Optimisation of a transcription-translation coupled *in vitro* system

Inaugural-Dissertation zur Erlangung des Doktorgrades des Fachbereichs Biologie, Chemie und Pharmazie der Freien Universität Berlin

Vorgelegt von
Madina Iskakova
aus Almaty, Kasachstan
2005

Gutachter:

Herr Prof. Dr. Knud. H. Nierhaus

Herr Prof. Dr. Mathias Ziegler

Datum der Dissertation:

25. Mai 2005

Summary

Cell-free protein synthesis exploits the catalytic machinery of the cell to produce active proteins. An *in vitro* system is flexible and well controlled, and it offers several advantages over conventional *in vivo* technologies such as easy ways for purification, synthesis of regulatory and/or toxic proteins, incorporation of artificial or modified amino acids that might be doted with isotopes required for NMR.

Here I describe experiments exploring optimisation possibilities concerning yield and quality of the synthesised protein. Some experimental strategies also include expression of eukaryotic genes in prokaryotic expression systems.

The following results have been achieved:

1: Quality criteria developed that allow a critical evaluation of parameters important for the coupled transcription/translation system or improving the yield and quality of the synthesized protein exploiting the features of the green fluorescent protein GFP.

2: The standard transcriptase used in overexpression studies *in vivo* and *in vitro* is the T7 polymerase. The fundamental difficulty with this enzyme is the fact that it is about six times faster than the *E. coli* transcriptase and thus uncouples transcription from translation, a possible reason for the fact that *in vitro* systems usually produce proteins with an activity of 30 to 60% only. We tested some slow mutants of T7 polymerase that approached the rate of the *E. coli* transcriptase and observed indeed a significant improvement up to 100% of the active fraction, although at the cost of lower yields.

3: A similar improvement of the active fraction was observed at lower incubation temperatures down to 20°C, again at the cost of lower yields.

4: According to literature data some amino acids are metabolised during *in vitro* incubations and thus could cause a limitation of protein synthesis. Indeed, we demonstrate that a second addition of amino acids in the middle of the incubation triggers a burst of further protein synthesis. Using this trick at 20°C

pushed the yield of protein to almost that seen at 30°C, but now with an active fraction of 100%. In contrast, our analysis revealed that NTPs are not limiting the gene expression *in vitro* in our system (modified Roche RTS).

5: It is known that the codon usage of highly and lowly expressed proteins in *E. coli* differs dramatically. When we examined this point with human genes, to our surprise a corresponding difference could not be observed. Due to this fact it was possible to identify 11 tRNAs the corresponding codon are quite often used in human genes but rarely in *E. coli* genes. Therefore, for a good expression of eukaryotic genes in *E. coli* systems these 11 tRNAs should be added (and not only the 7 tRNAs supplied in systems from Novagen).

6: I outlined some ways to improve further the expression system.

Zusammenfassung

Zellfreie Proteinsynthese benutzt den Translations- und manchmal auch den Transkriptionsapparat der Zelle zur Synthese aktiver Proteine. Ein *in vitro* System ist flexibel und gut kontrollierbar, und es birgt zahlreiche Vorteile gegenüber *in vivo* Techniken, Beispiele sind Synthese von regulativen und/oder toxischen Proteinen, Einbau von artifiziellen oder modifizierten Aminosäuren, die sogar mit seltenen Isotopen für NMR Untersuchungen dotiert sein können.

In dieser Arbeit beschreibe ich Experimente, die Optimierungs-möglichkeiten bezüglich Ertrag und Qualität der synthetisierten Proteine untersuchen. Einige Experimentalstrategien beziehen sich auch auf die Expression eukaryontischer Gene in prokaryontischen Systemen.

Folgende Ergebnisse wurden erreicht:

1: Qualitätskriterien wurden entwickelt, die eine präzise Bestimmung der synthetisierten Proteinmenge als auch die aktive Fraktion zu messen gestatten. Dabei wurden die Eigenschaften des grün-fluoreszierenden Proteins GFP ausgenutzt.

2: Als Standard-Transkriptase wird *in vivo* als auch *in vitro* T7 Polymerase benutzt. Die grundsätzliche Schwierigkeit mit dieser Transkriptase beruht darauf, dass sie etwa sechs-mal schneller ist als die *E. coli* Transkriptase und damit Transkription von der Translation entkoppelt. Das ist vermutlich der Grund, weshalb *in vitro* Systeme Proteine produzieren, die nur zu 30 bis 60% aktiv sind. Wir testeten einige langsame T7 Polymerasen-Mutanten und beobachteten tatsächlich eine Verbesserung bis zu 100% der aktiven Fraktion, allerdings auf Kosten des Ertrags.

3: Eine ähnliche Verbesserung der aktiven Fraktion wurde bei verminderter Inkubationstemperatur (bis 20°C) beobachtet, jedoch wieder auf Kosten der synthetisierten Menge.

4: Von Literaturdaten wissen wir, dass einige Aminosäuren während der Inkubation metabolisiert werden, was die Proteinsynthese limitieren könnte.

Tatsächlich konnten wir zeigen, dass eine zweite Zugabe von Aminosäuren in der Mitte der Inkubation einen dramatischen Schub der Proteinsynthese auslöste. Die Kombination dieses Tricks mit einer Inkubation bei 20°C vermehrte die Proteinmenge zu der, die bei 30°C gefunden wurde, jetzt aber mit 100% aktiver Fraktion.

5: Es ist bekannt, dass der Codon-Gebrauch bei hoch und niedrig exprimierten Proteinen in *E. coli* deutlich unterschiedlich ist. Eine entsprechende Untersuchung von hoch und niedrig exprimierten Genen im menschlichen Genom offenbarte zu unserer Überraschung keinen unterschiedlichen Codongebrauch. Deshalb war es möglich, 11 tRNA anzugeben, deren zugehörige Codone recht häufig in eukaryontischen mRNAs anzutreffen sind, aber selten in *E. coli* mRNAs vorkommen. Diese 11 tRNAs sollten zur *E. coli* Gesamt-tRNA zugegeben werden, um eine optimale Expression eukaryontischer mRNA in *E. coli* Systemen zu gewährleisten. Unser Ergebnis kontrastiert zu den 7 tRNAs in Zahl und Art, die von Novagen angegeben werden.

6: Ich gebe schließlich zusätzliche Hinweise zu einer weiteren Optimierung der *in vitro* Proteinsynthese.

Table of Contents

Abbreviations

AA	acrylamide
aa-tRNA	aminoacyl-tRNA
Å	Angstrom
Asn	asparagine
Asp	aspartic acid
ATP	adenosine triphosphate
BAA	bis-acrylamide
BPB	bromphenol blue
BSA	bovine serum albumine
cpm	counts per minute
Da	dalton
DNA	desoxyribonucleic acid
dsDNA	double strand DNA
DTT	dithiothreitol
EF-G	elongation factor G
EF-Ts	elongation factor thermo stable
EF-Tu	elongation factor thermo unstable
F.I.	fragmentation index
J	joule
GDP	guanine diphosphate
GFP	green fluorescent protein
Gly	glycine
GTP	guanine triphosphate
His	histidine
HPLC	high performance liquid chromatography
HSS	HEPES salts solution
HSWP	high-salt washed proteins

IF	initiation factor
IPTG	isopropyl-beta-D-thiogalactoside
IRES	internal ribosomal entry site
k-	kilo-, 1000
kb	kilobase, 1,000 nucleotide bases
Leu	leucine
m-	milli-, 0.001
mA	milliampere
µ	micro-, 0.000001
µCi	microcurie
M-	mega-, 1,000,000
MDa	megadaltons
MgAc	magnesium acetate
MQ	milliQ water
mRNA	messenger RNA
M.W.	molecular weight
NaAc	sodium acetate
NH_4Ac	ammonium acetate
nt	nucleotide(s)
NTP	nucleoside triphosphate
Ω	Ohm
PAA(G/gel)	polyacrylamide gel
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEP	phosphoenol pyruvate
PK	pyruvate kinase
Phe	phenylalanine
P.I.	protease inhibitors
Poly(A)	poly-adenine mRNA
Poly(U)	poly-uridine mRNA

PPi	inorganic pyrophosphate
PTF	peptidyl transferase centre
RBS	ribosomal binding site
RF	release factor
RNA	ribonucleic acid
RNAP	RNA polymerase
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal RNA
RRF	ribosome recycling factor
RTS	rapid translation system
S	Svedberg unit (sedimentation coefficient)
SCS	small compound solution
SD	Shine-Dalgarno sequence
SDS	sodium dodecylsulphate
TCA	trichloroacetic acid
tRNA	transfer RNA
Tyr	tyrosine
UTR	untranslated region
V	volt
v/v	volume/volume
WT	wild type
w/v	weight/volume

Chapter 1

INTRODUCTION

More than 100 different molecules participate in the translation of the genetic message. The way this plethora of molecules is orchestrated for protein synthesis is studied mainly by *in vitro* systems. The most global one is the coupled transcription/translation system, where the input is a gene and the output the corresponding protein. It is a homogenous system, which means that it has components from one and the same organisms. One essential component is the cell free extract, also called lysate and/or S30-extract, which contains most of the cellular cytoplasmic compounds necessary for protein synthesis itself, *i.e.* ribosomes, translational factors for initiation, elongation and termination (IFs, EFs, and RFs), all tRNA synthetases (RS), and tRNAs as well. Of course, it also contains the components important for transcription, but here not an *E. coli* RNA polymerase (RNAP), is utilized, but RNAP from the T7 bacteriophage. The gene of interest is introduced into the system on a plasmid or as a linearised double-stranded DNA flanked by the T7 RNAP promoter and terminator, in order to get a transcript – messenger RNA. Of course, the presence of the building blocks for both, mRNA and protein, are also present. These are nucleotides and amino acids. As any other process in cells, also in cell-free systems the energy supplying pathways are utilized e.g. regeneration of ATP, (Kim and Swartz, 2000).

Historically, these kind of cell-free systems have been studied for about 50 years. Several groups independently demonstrated that disruptured cells or their isolated fractions were capable of synthesizing proteins (Borsook *et al.*, 1950; Gale and Folkes, 1954; Winnick, 1950). Meanwhile, in cells ribonucleoprotein particles were observed and identified (Palade, 1955) whose ability to synthesise proteins was experimentally proven (Keller and Littlefield, 1957; Littlefield *et al.*, 1955). The word “ribosome” was coined by Horward Dintzis in 1958 (Rheinberger, 2004, p. 15) to designate their ribonucleoprotein entity. Programmed with endogenous mRNA, this particles were simply reading this

molecules to which they had been already attached at the time of cell disruption. Still the significance of these investigations was great, since they opened the way for reseach of protein biosynthesis on the molecular level, including all the large and small molecules involved into this process, as well as the mechanisms of their interaction with each other (Zamecnik, 1969).

A revolutionary step in the development of cell-free translational systems was the introduction of exogenous messangers. This was first done by Nirenberg and Matthaei in 1961 with a bacterial system (Nirenberg and Matthaei, 1961). They also seem to be the first ones to report dependence of the bacterial cell-free system on the presence of DNA (Matthaei and Nirenberg, 1961). A couped transcription-translation systems came into wide use after some major improvements made by two groups (Gold and Schweiger, 1969; Lederman and Zubay, 1967). In the case of prokariotic systems addition of a proper DNA (plasmid, isolated gene or a synthetic DNA fragment) to the DNA-free extract instead of mRNA results in synthesis of a corresponding mRNA. In this case, ribosomes start to translate the nascent chains of mRNA even prior to the completion of the RNA synthesis. Thus, translation is going on while mRNA is still synthesized and the rates of transcription and translation are coordinated: such systems are called coupled transcritpion-translation systems.

Several types of cell-free systems can be distinguished, the most often applied ones are (i) the so called batch system (Kim and Swartz, 1999; Kim and Swartz, 2000), where all the components are present in one and the same reaction vial, and (ii) a continuous-exchange cell-free (CECF) system (Alakhov *et al.*, 1995). The latter is characterized by a reaction chamber that contains molecules for the synthesis itself, and a supplementary chamber with building blocks for the synthesised product and regeneration of energy. A semi permeable membrane, thus allowing the concentration of the final product, separates these two.

The composition and energy supply of the bacterial batch systems was significantly improved during past 50 years. Kim and Swartz discovered that in

a conventional cell-free synthesis system derived from *E. coli,* phosphoenolpyruvate (PEP), the secondary energy source for ATP regeneration, and several amino acids are rapidly degraded during incubation of the reaction mixture (Kim and Swartz, 1999; Kim and Swartz, 2000). The degradation of such compounds takes place even in the absence of protein synthesis, thus reducing the capacity for it to take place. However, the lost potency is completely recovered when the reaction mixture is supplied with an additional PEP and amino acids, suggesting that catalytic activity is relatively stable.

In several papers Kim and Swartz have published their study of factors that cause early cessation of protein synthesis in cell-free system from *E. coli* (Kim and Swartz, 1999; Kim and Swartz, 2000). They have reported, that PEP, the secondary energy source for ATP regeneration, and several amino acids are rapidly degraded during the cell-free protein synthesis reaction, which severely reduces the capacity for protein synthesis. Of twenty amino acids, only arginine, cysteine, and tryptophan were required to restore system activity. Through coordinated addition of PEP, arginine, cysteine, tryptophan, and magnesium, the final concentration of cell-free synthesised chloramphenicol acetyltransferase (CAT) increased more than 4-fold. The figure 1.0 below schematically describes the conventional ATP regeneration system (A) and that one proposed by Kim and Swartz, 1999 (B).

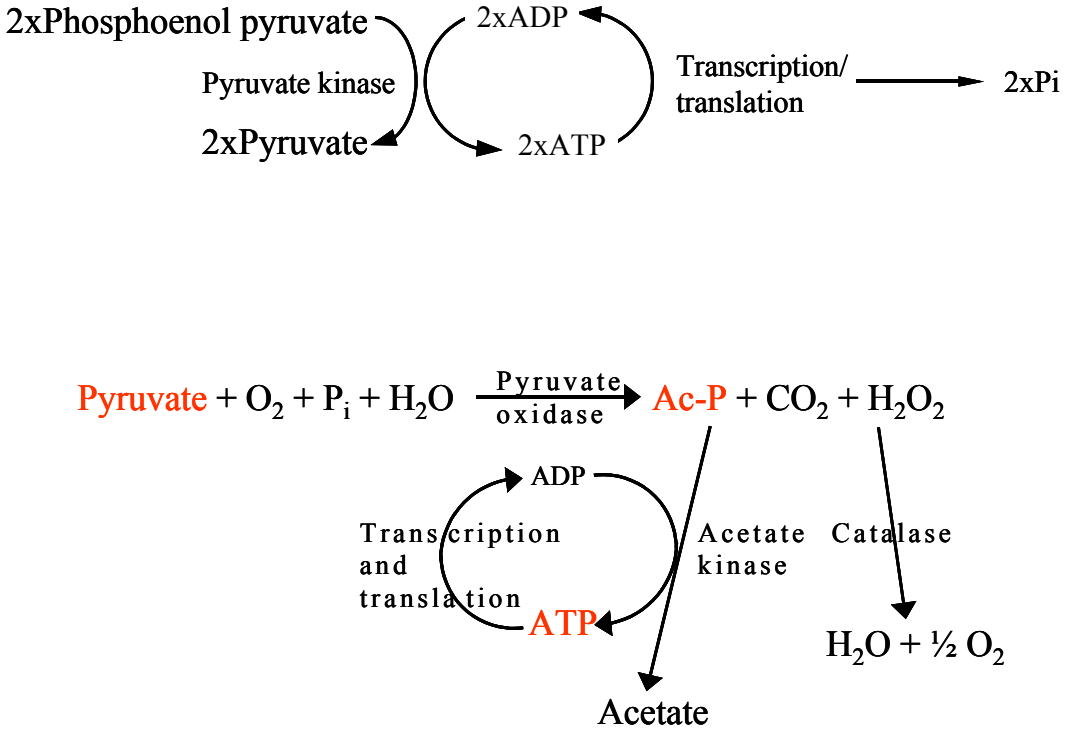


Figure 1.0 Regeneration of ATP in an in vitro *protein synthesis systems.* (**A**) Conventional scheme for energy regeneration using phosphoenol pyruvate and pyruvate kinase. (**B**) A new scheme proposed by (Kim and Swartz, 1999), using pyruvate, pyruvate oxidase and endogenous acetate kinase.

A system like this has a number of advantages compared to *in vivo* over-expression of a target protein. These are:

1. Direct expression from linear or circular DNA template and no need for purification of an mRNA after *in vitro* transcription, when a significant loss of a template can occur, as well as partial or complete degradation of mRNA due to RNases, present in the cell lysate. After all, when transcription occurs in such a coupled system, partial decay of mRNA is always restored by newly transcribed molecules.

2. Expression of outer membrane proteins, study of their assembly and processing in the *in vivo* near conditions can be allowed.

3. Proteins expressed at very low levels *in vivo*, toxic and alike, may be expressed at higher levels *in vitro*, also allowing a closer look on their function.

4. The incorporation of the added labelled or unnatural amino acids, modification of active centre and different protein chains may allow a deeper

understanding of a protein function, may also allow synthesis of novel proteins with a desired function. (Only, application of last is questionable)

5. Relative purity of a product allows studying proteins as they are when (if) not involved into macro-molecular complexes.

6. Studying processes of synthesis and factors involved into the regulation of transcription and translation still remains an up to date subject of interest.

7. Studies of co-translational folding pathways during protein synthesis.

8. Formation of truncated proteins during/after synthesis and means to suppress these unwanted products.

In order to understand all this it's necessary to understand the working principles of the protein synthesis machinery – the ribosome itself. At the moment quiet a lot is known about prokaryotic ribosome, and analysing this knowledge, (putting pieces into a puzzle), allows us to make some of the features of the ribosome to work for us in order to improve what already is achieved on the field of *in vitro* systems for protein synthesis. Because, there are still possibilities to go further in the "*in vivo* near" conditions, to increase protein yield. Besides that, not all proteins can be expressed *in vitro* at high levels, there are several difficulties which have been investigated during time-frame of this work and attempts to improve the last were taken.

First I will focus on the protein synthesis machinery – the ribosome.

1.1 Protein synthesis

The ribosome is a macromolecular complex which catalyses peptide bond formation – a process vital to all organisms. Information is transported from the genome via mediator molecules called messenger RNAs (mRNAs), to the ribosome and it translates the sequence of the codons on the mRNA into the corresponding sequence of amino acids, using adaptor molecules – transfer RNAs (tRNAs). A single ribosome can incorporate 10 to 20 amino acids per second (Bremer and Dennis, 1996) with an accuracy of about one misincorporation per 3000 amino acids incorporations (Bouadloun *et al.*, 1983).

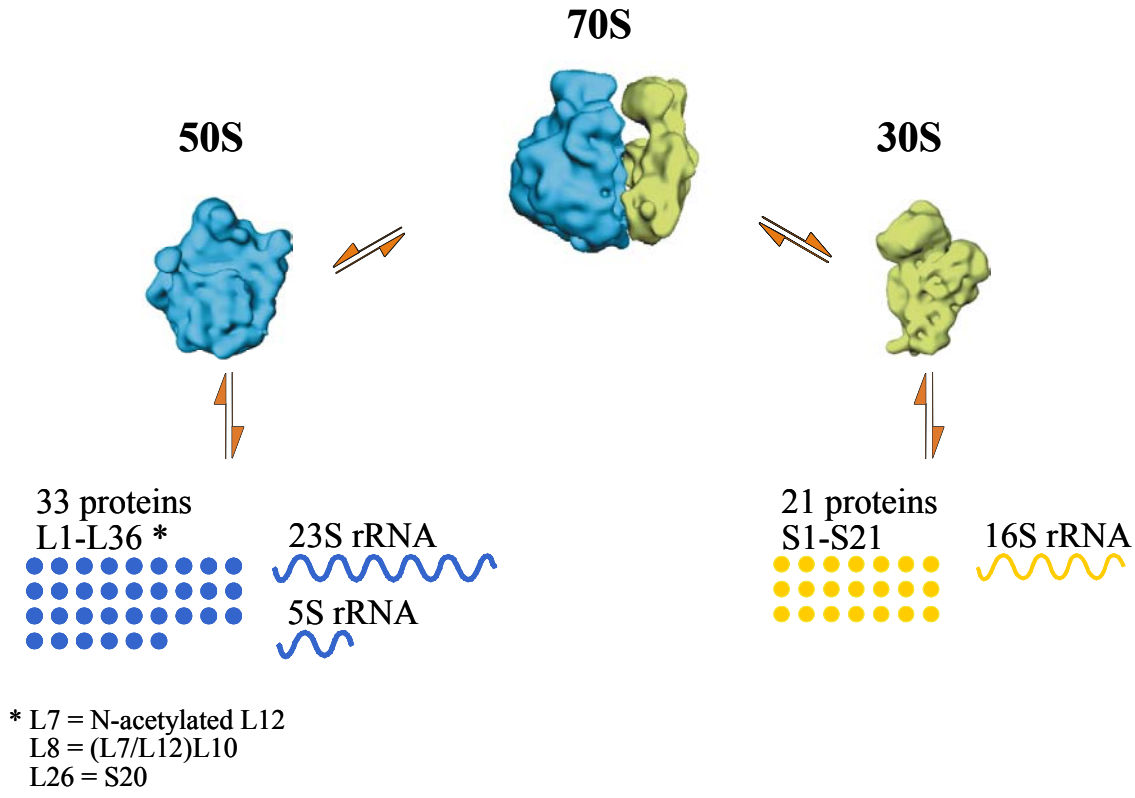


Figure 1.1-1 Composition of the E. coli *ribosome.*

The bacterial ribosome has a mass of approximately 2.6-2.8 MDa, a relative sedimentation of 70 S and a diameter of 200-250 Å. Under certain functional states the 70S ribosome falls apart into two unequal subunits: a large 50S subunit and a small 30S subunit. Each subunit is a ribonucleoprotein particle with one third of the mass consisting of protein and the other two thirds (Moore and Steitz, 2002; Ramakrishnan and Moore, 2001) of ribosomal RNA (rRNA): 33 proteins and two rRNA molecules – 5S (120 nts), and 23S (~2900 nts) rRNA in large subunit, and 20 proteins and a single 16S (~ 1500 nts) rRNA in the 30S small subunit (Figure 1.1-1).

Both subunits differ in functions as well. On the 30S subunit the codon-anticodon interaction between the mRNA and tRNA substrates, namely, process of decoding occurs. The large 50S subunit performs the central catalytic function of protein synthesis, in its active center for peptide bond formation between the nascent polipeptide chain and the incoming aminoacylated tRNA. One of the special features of the 50S subunit is the tunnel for peptide exit. It

runs from the peptidyl-transferase (PTF) centre at the foot of the central protuberance through the subunit down to the base of the cytoplasmic side of it with a length of about 100 Å and a width of 10 to 20 Å (Ban *et al.*, 2000; Stark *et al.*, 1995). Additionally, the 50S subunit has a factor-binding centre and all of the G-protein factors involved in protein synthesis interact with it during at least part of their duty cycles. Both subunits are involved in translocation of the mRNA by one codon in each cycle. It was mentioned above, that 2/3 of each subunit consists of rRNA, which in fact plays the main role in protein biosynthesis.

Structurally, the 50S subunit is spherical with three almost cylindrical extensions (Figure 1.1-2). These extensions are called L1 protuberance, the central protuberance and the L7/L12 stalk. A striking difference between the two subunits has to do with the relationship between the secondary structures of their RNAs and their overall morphology. The six secondary structure domains of 23S rRNA are intricately interwoven in the 50S subunit to form a monolithic structure. On the other hand, the 30S subunit is divided into three domains (head, body, and platform). Each of these domains contains one of the principal secondary structure domains of 16S rRNA: The 5' major domain represents the body, the central domain the platform and the 3' major domain the head of the small subunit. The 3' minor domain of the 16S rRNA forms an extended helix (h44 in *E. coli* helix numbering) and runs down the long axis of the 30S subunit surface that interacts with the 50S subunit. All four domains of the 30S particle join at a narrow neck region. The two active sites (the decoding centre and the peptidyl transferase centre) face each other across the subunit interface and are functionally linked by both the two ends of A-site tRNA and a prominent inter-subunit bridge (bridge 2B).

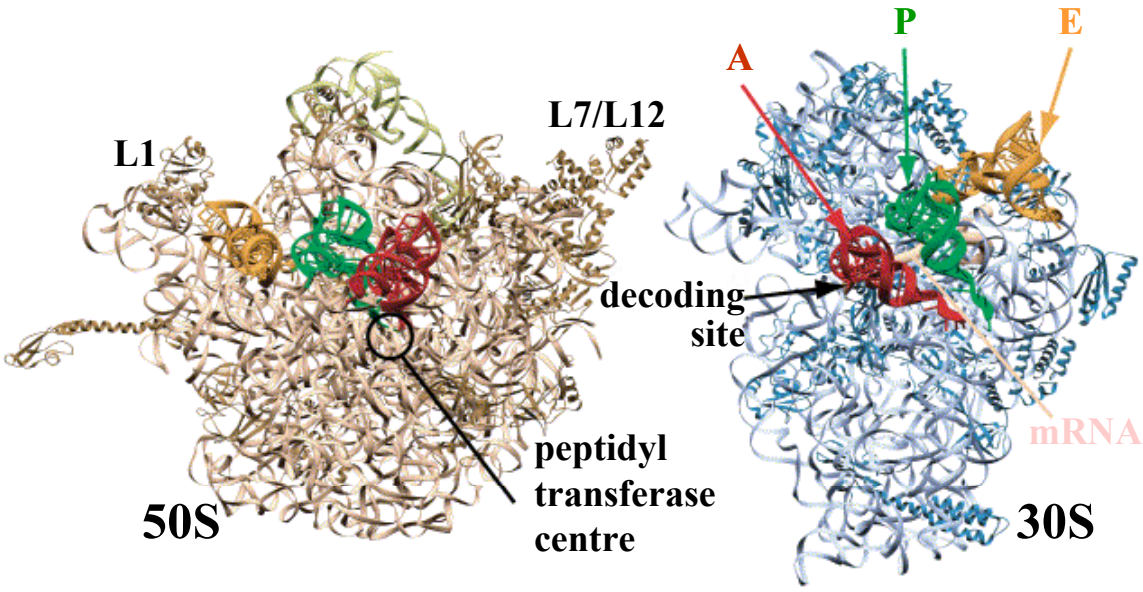


Figure 1.1-2 Crystal structure of the 50S and 30S subunits in the 70S ribosome. The location of the A-, P-, E-site tRNAs are indicated by arrows, respectively. The passage on the mRNA is indicated in pink. Essential landmarks of the 50S subunit are prescribed (Ramakrishnan, 2002).

The architectural difference between the subunits may reflect a greater functional need for flexibility on the part of the small subunit and, not necessarily alternative, a different evolutionary age of the subunits, where the large one is probably the older one (Sardesai *et al.*, 1999).

Cryo-EM and X-ray maps of the 70S ribosome show a number of highly conserved bridges connecting the subunits (Cate *et al.*, 1999; Frank *et al.*, 1995; Gabashvili *et al.*, 2000). The inter-subunit interface of both subunits, especially the part that binds mRNA and tRNAs, is largely free of protein.

Protein synthesis would not be able without participation of the adaptor molecules that ensure amino acids transfer to the ribosome – tRNAs. These molecules are on the border from RNA to protein world. Initially, two binding sites for tRNAs were proposed for the ribosome (Lipmann, 1963; Watson, 1963; Watson, 1964). The two sites of this model are the "A" site (for aminoacyl-tRNA or acceptor site) and the "P" site (for peptidyl-tRNA). However, functional studies at the beginning of 80's (Grajevskaja *et al.*, 1982; Lill *et al.*, 1984; Rheinberger and Nierhaus, 1980; Rheinberger *et al.*, 1981); have demonstrated a third tRNA binding site, the "E" site (E for the exit), from which

deacylated tRNA leaves the ribosome. Neutron scattering, cryo-electron microscopy and X-rays diffraction studies have confirmed the existence of the E-site on the ribosome (Agrawal *et al.*, 2000; Nierhaus *et al.*, 1998; Wadzack *et al.*, 1997; Yusupov *et al.*, 2001). This third ribosomal binding site has been found on ribosomes of all kingdoms and seems to be a universal feature of ribosomes (for review see Blaha and Nierhaus, 2001). However; disagreement exists on several points concerning the importance of the E site (for more details see Burkhardt *et al.*, 1998; Wilson and Nierhaus, 2003).

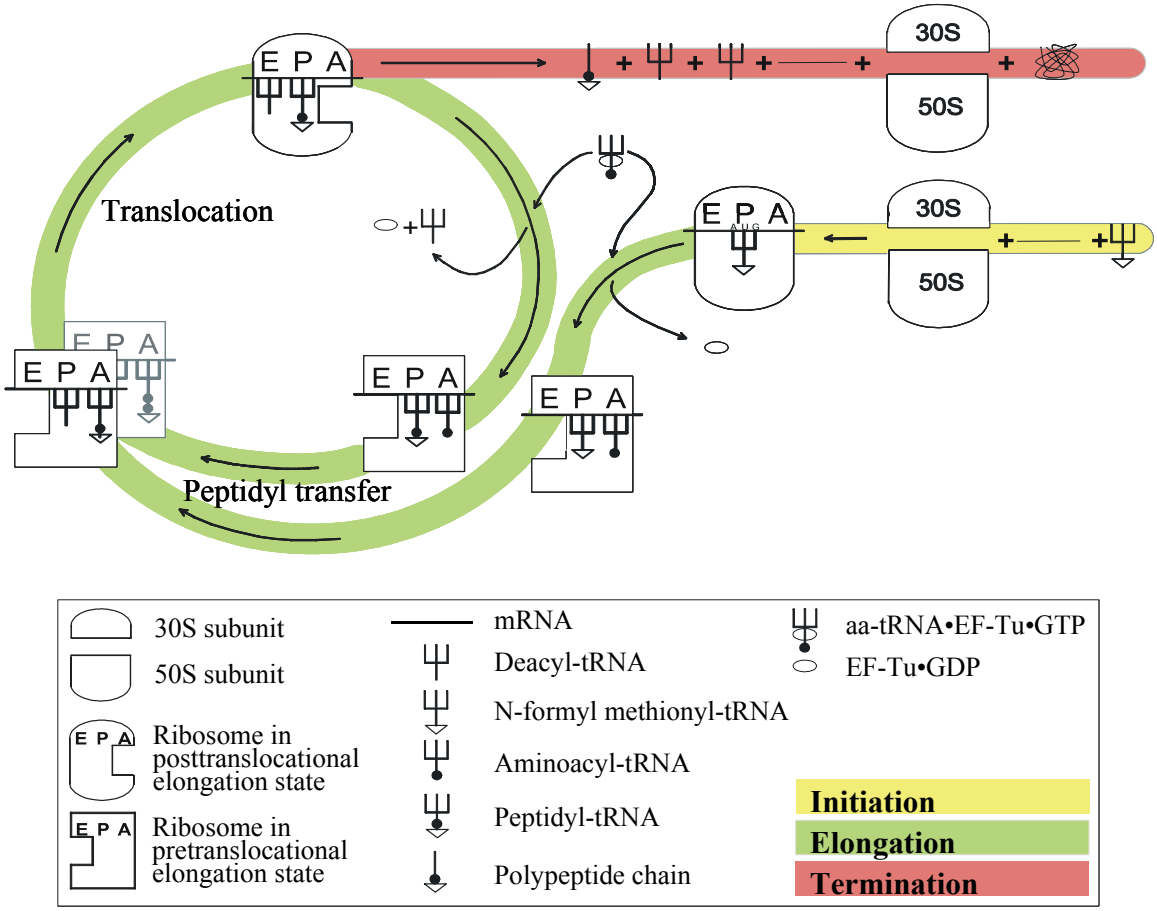


Figure 1.1-3 The functional phases of the ribosome during protein synthesis.

The whole process of protein synthesis is conventionally divided into three consecutive phases: initiation, elongation and termination (Figure 1.1-3). Briefly, during initiation mRNA is positioned with the AUG start-codon at the P-site on a small subunit, and large subunit associates. Within an elongation stage ribosome goes “codon-by-codon” along the mRNA synthesising the given

peptide until a stop-codon enters the A-site. This is a signal for translation termination followed by peptide release and subunits dissociation. Each stage is assisted by translational factors of initiation (IFs), elongation (EFs), and termination or release factors (RFs), respectively.

1.2 Initiation

The initiation phase of protein synthesis is one of the rate-limiting steps of translation. There are significant differences between translation-initiation events in eukaryotes, archaea and eubacteria; however, the final state of the ribosome following initiation is principally the same, namely, a ribosome programmed with an initiator tRNA and mRNA, such that the start codon and tRNA are both positioned at the P-site. Indeed, the production of functionally active proteins necessitates that translation initiates at the start codon within the mRNA. As well as the use of the correct codon as the start codon, the placement at the P-site of the ribosome must also be precise; since codons are composed of three bases, incorrect placement by one or two bases will result in a complete loss of the correct reading frame. There are two major contributors to ensure the fidelity of this process: (i) the mRNA itself and (ii) a subset of translation factors termed the initiation factors (IFs). Unlike eukaryotic protein synthesis the mechanism of translation initiation in prokaryotes is relatively simple and requires at least three factors for initiation, namely IF1, IF2 and IF3, which form an intermediate initiation complex consisting of an these factors, mRNA, initiator tRNA and the 30S subunit.

The exact role of the factors and the chronology of events surrounding initiation are still controversial. IF1 accelerates IF3-dependent 70S dissociation and stimulates IF2 and IF3 to form an initiation complex (Figure 1.2-1). Protection studies suggest that IF1 binds to the A-site, preventing binding of the elongator tRNA to the ribosomal A-site (or decoding site). The presence of IF1 increases the affinity of IF2 for the ribosome. IF2, as a binary complex with GTP, binds the 30S subunit and directs the initiator tRNA into the prospective

P-site on the 30S subunit. IF2 specifically recognises the blocked α-amino group of the initiator tRNA, thus excluding aminoacylated elongator tRNAs in the initiation step. IF3 participates in both the first and last steps of translational initiation, being involved in selection of the initiator tRNA as well as dissociation of the 70S ribosome into subunits. IF3 binds to the stem of the initiator tRNA and prevents the binding of the elongator tRNA to the P-site of the 30S initiation complex.

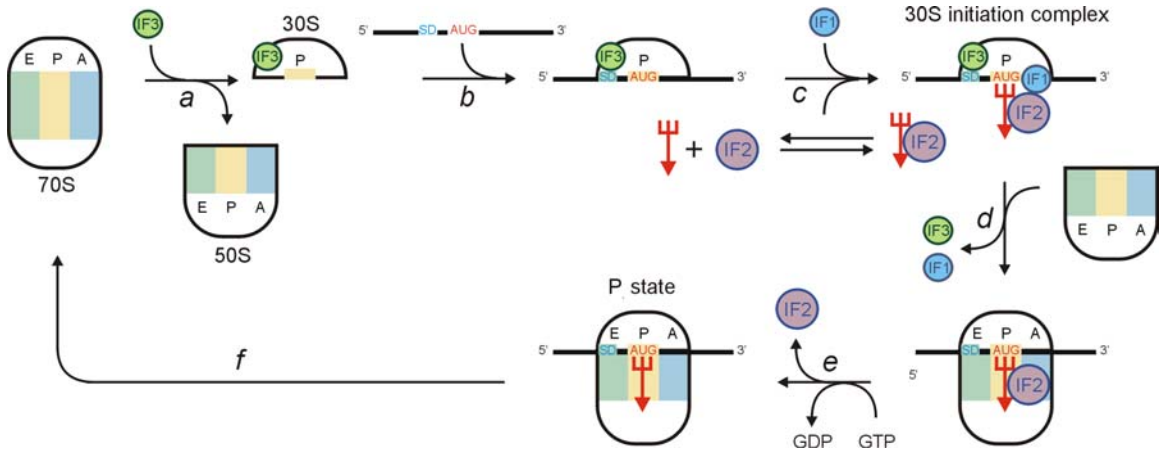


Figure 1.2-1 Schematic representation of the initiation of protein.

(**a**) Binding of IF3 to the 30S subunit dissociates empty 70S ribosomes into composing 30S and 50S subunits. (**b**) IF3 aids in positioning of the mRNA in a way that the AUG start codon is located at the P-site of the 30S subunit. The Shine–Dalgarno (SD) sequence of the mRNA is located in the vicinity of the binding position of IF3, and interacts with the anti-SD sequence of the 16S rRNA of the 30S subunit. (**c**) Binding of the initiator fMet-tRNA$_f^{Met}$ (red) can occur directly or in the form of a ternary complex with IF2 (purple) and GTP is stimulated by the presence of IF1 (blue) and results in the formation of the 30S initiation complex. (**d**) Association of the 50S subunit with 30S initiation complex results in the release of IF3 and IF1, but IF2 remains bound at the A-site. (**e**) The GTPase activity of IF2 is stimulated by 50S subunit and ultimately leads to the release of IF2·GDP from the ribosome, allowing full accommodation of the initiator tRNA at the P-site on the 50S subunit. This complex is termed the Pi state, i.e., P-site is occupied and the A- and E-sites are free. (**f**) Following the translation elongation, termination and ribosome recycling empty 70S ribosomes are ready to renter into translation-initiation phase.

The conventional mechanism of translation in bacteria includes complementary interactions between a purine rich sequence at the 5'-nontranslated region of mRNA (known as Shine-Dalgarno, SD), and the 3' end of 16S rRNA (anti-SD sequence) (Gold, 1988), facilitating the positioning of the correct AUG start codon at the P-site. There are data in the literature indicating

that the interaction between mRNA and the 30S ribosomal subunit is much more complex than thought before (Kozak, 1999; Mccarthy and Brimacombe, 1994; Sprengart and Porter, 1997). It has been found that besides the SD sequence other nucleotides located upstream of the initiation codon, are also involved in this interaction (Mccarthy and Brimacombe, 1994). The complex type of mRNA-ribosome interaction is illustrated as well by the existance of nucleotide sequences enhancing translation. Such sequences were found in several phage, viral and bacterial genes. The first enhancer of translation that was active in both prokaryotic and eukaryotic cells (Gallie and Kado, 1989) was identified at the 5'-nontranslated region of the tobacco mosaic virus (TMV) RNA. Enhancers of translation were also found in the genome of bacteriophages T7 (Olins *et al.*, 1988), λ (Wu and Janssen, 1996), Qβ (Ugarov *et al.*, 1994), and *Mycoplasma genitalium* (Loechel *et al.*, 1991). It is noteworthy that unlike the purine-rich SD sequence all bacterial enhancers of translation were poor in purines, and particularly guanine (G). Several studies show that besides their enhancing activity, some nucleotide sequences were also capable of independently (in the absence of SD sequence), initiating translation in *E. coli* cells (Ivanov *et al.*, 1992; Walz *et al.*, 1976).

1.2.1 70S initiation type

In bacteria most mRNAs result from the simultaneous transcription of a row of adjacent genes (operon) and therefore carry the information for the synthesis of several proteins, often functionally related. Such mRNAs are called polycistronic: they contain several translation initiation sites, one for each cistron. In *E. coli* polycistronic mRNAs contain four cistrons on average.

The recognition of the translation start sites on the mRNA is performed by an initiation complex including the small ribosomal subunit (30S), the initiator tRNA carrying the amino acid formylmethionine (fMet-tRNA$_f^{Met}$), and initiation factors. Theoretically, the various initiation codons of a bacterial polycistronic mRNA can be recognized independently of one another. Aided by the SD

sequences, the 30S initiation complexes can land on any of the available translation initiation sites (Figure 1.2-2).

In practice, and it is true for ribosomal proteins (r-proteins, it has been observed that in order to be accessible to the initiation complex an initiation region of mRNA (including the start codon and the SD motif) must be in a single-stranded, nonhydrogen-bonded state. Once the initiation site is recognized a translating ribosome may unfold the secondary structure of the mRNA.

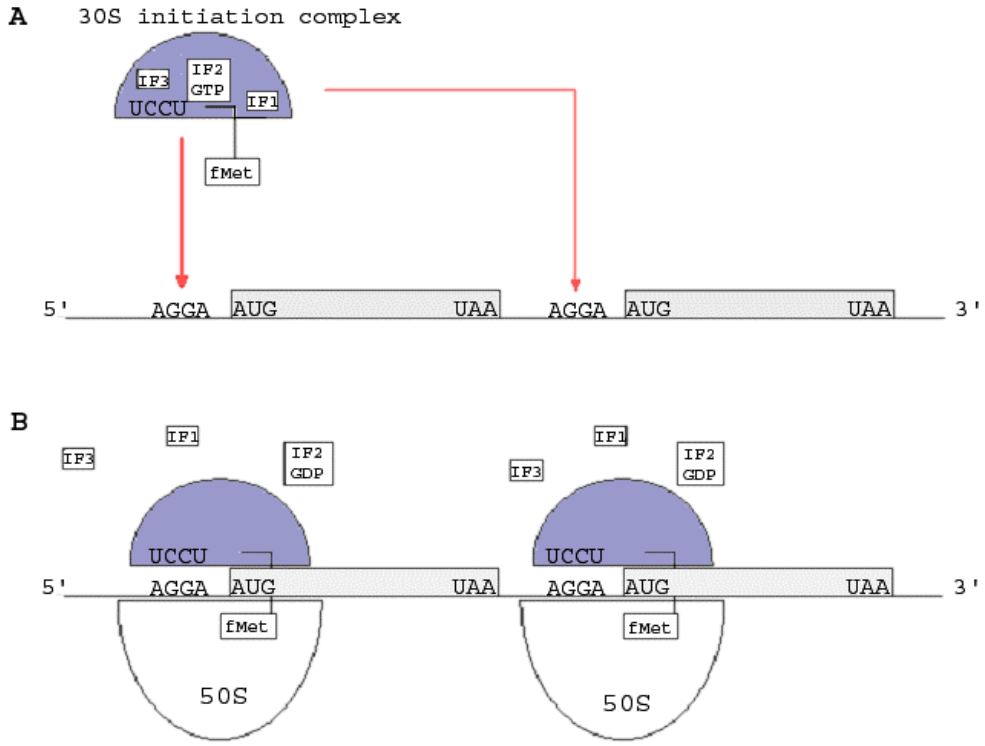


Figure 1.2-2 Translational initiation on a bacterial polycistronic mRNA. (**A**) The initiation complex, comprising the 30S ribosomal subunit, the indicated initiation factors (IF) and the initiator tRNA charged with the amino acid formylmethionine (fMet) recognizes two initiation codons on a bicistronic mRNA. The anti-Shine–Dalgarno sequence at the 3' end of the 16S ribosomal RNA (schematically indicated as UCCU) base-pairs with the Shine–Dalgarno sequence on the mRNA (AGGA), located a few nucleotides upstream of the initiation codons (AUG). The coding regions are illustrated as grey boxes, each terminating with a stop codon UUA. (**B**) Following recognition of the start codons by the initiation complex, the large ribosomal subunit (50S) joins the 30S subunit and the initiation factors dissociate from the ribosome.

This often creates a situation that renders the translation of the various cistrons in a polycistronic mRNA interdependent. In many cases, 'downstream' cistrons may not be expressed efficiently if those 'upstream' have not been

previously translated; for instance, when the initiation region of a downstream cistron is sequestered within a double-helical structure. Translational initiation at that site may take place only if a ribosome completing translation of the previous cistron unfolds the structured region thus unmasking the initiation codon. This phenomenon is termed *translational coupling*.

Sometimes, translational coupling is exploited to cause autogenous translational regulation; a repressor protein (usually a translation product of that same mRNA), will bind at a specific site on a polycistronic mRNA, thereby inhibiting the translation of most of the cistrons located downstream from the repressor-binding site. Presumably, the repressor hinders ribosome access to the initiation site of a critical upstream cistron whose translation is coupled to those downstream. Autogenous translational regulation is regularly observed with relatively stable bacterial mRNAs such as those encoding ribosomal proteins.

Often, the downstream cistron is translated by reinitiation, meaning that the ribosomes terminating translation of the upstream cistron do not dissociate from the mRNA but proceed directly to translate the next cistron, occasionally shifting the reading frame if this is required (Londei, 2001; http://www.els.net). What happens is that a 70S ribosome after termination on a stop codon of a preceding cistron and peptide release does not dissociate from such a polycistronic mRNA. An empty 70S (no peptide chain and tRNAs present) is capable of scanning in a one-dimensional manner upwards and downwards the mRNA up to 40 nucleotides, until it is "caught" by a nearby SD-sequence due to a base pairing of the last with a 3'-region of 16S rRNA anti-SD sequence. This helps to position a following AUG-start codon directly at the P-site. Petersen HU, Danchin A, Grunberg-Manago M., (Petersen *et al.*, 1976) had shown that the formylation of the methionyl initiator tRNA is only obligatory when polypeptide synthesis is initiated by non-dissociated 70S ribosomes. The site of IF2 binding on 70S ribosomes overlaps that of EF-Tu. Thus, the proper initiator tRNA, fMet-tRNA$_f^{Met}$, is delivered by the IF2 in the GTP form. GTP increases

the affinity of the IF2•fMet-tRNA$_f^{Met}$ complex to ribosome. This means that discrimination is on the factor and not on the ribosomal level.

1.3 Elongation

Elongation of the peptide chain (for reviews see Nierhaus *et al.*, 1998; Spahn and Nierhaus, 1998; Wilson and Noller, 1998) is the central event in protein synthesis. The elongation cycle directly involves two main protein elongation factors, EF-Tu and EF-G. EF-Tu forms a ternary complex with GTP and an aminoacyl-tRNA, directing the tRNA to the ribosomal A site. After peptidyl-transfer, EF-G is involved in translocating the mRNA•peptidyl-tRNA complex from the A site to the P site. This enables the next EF-Tu ternary complex access to the new codon present in the A site and thus the elongation cycle is repeated.

Three models for elongation have been proposed. The allosteric three-site model is based on functional experiments suggesting the A and E sites are allosterically linked – such that occupation of the E site by a deacylated tRNA (posttranslocation) creates a low affinity A site, and *vice versa* (Nierhaus, 1990). The hybrid-site model is based on structural experiments, which suggest that the tRNA movements during translocation occur in two steps, creating hybrid A/P and P/E sites (Moazed and Noller, 1989). The α-ε model is based on the observation that the ribosome contact patterns of the acylated tRNA before and after translocation in the A and P sites were similar, as were the patterns of the deacylated-tRNA in the P and E sites respectively (Nierhaus *et al.*, 1995). The main feature of this model is a ribosomal domain that is tightly bound to the tRNAs and moves with them during translocation.

The elongation cycle ends when translocation of a termination codon into the A site occurs. Simply, the termination of protein synthesis involves the hydrolytic release of the completed polypeptide from the peptidyl-tRNA in response to a stop codon appearing in the decoding site (A site) of the ribosome. It is this process which is mediated by the RFs.

1.4 Termination

The synthesis of the polypeptide chain continues until a stop codon (UAA, UAG or UGA), is invading the A-site. Protein factors, called release factors (RF), are in charge of the release of a nascent polypeptide chain from the ribosome and recycling of the ribosomes for the next initiation. Two classes of release factors are distinguished: *Class I* factors that do not consume energy and are specific decoding factors responsible for the hydrolysis of the peptidyl-tRNA. RF1 and RF2 belong to this group, and recognize UAG and UGA respectively; both RFs overlap in the recognition of the termination codon UAA. Class I RFs promote hydrolysis of the ester bond between the polypeptide and the P-site tRNA on the ribosome. RF1 and RF2 genes (*prfA* and *prfB*), have been shown to have a high similarity at the amino acid level (Caskey *et al.*, 1984; Craigen *et al.*, 1985; Weiss *et al.*, 1984). In eukaryotes and in archaea only a single class I factor, eRF1 (and aRF1) has been identified that recognizes all three-stop codons (reviewed by Wilson and Nierhaus, 2003).

The *Class II* release factors are non-decoding and energy consuming. The RF3 belongs to this class. It stimulates the termination process in a GTP dependent manner.

RF3 in *E. coli* is not essential; knockout strains of its *prfC* gene are viable. The main function of RF3 is to support removal of the class I RFs from the ribosome using GTP hydrolysis, once the peptide hydrolysis has taken place (Freistroffer *et al.*, 1997; Zavialov *et al.*, 2001). In other words, the post-termination complex is the substrate for RF3 that stimulates the recycling of RF1 and RF2 (Zavialov *et al.*, 2001).

Chapter 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals, enzymes and instruments suppliers

Amersham Biosciences, England

L-^{35}S-Methionine

L-^{14}C-phenylalanine

Spermidine

Spermine

FluorImager 595

Storm™ imaging system

Personal Densitometer™ SI

Storage Phosphor Screens

ImageQuant™ TL (software)

Beckman, Germany

Ready Value (scintillation liquid)

Ultracentrifuge tubes Ultra-Clean

BioRad, USA

Ammonium persulfate

SDS (sodium-dodecylsulphate)

Qiagen, Germany

Qiagen Maxi Prep Tip 500

Qiagen Midi Prep Tip 100

Qiagen Mini Prep

Qiaquick PCR Purification kit

Quiquick Nucleotide Removal kit

Difco BD, USA

Bacto agar

Bactotryptone, Yeast extract

EMD Biosciences, Germany

HEPES

Gibco-BRL, USA

Agarose (ultra-pure)

Urea (ultra-pure)

Sucrose (ultra-pure)

TEMED

Merck, Germany

All chemicals used in the laboratory and not extra listed here, and all essential amino acids

Bromphenol blue

Etidium bromide (1%)

Urea

Tris

Xylen cyanol

Microfluidics, USA

Microfluidizer® Processor M-110L

New England BioLabs, USA

Restriction endonucleases

Reaction buffers

T4 DNA ligase, 6 U/µl

Packard Instrument Company, USA

Filter Count (scintillation liquid)

Pharmacia, USA

BSA (DNase and RNase free, 2.6 mg/ml)

Long polyuridine (poly(U)

Sephadex

Roche, Germany

Adenosine-5'-triphosphate

Alkaline phosphatase (CIP, calf intestine phosphatase) 20 U/µl

Ampicillin

dNTPs

Guanosine-5'-triphosphate

Lysozyme

T4 Polynucleotide kinase

Poly-uridine (poly(U)

Phosphoenolpyruvic acid

Pyruvate kinase 10 mg/ml

RTS 100 *E. coli* HY Kit

RTS 500 *E. coli* HY Kit

RTS ProteoMaster Instrument

T4 DNA polymerase, 1U/µl

$tRNA^{bulk}$ (from *E. coli*)

Roth, Germany

Rotiphenol

Rotiphorese 30 (27,5:1)

Sartorius GmbH, Germany

Nitrocellulose filters (Nr. 11306)

Schleicher and Schuell, Germany

Selecta glass filter (Nr. 6)

Serva, Germany

Acrylamide

Alcoa A-305 Aluminium oxide

Bis-acrylamide

Sigma-Aldrich, USA

NTPs-Tris

$tRNA_f^{Met}$

Whatman Ltd., England

Paper Filters

2.1.2 Bacterial Strains and Plasmids

The new Rosetta™ (DE3) strain is derived from *lacZY* mutant of BL21 (*lon* and *ompT* proteases deletion), to enable precise control of expression levels by adjusting the concentration of IPTG. This strain is designed to alleviate codon bias when expressing eukaryotic proteins in *E. coli*.

Plasmid pET23c(+): The pET-23c(+) vector carries an N-terminal T7•Tag® sequence plus an optional C-terminal His•Tag® sequence. This vector has the "plain" T7 promoter instead of the T7*lac* promoter and the *lacI* gene is absent.

Plasmid pIVEX2.2GFPcyc3: this is an In Vitro EXpression vector that has a N-terminal *Strep*-tag prior to GFPcyc3 gene sequence. This plasmid was utilised as a general control for the study of *in vitro* expression system and for the comparison of the expression of some given proteins in this system. GFPcyc3 is a Green Fluorescence Protein that carries three point mutations, which allow fast maturation of the fluorophore in comparison to the wild type GFP.

2.2 Media

2.2.1 Luria-Bertani (LB) medium

Bacto-Tryptone/Peptone	10 g
Yeast Extract	5 g
NaCl	5 g
1M NaOH	1 ml
H_2O	ad 1 L
	pH 7.4

Media has to be sterilized via autoclaving. It is used for liquid overnight culture in order to obtain cells for isolation of given plasmid or for small scale

ribosome and S30 extract isolation. Cells are grown in the presence of suitable antibiotic.

2.2.2 L-Agar Plate

Agar added to LB	15 g/L

Media is sterilized *via* autoclaving, stirred well and distributed in 15-20 ml aliquots onto sterile Petry dishes while media is still liquid (when required, antibiotics were added just before the pouring to the plates).

2.2.3 Antibiotic Solution

Ampicillin	100 mg/ml

A measured amount of ampicillin powder was weighted and the amount of water was added necessary for the concentration of 100 mg/ml. After complete salvation the clear solution of ampicillin was filtrated through a sterile filter and small aliquots were prepared and stored at – 20°C.

2.3 Buffers

2.3.1 Acrylamide and staining solutions

Solution	Substance	Measure
30% AA/BAA (37.5:1) gel stock solution for protein gel	Acrylamide	300 g
	Bis-acrylamide	8 g
Agarose gel solution	10X TBE	5 ml
	Agarose	0.8-2% w/v
	Etidium Bromide (1%)	1.5 µl
	MQ-H_2O	ad 50 ml
APS solution 10%	Ammonium persulfate	10% w/v
Coomassie blue staining solution	Coomassie blue R-250	0.25% w/v
	Methanol	50% w/v
	Glacial Acetic acid	10% w/v
Distaining solution for Coomassie gels	Methanol	25% v/v
	Glacial acetic acid	8% v/v
	H_2O	67% v/v

Agarose gel sample buffer (5X) (for DNA)	EDTA	50 mM
	Ficoll 400	10%
	Bromphenol blue	0.25% w/v
	Xylencyanol	0.25% w/v
TBE (10X)	Tris (base)	108 g
	Boric acid	55 g
	EDTA	7.4 g
	MQ-H_2O	ad 1 L
Etidium bromide staining solution	Etidium bromide 1%	30 µl
	MQ-H_2O	300 ml
SDS-PAGE protein sample buffer (for protein)	SDS	2% w/v
	Tris-HCl (pH 6.8)	90 mM
	Glycine	10% w/v
	ß-mercaptoethanol	29 mM
	Bromphenol blue	0.1% w/v
SDS-PAAG separation buffer (pH 8.8); 1.5 M Tris	Tris (base)	180 g
	MQ-H_2O	ad 1 L
SDS-PAAG stacking buffer (pH 6.8); 0.5 M Tris	Tris (base)	60 g
	MQ-H_2O	ad 1 L
Tris-Glycine 5x running buffer; (Protein SDS-PAGE)	Tris (base)	30 g
	Glycin	144 g
	SDS	2 g
	MQ-H_2O	ad 1L
SDS-PAAG separating gel for proteins; (15% acrylamide)	1.5 M Tris-HCl, pH 8.8	2.5 ml
	30% Acrylamide, 37.5:1	5.0 ml
	10% SDS	0.1 ml
	10% APS	50 µl
	TEMED	10 µl
	MQ-H_2O	ad 10 ml
SDS-PAAG stacking gel for proteins; (5% acrylamide)	1.5 M Tris-HCl, pH 6.8	1.25 ml
	30% Acrylamide, 37.5:1	0.85 ml
	10% SDS	50 µl
	10% APS	15 µl
	TEMED	5 µl
	MQ-H_2O	ad 5 ml

2.3.2 Buffers for microbiological and molecular methods

Solution	Substance	Measure
P1 (Suspending agent)	Tris-HCl, pH 8.0	50 mM
	EDTA	10 mM
P2 (Cell-lysis)	NaOH	200 mM
	SDS	1% w/v
P3 (Neutralization buffer)	Potassium acetate, pH 5.5	3 M
QBT (Equilibration)	MOPS-KOH pH 7.0	50 mM
	NaCl	1000 mM
	Ethanol	15% v/v
QC (Washing buffer)	MOPS-KOH pH 7.0	50 mM
	NaCl	1250 mM
	Ethanol	15% v/v

QF (Elution buffer)	Tris-HCl, pH 8.5	50 mM
	NaCl	1250 mM
	Ethanol	15% v/v
TE buffer	Tris-HCl, pH 8.0	10 mM
	EDTA	1 mM
10 x Klenow buffer	Tris-HCl, pH 7.6	500 mM
	$MgCl_2$	100 mM
	DTT	10 mM
	BSA (DNase free)	500 µg/ml
10 x Ligation buffer	Tris-HCl, pH 7.6	660 mM
	$MgCl_2$	50 mM
	DTT	50 mM
	Spermidine	5 mM
	BSA (DNase free)	500 µg/ml
10X dephosphorylation buffer (for dephosphorylation of 5'phosphorylated ends of DNA or RNA)	Tris-HCl pH 7.6	500 mM
	$MgCl_2$	10 mM
	$ZnCl_2$	10 mM
	EDTA	5 mM
Glycerol storage solution	Tris-HCl, pH 8.0	25 mM
	$MgSO_4$	100 mM
	Glycerol	66% v/v

2.3.3 Buffers for the functional studies and ribosome preparation

Solution	Substance	Measure
Binding buffer ($H_{20}M_{4.5}N_{150}SH_4Spd_2Spm_{0.05}$)	Hepes-KOH, pH 7.5	20 mM
	MgAc	4.5 mM
	NH_4Ac	150 mM
	ß-mercaptoethanol	4 mM
	Spermidine	2 mM
	Spermine	0.05 mM
Tico-Buffer ($H_{20}M_6N_{30}SH_4$)	Hepes-KOH, pH 7.5	20 mM
	MgAc	6 mM
	NH_4Ac	30 mM
	ß-mercaptoethanol	4 mM
Mix I $H_{100}M_{21}N_{870}SH_{20}Spd_{12}Spm_{0,3}$ (For poly(U) dependent poly(Phe) synthesis assay)	Hepes-KOH pH 7,5	100 mM
	MgAc	21 mM
	NH_4Ac	870 mM
	ß-mercaptoethanol	20 mM
	Spermidine	12 mM
	Spermine	0,3 mM
Mix II/Charging Mix $H_{80}M_{15}N_{840}SH_{16}Spd_{12}Spm_{0.3}$ (For poly(U) dependent poly(Phe) synthesis assay and RF2 system)	Hepes-KOH pH 7.5	80 mM
	MgAc	15 mM
	NH_4Ac	840 mM
	ß-mercaptoethanol	16 mM
	Spermidine	12 mM
	Spermine	0.3 mM
Mix E $ATP_{45}GTP_{22.5}(AcPO_4)_{75}$	ATP	45 mM
	GTP	22.5 mM

energy mix/charging mix (for poly(U) dependent poly(Phe) synthesis)	Acetyl phosphate	75 mM
	KOH	360 mM
10X buffer for RNase assay (Binding Buffer 10X)	HEPES-KOH pH 7.5	200 mM
	MgAc	45 mM
	NH_4Ac	1500 mM
	ß-mercaptoethanol	40 mM
	Spermidine	20 mM
	Spermine	0.5 mM

2.3.4 Components of a batch in vitro *system for coupled transcription-translation*

Component	Final concentration	Components	Final concentration
HEPES-KOH (pH 8.2)	60 mM	GTP	1 mM
Ammonium acetate	80 mM	UTP	1 mM
Potassium glutamate	230 mM	PEG-8000	2% (w/v)
Sodium oxalate	3 mM	Methionine	2 mM
DTT	2 mM	Amino acids (19)	2 mM
Cycle-AMP	0.7 mM	PEP	35 mM
Folinic acid	35 µg/ml	Magnesium acetate	12 mM
tRNA	350 µg/ml	T7 RNA polymerase	100 µg/ml
NADH	0.35 mM	*E. coli* S30 extract	4-6 A_{260}
Coenzyme A	0.3 mM	Plasmid DNA	4 µg/60 µl
ATP	1.5 mM	Rifampicin	10 µg/ml
CTP	1 mM		

2.4 Analytic Methods

2.4.1 Photometric measurements

Spectrophotometric analysis of the amount of DNA or RNA

The concentrations of oligonucleotides, DNA and RNA were determined by absorption measurement in a Hitachi U-3000 spectrophotometer. The measurements were performed at 260 or 280 nm. The reading at 260 nm allows estimation of the concentration of nucleic acid in a sample, whereas the ratio between the reading at 260 nm and 280 nm (A_{260}/A_{280}) provides an estimate of the purity of the nucleic acid.

A_{260}/A_{280} = 1.8 for pure DNA*

A_{260}/A_{280} = 2.0 for pure RNA*

* These coefficients are valid for high molecular weight species (Berger, 1987). In the case of oligonucleotides (less than 100 bases), the base composition can have a significant effect in the extinction coefficient. This means that the A_{260}/A_{280} ratio will change with the sequence. Extreme cases are A rich sequences (A_{260}/A_{280} ratio larger than 2.2), and C rich sequences (A_{260}/A_{280} ratio smaller than 1.5).

Conversion factors used for DNA and RNA quantification are:

1 bp in double stranded DNA	660 g/mol
1 A_{260} unit of double stranded DNA	50 µg
1 A_{260} unit single stranded DNA or RNA (>100 bases)	40 µg
1 A_{260} unit of single stranded DNA (less then 25 bases)	20 µg
1 A_{260} unit of single stranded DNA (30-80 bases)	30 µg

Determination of ribosome and nucleic acid concentration

The concentrations of 70S ribosomes and 30S and 50S subunits and tRNAs were determined by photometry at 260 nm, using the following relationships:

70S	1 A_{260}/ ml = 24 pmol/ml
50S	1 A_{260} /ml = 36 pmol/ml
30S	1 A_{260} /ml = 72 pmol/ml
tRNA	1 A_{260} /ml = 1500 pmol/ml

Fluorometric analysis of GFP

Native PAGE with GFP in a native buffer was performed for 2.5 hours in a 15% protein polyacrylamide gel. Fluorescence was measured directly from the gel on a FluorImager™ 595 dual-excitation, laser-induced fluorescence scanner. The images were analysed using the ImageQuant program where the relative intensities of the GFP band from expression system were compared with that one of the GFP with known concentration (marker).

2.4.2 Radioactivity Measurements

The radioactivity measurements were performed using a liquid scintillation counter Wallac Rack Beta model 1209 or a Rack Beta model 1409. The treatment of the samples before measurements was done according to their physical nature and contents of an isotope.

Liquid samples

Radioactive liquid samples with a volume of 1 ml or less were put to the scintillation vials (plastic, 20 ml maximal volume), then 5 ml of Ready Value (Beckman) or Filter Count (Packard Bioscience), were applied. After quick and vigorous shaking (10 sec), the samples were immediately counted.

Quantification of amounts of [^{35}S]-Met labelled proteins in SDS-PAAG

Dried SDS protein gels were stored in Storage Phosphor Screens for 24-48 hours. Storage phosphor screens retain energy from beta particles, X-rays, and gamma rays. The lower limit of detection for a 1 h exposure is less than 2 dpm/mm^2 for ^{14}C (200 and 100 µ only). The lower limit of detection for ^{32}P is typically 5–10 times lower than the limit for ^{14}C. Storage phosphor screens are used in conjunction with a storage phosphor system such as PhosphorImager™ or Storm™.

The electrical signal is digitised to permit image display and analysis. Scan control software and ImageQuant Image Analysis Software are available for Windows and are included with purchase of the Storm system.

Cold trichloroacetic acid (TCA) precipitation assays

From each sample (normally two per assay) a 10 µl aliquot was delivered into a glass test-tube (8 x 10 cm) containing 20 µl of precipitation carrier solution (1% BSA). Two ml of ice cold TCA (10% w/v) were immediately added and mixed for 1-2 seconds. The polypeptide chains were precipitated at

90°C water bath for 15 min, cooled down on ice and the precipitation mix was then filtered through glass fibre filters. The filters were washed three times with 2 ml of cold TCA (10%) and once with 2 ml diethyl ether/ethanol (1:1 v/v). The radioactivity adsorbed on the filter was measured after application of 5 ml of ReadyValue scintillation solution and vigorous shacking.

2.4.3 Agarose gel electrophoresis of DNA and RNA

This technique was used for analysis of plasmid DNA after analytical and preparative isolations from cells, as well as for restriction reaction and polymerase chain reaction (PCR), product analysis, and for the qualitative evaluation of rRNAs from 70S ribosomes, 50S and 30S ribosomal subunits.

The samples were prepared according to the type and the expected size of the nucleic acids to be separated. Plasmid DNA samples were prepared in a total volume of 6-10 µl in 5X agarose gel sample buffer including SDS (2% final concentration), in cases where the preparation was not highly purified (e.g. miniprep DNA). The amount of DNA was 0.2-1 µg per lane if intact plasmid preparations were analysed, or 1.5-2 µg if fragments of 600 bp or less were expected. For rRNA analysis, 0.05 A_{260} units of the corresponding ribosomal particles were incubated 2 min at 65°C in 5X agarose gel sample buffer containing SDS (2% final concentration), and kept 10 min at room temperature before loading onto the gel.

The percentage of agarose used depended on the expected size of the nucleic acid to be separated: 0.8% for rRNA and 3 kb plasmids, and 1.5-2% for restriction fragments of 600 bp or less. The buffer system used were TBE (for minipreps and restriction analysis).

The agarose solution was heated in a microwave oven for not more than one minute. The solution was cooled to room temperature, 1.5 µl of etidium bromide was added and the mixture was poured into the gel chambers:

a) Mini-gels: 9 x 7 x 0.7 cm (total volume = 50 ml) with 14 or 28 wells (10 µl maximal sample capacity). Used mainly in routine checking of plasmid isolation, minipreps and restriction analysis. Run at 50-60 V.

b) Regular gels: 14 x 8 x 0.7 cm (total gel volume = 80 ml) with 14 or 28 sample wells (20 µl maximal sample capacity). Used in the analysis of small DNA fragments and rRNA. Run at 60-100 V.

The length of the nucleic acid molecules was estimated according to the migration of standards of known molecular weights.

2.5 Preparative methods

2.5.1 S30-extract preparation from E. coli

A large amount of fresh or frozen *E. coli* cells collected on a *lag* stage of growth was suspended in two volumes of Tico buffer. Suspension was centrifuged in a pre-weighted chilled GSA tubes for 15 min with 10,000g at 4°C. All the following steps were carried out in a cold room. The pellets (cell debris) then were weighted and dissolved in Tico buffer in proportion as 1 ml per g. Such a suspension should easily go through the tip of 10 ml glass pipette. When this condition was reached, everything was applied to the Microfluidizer apparatus, which was washed and equilibrated with Tico buffer prior. Within this apparatus cells are pushed under high pressure (18 atm), through a tiny channel; which leads to their breakage. Then, collected suspension was centrifuged and resulting supernatant was transferred to the clean tubes and centrifuged again. These two steps of centrifugation were carried out in HB4 rotor with 16,000 g for 45 min at 4°C. After final centrifugation from the clear supernatant fraction fresh aliquots were prepared, shock-frozen and stored at –80°C. A small amount enough for double determination was kept in order to measure A_{260} of S30-exstract.

2.5.2 Preparation of 70S ribosomes

In a typical preparation ~300 grams of frozen *E. coli* cells were thawed via suspending in ~600 ml (double amount), of Tico buffer, and recovered by centrifugation at 8,000 rpm for 15 minutes in a Sorvall GSA rotor at 4°C. The cell pellet was weighted, and a double amount of aluminium oxide (Alcoa-305), was added. This mixture was transferred to a cooled mortar, and the cells were ground for about 40 minutes. After addition of Tico buffer (1.5 ml per gram of cell), the cells were further mixed for 10 minutes. The homogenate was then twice centrifuged at 8,000 rpm for 10 minutes in a GSA rotor at 4°C in order to remove the Alcoa and the unbroken cells. The supernatant was centrifuged at 16,000 rpm (30,000 x g), for 45 minutes in a SA-600 rotor. The pellet was discarded and the supernatant (S-30), containing ribosomes and soluble enzymes was further centrifuged at 22,000 rpm (30,000 x g), during 17-20 hours in a 45 Ti rotor in order to sediment the 70S ribosomes. The pellet was suspended in Tico buffer and again centrifuged in a SA-600 rotor at 8,000 rpm during 10 minutes in order to eliminate the non-dissolved aggregates. The ribosomes in suspension (crude 70S) were then shock-frozen in liquid nitrogen in aliquots containing 6,000-9,000 A_{260} units and stored at – 80°C.

The yield of crude 70S ranged between 300 and 400 A_{260} units per g of wet cells processed.

2.5.3 Extraction of small RNAs from the cell (from the Ph. D. thesis of Ralf Jünemann)

A 30 ml overnight cell culture was centrifuged for 5 min with 5,000 rpm in HB4 rotor at 4°C. The cell pellet was washed with 10 ml Cell Wash buffer and centrifuged again. The resulting pellet was suspended in 1.5 ml of Cell Wash buffer and transferred to 2 ml pre-weighted tubes. Another centrifugation step carried out in a table centrifuge for 5 min with 5,000 rpm at 4°C to estimate the weight of cell pellets, to which 1 ml of $H_{20}M_{10}$ buffer per 0.25 g cell pellet was applied. When suspended well, one volume of 70% phenol (RNase free) was

added to each sample tube and vigorously shacking during 45 min at 4°C. Under such a treatment RNA goes into water phase while DNA and proteins go to the organic phase. It is important to vortex solution strongly and continuously; otherwise the output of small RNAs will be very poor. After the next centrifugation at 14,000 rpm at 4°C for 30 min in a table centrifuge water phase was removed to fresh tubes and the remaining phenol was extracted by adding one volume of chloroform - isoamyl alcohol (24:1) followed another centrifugation step. To the water phase from the last step 0.65 volumes of 5M NaCl were added, and samples were kept on ice for 30 min. During this incubation large RNAs like ribosomal precipitate, and small RNAs like tRNAs are still in solution. After a final centrifugation in a table centrifuge at 4°C with 14,000 rpm for 30 min, the supernatant containing tRNAs was kept, while the pellet was discarded. The small RNAs from the supernatant were precipitated by standard ethanol precipitation technique. The resulting pellets of RNAs were dissolved in water and small aliquots were shock-frozen and stored at –80°C. One small volume is kept for A_{260} measurements.

2.6 Genetic methods

2.6.1 Preparation of E. coli *competent cells for electroporation*

The *Escherichia coli* strain XL1-Blue strain allows blue-white colour screening for recombinant plasmids and is an excellent host strain for routine cloning applications using plasmid or lambda vectors. The XL1-Blue genotype is as follows: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*q*Z.M15* Tn*10* (Tetr)]. (Genes listed signify mutant alleles. Genes on the F′ episome, however, are wild type unless indicated otherwise). This cell strain is tetracycline resistant and was used for preparation of competent cells.

One-litre culture of the *E. coli* strain of interest was grown to an OD_{600} of 0.5-1.0. The cells were chilled on ice, and then centrifuged, at 4,000 rpm for 15 min at 4°C in a GSA rotor to obtain pellets, which were suspended in 1 volume of pre-chilled sterile MQ-H_2O. Bacterial cells were repeatedly centrifuged and

washed in decreasing volumes of cold MQ-H_2O: two washes with 0.5 volume, one wash with 0.02 volume and final resuspension in 2-3 ml filter sterilised 10% (v/v) glycerol. Finally, from the cell suspension in 40 µl aliquots were prepared, shock frozen in a dry ice-ethanol bath and stored at –80°C.

The competence of the cells was defined by transformation using the plasmid pIVEX2.2GFPcyc3; using 10 ng of plasmid plus 40 µl of competent cells (processed according to the standard protocol using LB/ampicillin plates). The transformation efficiency was determined to be in the range of 0.5-1.5 x 10^6 transformants per µg of plasmid DNA. This level of competence was considered to be sufficient for further experiments.

2.6.2. Cloning strategies

T7 RNA polymerase expression system

In the following section some rudimentary description of the T7 polymerase expression system is given. For a more detailed understanding refer to (Studier *et al.*, 1990). A vast number of diverse vectors and strains are available for use with the pET expression system, and can be accessed through the Novagen web site:

http://www.merckbiosciences.co.uk/g.asp?f=NVG/home.html

The bacteriophage T7 RNA polymerase

The T7 RNA polymerase expression system relies on the fact that the RNA polymerase of bacteriophage T7 has a very high specificity for its own promoter. Furthermore, there are no known *E. coli* sequences that resemble this region. The bacteriophage T7 RNA polymerase initiates rapidly and transcribes about five times faster than an *E. coli* RNA polymerase. Efficient transcription termination signals for T7 RNA polymerase are very rare.

The T7 RNA polymerase can be delivered to a cell by either induction or infection. For convenience the former was preferred (as seen in Figure 2.6.1-1).

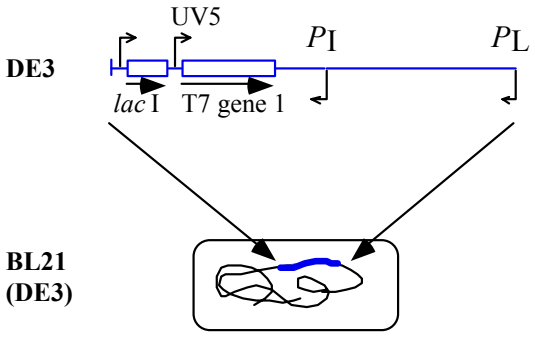


Figure 2.6.1-1 Formation of lysogenic BL21(DE3). The *E. coli* expression strain BL21(DE3) contains the *gene 1* encoding T7 RNA polymerase under control of an IPTG-inducible *lac*UV5 promoter.

pET vectors and expression

The plasmid vectors for expression using the T7 RNA polymerase are termed pET vectors (plasmid for expression by T7 RNA polymerase). This study utilizes pET23c(+). In pET23c(+), genes of interest are cloned directionally between an *Nde* I or *Xba* I and a *Xho* I site, placing the gene under the control of the T7 promoter.

It is important to note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the vector map (Figure 2.6.1-2). The f1 origin is oriented so that infection with helper phage will produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer.

pET23c(+) sequence landmarks

Name	Position on the plasmid
T7 promoter	303-319
T7 transcription start	302
T7•Tag coding sequence	207-239
Multiple cloning sites (*BamH* I – *Xho* I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
pBR322 origin	1450
bla coding sequence	2211-3068
f1 origin	3200-3655

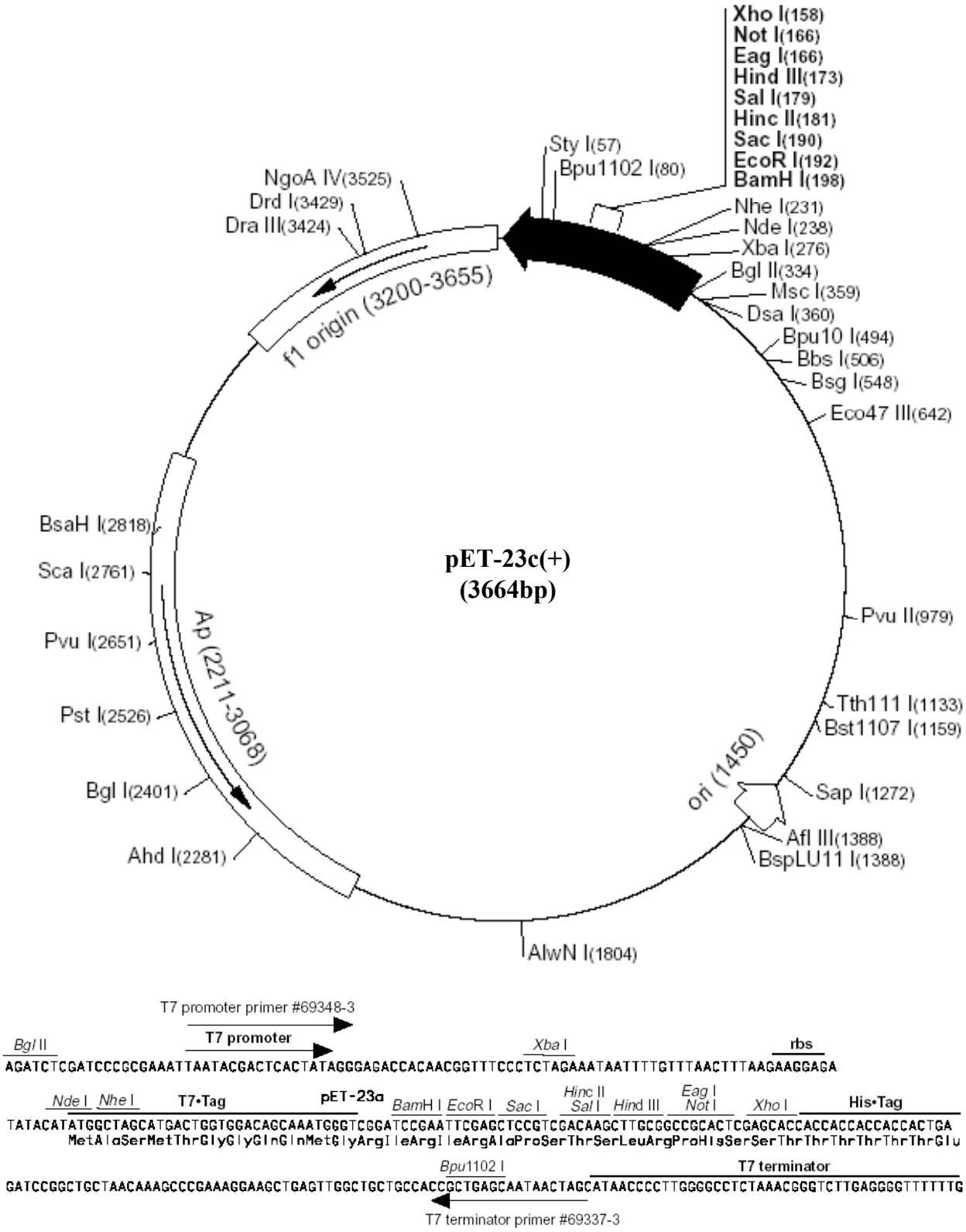
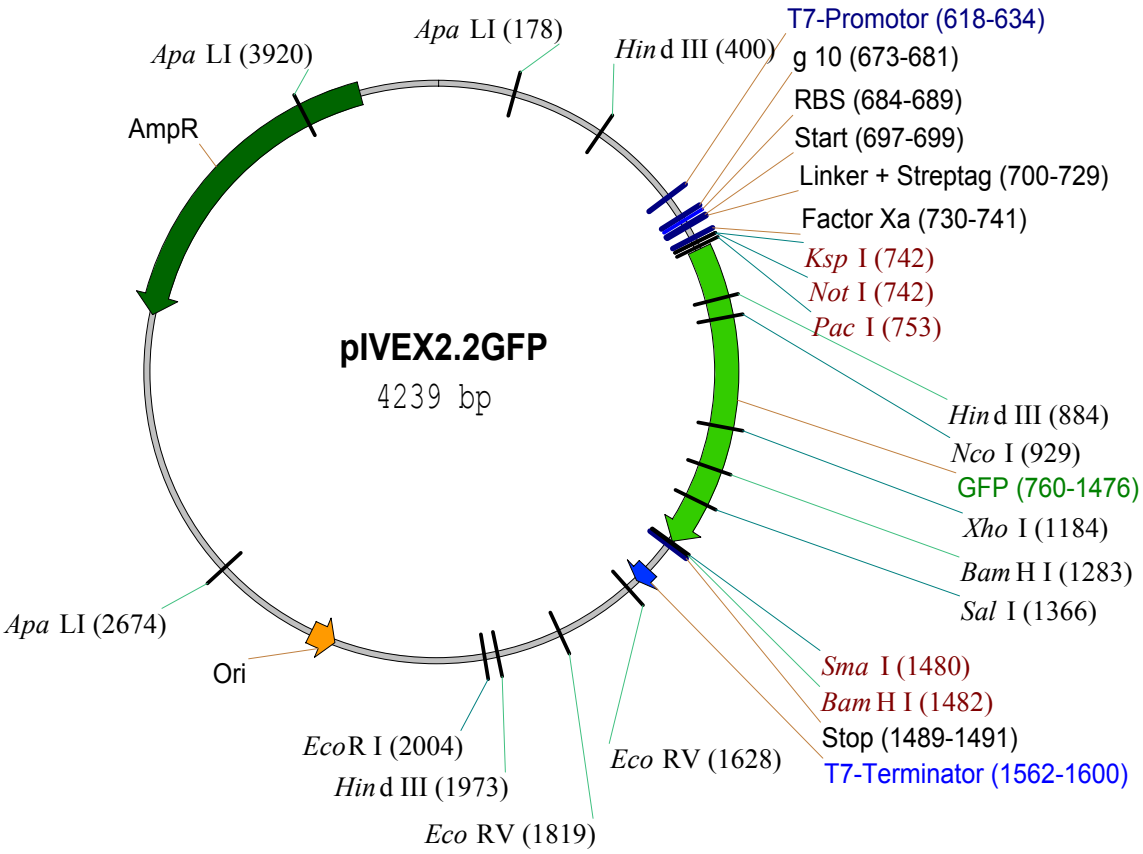


Figure 2.6.1-2 Cloning/expression region of the DNA plasmid pET-23c(+).

pIVEX vector and expression

Plasmid pIVEX2.2GFPcyc3 was utilised as a general control for the study of *in vitro* expression system and for the comparison of the expression of some given proteins in this system. GFPcyc3 is a Green Fluorescence Protein that carries three point mutations, which allow fast maturation of the fluorophore in comparison to the wild type GFP. The gene is introduced on the plasmid directly between *Pac* I and *Sma* I restriction sites.

The plasmid contains (i) a gene providing ampicillin resistance; (ii) the strong T7 promoter sequence (Studier and Moffatt, 1986); (iii) the T7 *gene* 10 translational enhancer (Olins *et al.*, 1988); (iv) the prokaryotic Shine-Dalgarno sequence (Shine and Dalgarno, 1974) as a binding site for prokaryotic ribosomes with an optimum distance to the start AUG codon (Chen *et al.*, 1994); (v) the T7 terminator that stops transcription and prevents 3'-terminal exonucleolytic degradation of the mRNA; (vi) *Strep*-tag® II (Voss and Skerra, 1997) at the N-terminal to allow purification of the expressed protein; (vii) restriction protease factor Xa cleavage site (see Figure 2.6.1-3 below).




```
                          T7-Promotor
601   TCTCGATCCC GCGAAATTAA TACGACTCAC TATAGGGAGA CCACAACGGT TTCCCTCTAG
      AGAGCTAGGG CGCTTTAATT ATGCTGAGTG ATATCCCTCT GGTGTTGCCA AAGGGAGATC

                     g10 ε       SD-Sequ.    NcoI   Linker Streptag
661   AAATAATTTT GTTTAACTTT AAGAAGGAGA TATACCATGA CCAGCTGGAG CCACCCGCAG
      TTTATTAAAA CAAATTGAAA TTCTTCCTCT ATATGGTACT GGTCGACCTC GGTGGGCGTC
                                            MetT hrSerTrpSe rHisProGln

                              KspINotI     PacI          NcoI       SalI   XhoI
SacI
721   TTCGAAAAAA TCGAAGGCCG CGGCCGCTTA ATTAAAACCA TGGCAGTCGA CTCGAGCGAG
      AAGCTTTTTT AGCTTCCGGC GCCGGCGAAT TAATTTTGGT ACCGTCAGCT GAGCTCGCTC
      PheGluLysI leGluGlyAr gGlyArgLeu IleLysThrM etAlaValAs pSerSerGlu

        PstI  SmaI BamHI
781   CTCTGCAGCC CGGGATCCGGC TGCTAACAA AGCCCGAAAG GAAGCTGAGT TGGCTGCTGC
      GAGACGTCGG GCCCTAGGCCG ACGATTGTT TCGGGCTTTC CTTCGACTCA ACCGACGACG
      LeuCysSerP roGlySerGly Cys***
```

Figure 2.6.1-3 Map of the pIVEX2.2GFPcyc3 expression vector. On the map GFPcyc3 insert is shown in green with markings of its position on the plasmid in brackets. Below the linker of the pIVEX2.2 vector is represented with corresponding markings; sites used for insertion of GFPcyc3 are underlined.

2.6.3 Vector construction and protein overexpression

Plasmid constructs

The pET23c(+) expression system was chosen to compare the efficiency of *E. coli* translation elongation factors (EFs) expression in the coupled transcription-translation system. The elongation factor EF-Tu (for temperature instability), is usually fragmented in the *in vitro* systems, unlike EF-Ts (temperature stable). Table 2.6.3-1 summarises the plasmids used in this study.

Table 2.6.3-1 Plasmid constructs

Plasmid	Description	Source
pET23c(+)	T7 pol promoter	Novagen
pET23c(+)_EF-Tu	EF-Tu gene in pET23c(+)	This work
pET23c(+)_EF-Ts	EF-Ts gene in pET23c(+)	This work
pIVEX2.2GFPcyc3	GFPcyc3 gene in pIVEX2.2	///////
pIVEX2.3GFPwt	GFPwt gene in pIVEX2.3	///////

Polymerase chain reaction

Plasmid DNA containing an EF-Tu or EF-Ts gene was used as a template for PCR with the appropriate sets of forward and reverse primer pairs (as seen in Table 2.6.3-2).

Table 2.6.3-2 PCR primers

Forward Primer	Sequence (5'-3')*
Tu_NdeI	CGTCACTTTAAGAAGAGATATA**CATATG**TCTAAAGAAA AGTTTGACC
Ts_NdeI	CGTCACTTTAAGAAGAGATATA**CATATG**GCTGAAATTA CCGC
Reverse Primer	**Sequence (5'-3')***
EF-Tu-XhoI	GT**CTCGAG**GCTCAGAACTTTTGCTACAA
EF-Ts-XhoI	GT**CTCGAG**AGACTGCTTGGACATCG

*The restriction sites are in bold and underlined type

PCR conditions

The optimal reaction conditions (incubation times and temperatures, concentration of the enzyme, template DNA, Mg^{2+} ions) depend on the template/primer pair and must be determined individually.

To increase fidelity of the PCR reaction use a thermo stable DNA-polymerase with 3'-5'-exonuclease activity like Pwo or the Expand™ High Fidelity PCR System.

For a standard PCR reaction in a total volume of 50 µl use 20 pmoles of each forward and reverse primers together with 100 ng of the template DNA, 400 pmoles of each dNTP, 2.6 units of Expand™ enzyme and 5 µl 10x Expand HF buffer with 15 mM Mg^{2+}.

To avoid unspecific products, reactions were cycled 25 times with an extension time of 90 sec at 72°C. Conveniently an annealing temperature of 50°C was employed successfully for each set of appropriate primer pairs listed in Table 2.6.3-2. Homogeneous PCR products of the expected size were analysed by 1% agarose gel electrophoresis (see Figure 2.6.3-1).

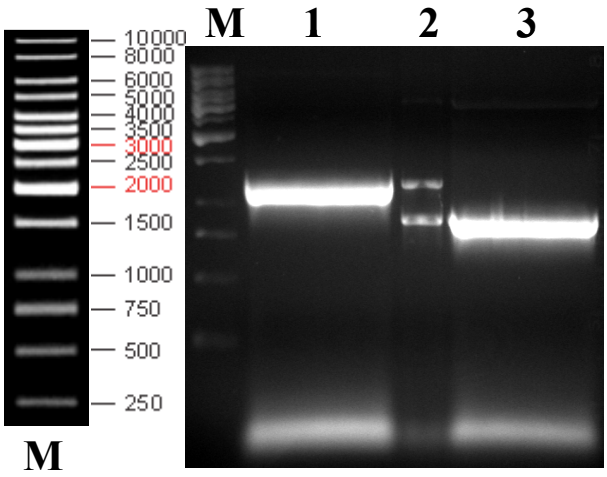


Figure 2.6.3-1 Gel electrophoresis of PCR products. PCR reaction products were analysed by electrophoresis through a 1.0 % agarose gel. Lane M contains GeneRuler™ 1kb DNA Ladder (bp). PCR products were prepared to produce approximately 1185 bp products of *Nde* I-EF-Tu- *Xho* I (Lane 1), and 852 bp *Nde* I-EF-Ts- *Xho* I (Lane 3). Lane 2 contains both EFs.

Cloning of PCR products

PCR products were digested with *Nde* I/*Xho* I and gel purified prior to ligation into *Nde* I/*Xho* I-digested and purified vector pET23c(+). Ligations were transformed into XL1-Blue cells and clones were analysed by restriction digestion for insert containing vectors (Figure 2.6.3-2).

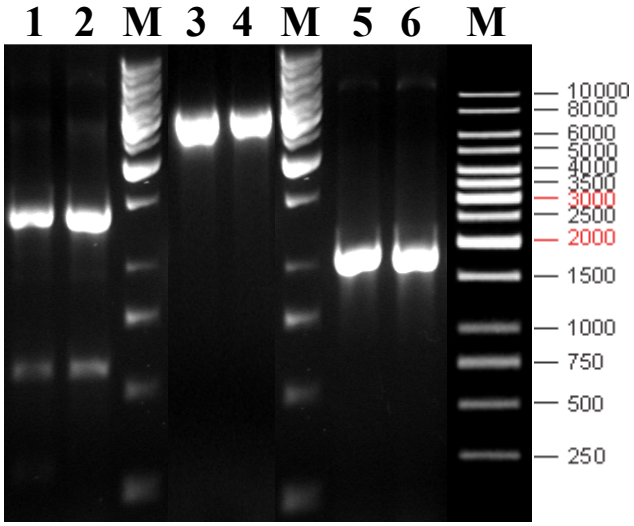


Figure 2.6.3-2 Restriction analysis of clones. Lanes 1 and 2 show *Nde* I/*Xho* I-digested EF-Tu (1.2 kb), lanes 5 and 6 show *Nde* I/*Xho* I-digested EF-Ts (0.8 kb). Pre-ligation pET23c(+) vector (3.6 kb) is shown in lanes 3 and 4. All restriction products were analysed on 1.0 % agarose gel. Lane M contains GeneRuler™ 1kb DNA Ladder; relevant marker sizes are indicated in bp.

Positive clones were sequenced in both directions to verify the absence of polymerase-introduced mutations and used in experiments discussed below.

2.7 *In vitro* protein synthesis systems based on *E. coli* lysate

Transcription and translation take place simultaneously in a reaction mix. Expression of genes behind a T7 promoter is possible from either circular or linear DNA template. Upon addition of the DNA template and T7 RNA polymerase to the DNA-free *E. coli* lysate transcription and translation are closely coupled in time and space: While the T7 RNA polymerase transcribes the template gene, the ribosomes provided by the *E. coli* lysate start to translate the 5'-end of the nascent mRNA.

The *E. coli* lysate is prepared with some modifications according to the method of Zubay (Zubay, 1973). An *E. coli* strain with the lowest protease activity was selected and growth conditions were optimised to allow optimum protein expression from linear (PCR-generated) and plasmid templates.

2.7.1 Batch system

Based on the Kim and Swartz system (Kim and Swartz, 1999; Kim and Swartz, 2000) the substances discussed in section 2.3.4 were used to prepare HEPES salts (HSS) and Small compounds (SCS) stock solutions. Along with substances displayed in Table 2.7.1-1 below solutions were pipetted together into LM-mix in corresponding volumes. The LM-mix is stabled for two to three weeks and can be prepared for multiple reactions assay in advance, shock frozen and stored at – 80°C. This LM-mix is pipetted into the Master-mix alone with essential salts and enzymes (see Figure 2.7.1-1, Master-mix preparation).

Table 2.7.1-1 Pipetting scheme for the LM-mix preparation

Substance	Volume
HSS	400 µl
SCS	80 µl
ATP, 100 mM	50 µl
CTP, 100 mM	40 µl
GTP, 100 mM	40 µl
UTP, 100 mM	40 µl
H_2O	70 µl
aa C, 55 mM	40 µl
aa DY, 55 mM	40 µl
PEP, 670 mM	200 µl

Final volume of 1000 µl was divided into 150µl aliquots and stored at – 80°C for up to three weeks.

The Master-mix contains among the rest [^{35}S]-Met and T7 RNA polymerase.

Master Mix Preparation

Sample	1:10 in H_2O [^{35}S]-Met	H_2O	100 mM Mg acet.	10 mM, ph 7.6 HEPES-KOH	LM-mix	T7 RNAP	240 µg/ml Rifampicin
Volume	**1µl**	4 µl	6 µl	4 µl	12.5 µl	1 µl	**1.5 µl**
1. DNA - 2. DNA - 3. DNA + 4. DNA +							
Volume x 5	5 µl	20 µl	30 µl	20 µl	62.5 µl	5 µl	7.5 µl

Preparation of the Reaction Mix

[^{35}S]-Met in H_2O	
[^{35}S]-Met	8.0 µl
H_2O	72.0 µl
	80.0 µl

Component	40% PEG	Master mix	DNA template	S30 extract
Volume	2.5 µl	30 µl	5 µl	12.5 µl
	VORTEX !!!			

Figure 2.7.1-1 Preparation of Master- and Reaction-mixes. The [^{35}S]-Met was diluted accordingly.

2.7.2 RTS 100 High Yield E. coli *Kit*

Here we scaled down the Roche batch system to 10 µl volume, where conditions could be easily altered for optimization; Roche is suggesting a volume of 50 µl.

According to the protocol we reconstituted the contents of the kit with a supplied Reconstitution buffer (Table 2.7.2-1). The reaction mixes were prepared according to the Table 2.7.2-2 and 10 µl were distributed to separate vials, DNA template was added after this. For radioactive labelling of elongation factors EF-Tu and EF-Ts the reactions were prepared according to the Table 2.7.2-3. Standard incubation temperature was 30°C. Samples were introduced into ProteoMaster instrument and incubated according to the assay requirements.

Table 2.7.2-1 Reconstitution of reaction components (RTS 100 HY E. coli *Kit)*

Solution	Reconstitution procedure
1. *E. coli* Lysate	Reconstitute the lyophilizate with 0.36 ml of Reconstitution Buffer, mix carefully by rolling or gentle shaking. DO NOT VORTEX!
2. Reaction Mix	Reconstitute the with 0.30 ml of Reconstitution Buffer, mix by rolling or shaking
3. Amino Acids	Reconstitute the lyophilizate with 0.36 ml of Reconstitution Buffer, mix by rolling or shaking.
4. Methionine	Reconstitute the lyophilizate with 0.33 ml of Reconstitution Buffer, mix by rolling or shaking.
5. Reconstitution Buffer	1.6 ml; Ready-to-use solution; stable at 2–8°C, can also be stored at – 15°C to – 25°C.
6. Control vector GFP	Briefly centrifuge down the content of the bottle and reconstitute the lyophilizate with 50 µl of sterile DNase- and RNase- free water The solution is stable at –15°C to –25°C

Table 2.7.2-2 Preparation of working solutions (RTS 100 HY E. coli *Kit)*

Contents	Preparation of working solution for one 50 µl reaction
Reaction Solution	Into one of the supplied reaction tubes pipette the following components: 1. 12 µl *E. coli* Lysate 2. 10 µl Reaction Mix 3. 12 µl Amino Acids 4. 1 µl Methionine 5. 5 µl Reconstitution Buffer 6. 0.5 µg of circular DNA template or 0.1–0.5 µg of linear template in 10 µl of water or TE-buffer. Mix carefully by rolling or gentle shaking. DO NOT VORTEX!

Table 2.7.2-3 Radioactive labelling (RTS 100 HY E. coli *Kit)*

Step	Action
1	Reconstitute bottles 1 to 4 and bottle 6 according to Table 2.4.5.2-1
2	Dilute 10 µl of the reconstituted Methionine solution of 3.2.2 with 990 µl of nuclease-free water to yield a 1 mM Methionine solution.
3	For one radioactive reaction prepare the following Reaction Solution in one of the reaction tubes: 1. 6 µl *E. coli* Lysate 2. 5 µl Reaction Mix 3. 6 µl Amino Acids 4. 1.25 µl 1 mM Methionine solution (see Note) 5. 2 µl of a L-[^{35}S]-Methionine (SJ 235 Amersham) 15 mCi/ml 6. 2.5 µl Reconstitution Buffer 7. 0.25 µg of the plasmid DNA or 0.05–0.25 µg of linear template in 2.25 µl of water or TE-buffer.
4	The reaction temperature: 30°C
5	Stop the reaction after 60 min.
6	Apply 2–5 µl of the reaction samples onto SDS-polyacrylamide gels. Note: For optimum results precipitate the proteins with cold acetone before applying onto SDS-polyacrylamide gels (see in the appendix section).
7	After the separation, dry the gel and place into Storage Phosphor Screens for autoradiography (3–20 hrs exposition time).

Note: Addition of unlabeled methionine to the labelling reaction is required to prevent premature termination for larger proteins or proteins with many methionine residues.

2.7.3 RTS 500 High Yield E. coli *Kit*

The principle of RTS 500 reaction is that transcription and translation take place simultaneously in the 1 ml reaction compartment, and substrates and energy components essential for the reaction are continuously supplied *via* a semipermeable membrane from the 10 ml feeding compartment (Figure 2.7.3-1). Through the same membrane potentially inhibitory reaction by-products are diluted *via* diffusion from reaction compartment into feeding compartment (RTS 500 HY *E. coli* Kit). The reaction device is supplied together with the kit.

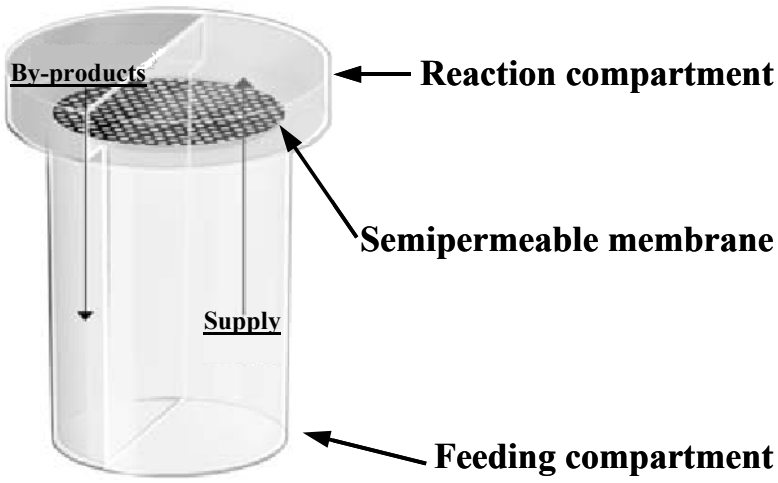


Figure 2.7.3-1 Reaction device for RTS 500. Principles of continuous exchange cell-free protein synthesis as suggested by Spirin (1988). Essential compartments and membrane are indicated. Arrows within the device indicate direction of the flow.

Lyophilised reagents provided with RTS 500 Kit were reconstituted according to the Kit directions (Table 2.7.3-1).

Table 2.7.3-1 Reconstitution of RTS 500 reaction components (RTS 500 HY E. coli *Kit)*

Content	Reconstitution/Preparation of working solution
E. coli Lysate	Reconstitute the lyophilizate with 0.525 ml of Reconstitution Buffer, mix carefully by rolling or gentle shaking. DO NOT VORTEX!
Reaction Mix	Reconstitute the lyophilizate with 0.25 ml of Reconstitution Buffer mix by rolling or shaking.
Feeding Mix	Reconstitute the lyophilizate with 8.1 ml of Reconstitution Buffer, mix by rolling or shaking.
Amino Acid Mix	Reconstitute the lyophilizate with without Methionine 3 ml of Reconstitution Buffer, mix by rolling or shaking.
Methionine	Reconstitute the lyophilizate with 1.8 ml of Reconstitution Buffer, mix by rolling or shaking.
Reconstitution Buffer	Ready-to-use solution stable at 2-8°C, but can also be stored at –15 to – 25°C.

The working solutions were prepared from reconstituted reagent according to the RTS 500 Kit protocol (Table 2.7.3-2). The following Reaction solution was loaded into 1 ml reaction compartment of the supplied reaction device. The

Feeding solution was loaded into feeding compartment with care avoiding air bubbles. The loaded reaction device was introduced into ProteoMaster instrument and incubated at 30°C if not otherwise indicated.

Table 2.7.3-2 Preparation of working solution for RTS 500 (RTS 500 HY E. coli *Kit)*

Content	Reconstitution/Preparation of working solution
Feeding Solution	Add 2.65 ml of the reconstituted Amino Acid Mix without Methionine and 0.3 ml of reconstituted Methionine to Feeding Solution 3. Mix by rolling or shaking. Total volume of Feeding Solution is 11 ml.
Reaction Solution	To the content of *E. coli* lysate, add 0.225 ml of the reconstituted Reaction Mix, 0.27 ml of the reconstituted Amino Acid Mix without Methionine and 30 µl of reconstituted Methionine. Add 10–15 µg of the DNA template in a maximum volume of 50 µl. Mix carefully by rolling or gentle shaking. Total volume of reaction solution is 1.1 ml. DO NOT VORTEX.

After stop of the reactions with GFP, 5 to 10 µl were taken for the SDS-PAAG electrophoretic analysis of a total fraction and the rest of the reactions was stored at + 4°C overnight to gain the active form. The native PAAG electrophoretic analyses were performed for 24-48 hours (see Chapter Result, section 3.4 for details).

Chapter 3

RESULTS

This chapter is organised along the processing steps of a coupled transcription-translation system. We start with breaking the cells and walk *via* optimisation of the *in vitro* system until the fractionation of the cell lysate.

3.1 Optimisation of the method for cell breakage

Liquid-based homogenisation is the most widely used cell disruption technique for small volumes and cultured cells. One of the effective methods to obtain cell lysate has always been a grinding of the cells together with hard and small particles, like Alcoa – aluminium oxide (Alcoa-305) - in a mortar, until the production of this reactive was stopped. This led us to the analysis of existing methods for isolation of cell lysate and selection of the best. Among others was French Press technique named for the inventor Stacey French, an instrument used to disrupt cells. The sample (volume of 40 to 250 ml) is placed in the bottom chamber of the machine, and the pressure on the sample is raised to a particular level. The cells are then forced into a different chamber, while the rapid change in depression between the two chambers causes the rupture of the cells. Only two passes are required for efficient lysis due to high pressures of 7-8,000 psi used with this process. The equipment is expensive, but the French press is often the method of choice for breaking bacterial cells mechanically. Nirenberg and Matthaei first used a mortar and pestle to grind the cells and release the cell sap, but soon moved to the French press to create their cell-free system of *E. coli* bacteria.

Operating Principle of the Microfluidizer Processor

Microfluidics' homogenisers contain air-powered intensifier pumps designed to supply the desired pressure at a constant rate to the product stream. As the pump travels through its pressure stroke, it drives the product at constant pressure through precisely defined fixed-geometry microchannels within the interaction chamber.

As a result, the product stream accelerates to high velocities creating shear rates within the product stream that are orders of magnitude greater than any other conventional means. Pressure (17,000 psi) and shear forces are the parameters that break the cells.

The Microfluidizer model M-110L (Figure 3.1-1) is a suitable apparatus for producing high yields in cell rupture with minimal processing and easy recovery. In this apparatus cell walls are ruptured by shearing forces that do not destroy cell contents, such as ribosomes, and allow for easy separation.

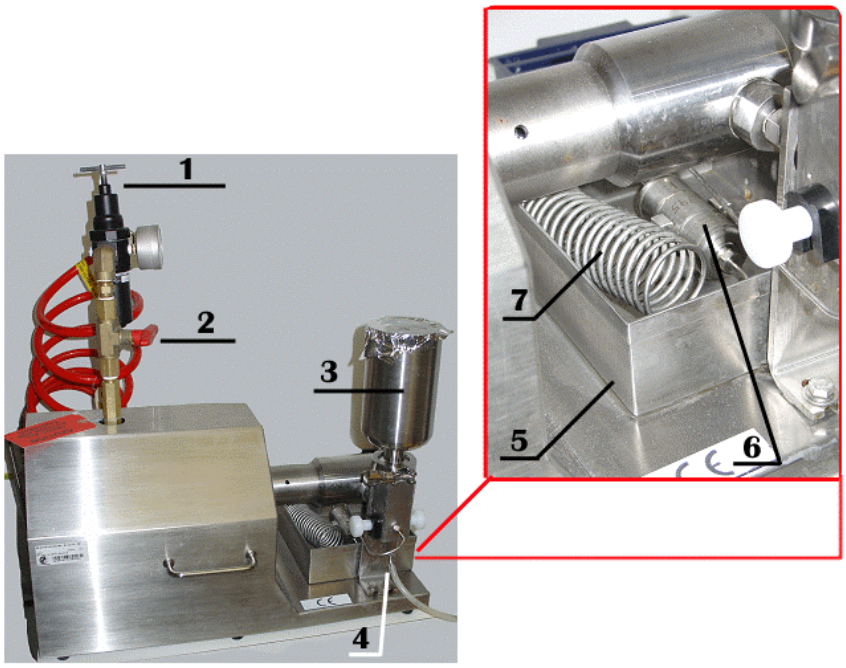


Figure 3.1-1 Microfluidizer Processor M-110L. High shear fluid processing system for lab-volume cell rupture from Microfluidics™. Designations: **1** – air pressure regulator, **2** – air valve, **3** – feed reservoir, **4** – continuous flow outlet, **5** – an ice-water box, with **6** – interaction chamber where cells are ruptured, and **7** – cooling coil.

After all it is suitable for several reasons:

sample volume – a wide range can be used, from 14 ml (12 ml recovery) up to 400 ml and/or 1 litre;

temperature of the S30 cell lysate – the pressure block can be cooled with ice water that leaves the sample temperature at low levels: the temperature is raised during the cell crashing from 0 to about 12°C for less than 2 min;

dilution – the S30 cell lysate has to be as concentrated as possible. This can be also a limiting step, because if cell suspension is too thick it can block the microchannel inside the pressure steel-block (interaction chamber, Figure 3.1-1, assigned as 6). We found out that a suspension of cells with 1 ml of buffer per 1 g of cells pellet that can go through a tip of 10 ml glass pipette by gravity flow can suffice. Therefore, the intracellular cell milieu is diluted slightly more than 2-fold.

3.2 Selection of the optimal *E. coli* strain for the batch system

We thought at the beginning of this project that a strain deficient in RNases would help to obtain a high protein yield output. However, we observed that mRNA degradation does not play a main role for the synthesis of a protein it codes for (see section 3.6). The synthesis of mRNA goes faster than its degradation and thus degradation does not affect the rate of protein synthesis. For example, when we compared the *E. coli* strains Can/20-12E (Zaniewski *et al.*, 1984) with BL21, the latter was superior; Can/20-12E lacks five RNases (I^-, II^-, D^-, BN^- and T^-) and BL21 was in contrast deficient in the protease genes *lon* and *ompT*. The *lon* gene codes for an aminopeptidase La, a heat shock protein and ATP-dependent Ser protease La with broad substrate specificity that degrades short-lived regulatory and abnormal proteins in the presence of ATP; it hydrolyses two ATPs for each peptide bond cleaved in the protein substrate. The *ompT* gene codes for an outer membrane protein T (OmpT) that is also called protease VII. This protease can cleave T7 RNA polymerase, ferric enterobactin receptor protein (FEP), antimicrobial peptide protamine and other proteins preferentially between two consecutive basic amino acids (Dekker *et al.*, 2001).

In a course of this study BL21 was used as a main source for cell-free lysate preparations. Also, a number of protease deficient strains were selected from the American Type Culture Collection (ATCC) and from the literature, a brief overview of their genotypes and the corresponding sources is given in

Table 3.2-1. The efficiencies of the corresponding lysates will be tested after this thesis project.

Table 3.2-1. Strains, selected for further study in cell-free system for protein synthesis.

Name	Genotype	Description	Source	References
Rosetta ™	BL21 *lacYZ* deletion, lacks *lon* and *ompT* proteases; Cam^r	Six tRNAs on a plasmid: *proL* tRNA, *leuW* tRNA, *argW* tRNA, *glyT* tRNA, *argU* tRNA, *ileX* tRNA	Novagen	
LC24	*snoC*; Tc^r	Protease-deficient	ATCC	(Huala *et al.*, 1991)
HA63	*snoB*; *lon*			
SF120	F^- Δ(lacIPOZY)X74 *galE galK thi rpsL* ΔPhoA(PvuII) ptr32::ΩCatR degP41(ΔPstI-KanR) Δ(*ompT*)			(Baneyx and Georgiou, 1991)
C41 (DE3)	F *dcm ompT hsdS* (rB mB) *gal* λ (DE3); Amp^r	Protease-deficient; derivatives of *E. coli* B strain, BL21(DE3)	Avidis	(Miroux and Walker, 1996; Studier and Moffatt, 1986)
C43 (DE3)	*F dcm ompT hsdS* (rB mB) *gal* λ (DE3); Amp^r; derivative of C41(DE3)			

3.3 Batch system: *in vitro* coupled transcription-translation

The *in vitro* coupled transcription-translation system has got its name from the fact that both processes of transcription and translation are performed in one and the same tube. In order to obtain the protein of interest there is no need to obtain and isolate the mRNA in a significant amount before giving it to the translation reaction. In a coupled system there are already components for both processes, and the only requirement is to add the gene (linear or circular DNA) under T7 RNA polymerase promoter. Thus, transcription is performed by T7 RNA polymerase and translation is performed by ribosomes and additional factors present in the cell extract. It may sound easy, but it is also necessary to add the "building blocks" for either process, nucleotide triphosphates (NTPs) for mRNA and amino acids for protein synthesis. It is also necessary to regenerate the general energy source, which is adenosine triphosphate (ATP), as well as to

buffer the whole system to an "*in vivo* near" pH and ionic condition and to block the activity of the intrinsic *E. coli* RNA polymerase in order to avoid unspecific products. All these aspects were taken into account, adjusted and published by several groups before (Zamecnik, 1969). Kim and Swartz, 2000 (Kim and Swartz, 2000) published a table with final concentrations of most of the components of the system, which seemed to us inconvenient due to the odd concentrations, e.g. 57.2 mM HEPES instead of 60 mM (see Table 3.3-1, central column). After the final concentration of each of the components was brought to even numbers (Table 3.3-1, column on the right hand), stock solutions were prepared, divided into suitable aliquots, shock-frozen and stored at –80°C. One of our aims in optimising the batch system for *in vitro* coupled transcription-translation was to determine volumes and final concentrations of mixes and solutions to ease up the pipetting scheme and storage of stock solutions.

In general, the reaction mix has to be mixed together in a few steps. Before I will talk about the steps I will consider the various mixes. The HEPES salts solution (HSS) is a mixture of stable components of the system, which buffers its pH (HEPES•KOH) and supports the ionic condition (NH_4^+ acetate, K^+ glutamate), as well as blocks the pathway for degradation of some amino acids (Na^+ oxalate, [Kim, 2000 #14604]); it can be stored for long time at – 80°C and is not affected by multiple thawing. The other mixture, small compound solution (SCS), which name says already that it contains relatively small molecules - source for energy regeneration and tRNAs, is less stable than HSS and needs fresh preparation about every half a year. Both stock-solutions mentioned are added to the LM-Mix together with NTPs, PEP (for ATP regeneration, too), PEG-8000 (to create "*in vivo* near" viscosity of the system) and amino acids (two of 20 amino acids, aspartic acid and tyrosine are prepared separately due to their solubility requirements). This LM-Mix is prepared on the same day, when the main experiment is planned and could be used for a set of reactions at once. Still, this LM-Mix is added to the Master-Mix along with a defined concentration of T7 RNA polymerase, Mg^{2+} acetate, labelled methionine (if

required) to detect synthesized proteins easily, and antibiotic rifampicin, which blocks activity of the *E. coli* RNA polymerase molecule (β subunit, Williams and Piddock, 1998). And still this Master-Mix is not yet a reaction, ready for incubation, because some of the essential components are missing. To the Master-Mix we added DNA with a gene of interest under T7 promoter and S30 cell lysate from *E. coli* BL21 strain. Eventually this is the reaction mix (R-Mix), ready for protein synthesis.

Table 3.3-1 Optimisation of the final concentration of the batch cell-free system for protein synthesis.

Component	Final concentration (Kim and Swartz, 2000)	Final concentration 19. Jan 02
HEPES-KOH (pH 8.2)	57,2 mM	60 mM
Ammonium acetate	80 mM	80 mM
Potassium glutamate	200 mM	230 mM
Sodium oxalate	2,7 mM	3 mM
DTT	1,76 mM	2 mM
Cycle-AMP	0,67 mM	0.7 mM
Folinic acid	34 µg/ml	35 µg/ml
tRNAs	340 µg/ml	350 µg/ml
NADH	0,33 mM	0.35 mM
Coenzyme A	0,27 mM	0.3 mM
ATP	1,2 mM	1.5 mM
CTP	0,86 mM	1 mM
GTP	0,86 mM	1 mM
UTP	0,86 mM	1 mM
PEG-8000	2% (w/v)	2% (w/v)
Methionine	2 mM	2 mM
Amino acids (19)	0.5 mM	2 mM
PEP	33 mM	35 mM
Magnesium acetate	15 mM	12 mM
T7 RNA polymerase	30 µg/ml	100 µg/ml
E. coli S30 cell lysate		4-6 A_{260}
Plasmid DNA		4 µg/60 µl
Rifampicin		10 µg/ml
^{3}H-Leu	1.2 µM	-----
^{35}S-Met	-----	1.5 µM

Figure 3.3-1 gives a schematic overview of the composition of the R-Mix together with the dilution factors (DF), which facilitate calculation of final concentrations.

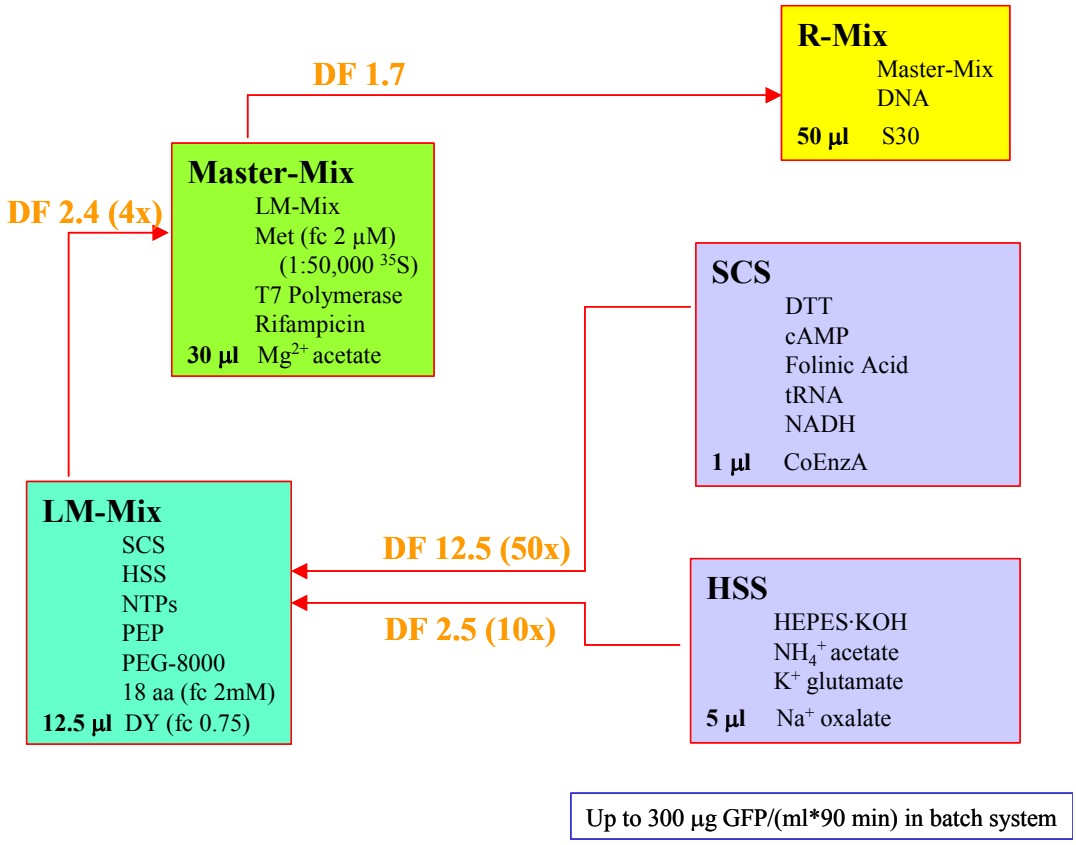


Figure 3.3-1. Schematic representation of the modified Swartz & Kim in vitro *system* (Kim and Swartz, 2000). The reaction mix (R-Mix) is prepared from the given stock solutions, which are added subsequently according to this scheme. Golden numbers stand for dilution factor (DF).

We started with the investigation of the parameters of the optimised system, measuring the levels of GFPcyc3 synthesis by TCA precipitation. Kinetics of the GFPcyc3 synthesis reflect the way this reaction goes, *i.e.* the rate, as well as the extent of the protein synthesis – how soon does the system reaches the GFPcyc3 saturation phase (Figure 3.3-2). According to our measurements the extent of the reaction can be observed after 120 minutes, while its rate after 15 minutes.

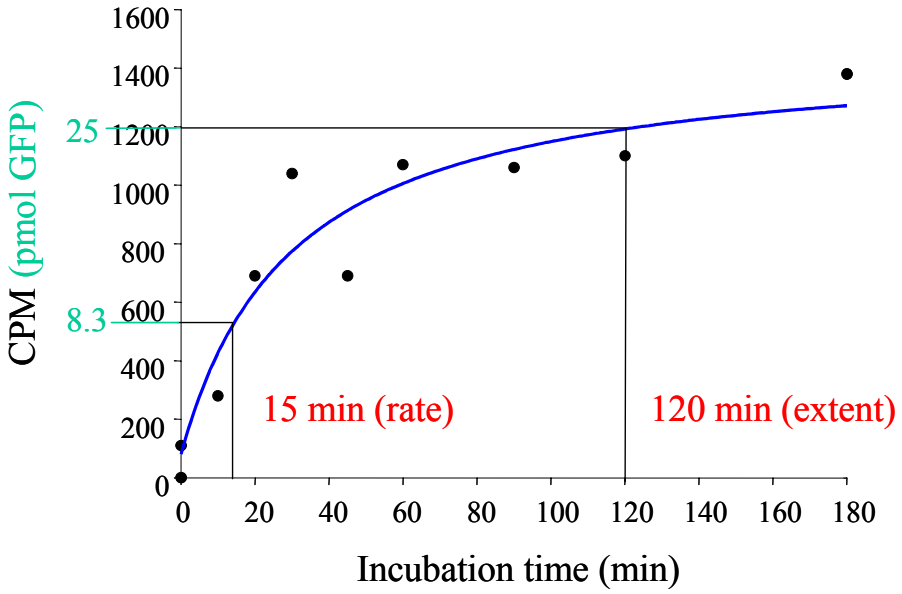


Figure 3.3-2 Kinetics of the GFPcyc3 synthesis. Characterising the batch system by the rate and extent of protein synthesis reaction.

The next step was the systematic modelling of the modified Swartz system, in order to have a reliable method to characterise the protein synthesis. The TCA precipitation analysis results always stroke us with a high background radioactivity counts. With the following example I will try to explain what that means. The sample reaction without plasmid DNA template would serve as the 'background' for those reactions that contain DNA template for protein synthesis. After TCA precipitation the minus DNA sample shows 400 cpm, which is taken as a background and subtracted from the 1600 counts of the sample with a DNA template. We looked for ways to reduce it, and examined whether a pre-incubation would diminish the large background. For this we pre-incubated reaction mixes in the absence of [^{35}S]-Met extending pre-incubation time from 10 to 30 and 90 minutes that was followed by a normal main incubation of 120 minutes in presence of [^{35}S]-Met (Figure 3.3-2). As the figure shows, the background level stayed same (Figure 3.3-2, green closed circles), indicating that it does not depend on the pre-incubation time, and the [^{35}S]-Met incorporation into GFPcyc3 synthesis in the presence of DNA template decreases (Figure 3.3-2, red closed circle). A possible explanation of the high

background might be the transfer of methionine from the Met-tRNA to the N-termini of proteins present in the S30 extract.

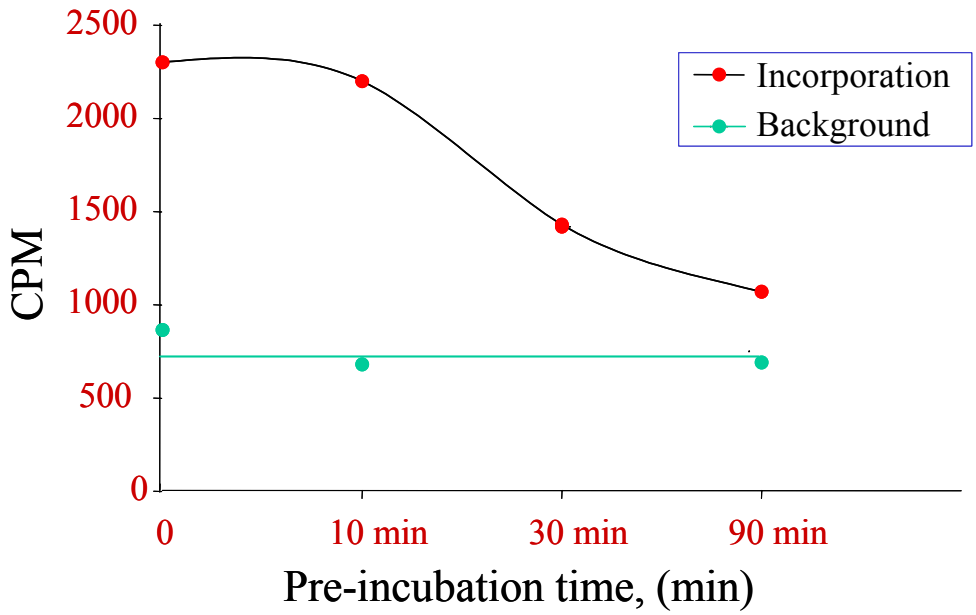


Figure 3.3-4 Pre-incubation of the reaction mixes in the absence of [^{35}S]-Met. Prior the main incubation samples with and without DNA (background) were incubated for 10, 30, and 90 minutes, followed by normal main-incubation of 120 minutes. Green closed circles indicate background level, and red closed circles indicate the level of [^{35}S].

Therefore, the large background cannot be removed by pre-incubation. Table 3.3-2 shows that background is independent of the presence of T7 RNA polymerase (compare cpm values of the assay #1 and assay #2, which are about the same). The background showed also, that it is not affected much by the presence of DNA template in the absence of the T7 RNA polymerase (*i.e.*, no mRNA production from the given DNA template, that results in the absence of protein synthesis). In the presence of both T7 RNA polymerase and plasmid DNA template the synthesis is only two fold above the background.

These data indicate that the TCA analysis is not a precise measure for this system that should reflect both given protein synthesis levels and the quality of the synthesised protein. If the hypothesis is correct that the background is caused by amino acid transferases that transfer aminoacyl residue from an aa-tRNA to the N-termini of mature proteins, we would expect the following: the background cpm would be scattered over a large MW-range, whereas the DNA

template dependent protein synthesis (GFP) would lead to a focusing of the corresponding cpm in the GFP product. Therefore, the background cpm would not be seen on an SDS-gel in contrast to the GFP counts. Figure 3.3-5 shows that our expectations were fulfilled: only a GFP band is seen.

Table 3.3-2 Investigation of the background in a TCA precipitation dependency on the presence of T7 RNA polymerase.

Assay, No.	T7 RNA polymerase	DNA plasmid	CPM (background)
1	-	-	800
2	+	-	940
3	-	+	1275
4	+	+	2300

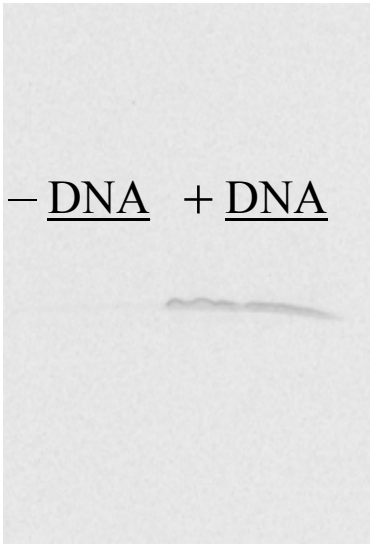


Figure 3.3-5 SDS gel with synthesised GFP scanned with PhosphoImager. Only proteins with [^{35}S]-Met label are seen as a defined band of the expected size coded by the added DNA template.

Another point was to improve the expressivity of the system itself in order to obtain higher levels of active and soluble fraction of the protein of interest. A number of approaches on this field are discussed in the sections below.

3.4 Quality criteria for the judgement of GFP expression

In order to characterize the activity of this cell-free system by analysing yield and activity of the synthesized protein, we used the fluorescent protein GFP (Chalfie *et al.*, 1994; Prasher, 1995; Prasher *et al.*, 1992) as a reporter protein, because this protein is easy to handle and to detect. Another good reason for it was that with this protein we could not only determine the total yield, but also the active fraction of the synthesized protein. In many cases the system can produce µg to mg amounts per ml of total protein synthesis. With GFP it is easy to identify active fraction of native protein in respect to the total protein synthesis. After the synthesis of GFP is finished, its fluorophore has to be folded properly, which is a criterion for its activity. The activity is determined by the fluorescence of GFP using ultraviolet-light (UV) excitation (Figure 3.4-1).

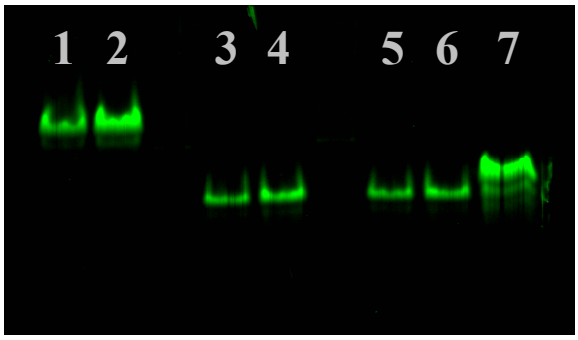


Figure 3.4-1 Green fluorescent protein in native PAA gel exposed to ultraviolet-light excitation. Lanes 1 and 2 contain GFPcyc3; lanes 3 and 4 – WT GFP; lanes 5 to 7 – WT GFP standard (with known concentration), in different amounts.

A reporter GFP cycle3 (GFPcyc3) has three point mutations that allow a fast maturation within 3-4 hours, whereas WT requires maturation overnight at 4°C (Crameri *et al.*, 1996). It was expressed from DNA plasmid for *in vitro* expression (pIVEX2.2) with T7 promoter and STREP-tag at the N-terminus. The latter one explains the different migration behaviour in the native gel as compared to the WT. Samples of the reaction mix during incubation were withdrawn at different time points (kinetics), and studied for rate and extent of protein synthesis reaction.

3.4.1 Denaturing SDS polyacrylamide gel

Analysing the GFP band within the S30 protein pattern we were lucky to detect that the GFP band hardly overlaps with any of the S30 extract bands. This opened the possibility to determine the total amount of the protein synthesised *in vitro*. Reference curves for the determination of the amount of synthesized GFP were made by applying defined GFP amounts to the gels. The input variations of the S30 reaction mixture per lane were normalized by scanning a well-defined band from the S30 pattern and taking into account the respective pixel numbers (see Figure 3.4-2 and Figure 3.4-3A).

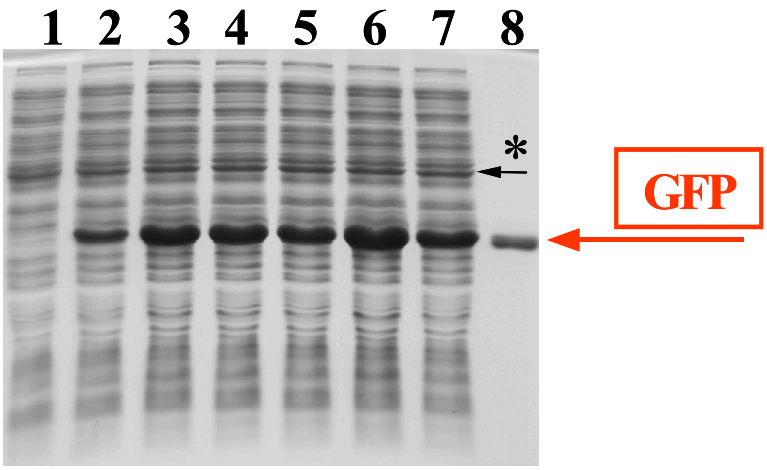


Figure 3.4-2 GFPcyc3 in SDS-PAAG. Lane 1 corresponds to sample in the absence of GFPcyc3 plasmid; lanes 2-7 – in the presence of GFPcyc3 plasmid; lane 8 – standard GFP. * – band used to normalize input.

3.4.2 Native polyacrylamide gel

We know already that GFP can emit green light when exited with ultra violet waves (UV-light), and this event indicates that the protein is folded correctly and its fluorophore is formed properly, and thus we can judge GFPcyc3 by quality. We examined whether this property is maintained after electrophoresis in the PAAG under non-denaturing conditions (*i.e.* in absence of SDS), thus conserving its active form, and whether radiation of the gels with UV-light will still result in fluorescence of this protein band. Indeed, in native PAAG conditions we could detect the fluorescence of GFP in the gel (Figure 3.4-3B), and it allows estimating the concentration of active molecules when

correlated to an active reference GFP of known concentration. Combining this method with the detection of total amount of GFP from one and the same reaction vial will allow us to identify the percentage of active GFP (Figure 3.4-3). Such analysis revealed that the active fraction of GFP is not higher than 60%, and to find a way to improve it up to 100% was one of our aims.

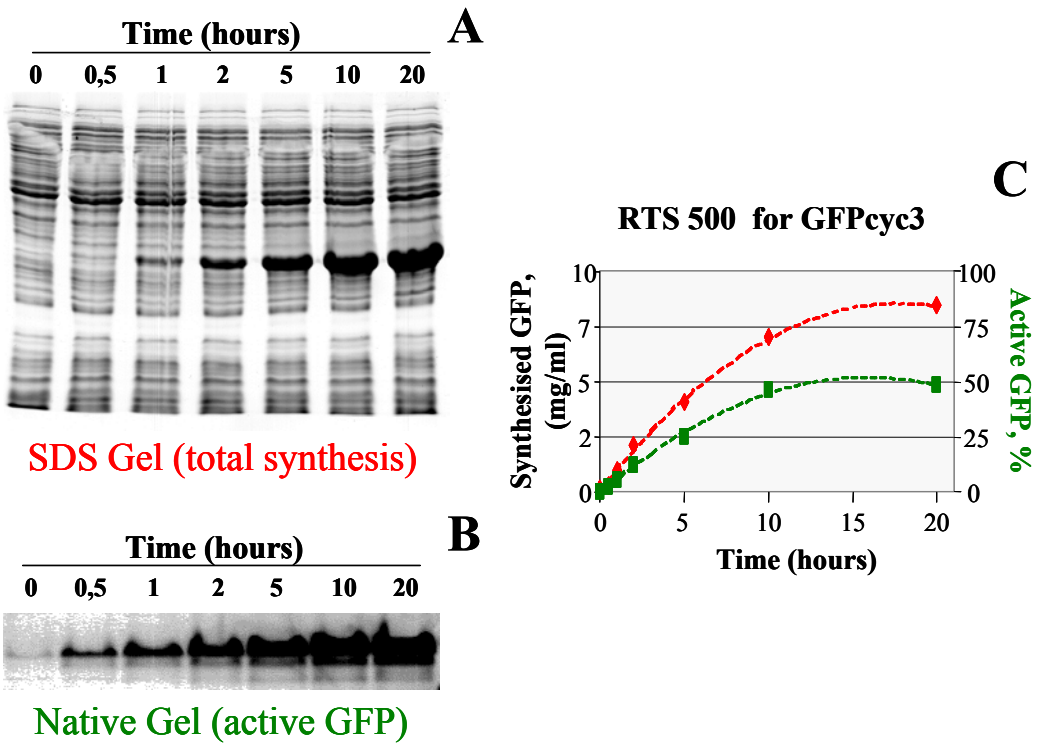


Figure 3.4-3 SDS and native PAA gel analysis. (**A**) GFPcyc3 total yield increase in course of time; (**B**) GFPcyc3 activity scanned in native gel; (**C**) processed data that reflect the GFPcyc3 synthesis in RTS 500 large-scale reaction set. Red symbols – total yield, green – active fraction, which is about 60%.

3.4.3 Detecting synthesized protein in small-scale reactions: incorporation of a labelled amino acid

The presence of [^{35}S]-Met in the batch or RTS reaction results in its incorporation into the nascent polypeptide chain and permits determination of the reporter protein band.

With use of scan control software and Image Analysis Software we determined relative band intensities (Figure 3.4-4), that could not be correlated to a GFP of known concentration in this assay due to absence of a [^{35}S]-Met labelled GFP standard. Thus, measurements were only relative, but they allow

identifying fragmented proteins. This also allows analysing synthesis of a protein, which band is overlapping with one of the S30-extract proteins and thus cannot be detected by Coomassie staining.

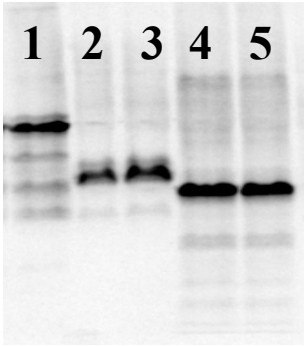


Figure 3.4-4 [^{35}S]-Met labelled proteins in SDS-PAAG Digitised image generated after scanning the gel by PhosphorImager™. Lane 1 – EF-Tu; lanes 2 and 3 – EF-Ts; lanes 4 and 5 – GFPcyc3.

TCA precipitation allows estimation of relative number of single protein chains synthesised by single ribosome, since we know the specific activity of [^{35}S]-Met, the number of Met residues per protein molecule and the number of 70S ribosomes per aliquot (we used the relationship 1 A_{260} unit of S30 is 24 pmol x 0.8 ≈ 18 pmol 70S ribosomes). This method was applied, when studying the T7 RNA polymerase mutants.

3.5 Synchronising the reactions of transcription and translation

3.5.1 Utilization of the "slow" T7 RNA polymerases (M. Dreyfus, Paris)

As mentioned earlier an *in vitro* coupled transcription-translation system utilises T7 RNA polymerase (RNAP, or transcriptase), derived from T7 bacteriophage, for mRNA transcription from the plasmid DNA, and the machinery for translation (ribosome, factors, tRNA synthetases, *etc*.) from the *E. coli* cell lysate.

In *E. coli* cells processes of transcription and translation are tightly coupled. It is known that *E. coli* RNAP proceeds with a speed of ~60 nucleotides per second, and while mRNA is synthesised *E. coli* ribosomes

initiate translation on the nascent chain of mRNA and proceed with a speed of ~20 amino acids per second (Bremer and Dennis, 1996), which is ~20 codons been deciphered per second. It follows that the first ribosome pursues immediately the transcriptase. These facts explain the tight coupling of transcription and translation leaving no room for a significant gap between the transcriptase and the following ribosome. Therefore, the nascent mRNA chain cannot form a secondary structure and thus complicate or even block translation elongation (Iost and Dreyfus, 1995). On the other hand, ribosomes protect mRNA from endonucleases that usually initiate the process of degradation from the 5' into 3' end direction.

It is known that wild type T7 RNAP synthesises about 5 times faster than the wild type *E. coli* RNAP at 37°C (Chamberlin and Ring, 1973; Lewicki *et al.*, 1993). Thus, using T7 transcriptase could influence tight coupling of transcription-translation process in a cell-free system as described in the preceding paragraph. This means that the situation now has pretty much changed for the ribosomes. Even though they might manage to initiate on a nascent mRNA without problems, the distance between T7 RNAP and the first translating ribosome will dramatically increase allowing formation of secondary structures that 70S ribosomes are not capable to resolve. The result is that ribosomes are stalled on the mRNA.

To overcome such a dissonance in this system, Prof. Dr. M. Dreyfus kindly provided us T7 RNAP mutants (Figure 3.5.1-1) that might solve the problem.

All constructs coding for T7 RNAPs are derivatives of pBH1161 (He *et al.*, 1997) and therefore they all encode the transcriptase with a His6 tag. Besides the wild type enzyme, these plasmids encode the following polymerases:

I810S – single mutant with substitution of isoleucine into serine in position 810, (Bonner *et al.*, 1994; Bonner *et al.*, 1992), a very slow RNA polymerase (1/6 as fast as WT *in vitro*; Bonner *et al.*, 1994), due to a relatively severe active site mutation. This enzyme is hardly faster than the *E. coli* RNA polymerase, but it is very poorly processive, which limits its usefulness.

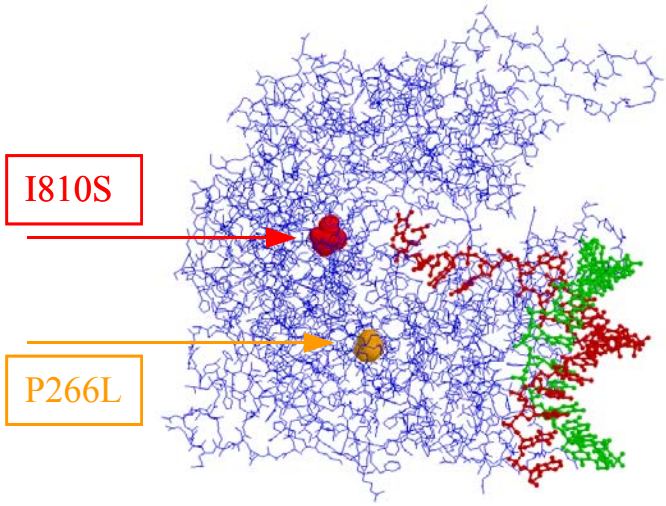


Figure 3.5.1-1 Structure of a transcribing T7 RNA polymerase initiation complex (Cheetham and Steitz, 1999). Blue colour indicates the T7 RNAP molecule; double helix represents the unwound DNA (green and red chains). Arrows direct the point mutation sites, isoleucine 810 to serine, and proline 266 to leucine.

I810S/P266L – double mutant, has the same active-site mutation and an additional mutation that is located far apart from the active site – a conserved proline at position 266 is substituted into leucine. As a result of this secondary mutation, the processivity is greatly improved.

I810N/P266L – where the I810N mutation appeared as a by-product of mutagenesis on I810S. According to estimates, this enzyme is ca. 2.5-fold slower than WT (this is an *in vivo* measurement; see (Iost *et al.*, 1992; Makarova *et al.*, 1995), but the cases examined in the *in vivo* and *in vitro* rate measurements are consistent), *i.e.* the active site change is less severe in this case than in the I810S case. In association with P266L, I810N may represent a useful compromise between the reduction of speed, and the maintenance of a high processivity.

P266L – this enzyme carries only the mutation increasing processivity. It is more active than the WT.

Taking into account the information about the T7 RNAP mutants we expected that the double mutant I810N/P266L would enable "*in vivo* near" tight coupling, resulting in the prevalence of a completed protein chains and fragmentation decrease. First we analysed the T7 RNA polymerase mutants for

mRNA production in the batch system with small-scale reactions using [^{35}S]-Met for GFPcyc3 detection in order to identify, whether there is any difference in the protein yield. From the results summarized in Table 3.5.1-1, based on TCA precipitation and relative number of molecules synthesised by single ribosome (assuming that 30% of the ribosomes were active), we could identify that:

(i) GFP synthesis by ribosomes from an mRNA transcribed by P266L mutant was about the same as compared to the wild type;

(ii) translation in the presence of the I810S T7 RNAP mutant resulted in two-fold decrease of GFP synthesis, compared to the wild type, and so was the translation of mRNA from I810S/P266L double mutant – in this case it may be due to severe active site mutation as reported above, which could result in defective or abortive mRNA transcripts;

(iii) translation in the presence of another double mutant, I810N/P266L, was just slightly improved above the discussed mutants.

Table 3.5.1-2 Activity of the T7 RNAP mutants in a batch system: [^{35}S]-Met incorporation and TCA precipitation results.

Mutant	activity (GFP per ribosome)
wt	**3.8**
I810S	**1.7**
P266L	**4.1**
I801S/P266L	**1.8**
I810N/P266L	**2.3**

Reaction volume – 50 µl; samples were withdrawn after 15 min of incubation at 37 °C.

Taking into account these date we selected T7 RNAP double mutants as objects for further investigation, and the WT T7 polymerase as a control. We checked whether the mutant T7 polymerases can improve the ratio active- / total- GFP synthesis. We note that from this kind of assay we cannot judge the activity of GFP molecules. To investigate both the total yield and the active fraction we performed the following experiment, where we increased the

reaction volume to 1 ml and run it in a special vial, where the reaction chamber is supplied by building blocks from the feeding chamber (10 ml) through a semipermeable membrane.

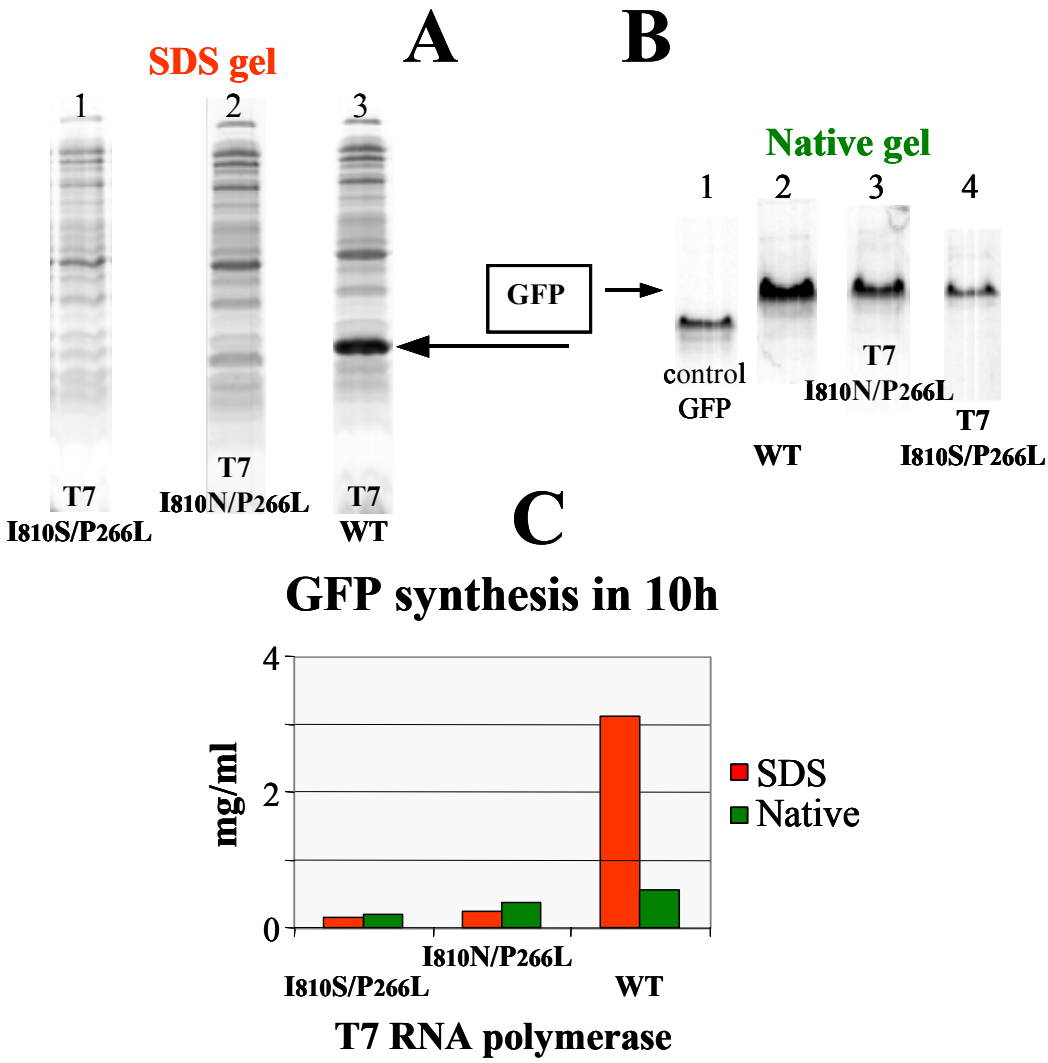


Figure 3.5.1-2 Fed-batch large-scale synthesis of GFP, (**not** RTS). T7 RNAP WT and slow mutants after 10h of incubation: SDS (**A**), and native (**B**) PAA gels, and comparison between total (**red**) and active (**green**) protein yield (**C**).

When GFPcyc3 gene was expressed in the large-scale reactions, its production from the mRNA transcribed by the double mutant I810N/P266L wins over GFPcyc3 production from mRNA molecules synthesised by the other double mutant I810S/P266L, for GFPcyc3 expression (compare lanes 1 and 2, Figure 3.5.1-2A). Both double mutants still run drastically behind in GFPcyc3 yield from the mRNA transcribed by the WT T7 RNAP (compare lanes 2 and 3, Figure 3.5.1-2A). In spite of the enormous difference in total yields seen with

the mutants versus wild-type T7 polymerase, the level of the active GFPcyc3 synthesized in the presence of the I810N/P266L mutant hardly differs to that in the presence of WT T7 RNAP (lanes 2 and 3 in Figure 3.5.1-2B; compare also red and green bars in Figure 3.5.1-2C). This means, that if the condition for tight coupling of transcription-translation processes is kept as "*in vivo* near", then even the protein folding is affected positively: the level of active GFPcyc3 is almost 100% (Figure 3.5.1-2C).

3.5.2 Varying the temperatures of incubation

Protein folding is a hierarchic process, sometimes pictured as an inverted funnel, in three fundamental stages. All proteins begin with a primary amino acid sequence, which folds into intermediate secondary shapes comprising the well known α helices and β sheets, and then into the final tertiary, or native form, in which they fulfil their function. Some proteins undergo a further phase, combining with other folded proteins to form quaternary structures.

In vivo expression of proteins often leads to inactive products. The reason is the following: if the synthesis of a single protein is so intensive that hydrophobic patches of neighboured nascent chains can contact before being buried and thus shielded inside the protein structure, the nascent chains will aggregate and eventually form inclusion bodies.

Decrease of incubation temperature slows down processes in cells (Lewicki *et al.*, 1993) and in cell-free systems, and the process of folding is among them. We carried out GFPcyc3 synthesis in RTS 100 and 500 under 37°C, 30°C, 25°C and 20°C to examine the effect on the protein synthesis (total yield), as well as on the amount of active protein molecules in this yield (active fraction). The results are summarized in Figure 3.5.2-1, where the left panel (Figure 3.5.2-1A) presents the comparison of total GFPcyc3 synthesis and a fraction of active GFPcyc3 at different temperatures. It is clearly seen that at 30°C the total fraction goes up to 2 mg/ml while the active fraction is about 60% of it, and as the incubation temperature was changed to 37°C the amount of active GFP cyc3

fraction is significantly reduced to less than 50% (0.6 mg/ml from 1.6 mg/ml of total protein). Decrease in temperature reduces the total yield of GFPcyc3 to 1 mg/ml at 25°C and to 0.6 mg/ml at 20°C, but for these cases almost 100% of GFP was active. There is a clear effect of temperature decrease both on the amount and quality of GFP, with best output of active fraction for 20°C where it reaches 100% (Figure 3.5.2-1B).

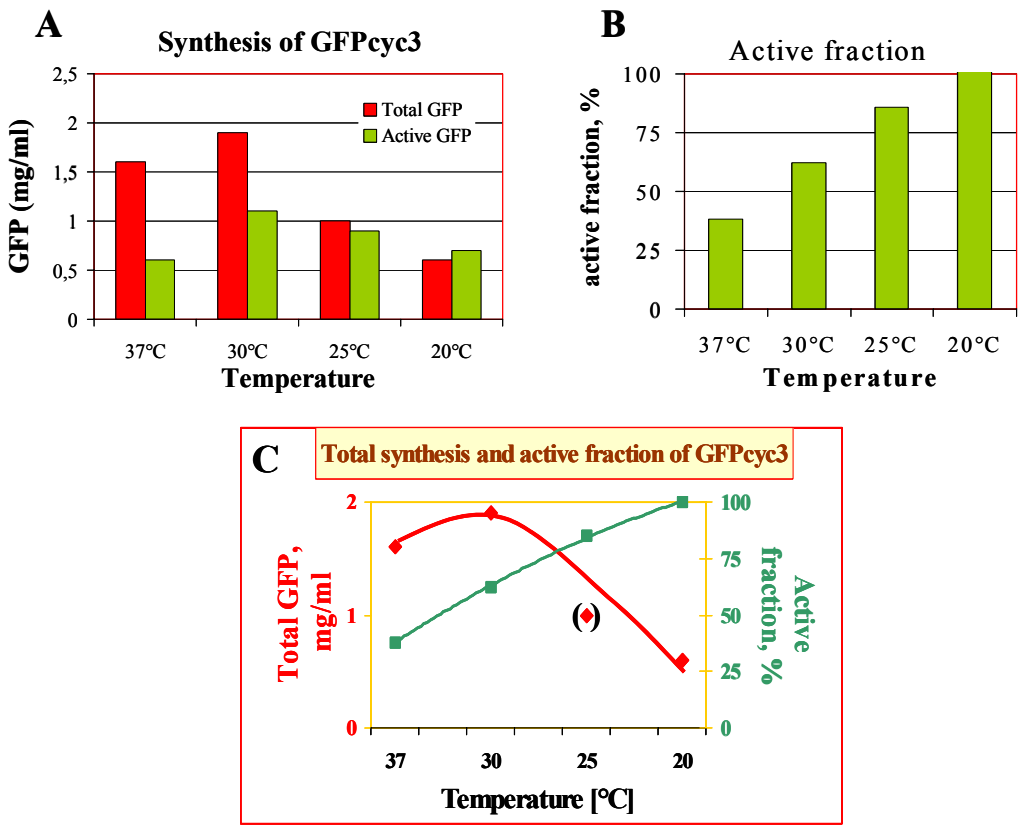


Figure 3.5.2-1 The effect of the temperature for incubation in RTS 500 on the active fraction of GFP. (**A**) comparison of the temperature effect on total protein synthesis and active fraction for different temperatures; (**B**) comparison of the behaviour of active fraction for different temperatures; (**C**) dependency curves of total yield (**red**) and active fraction (**green**) levels on the incubation temperature (X-axis).

Though the amount of GFPcyc3 goes down, the level of its activity significantly goes up (Figure 3.5.2-1C). In following experiments an incubation temperature of 20°C was often used as an alternative to incubation at 30°C.

3.6 An endeavour to increase the outcome of the given protein

3.6.1 Prolongation of the half-life of the mRNA: pseudo-circulation (J. Remme, Tartu)

In order to identify, whether the stability of a given mRNA in such an *in vitro* system is a critical factor for the expression of a given protein, we tested the expression of GFP from two different constructs described bellow, versus expression from pIVEX2.2GFPcyc3 (as a 100% control).

The idea for constructs mentioned came from literature data that point at the difference of stability for various RNAs in *E. coli* (or any) cells. Usually, the most stable RNAs against the activity of endonucleases are ribosomal ones (rRNAs). Half-life of naked rRNAs before assembly is longer due to pseudo-circulation, which is achieved by stem formation between 5' and 3' end of, for example, 23S rRNA (Figure 3.6.1-1). The spacer regions flanking the mature RNA sequences are also highly conserved. Precursor sequences at the 5' end and 3' end of 16S and 23S rRNA contain complementary sequence tracts that form strong base-paired stems enclosing the sequence of the mature species. The 146 nucleotides upstream of 16S rRNA include 131 involved in stem formation. Those and the 43 nucleotides immediately following mature 16S rRNA are identical in the four operons studied. The stem bracketing 23S rRNA involves 114 nucleotides on the 5' side and 71 nucleotides 3' to the 23S rRNA, and actually includes eight base pairs involving the 5' and 3' terminal nucleotides of mature 23S rRNA (see (Srivastava and Schlessinger, 1990) and references within).

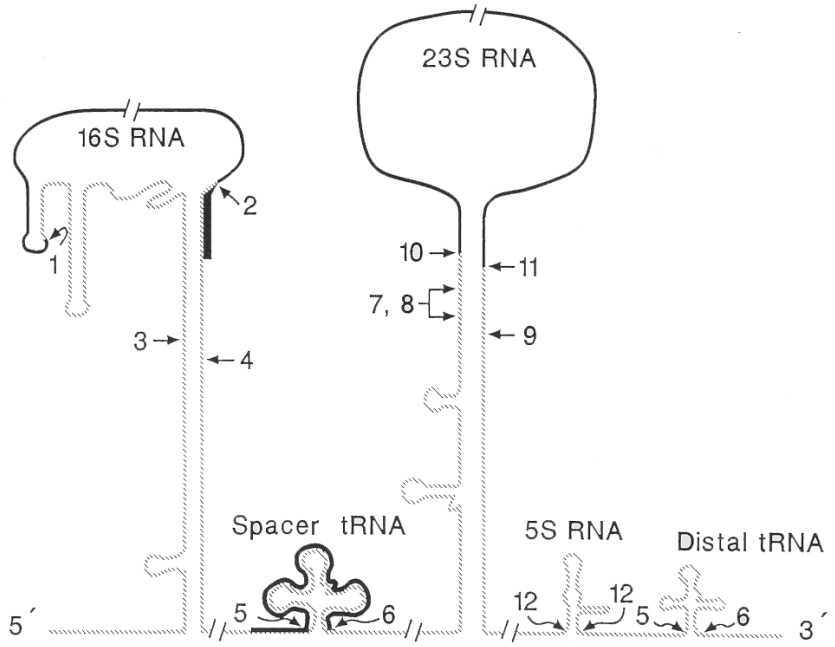


Figure 3.6.1-1 Schematic structure of an rrn operon and major processing steps for the 16S and 23S rRNA. The drawing is not to scale. Primary processing cleavages by RNase III (**3**, **4**, **7**, **8**, and **9**), and secondary processing to produce the mature termini of 16S rRNA (**1**, 5' end, and **2**, 3' end), 23S rRNA (**10**, 5' end, and **11**, 3' end), and 5S rRNA (**12**). The RNase P cleavage site (**5**) is shown at the 5' end, and RNase E site (**6**) at the 3' end of the tRNAs. Solid lines indicate mature 16S and 23S rRNA sequences; a hatched line indicates other precursor sequences (Srivastava and Schlessinger, 1990). In *E. coli* processing at the ends of the mature 16S and 23S rRNAs is coupled, since base pairing is required to generate the RNase III cleavage sites.

The stability determinant of pre-23S rRNA was used to increase the stability of mRNA (Liiv *et al.*, 1996). To this end, GFP gene was inserted into rRNA coding sequence (Figure 3.6.1-2).

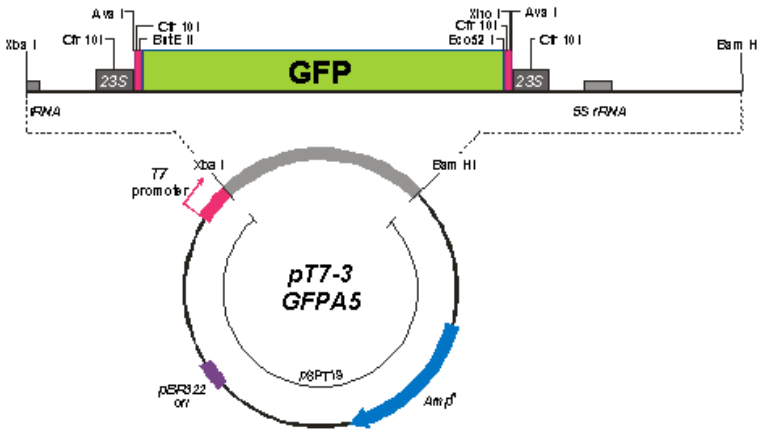


Figure 3.6.1-2 Map of pT7-3GFPA5 construct.

Ability of the chimerical RNA (5' region of 23S rRNA - GFP coding sequence - 3' region 23S rRNA) to form the processing stem was tested *in vitro* according to RNase III cleavage. This enzyme is known to recognize specifically and to cleave the processing stem (double helix) at position –7 in respect of 5' terminus of 23S rRNA. RNase III cleavage at position –7 was detected by reverse transcriptase primer extension. In order to test the specificity of the processing stem formation, the mutation V20 was introduced into 5' spacer sequence. This mutation blocks the formation of the double helix in the pre-23S rRNA (Liiv and Remme, unpublished).

The results of RNase III cleavage performed by the Remme group clearly showed that the processing stem is specifically formed in the chimerical constructs where both 5' and 3' spacers were present. Therefore, it is evident that the processing stem of 23S rRNA can be used to prepare pseudo-circular mRNA.

These complementary sequences used in the constructs from the Remme group to flank GFP sequence should result in stem formation and pseudo-circulation of mRNA molecule. This, in turn, should prolong the half-life of mRNA and, thus, its recycling by the ribosomes should possibly result in increase of the protein yield. This is the idea for construct named pXB-GFPΔA5. Another construct, made as a control and named pXB-GFPΔV20 carries a 20-nucleotide insert that disrupts complementarity between flanking sequences and does not produce a pseudo-circulated mRNA molecule, which, in turn, should be readily degraded and, as a result, should have poor or zero level of GFP synthesis.

Unexpectedly, both constructs yielded about the same total level of GFPcyc3 (Figure 3.6.1-3B), which is about twice less as the synthesis of GFP from our control plasmid, pIVEX2.2-GFPcyc3 (Figure 3.6.1-3A).

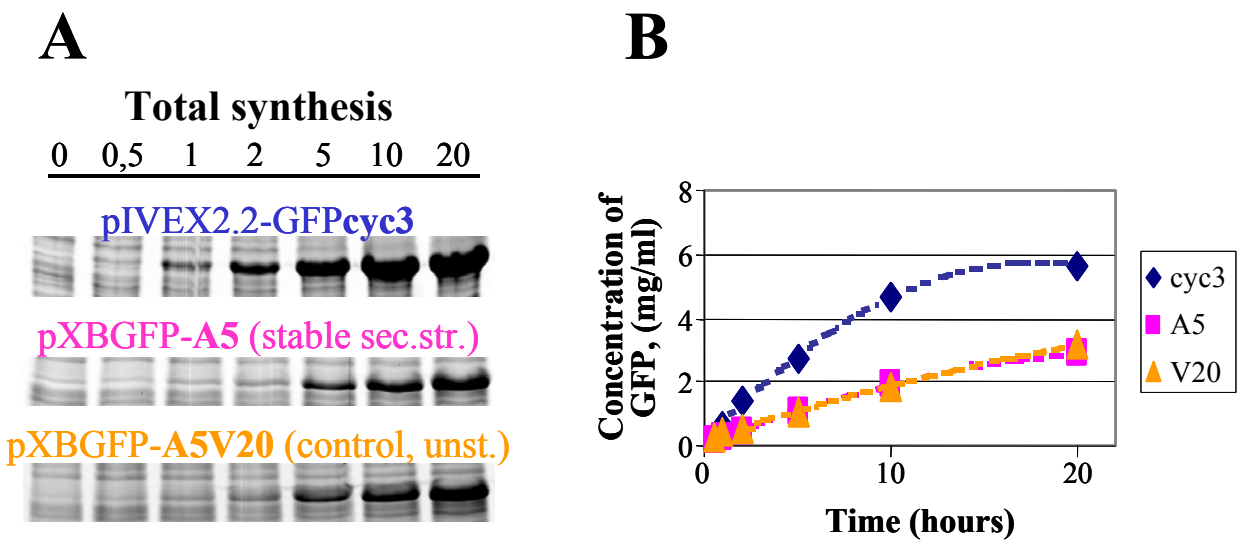


Figure 3.6.1-3 Synthesis of GFPcyc3 from pIVEX2.2, pXBGFPΔA5, and pXBGFPΔA5-V20 constructs in a fed large-scale reaction, SDS gel. (**A**) GFPcyc3 total synthesis analysed in SDS gel; (**B**) chart comparison of the levels of GFP total synthesis.

Obviously, the half-life of mRNA is not a limiting factor in our coupled transcription-translation system. Therefore, we decided to study the fate of mRNA during protein synthesis in the next experiment.

3.6.2 Fate of the transcribed mRNA during the protein synthesis

In order to identify what is the fate of mRNA molecule we measured in parallel the mRNA synthesized in the system *via* Northern blot, the total yield of GFP protein and its active fraction *via* polyacrylamide gel analysis. First, the study of the behaviour of mRNA transcripts (by Northern blot analysis), and GFP synthesis during incubation for 25 hours in the large-scale reaction system (RTS 500) from one and the same tube revealed that after one hour the mRNA amount decreases (Figure 3.6.2-1, lane corresponding to the second and fifth hours), whereas the rate of GFP synthesis is maximal until the seventh hour.

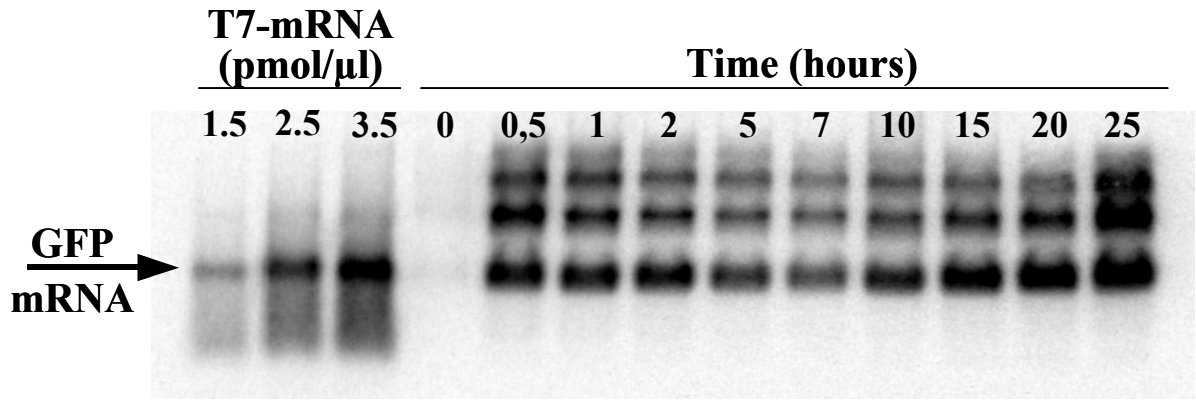


Figure 3.6.2-1 Northern blot describing levels of mRNA synthesis in the course of 25 hours of incubation for RTS 500 large-scale reaction. On the very left first three lanes represent pure mRNA transcripts of GFPcyc3 gene by T7 RNA polymerase. The other lanes are reaction samples from one and the same reaction vial taken at different time point, indicated above, (hours). The two bands above the GFP mRNA are products of incomplete transcription termination.

When the amount of GFP saturates, the rate of mRNA synthesis recovers, which is after the seventh hour of incubation (Figure 3.6.2-2) clearly saying that NTPs are not limiting the reaction of transcription. These results confirm that synthesis of mRNAs or their half-lives are not limiting factors for protein synthesis in our system.

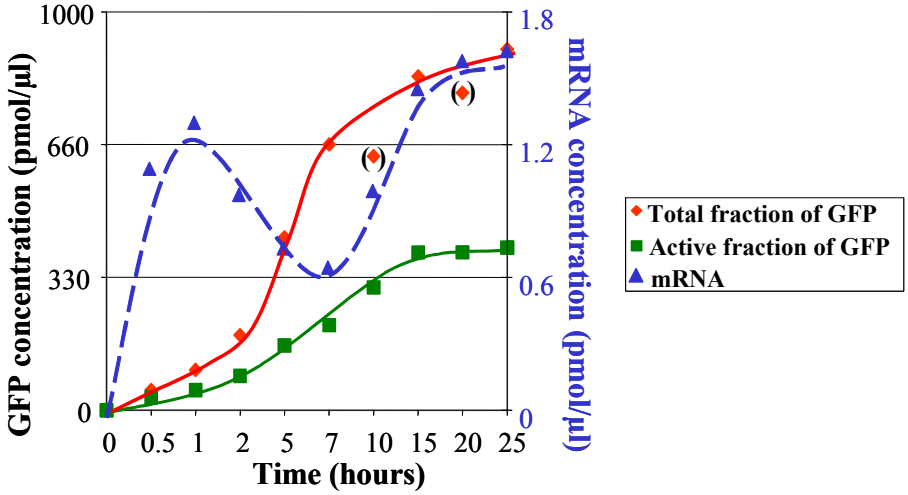


Figure 3.6.2-2 Tendency of mRNA and protein synthesis (total and active) in the RTS 500 large-scale reaction. Closed **blue** triangles represent behaviour of mRNA that is rapidly synthesised during first few hours, in contrast to the protein synthesis (both, total (**red**) and active (**green**) fractions). mRNA notably reduces during the time when protein synthesis is actively going (time points between two and seven hours), and goes up again as soon as protein synthesis level stabilises (after seventh hour).

3.6.3 Prevention of amino-acid shortage during the protein synthesis

As it was discussed in the Introduction already, Kim and Swartz reported that some of the twenty amino acids, namely arginine, cysteine, and tryptophan were metabolised in an *in vitro* cell-free system even in the absence of protein synthesis and therefore the shortage of these amino acids could impair protein synthesis (Kim and Swartz, 2000).

This encouraged us to add a mixture all twenty amino acids after five hours of incubation. The following Figure 3.6.3-1 presents the outcome of an amino acid addition that was analysed for total GFPcyc3 synthesis, and where samples were withdrawn before amino-acid addition and after (arrow, red line and squares), and compared to those of the reaction where no addition was made (blue line and diamonds). Indeed, a burst of protein synthesis is seen after the amino acids addition.

The next experiment was performed to observe the influence of the addition of the mixture of amino acids on total protein synthesis and active fraction of GFPcyc3 at different temperatures.

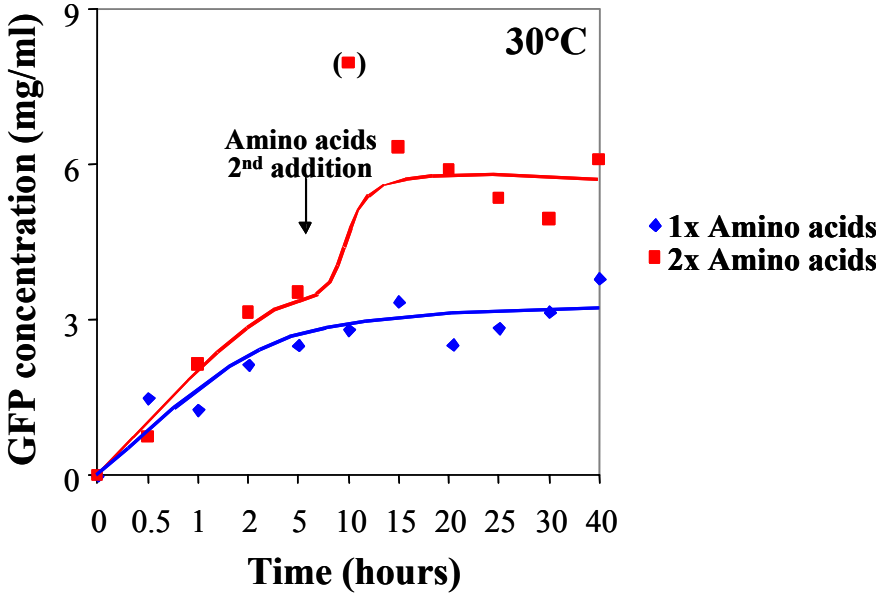


Figure 3.6.3-1 The effect of the second addition of the amino acids mix. After five hours of incubation a mixture of twenty amino acids was added (arrow).

Figure 3.6.3-2 presents the (i) comparison of GFP synthesis with and without a second addition of amino acids; this was performed at 30°C and 20°C; (ii) relation between total protein synthesis (blue bars) and level of active fraction (green bars) within one the mentioned conditions is also shown. This assay shows that indeed both decrease of incubation temperature to 20°C and a second addition of the mixture of all twenty amino acids improve the yield for about 20%. Furthermore, the GFP synthesized was 100% active!

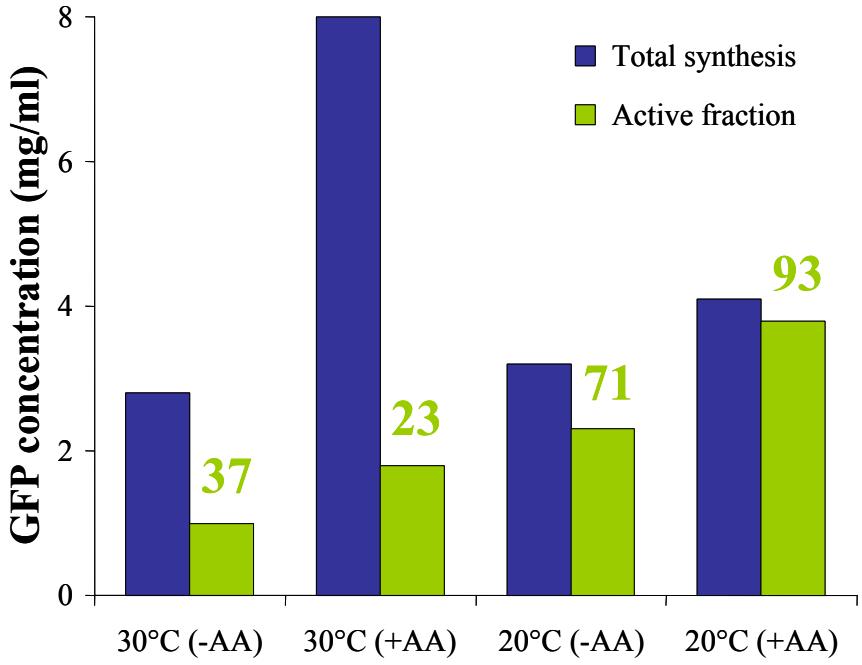


Figure 3.6.3-2 The effect of the second addition of the amino acids mix and decrease in temperature for incubation. A mixture of twenty amino acids was added as described (assigned with plus in brackets); large **green** numbers above indicate percentage of GFPcyc3 active fraction.

3.7 An attempt to improve the expression of eukaryotic genes in *E. coli* system

3.7.1 Addition of the tRNA fraction of the Rosetta ™ strain (Novagen)

The Rosetta™ (DE3) strain is derived from *lacZY* mutant of BL21 (*lon* and *ompT* proteases deletion), to enable precise control of expression levels by

adjusting the concentration of IPTG. This strain was designed to alleviate codon bias when expressing proteins of eukaryotic origin in *E. coli*.

When the mRNA of heterogeneous genes is over-expressed in *E. coli*, differences in codon usage can impede translation due to the demand for one or more tRNAs that may be rare or lacking in the population (Baca and Hol, 2000; Goldman *et al.*, 1995; Kane, 1995). It has been well established that insufficient tRNA pools can lead to translational stalling, premature translation termination, translation frame-shifting and amino acid misincorporation (Kurland and Gallant, 1996).

Six tRNAs were introduced into the commercially available Rosetta strain that carries the pRARE plasmid with the corresponding tRNA genes: *proL* tRNA, *leuW* tRNA, *argW* tRNA, *glyT* tRNA, *argU* tRNA, *ileX* tRNA. Under the IPTG induction these tRNA species are overexpressed in bacterial cell and their pool could be normalized to that of eukaryotes. The plasmid carries a chloramphenicol resistance (Figure 3.7.1-1).

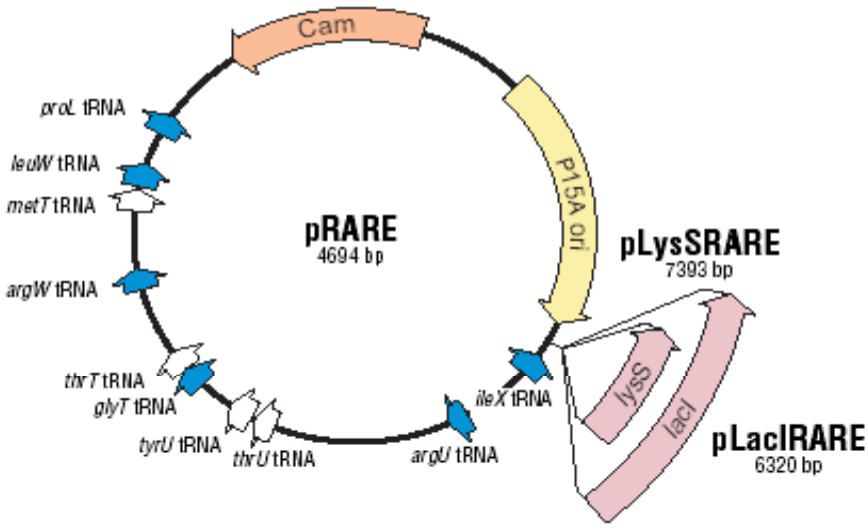


Figure 3.7.1-1 Map of pRARE plasmid family The basic structure of pRARE is indicated. pLysSRARE and pLacIRARE contain the genes encoding T7 lysozyme (LysS) and *lac* repressor (*lacI*), respectively. Also indicated are chloramphenicol resistance gene (Cam), replicon (P15A ori) and tRNA genes. tRNA genes corresponding to rare codons in *E. coli* are indicated in blue.

We isolated tRNAs from the Rosetta strain on a analytical scale after induction and checked two of them, which were predicted to be over-expressed. We selected $tRNA^{Leu}$ and $tRNA^{Ile}$ against the non-overexpressed $tRNA^{Val}$ used as a control and performed a charging assay (amino acylation) with either [^{14}C]-Leu or [^{14}C]-Ile *versus* [^{3}H]-Val amino acids. The ratio (cpm [^{14}C]/cpm [^{3}H]) derived from the Rosetta cells (overexpression) *versus* the ratio derived from wild-type cells reflects the degree of tRNA overexpression. With this assay we observed that the ratio $tRNA^{Leu}/tRNA^{Val}$ in the Rosetta strain is about twice as high as in wild type *E. coli* (= $tRNA^{bulk}$; 3.7 *versus* 1.7) indicating that $tRNA^{Leu}$ is twice as much abundant in the Rosetta strain as in the wild-type *E. coli* (Table 3.7.1-1).

Table 3.7.1-1 Charging of Rosetta™ tRNA with either [^{14}C]-leucine or [^{14}C]-isoleucine versus $tRNA^{bulk}$ with [^{3}H]-valine.

	[^{14}C]-Leu / [^{3}H]-Val	[^{14}C]-Ile / [^{3}H]-Val
tRNAs (Rosetta)	3.7x	0.85x
$tRNA^{bulk}$ (*E. coli*)	1.7x	0.93x

In contrast, the $tRNA^{Ile}/tRNA^{Val}$ ratios are about the same (0.85 and 0.93) indicating that Ile was not overexpressed in contrast to the claim of the plasmid producer Novagen.

For this reason we have a hesitation for overall usage of this strain for expression of eukaryotic genes, whether *in vivo* or *in vitro*.

3.8 Investigation of the fragmentation of a given protein

An object for this study was *E. coli* translational elongation factor EF-Tu, where Tu stands for temperature unstable property of this protein, because it denatures easily during and after purification. Furthermore, when it is translated *in vitro*, in addition to the mature protein band a few fragments are observed (Figure 3.8-1B, fragments assigned as 1 and 2).

As a control for a protein that is not fragmented, we used another *E. coli* translational elongation factor EF-Ts, for it is relatively stable towards temperature increase. Both factors were cloned into pET23c(+) vector, that has a T7 promoter, using *Nde* I and *Xho* I restriction enzymes (Materials and Methods, Figure 2.6.1-2 for vector map, page 32; part 2.6.3 for cloning strategy, page 35).

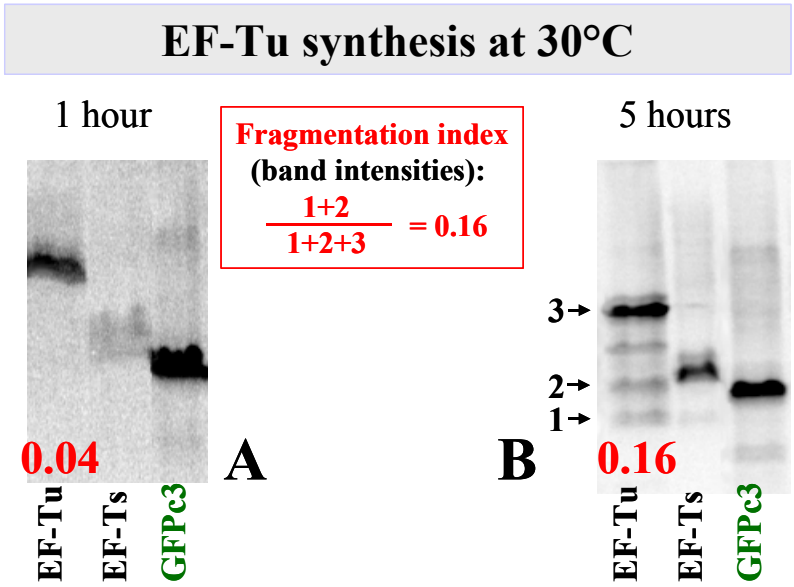


Figure 3.8-1 Synthesis of EF-Tu at 30°C with [^{35}S]-Met incorporation. Samples analysed after 1 hour (**A**) and 5 hours (**B**) of the reaction run start. Number **3** corresponds to the mature EF-Tu protein, number **2** and **1** correspond to the fragments of EF-Tu on the **EF-Tu** panel. Arrows indicate bands taken for calculation of the fragmentation index (F.I.) expressed by a formula shown in red frame. **Red** numbers represent the calculated F.I.

As a control for the efficiency of expression level, we used standard GFPcyc3 plasmid DNA. Expression of all three genes was carried out in RTS 100 small-scale reactions, incorporating [^{35}S]-Met, in order to monitor fragmentation levels. When expression reactions were incubated at 30°C, samples for SDS-PAAG analyses were withdrawn for two time points, after one and five hours of the reaction incubation. We observed that at the early stage of incubation (1 h) the fragmentation level of EF-Tu was relatively low (Figure 3.8-1A), and after the reaction stop four distinct bands were observed, one of which corresponded to the EF-Tu full-length protein and another three were fragments of the EF-Tu (Figure 3.8-1B). We selected two lower bands, indicated with numbers 1 and 2 on Figure 3.8-1B, and the band corresponding to mature

EF-Tu was indicated with number 3, to calculate the *fragmentation index* – a measure of the level of fragmentation for this protein (Figure 3.8-1B) based on band intensities. The yield of GFPcyc3 synthesis was used as a control for a mainly non-fragmented protein.

If temperature *via*, *e.g.* proteases, were the main reason for the fragmentation of EF-Tu, then reduction to 20°C should result in the reduction of fragmentation level (lessened number of fragments and their amount), as well as an increase to 37°C would result in the increase of fragmentation level (more fragments in number and amount). We therefore ran the next expression reactions at different incubation temperatures. We observed that level of fragmentation was lower at the beginning of each reaction (Figure 3.8-2, a 2 hours (**A**) and a 40 min (**B**) panels). Another thing that we observed was that at 37°C EF-Tu full-length protein synthesis increased (relative to GFPcyc3 synthesis) compared to that at 20°C, as well as that the number of fragments and amount of each had increased, too. In order to understand relative EF-Tu full-length protein synthesis level we compared its band intensity to that of GFPcyc3, at different incubation temperatures.

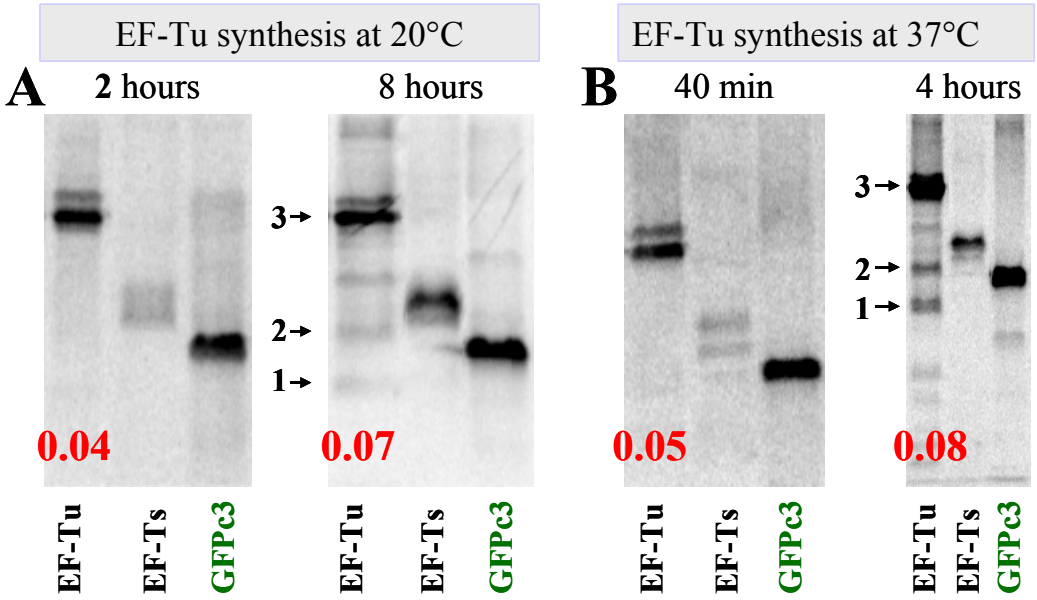


Figure 3.8-2 Investigating the temperature affect on the synthesis of EF-Tu. Synthesis of elongation factors and GFP: (**A**) at 20°C during eight hours, and (**B**) at 37°C during four hours. **Red** numbers indicate the fragmentation index.

The lower the number the lower the fragmentation is (Figure 3.8-2 and Table 3.8-1, F.I. grey row). Production of EF-Tu is better at 37°C, though for both, 20°C and 37°C EF-Tu full-length protein yield increases twice as much at the end of incubation reaction (Table 3.8-1, last row EF-Tu/GFP, compare values within green and red columns). The level of fragmentation is relatively the same at any incubation temperature tested here (Table 3.8-1, EF-Tu column, compare values of the grey row). We conclude that the reason for EF-Tu fragmentation is different and not temperature-produced or dependent.

Table 3.8-1 Fragmentation index of EF-Tu at different temperatures and EF-Tu over GFPcyc3 synthesis.

NAME	20°C		30°C		37°C	
	2h	8h	1h	5h	40 min	4h
EF-Tu, mature, pxl	447672	1361372	170041	3420355	157453	1624355
Frg. #1, pxl	7300	33713	2270	178573	3480	67817
Frg. #2, pxl	9298	75517	4595	371541	5106	64100
F.I.	0.04	0.07	0.04	0.16	0.05	0.08
EF-Tu/GFP	0.58	1.17	0.38	0.76	0.65	2.29

Data obtained from relative band intensities processed in ImageQuant and Excel software. Fragmentation index (F.I.) was calculated according to following: (Frg. #1 + Frg. #2)/(Frg. #1 + Frg. #2 + EF-Tu), all values in pixels for definite area.

The next investigation addressed the question, whether the presence of amino-acylated $tRNA^{Phe}$ ($tRNA^{Phe}$:Phe mixture) given in excess over all other tRNAs would stimulate the ternary complex formation (EF-Tu•Phe-$tRNA^{Phe}$•GTP) and then prevent degradation of EF-Tu. Here, we assume that free EF-Tu is a good target for proteases, unlike the ternary complex. To perform this assay we needed to know the relative amount of EF-Tu synthesised, so that we could give a $tRNA^{Phe}$:Phe mixture in excess, about three times as much as EF-Tu is expected to be synthesised during five hours of incubation in RTS 100 small-scale reaction. The amount of Phe amino acid should also be in excess over $tRNA^{Phe}$, about four times. We know that in the RTS100 small-scale reaction the yield of GFPcyc3 is about 500 µg/ml after five hours of incubation. And from previous experiments, when expressed in RTS 100 small-scale

reaction at 30°C for five hours, we estimated that EF-Tu synthesis relative to GFPcyc3 is ~50% after one hour of incubation and ~60% after five hours of incubation (Table 3.8-1, EF-Tu column, see values within orange rows). This means that ~ 250 µg/ml of EF-Tu is synthesised. The *tufB* gene is 1,185 nt long, resulting in a 395 amino acids protein, which is about 44.40 kDa in molecular weight. We assume that the upper range of EF-Tu synthesis in the system is 440 µg/ml = 440 mg/l, which is 10 pmol/µl and thus the amount of $tRNA^{Phe}$ we needed is 30 pmol. This, in turn, has to be added into reaction mix (25 µl final volume) in a volume not larger than 2.5 µl, *i.e.*, 750 pmol of $tRNA^{Phe}$ in 2.5 µl, and, in turn, the amount of Phe amino acid has to be 3000 pmol in the reaction mix.

The following set of RTS100 small-scale reactions was planned: EF-Tu or GFPcyc3 gene alone, in the absence of $tRNA^{Phe}$:Phe mixture; EF-Tu or GFPcyc3 gene in the presence of $tRNA^{Phe}$:Phe mixture from the beginning of reaction incubation; and EF-Tu of GFPcyc3 gene in the presence of $tRNA^{Phe}$:Phe mixture added one hour later, after the reaction incubation started already. Samples from each reaction tube were analysed after one and five hours of reaction incubation. Synthesised proteins were marked with [^{35}S]-Met, same as in previous experiments. If the EF-Tu protein in the form of ternary complex is more stable towards fragmentation, then an extra addition of the amino acylated tRNA would stimulate the ternary complex formation and, as a result, reduction of fragmentation level should follow. Because previous results on the influence of different temperatures of incubation showed little affect on the EF-Tu fragmentation, next reactions were performed at 30°C. Here, the three-fold excess of $tRNA^{Phe}$ and the phenylalanine amino acid in ratio 1:4 mixture over expected full-length EF-Tu protein synthesised, was added either from the beginning of reaction, or one hour later after the incubation had already started. According to the estimation (Table 3.8-2) of the fragmentation index and overall analysis of the full-length EF-Tu protein synthesis, presence of $tRNA^{Phe}$:Phe

mixture had a little positive effect on the reduction of EF-Tu fragmentation when was added one hour after incubation of the reaction had started.

Table 3.8-2 Influence of the addition of $tRNA^{Phe}$:Phe mixture to the protein synthesis reaction.

Name	Band int., 1 hour	EF-Tu / GFPcyc3	F.I.	Band int., 5 hour	EF-Tu / GFPcyc3	F.I.
EF-Tu	3835508.0	0.7	0.31	2832487.7	0.47	0.44
EF-Tu +, f.t.b.	1793103.4	2.9	0.39	1228826.4	0.72	0.53
EF-Tu +, a.1.h.	3568024.8	1.3	0.30	2731133.8	0.94	0.47

Data obtained from relative band intensities processed in ImageQuant and Excel software. F.I. – fragmentation index; + - stand for addition of $tRNA^{Phe}$:Phe mixture; f.t.b. – from the beginning of reaction; a.1.h. – after one hour of the reaction start. Band intensities are given in pixels.

Along with results discussed above we now consider, whether fragmentation of the EF-Tu protein might be due to activity of proteases present in the S30 lysate. We selected three protease inhibitor mixes, described below. The protease inhibitory (P.I.) mixes B and HP are from Serva and a Cocktail mix is from Sigma, respectively. The P.I. mix B is aimed to protect proteins in cell extracts isolated from prokaryotes, the P.I. mix HP to protect recombinant His-tagged proteins, both are directed towards proteolytic activity of aspartate proteases, metallo proteases and cysteine proteases as well as serine proteases; and the P.I. Cocktail mix is optimised for bacterial cell use and is a mixture of protease inhibitors with broad specificity for the inhibition of serine, cysteine, aspartic, metalloproteases, and aminopeptidases, too.

As in previous experiments, these mixtures were added from the beginning of reaction incubation, or one hour later, after the reaction had started already. Samples for SDS-PAAG analyses were withdrawn after one and five hours of incubation (Figure 3.8-3). In the case, when mixtures were added one hour later, the samples were taken just before the addition of protease inhibitor mixes (Figure 3.8-4A).

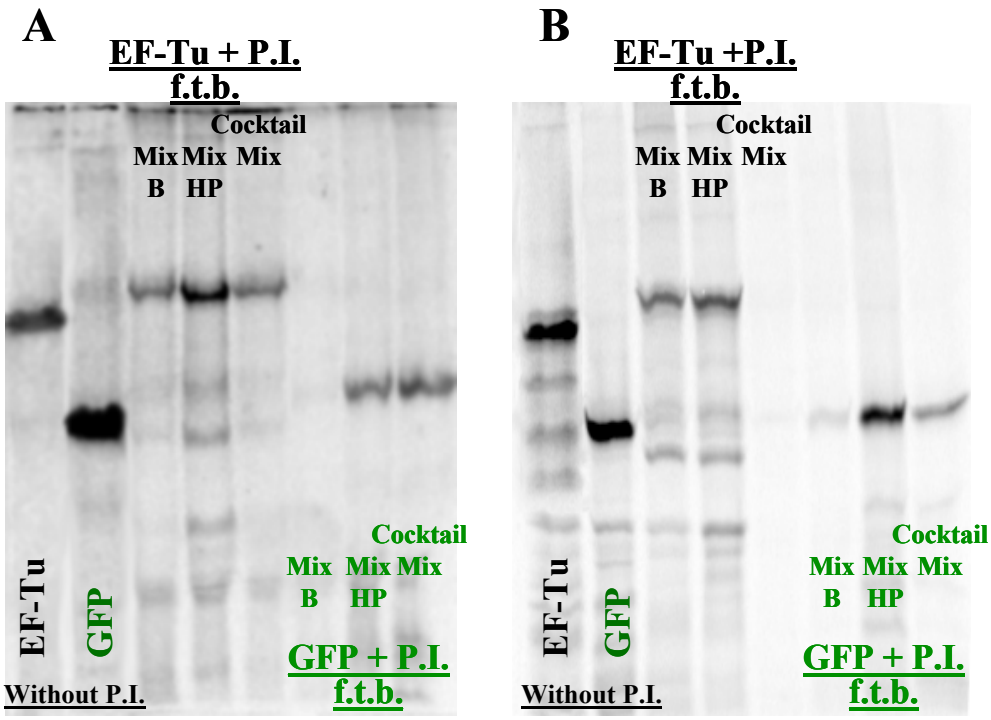


Figure 3.8-3 Investigating the effects of the Protease Inhibitor (P.I.) mix on the synthesis of EF-Tu. P.I. mix was added from the beginning (f.t.b.) of reaction. (**A**) samples analysed after one hour of reaction run. (**B**) Reactions were stopped after five hours of incubation.

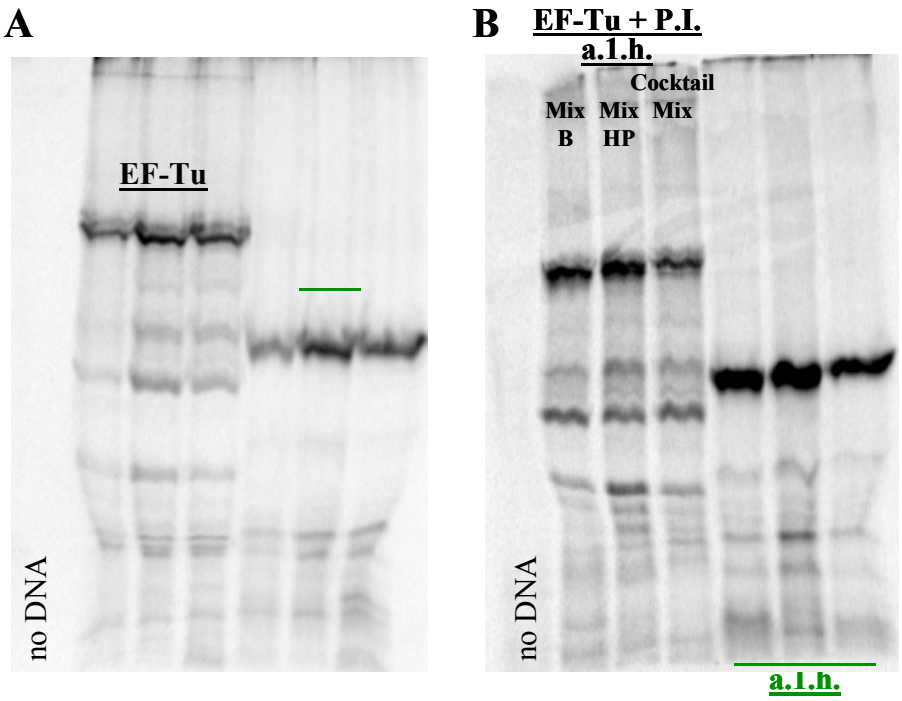


Figure 3.8-4 Investigating the effects of Protease Inhibitor (P.I.) mix on the synthesis of EF-Tu, when added after one hour (a.1.h.) of the reaction incubation start. (**A**) Samples analysed after one hour of reaction run, before P.I. mix was added. (**B**) Reactions were stopped after five hours of incubation.

The results allow one general conclusion: it looks like the addition of the P.I. mixes abolished not only the fragments of EF-Tu, but also the synthesis of a full-length protein (Figures 3.8-3 and 3.8-4, compare lanes in the absence of P.I. and lanes where any of P.I. mixes was added). Obviously, components of these mixes have negative effects on the process of protein synthesis itself and these mixes are more recommended to be used for protein storage, when isolated from cells. Though, considering that EF-Tu is a His-tagged protein the HP mix of protease inhibitors is more appropriate for use. This mix also did not impair the overall synthesis of both, EF-Tu and GFPcyc3, though reduced the yield of full-length product compared to the samples from reactions in the absence of any P.I. mix.

3.9 Division of the *E. coli* lysate into fractions that are simple in controlling

Previous studies used total cell lysates. Ganoza *et al.*, who used ribosome as an "affinity matrix", based on the fact that many of the proteins, which are required for each step of protein synthesis, bind to 70S ribosomal particles, described another possibility. By use of different concentrations of Mg^{2+} and varying concentration of NH_4^+ and K^+ ions it was possible to selectively elute sets of these proteins that yielded in: each of initiation (IF1, IF2, IF3) and elongation (EF-Tu and EF-G) factors, together with proteins that – according to the Ganoza group – are required to reconstruct synthesis, EF-P, W, and "rescue." Authors found that ribosomal eluate are also enriched with each aminoacyl-tRNA synthetase. The relative amounts of EF-Ts and EF-G are higher in the ribosome-free (also known as S100) supernatant (Ganoza *et al.*, 1996).

In the course of this study we aimed to analyse protein expression with a fractionated *E. coli* lysate system. Fractionation of the S30-lysate is outlined in Figure 3.9-1.

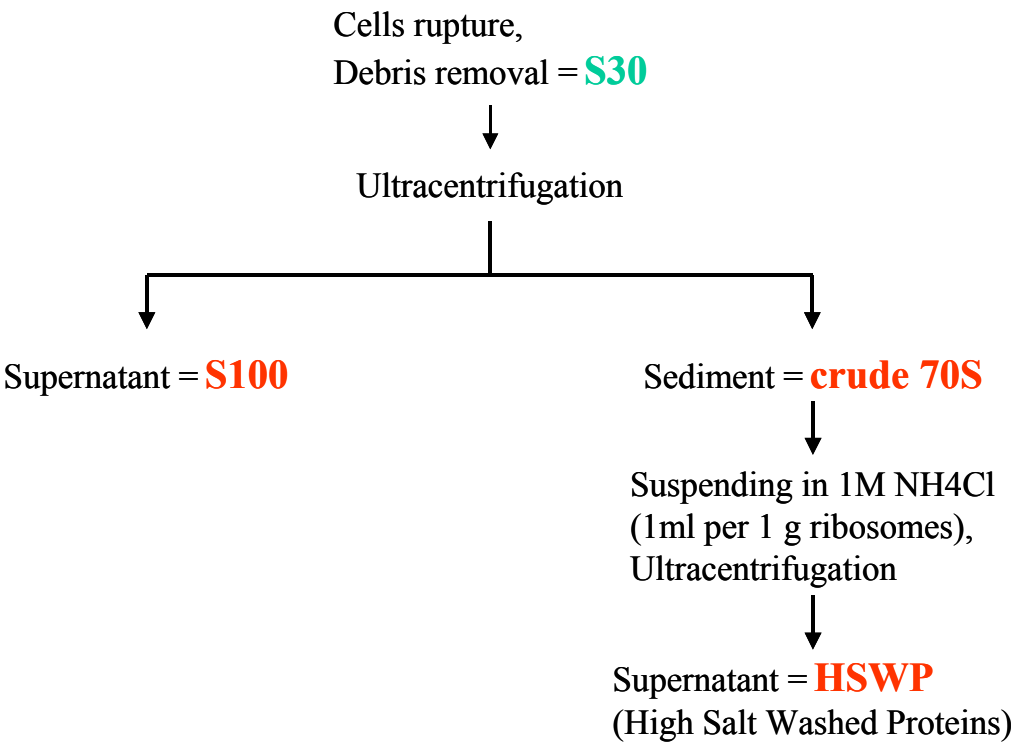


Figure 3.9-1 Schematic representation of the fragmentation procedure. Fractions used for further analysis are indicated **red**.

The idea was that with three fractions instead of one compound S30 preparation the system possibly could be better monitored and controlled. We examined the effects of the fractions in the batch system with GFPcyc3 expression. The results in Figure 3.9-2A indicate that the best yield is obtained in the presence of all three fractions, including crude 70S, high salt washed proteins and S100, which indicates that they all contain essential components for the protein synthesis. In Figure 3.9-2B we present the summary of the temperature effect in such a fractionated system.

According to the results with fractionated system, GFP synthesis in batch system with S30 extract or in the presence of HSWP (high-salt washed proteins, see Figure 3.9-1) together with crude 70S was similar at the incubation temperature of 30°C (lanes 5 and 2, respectively), and about five times higher in the presence of all fractions when incubated at 20°C (lanes 4 and 5).

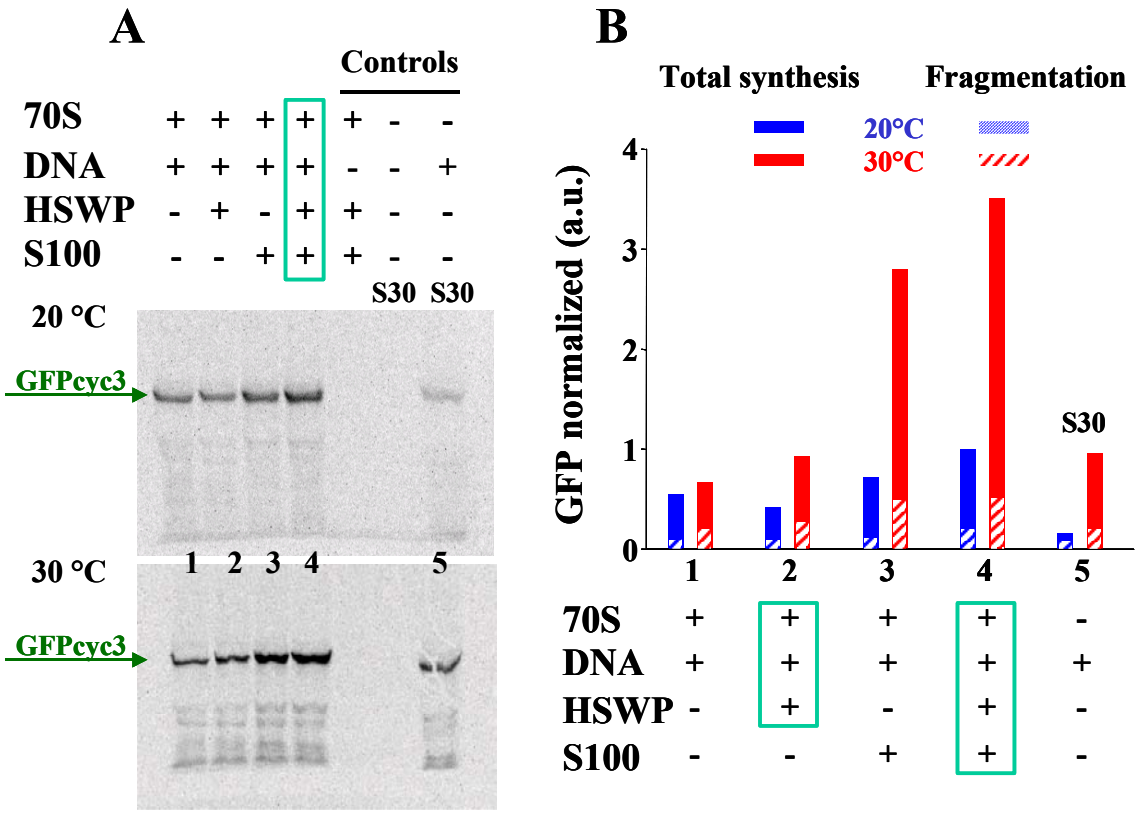


Figure 3.9-2 Synthesis of 2 GFPcyc3 in batch system (here, 55 µl reaction volume) with [^{35}S]-Met incorporation. (**A**) SDS PAA gel analyses, in the upper panel components and their combinations are indicated. (**B**) Summary of SDS gel results, including fragmentation level that was calculated as total field of a lane without the GFP band *vs.* the GFP band alone. Panel at the bottom corresponds to components and their combination. The boxed-in components indicate, that addition of S100 greatly increases total GFPcyc3 synthesis.

Figure 3.9-2B indicates clearly that addition of S100 greatly increases total GFPcyc3 synthesis, while in the absence of it the synthesis is as little as when only non-fragmented S30-extract is present.

Not only non-ribosomal proteins that are found on crude 70S are important for efficient protein synthesis *in vitro*, but also some additional proteins/factors from the ribosome-free supernatant (S100), among which EF-Ts and EF-G are probably involved (Ganoza *et al.*, 1996).

Chapter 4

DISCUSSION

4.1 Some comments to the optimisation procedures

For the *in vitro* protein synthesis reaction based on the utilization of a cell-free extract (often termed as S30), not only the initial cell strain and growth media, but also the method for isolation of a cell lysate is of great value.

In our days modern instruments provide an advantage to obtain cell lysate of high yields in cell rupture with minimal processing (one pass) and easy recovery for large (up to 1 litre) and small (14 ml with 12 ml recovery) volumes of cell suspensions. Our choice felt for the homogeniser "Microfluidizer M-110L" from Microfluidics, which is a fast and efficient machine, easy to handle and use. Fast rapture is important while producing cell lysate, for the longer is this lysate in a liquid state (if not used directly), the faster do degrade its main components, it looses the so necessary activity for protein synthesis due to fast RNAs degradation (tRNAs and rRNAs). Also it is important to work with cell lysate in a cold room, where temperature should not exceed +4°C – simply keep the suspension and later lysate on ice (water). Another remark is concerning the concentration of the cell lysate – it has to be high for better activity of the later. It means, that the density of ribosomes should be high for the best protein synthesis. Thorough suspending of cell pellets in 1 ml/g (wet weight) of Tico buffer is sufficient enough to obtain high A_{260} measurements of S30 lysate.

Concerning the cell strain, it was expected that strains deficient in major RNases would allow for obtaining of high protein yields, since mRNA molecules are the least stable ones of all RNAs. In the case of an *in vitro* coupled transcription-translation system the level of mRNA is relatively stable and, in fact, does not affect the protein yield (Figure 3.6.2-1, page 68). With respect to this, our selection was an *E. coli* BL21 strain, which is deficient in the protease genes *lon* and *ompT*. This strain was used for preparation of cell lysate and fractionation analyses. Besides this strain we selected different protease deficient strains from ATCC and literature reports. Also, a Rosetta™ strain from

Novagen, which is enriched with some tRNAs isoacceptors species was of great interest to us for reasons, discussed bellow.

We used a batch system to speed up the optimisation procedure using small volumes down to 10 µl. This kind of an *in vitro* coupled transcription-translation system generally utilizes high-energy phosphate compounds (like PEP) to regenerate the adenosine triphosphate (ATP), necessary to drive protein synthesis, as well as adequate substrates (particularly nucleoside triphosphates and amino acids). This system also requires buffering, pH and proper ionic conditions, as well as catalyst stability and avoidance of inhibitory by-products. Though this system attempts to mimic "*in vivo* near" conditions, it contains some unnatural components, such as pH buffers and polyethylene glycol (PEG), and is far away from physiological environment of the cytoplasm. We optimised the final concentrations of main components of the batch cell-free system for *in vitro* transcription-translation reactions. This was done in order to ease up calculation of numbers for stock solutions and monitoring of the system components activity. It is easy to tackle which of the main mixes or stocks had lost the activity for protein synthesis reaction due to one of the following reasons: presence of RNases, expiry of one of the "energy" sources for the protein synthesis, *e.g.* ATP, ionic conditions changed, *e.g.* Mg^{2+} concentration, as well as T7 RNA polymerase inactivation.

Recent literature data about a promising *in vitro* translation system, which is designed to mimic the cytoplasm (thus named Cytomim), reports that most unnatural components like PEG were removed, and ionic composition of the reaction was altered to more closely replicate that of the cytoplasm (Jewett and Swartz, 2004a).

An essential point of the coupled transcription/translation system was to establish reliable quality criteria for the synthesis of the reporter product green fluorescent protein (GFP). With use of this protein we established a reliable method for judgement of GFP expression, based on the electrophoretic analysis in denaturing and native PAA gel conditions.

We measure relative band intensities of GFPcyc3 in native and SDS (denaturing) protein gels from one and the same reaction vial and correlate to those bands, which contain known amounts of GFP. Thus, we identify total yield and native fraction in it, which is for GFP in the range of 40-60%.

In the case we would like to learn more on the fragmentation level of a given protein we performed a reaction with [^{35}S]-Met incorporation. Operating with exact number of Met-amino acids in a single protein chain we can estimate the relative number of protein molecules synthesised per single ribosome (assuming that only 30% are active), the concentration of which is known from A_{260} measurements. According to our estimation about 7 molecules of GFPcyc3 were synthesised by one ribosome in the reaction mix during 120 min incubation at 37°C.

Table 4.2-1 The TCA precipitation result for GFPcyc3 synthesis

Sample	cpm	Average	Minus background	Met incorp. in GFP prot. synth.	Amount of GFP, pmol	GFP made by one ribosome in 120 min	~30% of 70S participate in GFP synth.
1. DNA -	1271						
2. DNA -	1661	1466	0				
3. DNA +	3811						
4. DNA +	4952	4382	2916	295.5	49.3	2	**6.6**

The specific activity of [^{35}S]-Met was about 10 cpm/pmol. GFP contains 6 methionine residues.

4.2 Synchronising the reactions of transcription and translation

One of the problems in cell-free systems is the uncoupling of the naturally coupled processes of transcription and translation. This happens due to utilization of the bacteriophage T7 RNA polymerase (RNAP) instead of *E. coli* RNAP T7 RNAP is 5-6 times faster that *E. coli* (Iost *et al.*, 1992), and this affects negatively not only protein synthesis initiation and elongation steps, but also *in vivo* assembly of the *E. coli* ribosomes, when T7 RNAP transcripts of rRNA are used. The negative effect was the observation that only a tiny fraction,

within a few percent of the transcribed product (either mRNA or rRNA), was used for translation and ribosome assembly, respectively (Iost and Dreyfus, 1995; Lewicki *et al.*, 1993). In *E. coli* cells processes of transcription and translation are tightly coupled. While mRNA is synthesised *E. coli* ribosomes initiate translation on the nascent chain of mRNA: the *E. coli* RNAP proceeds with a speed of ~60 nucleotides per second, and ribosomes proceed with a speed of ~20 amino acids per second (Bremer and Dennis, 1996). It follows that the first ribosome pursues immediately the transcription, leaving no room for a significant gap in-between. Therefore, the nascent mRNA chain cannot form a secondary structure that could hinder or block elongation (Iost and Dreyfus, 1995; Iost *et al.*, 1992). On the other hand, ribosomes protect mRNA from endonucleases that usually initiate the process of degradation from the 5' into 3' end direction. We tested two approaches, (i) applying slow mutants of T7 polymerases, and (ii) lowering the incubation temperature. The latter step seemed to be a promising trial, since lowering the growth temperature to 25°C improved dramatically from 15 to 60% the rRNA fraction used for ribosomal assembly, which indicated that the rate of T7 RNAP goes down faster than the assembly rate (Lewicki *et al.*, 1993).

Mark Dreyfus (Ecole Normale Superieure – CNRS, Paris) kindly supplied us with different T7 RNAP mutants, which were either slower than the wild type (WT) T7 RNAP, thus reaching the rate of *E. coli* RNAP, or were more processive than the WT T7 RNAP (Bonner *et al.*, 1994; Makarova *et al.*, 1995). From these mutants, according to our results based on the judgement of GFP expression, we selected a double-mutant I810N/P266L, which has a reduced rate but an improved processivity. Translation from the transcripts produced by this mutant resulted in almost 100% active GFP production (although the yield was reduced to 10%, see Figure 3.5.1-2C, page 61), indicating that not only translation, but also folding was affected positively in this case. This finding and literature reports were furthermore supported by results described in the following chapter.

In another set of experiments we also achieved almost 100% active GFP output just by reducing temperature of incubation from 37°C and 30°C to 25°C and 20°C, though the overall protein production was reduced two to three times (Figure 3.5.2-1A, page 63). Thus, in further attempts to optimise cell-free protein synthesis this approach was combined with others in order to increase the output of fully synthesised and fully active proteins of interest, as will be discussed below.

4.3 Trials to increase the yield of expressed proteins

The turnover of mRNA is defined in terms of the mRNA half-life, which is the time it takes for 50% of the mRNA molecules in the cell to be degraded. In general, half-lives of mRNA are shortest in species with the shortest replication cycles, also compared to the half-life of ribosomal RNAs (rRNAs) within one organism. The half-life of a bacterial mRNA measures on average one (most labile mRNAs) to seven and a half minutes (ribosomal proteins mRNA, Mohanty and Kushner, 1999), with two to three minutes in average for most of the mRNAs in *E. coli* (Selinger *et al.*, 2003). It is essential that rapidly replicating organisms adapt quickly to changes in their environment by stabilising or destabilising certain mRNAs. Ribonuclease III (RNase III) is one of the mRNA-degrading enzymes in *E. coli*. It attacks duplex regions in mRNAs and in rRNA precursors (Ross, 2001).

Liiv, *et al.* (1996) found earlier that the major stability determinant of rRNA is the presence of a stable helix bracketing mature rRNA. The stability determinant of pre-23S rRNA was used to increase the stability of mRNA (Liiv *et al.*, 1996). In a trial to increase the stability of mRNA molecules by the stability determinants of the pre-23S rRNA, we used a construct that was kindly provided by Jaanus Remme. These stability determinants are flanking the GFP gene and upon transcription result in the pseudo-circularisation of an mRNA molecule, thus, in theory, prolonging the half-life of it. On the contrary, another construct with an insert of 20 nucleotides that would rupture the

complementarity of the flanking regions was used as a negative control for mRNA stability. However, the level of GFP expression from this two constructs was same and in comparison to our standard GFP expression construct these levels were two times less, indicating again, that mRNA stability is not a limiting factor in cell-free protein synthesis system.

These conclusions were supported by the results of an assay addressed to picture the fate of the transcribed mRNA (Figure 3.6.2-1, page 68). We observed a significant decrease of mRNA in the middle of incubation time, and after three to five hours the mRNA levels went up again, showing that the parallel reduction of protein synthesis was not caused by a depletion of the energy supply (NTPs).

Another possibility of an impaired synthesis could be shortage of amino acids during the incubation period. According to Kim and Swartz, shortage of the building block for protein synthesis could be observed for arginine, cysteine and tryptophan in an *in vitro* system even in the absence of protein synthesis (Kim and Swartz, 2000). Later on, Jewett and Swartz identified eight out of twenty amino acids, concentrations of which changed dramatically from the starting 2 mM. Aspartic acid increased to 20 mM during the reaction. Alanine increased to 4 mM, and then decreased to 1.5 mM. Glutamate, which was initially 160 mM, decreased as well, though the corresponding value was not quantified precisely. The following amino acids were depleted during the first hours of reaction: cysteine, serine, threonine, glutamine and asparagine (Jewett and Swartz, 2004b).

Our results indicated that energy resources of the system are sufficient for mRNA synthesis. A close study of the amounts of GFPcyc3 mRNA indicates a slight reduction of its concentration in the middle of reaction. This takes place exactly at the same time, when the process of protein synthesis enters its maximal synthesis rate. As soon as protein synthesis reaches saturation phase, synthesis of mRNA bursts again – an indication of the presence of sufficient amounts of NTPs (Figure 3.6.2-1, page 68). All together, these pieces fit well

into a puzzle that shortage in amino acids rather than in NTPs is a limiting factor for increase of protein synthesis. Indeed, when a mixture of twenty amino acids was added after the fifth hour of incubation (saturation is reached normally at this time), and reaction was incubated up to 40 hours, the total yield of GFPcyc3 increased twice. But still the level of active fraction was too low (<30%). This led to combination of two approaches with a very satisfying result, namely a combination of a decrease of the incubation temperature to 20°C and the addition of amino acids that allowed the gain of 4 mg/ml of 93% (~100%) of active GFP (Figure 3.6.3-2, page 70).

4.4 Trials and considerations to increase the expression of eukaryotic proteins in the bacterial *E. coli* system

The term codon usage means a "snap shot" of codon abundance within the mRNAs present at a given state. The codon abundance is roughly related to that of the corresponding tRNAs. Highly expressed genes are, for example, those, which products are directly used in translation. Lowly expressed genes are, for example, those for RNases, proteases, and down regulated genes that may change due to temperature or amino acid variation.

Data published on tRNA species abundance in different organisms were the basis for this research, as well as data from a single organism at different growth rates (*E. coli*; Dong *et al.*, 1996). In *E. coli* there are 86 genes coding for 46 different tRNA species. A tRNA species is defined as a tRNA with a unique anticodon, and their gene copy number ranges from four to one. The copy number of tRNA genes for codons frequently appearing in the reading frames of highly expressed genes is quite often four. Thus, gene dosage is involved in the regulation of tRNA concentrations (Ueda T. and K., 2001).

Dong *et al.* showed earlier that in *E. coli* the co-variation of tRNA abundance and codon usage is also growth rate-dependent (Dong *et al.*, 1996). They found a strikingly reduced codon usage in highly expressed genes *versus* lowly expressed genes, which were switched on under bad nutrient conditions. It

follows that many more codons are used under slow growth conditions. Accordingly, a corresponding biased distribution of tRNA abundance at various growth rates were observed, which could be roughly correlated with the values of codon frequencies in the mRNA pools calculated for bacteria growing at different rates. Another observation they made was a correlation of tRNA concentrations with the gene dose. The mechanisms that support the growth-rate dependent co-variation of tRNA abundance and codon usage are not known. They identified the fraction of tRNAs out of total tRNA (%) for each isoacceptor species in *E. coli* grown at low growth rates.

We took as a control the known *E. coli* codon usage in highly and lowly expressed genes (Figure 4.4-1; Bulmer, 1988): In case of some codons there is a sharp difference for their abundance within highly expressed genes, while those that show up to be rare in highly expressed genes turned out to be often in lowly expressed genes. For example, in case of codons coding for aspartic acid, GAC is found frequent and GAU rarely in highly expressed genes, whereas just the opposite is true in lowly expressed genes. The same is valid for codons coding for Tyr, His, Asn and Phe. In cases of some amino acids, for example, arginine, glycine, leucine, serine, proline and threonine, some codons almost do not appear in highly expressed genes (violet line, least point marked by red square). This is a clear codon bias correlated to gene expression in the *E. coli* genome.

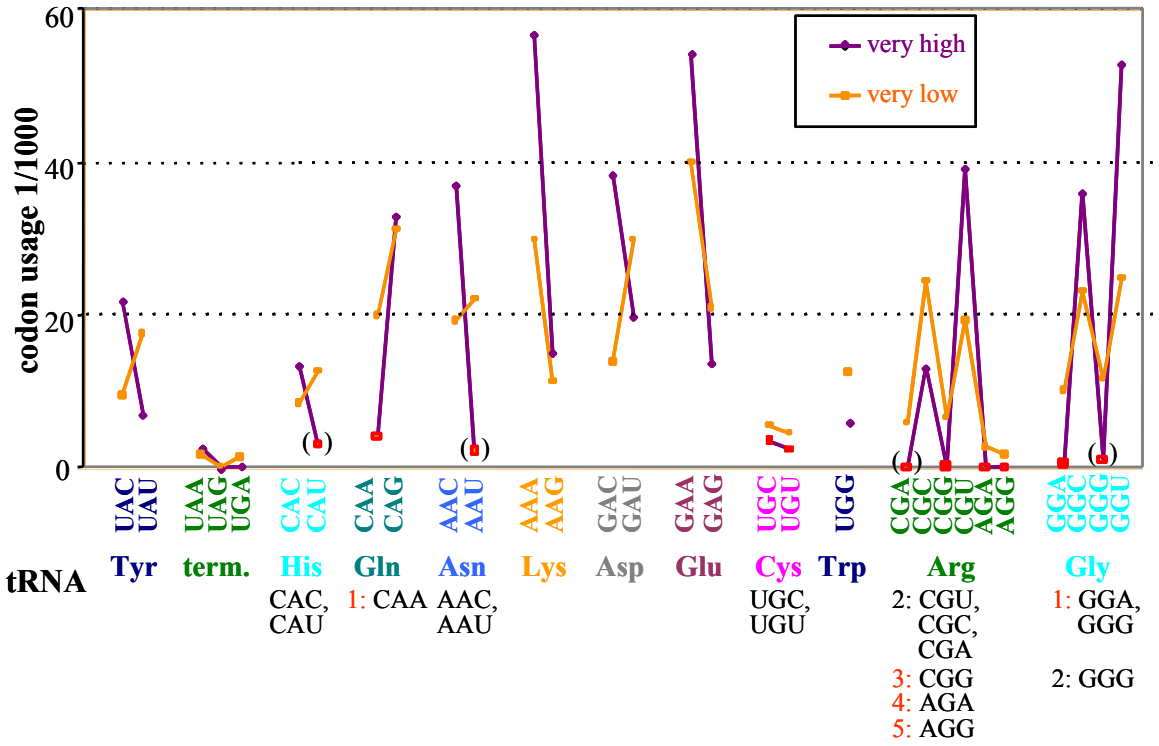


Figure 4.4-1 Distribution of different codons within E. coli genome. The genome was subdivided into two groups of genes corresponding to relatively **high** expressed (**violet**), and relatively **low** expressed (**orange**). Numbers stand for the tRNA isoacceptor species.

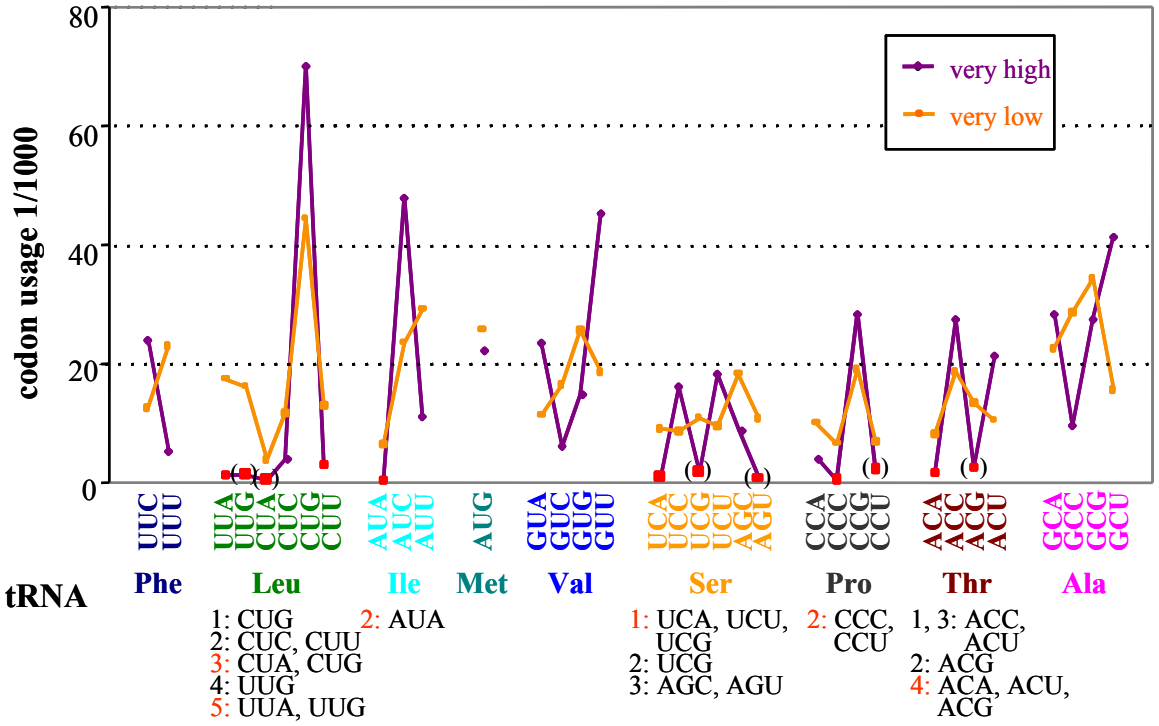


Figure 4.4-1 Continued…

When we performed a similar investigation in human genes, assuming that the ribosomal proteins and glycolysis enzymes are highly expressed genes, while genes for proteases and RNases are lowly expressed, none of such dramatic differences as seen in *E. coli* codons were detected. Also the total codon usage in the human genome does not deviate significantly from that in lowly or highly expressed genes. The probable reason is that in higher eukaryotes there is no need to give an overall preference to a certain codon within one or another group of genes due to homeostasis of the human cells, which are grouped in tissues and organs for better maintenance of their survival. In contrast, in *E. coli*, as in free-living bacteria, there is always a need to adapt the metabolism as fast as possible to an alteration in the environment.

The codon usage in human genes is shown in (Figure 4.4-2) and is, as explained, relatively stable and does not deviate from one group of genes to another, as well as for the total proteins group. This can be judged from the parallelism of most of the lines for both highly and lowly expressed genes, which directs to the fact that most codon/tRNA species for the same amino acid, for example, glutamine, lysine, leucine, valine, serine, threonine and alanine are represented equally within each group of genes, *i.e.*, if one codon is underrepresented in one group of genes (*e.g.,* ribosomal proteins), then it is underrepresented in all other groups of genes as well, with only slight difference in abundance, which is within the error bar.

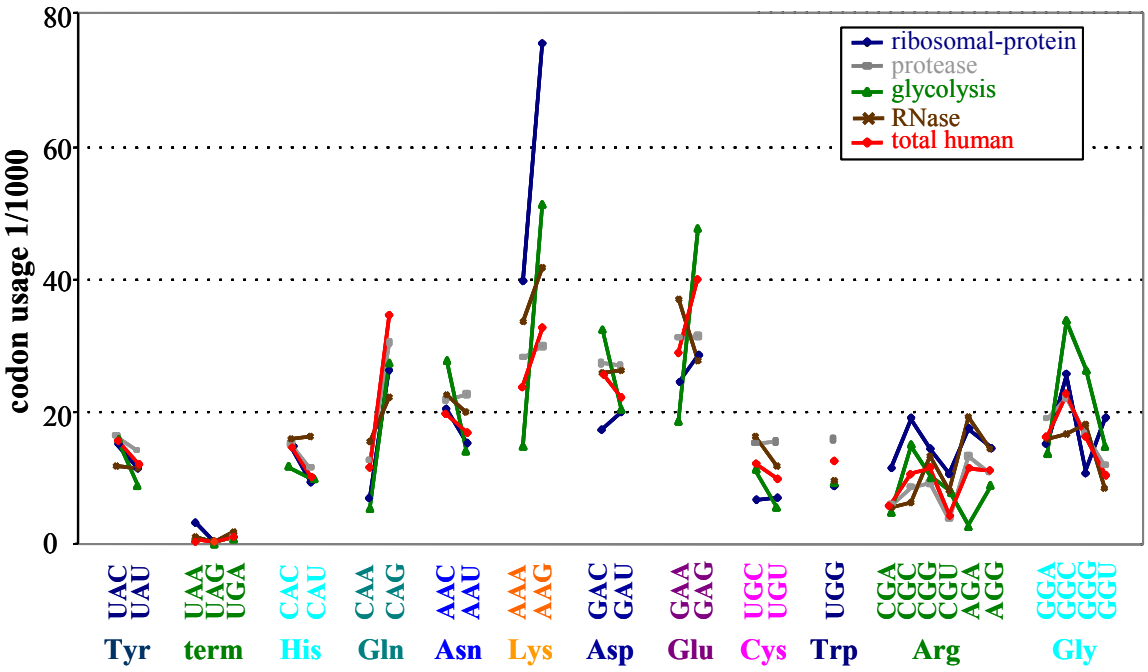


Figure 4.4-2 Distribution of different codons within human genome. Various gene families were selected from human genome and defined as highly expressed and lowly expressed ones. Codon distribution in overall human genes was also analysed.

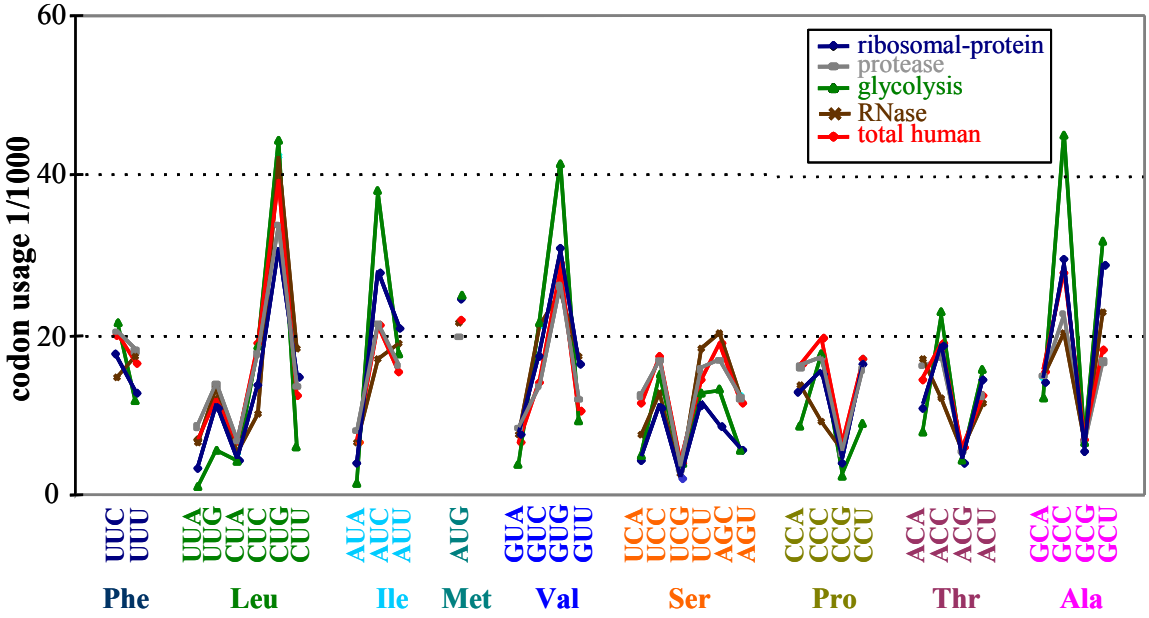


Figure 4.4-2 Continued…

It becomes clear from these results, that during expression of eukaryotic, particularly human genes in *E. coli* cells or cell-free extracts, there will be shortage of defined classes of tRNA species, which are underrepresented in *E. coli*.

We compared data from both analyses in order to identify those codons that are underrepresented in *E. coli*, but often used in the human genome, and we found out that there are at least 15 codons, which are of great importance for human protein synthesis and are almost not appearing in bacteria. Some of these codons were belonging to the same amino acid (*e.g.,* arginine, glycine, leucine, serine, proline and threonine group of codons; assigned as red squares in brackets on Figure 4.4-1). Since some tRNA species can recognize two to three different codons coding for the same amino acid, we determined the optimal set of tRNAs that should be enriched in the *E. coli* bulk tRNA to allow an easy translation for all human genes concerning the codon usage. See, for example, the Gly codons. The codons GGA and GGG are rare in *E. coli* (Figure 4.4-1) but relatively frequent in human genes (Figure 4.4-2). Since a single *E. coli* tRNA decodes both GGA and GGG, we only need to substitute the *E. coli* $tRNA^{bulk}$ with this $tRNA^{Gly}$ to decode both GGA and GGG in a human gene if expressed in an *E. coli* system (see orange number "1" below "Gly" in Figure 4.4-1). In this way we identified 11 tRNA species that should be enriched in *E. coli* $tRNA^{bulk}$ for a proper translation of a human gene in an *E. coli* system.

It is clear therefore that the tRNA concentrations in a bacterial system have to be adjusted to the eukaryotic codon usage.

Table 4.4-1 E. coli tRNAs and the codon recognition pattern (Dong et al., 1996).

tRNA	Anticodon (5'–3')	Codon recognition (5'–3')	№ of molecules per cell	Fraction of tRNA out of total tRNA (%)
Ala1B	UGC	GCU, GCA, GCG	3250(±223)	5.04
Ala2	GGC	GCC	617(±64)	0.95
Arg2	ACG	CGU, CGC, CGA	4752(±440)	7.37
Arg3	CCG	CGG	639(±63)	0.99
Arg4	UCU	AGA	867(±160)	1.34
Arg5	CCU	AGG	420(±69)	0.65
Asn	GUU	AAC, AAU	1193(±127)	1.85
Asp1	GUC	GAC, GAU	2396(±346)	3.72
Cys	GCA	UGC, UGU	1587(±126)	2.46
Gln1	UUG	CAA	764(±66)	1.18
Gln2	CUG	CAG	881(±94)	1.36
Glu2	UUC	GAA, GAG	4717(±411)	7.32
Gly1[a]	CCC	GGG		
Gly2	UCC	GGA, GGG	2137(±320)	3.31
Gly3	GCC	GGC, GGU	4359(±378)	6.76
His	GUG	CAC, CAU	639(±95)	0.99
Ile1	GAU	AUC, AUU	3474(±94)	5.39
Ile2[a]	CAU	AUA		
Leu1	CAG	CUG	4470(±346)	6.94
Leu2	GAG	CUC, CUU	943(±97)	1.46
Leu3	UAG	CUA, CUG	666(±94)	1.03
Leu4	CAA	UUG	1913(±190)	2.97
Leu5	UAA	UUA, UUG	1031(±117)	1.60
Lys	UUU	AAA, AAG	1924(±185)	2.97
Met f1	CAU	AUG	1211(±191)	1.88
Met f2	CAU	AUG	715(±107)	1.11
Met m	CAU	AUG	706(±96)	1.09
Phe	GAA	UUC, UUU	1037(±162)	1.60
Pro1	CGG	CCG	900(±150)	1.38
Pro2	GGG	CCC, CCU	720(±125)	1.11
Pro3	UGG	CCA, CCU, CCG	581(±95)	0.90
Sec	UCA	UGA	219(±73)	0.34
Ser1	UGA	UCA, UCU, UCG	1296(±94)	2.01
Ser2	CGA	UCG	344(±62)	0.53
Ser3	GCU	AGC, AGU	1408(±126)	2.18
Ser5	GGA	UCC, UCU	764(±127)	1.18
Thr1	GGU	ACC, ACU	104(±34)	0.16
Thr2	CGU	ACG	541(±94)	0.84
Thr3	GGU	ACC, ACU	1095(±62)	1.70
Thr4	UGU	ACA, ACU, ACG	916(±64)	1.42
Trp	CCA	UGG	943(±162)	1.46
Tyr1	GUA	UAC, UAU	769(±95)	1.19
Tyr2	GUA	UAC, UAU	1261(±126)	1.95
Val1	UAG	GUA, GUG, GUU	3840(±218)	5.96
Val2A	GAC	GUC, GUU	630(±98)	0.97
Val2B	GAC	GUC, GUU	635(±95)	0.98
4.5 S RNA			416(±63)	0.64

The number of tRNA molecules per cell and the fraction of tRNA out of total tRNA population in *E. coli* grown at 0.4 doublings per hour are shown as described in the text. ± stands for the standard deviations calculated from six independent measurements for each individual tRNA isoacceptor. The data on tRNA codon recognition patterns were obtained from (Björk, 1995; Garcia *et al.*, 1986; Ikemura, 1985; Ikemura and Ozeki, 1983; Komine *et al.*, 1990; Saxena and Walker, 1992).

[a] The tRNA isoacceptors Gly1 and Gly2 are treated collectively as are the data for Ile1 and Ile2. Highlighted with gold are the rows that correspond to our findings of the tRNA isoacceptors that are underrepresented in *E. coli* in respect of human genes expression.

Commercial suppliers have noticed a possible shortage already, and Novagen offers the Rosetta™ (DE3) strain that carries six tRNAs out of eight marked red in Table 4.4-2, in order to enable precise control of expression levels by adjusting the concentration of IPTG. The tRNAs introduced on a pRARE plasmid are *proL* tRNA2 (CCC), *leuW* tRNA3 (CUA), *argW* tRNA5 (AGG), *glyT* tRNA2 (GGA), *argU* tRNA4 (AGA and AGG), *ileX* tRNA2 (AUA).

Table 4.4-2. Rare codons in E. coli

Amino acid	Codon	Fraction in all genes	Fraction in Class II
Arg5	**AGG**	**0.022**	**0.003**
Arg4	**AGA**	**0.039**	**0.006**
Arg3	**CGG**	**0.098**	**0.008**
Arg2	**CGA**	**0.065**	**0.011**
Arg2	CGU	0.378	0.643
Arg2	CGC	0.398	0.330
Gly2	GGG	0.161	0.044
Gly2	**GGA**	**0.109**	**0.020**
Gly3	GGU	0.337	0.508
Gly3	GGC	0.403	0.428
Ile2	**AUA**	**0.073**	**0.006**
Ile1	AUU	0.507	0.335
Ile1	AUC	0.420	0.659
Leu5	UUG	0.129	0.034
Leu5	UUA	0.131	0.056
Leu3	CUG	0.496	0.767
Leu3	**CUA**	**0.037**	**0.008**
Leu2	CUU	0.104	0.056
Leu2	CUC	0.104	0.080
Pro3	CCG	0.525	0.719
Pro3	CCA	0.191	0.153
Pro2	CCU	0.159	0.112
Pro2	**CCC**	**0.124**	**0.016**

Codon usage is expressed here as a fraction of all possible codons for a given amino acid. "All genes" is the fraction represented in all 4,290 coding sequences in the *E. coli* genome (Nakamura *et al.*, 2000). "Class II" is the fraction represented in 195 genes highly and continuously expressed during exponential growth (Henaut and Danchin, 1996). Codons that are underrepresented in *E. coli* genes are marked in red. Numbers after amino acid represent tRNA isoacceptor species, according to Dong *et al.*

In order to homogenise the nomenclature of tRNA isoacceptors, I changed the names reported by Novagen to those used by Dong *et al.* This means that numbers after tRNAs in the following table and text correspond to the isoacceptor species, coding for a given amino acid.

We tested only two tRNAs as a simple check, $tRNA^{Leu}$ and $tRNA^{Ile}$. According to our analysis, only $tRNA^{Leu}$ of the two tRNAs that should be over-expressed proved the expectations: The $tRNA^{Leu}$ was charged to relatively higher levels as examined in an amino acylation assay, in comparison to $tRNA^{bulk}$ from standard non-induced *E. coli* cells, whereas $tRNA^{Ile}$ was not overexpressed (Table 3.7.1-1, page 72). Therefore, we are slow to use this strain for eukaryotic genes expression, whether *in vivo* or *in vitro*.

The important difference of our analysis to the tRNAs present in Rosetta™ (DE3) is the fact that 11 tRNAs should be supplied and not eight (highlighted red in Table 4.4-2, as suggested by Novagen; Henaut and Danchin, 1996; Nakamura *et al.*, 2000).

4.5 Design of an mRNA with an enhancer for high ribosome occupancy

Returning to initiation of translation, it is necessary to mention that 30S subunits, unlike eukaryotic 40S, are incapable of the process of scanning. A 30S subunit simply recognizes the ribosomal binding site (RBS). A sequence of the RBS of the mRNA, the Shine-Dalgarno sequence (SD), is known to be complementary to the sequence at 3' end of the 16S rRNA, known as anti-Shine-Dalgarno (anti-SD). The SD sequence is situated about 4 to 12 nucleotides upstream (in front) of the AUG start codon.

It is known that the very 3' end of 16S rRNA is in close proximity to the E-site of the 30S subunit, since the AUG start codon is positioned at the P-site – directed by the SD–anti-SD interactions, awaiting for IF2•GTP•Met-$tRNA_f^{Met}$ to come, and 50S subunit to assemble into the elongating 70S ribosome.

Imagine that RBS is “hidden” in a “forest” of secondary structures, which the 30S subunit alone cannot melt. As a result, we have no protein synthesis or one at very low levels. In fact, such a situation is much more usual for exogenous mRNAs, mRNAs of the genes with low expression levels, and for many eukaryotic mRNAs added to bacterial *in vitro* systems for protein synthesis. Even if the gene is introduced into *in vitro* system for transcription and translation, due to uncoupling of these processes by utilization of a T7 RNA polymerase, which is a fast molecule, mRNAs transcribed are still a hard task for 30S to initiate as well as for elongating 70S ribosomes. Therefore, we developed an mRNA with a cassette for an exogenous cistron that should allow high expression of this cistron. This construct shows the following features:

1: Weak secondary structures at the 5’-end

In early studies de Smit and van Duin presented data that expression is limited by mRNA structure or when mRNA has a low affinity for ribosomes. They also report experimental data on the energy of helix formation (ΔG^{0}_{f}) for mRNA, which in the range of –6.0 to –2.4 kcal/mol resulted in best expression levels, whereas a secondary structure of energy below –6.4 kcal/mol prevented ribosomal initiation (de Smit and van Duin, 1990). And another group reported few years earlier that the high preference for adenosine residues in true sites has indeed been ascribed to their low potential to form stable secondary structures (Looman *et al.*, 1987).

It is also known that both the eukaryotic and the prokaryotic ribosomes translate equally well artificial mRNAs such as poly-U and poly-A, without the aid of either a cap or IRES, which suggests that the basic RNA-binding properties of both kinds of ribosomes are the same. Another important feature for translation initiation is that an mRNA chain is single-stranded, which is a fundamental prerequisite for initiation to occur in both cell types: the unique ways in which such single-stranded molecule is ensured may suffice to account for most of the differences between prokaryotic and eukaryotic initiation (Londei, 2001).

It is important to consider the context of the 5'UTR sequences from natural enhancers and to include the starting coding sequence of the gene that follows. An example of a well-translated mRNA that does not have an SD sequence is the so-called epsilon sequence (UUAACUUUA) present in the leader region of *gene* 10 of the T7 bacteriophage. Is the absence of secondary structure responsible for its efficient translation? We analyzed the possible secondary structure of epsilon in the context of its WT mRNA by means offered on the mfold web server for nucleic acid folding and hybridization prediction developed by Zuker (Mathews *et al.*, 1999; Zuker, 2003), and found that the overall ΔG^0_f is relatively high (Figure 4.5-1), in comparison to the stability threshold identified as values below –6.0 kcal/mol.

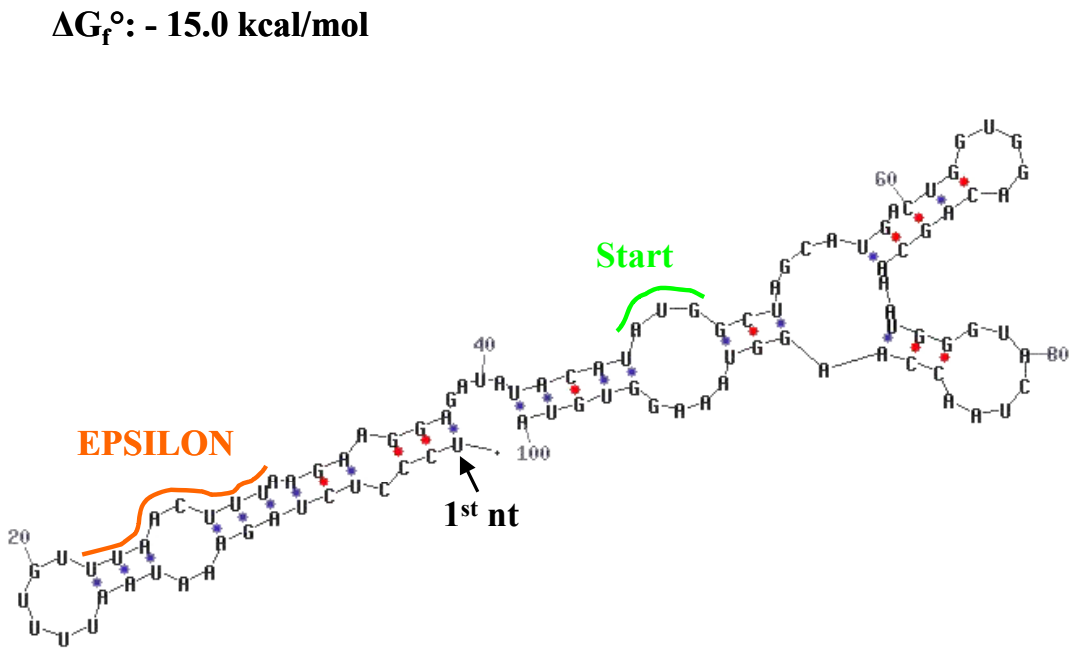


Figure 4.5-1 Epsilon sequence found in the leader region of gene 10 *of the T7 bacteriophage*. The $\Delta G_f°$ of the sequence that includes 5' untranslated region of *gene* 10, and about 50 nucleotides of sense codons. The 1st nucleotide (1st nt), the epsilon sequence and the AUG start codons are indicated.

However, secondary structures found in one mRNA molecule are not linked *via* stacking interactions. Therefore, we measured separately the stabilities of the secondary structures found in the case of *gene* 10 mRNA, and observed ΔG^0_f values, which can be solved by the 30S ribosomal subunit alone

(Figure 4.5-2A). Thus, a sequence, close to that of epsilon would be good if incorporated into the upstream region (5') of the mRNA molecule of interest.

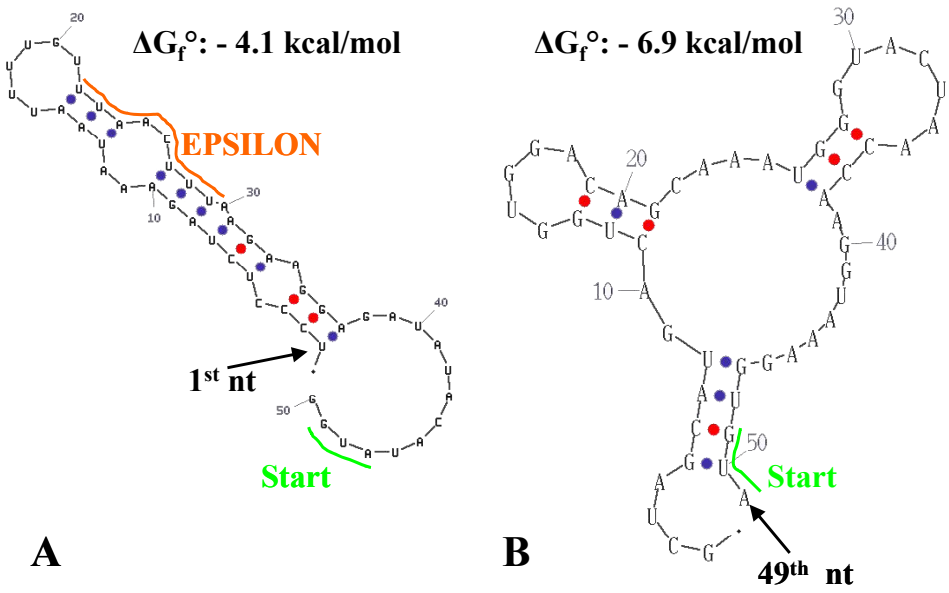


Figure 4.5-2 An individual secondary structures measurements of the epsilon gene 10 *T7 bacteriophage mRNA sequence*. (**A**) 5' untranslated region of *gene* 10, including AUG start codon; (**B**) secondary structure of AUG start codon followed by about 50 nucleotides of *gene* 10 coding sequence. The 1st nucleotide (1st nt), the epsilon sequence and the AUG start codons are indicated.

2: A short leader cistron preceding the exogenous cistron

We copied the codon sequence of the CAT leader peptide that precedes the chloramphenicol acetyl-transferase and is efficiently translated. This leader should warrant that the ribosomes easily initiate and thus lead to high ribosome occupancy of the mRNA.

3: Translation transition from the leader peptide to that of the exogenous cistron

The problem we would like to solve is that ribosomes present in high occupancies at the leader peptide should directly continue the translation of the following cistron without falling off the mRNA with a subsequent 30S *de novo* initiation. We applied a "trick" that is used during translation of polycistronic mRNAs coding for various ribosomal proteins. We included the transition site of the adjacent cistrons on a polycistronic mRNA coding for the ribosomal proteins

L29 and S17, where a short SD sequence of four nucleotides (for the downstream cistron) is followed by a stop codon UAA (for the upstream cistron) and the AUG-start codon under overlapping conditions: *GGU GCG* UAA **UG**. The stop codon UAA of the short leader is in red, the SD sequence of the following cistron is in italics, and the AUG start codon of the following cistron is in bold letters. The SD sequence is thus situated just four nucleotides upstream of this AUG-start codon and plays the role of an anchor that traps the empty (after termination and peptide release) ribosome exposing the AUG-codon near the P-site and reassuring the 70S initiation mode to occur. Thus, in front of the gene of interest we have a short open-reading frame that should facilitate the expression of the gene of interest.

The long row of sequence optimisation and energy comparisons of the respective secondary structures resulted finally in what we call *Berlin-sequence* (Figure 4.5-3 and Figure 4.5-4 for secondary structure), which we cloned into standard pET23c(+) vector, replacing the present T7 promoter with one carried by the insert.

Resuming everything mentioned above, this Berlin sequence can be described as a sequence based on successful examples from life, in order to create a good artificial non-SD enhancing sequence known for initiation of translation.

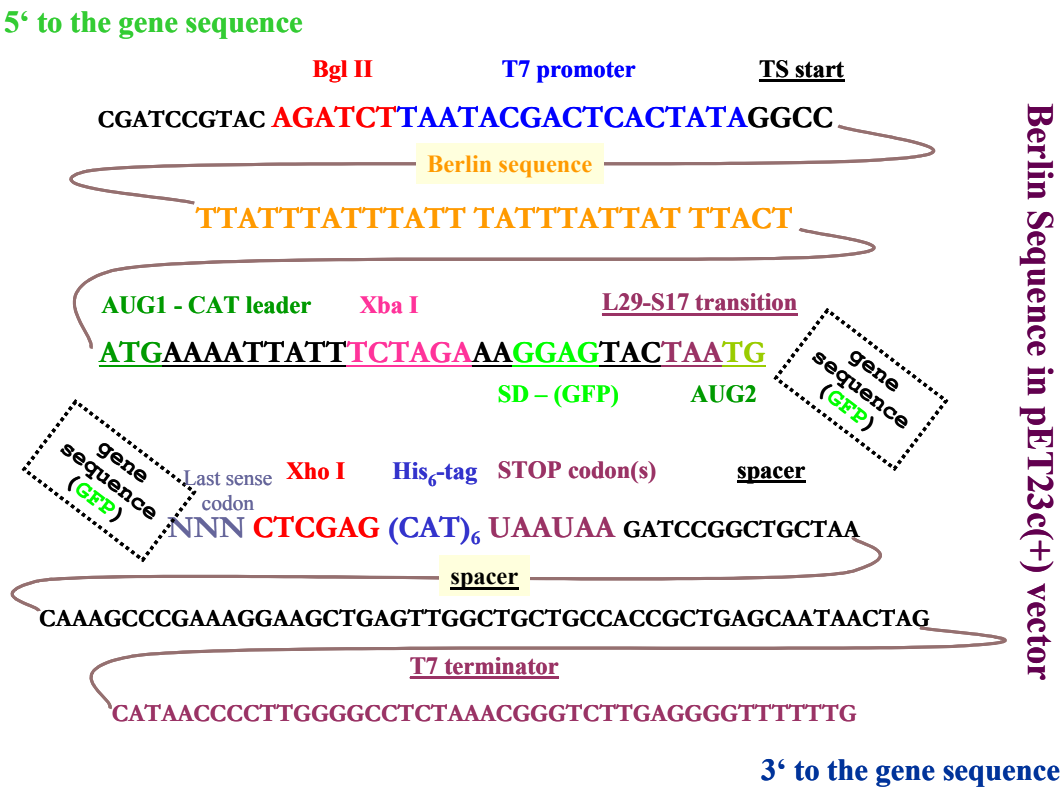


Figure 4.5-3 Sequence designed to enhance the initiation of translation. Marking of the important sites are given above and in colours.

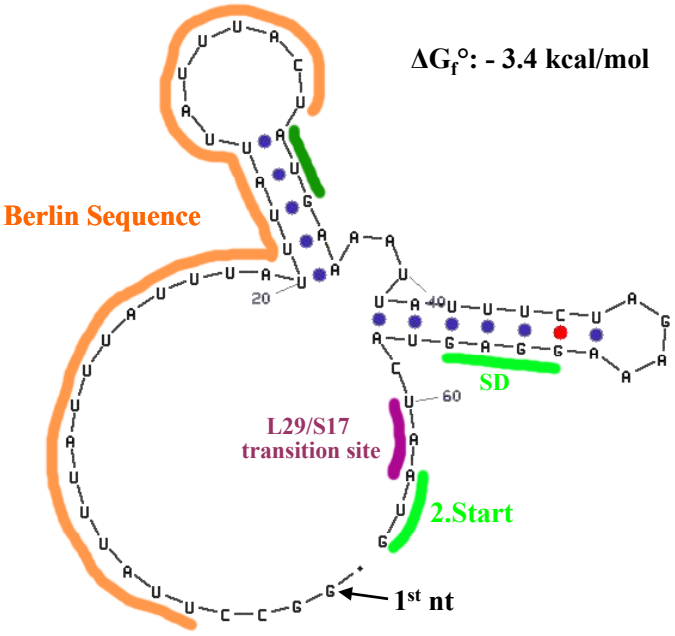


Figure 4.5-4 Secondary structure of the Berlin sequence designed to enhance initiation of translation. Marking are given according to the coloured regions that represent most significant sequences.

First results obtained by my colleague Witold Szaflarski in our group indicated that this construct triggers a super-expression *in vivo*, whereas the *in vitro* results are not yet satisfying. This part is an ongoing research program that includes further optimisation of the Berlin sequence.

4.6 Investigation of the fragmentation of a given protein

We aimed to reduce the level of fragmentation for *E. coli* translational elongation factor Tu (EF-Tu) by means that were already investigated and discussed in this work before. In our understanding the reduction of the incubating temperature may lead to the retardation or elimination of the unwanted fragments, which are observed during incubation of the reaction at 30°C. As a negative control for the temperature decrease effect we also ran a reaction with increased temperature that should have resulted in the increased level of fragmentation. The analysis was done for GFPcyc3 as well, assuming that this protein has almost zero level of fragmentation. We determined the *fragmentation index* (F.I.), in order to describe the level of fragmentation for EF-Tu. The calculation of this value was based on the relative intensities of the bands corresponding to full-length EF-Tu protein and two different and most pronounced fragments of EF-Tu, determined by radiolabelling (scanning of the SDS-PAAG), as indicated by the formula:

$$\text{F.I.} = (\text{frg. \#1} + \text{frg. \#2}) / (\text{frg. \#1} + \text{frg. \#2} + \text{EF-Tu})$$

According to this formula we calculated the F.I. of EF-Tu for different temperatures (Table 3.8-1, page 75) and found out that there is slight positive effect if at all.

In order to normalize the synthesis levels of full-length EF-Tu, we compared its band intensities to those of GFPcyc3, in kinetics at different incubation temperatures. We observed that the level of fragmentation towards the full-length protein at the beginning of each reaction was lower than at the end of the reactions. The other significant detail is that the total yield of EF-Tu at 37°C increased twice as compared to those at 20°C or 30°C (relative to

GFPcyc3 synthesis), and that the number and amount of fragments had increased, too (Table 3.8-1, page 75). We conclude that the reason for EF-Tu fragmentation is not related to the incubation temperature.

If the EF-Tu protein in the form of ternary complex is more stable towards fragmentation, then an extra addition of the amino acylated tRNA would stimulate the ternary complex formation and, as a result, reduce the fragmentation level. Because the previous results showed little affect of the incubation temperature on the EF-Tu fragmentation next reactions were performed at 30°C. Here, a three-fold excess of $tRNA^{Phe}$ and a twelve-fold excess of phenylalanine amino acid over the estimated synthesis of full-length EF-Tu protein was added either at the beginning of reaction, or one hour later after the incubation had already started. According to the estimation of F.I. and the overall analysis of the full-length EF-Tu protein synthesis (Table 3.8-2, page 77), the presence of $tRNA^{Phe}$:Phe mixture had again a little positive effect on the reduction of EF-Tu fragmentation when was added one hour later after incubation of the reaction had started.

In the presence of protease inhibitory mixes we observed in some cases a reduction of the EF-Tu fragments, but also an impairment of synthesis of the mature EF-Tu. This effect might be due to the components of the mixes, that could have acted as inhibitors of the protein synthesis. Besides this, the protease inhibitory mixes are recommended for storage of the preparative protein isolation from cell cultures.

Referring to the data obtained for incubation temperature affect on the EF-Tu synthesis, the major observation is that fragmentation of this protein is not much temperature dependent, but rather depends on the incubation time. The longer the incubation time, the more fragments are detected. According to our analysis, incubation up to two hours is recommended either at 30°C or at 20°C in order to obtain mainly mature EF-Tu protein. *In vivo*, additional factors might be involved. In fact, my colleague Yan Qin in our group has observed some

spectacular effects improving the output of fully active proteins by the addition of a universally conserved G-protein with a hitherto unknown function.

In the course of this thesis we have developed quality criteria that allow a critical evaluation of parameters important for the coupled transcription /translation system or improving the yield and quality of the synthesized protein. Improvements could be obtained by applying slower T7 polymerase and amino acid additions after half of the standard incubation time. Another example is that both lowering the incubation temperature to 25°C or 20°C plus the additional administration of amino acids improved the active fraction of the synthesized protein combined with a satisfying yield. We further identified 11 tRNAs that should be added to a bacterial system, *e.g.* from *E. coli*, for the optimal expression of eukaryotic genes. On the other hand, we could show that the shortage of NTPs or prolonging the half-life of mRNA does not improve the output of protein. But the optimisation coupled system as it stands after these analyses are not yet finished. This can be easily demonstrated by comparing the efficiency of the excellent RTS (Roche) with that of an *E. coli* cell. The reaction mix before synthesis contains about 40 mg per ml total proteins, and after 10 h an amount of GFP has been synthesized that comes to about 20% of the total proteins (8 mg/ml; Figure 3.4-3A, page 56). A continuation of this synthesis rate would lead to a doubling of the protein content in the reaction mixture (total proteins + synthesised GFP) after 50 h. *E. coli* has a doubling time of 20 min under reach medium conditions. It follows that the good RTS system is still 150-fold less efficient than protein synthesis *in vivo*. I have identified some ways to improve the system; others are under consideration of my colleagues in the Nierhaus group. Examples are a further optimisation of the Berlin sequence for *in vitro* expression and additions of newly found factors that improve the active fraction even at 30°C incubation temperature.

References

Agrawal, R.K., Spahn, C.M.T., Penczek, P., Grassucci, R.A., Nierhaus, K.H. and Frank, J. (2000) Visualization of tRNA movements on the *Escherichia coli* 70S ribosome during the elongation cycle. *J. Cell. Biol.*, **150**, 447-459.

Alakhov, Y.B., Baranov, P.V., Ovodov, S.J., Ryabova, L.A., Spirin, A.S. and Morozov, I.J. (1995) Method of preparing polypeptides in cell-free translation system. USA.

Baca, A.M. and Hol, W.G. (2000) Overcoming codon bias: a method for high-level overexpression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*. *Int. J. Parasitol.*, **30**, 113-118.

Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905-920.

Baneyx, F. and Georgiou, G. (1991) Construction and characterization of *Escherichia coli* strains deficient in multiple secreted proteases: protease III degrades high-molecular-weight substrates *in vivo*. *J. Bacteriol.*, **173**, 2696-2703.

Berger, S.L. (1987) Quantifying ^{32}P-labeled and unlabeled nucleic acids. *Methods Enzymol.*, **152**, 49-54.

Björk, G.R. (1995) Genetic dissection of synthesis and function of modified nucleosides in bacterial transfer RNA. In Cohn, W.E. and Moldave, K. (eds.), *Progress in Nucleic Acid Research and Molecular Biology, Vol 50*. Academic Press Inc., 525 B Street, Suite 1900, San Diego, CA 92101-4495, Vol. 50, pp. 263-338.

Blaha, G. and Nierhaus, K.H. (2001) Features and functions of the ribosomal E site. *Cold Spring Harbor Symposia on Quantitative Biology*, **65**, 135-145.

Bonner, G., Lafer, E.M. and Sousa, R. (1994) Characterization of a set of T7 RNA polymerase active site mutants. *J. Biol. Chem.*, **269**, 25120-25128.

Bonner, G., Patra, D., Lafer, E.M. and Sousa, R. (1992) Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure. *EMBO J.*, **11**, 3767-3775.

Borsook, H., Deasy, C.L., Haagensmit, A.J., Keighley, G. and Lowy, P.H. (1950) Incorporation *in vitro* of labeled amino acids into rat diaphragm proteins. *J. Biol. Chem.*, **186**, 309-315.

Bouadloun, F., Donner, D. and Kurland, C.G. (1983) Codon-specific missense errors *in vivo*. *EMBO J.*, **2**, 1351-1356.

Bremer, H. and Dennis, P.P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In Neidhardt, F.C., III, R.C., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikow, W.S., Riley, M., Schaechter, M. and Umbarger, H.E. (eds.), *Escherichia coli and Salmonella*. ASM Press, Washington D.C., Vol. 2, pp. 1553-1569.

Bulmer, M. (1988) Are codon usage patterns in unicellular organisms determined by selection-mutation balance? *J. evol. Biol.*, **1**, 15-26.

Burkhardt, N., Jünemann, R., Spahn, C.M.T. and Nierhaus, K.H. (1998) Ribosomal tRNA binding sites: Three-sites models of translation. *Crit. Rev. Biochem. Mol. Biol.*, **33**, 95-149.

Caskey, C.T., Forrester, W.C. and Tate, W. (1984) Peptide chain termination. *Gene Expression*

proceeding of The Alfred Benzon Symposium, Copenhagen Munkgard, **19**, 457-466.

Cate, J.H., Yusupov, M.M., Yusupova, G.Z., Earnest, T.N. and Noller, H.F. (1999) X-ray crystal structures of 70S ribosome functional complexes. *Science*, **285**, 2095-2104.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. and Prasher, D.C. (1994) Green fluorescent protein as a marker for gene expression. *Science*, **263**, 802-805.

Chamberlin, M. and Ring, J. (1973) Characterization of T7-specific ribonucleic acid polymerase. 1. General properties of the enzymatic reaction and the template specificity of the enzyme. *J. Biol. Chem.*, **248**, 2235-2244.

Cheetham, G.M. and Steitz, T.A. (1999) Structure of a transcribing T7 RNA polymerase initiation complex. *Science*, **286**, 2305-2309.

Chen, H.Y., Bjerknes, M., Kumar, R. and Jay, E. (1994) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res.*, **22**, 4953-4957.

Craigen, W.J., Cook, R.G., Tate, W.P. and Caskey, C.T. (1985) Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. *Proc. Natl. Acad. Sci. USA*, **82**, 3616-3620.

Crameri, A., Whitehorn, E.A., Tate, E. and Stemmer, W.P. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.*, **14**, 315-319.

de Smit, M.H. and van Duin, J. (1990) Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. *Proc. Natl. Acad. Sci. USA*, **87**, 7668-7672.

Dekker, N., Cox, R.C., Kramer, R.A. and Egmond, M.R. (2001) Substrate specificity of the integral membrane protease OmpT determined by spatially addressed peptide libraries. *Biochemistry*, **40**, 1694-701.

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

Frank, J., Zhu, J., Penczek, P., Li, Y.H., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R.K. and Agrawal, R.K. (1995) A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. *Nature*, **376**, 441-444.

Freistroffer, D.V., Pavlov, M.Y., MacDougall, J., Buckingham, R.H. and Ehrenberg, M. (1997) Release factor RF3 in *E. coli* accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *EMBO J.*, **16**, 4126-4133.

Gabashvili, I.S., Agrawal, R.K., Spahn, C.M.T., Grassucci, R.A., Svergun, D.I., Frank, J. and Penczek, P. (2000) Solution structure of the *E. coli* 70S ribosome at 11.5 angstrom resolution. *Cell*, **100**, 537-549.

Gale, E.F. and Folkes, J.P. (1954) Effect of nucleic acids on protein synthesis and amino-acid incorporation in disrupted staphylococcal cells. *Nature*, **173**, 1223-1227.

Gallie, D.R. and Kado, C.I. (1989) A translational enhancer derived from tobacco mosaic virus is functionally equivalent to a Shine-Dalgarno sequence. *Proc. Natl. Acad. Sci. USA*, **86**, 129-132.

Ganoza, M.C., Aoki, H., Burkhardt, N. and Murphy, B.J. (1996) The ribosome as 'affinity matrix': Efficient purification scheme for translation factors. *Biochimie*, **78**, 51-61.

Garcia, G.M., Mar, P.K., Mullin, D.A., Walker, J.R. and Prather, N.E. (1986) The *E. coli dnaY* gene encodes an arginine transfer RNA. *Cell*, **45**, 453-459.

Gold, L. (1988) Posttranscriptional regulatory mechanisms in *Escherichia coli*. *Ann. Rev. Biochem.*, **57**, 199-233.

Gold, L.M. and Schweiger, M. (1969) Synthesis of phage-specific alpha- and beta-glucosyl transferases directed by T-even DNA *in vitro*. *Proc. Natl. Acad. Sci. USA*, **62**, 892-898.

Goldman, E., Rosenberg, A.H., Zubay, G. and Studier, F.W. (1995) Consecutive low-usage leucine codons block translation only when near the 5' end of a message in Escherichia coli. *J Mol Biol*, **245**, 467-473.

Grajevskaja, R.A., Ivanov, Y.V. and Saminsky, E.M. (1982) 70-S ribosomes of *Escherichia coli* have an additional site for deacylated tRNA binding. *Eur. J. Biochem.*, **128**, 47-52.

He, B., Rong, M., Lyakhov, D., Gartenstein, H., Diaz, G., Castagna, R., McAllister, W.T. and Durbin, R.K. (1997) Rapid mutagenesis and purification of phage RNA polymerases. *Protein Express. Purif.*, **9**, 142-151.

Henaut, A. and Danchin, A. (1996) Analysis and predictions from *Escherichia coli* sequences, or *E. coli* in silico. In Neidhardt, F.C. (ed.) *Escherichia coli and Salmonella typhimurium cellular and molecular biology*, Vol. 2, pp. 2047-2066.

Huala, E., Moon, A.L. and Ausubel, F.M. (1991) Aerobic inactivation of *Rhizobium meliloti* NifA in *Escherichia coli* is mediated by *lon* and two newly identified genes, *snoB* and *snoC*. *J. Bacteriol.*, **173**, 382-390.

Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, **2**, 13-34.

Ikemura, T. and Ozeki, H. (1983) Codon usage and transfer RNA contents: organism-specific codon-choice patterns in reference to the isoacceptor contents. *Cold Spring Harb. Symp. Quant. Biol.*, **47 Pt 2**, 1087-1097.

Iost, I. and Dreyfus, M. (1995) The stability of *Escherichia coli lacZ* mRNA depends upon the simultaneity of its synthesis and translation. *EMBO J.*, **14**, 3252-3261.

Iost, I., Guillerez, J. and Dreyfus, M. (1992) Bacteriophage T7 RNA polymerase travels far ahead of ribosomes *in vivo*. *J. Bacteriol.*, **174**, 619-622.

Ivanov, I.G., Alexandrova, R., Dragulev, B., Leclerc, D., Saraffova, A., Maximova, V. and Abouhaidar, M.G. (1992) Efficiency of the 5'-terminal sequence (omega) of tobacco mosaic virus RNA for the initiation of eukaryotic gene translation in *Escherichia coli*. *Eur J Biochem*, **209**, 151-156.

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

Jewett, M.C. and Swartz, J.R. (2004b) Substrate replenishment extends protein synthesis with an *in vitro* translation system designed to mimic the cytoplasm. *Biotechnol. Bioeng.*, **87**, 465-472.

Kane, J.F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.*, **6**, 494-500.

Keller, E.B. and Littlefield, J.W. (1957) Incorporation of ^{14}C-amino acids into ribonucleoprotein particles from the Ehrlich mouse ascites tumor. *J. Biol. Chem.*, **224**, 13-30.

Kim, D.M. and Swartz, J.R. (1999) Prolonging cell-free protein synthesis with a novel ATP regeneration system. *Biotechnol. Bioeng.*, **66**, 180-188.

Kim, D.M. and Swartz, J.R. (2000) Prolonging cell-free protein synthesis by selective reagent additions. *Biotechnol. Prog.*, **16**, 385-390.

Kim, D.M., Swartz, J. R. (2000) Oxalate improves protein synthesis by enhancing ATP supply in cell-free system derived from *E. coli*. *Biotechnol. Lett.*, **22**, 1537-1542.

Komine, Y., Adachi, T., Inokuchi, H. and Ozeki, H. (1990) Genomic organization and physical mapping of the transfer RNA genes in *Escherichia coli* K12. *J. Mol. Biol.*, **212**, 579-598.

Kozak, M. (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene*, **234**, 187-208.

Kurland, C. and Gallant, J. (1996) Errors of heterologous protein expression. *Curr. Opin. Biotechnol.*, **7**, 489-493.

Lederman, M. and Zubay, G. (1967) DNA-directed peptide synthesis. 1. A comparison of T2 and *Escherichi coli* DNA-directed peptide synthesis in two cell-free systems. *Biochim. Biophys. Acta*, **149**, 253-258.

Lewicki, B.T.U., Margus, T., Remme, J. and Nierhaus, K.H. (1993) Coupling of rRNA transcription and ribosomal assembly *in vivo* - formation of active ribosomal subunits in *Escherichia coli* requires transcription of rRNA

genes by host RNA polymerase which cannot be replaced by bacteriophage-T7 RNA polymerase. *J. Mol. Biol.*, **231**, 581-593.

Liiv, A., Tenson, T. and Remme, J. (1996) Analysis of the ribosome large subunit assembly and 23S rRNA stability *in vivo*. *J. Mol. Biol.*, **263**, 396-410.

Lill, R., Robertson, J.M. and Wintermeyer, W. (1984) tRNA binding sites of ribosomes from *Escherichia coli*. *Biochemistry*, **23**, 6710-6717.

Lipmann, F. (1963) Messenger ribonucleic acid. *Prog. Natl. Acid. Res.*, **1**, 135-161.

Littlefield, J.W., Keller, E.B., Gross, J. and Zamecnik, P.C. (1955) Studies on cytoplasmic ribonucleoprotein particles from the liver of the rat. *J. Biol. Chem.*, **217**, 111-123.

Loechel, S., Inamine, J.M. and Hu, P.C. (1991) A novel translation initiation region from *Mycoplasma genitalium* that functions in *Escherichia coli*. *Nucleic Acids Res.*, **19**, 6905-6911.

Londei, P. (2001) Translation initiation models in prokaryotes and eukaryotes. *Encyclopedia of Life Sciences*.

Looman, A.C., Bodlaender, J., Comstock, L.J., Eaton, D., Jhurani, P., de Boer, H.A. and van Knippenberg, P.H. (1987) Influence of the codon following the AUG initiation codon on the expression of a modified lacZ gene in *Escherichia coli*. *EMBO J.*, **6**, 2489-2492.

Makarova, O.V., Makarov, E.M., Sousa, R. and Dreyfus, M. (1995) Transcribing of *Escherichia coli* genes with mutant T7 RNA polymerases: Stability of *lacZ* mRNA inversely correlates with polymerase speed. *Proc. Natl. Acad. Sci. USA*, **92**, 12250-12254.

Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911-940.

Matthaei, J.H. and Nirenberg, M.W. (1961) Characteristics and stabilization of DNAase-sensitive protein synthesis in *E. coli* extracts. *Proc. Natl. Acad. Sci. USA*, **47**, 1580-1588.

Mccarthy, J.E.G. and Brimacombe, R. (1994) Prokaryotic translation: The interactive pathway leading to initiation. *Trends Genet.*, **10**, 402-407.

Miroux, B. and Walker, J.E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.*, **260**, 289-298.

Moazed, D. and Noller, H.F. (1989) Intermediate states in the movement of transfer RNA in the ribosome. *Nature*, **342**, 142-148.

Mohanty, B.K. and Kushner, S.R. (1999) Analysis of the function of *Escherichia coli* poly(A) polymerase I in RNA metabolism. *Mol. Microbiol.*, **34**, 1094-1108.

Moore, P.B. and Steitz, T.A. (2002) The involvement of RNA in ribosome function. *Nature*, **418**, 229-235.

Nakamura, Y., Gojobori, T. and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.*, **28**, 292.

Nierhaus, K.H. (1990) The allosteric three-site model for the ribosomal elongation cycle: features and future. *Biochemistry*, **29**, 4997-5008.

Nierhaus, K.H., Beyer, D., Dabrowski, M., Schäfer, M.A., Spahn, C.M.T., Wadzack, J., Bittner, K.-U., Burkhardt, N., Diedrich, G., Jünemann, R., Kamp, D., Voss, H. and Stuhrmann, H.B. (1995) The elongating ribosome: Structural and functional aspects. *Biochem. Cell Biol.*, **73**, 1011-1021.

Nierhaus, K.H., Wadzack, J., Burkhardt, N., Jünemann, R., Meerwinck, W., Willumeit, R. and Stuhrmann, H.B. (1998) Structure of the elongating ribosome: Arrangement of the two tRNAs before and after translocation. *Proc. Natl. Acad. Sci. USA*, **95**, 945-950.

Nirenberg, M.W. and Matthaei, J.M. (1961) The dependance of cell-free protein synthesis in *E. coli* upon naturally occuring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA*, **47**, 1588-1602.

Olins, P.O., Devine, C.S., Rangwala, S.H. and Kavka, K.S. (1988) The T7 phage *gene 10* leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. *Gene*, **73**, 227-235.

Palade, G.E. (1955) A small particulate component of the cytoplasm. *J. Biophys. Biochem. Cytol.*, **1**, 59-68.

Petersen, H.U., Danchin, A. and Grunberg-Manago, M. (1976) Toward an understanding of the formylation of initiator tRNA methionine in prokaryotic protein synthesis. I. *In vitro* studies of the 30S and 70S ribosomal-tRNA complex. *Biochemistry*, **15**, 1357-1362.

Prasher, D.C. (1995) Using GFP to see the light. *Trends Genet.*, **11**, 320-323.

Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G. and Cormier, M.J. (1992) Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene*, **111**, 229-233.

Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation. *Cell*, **108**, 557-572.

Ramakrishnan, V. and Moore, P.B. (2001) Atomic structures at last: the ribosome in 2000. *Curr. Opin. Struct. Biol.*, **11**, 144-154.

Rheinberger, H.-J. (2004) *A history of protein biosynthesis and ribosome research*.

Rheinberger, H.-J. and Nierhaus, K.H. (1980) Simultaneous binding of the 3 tRNA molecules by the ribosome of *E coli*. *Biochem. Internatl.*, **1**, 297-303.

Rheinberger, H.-J., Sternbach, H. and Nierhaus, K.H. (1981) Three tRNA binding sites on *Escherichia coli* ribosomes. *Proc. Natl. Acad. Sci. USA*, **78**, 5310-5314.

Ross, J. (2001) mRNA turnover. *Encyclopedia of Life Sciences*.

Sardesai, N.Y., Green, R. and Schimmel, P. (1999) Efficient 50S ribosome-catalyzed peptide bond synthesis with an aminoacyl minihelix. *Biochemistry*, **38**, 12080-8.

Saxena, P. and Walker, J.R. (1992) Expression of *argU*, the *Escherichia coli* gene coding for a rare arginine tRNA. *J. Bacteriol.*, **174**, 1956-1964.

Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M. and Rosenow, C. (2003) Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.*, **13**, 216-223.

Shine, J. and Dalgarno, L. (1974) The 3'-terminal sequence of *E. coli* 16S rRNA: Complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl Acad. Sci. USA*, **71**, 1342-1346.

Spahn, C.M.T. and Nierhaus, K.H. (1998) Models of the elongation cycle: An evaluation. *Biol. Chem.*, **379**, 753-772.

Sprengart, M.L. and Porter, A.G. (1997) Functional importance of RNA interactions in selection of translation initiation codons. *Mol. Microbiol.*, **24**, 19-28.

Srivastava, A.K. and Schlessinger, D. (1990) Mechanism and regulation of bacterial ribosomal RNA processing. *Annu. Rev. Microbiol.*, **44**, 105-129.

Stark, H., Mueller, F., Orlova, E.V., Schatz, M., Dube, P., Erdemir, T., Zemlin, F., Brimacombe, R. and Heel, M.v. (1995) The 70S *Echerichia coli* ribosome at 23 Å resolution: Fitting the ribosomal RNA. *Structure*, **3**, 815-821.

Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.*, **189**, 113-130.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60-89.

Ueda T. and K., W. (2001) tRNA synthesis and regulation. *Enciclopedia of Life Sciences*.

Ugarov, V.I., Morozov, I.Y., Jung, G.Y., Chetverin, A.B. and Spirin, A.S. (1994) Expression and stability of recombinant RQ-mRNAs in cell-free translation systems. *FEBS Lett.*, **341**, 131-134.

Voss, S. and Skerra, A. (1997) Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein. Eng.*, **10**, 975-982.

Wadzack, J., Burkhardt, N., Jünemann, R., Diedrich, G., Nierhaus, K.H., Frank, J., Penczek, P., Meerwinck, W., Schmitt, M., Willumeit, R. and Stuhrmann, H.B. (1997) Direct localization of the tRNAs within the elongating ribosome by means of neutron scattering (Proton-spin contrast-variation). *J. Mol. Biol.*, **266**, 343-356.

Walz, A., Pirrotta, V. and Ineichen, K. (1976) Lambda repressor regulates the switch between PR and Prm promoters. *Nature*, **262**, 665-669.

Watson, J.D. (1963) Involvement of tRNA in the synthesis of proteins. *Science*, **140**, 17-26.

Watson, J.D. (1964) The synthesis of proteins upon ribosomes. *Bull. Soc. Chim. Biol.*, **46**, 1399-1425.

Weiss, R.B., Murphy, J.P. and Gallant, J.A. (1984) Genetic screen for cloned release factor genes. *J. Bacteriol.*, **158**, 362-364.

Williams, K.J. and Piddock, L.J. (1998) Accumulation of rifampicin by *Escherichia coli* and Staphylococcus aureus. *J. Antimicrob. Chemother.*, **42**, 597-603.

Wilson, D.N. and Nierhaus, K.H. (2003) The ribosome through the looking glass. *Angew. Chem. Int. Ed. Engl.*, **42**, 3464-3486.

Wilson, K.S. and Noller, H.F. (1998) Mapping the position of translational elongation factor EF-G in the ribosome by directed hydroxyl radical probing. *Cell*, **92**, 131-139.

Winnick, T. (1950) Studies on the mechanism of protein synthesis in embryonic and tumor tissues. II. Inactivation of fetal rat liver homogenates by dialysis, and reactivation by the adenylic acid system. *Arch. Biochem.*, **28**, 338-447.

Wu, C.J. and Janssen, G.R. (1996) Translation of vph mRNA in *Streptomyces lividans* and *Escherichia coli* after removal of the 5' untranslated leader. *Mol. Microbiol.*, **22**, 339-355.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H. and Noller, H.F. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science*, **292**, 883-896.

Zamecnik, P.C. (1969) An historical account of protein synthesis, with current overtones - a personalized view. *Cold Spring Harbor Symposia on Quantitative Biology*, **34**, 1-16.

Zaniewski, R., Petkaites, E. and Deutscher, M.P. (1984) A multiple mutant of *Escherichia coli* lacking the exoribonucleases RNase II, RNase D, and RNase BN. *J. Biol. Chem.*, **259**, 11651-11653.

Zavialov, A.V., Buckingham, R.H. and Ehrenberg, M. (2001) A posttermination ribosomal complex is the guanine nucleotide exchange factor for peptide release factor RF3. *Cell*, **107**, 115-124.

Zubay, G. (1973) *In vitro* synthesis of protein in microbial systems. *Annu. Rev. Genet.*, **7**, 267-287.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridisation prediction. *Nucleic Acids Res.*, **31**, 3406-3415.

Acknowledgements

I am glad to thank Prof. Knud Nierhaus for giving me the opportunity to work in his group, for his supervision and kind concern. I am very thankful for the knowledge I have gained on the ribosome during the timeframe of my thesis.

I would like to thank the entire group, who are more than colleagues to me and also for their kind collaboration I thank Detlev, Edda, Oliver, Pavel, Viter, Witold, Yan, and Yoshika. Together with Witold we were working on the optimisation of the *in vitro* systems and sharing the tricks we learned. I thank him for his participation and comments on the results (sections 3.6.2 and 3.6.3).

I am grateful to all other ribosomologists on our floor, to Daniel, Joerg, Klaus, Paola, Sean and Uschi.

I am glad that Evelyn, Tanya and Pasha, Alix, Diana, Julie and Handan visited our laboratory. My special gratitude for their visits to Andrei, Olga and Peter; who are coming from the laboratory of Prof. Dr. Olga Dontsova in Moscow.

I am thankful to Gabriele for her help and advices from the very beginning of my arrival to Berlin. Thankfulness to librarians for their attention in providing the required literature during the work frames of my thesis. I also would like to thank all people at the workshop.

I am grateful to all my friends and neighbours in the fraternity "Turnerschaft Berlin im CC" for their hospitality and kind understanding.

A very special thank is to my parents who cheered me a lot, and to my younger brother Tair for his joyfulness. I was very happy that they could visit me in Berlin.

CV OF MADINA ISKAKOVA

Max-Planck-Institut für Molekulare Genetik

AG Ribosomen, Prof. Knud H. Nierhaus

Ihnestr. 73 Tel.: + 49 – 8413 – 1587

D-14195 Berlin, Germany Fax: + 49 – 8313 – 1690

iskakova@molgen.mpg.de

PERSONAL INFORMATION

- Birth date: Jan. 14, 1979
- Birthplace: Alma-Ata, USSR
- Citizenship: Republic of Kazakhstan
- Fluent in Russian and in English, proficient in German

EDUCATION

- Ph. D. in Molecular Biology. Under the supervision of Prof. Knud H. Nierhaus. Dissertation entitled "Optimisation of a transcription-translation coupled *in vitro* system." Max-Planck-Institute for Molecular Genetics, Berlin, Germany, May 2005.
- M. D. in Biotechnology. Thesis title: "Cloning and structural and functional analysis of potato virus Y genomic RNA 3'-untranslated region." Kazakh National State University of Al-Farabi, Almaty, Kazakhstan, July 2001.
- Training Course in Biodiversity organised for six months by the United Nations University in Gent University, Gent, Belgium, May 2000.
- B. S. in Biotechnology. Kazakh National State University of Al-Farabi, Almaty, Kazakhstan, July 1999.
- Secondary school No. 120 with deep study of English, Almaty, Kazakhstan, June 1995.