synthesis could simply be restarted by introducing fresh ribosomes to the sample. In this manner a protein yield of 180% of the yield without ribosome replenishment could be

achieved. This insight might prove to be valuable as a strategy for improved protein yield: it is more cost efficient to add fresh ribosomes to a sample with correct timing instead of doubling the sample volume. However, it was found that this is not necessarily true for any cell-free system. Fresh ribosomes did not restart GFP synthesis in two systems that were tested (based on rabbit blood and *E. coli* extract respectively).

The combination of a cell-free system with properly characterized transcription/translation kinetics and a predictive mathematical model is a valuable tool for quantitative studies of gene expression kinetics, regulatory networks, modifications of individual components or the specific effects of drugs. It should be noted that the measurements presented here were performed in bulk format (on the order of ten to twenty micro liters) meaning that even at the lowest concentration presented here on the order of 10^3 DNA or mRNA molecules were present in a sample. This is markedly more than the numbers found *in vivo* (using micro containers in a microfluidics setup these numbers could be downscaled to typical cellular values, compare the outlook). Furthermore, the components of cell-free systems compared to living cells are diluted by a factor of two and the composition is of course much less complex. Therefore, in the following chapter an artificial system is studied that is more alike to natural cells: the GFP expression kinetics of mRNA and pDNA transfected cells. Than in chapter 6 a modification is introduced into the cell-free system and the GFP expression kinetics of the modified system are compared to the unmodified one.

Chapter 5

Modeling of exogenous gene expression in eukaryotic cells

The content of this chapter is the subject of the manuscript [M1]. Flow cytometry and quantitative fluorescence microscopy measurements reported in this chapter were performed by Carolin Leonhardt and Gerlinde Schwake.

5.1 From *in vitro* to *in vivo* modeling of gene expression

Cell-free protein synthesis is overall an artificial process despite the fact that these systems are composed of components that are mostly part of the cellular gene expression machinery. It is therefore of great interest to verify that the modeling approach described in chapter 4 is also applicable *in vivo*. Transfection of cells is a natural choice for a rate equation based modeling approach. The expression of such exogenous genes is not controlled by the cellular gene regulation network. Their expression dynamics therefore follows a deterministic profile as it is described by a rate equation model. Furthermore, transfection of cells can be performed both with DNA as well as mRNA vectors. From a modeling perspective this is advantageous because it allows examination of the translation process independently of transcription.
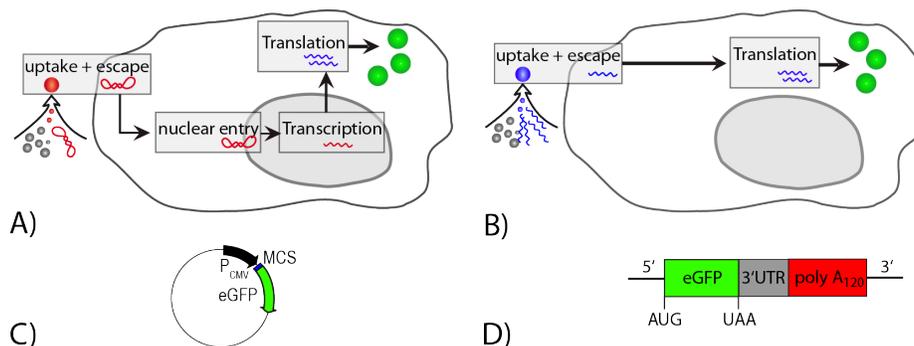
Figure 5.1: Transfection of an eukaryotic cell with exogenous genetic information using a cationic lipid agent. A) pDNA loaded vectors need to pass both the cell membrane as well as enter the nucleus to deliver their cargo. B) mRNA vectors on the other hand deliver their cargo directly into the cytosol. C) The pDNA vector used in this study carries the eGFP sequence under control of a CMV promoter. D) The mRNA was synthesized *in vitro* using a plasmid which was specifically designed with maximal possible analogy to the pDNA vector. The biggest difference between both mRNAs is that the *in vitro* mRNA carries an anti-reverse cap analog (ARCA) at its 5'-end as well as a poly(A) tail and an untranslated region (UTR) at its 3'-end. This enhances its translation rate as well as its stability against enzymatic degradation respectively.

5.2   Transfection of cells with synthetic pDNA and mRNA lipid vectors

Transfection of cells that is the introduction of an exogenous gene into the cell, is usually done with plasmid DNA (pDNA) as carrier of the genetic information. However, it is also possible to use mRNA instead which has certain advantages. In the case of eukaryotic cells a mRNA vector only has to deliver its cargo to the cytosol and does not need to enter the nucleus as shown in Fig. 5.1 (compare also Sec. 2.2). The probability of nuclear entry depends on the cell cycle, it is highest when the cell is close to mitosis. pDNA transfection is therefore inefficient for non-dividing cells whereas mRNA transfection is independent of the cell cycle [96–98]. Furthermore, with mRNA transfection there is no risk that the new genetic information is permanently integrated into the genome of the cell. A third advantage is that the immunoreaction of cells to mRNA transfection was observed to be less severe compared to pDNA transfection [99, 100].

Especially in the context of clinical applications gene delivery to eukaryotic cells remains a challenging task. Synthetic cationic lipid based vectors have been developed as a quantitative and predictive alternative to viral gene delivery. Mathematical modeling of

the statistical gene delivery process and the gene expression kinetics can help to further the understanding of transcriptome dynamics as well as RNA based devices.

Here transfection of a cell culture with pDNA and mRNA vectors (Fig. 5.1 C, D) respectively was quantitatively measured and compared. Identical preparation protocols were followed for both vector types in order to guarantee comparability of the results. Both pDNA and mRNA encoded eGFP so that fluorescence microscopy in combination with automated image analysis could be used to measure the gene expression kinetics at the single cell level. Arbitrary fluorescence units were converted into numbers of GFP per cell using the calibration standard described in Sec. 3.4.

Figure 5.2 A)-C) shows flow cytometry data of three different cell lines 22 hours after transfection (sidewards scatter vs. fluorescence signal). All three cell lines feature non-transfected as well as transfected cells after both pDNA and mRNA transfection. These populations correspond to the two different "clouds" of cells visible in each graph. pDNA transfection uniformly generates cells with higher eGFP fluorescence compared to mRNA transfection. That is the cloud of pDNA transfected cells is shifted farther to the right than the cloud of mRNA transfected ones. This effect can be more clearly seen in a histogram representation of the fluorescence data (Fig. 5.2 D)-F), pDNA data in red and mRNA data in blue).

Flow cytometry measurements offer the advantage that a great number of cells can be studied at once (here 15'000 cells per measurement) which yields good statistics. However, this comes at the prize that data are only taken at a specific point of time (here $22\,h$ after transfection). For this reason quantitative fluorescence microscopy was used as an additional technique in order to record single cell time traces of pDNA and mRNA transfected cells (see Fig. 5.3). These data more clearly disclose differences between pDNA and mRNA transfection than flow cytometry data. In Fig. 5.3 A, B an overlay of eGFP fluorescence and brightfield images of transfected cells is shown. The fraction of transfected cells was 61% for mRNA and 40% for pDNA transfection. In Fig. 5.3 C, D single cell time traces are depicted. These reveal that the onset time of GFP expression t_{on} after mRNA transfection is strongly centered around $2\,h \pm 1\,h$ after transfection whereas expression after pDNA transfection starts anywhere from $2\,h$ to $20\,h$ after transfection. mRNA transfection yields a steady rise in eGFP fluorescence levels until a maximum is reached and the fluorescence levels begin to decline due to mRNA degradation. pDNA transfection on the other hand yields sigmoidal shaped fluorescence time courses which result in a steady state of cellular eGFP levels.
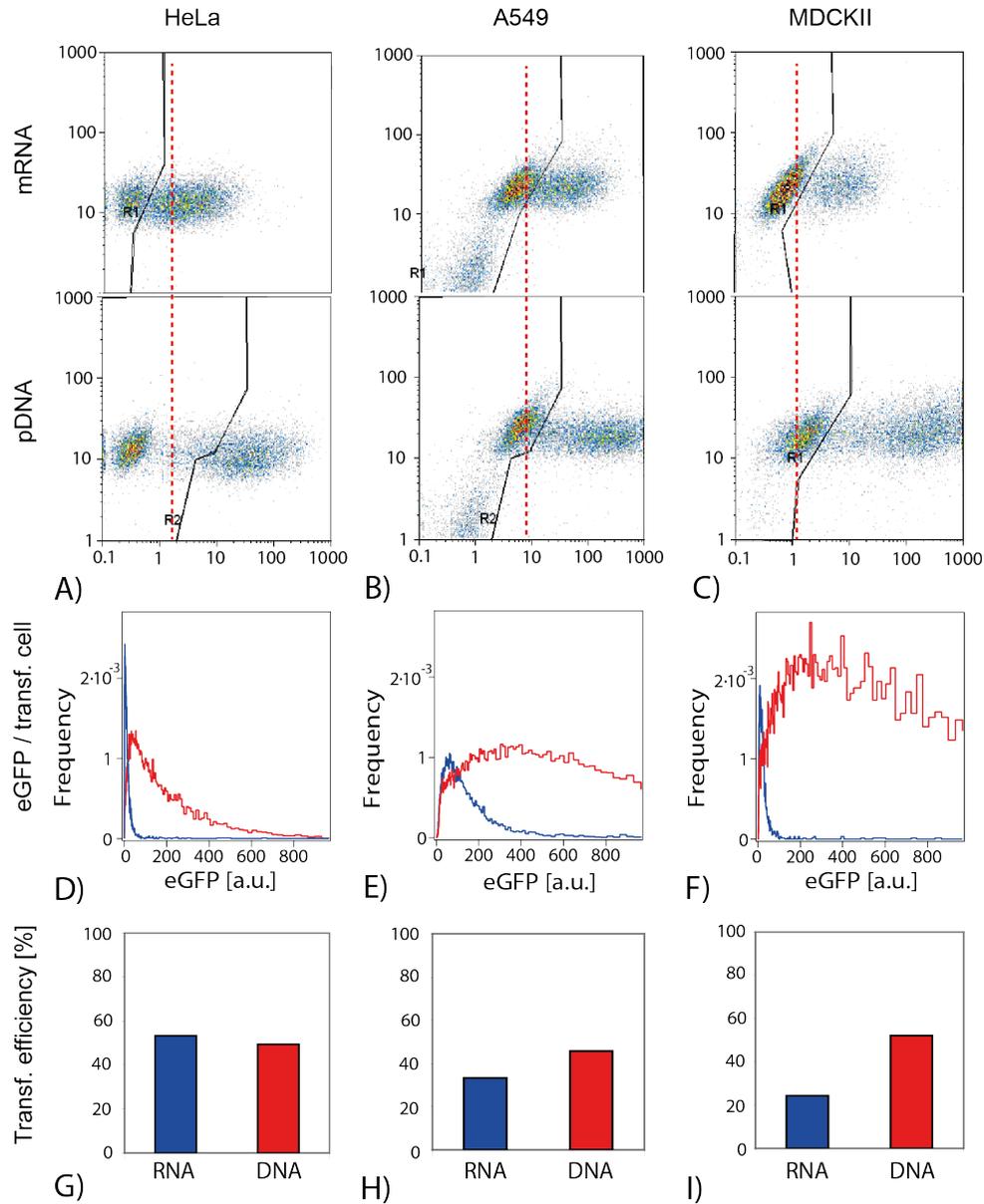
Figure 5.2: A)-C) show scatter plots of three different cell lines created from flow cytometry data of pDNA and mRNA transfection (sidewards scatter vs. fluorescence signal). The two distinct "clouds" of cells in each graph correspond to non-transfected cells and transfected cells respectively. The red line indicates the corresponding signal of a fluorescence standard at the various gain settings. D)-F) show histograms of fluorescence data. pDNA transfection (red) clearly generates higher fluorescence per transfected cell than mRNA transfection (blue). G)-I) Comparison of the percentage of transfected cells of pDNA and mRNA transfection for the various cell lines.
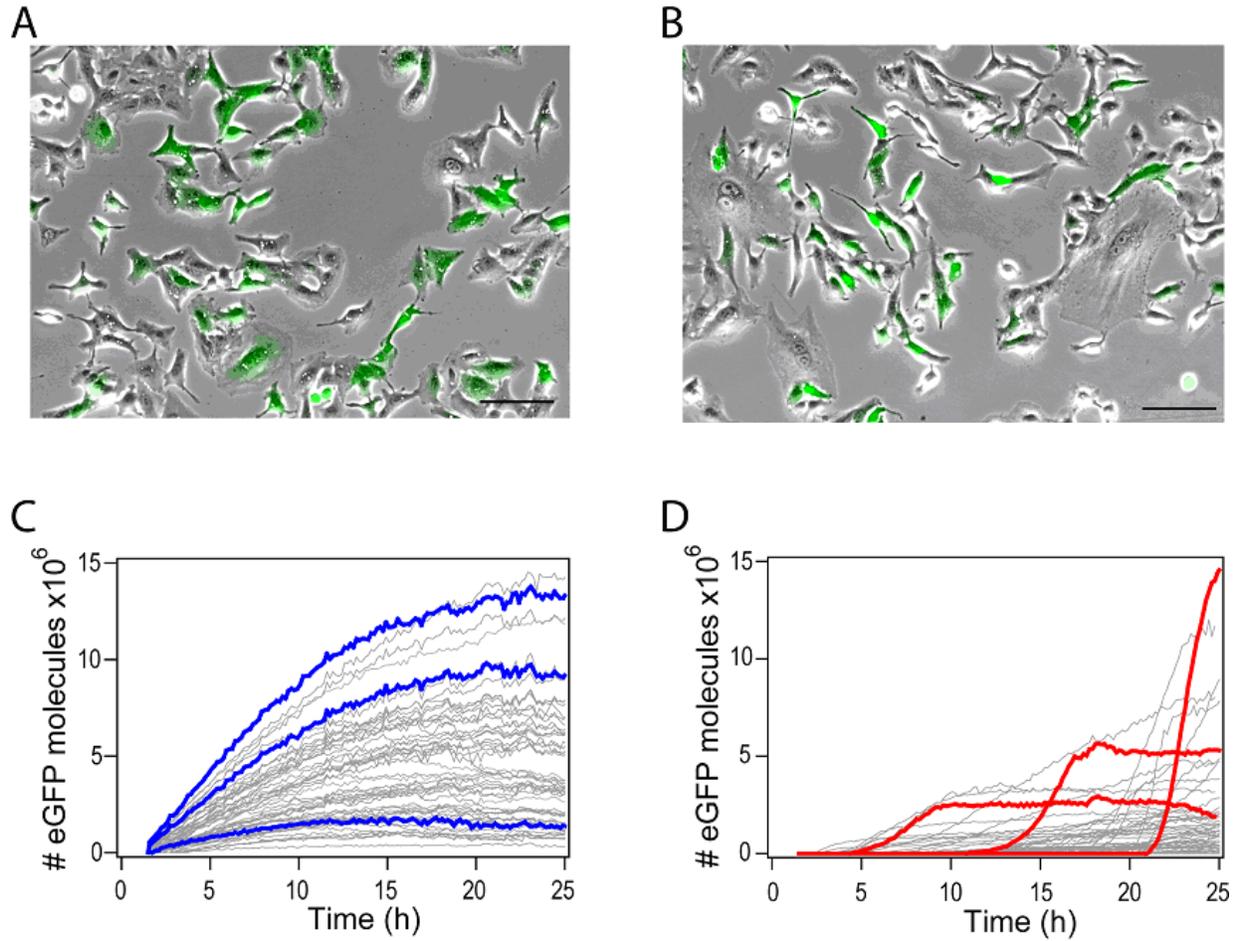
Figure 5.3: A) and B) show an overlay of bright field and fluorescence images of mRNA (A) and pDNA (B) transfected cells about $25\,h$ after transfection. Scale bars are $100\,\mu m$. C), D) The GFP expression kinetics of pDNA and mRNA transfected cells respectively show clearly distinct behavior. mRNA transfected cells uniformly start expression at a reasonably well defined time point about $2\,h$ after transfection. The onset time of expression in pDNA transfected cells on the other hand exhibits a considerably wider distribution. Note also that the maximal slope of pDNA transfected cells is markedly higher than in the case of mRNA transfection.

Rates in h^{-1}	mRNA vector	pDNA vector	Ref.
Transcription rate k_{TS}	n/a	180	[56, 101]
Translation rate k_{TL}	180	100	[25, 56]
mRNA degradation rate δ	-	0.1	[56, 102]
GFP degradation rate β	0.035	0.035	[56, 102]

Table 5.1: Literature values for protein production and degradation rates. The mRNA vector was designed in the framework of this study therefore no literature value is available for its degradation rate. The 1.8 times higher translation rate of the *in vitro* mRNA compared to the *in vivo* mRNA is due to its ARCA cap.

5.3   Analytical expression of GFP synthesis

The synthesis of eGFP in transfected cells can be described using a differential equation based model with continuous rates for transcription, translation, maturation and degradation. Literature values of these rates are depicted in Tab. 5.1. The *in vitro* transcribed mRNA that was used in the mRNA vector carries a cap analog on the $5'$ end which increases the translational efficiency [103]. The *in vitro* mRNA is also more stable against degradation therefore the literature value for δ cannot be applied. The simplest mathematical model of eGFP synthesis using ordinary differential equations (ODEs) looks as follows:

$$\frac{d}{dt}m = k_{TS}P - \delta m \tag{5.1}$$

$$\frac{d}{dt}G = k_{TL}m - \beta G \tag{5.2}$$

Here m, P and G denote the number of mRNA, pDNA and eGFP respectively. Note that degradation of pDNA is not included in the model as pDNA is expected to be stable over the course of one measurement (measurement time is up to $25\,h$) [104]. Furthermore, a eGFP maturation term was omitted from the model in order to keep the mathematical expression as simple as possible. This was validated by fitting exemplary traces with a model containing the maturation step and the one without described here. The discrepancy in the resultant fit parameters was found to be negligible. This model can be solved analytically. Using the initial condition that at time point zero no mRNA and no eGFP has yet been synthesized that is $m(0) = G(0) = 0$ one obtains the following solution for

eGFP synthesis as a function of time for the case of pDNA transfection:

$$G_{DNA}(t) = \frac{k_{TS}k_{TL}P}{\beta\delta(\delta-\beta)} \left[\left(1 - e^{-\beta(t-t_{on})}\right)\delta + \left(-1 + e^{-\delta(t-t_{on})}\right)\beta \right] \qquad (5.3)$$

Note that in equation (5.3) the time axis was shifted by t_{on} which accounts for the onset time of eGFP expression. Equation (5.3) reaches a steady state between eGFP synthesis and degradation over long times due to the lack of pDNA degradation:

$$\lim_{t\to\infty} G_{DNA}(t) = \frac{k_{TS}k_{TL}P}{\beta\delta} \qquad (5.4)$$

In the case of mRNA transfection the transcription term in equation (5.1) can simply be omitted. Using the initial condition $m(0) = m_0$ and $G(0) = 0$ where m_0 denotes the number of mRNA molecules that were delivered to the cell the resulting expression for eGFP synthesis after mRNA transfection reads:

$$G_{mRNA}(t) = \frac{k_{TL}m_0}{\delta-\beta} \left[1 - e^{-(\delta-\beta)(t-t_{on})}\right] e^{-\beta(t-t_{on})} \qquad (5.5)$$

The initial slope of equation (5.5) correlates to the maximal GFP synthesis rate in the cell. It is given by:

$$G'(0) = k_{TL}m_0 \qquad (5.6)$$

5.4 Mathematical analysis of pDNA and mRNA vector expression kinetics

pDNA model   The expression (5.3) was applied to fit eGFP expression time traces of pDNA transfected cells. However, it was found that the dynamics could not be fitted using rates for protein production and degradation that are in accordance with the literature values listed in Tab. 5.1. Using fit parameters with values that are in accordance with previous literature the resulting model curve rises markedly shallower than the experimental data. This is depicted in Fig. 5.4 A. The red curve was created applying equation (5.3) with fixed parameter values from Tab. 5.1 and the mean value of pDNA molecules per nucleus for these curves obtained from equation (5.4). In order to fit the steep rise in GFP levels unrealistically high values for eGFP production and degradation rates would have

$k_{TS}[h^{-1}]$	1962
$k_{TL}[h^{-1}]$	1090
$\delta\ [h^{-1}]$	0.64
$\beta\ [h^{-1}]$	0.64
P	5
$\tau\ [h]$	9.8
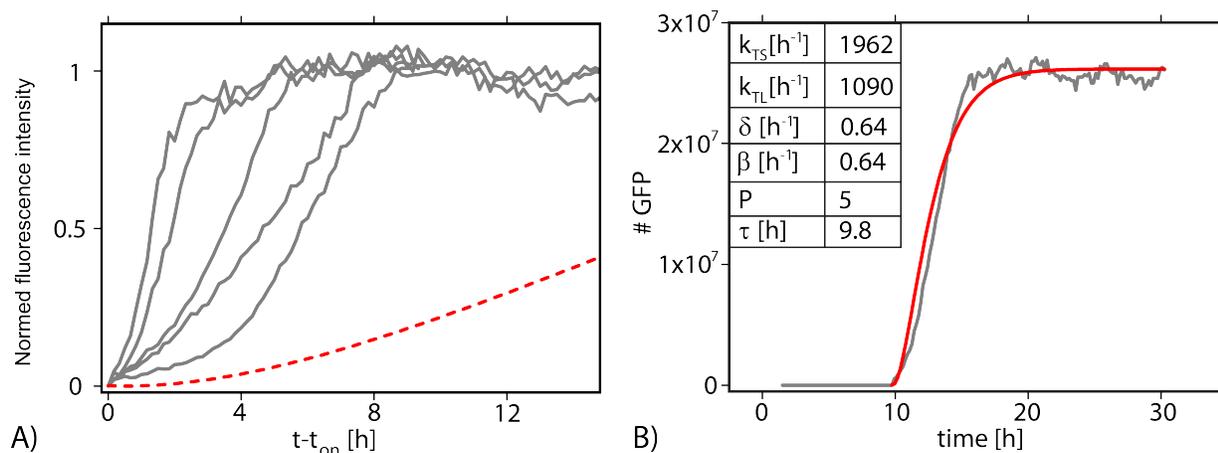
Figure 5.4: A) The pDNA model does not fit the data within the constraints of the literature values of Tab. 5.1. With these parameter values the steepness of the resulting curve is much too shallow to adequately fit the data. Note that the fluorescence levels were normalized and all curves were shifted by t_{on}. B) An acceptable fit as shown here can only be obtained using unrealistically high rates for GFP synthesis and degradation (about one to two orders of magnitude higher than literature values, compare the parameter values depicted in the box with the literature values in Tab. 5.1).

to be employed (about one to two orders of magnitude higher than literature values, see Fig. 5.4 B). However, despite its failure to accurately capture the early dynamics of GFP synthesis the expression for large t (equation (5.4)) can be used to fit the steady state between eGFP production and degradation and thus obtain the number of transfected pDNA molecules per cell. This number was found to be on the order of ten which is in good agreement with previously published data [56].

mRNA model Equation (5.5) was used to fit single cell traces of mRNA transfected cells. Unlike with pDNA mentioned above here the eGFP expression kinetics could be adequately fitted with parameter values that were in the established range of previous literature. Free fit parameters were the product of the translation rate and the number of transfected mRNA molecules $k_{TL} \cdot m_0$, the degradation rates for mRNA and eGFP δ and β as well as the onset time t_{on}. The mRNA model captures well both the initial rise of GFP levels and the late decrease due to degradation of mRNA and eGFP as shown by the exemplary fits depicted in Fig. 5.5 A. Normalizing the fluorescence levels and shifting by t_{on} reveals that the time courses of mRNA transfected cells collapse on one master curve whereas normalized and shifted pDNA curves show no such generic behavior (compare

Figures 5.4 A and 5.5 B).

The fit parameter distributions obtained by fitting traces of 283 mRNA transfected cells are shown in Fig. 5.5 C-F. In Fig. 5.5 C the distribution of t_{on} for pDNA transfected cells (in red) is given for comparison. The distributions of the degradation rates δ and β were fitted with a Gaussian[1]. This yields $\delta = 0.056 \pm 0.021 \, h^{-1}$ and $\beta = 0.051 \pm 0.022 \, h^{-1}$. This is slightly bigger than a value of $0.028 \, h^{-1}$ for δ reported previously [105]. The result is smaller than the degradation rate of endogenous mRNA $(0.1 \, h^{-1})$ which was to be expected due to the stabilizing cap that the *in vitro* mRNA carries. Note that in the model k_{TL} and m_0 are linked and therefore only the product of both appears as a fit parameter (Fig. 5.5 D). However, the number of mRNA molecules that are successfully transfected per cell would be of great interest. Since transfection is an inherently stochastic process it can be assumed that the greater part of the observed variance of $k_{TL} \cdot m_0$ is due to an underlying distribution of m_0 and not due to great variability of k_{TL}. Therefore, in the following section a stochastic mRNA delivery model will be developed to account for this variance.

5.5 Stochastic mRNA delivery model

The assumed mRNA delivery process to transfected cells is depicted in Fig. 5.6. mRNA and the lipid mixture lipofectamine form complexes with on average m mRNA molecules per complex. These complexes sediment on to the cells and are taken up via the endocytotic pathway with on average N endosomes per cell. The complexes then escape the endosomes with a lysis probability k and are unpacked with probability q. Afterwards the freed mRNA molecules are available for translation.

The number of endosomes as well as the number of complexes in an endosome are assumed to be small and to arise from stochastic and independent processes. Therefore, Poisson distributions are adopted for both. The probability to have N endosomes in a given cell is then described by:

$$P_N(k) = \frac{N^k}{k!} e^{-N} \tag{5.7}$$

Endosomes lyse with probability k. The probability distribution of N' lysed endosomes is a convolution of eq. (5.7) with a binomial with probability k since lysis can happen for

---

[1]The Gaussian or normal distribution is given by $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ where μ is the mean and σ the standard deviation.
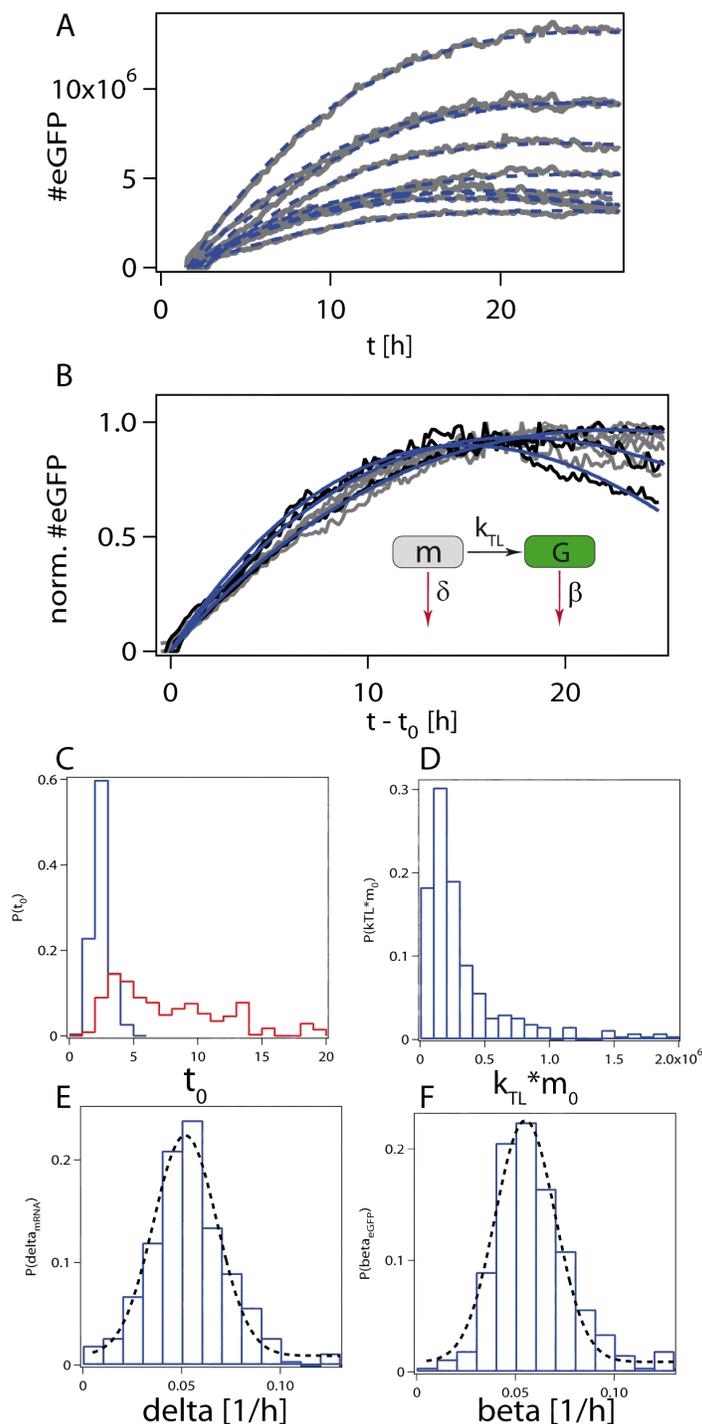
Figure 5.5: A) single cell traces of mRNA transfected cells; B) traces from A) shifted by t_{on} and normalized; C) distribution of t_{on} for pDNA (red) and mRNA (blue) transfected cells; D)-F) distributions of independent fit parameters of 283 analyzed A549 single cell traces (mRNA transfection). Refer to the text for details.
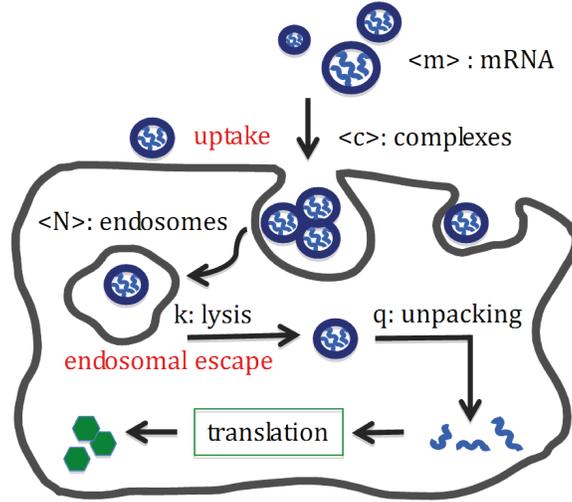
Figure 5.6: It is assumed that lipoplexes are taken up by the cell via the endocytotic pathway. First, lipoplexes sediment on to the cell membrane. Then the membrane invaginates around these lipoplexes and is chocked off creating an endosome with the lipoplexes inside. Subsequently the endosome lyses and the mRNA molecules are released into the cytosol. In the context of the stochastic delivery model described here it is assumed that the average number of endosomes per cell N as well as the average number of lipoplexes per endosome c are small. Therefore, N and c follow Poisson distributions.

all $N \geq N'$:

$$P(k|N', N) = \binom{N}{N'} k^{N'} (1 - k)^{N-N'} \cdot P_N(k) \tag{5.8}$$

Equation (5.8) can simply be reduced to a Poissonian with the parameter Nk. As motivated above the probability distribution of complexes per endosome is also a Poissonian with parameter c. The probability to have c' complexes from lysed endosomes in a cell is then a convolution of both processes. Using the fact that a convolution of N' Poisson distribution with parameter c is itself a Poisson distribution with parameter $N'c$ and accounting for all possible values of N' yields:

$$P(c') = \sum_{N'=0}^{\infty} \frac{(Nk)^{N'}}{N'!} e^{-Nk} \frac{(N'c)^{c'}}{c'!} e^{-N'c} \tag{5.9}$$

In the above equation it is implied that all complexes are identical which is an unrealistic assumption. One would rather expect that the complex size and therefore mRNA payload to follow a normal distribution. This means a convolution of equation (5.9) with a normal
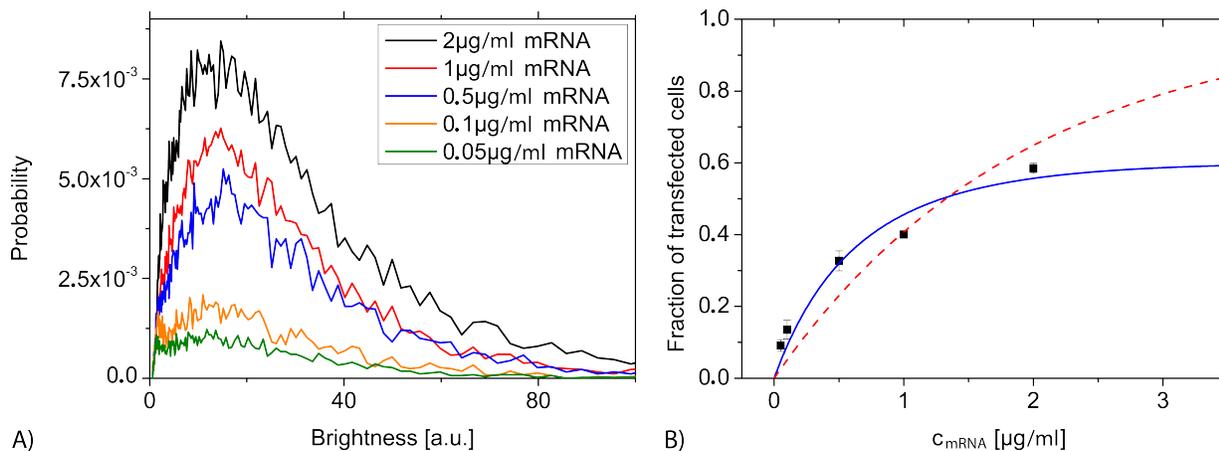
Figure 5.7: A) Probability distribution of mRNA transfected A549 cells at various mRNA concentrations generated from flow cytometry data. At higher concentrations the fraction of transfected cells increases but not the average fluorescence per cell (the position of the peak does not shift to the right). B) Fraction of transfected cells generated by integrating the data from (A). A stochastic mRNA delivery model with only one Poisson process does not satisfactorily fit the data (dashed red line). However, the model with two independent consecutive Poisson processes (equation (5.10)) presented in the text does (blue line). Here the fit parameters are $N' = 0.9 \pm 0.2$ and $c' = 1.1 \pm 0.5$.

distribution.

A further quantity of interest is the transfection ratio, that is the percentage of transfected cells at a given concentration of lipoplexes (that is mRNA) in the cell culture. This is the probability that in a cell at least one lipoplex is released from the endosomes (note that here the unpacking probability of complexes q in Fig. 5.6 is assumed to equal one for simplicities sake). The transfection ratio is obtained by summing over eq. (5.9) from $c' = 1$ to infinity:

$$P(c' \geq 1) = 1 - \exp[N'(e^{-c'} - 1)] \tag{5.10}$$

In Fig. 5.7 A the probability distributions of A549 cells transfected with various concentrations of mRNA are depicted. The data were recorded using flow cytometry. Higher concentrations of mRNA increase the number of transfected cells. However, interestingly the average brightness of the cells does not increase (that is, the peak in Fig. 5.7 A is not shifted to the right at higher mRNA concentrations). The dose response relationship between mRNA concentration and fraction of transfected cells is depicted in Fig. 5.7 B. It was obtained by integrating the probability distributions in Fig. 5.7 A. The fact that only
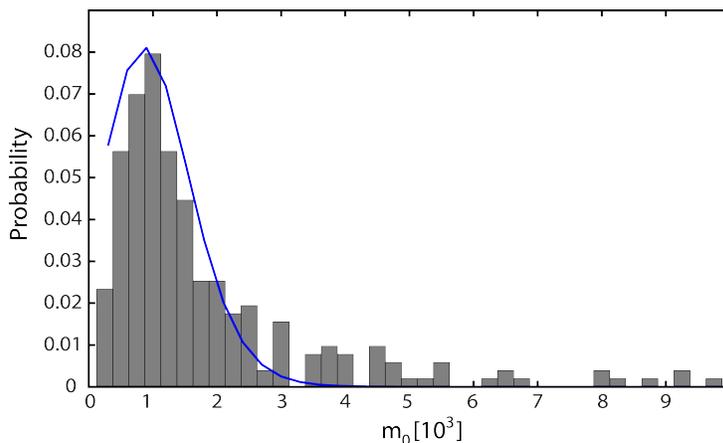
Figure 5.8: Distribution of mRNA molecules per transfected cell m_0 obtained by dividing the distribution of $k_{TL}m_0$ in Fig. 5.5 D by the literature value $k_{TL} = 180\,h^{-1}$. The fit (blue line) was obtained by inserting $N' = 0.9$ and $c' = 1.1$ into equation (5.9) and convolving with a normal distribution. This is necessary to account for a distribution in complex size and therefore mRNA payload. The fit parameters are the standard deviation $\sigma = 2$ and the average number of mRNA molecules per complex $N_{mRNA} = 300$.

up to 60% of all cells are transfected indicates that some bottleneck in the transfection process limits this number. The simplest model of such a bottleneck would be a single Bernoulli process[2] with a low probability of success. The resulting probability distribution would then be a Poissonian and the dose response relationship would therefore feature a simple exponential increase. This case is depicted with the doted red line in Fig. 5.7 B. Obviously, this model does not yield an acceptable fit to the data. An acceptable fit however can be obtained by inwoking equation (5.10) of the double Poissonian model discussed above (blue line in Fig. 5.7 B). The fit returns the average number of lysed endosomes $N' = 0.9 \pm 0.2$ as well as the average number of unpacked complexes $c' = 1.1 \pm 0.5$.

Using equation (5.9) the model can furthermore reproduce the probability distribution of successfully transfected mRNA molecules per cell m_0. In Fig. 5.8 the distribution of m_0 of 283 A549 cells obtained with quantitative fluorescence microscopy and image analysis is depicted. Note that these are the same data as shown previously in Fig. 5.5 D albeit divided by $k_{TL} = 180\,h^{-1}$ (see Tab. 5.1). When comparing these data with the predicted distribution of equation (5.9) with $N' = 0.9$ and $c' = 1.1$ it was found that the predicted distribution is narrower than the one obtained from single cell data. However, if one assumes that the complex size follows a normal distribution one obtains the standard

---

[2]In probability theory a Bernoulli process is a discrete stochastic process which has only two possible states – success and failure.

deviation σ as a fit parameter that broadens the distribution. The second fit parameter is the average number of mRNA molecules per complex N_{mRNA}. The fit depicted in Fig. 5.8 was obtained with $\sigma = 2$ and $N_{mRNA} = 300$.

5.6   Analysis of mRNA and pDNA vectors with FCS

In this section it shall be explained that the mRNA package size of 300 mRNA molecules obtained above is reasonable. To this end the average size of the lipoplexes was measured with FCS. Subsequently, an estimation of mRNA payload based on charge neutrality between negatively charged nucleotides and positively charged lipid analogues is presented.

mRNA was complexed with lipofectamine according to the same protocol as with transfection experiments with the modification that Sybr Gold was added as a fluorescent tag. After $30\,min$ of incubation at room temperature the average hydrodynamic radius of the lipoplexes was determined with FCS. The hydrodynamic radius of free mRNA was measured for comparison (see Fig. 5.9 A, B). For the formulas used here compare section 3.2. The average hydrodynamic radius of mRNA lipoplexes was $46\,nm - 70\,nm$ ($8\,nm$ for free mRNA). The same analysis was performed for pDNA lipoplexes and similar values were measured meaning that mRNA and pDNA lipoplexes have on average the same size.[3]

For simplicity's sake it is assumed that the lipoplexes have a cubic shape with an edge length twice the hydrodynamic radius. The concentration of cationic lipid analog DOTAP in the commercial transfection medium Lipofectamine is not known but typical transfection media have a concentration of about 50%. Therefore, this value is also used here. Previous studies established that cationic lipid/nucleic acid complexes exhibit alternating layers of lipids and nucleic acids with a repeat distance of $65\,\text{Å}$ as shown in Fig. 5.9 C [106]. Furthermore, the surface area of two phospholipids' head groups is about $140\,\text{Å}^2$ and DOTAP carries a single positive charge. From the requirement for charge neutrality of the complexes it follows that the number of nucleotides (one negative charge) equals the number of DOTAP molecules. The *in vitro* mRNA used here is about 1200 nucleotides long therefore a complex carries on average

$$m = \frac{2d^3}{65 \cdot 140\,\text{Å}^3 \cdot 1200} = 130 - 510 \tag{5.11}$$

---

[3]Note that FCS can also be used to measure the concentration difference between two samples. Here the difference between free mRNA and lipoplexes is $17\,nM/1.1\,nM = 15$ which would imply that lipoplexes carry on average 15 mRNA molecules. This however is misleading because such low sample concentrations are not in the linear regime of FCS.
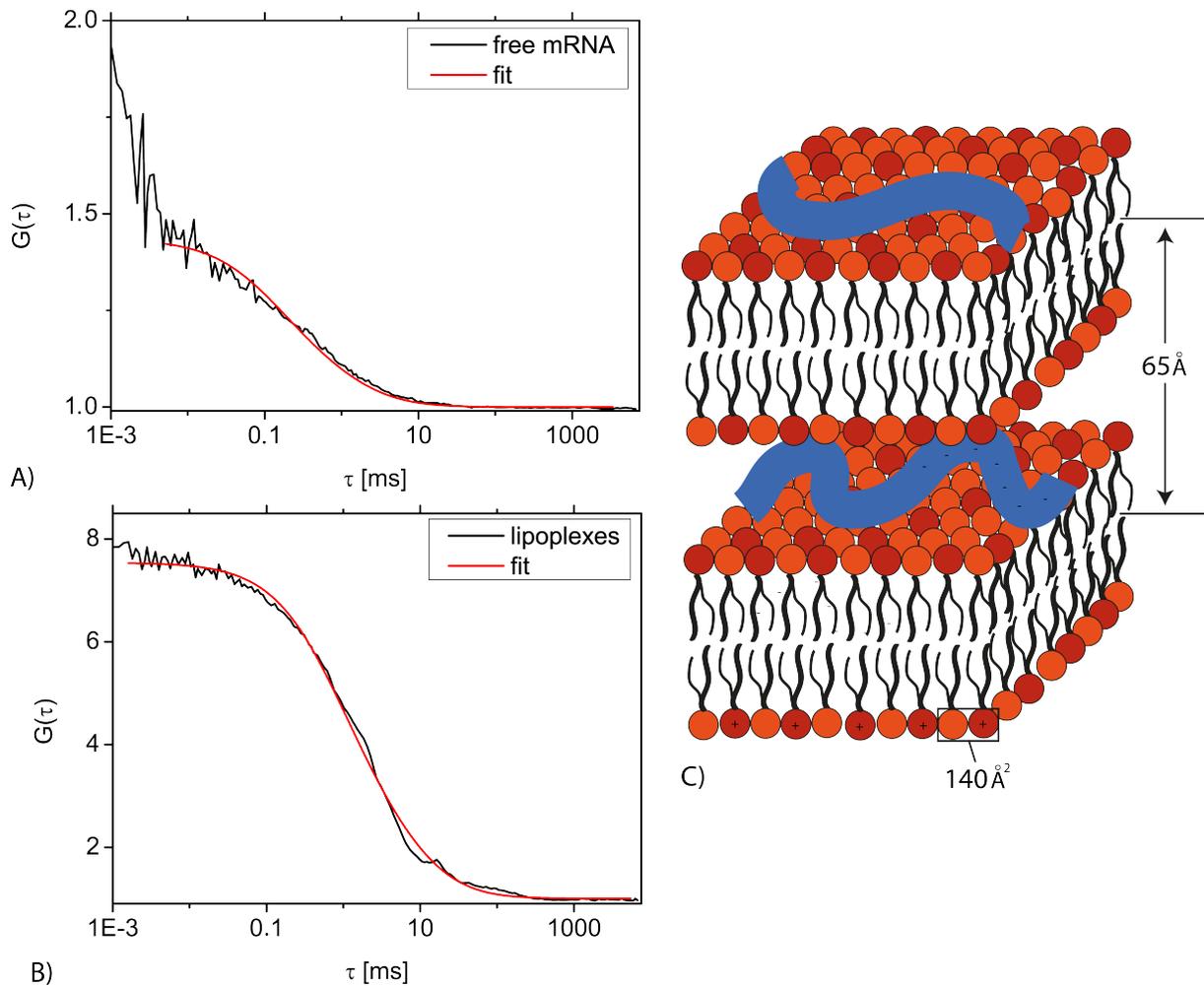
Figure 5.9: A) Autocorrelation curve of freely diffusing mRNA labeled with Sybr Gold. B) Autocorrelation curve after complexation with lipofectamine. The hydrodynamic radius of the complexes is $46\,nm - 70\,nm$ ($8\,nm$ for free mRNA). C) Cationic liposomes that are complexed with nucleic acids assume a shape which exhibits alternating layers of nucleic acids and lipid bilayers with a repeat distance of $65\,\text{Å}$. Charge neutrality of a complex demands that the single negative charge of each nucleotide is compensated by the single positive charge of a DOTAP head group. Here it is assumed that 50% of all lipids in lipofectamine are charged therefore the surface area of one positive charge is $140\,\text{Å}^2$.

mRNA molecules depending on its size. Here d is the edge length of the complex and the factor two stems from the fact both leaflets of the lipid bilayers need to be considered. Therefore the value of 300 mRNA molecules per lipoplex obtained from the fit in Fig. 5.8 seems reasonable.

5.7 Discussion

Various cell lines were transfected with pDNA and mRNA. Universal behavior was found for both transfection systems: onset times of pDNA transfected cells were widely distributed and eGFP expression kinetics did not fall upon a master curve after normalization and shifting by t_{on}. Onset times of mRNA transfected cells on the other hand were narrowly centered at two hours after transfection and expression kinetics followed a master curve.

A simple rate equation model for GFP synthesis of pDNA and mRNA transfected cells was developed. The model captures the dynamics in the case of mRNA transfection accurately but it fails to reproduce the steep rise of GFP levels in pDNA transfection when using meaningful fit parameter values. On the other hand in chapter 4 it is shown that this ODE model is fully capable of reproducing the kinetics of GFP transcription and translation in a cell-free system. It can therefore be theorized that the discrepancy to the data in living cells stems from the time delay that arises from the mRNA leaving the nucleus prior to translation initiation as well as some form of cooperativity effects which are not implemented in the model. A comparison of the kinetic rate constants for the cell-free system studied in chapter 4 and the transfected A549 cells is given in Tab. 5.2. The fact that the dynamics of mRNA translation are well captured shows that the ansatz of using a deterministic ODE model is in principle valid for this kind of question. Further investigation will have to be made in order to reveal the exact detail in which the current model oversimplifies the transcription step. In this regard a simplified model system of the gene expression mechanism in living cells like it is realized by cell-free systems may be an invaluable tool. For example Noireaux et al. showed how cell-free systems can be enclosed in cell like compartments. They studied compartmentalized gene expression in a cell-free system by encapsulating the system in lipid vesicles as well as by immobilizing DNA in dense brushes on a substrate [7, 107].

The mRNA model described here is a convenient tool for the determination of the number of mRNAs taken up by a cell as well as the mRNA degradation rate. It proved to be robust in the fitting procedure, that is no adjustment of initial values of the fit

	cell-free system	A549 cells
transcription rate	$11\,h^{-1}$	$180\,h^{-1}$
translation rate	$1\,h^{-1}$	$100\,h^{-1}$
mRNA degradation rate	$4.7 \cdot 10^{-2}\,h^{-1}$	$5.6 \cdot 10^{-2}\,h^{-1}$
GFP degradation rate	0	$5.1 \cdot 10^{-2}\,h^{-1}$
DNA/mRNA	$10^{-1} - 10^{4}$	10^{2}

Table 5.2: Comparison of kinetic parameters and DNA/mRNA content of PURExpress cell-free system and A549 cells obtained from data fitting as described in the text. Polymerases and ribosomes work one to two orders of magnitude slower *in vitro* than they do *in vivo*. For the estimation of the number of DNA/mRNA molecules in cell-free system the volume of a A549 cell was estimated to be $10\,pl$.

parameters or constraint of the fit interval was necessary for individual GFP traces. It can therefore be readily incorporated into automated data analysis. This way the model may prove useful for quantitative studies of translation such as in gene silencing or epigenetics.

Furthermore, a stochastic model for the delivery of lipid/mRNA complexes to the cytoplasm of a cell was developed. The model assumes that there are two bottleneck processes in the delivery: the number of complexes taken up by one endosome and the number of endosomes that lyse. This model with two independent consecutive Poisson processes fits the transfection probability as a function of dose unlike a simpler model with only one bottleneck process. It should be a convenient tool for the quick calculation of the expected fraction of transfected cells when planning mRNA transfection measurements.

Chapter 6

Cell-free gene expression in heavy water

The content of this chapter is the subject of manuscript [M2].

6.1 Effects of heavy water on gene expression

So far a mathematical rate equation model of gene expression has been presented. This modeling ansatz adequately reproduces protein synthesis *in vitro*. An adaption of this ansatz was applied to an *in vivo* system, albeit here the transcription step had to be omitted in order to receive satisfying results from the model. It is further of interest how the behavior of the *in vitro* system changes when a small disturbance is introduced. Using deuterium oxide instead of normal water falls into this category because its physical properties are slightly different from ordinary H_2O (Fig. 6.1 A, see also section 2.5 for details). Especially interesting about heavy water as a disturbance is that it affects all steps of the gene expression process. The sum of these individual disturbances is then reflected in the overall variation of protein synthesis (Fig. 6.1 B).

The differences between H_2O and D_2O can be classified into two categories: the "solvent isotope effect" (SIE) and the "kinetic isotope effect" (KIE) [108]. The SIE is due to the properties of D_2O as a solvent, i. e. its higher viscosity and density at $20°C$ compared to H_2O. The KIE is due to its slightly different chemical bonds. This effect is much more pronounced in hydrogen than in other elements because its isotope deuterium is twice as heavy as protium. In other elements the mass difference is only a small percentage of the element's mass.
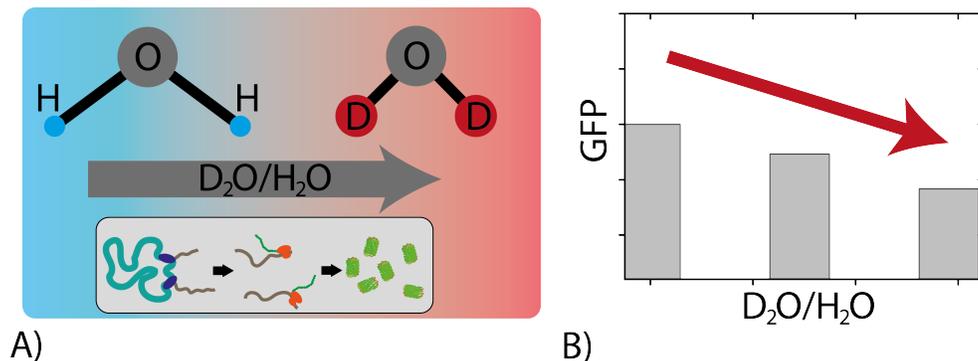
Figure 6.1: A) Deuterated water has slightly different physical properties from normal water. This affects all steps of protein synthesis: transcription, translation and maturation. B) The overall effect of deuteration on GFP synthesis arises from the combination of the effects on the individual steps.

The potentials of protium bonds and deuterium bonds have basically the same shape. The fundamental vibrational frequency of such chemical bonds follows from Hook's law and is given by

$$\nu = \frac{1}{2\pi c}\sqrt{\frac{k}{\mu}} \tag{6.1}$$

where k is the force constant and μ the reduced mass [76]. According to quantum mechanics the n-th energy state of a harmonic oscillator is given by

$$E_n = h\nu\left(n + \frac{1}{2}\right). \tag{6.2}$$

From this follows that deuterium bonds have lower zero-point energy and conversely require a higher activation energy in order to be broken. This energy difference is about $1.2 - 1.5\,kcal/mol$. Note that even at room temperature about 99% of all bonds are in the zero-point state. With this it can be shown that the KIE for deuterium and protium is with good approximation:

$$\frac{k_H}{k_D} = e^{(h\nu_H - h\nu_D)/2RT} \tag{6.3}$$

where k_H and k_D are reaction rates and R is the gas constant. At room temperature $(25°C)$ this results in $kH/kD \approx 7$ for C–H bonds [76].

From these observations immediately the question arises how living organisms are affected when exposed to deuterium oxide. Gilbert Lewis was the first to successfully purify deuterium oxide in 1933 and also the first to study its biological effects [75, 109, 110]. He

found that tobacco seeds did not sprout at all in samples containing pure deuterium oxide and respectively sprout considerably slower in samples containing 50% deuterium oxide than in light water. Mice and rats that were given deuterium oxide as drinking water died after seven days when the D_2O concentration in their body reached 50% and 90% D_2O was reported to be lethal for fish, tadpoles and drosophila [108, 111–113]. A difference in smell or taste between heavy water and light water is not discernable by humans. Rats however will avoid drinking heavy water if given the choice [114]. On the intracellular level D_2O was found to affect protein structure, functionality and folding kinetics, microtubule formation, mitosis and heat resistivity of the cellular membrane and cytoskeleton [115].

However, the effect of heavy water on gene expression kinetics has still not been elucidated in detail. This point is of interest for two reasons: First, heavy water could potentially be used as an ingredient of anti cancer therapy. It was demonstrated that D_2O hampers human tumor growth in animal models [116, 117] and represses proliferation in prokaryotes and eukaryotes [118, 119]. It is of interest to determine how far this is due to an effect on protein synthesis. Second, NMR studies which provide the bulk of our knowledge of protein structures require deuteration of proteins for enhanced contrast [77, 120, 121]. Deuterated proteins are usually produced in bacterial cultures. New insights into the impact of D_2O on the cell-free gene expression mechanism might help to improve yield strategies in cell-free systems.

Here the effect of D_2O on the individual steps comprising gene expression, transcription, translation and protein maturation was studied *in vitro* using GFP expression in a cell-free system. The overall effect of D_2O on protein synthesis was also measured in *E. coli* in order to compare the results of the *in vitro* system with the *in vivo* value.

6.2 Experimental results

First, the overall effect of D_2O on GFP synthesis in the PURE cell-free system was studied. Samples were prepared by mixing PURE kit with DNA and filling up with a D_2O/H_2O mix to the desired total concentration of D_2O. In order to achieve high D_2O concentrations it was necessary to dilute samples of cell-free kit more than it is recommended by the manufacturer. This resulted in a greatly reduced GFP yield. For this reason samples could only be diluted to a point where close to 60% of the total sample volume consisted of D_2O. Data were recorded fluorescently on a plate reader. Fig. 6.2 shows GFP synthesis rate and total yield in the cell-free system as a function of D_2O concentration. As can be
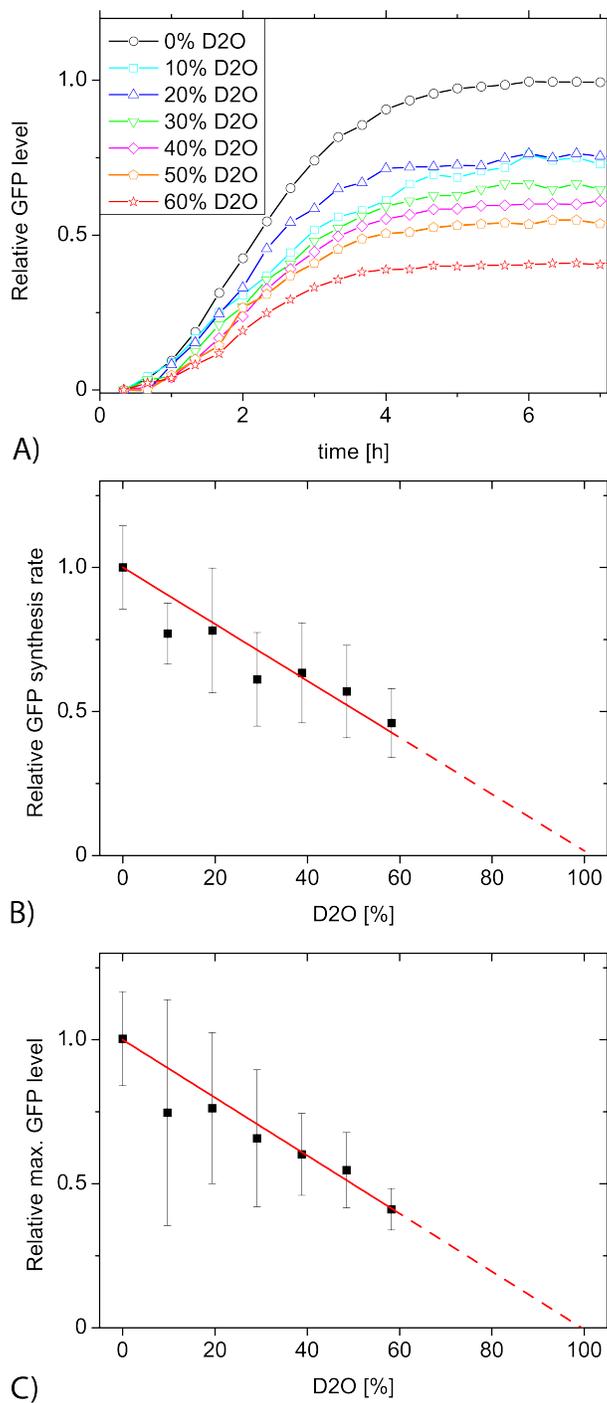
Figure 6.2: A) Exemplary curves of the kinetics of GFP synthesis in a cell-free system as a function of D_2O concentration; B) relative synthesis rate and C) relative final yield. The red line represents a linear fit to the data and was inserted to help guide the eye.
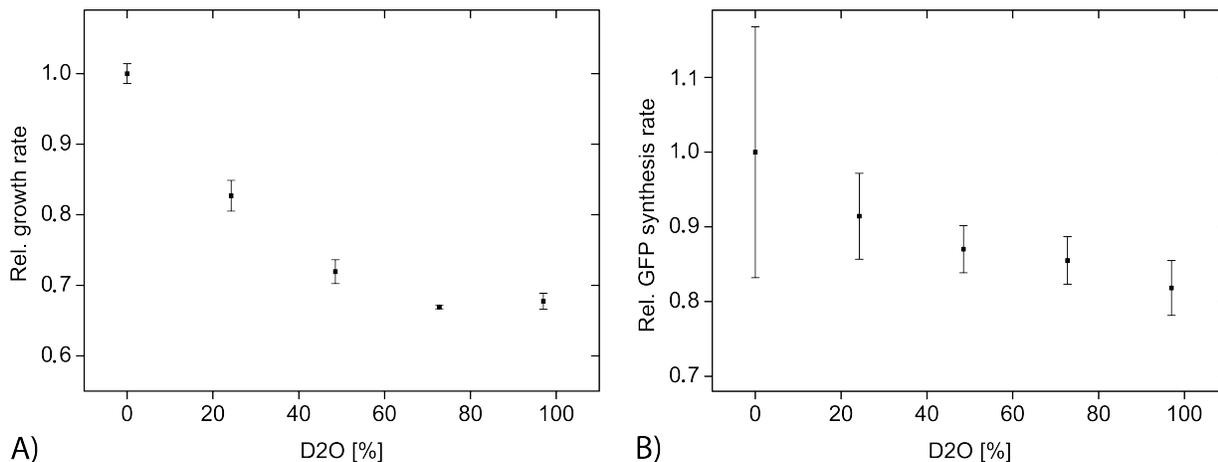
Figure 6.3: A) Proliferation and B) GFP synthesis rate of *E. coli* samples in M63 medium with various D_2O concentrations. Both proliferation and protein expression are inhibited by D_2O. However the inhibitory effect on GFP expression is less pronounced than in the *in vitro* system.

seen D_2O inhibits GFP synthesis in the *in vitro* system. Both parameters decrease linearly with increasing D_2O content. At the maximum possible D_2O concentration in the samlpes (about 60%) the synthesis rate and total yield of GFP were reduced by half the maximum value. Note that the observed reduction of fluorescence shown here and in the following figures is not due to D_2O affecting the fluorophores. This was verified by comparative analysis of fluorophore (GFP and Sybr Green mentioned further below) spectra in D_2O and H_2O.

In the next step the results gained from the *in vitro* system were compared with those for a living organism. Living cells feature a much more complex composition of molecules and also exhibit feedback mechanisms as part of the gene expression machinery. *E. coli* was put in M63 minimal medium containing various concentrations of D_2O and cellular proliferation and GFP synthesis were measured on a plate reader using brightfield and fluorescence settings respectively. Fig. 6.3 A shows that D_2O inhibits bacterial growth as described in previous literature. Fig. 6.3 B depicts the effect of D_2O on bacterial GFP synthesis. The inhibitory effect is less pronounced than in the *in vitro* system, dropping to about 80% in the fully deuterated medium.

A big advantage of the *in vitro* system is that it is possible to readily study the individual steps of the gene expression process as described in previous chapters. The effect of D_2O on transcription was measured as follows: cell-free expression was conducted for $4\,h$ at various D_2O concentrations. Then the synthesized mRNA was purified and used to produce
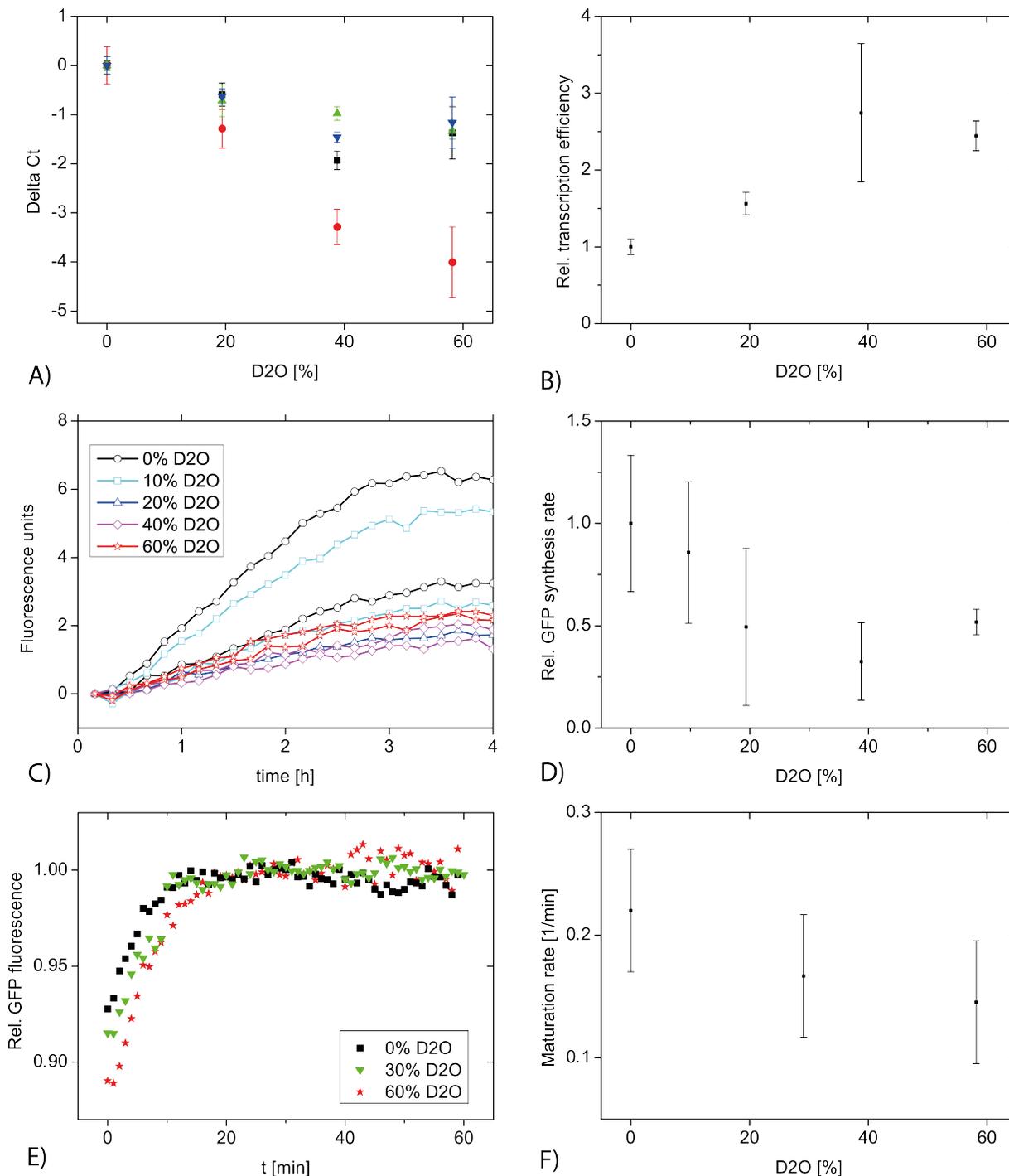
Figure 6.4: The effect of D$_2$O on GFP transcription, translation and maturation was measured independently in the cell-free system. A) Threshold cycle and B) the resulting transcription efficiency. C) Kinetics of translation and D) the resulting translation efficiency. E) Kinetics of maturation and F) the resulting maturation rates.
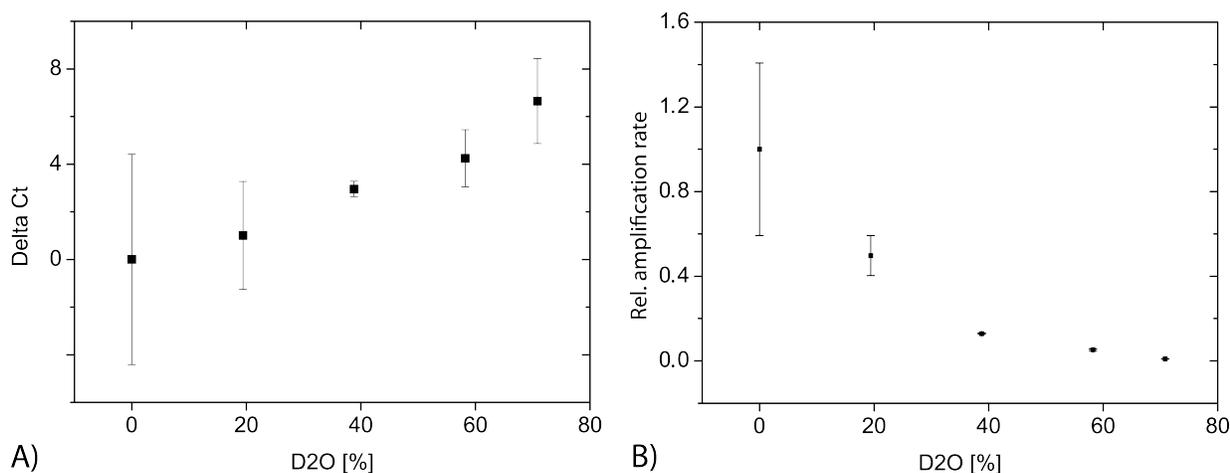
Figure 6.5: The PURE system contains T7 RNA polymerase which is a viral polymerase. For a comparison Taq DNA polymerase was used as a model prokaryotic polymerase. A) shows the threshold cycle and B) the resulting DNA amplification rate as a function of D_2O. Contrary to the effect seen with T7 RNA polymerase (Fig. 6.4 A, B) here D_2O leads to an inhibition of the synthesis rate.

cDNA using a reverse transcriptase protocol. The amount of synthesized cDNA is linearly dependent on the amount of template mRNA. Subsequently a real-time PCR was executed to determine the amount of cDNA and therewith mRNA template. The result of the PCR measurements is depicted in Fig. 6.4 A, B (for details on the interpretation of real-time PCR data see section 3.7). Interestingly the data do not show an inhibitory effect of D_2O on cell-free transcription but instead an increase of the transcriptional efficiency by a factor of about two. Real-time PCR melting curves were verified to not show primer dimers.

In this regard it was also checked if D_2O affects mRNA degradation: purified mRNA was put in H_2O and D_2O respectively, heated to $37°C$ and left standing for $3\,h$. Afterwards cDNA was produced by reverse transcription and the respective amount of cDNA was determined by real-time PCR. No D_2O effect on RNAse activity was detected (data not shown).

In order to study the effect on translation mRNA was expressed in cell-free system and purified. Subsequently this mRNA was used as template in fresh samples of cell-free system containing various concentrations of D_2O. GFP synthesis was measured fluorescently on a plate reader (Fig. 6.4 C, D). The GFP synthesis rate decreases with increasing concentrations of D_2O until 40% D_2O in the sample. At this point the data suggest that the synthesis rate reaches a plateau at about half its maximum value.

Previous literature describes that D_2O can affect protein folding, therefore GFP mat-

k_{syn}^{rel}	0.5	[Fig. 6.2 B]
k_{ts}^{rel}	2.4	[Fig. 6.4 B]
k_{tl}^{rel}	0.5	[Fig. 6.4 D]
k_{mat}^{rel}	0.7	[Fig. 6.4 F]

Table 6.1: Modified reaction rates in a PURE cell-free system with 60% D_2O. k_{syn}^{rel} denotes the total relative GFP synthesis rate and k_{ts}^{rel}, k_{tl}^{rel}, k_{mat}^{rel} denote relative rates for transcription, translation and maturation respectively. Here k_{syn}^{rel} is not simply the product of k_{ts}^{rel} and k_{tl}^{rel} because in the PURE system there is an abundance of transcribed mRNA compared to available ribosomes (compare Fig. 4.5). Note that the decreased maturation rate does not result in less GFP being synthesized but simply in a longer maturation time. Therefore the effect of D_2O on GFP synthesis observed in Fig. 6.2 B corresponds to the effect on translation and $k_{syn}^{rel} \approx k_{tl}^{rel}$.

uration was measured independently. As already described in section 4.2 the antibiotic chloramphenicol was added to samples of cell-free system 2 h after start to deactivate the ribosomes. Residual increase in GFP fluorescence after this point is due to maturation (see Fig. 6.4 E, F). GFP maturation is slowed down from $0.22 \, min^{-1}$ in light water to about $0.13 \, min^{-1}$ in 60% deuterated water.

The PURE cell-free system contains the viral polymerase of the T7 bacteriophage. One can assume that the effect of D_2O on transcription is focused on the participating polymerase. In order to compare *in vitro* results with those for *E. coli* shown above it is therefore of interest to test a prokaryotic polymerase. Taq DNA polymerase from *thermophilus aquaticus* is a prokaryotic polymerase used for *in vitro* cDNA synthesis in PCR. The effect of D_2O was studied by preparing real-time PCR samples with the same amount of cDNA template but various concentrations of D_2O. The results are shown in Fig. 6.5. Unlike with T7 RNA polymerase a strong inhibitory effect was found. At 70% D_2O in the samples the cDNA amplification rate approaches zero. The melting curves did not indicate mispaired primers.

6.3 Discussion

Heavy water affects chemical reactions in living organisms due to its slightly different physical properties compared to light water. Here the effect of D_2O on the individual steps of protein synthesis was studied in an *in vitro* system of purified components based on the gene expression system of *E. coli*. It was found that the inhibitory effect on cell-free GFP synthesis is linearly dependent on the concentration of D_2O. Comparison with

GFP expression in *E. coli* showed that the inhibitory effect is less pronounced *in vivo*. This indicates that living organisms possess some sort of compensation mechanism which makes them less susceptible to such perturbations.

Looking at the transcription step it was found that D_2O leads to enhanced efficiency of T7 RNA polymerase. At the first glance this seems to contradict the observed overall inhibition of GFP synthesis: translation is a downstream process of transcription. Inhibition of translation alone should not be enough to counteract an increased transcriptional efficiency. However, as shown in section 4.2 there is an overproduction of mRNA compared to the amount of ribosomes in the PURE system. Therefore an increase of the mRNA amount through D_2O does not result in more GFP as long as the ribosomes are saturated anyway. Furthermore a cell-free kit is an artificial system and it is imaginable that in such an environment D_2O produces a different effect than *in vivo*. What is more the results gained with Taq polymerase suggest that the target of D_2O in the transcription step is the participating polymerase and that different polymerases are affected differently. Apart from the polymerases the remaining ingredients are nearly identical in the measurements shown in Fig. 6.4 A, B and Fig. 6.5. Still D_2O amplifies mRNA synthesis by T7 RNA polymerase and inhibits DNA synthesis by Taq DNA polymerase. The observed inhibition of cell-free translation by D_2O matches the observed inhibition of the complete gene expression process reasonably well (table 6.1).

The results of this work suggest that D_2O could be used as part of a cancer therapy since it can hamper the expression of vital proteins in cells. Cancer cells would be especially susceptible to this effect because of their increased metabolism. However, for this it would be necessary to find a method to transfer D_2O specifically into cancer cells in a human body. Another consequence of this work is probably of more immediate practical use: deuterated proteins like they are necessary for NMR studies can be readily synthesized in a cell-free system. Protein synthesis in cell-free systems is generally much less cumbersome and quicker than it is in cell cultures while still yielding viable amounts of protein. In this study the cell-free kit had to be diluted in D_2O in order to achieve high concentrations. This negatively affected the performance of the system and therewith the protein yield. Therefore the maximum attainable D_2O concentration was restricted to 60%. However, if a cell-free system which itself is dissolved in D_2O was available 100% concentration would be possible. It was found that it does take longer for GFP to mature in D_2O but that it still folds correctly. As described in section 2.4.2 GFP maturation is an autocatalytic reaction that is it does not involve molecular chaperons. It has to be tested if such chaperons work

in a deuterated cell-free system in order to synthesize such proteins that require their help to fold correctly. The insights on the effect of D_2O on the individual steps of protein synthesis gained here might help to improve yields of deuterated protein.

Putting H_2O instead of D_2O into the samples can be interpreted as a small perturbation that globally affects all constituents of the gene expression reaction. It is interesting to note that the *in vivo* system, although much more complex, is more robust to this perturbation than the *in vitro* system. This might be due to some sort of regulatory mechanism that helps cells to adapt to changes in environmental conditions. Identifying this mechanism would be of fundamental theoretical interest and could also help to improve cell-free systems as an *in vitro* model platform.

Chapter 7

Cell-free expression of transmembrane protein ephrin-B2

Part of this chapter is the subject of publication [P1] (reproduced by permission of World Scientific Publishing Company).

7.1 The role of membrane proteins

In some synthetic biology approaches to the construction of an artificial minimal cell small volumes of cell-free expression system are encapsulated in lipid vesicles. In order to enable nutrient uptake, waste removal, signaling or binding to a substrate it is necessary to expand such "proto-cells" with integral membrane proteins. However, cell-free expression of membrane proteins is usually problematic. So far cell-free expression of soluble proteins, specifically GFP, was discussed. However, about 25% of the proteome are membrane proteins and as such they possess a lipophilic part (the part that incorporates into the lipid bilayer) in their structure. They are therefore not soluble and tend to form aggregates in solution. Furthermore, over-expressed membrane proteins tend to misfold and the yield is low [12]. For this reason structural and functional studies of membrane proteins are still scarce [122]. This is a matter of great concern because membrane proteins are important components of many cellular processes and they are the target of about half of all current pharmaceutical drugs [123, 124].

In recent years experimental strategies were developed with which correctly folded membrane proteins can be synthesized in cell-free systems. Membrane proteins can be expressed as a precipitate in a cell-free system and solubilized afterwards using a detergent.

It is further possible to add the detergent at the beginning of protein synthesis which yields already solubilized proteins [125]. A third method is to carry out the expression in the presence of an artificial lipid membrane system which is advantageous for certain applications. It has been reported that membrane proteins expressed in a cell-free system can integrate into lipid membranes *in situ* and fold correctly [10–12, 124, 126–128]. However, so far this method is not guaranteed to work for all membrane proteins and has to be tested for each protein species individually. One also needs to consider that proteins which require posttranslational modifications like glycosylation can not be expressed in contemporary cell-free systems (compare the respective paragraph in section 2.1.1). Incorporating membrane proteins into artificial lipid membranes has possible applications in the creation of artificial carrier systems: colloidal mesoporous silica (CMS) nanoparticles coated with a lipid bilayer were recently proposed as a core-shell type system that can store a drug in its interior and deliver it to a cell [129]. The lipid membrane prevents premature escape of the agent. Incorporating the proper signaling molecules into that membrane might facilitate cellular uptake or even enable delivery to specific cells. In the context of synthetic biology an artificial lipid vesicle with membrane proteins incorporated into its lipid bilayer is a first step on the way to creating an artificial cell.

It is also important to note that the role of the cellular membrane can not simply be reduced to some two-dimensional fluid in which the membrane proteins diffuse freely. Over the last few decades the understanding emerged that membranes exhibit a heterogeneous composition which plays an important role in the spatial arrangement of membrane adhering proteins [130]. Specifically, depending on the local lipid composition, membranes posses a varying local thickness and can deform in three dimensions. This property has been studied in detail in model membrane systems. It was found that the local membrane curvature is determined by a competition between the bending rigidity of lipid domains and minimization of the line tension along their boundaries [131, 132]. This opens up the possibility to manipulate the lateral organization of lipid domains with a suitably microstructured substrate [133, 134]. In extension this also applies to the proteins adhering to this membrane. In this manner it is for example possible to create well defined cellular binding regions on top of a biologically favorable surface, namely a lipid bilayer.

In this chapter it is shown how cell-free expression was used to synthesize a fusion protein of the integral membrane protein ephrin-B2 and GFP. The expression was carried out in the presence of model membranes and ephrin:GFP successfully incorporated into these membranes. This work has to different future applications in mind: first, lipid vesicle en-
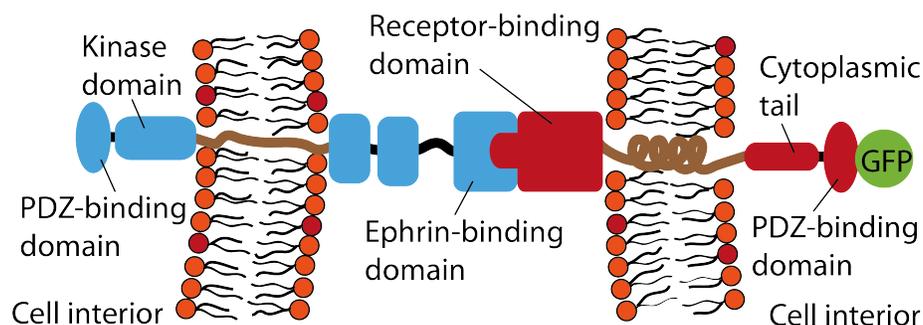
Figure 7.1: Ephrin-B2 is a transmembrane protein that interacts with the EphB class of receptor proteins. It plays a role in cell-cell signaling for axonal growth or cell migration. Ephrin-B2 has a single transmembrane region as well as a PDZ binding domain at its cytosolic C-terminus. The N-terminus carries the receptor binding domain with a hydrophobic motif that fits into a cleave in the receptor's binding domain. For this study a fusion protein was created in which GFP was attached to the C-terminus of ephrin-B2.

capsulated cell-free system that expresses and successfully incorporates a transmembrane protein is one step in the construction of a more advanced artificial cell. Second, incorporation of a binding protein like ephrin-B2 into a supported bilayer shall be used in future work for the construction of defined cellular binding spots. In addition, spatial organization of lipid domains in substrate bound model membranes by suitably microstructured substrates is demonstrated.

7.2 Ephrin-B2

Ephrins are a family of membrane proteins that play important roles in cell adhesion, migration and segmentation. Ephrins act as ligands for the Eph receptor family of receptor tyrosine kinases (see Fig. 7.1). The ephrin protein family can be subdivided into two classes, ephrin-A and ephrin-B. The five members of the ephrin-A class bind to the cellular membrane via a glycosylphosphatidylinositol (GPI) anchor and do not posses a cytoplasmic domain. The ephrin-B class has three members. They exhibit a single transmembrane domain as well as a cytoplasmic PDZ domain[1]. For this study a fusion protein of ephrin-B2 and GFP at its C-terminus was created. Previous work shows that such a fusion protein is functional, that is both the ephrin-B2 and the GFP part fold correctly [135]. The complete sequence and vector map of the ephrin:GFP fusion protein is shown in the appendix.

---

[1]PDZ domains are common protein interaction sites found in a variety of proteins.
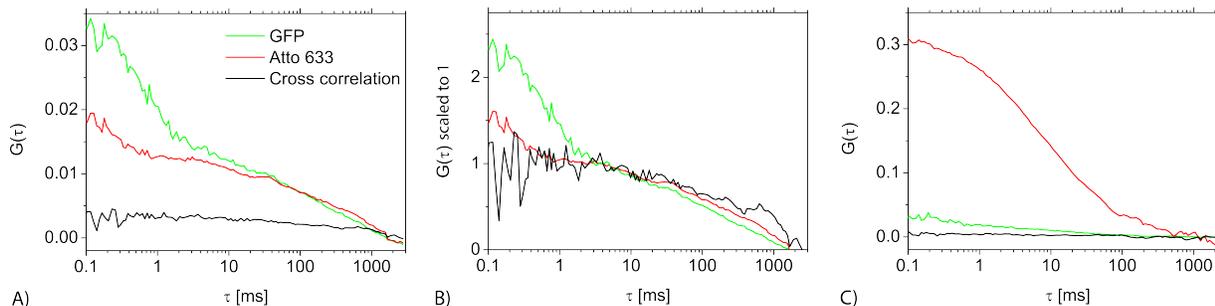
Figure 7.2: Measurement of ephrin:GFP bound to a membrane coated colloidal mesoporous silica (CMS) nanoparticle using FCCS. A) The observed cross-correlation indicates that ephrin:GFP (green GFP fluorescence) is incorporated into the membrane (red Atto 633 fluorescence). From the relatively small value of the cross-correlation follows that only a fraction of the ephrin:GFP is incorporated whereas the remainder diffuses independently of the nanoparticles. B) Here the data shown in (A) were scaled to 1 at the plateau of the correlation function at $3-5\,ms$. C) Negative control: signal of CMS nanoparticles without ephrin:GFP. Note that due to hardware constraints of the correlation unit the correlation curves could not be measured longer than shown here. The long diffusion times are most likely due to optical trapping.

7.3   Cell-free expression of ephrin-B2

Ephrin:GFP incorporation was conducted in the presence of two different artificial membrane systems: small unilamellar vesicles (SUVs) and CMS nanoparticles coated with a lipid membrane. SUVs were fabricated with various diameters ($100\,nm$, $200\,nm$ and $1\,\mu m$) and consisting of various lipid compositions (egg-PC, chol/bSM/DOPC, chol/DOPC and chol/DOPC/DOTAP). CMS nanoparticles with a diameter of $100\,nm$ were labeled with the fluorescent dye Atto 633 and subsequently coated with a lipid bilayer membrane consisting of the phospholipid DOPC. SUVs or nanoparticles respectively were added to the PURE cell-free system and ephrin:GFP was expressed in their presence (compare section 3.8 for the exact experimental protocol). Initially FCS was tried to confirm that ephrin:GFP bound to the SUVs' membrane. However, it was found that the average hydrodynamic radius (established via equation (3.5)) was only $30\,nm$ instead of the expected $100\,nm$ SUV radius. This result implies that at least a considerable part of the ephrin:GFP did not incorporate into the membranes. Instead it agglomerated in clusters which on average had a smaller size than the SUVs. This is not unexpected as it was reported in previous work that a good portion of membrane protein thus expressed does not successfully incorporate into a membrane (see for example [10, 12]). In order to corroborate the incorporation of
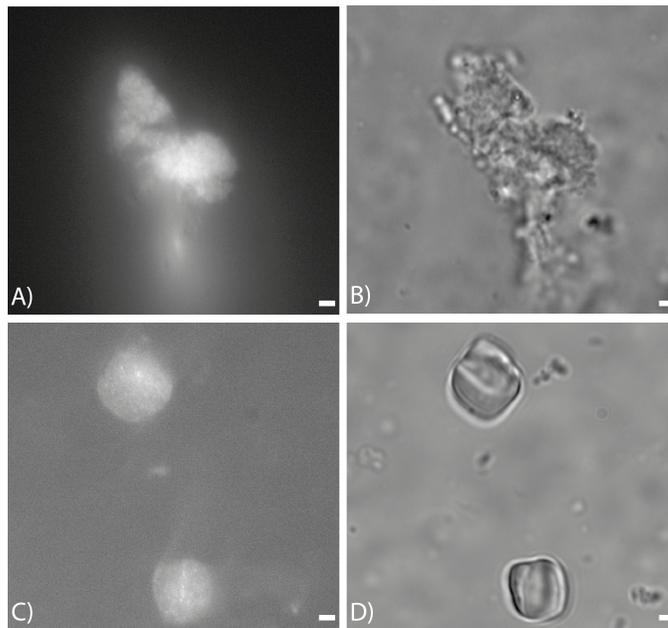
Figure 7.3: Fluorescence microscopy measurements of SUVs with ephrin:GFP performed with a 100× objective. A) Agglomeration of ephrin:GFP on the bottom of the sample chamber in fluorescence and B) bright field. C) Ruptured SUVs with incorporated ephrin:GFP in fluorescence and D) bright field. Scale bars are $2\,\mu m$.

ephrin:GFP into the artificial membranes the cross correlation signal of ephrin:GFP and Atto 633 labeled nanoparticles was measured with Fluorescence Cross Correlation Spectroscopy (FCCS). In the case of SUVs Texas Red/GFP cross correlation and in the case of CMS nanoparticles Atto 633/GFP cross correlation was measured using the Two-Photon FCS setup. The measured cross correlation amplitude is markedly smaller than the correlation amplitudes of the red and green channel (Fig. 7.2 A, B). This result suggests that part of the ephrin:GFP did indeed incorporate into the artificial membrane system. Cross talk between the red and green channel was excluded as explanation for the observed cross correlation (Fig. 7.2 C).

Ephrin:GFP incorporation into artificial membranes was further verified using fluorescence microscopy. SUVs with a diameter of $1\,\mu m$ were fabricated with the extrusion method. Individual SUVs of this size can be visualized with a 100× objective. Cell-free expression of ephrin:GFP was performed in the presence of SUVs of $1\,\mu m$ diameter comprised of chol/DOPC/DOTAP 30:69:1. After three hours of expression at $37°C$ the sample was examined in a ibidi VI slide (poly-l-lysine treated) with a fluorescence microscope (Fig. 7.3). Freely diffusing SUVs with incorporated ephrin:GFP could not be observed. This
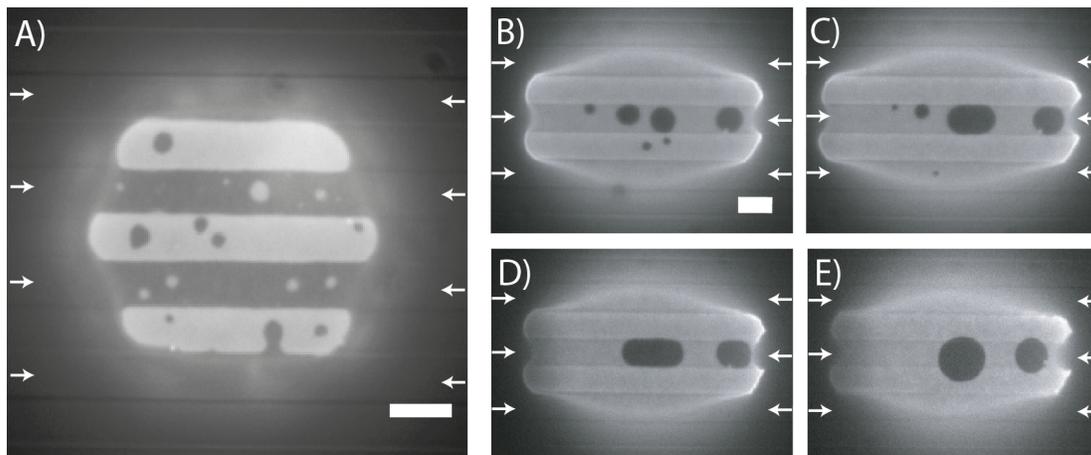
Figure 7.4: A) L_o (dark) and L_d (bright) domains are almost completely separated by the substrate induced energy barriers. L_o domains are trapped over the smooth stripes whereas L_d domains accumulate over the rough ones. Here the domains are completely immobile. B)-E) show a time series of trapped but still mobile L_o domains. The domains grow until they touch the edges of the strip. At this point the domains elongate along the energy barriers (the edges) until the line tension overcomes the attraction between GUV and substrate. Subsequently the GUV is locally lifted from the substrate and the L_o domains become circular again. Scale bars are $4\,\mu m$.

is most likely caused by the overall low concentration of expressed ephrin:GFP. Therefore only ephrin:GFP immobilized on the bottom of the slide could be observed. Numerous sedimented agglomerations of ephrin:GFP were found (Fig. 7.3 A, B) as well as numerous lipid patches of burst SUVs which clearly exhibited ephrin:GFP incorporation (Fig. 7.3 C, D).

It is possible that ephrin:GFP did not incorporate into SUV and nanoparticle membranes in its correctly folded, that is functional form. In future work antibody labeling will be used to corroborate this point.

7.4   Lateral organization of lipid domains on microstructured surfaces

One major aim of cell-free ephrin-B2 expression as shown here is to develop a reliable and convenient method for the fabrication of defined cellular binding sites on a substrate. In the following it will be shown that a microstructured substrate can lead to lateral alignment of coexisting lipid phases in a substrate adhering membrane. Membrane proteins with a
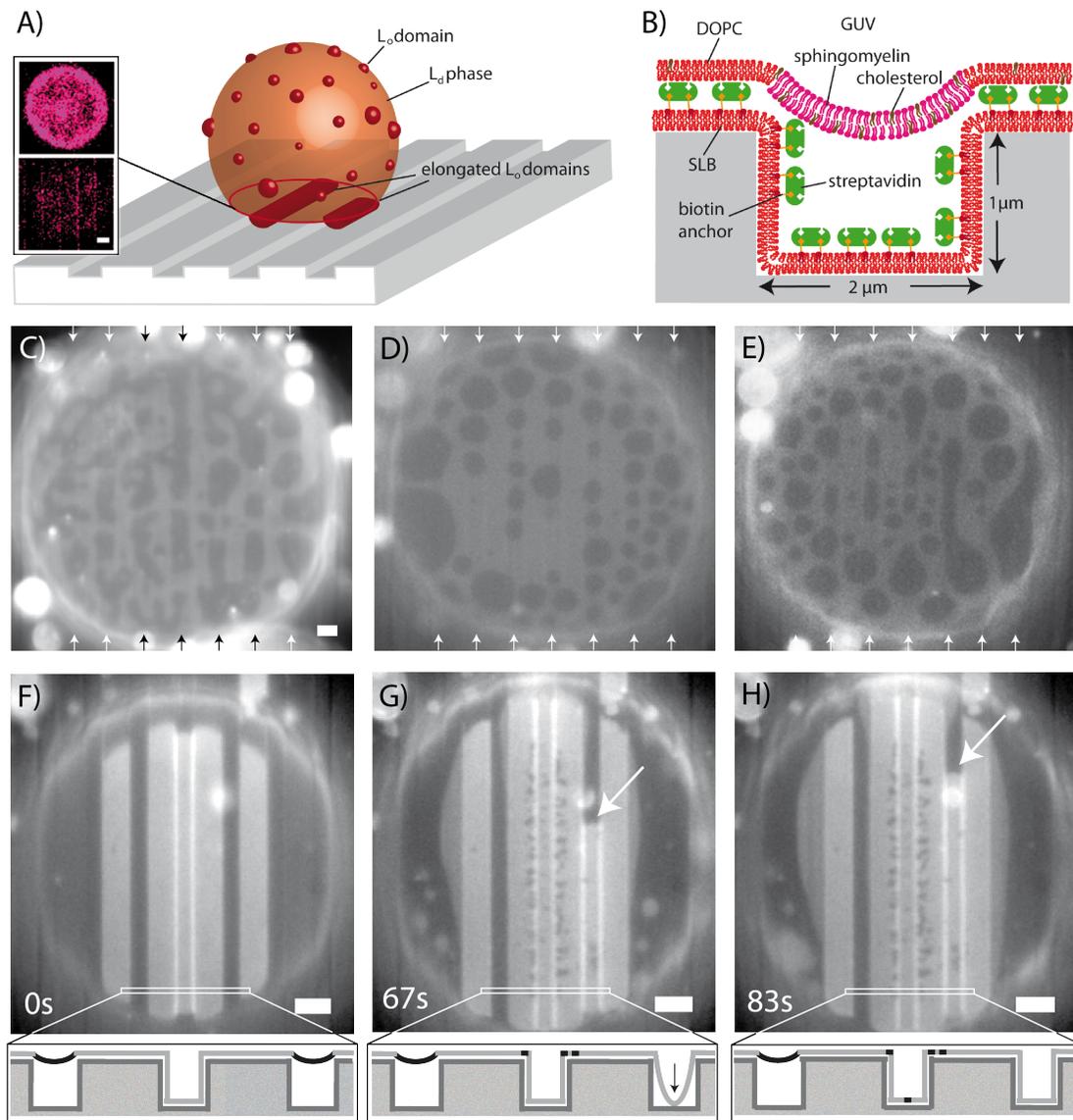
Figure 7.5: A) Schematics of a GUV on top of a substrate with grooves. The insert shows LSM images of the plane of the plateaus (top) and the grooves (bottom): the membrane is uniformly located in the plane of the plateaus meaning that it spans over the grooves. B) Schematics of a GUV on top of a SLB coated substrate. Only part of the contact area of the GUV is drawn. Streptavidin can be used to increase the interaction force between GUV and SLB. C) L_o domain alignment and deformation in a GUV on top of a BSA coated substrate (no SLB). D) L_o alignment but no deformation in a GUV on top of a SLB coated substrate (no streptavidin). Illumination disturbs the lipid domains and leads to L_o domain coalescence and subsequent outflow from the bottom area of the GUV (E). F)-H) Time series of nearly perfect domain separation and alignment in a GUV on top of a substrate with SLB and streptavidin. L_o domains remain mobile. Schematic cross section on the bottom. Scale bars are $4\,\mu m$.

preference for one of the lipid phases can thus be arranged in a defined pattern.

Giant unilamellar vesicles (GUVs) comprised of chol/bSM/DOPC 20:40:40 were prepared using the electroformation method (compare section 3.8). The lipid DHPE carrying the dye Texas Red was added to the composition in concentrations lower than one mole percent for visualization under a fluorescence microscope. Membranes with this lipid composition exhibit two coexisting domains at room temperature which are distinguished by the degree of order of their phospholipids' tails. These phases are therefore termed "liquid ordered" (L_o) and "liquid disordered" (L_d) respectively. Cholesterol predominantly accumulates in the L_o phase whereas TR-DHPE accumulates in the L_d phase. Thus the coexisting phases can be differentiated by fluorescence [136, 137]. Glass substrates were microstructured with a stripe pattern of alternating smooth and rough surfaces using standard reactive ion etching methods as described in section 3.9. A cover layer of the protein bovine serum albumin (BSA) was applied to the substrates before measurements to prevent GUVs from rupturing when they sedimented. Sedimentation occurred due to the higher density of the sucrose solution inside the GUVs compared to the glucose solution outside. Lipid phases on the bottom of sedimented GUVs arranged according to the stripe pattern (Fig. 7.4). L_o domains were trapped over the smooth regions and started to accumulate and coalesce whereas L_d accumulated over the rough regions. In some cases the trapped L_o domains were found to be completely immobilized (Fig. 7.4 A) whereas in others they retained their mobility (Fig. 7.4 B-E). In the latter case the growing domain elongated along the pattern boundary until the line tension overcame the energy barrier. Thereby the GUV membrane was locally detached from the substrate and the domain became circular again.

Lipid phase demixing was also achieved with substrates that featured a micro pattern of grooves instead of rough and smooth stripes (Fig. 7.5 A). Sedimented GUVs freely span over the grooves due to the large hight to width aspect ratio of the micro pattern. This was controlled with laser scanning microscopy as shown in the inlet in Fig. 7.5 A: the GUV membrane is located in the plane of the plateaus whereas in the plane of the grooves only light scattering is detected. On these substrates the surface was passivated either with a BSA layer as mentioned above or with a supported lipid bilayer (SLB). The SLB could be functionalized with streptavidin to increase the binding strength between the GUVs and the substrate (Fig. 7.5 B). In all cases the L_o domains remained fluid and accumulated and elongated in the freely suspended membrane parts over the grooves. In GUVs atop of BSA passivated substrates (Fig. 7.5 C) and atop of SLB coated substrates without streptavidin (Fig. 7.5 D-E) the substrate induced energy barrier was small. Continued illumination

disturbed elongated L_o domains and they adopted a circular shape and diffused out of the contact area. In the case of GUVs atop of SLB covered substrates with streptavidin the energy barrier was significantly higher and the lipid domains arranged precisely along the groove edges Fig. (7.5 F). If the interaction between GUV and the streptavidin layer became too strong the GUV membrane was dragged down into the grooves and the L_o domain was pressed out (see arrows in Fig. 7.5 G-H and the cartoon below). Note that the streptavidin layer was bound to the SLB via biotinylated lipids whereas the interaction between GUVs and streptavidin was non-specific. It was found that even very low concentrations of biotinylated lipids in the GUVs increased the interaction thus that GUVs quickly ruptured.

7.5   Discussion

Cell-free expression systems can readily be used to synthesize a wide variety of soluble proteins. However, the synthesis of membrane proteins is still no trivial matter and special protocols have to be followed to achieve correct folding and adequate yields. This is the case for both cellular as well as cell-free expression. For this study a fusion protein of ephrin-B2 and GFP was constructed and expressed in cell-free system in the presence of artificial lipid membranes. Using two complimentary methods it is shown that ephrin:GFP most likely incorporated into the membranes: with FCCS for freely diffusing vesicles and with fluorescence microscopy for vesicles immobilized on a substrate. Control measurements using antibody labeling need yet to be done to ensure this point.

Incorporating the proper receptor molecules into the lipid membrane surrounding a CMS nanoparticle could facilitate cellular uptake and enable targeting of specific cell types. This might have great potential for example in anti cancer therapies were such nanoparticles could be used as drug ferries. Vesicles with incorporated membrane proteins are a ready model platform for the development of an artificial minimal cell, a major goal of synthetic biology research.

A major goal of this project is to use ephrin incorporated into a SLB to create binding sites for cells. Especially with regard to automated measurements it is important to create a defined pattern of such binding sites. That is ephrin anchors should only accumulate in specific regions of the substrate. Here it is shown that microstructures on a substrate can induce alignment and deformation of coexisting lipid phases in adhering membranes. This could be used as a possible sorting mechanism for membrane proteins with a preference

for a specific lipid phase. Apart from the static patterning approach described here a remarkable alternative would be to use an active substrate to create ephrin patterns. It was shown that surface acoustic waves (SAWs) can be used to create reversible patterns of SLB bound macromolecules [138, 139]. An interesting application for example would be to use propagating SAWs to create a density gradient of ephrin and to study cellular behavior under such conditions.

Chapter 8

Outlook

The thesis at hand concerns cell-free gene expression as a biophysical model system for the intricate protein synthesis reactions of living cells. Cell-free systems are dilutions of the minimal components of the cellular transcription and translation machinery which allow biochemical synthesis of specific target proteins *in vitro*. The main objective of this thesis was to quantitatively study gene expression kinetics *in vitro* and to create a predictive mathematical model of these kinetics. Measurements were carried out fluorescently on a micro plate array which permits real-time read-out and parallelization. Due to sample volumes of a few micro liters all components were present in high copy numbers therefore stochastic noise was insignificant. Such a properly characterized *in vitro* model system in combination with a stringent mathematical modeling is a valuable tool for studies of transcription and translation mechanics, gene interaction networks, modifications of individual reaction components as well as the construction of artificial cellular structures.

To this end a quantitative and predictive rate equation model of the cell-free protein synthesis kinetics was developed. The model accurately fits time courses of mRNA as well as protein synthesis from beginning to end. This was achieved by taking into account finite resources as well as degradation of the cell-free machinery over time. In order to check how far this completely artificial system can be compared to one that more closely resembles the situation *in vivo* a similar modeling approach was applied to the GFP synthesis kinetics of mRNA and DNA transfected cells respectively. It was found that the time courses of mRNA transfected cells but not DNA transfected cells can be fitted with this basic ansatz. This finding hints at a mechanism in cellular gene expression, possibly mRNA transport from the nucleus to the cytosol or some sort of cooperativity, that is so far not completely understood. Comparative studies with an *in vitro* model might help to elucidate this
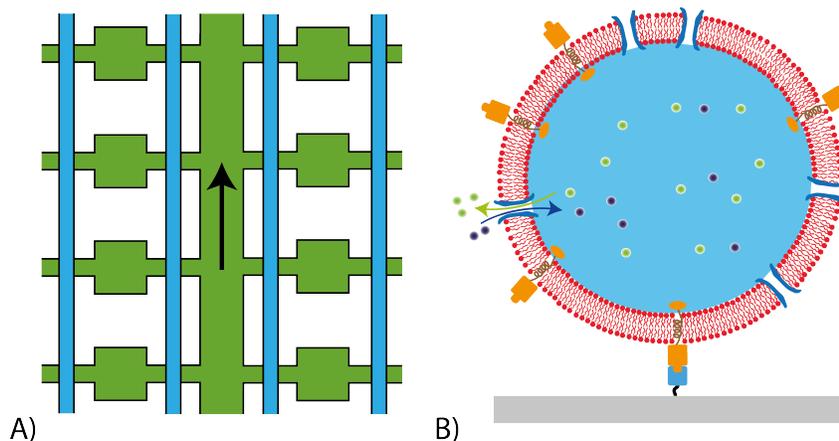
Figure 8.1: A) Cell-free expression can be readily combined with microfluidics chips. Here the schematics of a chip that was developed at the chair is shown. The quadratic sample chambers have a volume of $0.1\,nl$ each. They could be filled with $1 - 10^4$ DNA or mRNA molecules for cell-free expression depending on the respective concentration. A second layer of channels (drawn in blue) atop the first one can be used as valve that seals the in- and outlets of the chambers. B) Lipid vesicles filled with cell-free system can be used as a platform for the construction of an artificial minimal cell. Especially the extension of this concept to transmembrane proteins would be interesting. This would for example allow to emulate cellular nutrient uptake *in vitro*.

point: cell-free kits are basically a modular system and they could be expanded with new functionalities like compartmentalization, an expression network with several stages [5] or feedback loops.

In this regard it would be of great interest to reliably measure the kinetics of mRNA and protein synthesis simultaneously. Molecular beacons as applied in this thesis reliably label mRNA but they also impede translation because ribosomes can not process mRNA/molecular beacon hybrids. This problem could be solved by using the new RNA aptamers introduced by Paige et al. [85] (compare section 3.5). Adding the aptamer sequence to the GFP sequence on the mRNA both transcription and translation kinetics could be visualized simultaneously.

One of the big advantages of cell-free gene expression is that it can be readily downscaled and combined with microfluidics. In this thesis the typical sample volume was of the order of ten micro liters. Various microfluidics approaches were tested to downscale the sample volume (down to nano liters) but this resulted in significant loss of signal to noise. However, this was most likely due to deviations in the filling of the nano wells used here (compare section 3.3). It should be possible to avoid this problem with a more advanced microfluidics

setup like the one described by Maerkl et al. [18] or a chip that was recently developed at the chair (Fig. 8.1 A). These chips exhibit a two story layout of the microfluidics channels where the upper channels can be used as valves that seal the sample volumes. Such chips would allow for a massive increase in measurement parallelization on the order of hundreds of parallel samples instead of only about ten. A further point of interest that could be studied using small volumes as described here ($0.1\,nl$) would be stochasticity in gene expression due to low copy numbers of participating components. Depending on the concentration of the DNA used it is possible to fill such volumes with only one DNA or mRNA molecule each. Single molecule sensitivity on the fluorescence microscope would be necessary [140, 141] to measure the exact number of synthesized mRNAs or proteins respectively.

Furthermore, with such a microfluidics chip the unpacking mechanics of complexed DNA, specifically polyethylenimine (PEI) complexed DNA could be studied. PEI is a cationic polymer that is used in transfection. So far it is not completely understood how exactly the complexed DNA is processed by the cell, that is if it is unpacked or if polymerases manage to transcribe complexed DNA. This could be elucidated with cell-free expression: unpacking would become noticeable as a step function in the GFP fluorescence as more and more DNA molecules become accessible. Processing of complexed DNA on the other hand should be verifiable by a simple decrease in the transcription rate compared to free DNA. Especially for the verification of the first case working with small copy numbers and massive parallelization using microfluidics would most likely be necessary. Some bulk experiments were performed in the course of this thesis but it was not possible to distinguish the two cases from these results alone.

The creation of a synthetic minimal cell is a long standing goal of synthetic biology. Encapsulation of a cell-free system inside a vesicle is one approach to this end. Such encapsulation methods were tested successfully in the course of this thesis but not further pursued. This was because the focus shifted to the point that all samples should have the same volume in order to make parallel quantitative measurements of the GFP synthesis kinetics. However, it would probably be interesting to pursue this approach for transmembrane proteins like ephrin-B2 as well as nutrient uptake proteins or membrane pore proteins. This way an artificial cell could be created that could be immobilized on a substrate, be fed from the environment and release its waste products (Fig. 8.1 B).

Miniaturization, automatization and "lab on a chip" are typical keywords which are often used to describe the expected future development of current laboratory methods. In

the course of this process the importance of cell-free expression systems can be expected to grow further as they suit these requirements well.

Appendix A

List of publications

P1 Tobias Stögbauer, Martin Hennig, Joachim O. Rädler. Alignment and deformation of lipid bilayer domains in vesicles adhering to microstructured substrates. *Biophysical Reviews and Letters*, 5:153-161, 2010.

P2 Tobias Stögbauer, Lukas Windhager, Ralf Zimmer, Joachim O. Rädler. Experiment and Mathematical Modeling of Gene Expression Dynamics in a Cell-Free System. *Integrative Biology*, 4:494-501, 2012.

M1 Gerlinde Schwake, Carolin Leonhardt, Tobias Stögbauer, Susanne Rappl, Jan-Tim Kuhr, Thomas S. Ligon, Joachim O. Rädler. Quantitative modeling of mRNA transfection based on single-cell expression time-courses. In preparation.

M2 Luisa Hohlefelder*,Tobias Stögbauer*, Madeleine Leisner, Thomas M. Bayerl, Joachim O. Rädler. D2O inhibits GFP expression in prokaryotic cell-free assays and bacteria. In preparation. *shared first authorship

Appendix B

Ephrin-B2:GFP sequence and vector map

```
LOCUS         dna                      5409 bp
FEATURES              Location/Qualifiers
     Promoter         21..39
                      /gene="T7 prom"

     misc_binding     98..103
                      /dbxref="REBASE:NdeI"

     other_gene       100..1172
                      /gene="EFNB2"

     Tag              101..133
                      /gene="T7_leader tag"

     Other Gene       101..133
                      /gene="T7_gene10_leader other"

     ORF              101..1909
                      /sequence="ORF_1 rf(2)"

     misc_binding     136..141
                      /dbxref="REBASE:BamHI"

     misc_binding     142..147
                      /dbxref="REBASE:KpnI"

     misc_binding     250..255
                      /dbxref="REBASE:ClaI"

     misc_binding     491..496
                      /dbxref="REBASE:EcoRV"
```

```
     Reporter          1190..1906
                       /gene="EGFP reporter"
     misc_binding      1911..1918
                       /dbxref="REBASE:NotI"
     misc_binding      1920..1925
                       /dbxref="REBASE:XhoI"
     misc_binding      1920..1925
                       /dbxref="REBASE:AvaI"
     misc_binding      1920..1925
                       /dbxref="REBASE:AvrI"
     Terminator        1954..2082
                       /gene="T7 term"
     Rep_Origin        2111..2417
                       /gene="f1 origin"
     Promoter          2611..2639
                       /gene="amp prom"
     Marker            2681..3541
                       /gene="amp marker"
     ORF               2681..3541
                       /sequence="ORF_2 rf(2)"
     misc_binding      3222..3227
                       /dbxref="REBASE:PstI"
     misc_binding      3344..3354
                       /dbxref="REBASE:BglI"
     Rep_Origin        3696..4315
                       /gene="pBR322 origin"
     misc_binding      3942..3950
                       /dbxref="REBASE:AlwNI"
     misc_binding      4587..4592
                       /dbxref="REBASE:AccI"
     Other Gene        4730..4921
                       /gene="ROP other"
     misc_binding      4767..4772
                       /dbxref="REBASE:PvuII"
BASE COUNT     1360 a   1422 c   1386 g   1241 t     0 others
ORIGIN
       1 agatctcgat cccgcgaaat taatacgact cactataggg agaccacaac ggtttccctc
```

```
  61 tagaaataat tttgtttaac tttaagaagg agatatacat atggctagca tgactggtgg
 121 acagcaaatg ggtcgggatc cggtaccgag gagatctgcc gccgcgatcg ccatggctgt
 181 gagaagggac tccgtgtgga agtactgctg gggtgttttg atggtttttat gcagaactgc
 241 gatttccaaa tcgatagttt tagagcctat ctattggaat tcctcgaact ccaaatttct
 301 acctggacaa ggactggtac tatacccaca gataggagac aaattggata ttatttgccc
 361 caaagtggac tctaaaactg ttggccagta tgaatattat aaagtttata tggttgataa
 421 agaccaagca gacagatgca ctattaagaa ggaaaatacc cctctcctca actgtgccaa
 481 accagaccaa gatatcaaat tcaccatcaa gtttcaagaa ttcagcccta acctctgggg
 541 tctagaattt cagaagaaca aagattatta cattatatct acatcaaatg ggtctttgga
 601 gggcctggat aaccaggagg gaggggtgtg ccagacaaga gccatgaaga tcctcatgaa
 661 agttggacaa gatgcaagtt ctgctggatc aaccaggaat aaagatccaa caagacgtcc
 721 agaactagaa gctggtacaa atggaagaag ttcgacaaca agtccctttg taaaaccaaa
 781 tccaggttct agcacagacg gcaacagcgc cggacattcg gggaacaaca tcctcggttc
 841 cgaagtggcc ttatttgcag ggattgcttc aggatgcatc atcttcatcg tcatcatcat
 901 cacgctggtg gtcctcttgc tgaagtaccg gaggagacac aggaagcact cgccgcagca
 961 cacgaccacg ctgtcgctca gcacactggc cacacccaag cgcagcggca acaacaacgg
1021 ctcagagccc agtgacatta tcatcccgct aaggactgcg gacagcgtct tctgccctca
1081 ctacgagaag gtcagcgggg actacgggca cccggtgtac atcgtccagg agatgccccc
1141 gcagagcccg gcgaacattt actacaaggt cacgcgtccg gtcgccacca tggtgagcaa
1201 gggcgaggag ctgttcaccg gggtggtgcc catcctggtc gagctggacg gcgacgtaaa
1261 cggccacaag ttcagcgtgt ccggcgaggg cgagggcgat gccacctacg gcaagctgac
1321 cctgaagttc atctgcacca ccggcaagct gcccgtgccc tggcccaccc tcgtgaccac
1381 cctgacctac ggcgtgcagt gcttcagccg ctaccccgac cacatgaagc agcacgactt
1441 cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc accatcttct tcaaggacga
1501 cggcaactac aagacccgcg ccgaggtgaa gttcgagggc gacaccctgg tgaaccgcat
1561 cgagctgaag ggcatcgact tcaaggagga cggcaacatc ctggggcaca agctggagta
1621 caactacaac agccacaacg tctatatcat ggccgacaag cagaagaacg gcatcaaggt
1681 gaacttcaag atccgccaca acatcgagga cggcagcgtg cagctcgccg accactacca
1741 gcagaacacc cccatcggcg acggccccgt gctgctgccc gacaaccact acctgagcac
1801 ccagtccgcc ctgagcaaag accccaacga gaagcgcgat cacatggtcc tgctggagtt
1861 cgtgaccgcc gccgggatca ctctcggcat ggacgagctg tacaagtaaa gcggccgcac
1921 tcgagcacca ccaccaccac cactgagatc cggctgctaa caaagcccga aaggaagctg
1981 agttggctgc tgccaccgct gagcaataac tagcataacc ccttggggcc tctaaacggg
2041 tcttgagggg ttttttgctg aaaggaggaa ctatatccgg attggcgaat gggacgcgcc
2101 ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga ccgctacact
2161 tgccagcgcc ctagcgcccg ctcctttcgc tttcttccct tcctttctcg ccacgttcgc
2221 cggctttccc cgtcaagctc taaatcgggg gctcccttta gggttccgat ttagtgcttt
```

```
2281 acggcacctc gaccccaaaa aacttgatta gggtgatggt tcacgtagtg ggccatcgcc
2341 ctgatagacg gttttttcgcc cttttgacgtt ggagtccacg ttctttaata gtggactctt
2401 gttccaaact ggaacaacac tcaaccctat ctcggtctat tcttttgatt tataagggat
2461 tttgccgatt tcggcctatt ggttaaaaaa tgagctgatt taacaaaaat ttaacgcgaa
2521 ttttaacaaa atattaacgt ttacaatttc aggtggcact tttcggggaa atgtgcgcgg
2581 aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca tgagacaata
2641 accctgataa atgcttcaat aatattgaaa aaggaagagt atgagtattc aacatttccg
2701 tgtcgccctt attcccttttt ttgcggcatt ttgccttcct gtttttgctc acccagaaac
2761 gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca cgagtgggtt acatcgaact
2821 ggatctcaac agcggtaaga tccttgagag ttttcgcccc gaagaacgtt ttccaatgat
2881 gagcactttt aaagttctgc tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga
2941 gcaactcggt cgccgcatac actattctca gaatgacttg gttgagtact caccagtcac
3001 agaaaagcat cttacggatg gcatgacagt aagagaatta tgcagtgctg ccataaccat
3061 gagtgataac actgcggcca acttacttct gacaacgatc ggaggaccga aggagctaac
3121 cgctttttttg cacaacatgg gggatcatgt aactcgcctt gatcgttggg aaccggagct
3181 gaatgaagcc ataccaaacg acgagcgtga caccacgatg cctgcagcaa tggcaacaac
3241 gttgcgcaaa ctattaactg gcgaactact tactctagct tcccggcaac aattaataga
3301 ctggatggag gcggataaag ttgcaggacc acttctgcgc tcggcccttc cggctggctg
3361 gtttattgct gataaatctg gagccggtga gcgtgggtct cgcggtatca ttgcagcact
3421 ggggccagat ggtaagccct cccgtatcgt agttatctac acgacgggga gtcaggcaac
3481 tatggatgaa cgaaatagac agatcgctga gataggtgcc tcactgatta agcattggta
3541 actgtcagac caagtttact catatatact ttagattgat ttaaaacttc attttttaatt
3601 taaaaggatc taggtgaaga tcctttttga taatctcatg accaaaatcc cttaacgtga
3661 gttttcgttc cactgagcgt cagaccccgt agaaaagatc aaaggatctt cttgagatcc
3721 ttttttttctg cgcgtaatct gctgcttgca aacaaaaaaa ccaccgctac cagcggtggt
3781 ttgtttgccg gatcaagagc taccaactct ttttccgaag gtaactggct tcagcagagc
3841 gcagatacca aatactgtcc ttctagtgta gccgtagtta ggccaccact tcaagaactc
3901 tgtagcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg ctgccagtgg
3961 cgataagtcg tgtcttaccg ggttggactc aagacgatag ttaccggata aggcgcagcg
4021 gtcgggctga acggggggtt cgtgcacaca gcccagcttg gagcgaacga cctacaccga
4081 actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaag ggagaaaggc
4141 ggacaggtat ccggtaagcg gcagggtcgg aacaggagag cgcacgaggg agcttccagg
4201 gggaaacgcc tggtatcttt atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg
4261 attttttgtga tgctcgtcag ggggggcggag cctatggaaa aacgccagca acgcggcctt
4321 tttacggttc ctggcctttt gctggccttt tgctcacatg ttctttcctg cgttatcccc
4381 tgattctgtg ataaccgta ttaccgcctt tgagtgagct gataccgctc gccgcagccg
4441 aacgaccgag cgcagcgagt cagtgagcga ggaagcggaa gagcgcctga tgcggtattt
```

```
4501 tctccttacg catctgtgcg gtatttcaca ccgcatatat ggtgcactct cagtacaatc
4561 tgctctgatg ccgcatagtt aagccagtat acactccgct atcgctacgt gactgggtca
4621 tggctgcgcc ccgacacccg ccaacacccg ctgacgcgcc ctgacgggct tgtctgctcc
4681 cggcatccgc ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt cagaggtttt
4741 caccgtcatc accgaaacgc gcgaggcagc tgcggtaaag ctcatcagcg tggtcgtgaa
4801 gcgattcaca gatgtctgcc tgttcatccg cgtccagctc gttgagtttc tccagaagcg
4861 ttaatgtctg gcttctgata aagcgggcca tgttaagggc ggttttttcc tgtttggtca
4921 ctgatgcctc cgtgtaaggg ggatttctgt tcatgggggt aatgataccg atgaaacgag
4981 agaggatgct cacgatacgg gttactgatg atgaacatgc ccggttactg gaacgttgtg
5041 agggtaaaca actggcggta tggatgcggc gggaccagag aaaaatcact cagggtcaat
5101 gccagcgctt cgttaataca gatgtaggtg ttccacaggg tagccagcag catcctgcga
5161 tgcagatccg gaacataatg gtgcagggcg ctgacttccg cgtttccaga ctttacgaaa
5221 cacggaaacc gaagaccatt catgttgttg ctcaggtcgc agacgttttg cagcagcagt
5281 cgcttcacgt tcgctcgcgt atcggtgatt cattctgcta accagtaagg caaccccgcc
5341 agcctagccg ggtcctcaac gacaggagca cgatcatgcg cacccgtggc caggacccaa
5401 cgctgcccg
```
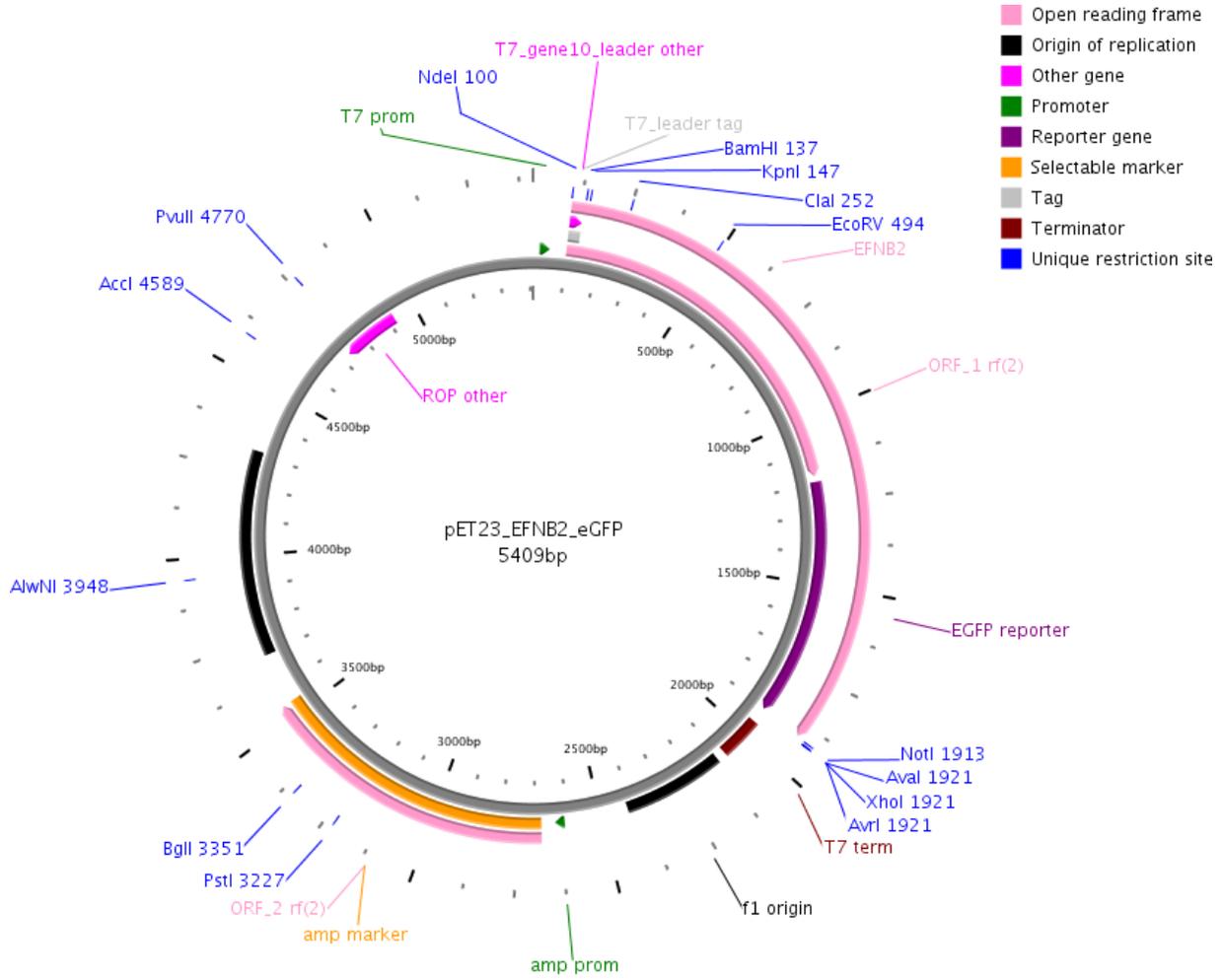
Created using PlasMapper

Bibliography

[1] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond.)*, 117(4):500–544, Aug 1952.

[2] Ralf Jungmann, Stephan Renner, and Friedrich C Simmel. From DNA nanotechnology to synthetic biology. *HFSP journal*, 2(2):99–109, April 2008.

[3] Petra Schwille and Stefan Diez. Synthetic biology of minimal systems. *Critical Reviews in Biochemistry and Molecular Biology*, 44(4):223–42, 2009.

[4] Michael B Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[5] Vincent Noireaux, Roy Bar-Ziv, and Albert Libchaber. Principles of cell-free genetic circuit assembly. *Proc. Natl. Acad. Sci. U.S.A.*, 100(22):12672–7, October 2003.

[6] D S Dimitrov and M I Angelova. Lipid swelling and liposome formation mediated by electric fields. *Bioelectrochemistry and Bioenergetics*, 19:323–336, 1988.

[7] Vincent Noireaux and Albert Libchaber. A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl. Acad. Sci. U.S.A.*, 101(51):17669–74, December 2004.

[8] Vincent Noireaux, Roy Bar-Ziv, Jeremy Godefroy, Hanna Salman, and Albert Libchaber. Toward an artificial cell based on gene expression in vesicles. *Physical Biology*, 2(3):P1–P8, September 2005.

[9] Christian Klammt, Frank Lohr, Birgit Schafer, Winfried Haase, Volker Dotsch, Heinz Ruterjans, Clemens Glaubitz, and Frank Bernhard. High level cell-free expression and specific labeling of integral membrane proteins. *European Journal of Biochemistry*, 271(3):568–580, February 2004.

[10] Rolf Kalmbach, Igor Chizhov, Miria C Schumacher, Thomas Friedrich, Ernst Bamberg, and Martin Engelhard. Functional cell-free synthesis of a seven helix membrane protein: in situ insertion of bacteriorhodopsin into liposomes. *Journal of Molecular Biology*, 371(3):639–48, August 2007.

[11] Rudolf Robelek, Eva S Lemker, Birgit Wiltschi, Vinzenz Kirste, Renate Naumann, Dieter Oesterhelt, and Eva-Kathrin Sinner. Incorporation of in vitro synthesized GPCR into a tethered artificial lipid membrane system. *Angewandte Chemie (International ed. in English)*, 46(4):605–8, January 2007.

[12] Jenny A Cappuccio, Craig D Blanchette, Todd A Sulchek, Erin S Arroyo, Joel M Kralj, Angela K Hinz, Edward A Kuhn, Brett A Chromy, Brent W Segelke, Kenneth J Rothschild, Julia E Fletcher, Federico Katzen, Todd C Peterson, Wieslaw A Kudlicki, Graham Bench, Paul D Hoeprich, and Matthew a Coleman. Cell-free co-expression of functional membrane proteins and apolipoprotein, forming soluble nanolipoprotein particles. *Molecular & Cellular Proteomics*, 7(11):2246–53, November 2008.

[13] Petra S Dittrich, Michael Jahnz, and Petra Schwille. A new embedded process for compartmentalized cell-free protein expression and on-line detection in microfluidic devices. *ChembioChem*, 6:811–814, May 2005.

[14] Takatoki Yamamoto, Takahiko Fujii, and Teruo Nojima. PDMS-glass hybrid microreactor array with embedded temperature control device. Application to cell-free protein synthesis. *Lab on a Chip*, 2(4):197–202, November 2002.

[15] Qian Mei, Carl K Fredrickson, Shouguang Jin, and Z Hugh Fan. Toxin detection by a miniaturized in vitro protein expression array. *Analytical Chemistry*, 77(17):5494–500, September 2005.

[16] Ruba Khnouf, David J Beebe, and Z Hugh Fan. Cell-free protein expression in a microchannel array with passive pumping. *Lab on a Chip*, 9(1):56–61, January 2009.

[17] Linas Mazutis, Jean-christophe Baret, Patrick Treacy, Yousr Skhiri, Fallah Araghi, Michael Ryckelynck, and Andrew D Griffiths. Multi-step microfluidic droplet processing: kinetic analysis of an in vitro translated enzyme. *Lab on a Chip*, 9:2902–2908, 2009.

[18] Sebastian J Maerkl and Stephen R Quake. A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. *Science*, 315(233), 2007.

[19] Doron Gerber, Sebastian J Maerkl, and Stephen R Quake. An in vitro microfluidic approach to generating networks. *Nature Methods*, 6(1):71–74, 2009.

[20] Yoshihiro Shimizu, Inoue Akio, Yukihide Tomari, Tsutomu Suzuki, Takashi Yokogawa, Kazuya Nishikawa, and Takuya Ueda. Cell-free translation reconstituted with purified components. *Nature Biotechnology*, 19:751–5, August 2001.

[21] Yoshihiro Shimizu, Takashi Kanamori, and Takuya Ueda. Protein synthesis by pure translation systems. *Methods*, 36:299–304, July 2005.

[22] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Rob Phillips, and Terence Hwa. Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development*, 15:116–124, 2005.

[23] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Thomas Kuhlman, Rob Phillips, and Terence Hwa. Transcriptional regulation by the numbers: applications. *Current Opinion in Genetics & Development*, 15:125–135, 2005.

[24] Francis H. C. Crick. On protein synthesis. *Symp. Soc. Exp. Biol*, 12:138–163, 1958.

[25] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter, editors. *Molekularbiologie der Zelle*. VILEY-VCH Verlag GmbH & Co. KgaA, Weinheim, 4 edition, 2004.

[26] Christian B Anfinsen, E Haber, M Sela, and F H White. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.*, 47(9):1309–1314, 1961.

[27] Christian B Anfinsen. Principles that Govern the Folding of Protein Chains. *Science*, 181(4096):223–230, 1973.

[28] Anmoll at de.wikipedia. Sekundärstruktur eines RNA bindenden Heterodimers. accessed 4 May 2012, `http://de.wikipedia.org/wiki/Sekundarstruktur`.

[29] Wikipedia. Posttranslational modification. accessed 11 Feb 2012, `http://en.wikipedia.org/wiki/Posttranslational_modification`.

[30] Alexander S Spirin. Cell-free protein synthesis. In *RTS application manual for cell-free protein synthesis*, volume 9, pages 7–12. May.

[31] John Littlefield, Elizabeth Keller, Jerome Gross, and Paul Zamecnik. Studies on cytoplasmic ribonucleoprotein particles from the liver of the rat. *Journal of Biological Chemistry*, 217:111–124, 1955.

[32] D Schachtschabel and W Zillig. Untersuchungen zur Biosynthese der Proteine , I Über den Einbau 14C-markierter Aminosäuren ins Protein zellfreier Nucleoproteid-Enzym-Systeme aus Escherichia coli B. *Hoppe-Seyler's Zeitschrift für physiologische Chemie*, 314:262–275, 1959.

[33] Marvin R Lamborg and Paul C Zamecnik. Amino acid incorporation into protein by extracts of E. Coli. *Biochimica et Biophysica Acta*, 42:206–211, 1960.

[34] A Tissières, D Schlessinger, and Gros Francoise. Amino acid incorporation into proteins by E. coli ribosomes. *Proc. Natl. Acad. Sci. U.S.A.*, 46:1450–1463, 1960.

[35] K. McQuillen, R. B. Roberts, and R. J. Britten. Synthesis of nascent protein by ribosomes in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, 45:1437–1447, Sep 1959.

[36] M. W. Nirenberg and J. H. Matthaei. 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, Oct 1961.

[37] Hideo Nakano and Tsuneo Yamane. Cell-free protein synthesis systems. *Biotechnology Advances*, 16:367–384, 1998.

[38] Geoffrey Zubay. In vitro synthesis of protein in microbial systems. *Annu. Rev. Genet.*, 7:267–287, 1973.

[39] H. R. Pelham and R. J. Jackson. An efficient mRNA-dependent translation system from reticulocyte lysates. *Eur. J. Biochem.*, 67:247–256, Aug 1976.

[40] B. E. Roberts and B. M. Paterson. Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. *Proc. Natl. Acad. Sci. U.S.A.*, 70:2330–2334, Aug 1973.

[41] Federico Katzen, Geoffrey Chang, and Wieslaw Kudlicki. The past, present and future of cell-free protein synthesis. *Trends in Biotechnology*, 23(3):150–156, March 2005.

[42] L. A. Ryabova, D. Desplancq, A. S. Spirin, and A. Pluckthun. Functional antibody production using cell-free translation: effects of protein disulfide isomerase and chaperones. *Nat. Biotechnol.*, 15:79–84, Jan 1997.

[43] XiuPing Jiang, Yuji Ookubo, Ikuo Fujii, Hideo Nakano, and Tsuneo Yamane. Expression of fab fragment of catalytic antibody 6d9 in an escherichia coli in vitro coupled transcription/translation system. *FEBS Letters*, 514:290 – 294, 2002.

[44] V. A. Kolb, E. V. Makeyev, and A. S. Spirin. Co-translational folding of an eukaryotic multidomain protein in a prokaryotic translation system. *J. Biol. Chem.*, 275:16597–16601, Jun 2000.

[45] H. Tarui, M. Murata, I. Tani, S. Imanishi, S. Nishikawa, and T. Hara. Establishment and characterization of cell-free translation/glycosylation in insect cell (Spodoptera frugiperda 21) extract prepared with high pressure treatment. *Appl. Microbiol. Biotechnol.*, 55:446–453, May 2001.

[46] Z. Zhang, J. Gildersleeve, Y. Y. Yang, R. Xu, J. A. Loo, S. Uryu, C. H. Wong, and P. G. Schultz. A new strategy for the synthesis of glycoproteins. *Science*, 303:371–373, Jan 2004.

[47] CyberCell Database: CCDB. *E. Coli* Statistics. accessed 27 march 2009, `http://www.wishartlab.com/CCDB/`.

[48] Monya Baker. Hollywood gives biologists a helping hand. *Nature News*, 2009.

[49] R G Kim and C Y Choi. Expression-independent consumption of substrates in cell-free expression system from Escherichia coli. *Journal of Biotechnology*, 84:27 – 32, 2000.

[50] Alexander S Spirin, Vladimir I. Baranov, Lubov A. Ryabova, Sergey Yu. Ovodov, and Yuly B. Alakhov. A Continuous Cell-Free Translation System Capable of Producing Polypeptides in High Yield. *Science*, 242:1162–1164, 1988.

[51] D M Kim and J R Swartz. Prolonging cell-free protein synthesis with a novel ATP regeneration system. *Biotechnology and Bioengineering*, 66(3):180–8, January 1999.

[52] Alexander S Spirin. High-throughput cell-free systems for synthesis of functionally active proteins. *Trends in Biotechnology*, 22(10):538–545, 2004.

[53] Arjun Raj, Charles S Peskin, Daniel Tranchina, Diana Y Vargas, and Sanjay Tyagi. Stochastic mRNA synthesis in mammalian cells. *PLoS Biology*, 4(10):1707–1719, October 2006.

[54] Jonathan M Raser and Erin K O'Shea. Control of Stochasticity in Eukaryotic Gene Expression. *Science*, 304:1811–1814, June 2004.

[55] Mads Kaern, Timothy C Elston, William J Blake, and James J Collins. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews. Genetics*, 6(6):451–64, June 2005.

[56] Gerlinde Schwake, Simon Youssef, Jan-Timm Kuhr, Sebastian Gude, Maria Pamela David, Eduardo Mendoza, Erwin Frey, and Joachim O Rädler. Predictive modeling of non-viral gene transfer. *Biotechnology and Bioengineering*, 105(4):805–13, March 2010.

[57] Jeff Hasty, David McMillen, Farren Isaacs, and James J Collins. Computational Studies Of Gene Regulatory Networks: In Numero Molecular Biology. *Nature Reviews. Genetics*, 2(April):268–279, 2001.

[58] D. Davenport and J. A. C. Nicol. Luminescence in Hydromedusae. *Royal Society of London Proceedings Series B*, 144:399–411, November 1955.

[59] Osamu Shimomura, Frank H Johnson, and Yo Saiga. Extraction , Purification and Properties of Aequorin , a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *J Cell Comp Physiol.*, 59:223–239, 1962.

[60] James G Morin and J W Hastings. Energy Transfer in a Bioluminescent System. *Journal of Cellular Physiology*, 77:313–318, 1970.

[61] Hiroshi Morise, Osamu Shimomura, Frank H Johnson, and John Winant. Intermolecular Energy Transfer in the Bioluminescent System of Aequorea. *Biochemistry*, 13(12):2656–2662, 1974.

[62] Osamu Shimomura. Structure of the chromophore of aequorea green fluorescent protein. *FEBS Letters*, 104(2):2–4, 1979.

[63] Chris W Cody, Douglas C Prasher, William M Westler, Franklyn G Prendergast, and William W Ward J. Chemical Structure of the Hexapeptide Chromophore of the Aequorea Green-Fluorescent Protein. *Biochemistry*, 32:1212–1218, 1993.

[64] Fan Yang, Larry G. Moss, and George N. Jr. Phillips. The molecular structure of green fluorescent protein. *Nature Biotechnology*, 14:1246–1251, 1996.

[65] Douglas C Prasher, Virginia K Eckenrode, William W Ward, Frank G Prendergast, and Milton J Cormier. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene*, 111:229–233, 1992.

[66] Martin Chalfie, Yuan Tu, Ghia Euskirchen, William W Ward, and Douglas C Prasher. Green Fluorescent Protein as a Marker for Gene Expression. *Science*, 263(1988):802–805, 1994.

[67] Roger Heim, Andrew B. Cubitt, and Roger Y. Tsien. Improved green fluorescence. *Nature*, 373:663–664, 1995.

[68] Nobelprize.org. The nobel prize in chemistry 2008. accessed 7 Jan 2012, `http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2008/`.

[69] Osamu Shimomura. The discovery of aequorin and green fluorescent protein. *Journal of Microscopy*, 217:3–15, 2005.

[70] GFP – Green fluorescent protein. accessed 13 Jan 2012, `http://gfp.conncoll.edu/tsien.html`.

[71] Roger Heim, Douglas C Prasher, and Roger Y Tsien. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.*, 91(December):12501–12504, 1994.

[72] Olympus. Formation of the gfp fluorophore. accessed 15 Jan 2012, `http://www.olympusconfocal.com/java/fpfluorophores/gfpfluorophore/index.html`.

[73] Wikipedia. Gfp and fluorophore. accessed 15 Jan 2012, `http://upload.wikimedia.org/wikipedia/commons/f/f6/Gfp_and_fluorophore.png`.

[74] Harold C Urey, F G Brickwedde, and G M Murphy. A Hydrogen Isotope of Mass 2. *Phys. Rev.*, 31(1):164–165, 1932.

[75] Gilbert N Lewis and Ronald T Macdonald. Concentration of H2 Isotope. *J. Chem. Phys.*, 1(6):341–344, 1933.

[76] Kenneth B Wiberg. The Deuterium Isotope Effect. *Chem. Rev.*, 55(4):713–743, 1955.

[77] Xinji Guo and James G Kempf. Enhanced , simplified expression of perdeuterated hemoglobin for NMR structure and dynamics. *Protein Expression and Purification*, 72:8–18, 2010.

[78] SNO – The Sudbury Neutrino Observatory. accessed 22 May 2012, `http://www.sno.phy.queensu.ca/`.

[79] Douglas Magde, Elliot L Elson, and Watt W Webb. Thermodynamic Fluctuations in a Reaction System - Measurement by Fluorescence Correlation Spectroscopy. *Physical Review Letters*, 29(11):705–708, 1972.

[80] Elliot L Elson and Douglas Magde. Fluorescence Correlation Spectroscopy . I . Conceptual Basis and Theory. *Biopolymers*, 13:1–27, 1974.

[81] Douglas Magde, Elliot L Elson, and Watt W Webb. Fluorescence Correlation Spectroscopy. II. An Experimental Realization. *Biopolymers*, 13:29–61, 1974.

[82] Eugene P Petrov and Petra Schwille. Translational Diffusion in Lipid Membranes beyond the Saffman-Delbrück Approximation. *Biophysical Journal*, pages L41–L43, 2008.

[83] S Tyagi and F R Kramer. Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotechnology*, 14:303–308, 1996.

[84] Theodor Förster. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik*, 437(1-2):55–75, 1948.

[85] Jeremy S Paige, Karen Y Wu, and Samie R Jaffrey. RNA Mimics of Green Fluorescent Protein. *Science*, 333:642–646, 2011.

[86] Randall K Saiki, Stephen J Scharf, Fred Faloona, Kary B Mullis, Glenn T Horn, Henry A Erlich, and Norman Arnheim. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230:1350–1354, 1985.

[87] Randall K Saiki, David H Gelfand, Susanne Stoffel, Stephen J Scharf, Russell Higuchi, Glenn T Horn, Kary B Mullis, and Henry A Erlich. Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science*, 239:487–491, 1988.

[88] Eyal Karzbrun, Jonghyeon Shin, Roy Bar-Ziv, and Vincent Noireaux. Coarse-Grained Dynamics of Protein Synthesis in a Cell-Free System. *Physical Review Letters*, 106(4):1–4, January 2011.

[89] Pierre C Jelenc and C G Kurland. Nucleoside triphosphate regeneration decreases the frequency of translation errors. *Proc. Natl. Acad. Sci. U.S.A.*, 76(7):3174–3178, 1979.

[90] Michael C Jewett and James R Swartz. Substrate replenishment extends protein synthesis with an in vitro translation system designed to mimic the cytoplasm. *Biotechnology and Bioengineering*, 87(4):465–72, August 2004.

[91] Kensuke Osada, Hiroki Oshima, Daigo Kobayashi, Motoyoshi Doi, Manabu Enoki, Yuichi Yamasaki, and Kazunori Kataoka. Quantized Folding of Plasmid DNA Condensed with Block Catiomer into Characteristic Rod Structures Promoting Transgene Efficacy. *JACS*, 132(35):12343–12348, 2010.

[92] Judith A Megerle, Georg Fritz, Ulrich Gerland, Kirsten Jung, and Joachim O Radler. Timing and Dynamics of Single Cell Gene Expression in the Arabinose Utilization System. *Biophysical Journal*, 95(August):2103–2115, 2008.

[93] Srinivas S Sastry and Barbara M Ross. Nuclease Activity of T7 RNA Polymerase and the Heterogeneity of Transcription Elongation Complexes. *Biochemistry*, 272(13):8644 –8652, 1997.

[94] Maribeth Maslak and Craig T Martin. Kinetic Analysis of T7 RNA Polymerase Transcription Initiation from Promoters Containing Single-Stranded Regions. *Biochemistry*, 32:4281–4285, 1993.

[95] Shuntaro Takahashi, Ryoko Akita, Hisao Matsuno, Hiroyuki Furusawa, Yoshihiro Shimizu, Takuya Ueda, and Yoshio Okahata. 70 S Ribosomes Bind to Shine-Dalgarno Sequences without Required Dissociations. *ChemBioChem*, 9:870 – 873, 2008.

[96] J. Zabner, A. J. Fasbender, T. Moninger, K. A. Poellinger, and M. J. Welsh. Cellular and molecular barriers to gene transfer by a cationic lipid. *J. Biol. Chem.*, 270(32):18997–19007, Aug 1995.

[97] M. Wilke, E. Fortunati, M. van den Broek, A. T. Hoogeveen, and B. J. Scholte. Efficacy of a peptide-based gene delivery system depends on mitotic activity. *Gene Ther.*, 3(12):1133–1142, Dec 1996.

[98] S. Brunner, T. Sauer, S. Carotta, M. Cotten, M. Saltik, and E. Wagner. Cell cycle dependence of gene transfer by lipoplex, polyplex and recombinant adenovirus. *Gene Ther.*, 7(5):401–407, Mar 2000.

[99] J. F. Arthur, L. H. Butterfield, M. D. Roth, L. A. Bui, S. M. Kiertscher, R. Lau, S. Dubinett, J. Glaspy, W. H. McBride, and J. S. Economou. A comparison of gene transfer methods in human dendritic cells. *Cancer Gene Ther.*, 4(1):17–25, 1997.

[100] I. Strobel, S. Berchtold, A. Gotze, U. Schulze, G. Schuler, and A. Steinkasserer. Human dendritic cells transfected with either RNA or DNA encoding influenza matrix protein M1 differ in their ability to stimulate cytotoxic T lymphocytes. *Gene Ther.*, 7(23):2028–2035, Dec 2000.

[101] David A Hume. Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. *Blood*, 96(7):2323–2328, 2000.

[102] Andrea Sacchetti, Tarek El Sewedy, Ashraf F Nasr, and Saverio Alberti. Efficient GFP mutations profoundly affect mRNA transcription and translation rates. *FEBS letters*, 492:151–155, 2001.

[103] E W A Grudzien, Janusz Stepinski, Marzena Jankowska-Anyszka, Ryszard Stolarski, Edward Darzynkiewicz, and Robert E Rhoads. Novel cap analogs for in vitro synthesis of mRNAs with high translational efficiency. *RNA*, 10:1479–1487, 2004.

[104] Shyamsundar Subramanian and Friedrich Srienc. Quantitative analysis of transient gene expression in mammalian cells using the green fluorescent protein. *Journal of Biotechnology*, 49:137–151, 1996.

[105] Silke Holtkamp, Sebastian Kreiter, Abderraouf Selmi, Petra Simon, Michael Koslowski, Christoph Huber, Özlem Türeci, and Ugur Sahin. Modification of antigen-encoding RNA increases stability , translational efficacy , and T-cell stimulatory capacity of dendritic cells. *Blood*, 108(13):4009–4017, 2006.

[106] Joachim O Rädler, Ilya Koltover, Tim Salditt, and Cyrus R Safinya. Structure of DNA - Cationic Liposome Complexes : DNA Intercalation in Multilamellar Membranes in Distinct Interhelical Packing Regimes. *Science*, 810(1997), 1997.

[107] Shirley S Daube, Dan Bracha, Amnon Buxboim, and Roy H Bar-Ziv. Compartmentalization by directional gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 107(7):2836–2841, 2010.

[108] Kaname Hirai, Mihoko Tomida, Yuichiro Kikuchi, Ohmi Ueda, Hiroshi Ando, and Naokazu Asanuma. Effects of Deuterium Oxide on Streptococcus mutans and Pseudomonas aeruginosa. *Bull Tokyo Dent Coll*, 51:175–183, 2010.

[109] Gilbert N Lewis. The Biochemistry of Water Containing Hydrogen Isotope. *J. Am. Chem. Soc.*, 55(8):3503–3504, 1933.

[110] Gilbert N Lewis. The Biology of heavy water. *Science*, 79(2042):151–153, 1934.

[111] M. Ikeda, M. Hirono, M. Kishio, J. Matsuura, M. Sakakibara, and T. Yoshioka. Examination of microgravity effects on spontaneous Ca2+ oscillations in AtT20 pituitary cells using heavy water. *J Gravit Physiol*, 7(2):P63–64, Jul 2000.

[112] Masayuki Ikeda, Shigeru Suzuki, Masahiro Kishio, Moritoshi Hirono, Takashi Sugiyama, Junko Matsuura, Teppei Suzuki, Takayuki Sota, Charles N Allen, Shiro Konishi, and Tohru Yoshioka. Hydrogen-Deuterium Exchange Effects on beta-Endorphin Release from AtT20 Murine Pituitary Tumor Cells. *Biophysical Journal*, 86:565–575, 2004.

[113] H V Samis, M B Baird, and H R Massie. Deuterium Oxide Effect on Temperature-Dependent Survival in Populations of Drosophila melanogaster. *Science*, 183:427–428, 1973.

[114] Inglis J Miller and Gregory Mooser. Taste Responses to Deuterium Oxide. *Physiology & Behavior*, 23:69–74, 1979.

[115] A S Goryunov. H / D Isotope Effects on Protein Hydration and Interaction in Solution. *Gen. Physiol. Biophys.*, 25:303–311, 2006.

[116] Hans J Altermatt, Jan-Olaf Gebbers, and Jean A Laissue. Heavy Water Delays Growth of Human Carcinoma in Nude Mice. *Cancer*, 62(3):462–466, 1988.

[117] Hans Jörg Altermatt, Jan-Olaf Gebbers, and Jean Albert Laissue. Heavy Water Enhances The Antineoplastic Effects Of 5-Fluoro-Uracil And Bleomycin In Nude Mice Bearing Human Carcinoma. *Int. J. Cancer*, 45:475–480, 1990.

[118] Hiroshi Et Al. Takeda. Mechanisms of cytotoxic effects of heavy water on cancer cells. *Anti-Cancer Drugs*, 1998.

[119] Johannes Hartmann, Yvonne Bader, Zsuzsanna Horvath, Philipp Saiko, Michael Grusch, Christoph Illmer, Sibylle Madlener, Monika Fritzer-Szekeres, Nicole Heller, Rudolf-Giesbert Alken, and Thomas Szekeres. Effects of Heavy Water (D2O) on Human Pancreatic Tumor Cells. *Anticancer Research*, 25:3407–3411, 2005.

[120] Koh Takeuchi, Elise Ng, Thomas J Malia, and Gerhard Wagner. 1- 13 C amino acid selective labeling in a 2 H 15 N background for NMR studies of large proteins. *Journal of Biomolecular NMR*, 38:89–98, 2007.

[121] Kithsiri Herath, Gowi Bhat, Paul L Miller, Sheng-ping Wang, Alison Kulick, Genevieve Andrews-kelly, Christopher Johnson, Rory J Rohm, Michael E Lassman, Stephen F Previs, Douglas G Johns, Brian K Hubbard, and Thomas P Roddy. Equilibration of 2 H labeling between body water and free amino acids : Enabling studies of proteome synthesis. *Analytical Biochemistry*, 415:197–199, 2011.

[122] Alexander Koglin, Christian Klammt, Nikola Trbovic, Daniel Schwarz, Birgit Schneider, Birgit Schäfer, Frank Löhr, Frank Bernhard, and Volker Dötsch. Combination of cell-free expression and NMR spectroscopy as a new approach for structural investigation of membrane proteins. *Magn. Reson. Chem.*, 44:S17–S23, July 2006.

[123] Andrew L Hopkins and Colin R Groom. The druggable genome. *Nature Reviews. Drug Discovery*, 1(September):727–730, 2002.

[124] Jessica J Wuu and James R Swartz. High yield cell-free production of integral membrane proteins without refolding or detergents. *Biochimica et Biophysica Acta*, 1778:1237 – 1250, 2008.

[125] Christian Klammt, Daniel Schwarz, Frank Löhr, Birgit Schneider, Volker Dötsch, and Frank Bernhard. Cell-free expression as an emerging technique for the large scale production of integral membrane protein. *FEBS Journal*, 273:4141–4153, 2006.

[126] Eva-Kathrin Sinner, Ute Reuning, Fatma Nese Kök, Barbara Saccà, Luis Moroder, Wolfgang Knoll, and Dieter Oesterhelt. Incorporation of integrins into artificial planar lipid membranes: characterization by plasmon-enhanced fluorescence spectroscopy. *Analytical Biochemistry*, 333(2):216–24, October 2004.

[127] Shin-ichiro M Nomura, Satoshi Kondoh, Wakiko Asayama, Akikazu Asada, Shigemichi Nishikawa, and Kazunari Akiyoshi. Direct preparation of giant proteoliposomes by in vitro membrane protein synthesis. *Journal of Biotechnology*, 133:190–195, 2008.

[128] Takuma Ohtsuka, Satoshi Neki, Tamotsu Kanai, Kazunari Akiyoshi, Shin-ichiro M Nomura, and Takashi Ohtsuki. Synthesis and in situ insertion of a site-specific fluorescently labeled membrane protein into cell-sized liposomes. *Analytical Biochemistry*, 418:97–101, 2011.

[129] Valentina Cauda, Hanna Engelke, Anna Sauer, Delphine Arcizet, Joachim Rädler, and Thomas Bein. Colchicine-Loaded Lipid Bilayer-Coated 50 nm Mesoporous Nanoparticles Efficiently Induce Microtubule Depolymerization upon Cell Uptake. *Nano Lett.*, 10:2484–2492, 2010.

[130] Raghuveer Parthasarathy and Jay T Groves. Curvature and spatial organization in biological membranes. *Soft Matter*, 3:24–33, 2007.

[131] Frank Jülicher and Reinhard Lipowsky. Shape transformations of vesicles with intramembrane domains. *Phys. Rev. E*, 53(3):2670–2683, 1996.

[132] Tobias Baumgart, Samuel T Hess, and Watt W Webb. Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. *Nature*, 425:821–824, 2003.

[133] Tae-Young Yoon, Cherlhyun Jeong, Sang-Wook Lee, Joon Heon Kim, Myung Chul Choi, Sung-Jin Kim, Mahn Won Kim, and Sin-Doo Lee. Topographic control of lipid-raft reconstitution in model membranes. *Nature Materials*, 5:281–285, 2006.

[134] Raghuveer Parthasarathy, Cheng-han Yu, and Jay T Groves. Curvature-Modulated Phase Separation in Lipid Bilayer Membranes. *Langmuir*, 22:5095–5099, 2006.

[135] Nicole K Noren, Mark Lu, Andrew L Freeman, Mitchell Koolpe, and Elena B Pasquale. Interplay between EphB4 on tumor cells and vascular ephrin-B2 regulates tumor growth. *Proc. Natl. Acad. Sci. U.S.A.*, 101(15):5583–5588, 2004.

[136] Christian Dietrich, Zoya N Volovyk, Moshe Levi, Nancy L Thompson, and Ken Jacobson. Partitioning of Thy-1, GM1, and cross-linked phospholipid analogs into lipid rafts reconstituted in supported model membrane monolayers. *Proc. Natl. Acad. Sci. U.S.A.*, 98(19):10642–10647, 2001.

[137] Luis A Bagatolli. To see or not to see : Lateral organization of biological membranes and fluorescence microscopy. *Biochimica et Biophysica Acta*, 1758:1541–1556, 2006.

[138] Martin Hennig, Jürgen Neumann, Achim Wixforth, Joachim O Rädler, and Matthias F Schneider. Dynamic patterns in a supported lipid bilayer driven by standing surface acoustic waves. *Lab on a Chip*, 9:3050–3053, 2009.

[139] J Neumann, M Hennig, A Wixforth, S Manus, J O Rädler, and M F Schneider. Transport , Separation , and Accumulation of Proteins on Supported Lipid Bilayers. *Nano Lett.*, 10:2903–2908, 2010.

[140] Long Cai, Nir Friedman, and X Sunney Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440(7082):358–62, March 2006.

[141] Gene-Wei Li and X Sunney Xie. Central dogma at the single-molecule level in living cells. *Nature*, 475:308–315, 2011.