universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

A novel transcriptional termination signal capable to terminate T7 polymerase mediated transcription more efficiently

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer.nat.)

Verfasser:	Wittwer Alexander
Studienrichtung /Studienzweig:	A 441 Diplomstudium Genetik-Mikrobiologie
Betreuerin / Betreuer:	Dipl.-Ing. Dr.nat.techn. Univ.-Prof. i.R. Karl Bayer, Ao. Univ.-Prof. Dipl.-Biol. Dr. Angela Witte

Wien, am 10/03/10

Abstract

High -level protein production is the key objective of many biological applications and plays a central role in production of pharmaceutically active substances in an industrial scale. For many reasons *E. coli* is still one of the widely used hosts for recombinant protein production and plasmid based expression systems are very common. Most of the expression systems rely on genetic elements originally stemming from bacteriophage T7 and the commercially available pET30a system has gained increasing popularity in recent years. Foreign gene sequences are cloned under the control of the very strong T7 promoter and transcription mediated by the highly processive T7 polymerase gets stopped by a downstream located T7 termination signal. Plasmid encoded and T7 based expression systems provide high gene dosage and a high transcript level of cloned gene sequence. That often accompanies with enhanced recombinant protein yield, resulting in the accumulation of foreign protein making up 10%-30% or even more of total cellular protein content.

Such strong plasmid based T7 expression systems also bear some drawbacks, resulting in a decrease of overall protein yield. The highly induced expression of recombinant proteins is enforced at the expense of the translation of inherent host proteins, leading to a cellular state similar to that observed under starvation conditions. Further recombinant protein production finally results in a total overburden of the host metabolism ending up in cell death. Thus, high production rates can only be maintained for a very short time period, resulting in limited amount of recombinant protein yield. Continuous research strongly addresses a better exploitation of host cell's potential to produce recombinant protein in order to increase product yield, making the whole process more cost-efficient. The key objective of the underlying work was the increase of transcriptional termination occurring at the plasmid encoded T7 terminator (TE of 80%) in order to further improve the efficiency of recombinant plasmid encoded protein production.

The combination of an altered artificially synthesised T7 termination signal with two well known transcriptional terminators (rrnBT1 and T7) showed a remarkable increase of termination efficiency (TE). That novel termination region was called tZENIT and revealed a TE of nearly 100%. To characterize putative effects of an enhanced termination signal on product yield and process stability, several fed batch cultivations were carried out. Fermentation of a *E. coli* HMS174(DE3) strain carrying a pET30a derivative containing the tZENIT region showed a tremendous decrease of plasmid copy number (PCN) under standard conditions. That reduced plasmid level was maintained over the whole process. Bioreactor cultivation of the same production strain operated under reduced induction strength putatively revealed higher specific production rates and an increase of cell densities, both resulting in a higher yield of recombinant protein.

Summing up, a more efficient termination signal positively influenced plasmid stability and was shown to bare the potential to influence plasmid replication control. Lowering the gene dosage after induction also enhanced recombinant protein yield and allowed cultivation at higher cell densities.

Kurzfassung

Für viele biologische Fragestellungen ist die Bereitstellung von ausreichenden Mengen eines bestimmten Proteins von zentraler Bedeutung. Insbesondere für die Herstellung von pharmazeutisch aktiven Substanzen in großen Mengen ist die rekombinante Proteinerzeugung eine äußerst wichtige Technologie. Bis heute ist *E. coli* der am häufigsten eingesetzte Wirtsorganismus und Plasmid basierende Expressionssyteme finden noch immer weite Verbreitung. Die meisten Systeme basieren auf genetischen Elementen, die ursprünglich vom Bakteriophagen T7 stammen. Das kommerziell vertriebene pET Expressionssytem erfreut sich immer größerer Beliebtheit und war auch Gegenstand dieser Arbeit. Bei dem Plasmid basierenden pET-System wird das fremde Gen unter den Einfluss des starken T7 Promoters gestellt, die Transkription wird durch die äußerst aktive T7 Polymerase bewerkstelligt und eine T7 Terminationssequenz am 3' Ende bewirkt das Abfallen des Syntheseenzyms. Diese Anordnung führt in den meisten Fällen zu einer sehr hohen Transkriptionsrate des fremden Genes, und durch die Lokalisation der fremden Sequenzen auf Plasmiden wird auch ein hohe Gendosis erreicht. Diese Besonderheiten gehen in aller Regel auch mit einer hohen rekombinanten Proteinausbeute einher und erlauben eine starke Anreicherung des fremden Proteins von bis zu 10%-30% der gesamten zellulären Proteinfraktion.

Obwohl noch immer sehr weit verbreitet, weisen diese starken Plasmid basierenden Expresionssysteme auch eine Reihe von Nachteilen auf, die schlussendlich auch die Ausbeute an rekombinanten Protein stark vermindern können. Eine solch starke Expression von zellfremden Proteinen wirkt sich meist nachteilig auf die Herstellung von zelleigenen und essentiellen Proteinen aus, sodass die Wirtszelle aufgrund dieser übermäßigen Belastung sogar in einen Zustand versetzt wird, der jenem durch Hunger induzierten Stresszustand sehr stark ähnelt. Eine andauernde Ausbeutung der Zellmaschinerie zur rekombinanten Proteinerzeugung kann schlussendlich zu einem totalen Zusammenbruch des Zellmetabolismuses führen und in letzter Konsequenz bedingen all diese Ereignisse den Zelltod. Aufgrund dieser übermäßigen metabolischen Belastung der Wirtszelle kann die Herstellung des Fremdproteins nur für relativ kurze Zeit aufrecht erhalten werden und dies mindert daher die gewonnene Ausbeute an Fremdprotein. Die stetige Forschung in diesem Bereich fokussiert sich daher sehr stark auf eine bessere Ausnutzung des Zellpotentials zur rekombinanten Proteinsynthese, um dadurch höhere Ausbeuten zu erzielen. Aufgrund des Kostendrucks im freien Markt wird auch von Seiten der Industrie eine Effizienzsteigerung gewünscht, um dadurch die rekombinante Proteinherstellung ökonomischer zu gestalten.

Die zugrunde liegende Arbeit ist um eine Verbesserung des Plasmid basierenden pET Systems bemüht. Ziel der Bestrebungen war es die transkriptionelle Termination am Plasmid kodierten T7 Terminator (TE 80%) zu steigern. Dabei zeigte eine Kombination aus einem künstlich hergestellten T7 Abkömmling mit zwei bereits bekannten intrinsischen Terminatoren (rrnBT1 und T7) eine bemerkenswerte Erhöhung der Terminationseffizienz (TE). Diese neuartige Terminationsregion wurde tZENIT Region genannt und zeigte ein TE von nahezu 100% auf. Um nun mögliche Auswirkungen eines verbesserten Terminationssignales auf Aspekte wie Produktausbeute und Systemstabilität zu bestimmen, wurden mehrere fed batch Kultivierungen durchgeführt. Die Fermentation eines *E. coli* HMS174(DE3) Produktionsstammes, welcher einen pET30a Vektor mit der entsprechenden tZENIT Region enthielt, zeigte unter Standardbedienungen eine drastische Reduzierung der Plasmidkopienzahl (PCN) pro Zelle und dieser erniedrigte Level an plasmidischer DNA

wurde über den gesamten Prozess aufrecht erhalten. Der selbe Produktionsstamm wurde einer weiteren Kultivierung unter reduzierter Induktionsstärke unterzogen. Die daraus resultierenden Ergebnisse zeigten neben der bereits beobachteten PCN Reduktion auch eine deutliche Steigerung des spezifischen Produktgehaltes, und die Wirtszellen waren in der Lage weit höhere Zelldichten zu erreichen. All diese hinsichtlich einer Produktsteigerung positiven Effekte führten schlussendlich zu einer deutlichen Erhöhung des Ertrages an rekombinanten Protein.

Acknowledgement

Als es daran ging eine Arbeitsgruppe zur Durchführung meiner Diplomarbeit zu finden, bin ich das kleine Risiko eingegangen, die Arbeit unter der Schirmherrschaft eines mir aus Vorlesungen oder Übungen nicht bekannten Professors an einer anderen Universität zu verfassen. Ziel dieser Bestrebungen war es näher an für die Industrie relevante Themengebiete zu forschen und zu arbeiten. Nach über einem Jahr intensivster Arbeit kann ich mich nur bei Prof. Karl Bayer und der gesamten Arbeitsgruppe mikrobielle Fermentation von ganzem Herzen bedanken, dieser Schritt hat sich in mehrfacher Hinsicht ausgezahlt. Aufgrund der einzigartigen Einbettung der Arbeitsgruppe in das Forschungsprojekt 'Austrian Center of Biopharmaceutical Technology (ACBT)' konnte ich tiefe Einblicke in industrienahe Forschungstätigkeiten gewinnen. Auch in einer mir zu Beginn nahezu gänzlich unbekannten Materie, der mikrobiellen Fermentation und deren Abläufe habe ich mir im Laufe meiner Tätigkeit zusätzliches Wissen aneignen können.

Nicht nur hinsichtlich der Arbeitsgruppe und Arbeitskollegen hatte ich großes Glück, sondern auch die mir gestellte Aufgabe entpuppte sich als Glücksgriff. Obwohl ich schon von Anfang an ein gewisses Potential zur Verbesserung des bestehenden Systems mit Hilfe dieses Ansatzes sah, konnte keiner ahnen, dass solche Steigerungen hinsichtlich Plasmidstabilität und Produktausbeute möglich sind. Zu allem Überdruss wird momentan auf Grundlage der Ergebnisse dieser Diplomarbeit ein Patent eingereicht und scheint auch relativ große Chancen einer tatsächlichen Realisierung zu haben, eine Tatsache die ich zu Beginn meiner Arbeit nicht für möglich gehalten hätte.

Zu guter Letzt möchte ich mich noch bei zwei Personen explizit bedanken. Zum einem gilt ein großes Dankeschön Jürgen Mairhofer, der für alle molekularbiologischen Fragen der erste Adressant war und mir immer wieder helfend zur Seite stand. Des Weiteren möchte ich mich bei Gerald Striedner bedanken, der mir in meiner Arbeit sehr viele Freiheiten gewährte und mir auch eine fixe Anstellung ermöglichte.

Table of contents

List of Figures

List of Tables

1 Introduction

1.1 Recombinant protein production

For many applications like biochemical studies, crystallographic analysis or for biotechnological and medical use, a sufficient and highly pure amount of distinct protein is necessary. Nowadays a range of expression systems and a variety of hosts ranging from prokaryotes, over low eukaryotes to mammalian or insect cells are available and ensure high level production of recombinant proteins. The choice of an adequate system strongly depends on protein encoded properties, thus that decision has to be done anew in regard to different proteins. For many reasons *E. coli* is still one of the widely used hosts for recombinant protein production. Intensive research over the last century have evoked deep insight in its genetics and metabolism, making that prokaryote to the best understood organism. Since 1997 the whole genomic sequence is accessible and most of the coding sequences are also functionally characterized. A set of mutant strains aid to circumvent some problems resulting from the exertion of *E. coli* as host organism. For instance strains deficient to produce functional proteases avoid the proteolytic cleavage of foreign proteins. Mutant strains over-expressing chaperons aid in correct folding of produced recombinant protein. Further advantages for the application of prokaryotes as host organisms are their minor requirements on nutrition media composition and their rapid growth rates.

Nevertheless the use of prokaryotic production systems is mainly restricted by the absence of adequate post-translational modification mechanisms in *E. coli* or other prokaryotes. Especially a lot of eukaryotic proteins have to get modified by addition of glycosyl groups in order to adopt their fully biological function. The recombinant production of complex glycoproteins has to be carried out in lower eukaryotes like yeasts (e.g. *Pichia pastoris)* or even in mammalian or insect cells. The remaining introduction part is directed to plasmid based expression systems for recombinant protein production in *E. coli*, the kind of system also used in underlying diploma thesis.

1.2 Plasmid based Expression systems

One of the great advantage of E. Coli as host for heterologous protein production is the variety of available plasmids. Plasmids are categorized into four classes according to their average copy number within cell. Very high copy number vectors are existent in more than 100 copies per cell, high copy number vectors in 15-60 copies, medium copy number vectors in about 10 copies, and low copy number plasmids in 1 or 2 copies per cell.

The crucial attribute of a plasmid to ensure a stable plasmid copy number is the origin of replication (ORI). In modern biotechnology mainly two replicons, namely ColE1 and p15A, are used in expression vectors (Baneyx, 1999). Replicon p15A is derived from plasmid pACYC184 and ensures a stable copy number of 10-12 plasmids per cell. The widely dispersed ColE1 replicon is present on pBR322 derivatives (copy number 15-20) or on pUC vector series. When stemming from pUC plasmid family, ColE1 ensures a plasmid copy number (PCN) of 500-700. Under laboratory conditions (selective pressure) plasmids are stably replicated, and therefore plasmids carrying the gen sequence of desired recombinant protein assure high gene dosage, and that often accompanies with high production yields.

The relative low loss rate (10^{-5}-10^{-6} per generation) of plasmids carrying one of the both mentioned replicons can increase tremendously when conditions strongly favour cells carrying no plasmid. That for example could be the case, when the cloned gene encodes a protein which is toxic for the host, or otherwise negatively interferes with the host metabolism, resulting in reduced growth rate. For instance over expression of membrane proteins often results in cell death.
Plasmid loss also plays an important role, when cells are cultivated in a continuous process or other conditions resulting in high cell densities. A simple way to overcome that problem is the addition of antibiotics into growth medium, and the simultaneous use of antibiotic resistance genes on expression vectors. In that case only cells carrying plasmids will survey in an antibiotic saturated environment.

For laboratory practice the utilization of antibiotics is a very powerful and in most cases sufficient tool, but the system also exhibits some drawbacks, especially in regard to recombinant protein production of pharmaproteins in an industrial scale. The effect of antibiotics decreases with time, because the substances get degraded or even inactivated. Especially for medical use a contamination of the product with antibiotics is unacceptable.

To circumvent the use of antibiotics a number of alternatives have been developed. Most of them rely on genetic complementation (Baneyx, 1999), whereby host cells carry gene knock outs of essential genes, and only can survive when harbouring plasmids supplemented with the essential gene sequence. A similar strategy is the use of repressor carrying cloning vectors. Upon plasmid loss repressor molecules are missing, expression of toxic proteins is induced, and finally cell death takes place. Whether all these systems are functionally, all those alternatives require the correct expression of putative plasmid encoded genes, therefore increase the total metabolic burden of the host, and that often results in lower recombinant protein yield. Most of those systems also require special growth medium composition and specifically designed host strains, making these alternatives not very attractive for an industrial purpose (Miroux and Walker, 1996).

For a sufficient yield of recombinant protein one have to consider several aspects when designing a recombinant expression system. Aspects like the type of used replicon, the transcriptional promoter, the kind of host strain, mRNA stability, translational issues like the sequence of ribosome binding site (RBS), and finally the properties of the translated recombinant protein themselves have a great impact on total product yield. It would go beyond the scope of that diploma thesis to deal with all those aspects, but in the following the origin of replication and the promoter are mentioned in more detail.

1.2.1 Plasmid encoded origin of replication

Each cloning vector contains a special kind of replication origin, which determines the plasmid copy number and therefore the gene dosage of the recombinant protein. The terms replicon and ori are often used as synonyms. Replicons contain the information necessary to start and end the replication of DNA (Benjamin Lewin. Genes IX. Jones and Barlett Publishers, Inc. Ninth edition. 2008. S 422ff). Within each replicon the synthesis of new DNA (replication) starts at a defined location called origin of replication (oriV or ori). The origin of replication is usually about 300 bp long and is recognized by specific proteins that initiate replication. These proteins may be chromosomally or plasmid encoded. All DNA polymerases require a primer, a short chain of single stranded DNA or RNA, before replication can begin. In most cases this primer is a short chain of RNA nucleotides that is produced by a RNA polymerase. This primer hybridizes to the single stranded DNA at the oriV and afterwards gets extended by a DNA polymerase in the 5' to 3' direction. Replication can be either unidirectional with only one replication fork proceeding from the oriV or bidirectional with two replication forks proceeding from the oriV in opposite directions. Plasmids with the same replicon belong to the same incompatibility group and cannot coexist in the same cell.

Plasmid DNA, phage DNA as well as chromosomal DNA of bacteria are characterized by just a single origin of replication. In contrast chromosomes of eukaryotes contain a set of replication origins. The strict regulation of initiation of plasmid replication ensures a tight control of plasmid copy number. As regulation of plasmid initiation by restricting the amount of available primer for the initiation of DNA replication is realized by ColE1 derived replicons, and many of the most popular cloning vectors exhibit that replicon, in the following the control mechanism of ColE1 replicon is explained in detail.

1.2.1.1 Regulation of ColE1/pMB1 plasmid copy number

Many of the plasmids currently in use carry a replicon from the naturally occurring plasmid pMB1 or plasmid ColE1 (http://homepage.mac.com/gmunson/VooDoo/plasmidreplicons.html). Bacterial cells harbouring the ColE1 plasmid are able to produce a special type of bacteriocin, the so called colicins. Those substances are produced by some E. Coli strains and are able to kill other closely related E. Coli strains. Both types of plasmids were independently isolated, but have extensive functional and structural homologies, including the same origin of replication and thus the same mode of replication. Plasmids with the pMB1 or ColE1 replicon are maintained at 15-30 copies per cell. The pMB1 replicon, carried by pBR322 and its derivatives, is incompatible with the ColE1 replicon because of their near identity.

The pMB1/ColE1 replicon consists of the origin of replication (oriV), and three genes encoding two regulatory RNAs (RNA II and RNA I) and the replication protein Rop (see Fig. 1.1). RNA II function as activator of replication by serving as primer at which DNA synthesis is initiated. Plasmid replication starts with the transcription of a 555bp long so called pre-RNA II molecule, which is encoded directly upstream of the replication origin. When RNA polymerase reaches oriV, the transcript gets cleaved by the enzyme RNAse H resulting in the generation of a transcript with a free 3'-OH end (RNA II). The generated RNA II fragment is tightly bound to DNA and is located directly upstream of the origin (around position -20). The stable hybrid of DNA and RNA forms a short double stranded region at which DNA synthesis is initiated.

In contrast RNA I negatively regulates plasmid replication. The molecule shows a length of 108 bases and is encoded within the same DNA section as RNA II, but is located in opposite direction. Because of their position on the same DNA fragment RNA I and RNA II are complementary to each other, thus are capable to undergo base pairing.

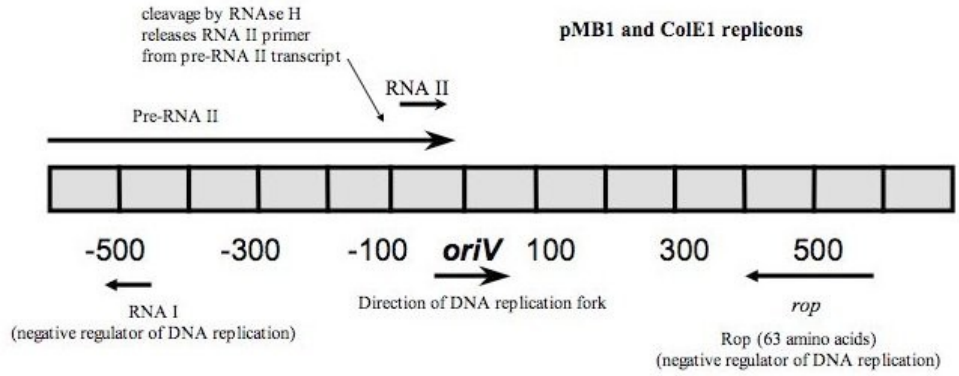


Fig. 1.1. ColE1/pMB1 type replicon. The most common replicon in modern gene technology, utilized in all pBR322 derived plasmids. It consists of the origin of replication, oriV, and three genes encoding two regulating RNAs (RNA II and RNA I) and the replication protein Rop.

Through that capability of RNA I to bind at the 5' region of RNA II the control of plasmid copy number is carried out. Because of their complementary sequence region of 108 bases, both molecules would interact immediately after their transcription. To prevent base pairing of the two counter-transcripts, both molecules form characteristic secondary structures during their transcription. Each RNA molecule constitutes three characteristic hairpin structures, thus possesses stem regions of double stranded RNA. Those stem structures reduce the number of bases, capable to undergo base pairing.

Both regulatory RNAs are constitutively transcribed from their own promoter sequence ($P_{RNA\ I}$ and $P_{RNA\ II}$). RNA I function in *trans* and diffuses into the cytoplasm. In contrast RNA II function in *cis* as primer for replication. A low plasmid copy number (PCN) accompanies with a low amount of the inhibitor RNA I, thus PCN increases. Initiated plasmid replication leads to an increment of RNA I gene dosage, and therefore the amount of available inhibitor RNA increases with time. A higher concentration of RNA I molecules within cell makes a binding event to the 5' part of RNA II more probable. That binding causes a change in RNA II and that structural change propagates down to the 3' end of RNA II and prevents cleavage of the primer RNA by Rnase H. As result transcription runs on, and plasmid replication gets repressed (Benjamin Lewin. Genes IX. Jones and Barlett Publishers, Inc. Ninth edition. 2008. S 422ff).

Proliferating bacteria are characterized by a high rate of cell division events. Before division into two separate daughter cells occurs, the original cell increases its cell volume, thus less RNA I molecules per defined cell volume are present. That thinning out of

inhibitor RNA I leads to an increment of unpaired RNA II primer molecules, and that is responsible for the raise of plasmids during cell growth. Finally that mechanism guarantees a stable plasmid copy number of about 15 within both daughter cells.

The rop protein is also encoded within ColE1 replicon and gets constitutively transcribed from its own promoter Prop. The Rop (repressor of primer) protein comprises 63 amino acids and is completely conserved in pMB1 and ColE1. Rop acts in trans and stabilizes the binding of RNA I to RNA II, thus reduces replication probability and copy number of plasmid. Rop isn't essential for a functional replicon and indeed the pUC family of expression vectors lacks that protein. The absence of the rop gene in pUC vector series increases plasmid copy number 3 to 4 fold. In addition to the missing of rop gene, pUC plasmids exhibit a single nucleotide substitution that reduces RNA I function, resulting in a putative increase in plasmid copy numbers of 500-700 copies per cell.

1.2.2 The transcriptional promoter

Promoters in systems for recombinant protein production require several features. Transcription at these sequences should be initiated very frequently in order to assure a high mRNA level of the cloned gene fragment. A high amount of mRNA molecules often directly correlates with high yields of recombinant protein. These strong promoters used in modern biotechnology should be easily inducible with preferable cheap and non toxic substances. In the absence of an inducer there should be no remarkable basal transcription. Leaky promoters are especially inapplicable when the cloned gene encodes toxic proteins. But even non-toxic proteins can have harmful effects on host's viability when expressed in higher levels. In that regard membrane proteins are often mentioned. A high concentration of membrane spanning proteins can lead to cell death, because they destroy the fluidity of host's inner membrane, resulting in its clogging.

Because of the fact that the genetic regulation of the lac operon has been intensively investigated since the late fifties, the underlying genetic mechanisms are well known. Hence many of the promoters used in gene technology have been constructed from lac-derived regulatory elements. Also the expression system used in this diploma thesis utilizes lac-derived genetic elements, and therefore the lac operon as paradigm of prokaryotic regulation is explained in the following.

1.2.2.1 The *lac* operon

The organization of several genes in a linear progression within operons is a very typical gene arrangement in prokaryotic genomes. Operons often comprise genes coding for enzymes of one metabolic pathway, thus are functionally related and are needed in similar amounts. All genes within a operon are commonly transcribed as part of one polycistronic mRNA. Despite structural genes a operon also comprises elements, responsible for gene expression regulation of the whole operon. Thus the activity of the operon is tightly controlled by regulatory gene products interfering with cis-acting control elements.

Among the prokaryotic operons the catabolic operons like the gal, lac, and ara operons are well studied. Genes encoded in those operons enable E. Coli to utilize alternative sugar sources despite glucose. To avoid a waste of energy by producing enzymes for the utilization of alternative sugar substrates in the presence of glucose, the expression of those catabolic operons is tightly regulated.

a) The structure of the *lac* operon

The lac operon comprises 6000bp in total and encodes three structural genes, namely lacZ, lacY, and lacA (see Fig. 1.2). lacZ encodes β-galactosidase (LacZ), an intracellular enzyme that cleaves β-galactosidic bounds of sugar components. Its main role is to cleave the disaccharide lactose into glucose and galactose. As an important by product the enzyme also produces the global regulator β-1, 6-allolactose.
lacY encodes for the transport protein β-galactoside permease, a 30-kD membrane-bound transport protein that pumps lactose into the cell. lacA encodes β-galactoside transacetylase, an enzyme that transfers an acetyl group from acetyl-CoA to β-galactosides.

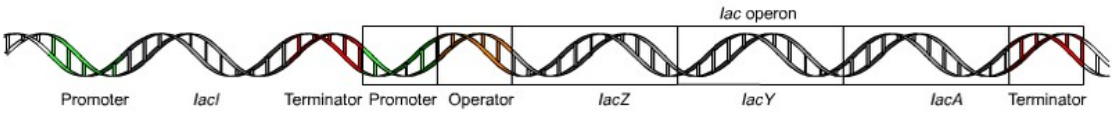


Fig. 1.2. Organization of *lac* operon. The operon comprises about 6000bp in total. The structural genes *lac*Z, *lac*Y, and *lac*A are organized in this order within the operon, and are commonly transcribed from one promoter at the 5' end. Transcription of those genes is stopped by a common transcriptional terminator and results in a long polycistronic mRNA. The operator sequence directly follows the promoter sequence of structural genes. The gene encoding the repressor molecule *LacI* is located upstream of the structural gene cluster. It comprsises its own transcriptional promoter and terminator.

All three structural genes are transcribed starting from one common promoter (P_{lac}), resulting in a long polycistronic mRNA (Benjamin Lewin. Genes IX. Jones and Barlett Publishers, Inc. Ninth edition. 2008. S 304ff). Transcription of structural gene cluster is stopped by a common transcriptional terminator. At basic state transcription of structural genes is turned on, but expression of the lac operon is negatively regulated, because the repressor molecule LacI turns off the expression of structural gene cluster. LacI is also encoded within the operon, lies directly upstream of the structural genes, but has its own transcriptional promoter and terminator. *lacI* is constitutively transcribed equal if lactose is present or not. The proximity of *lacI* and the genes coding for the structural genes isn't required at all, because *lacI* encodes for a diffusible and *trans*-active regulator, thus could be located elsewhere within bacterial genome. The functional repressor molecule is constituent of a tetramer of identical subunits. Under normal conditions about 10 repressor tetramers are present in wild-type cells.

b) Genomic region of *lac* promoter

In its active tetrameric form LacI is able to bind to a palindromic sequence, the so called operator (O1), which is in proximity of P_{lac}. In detail the operator sequence overlaps with the structural gene promoter (see Fig. 1.6), thus both regulatory elements share a common DNA fragment. In the past it was thought, that binding of the repressor inhibits RNA polymerase to interact with the promoter, because part of it is occupied by the repressor molecule. Further investigation revealed, that the bound repressor doesn't impair the binding of RNA polymerase to P_{lac}, in contrast the interaction even gets reinforced. In fact a repressor molecule bound to its operator impairs the melting of both nucleotide strands at the start point of transcription, thus prevents RNA polymerase from initiating transcription at the promoter. Despite O1 the operon exhibits putative operator sequences, namely O2 and O3. When designating the base at which transcription starts with +1, O2 is located more upstream at position +410 and O3 (-83) is positioned more downstream relative to O1 (see Fig. 1.3). Both alternative operator sequences exhibit the ability to bind repressor tetramers. Strongest repression takes place when a repressor tetramer simultaneously binds to O1 and to either O2 or O3.

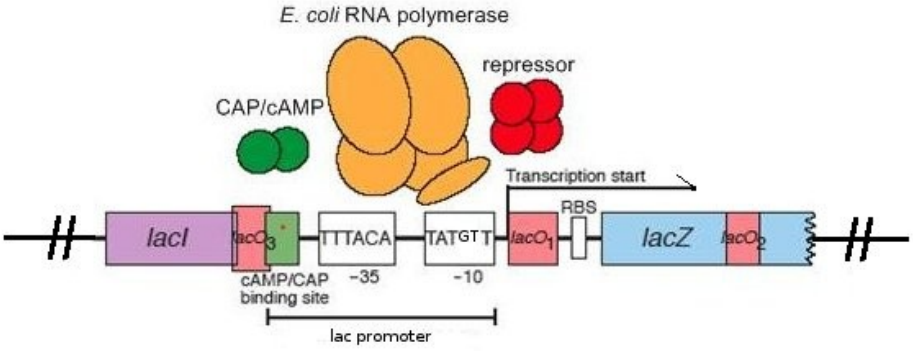


Fig. 1.3. Genomic region of the *lac* promoter. The *lac* promoter region containing -35 and -10 box is recognized by RNA polymerase. Melting of DNA strands is inhibited by bound repressor tetramers. *LacI* primarily binds to *lac*O1 operator sequence. One putative operator element (O2) is present more downstream within *lac*Z gen and a second more upstream in near proximity to CAP-binding site.

c) Induction of the *lac* operon

In the absence of β-galactosides (mainly lactose) in growth medium it would be a waste of energy to produce β-galactosidase, thus cells growing under this condition contain very few molecules (basal transcription) of that enzyme. Only in the presence of lactose and in the coevally absence of the main sugar source glucose it would make sense to express β-galactosidase. In that case the substrate of the enzyme itself has the ability to turn on expression of the enzymes, capable to metabolize it. Those substances are called inducers. The natural inducer of the *lac* operon isn't lactose, rather it's the by product of β-galactosidase activity, the global regulator β-1, 6-allolactose. Because small amounts of β-galactosidase are even present in the absence of lactose, the inducer is rapidly formed when lactose suddenly appears in medium. Allolactose in turn doesn't accumulate, because it also gets metabolised by β-galactosidase. Some similar substances, primarily the analogue isopropyl-β-D-thiogalactopyranoside (IPTG), also have the ability to function as inducer for the *lac* operon, but in contrast those substances aren't metabolised by β-galactosidase. Hence those alternative substances are excellent inducers, because they are needed in relative low concentration for fully inducing the operon. Nevertheless especially in regard to large scale industrial fermentation processes IPTG is relatively expensive.

Target of the inducer molecules is the tetrameric repressor *LacI,* which comprises an additional binding site for the inducer substance. Attachment of the inducer molecules induces a conformational change within the protein, finally resulting in reduced affinity

of the repressor to the operator. Such kind of control is also called allosteric control. As result the repressor diffuses from the control element and therefore enables the RNA polymerase mediated transcription of the structural operon genes.

d) The cAMP/CAP complex

In the case of the simultaneous presence of glucose and lactose in the growth medium it wouldn't make sense to fully induce the *lac* operon (see Fig. 1.4). The fine tuning of β-galactosidase gene expression is accomplished by a putative positive regulation of the operon. In near proximity of P_{lac} a further regulatory DNA sequence is present, namely the CAP site (see Fig. 1.3). Its consensus sequence is able to bind the most widely acting activator in *E. coli*, the catabolite activator protein (CAP or sometimes also named CRP (cAMP receptor protein)). In its active form the CAP protein attaches to the CAP site. Furthermore bound CAP interact with the carboxy terminal part of RNA polymerase, ending up in an increased affinity of the polymerase to its promoter. For its active form CAP has to bind to the very small signal molecule cAMP. *E. coli* uses the global starvation molecule cAMP to sense the availability of the main sugar resource glucose. Cyclic AMP (cAMP) is synthesized by the enzyme adenylate cyclase, which utilizes ATP to produce cAMP. The absence of glucose leads to stimulation of adenylate cyclase, resulting in a higher level of cAMP. In the presence of glucose the adenylate cyclase enzym isn't efficient in producing cAMP, thus cAMP level decreases. Hence *lac* operon is just fully induced when glucose is absent and a simultaneous presence of the alternative sugar compound lactose is warranted (see Fig. 1.4). When that is the case, cAMP level is high, active CAP (activator of transcription) binds to CAP site, and because of the existence of inducer allolactose *LacI* isn't able to bind to the operator.

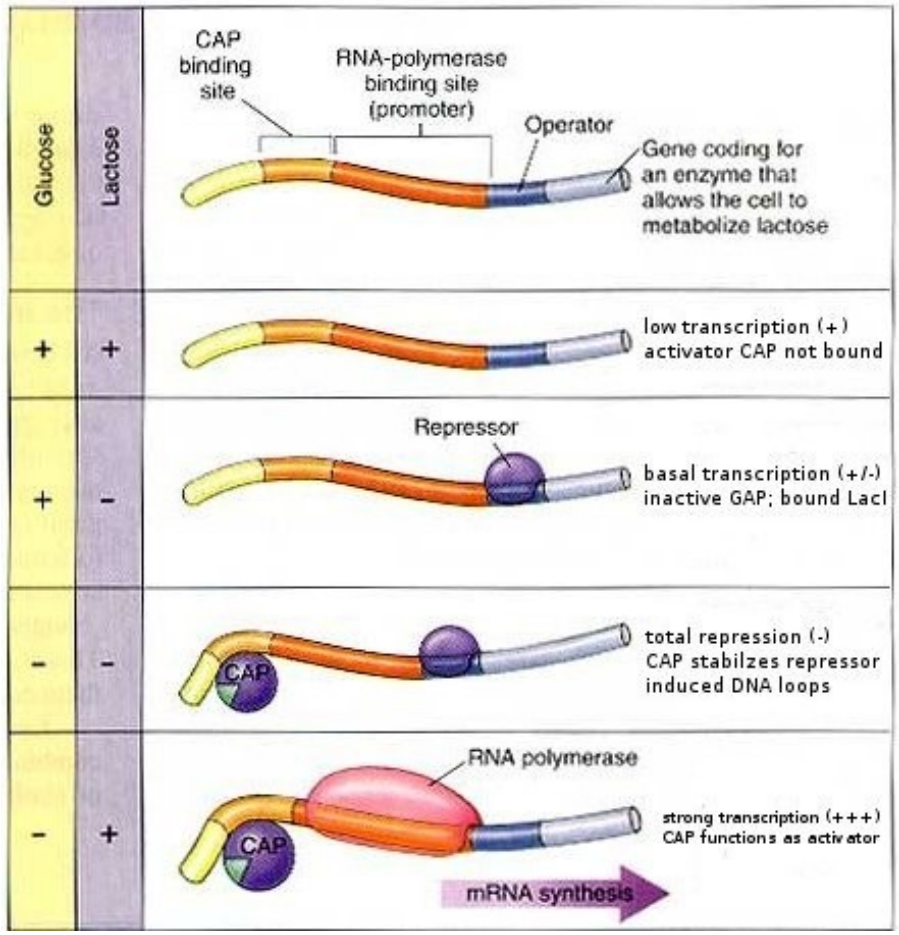


Fig. 1.4. Regulation of the *lac* operon depending on the availability of distinct sugar components. In the presence of lactose and the simultaneous absence of glucose the structural genes of the operon are highly transcribed. Fully repression of the operon is achieved when both sugar compounds are absent.

e) The fully repressed configuration of the *lac* operon

In the case that neither glucose nor lactose exists in the medium the strongest repression of the *lac* operon is observed. That's somewhat surprising, because under that conditions the active form of activator CAP should be available. That phenomenon could be explained by the two putative operator elements mentioned above. When examining the tetrameric form of *LacI* in more detail an assemble of two dimers forming the tetramer could be observed. Each dimer has the ability to bind to one operator. Thus one repressor can interact with two operator sequences. To achieve such a multiple operator-interaction a bend within the DNA between those regulatory sequences has to be introduced. As result a DNA loop is formed, and that structure gets stabilized through active

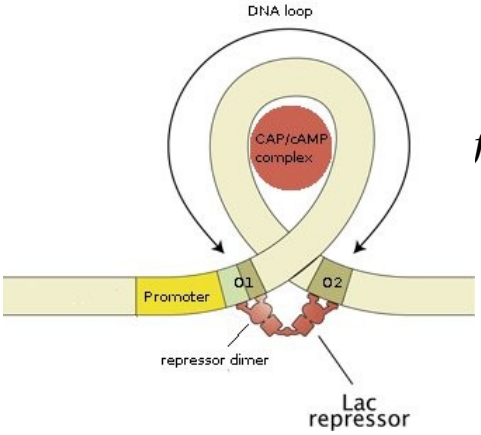


Fig. 1.5. *lac* promoter region in the absence of glucose and lactose. One repressor tetramer consisting of two dimers binds coevally two operator sequences, resulting in a DNA loop. The structure gets stabilized by an active cAMP/CAP complex.

CAP molecules (see Fig. 1.5). Formation of such a compact DNA arrangement impairs even basal transcription at P_{lac}.

f) Catabolite repression

The sensitiveness of the original *lac* promoter to a cAMP dependent regulation known as catabolite repression limits its usefulness in modern biotechnology. In case of the *lac* operon the presence of glucose reduces the transcription rate of operon encoded structural genes due to a decrease in active cAMP/CAP complexes. As result of that regulation gene expression of genes under the control of the *lac* promoter strongly differ according to different growth rates and medium composition. Catabolite repression is exerted on a variety of operons encoding catabolic enzymes and is mostly accomplished by intervening into cAMP-CAP system, thus the term cAMP-mediated catabolite repression was coined. Because of its effect on a variety of different operons, the underlying mechanism can't be restricted to distinct regulatory sequences unique for one operon. In all known operons sensitive for catabolite repression, the regulation is exerted within promoter region.

Wild-type *E. coli* cells grown under different conditions can show a high variation of β-galactosidase level, mainly exerted by catabolite repression as mentioned above (Wanner et al., 1978). That variety gets eliminated when cAMP is added to growth medium, suggesting that cAMP is the main mediator of that kind of regulation. Mutating the CAP-site within original *lac* promoter also reduces the observed variety in β-galactosidase expression and makes the promoter insensitive to catabolite repression (Wanner et al., 1977). That kind of insensitive promoters are widely used in biotechnology, because they allow the adjustment of well defined transcription rates, making the whole process more predictable. In the 70th a *lac* promoter mutant, namely *lac*UV5 was isolated. That promoter just carries two single mutations within -10 box (Pribnow-box) compared to original *lac* promoter (see Fig. 1.6), remaining nucleotides are unaffected, thus also the CAP-binding site is the same as in wild-type promoter. The exchange of the two nucleotides within -10 box (see Fig. 1.6) results in a conversion into consensus sequence 5'-TATAAT-3'. Because of its accordance with consensus sequence the strength of *lac*UV5 increased about 2.5 fold. Interestingly the point mutations also lead to insensitivity of

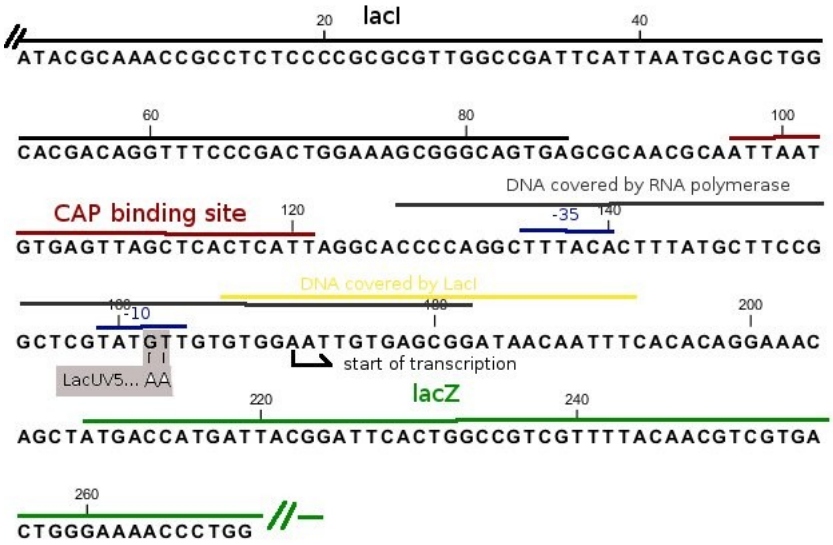


Fig. 1.6. Nucleotide sequence of *lac* promoter region. Important sequences are over lined. The *lac* promoter is located upstream of *lacZ*. As depicted, the DNA sequence covered by RNA polymerase and the binding site for repressor *LacI* share a common sequence fragment, thus the two regulatory regions overlap. The derivative *lacUV5* just exhibits two single mutations within -10 box. The two divergent nucleotides are shadowed grey.

the promoter to catabolite repression. Transcription initialized at the *lac*UV5 promoter is unaffected by cAMP and is largely independent from growth rate or medium composition. Both properties are very useful in modern biotechnology and lead to its intensive application in many cloning vectors.

1.2.3 The pET expression system

Over the last few years the commercially available pET expression (by Novagen, Merck KgaA, HE, Germany) system has gained increasing popularity and importance (see Fig. 1.7). The system was generated by Studier and colleagues (Studier and Moffatt, 1986) and takes advantage of the highly processive DNA directed RNA polymerase stemming from bacteriophage T7. That polymerase is extremely accurate and just transcribes DNA sequences provided with a T7 promoter sequence. Compared to *E. coli* RNA polymerase, T7 polymerase transcribes up to five times faster, resulting in a large amount of mRNA. T7 RNA polymerase is encoded within the host's genome, thus is supplied in *trans*. The gene coding for T7 polymerase is part of the genome (~50kb) of a lysogenized prophage (λDE3). Like other λ phages also λDE3 prefers a distinct attachment site (attBB') for integration into *E. coli* genome.

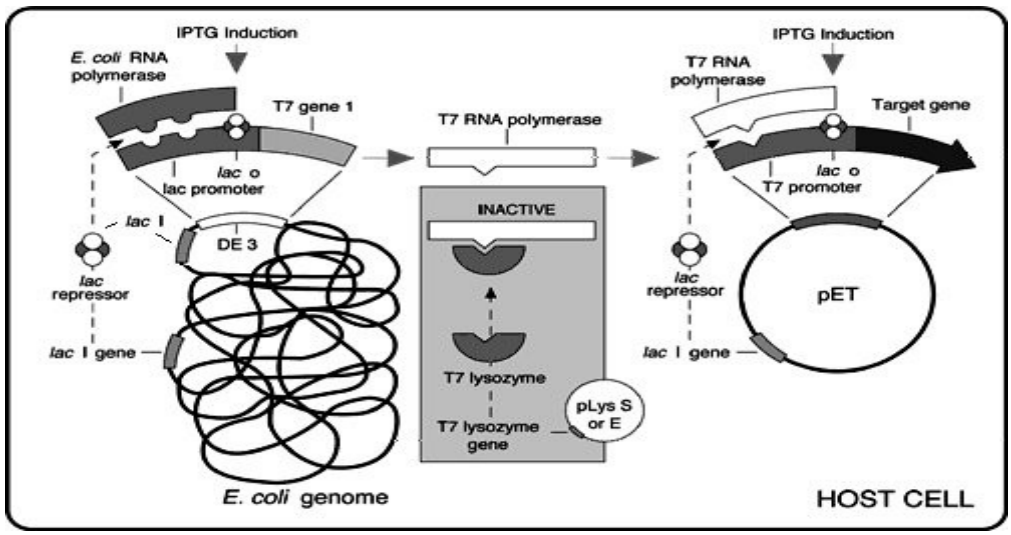


Fig. 1.7. pET expression system. The essential T7 RNA polymerase is delivered in *trans* as part of the λDE3 fragment. Transcription of the gene (T7 gene 1) is under the control of the *lac* promoter derivative *lac*UV5, thus is inducible by addition of IPTG. The gene coding for the recombinant protein is cloned downstream of the *lac*/T7 hybrid promoter, thus also respond to IPTG. To suppress basal transcription the host often gets co-transformed with plasmid pLysS or pLysE, carrying a gene coding for T7 lysozyme.

a) Structure and genomic location

The attachment site attBB' (sequence: 5'-CCGGGCTATGAAATAGAAAAATGAATCCGTTG AAGCCTGCTTTTTTATACTAACTTGAGCGAAACGGGAAGGTAAAAAGACAAAAAGTTGTTTTTAATA -3') lies between the galactose(*gal*)-operon and the biotin(*bio*)-operon. According to the gene annotation of the genome of *E. coli* strain K12 substrain MG1655 (Refseq: NC_000913) the attachment site lies between genes ybhB and ybhC in near proximity to gene bioA. MG1655 is a non pathogenic E. Coli strain which carries very few genetic modifications. The strain just has been cured from lysogenic bacteriophage λ and from F-Plasmid, thus strongly resembles wild type E. Coli. MG1655 in turn derives from substrain W1485, which was obtained by Joshua Lederberg from the original K-12 isolate gained from a patient in 1922.

To immediately turn on transcription of cloned genes, the gene coding for T7 polymerase lies under the control of the IPTG inducible *lacUV5* promoter. Genetic elements surrounding the T7 gene 1 resemble the arrangement within the *lac* operon (see Fig. 1.8). The gene is preceded by a sequence coding for the 5' end of the *lac*Z gen. Although that shortened *lac*Z gene doesn't encode a functional β-galactosidase protein, the fragment contains the important regulatory element O2. As in the original *lac* operon the structural unit contains a upstream located CAP binding site and a putative l*acI* gene, comprising the second alternative operator sequence (O3). In total pET production strains possess three copies of the *LacI* repressor molecule encoding gene. One is present within DE3 phage fragment, one is part of the origianl *lac* operon and the third one is encoded on most of the 40 different available pET vectors (see Fig. 1.7). That high gene dosage is inalienable, because *lac*I is weakly expressed, but a sufficient amount of repressor tetramers is necessary to reduce basal transcription. A 10-fold increase of transcription can be achieved when utilizing the promoter mutant variant $LacI^q$ (Miroux and Walker, 1996).

Genes coding for the desired recombinant protein product are cloned downstream of a T7/*lac* promoter hybrid, encoded by pET expression vectors (medium copy number plasmids). The combining of sequences stemming form the *lac*- and T7-promoter has some advantages (Baneyx, 1999). On the one hand initiation of transcription can be tightly controlled by the well known repressor *LacI* and on the other hand the component stemming from T7 promoter assures, that the promoter is specifically recognized by the highly processive T7 RNA polymerase. Thus the promoter is invisible for *E. coli* RNA polymerase and the leakiness of the original *lac* promoter is extremely reduced. All together that arrangement enables the extraordinary high accumulation of mRNA ori-

ginating from the cloned gene. That high amount of mRNA often accompanies with high product yield, resulting in an accumulation of foreign protein making up 10%-30% or even more of total cellular protein content.

Inducing the pET system with IPTG leads to dissociation of *LacI* repressor molecules from the *lac* operator preceding the T7 RNA polymerase gen at the λDE3 fragment as well as from the operator sequence encoded on the expression plasmid. As result T7 RNA polymerases are available and those enzymes initialize transcription at the plasmid encoded T7 promoter.

b) Drawbacks

Although widely used, the T7 based expression system also exhibits some drawbacks and limitations. As a derivative of the *lac* promoter also *lac*UV5 allows some basal transcription in non-induced state. Because of the extraordinary high processivity of T7 RNA polymerase even few molecules can result in a noticeable amount of recombinant protein in non-induced state, and that could have serious effects on the host's viability, especially in regard to expression of toxic polypeptides (for instance membrane proteins). To avoid background expression the host often gets co-transformed with plasmid pLysS or pLysE (see Fig. 1.7). Those plasmids carry a gene coding for T7 lysozome, which inhibits the action of T7 polymerase. In addition the lysogenized λ phage also contains the T7 gene 3.5 amidase, whose product is a natural counterpart of T7 polymerase.

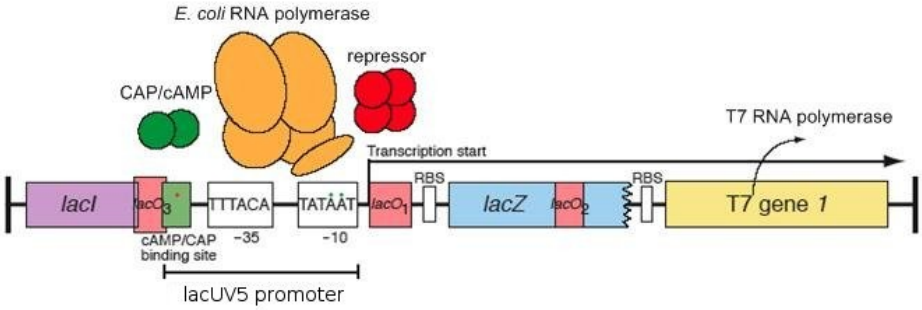


Fig. 1.8. The important part of λDE3 genome containing the T7 RNA polymerase encoding gene under the control of the *lacUV5* promoter. The arrangement is similar to the original *lac* operon. The fragment contains all important regulatory sequences like operators, CAP binding site, -10 and -35 box. As in *lac* operon the regulatory unit is preceded by a putative l*acI* gene.

In many cases the over-expression of the cloned gene leads to cell destruction, resulting in cell death. That phenomenon can be explained by an uncoupling of transcription from translation, caused by the high processivity of T7 RNA polymerase and the concomitant accumulation of mRNA. Translation of produced mRNA has to be accomplished by *E. coli* encoded proteins, and those enzymes aren't designed to translate such an amount of mRNAs. Hence mRNA accumulates within cytoplasm and in previous studies it was shown, that naked RNA stretches cause destruction of ribosomal RNAs (Dong et al., 1995), destroying the host's cell translation machinery. To overcome that problem, mutant strains producing a less active T7 polymerase have been selected. Diminished polymerase activity leads to a smoother increase of recombinant mRNA, thus the outstrip of transcription over translation gets delayed, resulting in an enlarged recombinant protein production period. The same effect can be achieved, when mutations affect the number of available T7 polymerase enzymes.

Surprisingly even pET vectors containing no foreign DNA fragments are toxic for the host when induced with IPTG (Miroux and Walker, 1996). Such empty plasmids just expose the original expression vector, thus, depending on the type of used pET plasmid, different tags (S-tag, His-tag, T7-tag, etc.) are present surrounding the MCS. That would suggest, that also the vector and its associated tags contribute to toxicity of pET expression system. Due to the presence of a functional T7 promoter and sequences coding for small peptides (tag-sequences) the accumulation of short mRNA molecules cannot be excluded. Hence the toxic effect of empty pET vectors may underlie the same mechanism as described above (outstrip of transcription over translation).

1.2.4 A different up-regulation of RNA I and RNA II post induction

Non published microarray data (Striedner et al.), gained from a bioreactor cultivation of *E. coli* HMS174(DE3) cells carrying the expression vector pET30a, revealed a different up-regulation of the two regulatory RNAs, namely RNA I and RNA II (see Fig. 1.9). As described in 1.2.1.1 those two small RNAs accomplish the tight regulation of plasmid replication originating from ColE1/pMB1 origins. Immediately after induction of recombinant protein synthesis both types of RNAs get up-regulated, whereby activator RNA II is transcribed in nearly the same extent as the gene encoding the recombinant protein. In contrast the replication inhibitor RNA I rises more smoothly, resulting in nearly a doubling of initial RNA I amount. A putative explanation for that observed imbalance between RNA I and RNA II transcription could be an insufficient transcriptional termination at the upstream located intrinsic T7 termination signal. Read through events at the T7 terminator may result in an increased transcription of activator RNA II, which is encoded in the same orientation as gene sequences cloned into the MCS.

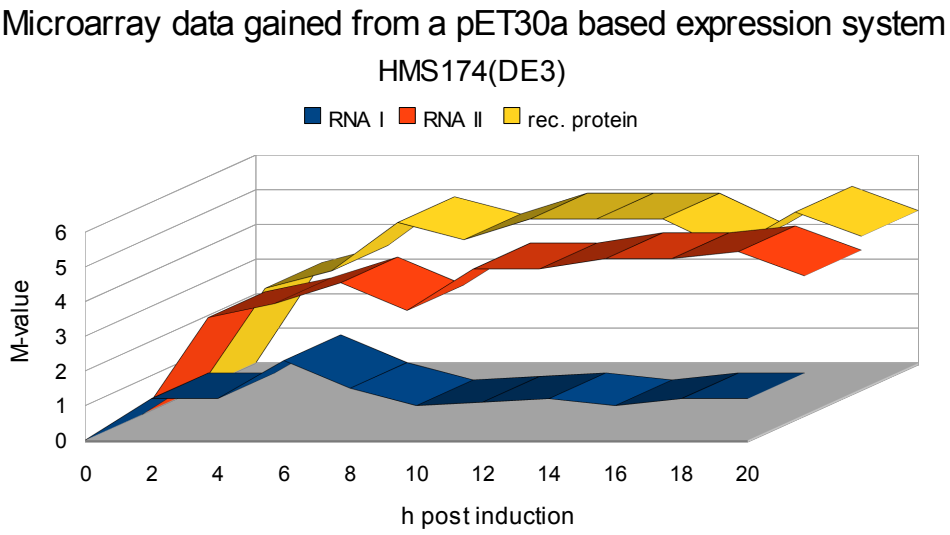


Fig. 1.9. Microarray data from a bioreactor cultivation of *E. coli* HMS174 (DE3) cells carrying the expression plasmid pET30a. After induction you can see a simultaneous increase of RNA II and mRNA encoding the recombinant protein, but just a slight increase of counterpart RNA I. The M-value indicates the extent of up-regulation of a distinct gene at a transcriptional level (M-value=ld(x)). For example a M-value of 4 of a distinct gene span sixteen orders of magnitude in comparison to transcription rate in not induced state.

1.3 Transcriptional termination

Transcription lies at the heart of cellular gene expression, and thus may present the most powerful option to manipulate the expression rate of a single gene or group of genes. Once the ternary complex is build up, it must be stable enough to allow the incorporation of up to hundred bases per second without dissociation of the RNA polymerase during non terminating transcriptional pauses or delays. Thus a tight connection of the elongating RNA polymerase with the template DNA and the producing RNA transcript is essential for the ability to produce mRNAs with a length of several hundred or thousand nucleotides.

After transcriptional initiation and the building up of an extraordinary stable ternary complex the enzyme moves along the template, incorporates nucleotides one by one and produces the desired RNA chain. The synthesis of RNA and the release of the mRNA of a single gene or transcriptional operon have to be stopped at distinct sites on the template. This process is called transcriptional termination and resembles the events during transcriptional initiation but in reversed order, resulting in the dissociation of RNA polymerase and the release of transcribed RNA. Termination occurs in response to well defined signals within the template DNA, the so called transcriptional terminators. These terminators have been identified as sequences within the DNA template which are absolutely required for a termination event. The deletion of the whole terminator sequence or part of it results in a non dissociation of the RNA polymerase and therefore the incorporation of nucleotides into the growing RNA chain doesn't stop, and a extraordinary long transcript is produced. This non terminating event is also described as read through of the polymerase in literature. Thus the information for transcriptional termination must lie within template DNA.

Many terminators require the formation of a characteristic secondary structure within the currently transcribed RNA. According to their characteristic appearance these structures are often named 'hairpins'. Their occurrence within the transcript indicates that transcriptional termination is not simply determined by template sequence. Like every biological process even termination is not a make or break decision, therefore doesn't happen in an extent of 100%. Indeed terminators vary widely in their efficiencies of termination.

Transcriptional termination doesn't solely comprise the generation of a transcript end, rather transcriptional termination displays an object for regulation of gene expression. Transcriptional termination as well as initiation strongly rely on putative control sequences and auxiliary factors, thus are tightly controlled. The dissociation of the RNA

polymerase and therefore the release of transcripts defines the available amount of mRNA, thus accompanies with the strength of gene expression and that release is tightly controlled through interactions between RNA polymerase and sequences within synthesised RNA and template DNA.

1.3.1 Rho-independent termination

Terminators are distinguished in *E. coli* according to whether RNA polymerase requires additional factors to terminate *in vitro*. At the first group of termination signals the core enzyme can terminate *in vitro* at certain sites in the absence of any other factors. These sites of termination are called **intrinsic terminators** or also **class I terminators.**

1.3.1.1 Intrinsic terminators (Class I terminators)

Intrinsic terminators share one common structural feature, the so called hairpin structure. On the one hand the hairpin comprises a stem structure, encoded by a dG-dC rich sequence of dyad symmetrical structure. On the other hand the structure also exhibits a dA-dT rich region at the 3'end directly following the stem structure (see Fig. 1.10), encoding a continuous tract of rU in the transcript. Comparison of the about ~1100 intrinsic termination signals within *E. coli* revealed a consistent distance of seven to nine bases between those two hairpin features. About half of the genes within *E. coli* genome possess intrinsic termination signals, thus are widely used regulatory elements mediating transcriptional termination. Hairpin formation is thought to force RNA polymerase to slow down or even to pause. That delay in RNA polymerase's progression may display an opportunity for termination to occur. Sequences highly similar to intrinsic termination signals but differing in the number of base pairs between the hairpin and the U-tract (typically 10-11 rU) were shown to induce pausing of RNA polymerase but fail to induce dissociation of the enzyme. Pausing at those hairpin sequences normally lasts in a range of ~60 sec (Benjamin Lewin. Genes IX. Jones and Barlett Publishers, Inc. Ninth edition. 2008. S 287ff).

The uridine rich region at the 3' end is thought to facilitate transcript release when RNA polymerase pauses at hairpin structures. The breakage of a dA-rU base pair interaction requires the least energy compared to any other base pair bindings within a RNA-DNA hybrid, thus is believed to facilitate the unravel of the hybrid structure finally ending in transcript release.

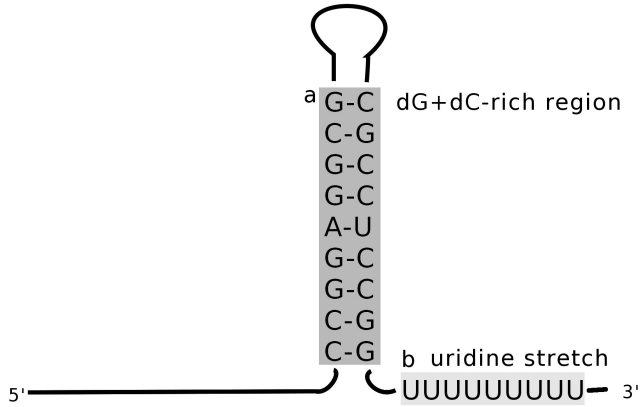


Fig. 1.10. Features common in intrinsic termination signals. Terminators mainly contain two separate structural features. (a) A sequence region rich in dG and dC of dyad-symmetrical structure to form the extraordinarily strong stem structure of the hairpin. (b) A consecutive run of thymidines (according to the non-template strand) at the 3'end, which gets transcribed into a uridine stretch at transcript end. Termination occurs within this rU rich region.

Alternatively, Yager and von Hippel (von Hippel and Yager, 1991) explained the termination event by considering the contribution of the relative free energy. In their model transcriptional termination is likely to occur when the relative free energies (ΔG) of a set of events is high enough. Events like the formation of unwound DNA in the transcription bubble, the dissolving of the hybrid between template DNA and produced RNA, the building up of a contact between RNA polymerase and nearby template DNA and respectively nascent RNA are all together crucial for the occurrence of a termination event. The relative free energies values of all these events are highly dependent on the nucleotide sequence in the template DNA and vary at every single nucleotide the RNA polymerase is sliding along. That means, that at every template position a decision, whether the polymerase elongates or terminates, is made. Termination occurs, when the summation of relative free energies of all those events is high enough, resulting in an enhanced instability of the elongation complex, finally leading to a dissolving of RNA polymerase from template DNA.

From a kinetically point of view transcriptional termination or elongation represent two competitive pathways and the decision which pathway is favoured depends on template sequence and is made at every template position. Theoretically termination can occur at each position along the template, but normally transcriptional termination is a very rare event. Kinetic analysis of elongation complexes have shown, that the half-times for complex dissociation and RNA release lie within a range of several hours or even days, indicating that these elongating enzyme arrangements are very stable. When consider-

ing that the average dwell-time for the incorporation of a nucleotide at a given template position ranges within 10-50msec (at saturated NTP concentrations) (Rhodes and Chamberlin, 1974), we can see, that most of the time a decision in favour of transcript elongation is made. Characteristically the sequences of Rho independent terminators are able to destabilize the elongating polymerase in such an extent that termination can occur. Due to this kinetically approach it should be possible to predict the likelihood of a termination event when the template sequence is known.

But all the efforts to predict the occurrence of a termination event or the exact position of transcript termination are complicated by the finding that DNA sequences surrounding the proper termination sequence also effect transcriptional termination mediated by E. coli RNA polymerase (Reynolds and Chamberlin, 1992) (Telesnitsky and Chamberlin, 1989). Interestingly the termination efficiency (TE) of a distinct terminator can vary extremely when the sequence gets transcribed by a set of different RNA polymerases. Hence the enzyme by itself also plays an important role in this whole termination process (Jeng et al., 1992).

As described above the stability of the elongating ternary complex is very high and when once build up the enzyme complex is very processive, resulting in a low average dwell-time. Strategies to delay the enzyme by its incorporation of new nucleotides lead to an increase in termination efficiency. For example you can reduce the supply of free UTP nucleotides, which are needed in high amount during transcription, because the stretch of adenosines (dA) in the template DNA cause the incorporation of a unusually high number of rUTPs. Shortening the supply of this essential nucleotide slows down the whole process of nucleotide incorporation, because the enzyme has to 'wait' until this rare nucleotide is available again. Therefore the whole average dwell time of the enzyme is increased, especially at the 3' ends of the transcripts. All together this makes a termination event more probable (Wilson and von Hippel, 1994).

A lowering of transcription rate could also explain the positive effect of the general elongation factor NusA on termination. After dissociation of the σ- subunit this factor binds to the remaining core enzyme and is believed to slow down the elongating polymerase. That delay leads to an enlargement of the time window within a termination event can occur. Thus TE may increases at most intrinsic as well as Rho-dependent termination sequences (Gill et al., 1991).

Early transcriptional termination has been an object of molecular biological research and this research over several decades originated a deep insight in the whole termination event. A large number of models and hypothesis trying to explain transcriptional ter-

mination can be seen as proof of that intensive investigation during a long time period, but all together a complete understanding of the whole process is far away of realization. Indeed only few general rules concerning this essential process can bee drawn. So it seems, that the length of the uridine tract encoded by the deoxyadenosine stretch at the 3' end as well as the stability and sequence of the RNA hairpin formed by the G + C-rich region are the greatest contributors which influence termination efficiency. But not only the stability of the hairpin or the number of uridines per se are the crucial factors, also structural features of the whole terminator play an important role. Therefore the hairpin structure as well as the U-tract are essential for termination, but these two sequences alone are not sufficient to induce termination, other interactions, accomplished by sequences more downstream or even upstream of the proper terminator, interact with the polymerase and account for overall TE.

a) The role of the RNA hairpin in intrinsic termination

Nowadays a set of base analogues, like5-bromo-UMP, 5-iodo-CMP, or 4-thio-UMP represent a formidable tool to invent many molecular biological questions. As described in 1.3.1.1, the formation of a hairpin like secondary structure within the nascent RNA transcript plays an important role in termination. To examine that importance, the base analogue inosine-triphosphate (rITP) (see Fig. 1.11) was incorporated instead of the normal ribonucleotide guanine-triphosphate (rGTP). These ribonucleotides weaken normal base pairing and prevent the formation of secondary structures within the transcript. Utilization of ITPs in an *in vitro* transcription experiment abolished termination (Hartvig and Christiansen, 1996), indicating that the formation of a hairpin structure is essential for a termination event at intrinsic (class I) terminators.

Termination mediated by T7 RNA polymerase in the presence of intrinsic termination (class I) sites was shown to be a single strand dependent process. T7 RNA polymerase is able to use templates consisting of the single stranded non-coding strand (template strand) as long as the promoter region is double stranded (Milligan and Uhlenbeck, 1989). Transcription originating from these single stranded templates is successfully abolished, when an intrinsic terminator is utilized, whether overall TE is diminished. A possible explanation for this decrease in TE could be a suggested interaction between the elongating enzyme and the non-template strand, as reported in recently accomplished studies.

For a long time the stability of the hairpin mediated by G-C pairs within the stem structure was believed to be the most essential compartment of the hairpin structure to affect TE. Insertion of putative bases into the stem structure should theoretically result in a

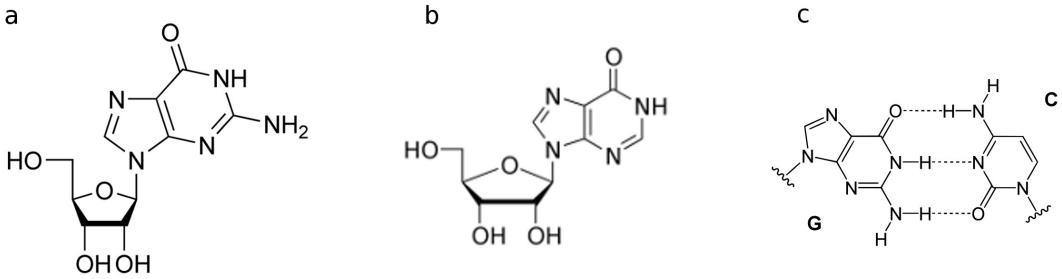


Fig. 1.11. Chemical structure of nucleosides guanosine and inosine. (a) Chemical structure of guanosine. Guanosine consists of the purine base guanine linked (N-glycosidic bond) to the sugar component ribose. (b) The base analogue inosine is a nucleoside similar to guanosine. It consists of ribose as sugar component and is linked to the base hypoxanthine. Inosine is commonly found in tRNAs and is essential for proper translation of the genetic code in wobble base pairs. Inosine is capable to pair with all four bases (cytosine, adenine, thymine and guanine). (c) Normal base pairing between cytosine and guanine includes three hydrogen bounds. Replacement of guanine with inosine leads to a reduction of possible hydrogen bounds, because one essential amino group is missing, leading to a destabilization of possible secondary structures.

higher overall ΔG value, and therefore the overall TE should increase (Wilson and von Hippel, 1994). Surprisingly the increase of thermodynamic stability by inserting G-C pairs didn't result in higher TE, indicating that the stability of the hairpin structure isn't the only essential determinant of termination. It's assumed that in addition to stability the three dimensional structure of the hairpin plays an important role in termination. For the most characterized intrinsic terminators the distance from the first closing base pair of the stem structure to the first termination position is conserved. That invariance could be seen as a putative evidence for the importance of the three dimensional structure (McDowell et al., 1994) (Reynolds et al., 1992). As a conclusion it seems that the hairpin has to assume a distinct three dimensional shape, in order to interact with the elongating polymerase. As a consequence of this interaction the conformation of the polymerase gets changed, upon the polymerase is more capable for a termination event. In presence of appropriate A+T rich sequences downstream of the hairpin termination can occur.

The decisive contribution of conformational changes in a biological process is well known from transcriptional initiation and indeed nowadays transcriptional termination is seen as an inversion of incidences happening during early stages of transcription. After recognition of promoter sequences and binding of the holo-enzyme, RNA polymerase doesn't begin to synthesize RNA molecules of the desired length immediately. In a process called abortive initiation the enzyme transcribes only short RNA molecules of minor length, releases the short transcripts and reinitiates without dissociation form the promoter. This abortive initiation complex is characterized by short life times and therefore is totally contrary to the highly stable and processive elongating polymerase. The

transition into a stable elongation complex occurs in response to the binding of the nascent RNA to a specific binding site within the N-terminal portion of the enzyme. Upon interaction conformational changes are induced, leading to a release of contacts upstream of the promoter and finally ending in an isomerization to a highly stable elongating state (Muller et al., 1988) (Ling et al., 1989) (Diaz et al., 1996).

Such conformational rearrangements are also believed to play an essential role during transcriptional termination, but occur in an inverted order. Due to the fact that the structure of the hairpin is important for termination events at intrinsic terminators, it's assumed that specific interactions between the RNA polymerase and the nascent RNA occur. As a consequence conformational changes are induced resulting in a more termination capable enzyme, evoking the dissociation of the enzyme and release of the immediately synthesized RNA molecule.

b) A deoxythymidine tract downstream of the stem structure

As mentioned above (see 1.3.1.1a) hairpins alone are able to induce conformational changes making the polymerase more capable for a termination event, but termination only occurs when A+T rich sequences are presented downstream. Deletion mutation analyses reinforced the importance of these downstream A+T rich regions. The consecutive shortening of the 9 thymidines consisting 3' end of the *thr* attenuator (see Fig. 1.12) by deletion of adenosines in the template strand showed different consequences on TE (Lynn et al., 1988). Reducing the number of thymidines following the stem structure of this intrinsic terminator to a value of 6, didn't lead to a noticeable effect on TE. Further deletion of up to three thymidines showed a linear decrease in TE. Templates consisting of only one or two thymidines didn't result in termination of transcription and lead to a complete read through of the polymerase. These results indicate that at least 4 consecutive thymidines must be present in template to notice transcriptional termination. TE can be increased by introducing additional nucleotides but at the presence of 6 thymidines the addition of putative bases doesn't lead to an increase in TE.

According to the original hybrid destabilization model the thymidine residues (respectively uridine residues in the RNA transcript) facilitate dissociation of the newly synthesized transcript from template DNA, because rU-dA base pairs are the weakest possible base pairing, mediated by only two hydrogen bounds. Whether it seems that the U tract has putative roles in termination, incorporation of base analogues confirmed the helping function of the U tract in dissociation of the transcript. *In vitro* experiments in presence of base pair stabilizing 2,6- diaminopurine reduced termination (Farnham and Platt, 1982). This base analogue is able to form stronger base pairs with dA or dT and as a consequence the transcript isn't easily released anymore, resulting in an increased read through of the polymerase.

Newer hypothesis concerning transcriptional termination at intrinsic termination regard the role of A+T rich sequences not only in facilitating the release of the transcript. Nowadays these sequences are also thought to induce a so called inchworming movement of the RNA polymerase (Nudler et al., 1994). It was believed that once the enzyme has started elongation the structure of the enzyme complex is more or less invariant and the polymerase slides along the template in a monotonic movement. Studies of elongating polymerases revealed, that indeed the enzyme maintain the same conformation and moves along the template uniformly most of the time, but this monotonic movement can be interrupted at characteristic template positions. Some template sequences are able to hamper the movement of the polymerase at the front end, whereas the delay isn't transmitted to the back end yet. Upon this time shift in realization of the reduced elongation rate a tension within the enzyme complex accumulates. When the internal tension is high enough the compressed enzyme relaxes again and thereby leaps several bases downstream of the template. This cycle of strain accumulation within the elongating polymerase and succeeding relaxation is called inchworming process. Interestingly dA+dT rich sequences are assumed to be able to induce inchworm-like movement of RNA polymerases.

As an essential compartment of intrinsic terminators the role of dT-dA rich sequences in inducing inchworming like movements and a putative contribution on the whole termination event has to be investigated. Perhaps an inchworming cycle resulting in haltering of the front enzyme part is necessary for hairpin formation especially with regard to the competitive DNA-RNA hybrid formation.

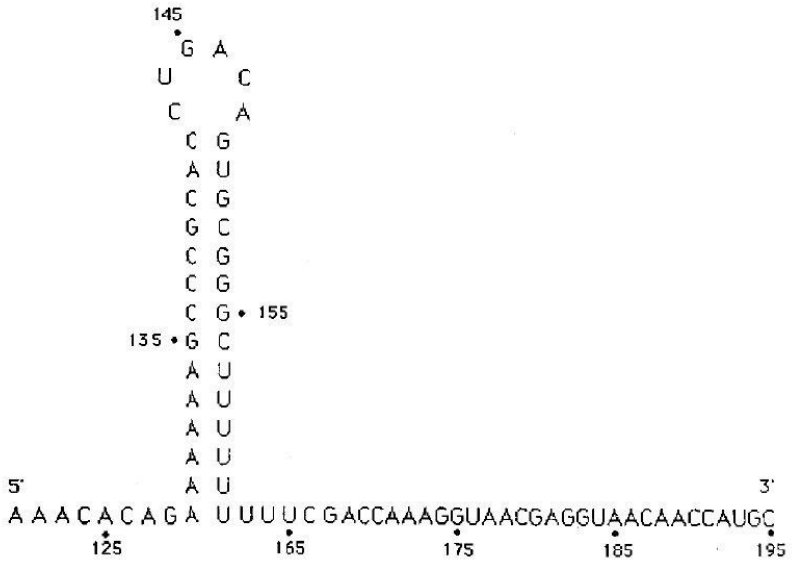


Fig. 1.12. The secondary RNA structure of the thr attenuator. It's a typical class I terminator, consisting of a G-C rich stem structure and a following consecutive run of thymidines (according to the non-template strand) at the 3'end. Deleting the number of thymidines resulted in a decrease of TE, indicating that this sequential features influence transcriptional termination (Lynn et al., 1988).

c) A deoxythymidine tract upstream of the G-C region

As mentioned above the *thr* attenuator (see Fig. 1.12) is a very representative intrinsic terminator consisting of a dG+dC rich dyad-symmetrical region forming the stem structure of the hairpin and a dA-dT rich sequence section more downstream. As a putative feature the *thr* attenuator also exhibits a dT-dA rich region directly upstream of the stem structure (Gardner, 1982). This consecutive run of adenosines (according to the non template strand) was found in over one third of intrinsic terminators. The frequency of this sequence feature suggests that this run of adenosines could have an important biological function, and therefore could assist transcriptional termination.

The first assessment of the function of this upstream adenosine run was a potential base pairing with the more distal uridine containing tail at the 3' transcript end. On the one hand these putative base pairings could increase the overall stability of the hairpin and therefore making hairpin formation more probable. On the other hand an interaction of the upstream located adenosines of the transcript with uridines within the 3' end reduces the number of base pairs within the RNA-DNA hybrid, and may assist in transcript release. In order to investigate the role of this relative common feature, deletion mutants lacking one to six thymidines (according to the template strand) were analysed

in an *in vitro* transcription assay. The consecutive deletion of thymidines leads to a small linear decrease in TE. The introduction of up to 3 thymidines into the wild type template maximizes putative base pairings, because now all nine uridines of the transcript tail can undergo base pairing. Surprisingly putative thymidines only moderately increased TE (Yang et al., 1995).

Taking together it seems that dA-dT sequences upstream of hairpin structures aren't absolutely required for efficient termination, but the presence of such elements is able to assist termination. Sequence features like the dG+dC rich elements within the stem structure responsible for the stability of the hairpin and a run of uridines at the 3' end are by far more crucial for the occurrence of a termination event. Because of the fact that over one third of intrinsic terminators possess these upstream located dA-dT rich sequences, it's hard to believe, that such a common feature doesn't have an evident biological function. Maybe these sequences are involved in biological processes despite transcriptional termination, being advantageous for the organism and therefore got enriched during evolution. But the alternative function of these dA-dT rich sequences is still unknown and has to be investigated.

d) Sequences more distal to the proper terminator can influence TE

Sequences directly following the terminator, but also sequences more downstream of the intrinsic terminator can have a great influence on TE (Telesnitsky and Chamberlin, 1989). Surprisingly it was found that even sequences near the promoter, therefore sequences several hundred or thousand nucleotides upstream of the intrinsic terminator are able to have an impact on TE. The cloning of a 51bp long DNA fragment originally stemming from bacteriophage λ in the proximity of a promoter resulted in a higher TE at the downstream lying *thr* attenuator (Jeng et al., 1992). This finding includes the obscure conclusion, that promoter near sequences can influence an event which is strongly driven by processes within the nascent transcribed RNA, but the transcript parts essential for termination (3'-end) aren't present at the moment, when the polymerase slides along these promoter near sequences. A possible explanation could be, that while transcribing promoter near sequences, the structure of the elongation complex is changed in such a manner, that a termination event more distal gets more possible or even impaired. It's also thinkable that the mRNA part of this 51bp long fragment can interact with the terminating RNA polymerase whereupon transcriptional termination gets enhanced.

First evidence, that sequences downstream of the proper termination signal can influence termination, stems from analysis of hybrid terminators. Adjacent to the complete T7 terminator, including the crucial structures of the dG-dC rich region forming the stem of the hairpin and the run of thymidines at the 3'-end, the downstream sequence of the similar T3 terminator was cloned. T7 alone exhibit a termination efficiency of 74% (according to Telesnitsky and Chamberlin, 1989). *In vitro* transcription analysis of the hybrid terminator revealed a decrease in TE to a value of only 7%(Telesnitsky and Chamberlin, 1989). This drastic decay in probability of a termination event is very surprising, because all the structures important for stopping the elongating polymerase should be present. Therefore sequences downstream of the U-rich region of the transcriptional termination site play an important role. This conclusion also means, that not yet transcribed sequences within wild type T7 terminator, influence the capability of the *E. coli* polymerase to terminate at intrinsic terminators. This finding reveals, that termination is a complex process and signals additional to transcript encoded structures effect RNA polymerase mediated termination.

As a simple assumption one could believe that a combination of two terminators stops polymerase more efficiently than only one terminator. Such a combination was done by cloning the termination signal of the rrnC operon downstream of the well known *thr* attenuator. By considering the results, gained from analysis of the upper mentioned hybrid termination signal, a less surprising decrease in TE at the original *thr* attenuator was observed. The insertion of the ribosomal terminator lead to a nearly bisection of the TE at the upstream located *thr* attenuator (Jeng et al., 1992).

Ultimately it was found, that even lesser rearrangements including only three to seven nucleotides following the transcript termination site are sufficient to diminish the capability of RNA polymerase to terminate, resulting in a decrease in TE from 74% to 3% (Telesnitsky and Chamberlin, 1989). The observation, that the alteration of only few nucleotides can drastically influence TE, makes it hard to compare results gained from different *in vitro* transcription studies. Depending on the used cloning vector, one distinct termination signal could exhibit different calculated TE.

e) An extraordinary stable tetraloop sequence

RNA hairpin structures are the essential structural feature within intrinsic termination signals, but also are very common secondary structures in rRNAs. The sequences of eubacterial 16S like rRNAs reveal numerous hairpins, especially hairpins with loop sequences containing four nucleotides are very common. Lots of these hairpins having a tetranucleotide loop, exhibit the loop sequence UUCG. Mutational analysis showed that base changes within the helical stem structure of the hairpins had no functional effect as long complementary exchanges restored base pairing within the stem structure. That means, that for hairpin function the primary sequence is not as essential as the stability of the hairpin (Woese et al., 1983). In contrast changing the loop sequence lead to serious effects on ribosome mediated translation. The importance of maintaining the integrity of loop sequences, expose them as possible interaction sites between rRNAs and ribosomal proteins, or between rRNAs and surrounding rRNAs parts (Carey and Uhlenbeck, 1983).

The fact, that loop sequences serve as nucleation site immediately after their transcription, could explain their highly conserved sequences. RNA loops may induce hairpin folding and therefore serve as nucleation site. The ability to facilitate hairpin formation and stabilizing the structure of the hairpin depends on loop sequence, and therefore even changes of one nucleotide can have serious effects on hairpin stability.

In regard to the abundant occurrence of hairpin loops with the primary sequence UUCG within 16S like sequences the genome of bacteriophage T4 was analysed (Tuerk et al., 1988). Indeed the hexanucleotide CTTCGG was found 13 times within the non-template strand (those strand whose sequence is equal to that of the mRNA), and 10 times within the template strand. According to the genome size CTTCGG occurred twice the expected number of a statistically distributed hexanucleotide of the same sequence. Interestingly the sequence never occurred between genes at the template strand. In contrast the hexanucleotide within the non-template strand always was located between two ORF (intercistronic), to boot in 12 of 13 occurrences the sequence was surrounded by sequences of dyad symmetry, indicating potential hairpins with UUCG as loop sequence and a C-G closing base pair. Because of the presence of these putative hairpins within genes these hairpins may function as transcriptional terminators. *In vitro* transcription analysis with avian myeloblastosis virus reverse transcriptase showed, that all 12 hairpins are able to terminate transcription. Thermal denaturation experiments also revealed, that hairpins harbouring UUCG sequence in the loop, are more stable compared to control hairpins only different in loop sequence (Tuerk et al., 1988).

Analysing the T4 genome it is conspicuous, that there is a statistically preference of C-G closing base pairs within hairpins over any other closing base pairs. Even reversing the closing base pair into G-C lead to diminished hairpin stability. Among the highly phylogenetic conserved tetraloops within 16S like rRNAs about half of them have C-G closing base pairs (G-U pairs also allowed) (Tuerk et al., 1988). E. Coli genome analysis revealed the occurrence of 43 hairpins having UUCG loops and a stem structure containing at least 5 base pairs. 22 of these UCCG tetraloop sequences have a C-G closing base pair. All together the frequently used tetraloop sequence UUCG and C-G as closing base pair doesn't seem to be coincidental and therefore must rely on a physical or biological benefit (d'Aubenton Carafa et al., 1990). But until now the definite physical or biological factor making the CUUCG loop sequence such a likely nucleation site is not known. Both the loop sequence as well as the type of the closing base pair influences the probability of hairpin formation before the descending part of the stem is transcribed. The sequence CUUCG serves as strong nucleation site and reduces ΔG for loop closure, making that sequence widely distributed among hairpin structures.

f) Characterization of the T7 terminator

Nowadays most of the expression systems for producing recombinant proteins issue from bacteriophage T7. Genes coding for the desired protein are cloned downstream of the highly processive T7 promoter. Transcription, executed by T7 RNA polymerase and originated from this strong promoter, stops at an intrinsic terminator, which also stems from T7 bacteriophage's genome. The terminator lies within the late transcribed region at map unit 60.62 in the genome of bacteriophage T7 and is located at the 3' end of the minor capsid protein (Sengupta et al., 1989). This protein is an essential component of the bacteriophage’s protein capsid and is needed in a high amount.

Genes coding for proteins which are required in great quantities are good candidates for having strong terminators, because the cell has to optimize the expression rate, in order to prevent a waste of energy, originating from the production of not desired side products. The T7 terminator forms an extraordinary stable hairpin structure with an extended stem structure, resulting in a high ΔG value of -30 kcal (see Fig. 1.13). With a TE of about 80% the T7 terminator is a very strong terminator. Nevertheless read through occurs in an extent of 20%. This inefficiency in transcriptional termination was assumed to be essential for phage viability (Dunn and Studier, 1983). Downstream from T7 terminator located genes 1 and 12 encode for proteins incorporated into the tail structure of the phage and therefore are essential for an intact and infectious phage particle. But these two genes aren't provided with appropriate promoter sequences hence the two genes aren't part of a separate functional transcription unit. Despite the absence of a

putative promoter, a AUG startcodon and a ribosome binding site is presented on the readthrough transcript, and this should enable the translation of the two genes. Thus the essential tail proteins can only be expressed when readthrough at the upstream located T7 terminator occurs. It's thought that the T7 terminator is naturally 'designed' to allow readthrough in a small extent to enable expression of downstream located genes (Dunn and Studier, 1983).

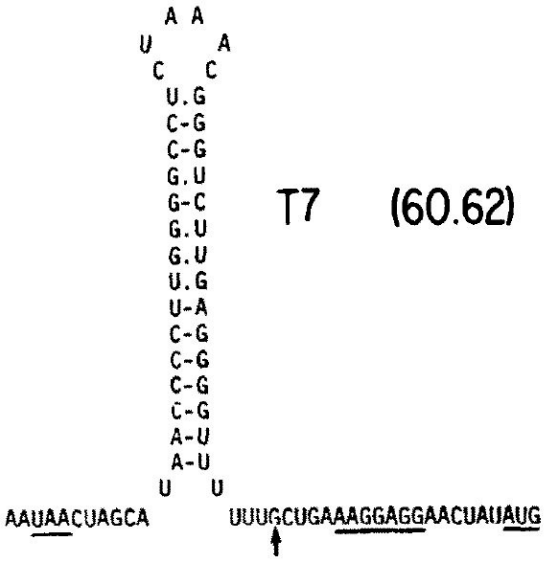


Fig. 1.13. RNA secondary structure of T7 termination signal within the late transcribed region at map unit 60.62. The T7 terminator forms an extraordinary stable hairpin structure with a calculated ΔG value of -30 kcal. The resulting transcript is terminated at the guanosine, indicated by an arrow. The termination signal lies downstream of a gen coding for the minor capsid protein and its translational stop codon UAA is underlined. Two genes are located directly downstream of the T7 termination signal. Both genes aren't provided by a promoter sequence, but exhibit a ribosome binding site and the translational startcodon AUG (both sequence features are underlined) (Dunn and Studier, 1983).

1.3.1.2 Class II terminators

Surprisingly a signal within the human preproparathyroid hormone (PTH) gene (see Fig. 1.14), completely different from usual rho independent termination signals, was found to be capable of terminating transcription, mediated by bacteriophage T7 RNA polymerase (Mead et al., 1986). Further studies of this unusual termination signal revealed an absolutely different structure, compared to known rho independent termination signals with their predominant hairpin structure. The signal doesn't show any notable secondary structures and therefore lacks expanded stem-loop structures. In contrast to usual hairpin like terminators the termination event, induced by this new signal, needs the presence of the non-coding strand, thus is a double strand driven process (He et al., 1998). Further evidence, that transcriptional termination at these novel sequences, so called class II termination signals, is mediated through a mechanism independent from class I termination pathway, stems from *in vitro* transcription analyses with a mutated T7 RNA polymerase. Mutations between the 20kDa N-terminal portion and the 80kDa C-terminal fragment lead to a 'nicked' form of the polymerase, which isn't able to terminate at class II termination signals, but is still able to stop elongation when transcribing hairpin containing class I termination signals (Lyakhov et al., 1997). It seems that there are two different mechanism which can accomplish transcriptional termination in *E. coli,* and both pathways operate independently from each other.

Sequence analyses of putative organisms revealed the presence of further class II termination signals. These unusual sequences are present within the *E. coli* rrnB T1 termination site, in a cDNA copy of vesicular stomatitis virus (VSV), in adenovirus DNA, and within bacteriophage lambda DNA (Zhang and Studier, 1995) (Sousa et al., 1992). All these newly found sequences share a 7bp long consensus sequence (ATCTGTT according to the non-template strand) and an unequally long run of uridines at the 3'end, comparable to the U-run at intrinsic termination signals (see Fig. 1.14).

A putative class II termination site was also found within the genome of bacteriophage T7. The consensus sequence lacking the adjacent run of uridines is present at the right end of the concatemer junction (CJ) of replicating T7 DNA (Lyakhov et al., 1997). Because of the absence of an uridine tract, the importance of this 7bp long sequence was analysed by mutating the intrinsic T7 polymerase into the “nicked” form. Phages expressing the nicked T7 polymerase aren't able to recognize potential class II termination sites. Those phages failed to package replicated T7 DNA, and showed diminished maturation of phage particles, thus even the shortened class II signal seems to play an important biological role.

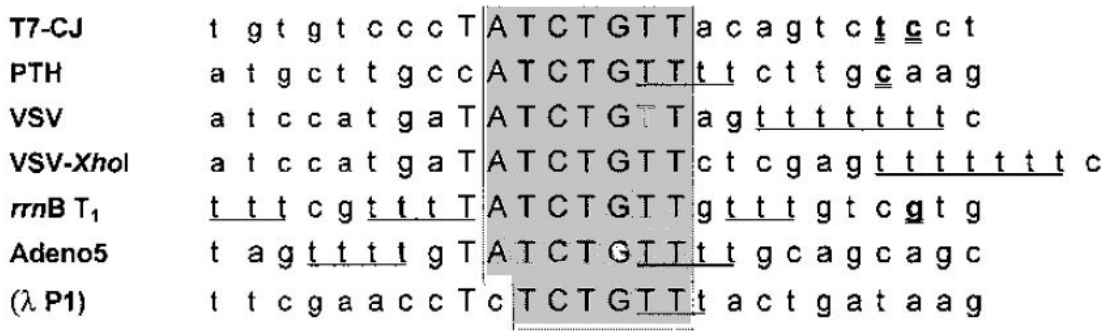


Fig. 1.14. Collection of all known class II termination signals. The consensus sequence ATCTGTT (according to the non-template strand) is shared by six out of seven class II sites and is shaded. All sequences capable to terminate transcription of T7 polymerase contain a varying stretch of dT. Consecutive runs of more than 2 thymidines are underlined. The T7-CJ signal lacks this run of T, and was previously shown to serve as pausing signal rather than a termination signal. The exact point of transcript termination is known for just a few signals and is marked by double underlined letter (Lyakhov et al., 1998).

In vitro transcription assays using a template containing the class II termination site present in PTH gene, revealed a termination efficiency of about 55%, thus that terminator is less efficient than the hairpin forming T7 terminator (T7 terminator shows a TE of about 80%). Both determined TE are quite stable when measured at any time between 5 seconds or 10 minutes (Lyakhov et al., 1998). Aliquots of mRNA taken from a single round *in vtiro* transcription reaction utilizing a T7-CJ signal containing template showed, that in later taken aliquots the portion of read through products increases, leading to a overall decrease in TE. In detail TE declined from a value of about 60% percent when measured 5 seconds after release of the initiation complex to 25% at aliquots taken 120 seconds post transcriptional induction. In contrast to other class II termination signals the sequence within the right end of the concatemer junction shows a time dependent variance in TE, which raises the suspicion that the signal may function as a pausing signal. Polymerases encountering that signal may arrest for 5sec or even longer, but after pausing they are still able to continue transcription.

The class II termination signals T7-CJ and PTH have a 7bp long consensus sequence in common, but differ in the presence of an adjacent run of uridines at the 3'end (see Fig. 1.14). To get putative sequence features essential for the occurrence of a termination event, mutations within the PTH termination signal adjacent to the consensus sequence were introduced (Lyakhov et al., 1998). Those changes barely effected termination, unless the run of four uridines was still maintained. The deletion of only one of the four uridines at the 3'end significantly diminished TE. Shortening the length of the downstream positioned uridines to a value of three, makes the sequence highly similar to the

sequence within the concatemer junctions of replicating bacteriophage T7 DNA, which consists of only two uridines in the downstream part. All known class II termination sites exhibit the consensus sequence ATCTGTT in the non-template. Putative to that common feature the presence of a downstream lying consecutive run of uridines determines, if the whole sequence function as a transcriptional terminator or as pausing signal. The later one is capable to interrupt the transcribing polymerase and therefore delays the enzyme's progress for some seconds.

All together, when we talk about transcriptional termination, we have to distinguish between two totally different mechanisms, both able to terminate transcription, mediated through a set of polymerases, stemming from bacteriophages or bacteria. Intrinsic terminators depend on the formation of a complex secondary structure and are able to terminate transcription even in the presence of only one strand (template strand). In contrast termination at class II sites doesn't rely on extended secondary structures within the newly transcribed RNA and the presence of the second strand is essential for the occurrence of a termination event. Both signals also can be distinguished because of the ability to be recognized by a mutated form of T7 polymerase. The nicked form isn't able to terminate at class II sites, but still terminates at templates, containing intrinsic terminators (class I terminators).

a) The complex transcription region of the E. Coli rrnB operon

Proliferating bacteria cultures consume most of their time and energy for the synthesis of proteins. Therefore they need a large amount of ribosomes as well as a high level of rRNA molecules. Under optimal growth conditions more than one fourth of the total cell mass consists of ribosomes (more than 20 000 ribosomes) (Rolf Knippers. Molekulare Genetik. Georg Thieme Verlag, 8. Auflage. 2001. S106f). Ribosomal RNA is the predominant product of transcription, consisting of about 80%-90% of the total mass of cellular RNA in both eukaryotes and prokaryotes. The number of major rRNA genes varies from seven in *E.Coli*, 100 to 200 in lower eukaryotes, to several hundreds in higher eukaryotes. The genes for the large and small rRNA usually form a tandem repeat. In *E.Coli* there are seven genes coding for ribosomal genes (rrnA, rrnB, rrnC, rrnD, rrnE, rrnG, rrnH).

All those genes are dispersed on the bacterial genome, but all are regulated in common. The rRNA genes are similar structured. All genes consists of three segments coding for the 16S rRNA, the 23S rRNA and the 5S rRNA and are oriented in that succession one after another, the so called tandem repeat. Those segments are separated by short DNA

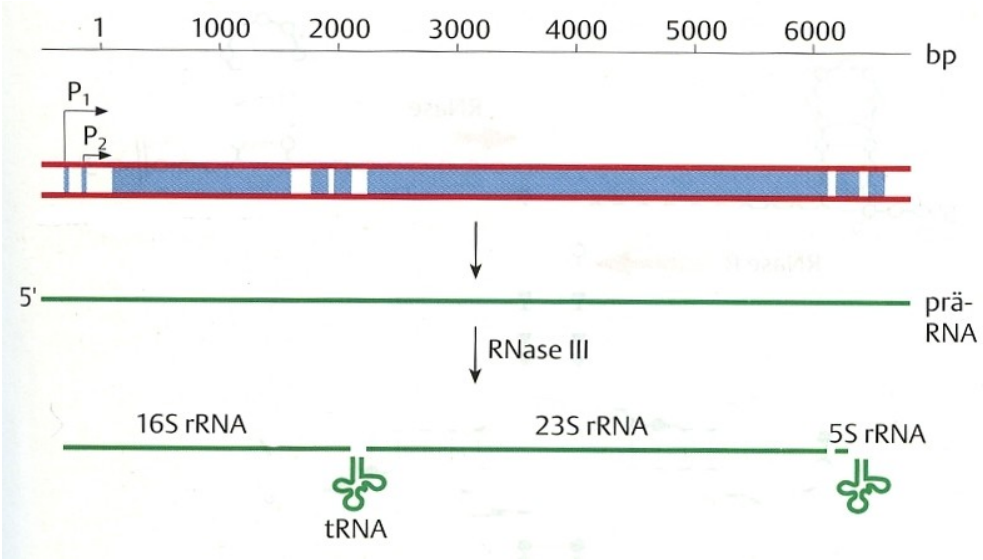


Fig. 1.15. Composition of a bacterial ribosomal gene. Length in base pairs is indicated above. A long transcript originating from one of the two promoters is generated. The long polycistronic mRNA contains the 16S rRNA, 23S rRNA and the 5S rRNA. That organization is called tandem repeat. The three rRNAs are separated by spacer DNA. Some spacer code for tRNAs. After transcription the whole transcript gets restricted by RNase III and this restriction ends up in functionally active rRNA molecules, ready to get incorporated into ribosomes.

fragments (spacer). In the spacer between the 16S rRNA gen fragment and the 23 S rRNA fragment are one or two genes coding for tRNA. Some rRNA genes also have a putative tRNA gene in the 3'-region downstream of the 5S rRNA gen.

The rRNA genes have two promoters, the first promoter (P1) lies about 300bp and the second promoter (P2) about 200bp upstream of the 16rRNA gene fragment (see Fig. 1.15). The promoter P1 is much stronger than P2. Originating from those promoters the whole rRNA gen gets transcribed, producing a long RNA, the so called primary transcript. Afterwards ribonucleases digest the primary transcript into the different rRNAs and tRNAs. The analysis of six ribosomal genes revealed, that all six genes contain intrinsic transcriptional terminators at their respective transcript end (Friedman et al., 1987).

Among the investigated rrn genes, the transcript termination region of the rrnB gene has a unique extended sequence, capable to terminate transcription. That enlarged region contains two segments likely to form intrinsic termination hairpins (T1 and T2), two additional smaller inverted repeats (IR1and IR2), and finally a pair of direct repeats (see Fig. 1.16) (Brosius et al., 1981). Because of the fact that the transcription of ribosomal genes lead to a biological functional RNA molecules, the survival of the whole cell

strongly depends on the integrity of those RNA molecules. Only when the molecule is correctly transcribed according to length and sequence its incorporation in a functional ribosome is guaranteed. Thus such genes are good candidates to have strong transcriptional terminators in order to produce transcripts of distinct and invariant length. Indeed rrn terminators are frequently used in several cloning vectors to achieve an efficient transcriptional termination of the cloned gene (Brosius, 1984).

The enlarged termination region within the rrnB gene arises the question, if the high complexity accompanies with a high ability to terminate transcription. Deletion studies with a shortened template only consisting of the first intrinsic T1 terminator without other downstream positioned inverted repeats (including T2) revealed, that termination is as efficient as with the original extended termination region (Orosz et al., 1991). That result indicates that the first intrinsic terminator in isolated form is sufficient to terminate transcription. Further analyses with templates containing only one of the two intrinsic terminators in isolated form showed that both signals are able to terminate transcription and exhibit nearly the same TE, whereas T2 seems to be a little bit more efficient. That's somewhat surprising because T1 exhibits larger secondary structures and has a strong homology to the intrinsic terminator present in rrnD. Both small inverted repeats only show a slight ability to terminate transcription, but TE markedly increases when the sequences get reversed.

Summing up, indications like the ability of both intrinsic terminators to efficiently stop elongating RNA polymerases in isolated form, the inability of the putative inverted repeats to terminate transcription, the high homologies of the T1 and T2 with single terminators present in rrnD and rrnG, suggest, that the high complexity of the termination region present in rrnB gene has no particular function. Rather the enlargement of that region could be the consequence of genetic rearrangements like transposition events within adjacent rrn genes (Brosius et al., 1981).

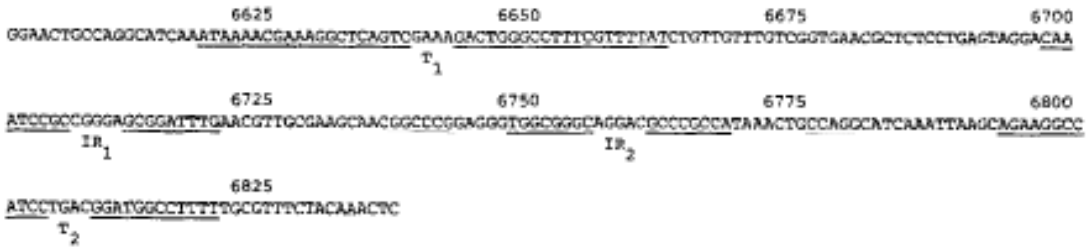


Fig. 1.16. The complex termination region of rrnB gene. The sequence contains two segments likely to form intrinsic termination hairpins, T1 and T2 respectively. Sequences of dyad symmetrical structure are underlined. The rrnB gene also involves two inverted repeats, indicated as IR1 and IR2 (Orosz et al., 1991).

The sequential comparison of seven known intrinsic ribosomal terminators reveal some common features (see Fig. 1.17). One of the obvious shared characteristics results from their organization in tandem repeats. All terminators are preceded by highly structural RNAs, which get not translated into proteins in order to acquire their biological function. Four rrn termination signals are prefixed by 5S rRNA, and in two cases tRNAs are positioned before the intrinsic terminator. All sequences code for an extraordinary large hairpin structure. The stem structure always consists of two parts. One of them is rich in A-T base pairs, whereas adenosines are predominant on the ascending and tyrosines on the descending strand. The second part of the stem structure is positioned adjacent to the loop and consists of 7-9 nucleotides mainly guanosines or cytosines. All ribosomal terminators possess a purine rich tetraloop, whereas the third position is conserved and consists of an adenosine. The lower A-T rich part of the stem region is directly followed by the conserved sequence motif TYTG, whereas Y is a place holder for a pyrimidine nucleotide (Orosz et al., 1991).

Fig. 1.17. Collection of *E. coli* ribosomal termination signals. Sequences of dyad symmetrical structures are underlined and potentially undergo base pairing, building up the stem structure. The stem regions mainly containing G-C base pairs and nearly located to the loop region are in lower case letters. All termination signals posses the conserved sequence motif TYTG directly following the A-T rich region of the stem structure and are depicted in parenthesis (Orosz et al., 1991).

b) Terminator T1 of the rrnB operon

As already mentioned the enlarged termination region of the rrnB operon contains two sequence regions capable to form very complex hairpin structures and are assumed to function as ordinary intrinsic termination signals. *In vitro* transcription studies revealed that termination at terminator T1 occurs through two totally different mechanisms through two totally different mechanisms (Hartvig and Christiansen, 1996) at two separate parts of the large secondary structure (see Fig. 1.18). SP6 RNA polymerase terminates elongation at the upper site even in the absence of the non-template strand, designating the site as intrinsic termination signal (Uptain and Chamberlin, 1997). More precise analyses of transcript length showed that termination takes place at two directly adjacent adenosines (according to the template strand) within the upper termination site.

In contrast transcriptional termination at the termination site closer to the 3' end depends on the presence of the non-template strand and isn't abolished due the incorporation of IMP, which destabilizes secondary structures (Christiansen, 1988). All these attributes reveal this downstream termination region in proximity to class II terminators, and indeed on closer examination that signal also contains the consensus sequence -ATCTGTT- in the non-template strand, a typical feature of class II termination sites. As reported in 1.3.1.2, all known class II signals, regardless if they function as pausing signal or termination signal, share that common nucleotide sequence. As other class II termination signals even the rrnBT1 downstream signal isn't able to stop elongating polymerase, when the *in vitro* transcription assay is carried out with the nicked form of SP6 RNA polymerase. The mutated polymerase reads through the downstream signal without exception, yet TE at the upper positioned intrinsic termination signal increases (Kwon and Kang, 1999). Therefore the ribosomal terminator T1 of the rrnB operon possesses the noticeable feature to combine class I and class II terminators within a common structure. That combination could be the reason for the tremendous ability to terminate elongating RNA polymerase while transcribing rrnB operon, therefore achieving rRNA molecules of distinct length, in order to ensure their right biological function.

In vitro transcription experiments using the SP6 polymerase revealed that the major part of transcriptional termination events occur at the downstream termination signal of rrnBT1. Namely about 73% of overall termination take place at the downstream signal and almost uniquely at one U residue boxed in Fig. 1.18. Termination at the upper positioned intrinsic termination signals contribute to total TE to a lesser extent, and was calculated with a value of 22%, taking both termination relevant U residues together (Kwon and Kang, 1999).

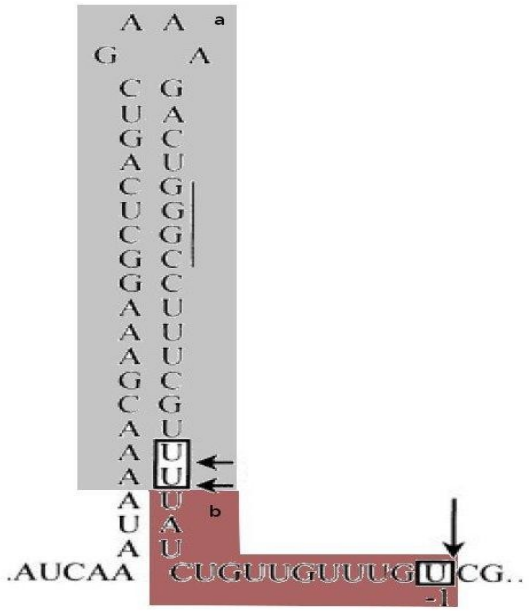


Fig. 1.18. RNA secondary structure of the first intrinsic terminator within the complex termination region of ribosomal rrnB gene. SP6 RNA polymerase terminates at three sites in total, and those sites are indicated by boxes and arrows. The downstream lying termination site is marked with -1. (a) The shaded box designates the upstream termination site which consists of a hairpin forming so called intrinsic terminator . (b) The highlighted sequence shows the downstream termination signal which is a class II termination signal. The sequence includes a conserved region (ATCTGTT in the non-template strand) which is also present in other class II terminators (e.g. in the PTH gene, VSV, etc.)

More detailed analyses of the apparent class II termination signal within the rrnBT1 signal confirmed that termination is totally independent of secondary structures within the recently transcribed RNA molecule. Deleting part of the hairpin structure maintained transcriptional termination at the downstream site, nevertheless TE showed a slight decrease, compared to the original highly structured terminator. The essential role of the consensus sequence ATCTGTT (non-template strand) shared by all known class II signals was affirmed with *in vitro* transcription assays using templates lacking that common feature. The absence totally abolished termination at the rrnBT1 site. Even the disturbance of integrity of the consensus sequence by addition of putative nucleotides lead to a totally loss of termination.

As depicted in Fig. 1.19 the downstream rrnBT1 termination signal contains a consecutive run of 3 thymidines (respectively 3Us in the RNA transcript) between the consensus sequence and the uridine at which termination occurs. Increasing the number of thymidines within this thymidine stretch accompanied by an increase in TE, but also

enabled termination to take place at multiple sites, producing transcripts of variant length. That variance in length is caused by a so called slippage of polymerase when transcribing templates containing highly repetitive nucleotide sequences.

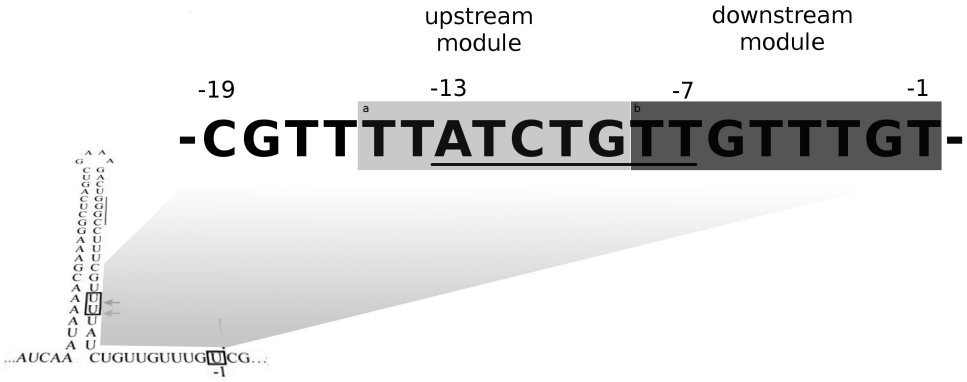


Fig. 1.19. rrnBT1 downstream termination site. The signal contains the consensus sequence -ATCTGTT- (underlined) and is present in all known class II termination signals. When designating the site of termination -1, the termination signal could be divided into functional parts, depending on their requirement of double stranded regions. The upstream module reaches from nucleotide -15 to -9 and has to be double stranded in order to allow termination. The downstream module merges nucleotide -8 to -1 and mediates the release of the newly synthesized RNA molecule.

The initial statement, that the downstream positioned class II termination signal of rrnBT1 is only functional when present in double stranded mode isn't true at all. Analyses utilizing heteroduplex DNA templates revealed, that small single strand regions, evoked through mismatches, only decreased efficient termination when present in consensus sequence or directly upstream of that conserved region. Whereas same mismatches directly adjacent to the 3' end of the essential consensus region didn't decrease TE, suggesting that for efficient termination within the more downstream lying part of the class II signal, the existence of double stranded template regions isn't essential. Because of that dependence for either single or double stranded DNA templates it's suggested to distinguish two functionally different parts within the downstream part of rrnBT1. A region from nucleotide -15 to -9, (when designating the proper site of termination -1), including most of the consensus sequence, is totally dependent on its integrity and its existence in double stranded form, is called the 'upstream module'. Hence the adjacent sequence reaching from nucleotide -8 to -1 is called the 'downstream module' and seems to mediate the release of RNA transcripts comparable to 3' regions of classical intrinsic termination signals.

The ability to release newly transcribed mRNA at the downstream module is thought to be achieved by its consecutive run of thymidines (non-template strand). Transcription of that region results in weakest possible base pairings (rU:dA) and therefore to a destabilization of the DNA:RNA hybrid. Indeed enlarging the thymidine rich sequence by insertion of putative T nucleotides lead to an increase in TE, but as side effect slippage of RNA polymerase evoked. As seen in original sequence (Fig. 1.19), the run of thymidines within the downstream module is punctuated by G nucleotides, which may be necessary in order to truncate slippage of RNA polymerase.

The upstream module, mainly existing of the shared consensus sequence, is responsible for the observed pausing of RNA polymerase. Stopping polymerase in its progress may be the reason for the appearance of slippage at relatively short run of thymidines. Slippage also occurs at intrinsic terminators, but due to the absence of a presented pausing signal, the consecutive run of uridines at the 3'prime end has to be enlarged up to 11 thymidines. Inducing pausing due to incorporation of modified nucleotides wasn't sufficient enough to mediate termination, even in the presence of an intact downstream module. The occurrence of a termination event depends on the presence of both modules, especially the entirely integrity of the consensus sequence is of essential importance. A possible explanation could be an induced conformational change within the elongating polymerase while transcribing the pausing signal. Those structural rearrangements enhance the probability of the polymerase to terminate transcription at the following hybrid destabilizing downstream module.

1.3.2 Rho-dependent termination

For the sake of completeness also rho-dependent should be shortly mentioned although that kind of termination isn't treated in underlying diploma thesis. Termination at that kind of termination sites requires the hexameric protein rho, which belongs to ATP-dependent helicases protein family (Benjamin Lewin. Genes IX. Jones and Barlett Publishers, Inc. Ninth edition. 2008. S 289f). Rho attaches to the 5' end of newly transcribed mRNA at distinct sequences, called *rut* sites. Starting from that recognition site rho moves along the transcript driven by catalytic ATP cleavage until it catches up transcribing RNA polymerase. When the polymerase pauses, enforced through a termination site at the 3'end of the transcript, rho actively destabilizes the remaining RNA-DNA hybrid within transcription bubble, resulting in dissociation of polymerase and release of the transcript. About half of the total amount of termination signals in *E. coli* respond to rho-dependent termination mechanism. So far rho-dependent termination isn't implemented in any expression system for recombinant protein production, thus isn't object of the following investigations.

2 Aim of the work

The main goals of the underlying diploma thesis were the creation of a more efficient transcriptional termination signal and the investigation of a putative correlation between transcriptional termination and plasmid replication control. Microarray data presented in 1.2.4 strongly indicate, that readthrough events throughout the origin of replication are able to deregulate plasmid replication, a phenomenon normally addressed to uncharged tRNAs.

If a stabilization of plasmid copy number is reached its effect on recombinant product yield has to be investigated. It's conceivable, that strong expression systems in combination with an increased gene dosage as result of enhanced PCN trigger a recombinant protein formation, that is usually too high for the host cell metabolism to cope with, and may result in a shortened product formation period and loss of total protein yield. A lowering and stabilization of plasmid copy number throughout the whole process, may issue in an enlargement of cell viability, finally resulting in an increase of total protein yield.

To investigate putative consequences of an enhanced termination signal, bioreactor cultivations of strains carrying an appropriate pET30a derivative have to be carried out, and obtained results have to be compared with cultivation data received from standard system.

3 Material and Methods

3.1 Bacterial strains

For cloning procedures E. coli strain DH5α (F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–) was used as host and E. coli HMS174(DE3) (F- recA1 hsdR(rK12- mK12+) (DE3) (Rif R)) (Novagen, Merck KgaA, HE, Germany) was used as production strain (http://openwetware.org/wiki/E._coli_genotypes). The expression host is lysogenized by λ DE3 phage as described in 1.2.3. Thus expression of T7 polymerase is inducible by addition of IPTG.

3.2 Plasmids

As standard expression vector pET30a (Novagen, Merck KgaA, HE, Germany) (see Fig. 3.1) was utilized. The plasmid carries a kanamycin resistance gen and a putative gen coding for repressor molecule *LacI*. For cloning of foreign DNA numerous restriction sites within the multiple cloning site (MCS) are available. Inserted gen is under the control of the ineducable T7/lac hybrid promoter and transcription, mediated by T7 polymerase, gets stopped by a single T7 termination signal at the 3'end. Replication is mediated through a ColE1 like origin of replication as described in 1.2.1.1. The plasmid is maintained at about 15 copies per cell, thus belongs to middle copy plasmid category.

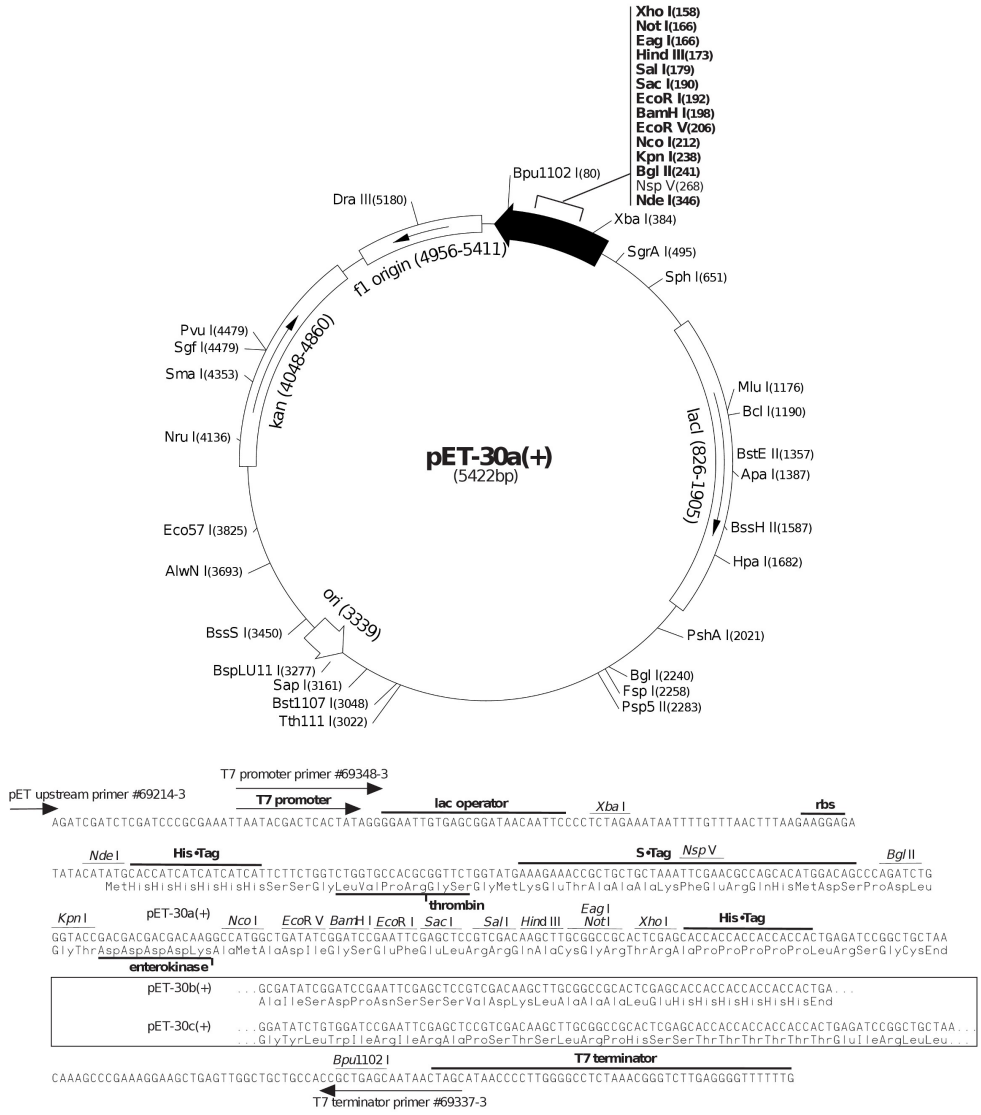


Fig. 3.1. Plasmid map of expression plasmid pET30a (Novagen, Merck KgaA, HE, Germany)**.**

3.3 Human superoxiddismutase (hSOD)

Recombinant human superoxiddismutase (rhSOD) was chosen as model protein. Up to now no toxic effect of hSOD when expressed in *E. coli* could be observed, thus the cellular burden on the host, triggered by expression of the foreign gen, is solely caused by the level of production rate (qP).

hSOD complexed with Cu/Zn in its active centre participates in antioxidant defence in the cytoplasm of nearly all cells exposed to oxygen. Especially erythrocytes and liver

cells possess a high level of active hSOD enzymes, catalysing the reduction of two oxygen radicals, resulting in the formation of hydrogen peroxide. The protein has a molecular weight of 32kDa and in its active form the enzyme consists of two not covalently bound identical subunits (153 AA each), complexed with one Zn and one Cu atom in its active centre.

3.4 Cloning methods

Cloning procedures were done according to (Sambrook and Russel, 2001). Restriction enzymes and other modifying enzymes were purchased from New England Biolabs (Ipswich, MA, USA) and applied according to the manufacturer's recommendations. All primers and synthetic oligonucleotides were acquired from Sigma Aldrich (St. Louis, MO, USA).
Primer design, *in silico* cloning, sequence analysis and secondary structure prediction was done with CLC main workbench. Correct insertion of fragments was confirmed using PCR over the adequate sequence region and amplification products were verified with regard to length through agarose gel electrophoresis. Clones possessing a correct cloning product were amplified in liquid medium and the respective plasmid was isolated using a standard plasmid preparation kit (Wizard Plus SV Minipreps DNA purification System from Promega, Cat. No: A1330). Purified plasmids were sent to AGOWA (AGOWA genomics GmbH, B, Germany) for sequencing. Due to the fact that most of the samples contain sequence regions showing pronounced secondary structures, sequencing was done in the presence of denaturing reagents in single economy read mode.

3.4.1 Quantification of DNA and RNA

For nucleic acid concentration determination the NanoDrop ND-1000 Spectrophotometer (Nanodrop Technologies, Inc. USA) was used for both RNA and DNA measurement. That kind of measuring is characterised by an extremely low amount of needed sample volume. 2µl of a DNA/RNA solution are enough for measuring concentrations ranging from 20ng/µl up to 3700ng/µl. Measurement is conducted by a spectrum range of λ values starting from 220nm to 750nm. DNA or RNA absorb ultraviolet light with an absorption peak of 260nm. Protein contamination is detected by absorbance of UV light at a wavelength of 280nm, because at that wavelength the aromatic residues of amino acids absorb light.

3.4.2 Agarose gel electrophoresis

Agarose gels were used to determine the length and concentration of DNA samples. DNA carries a negatively charged phosphate backbone, thus migrates to the positively charged pole after application of an electric field. The velocity of migration in a closed meshed agarose gel strongly depends on the length of the dsDNA strand. For visualization the gel gets supplied with the dye ethidium bromide, which resembles nucleic bases and gets incorporated within dsDNA. Irradiation with UV light results in fluorescence of ethidium bromide, hence DNA gets visible.

For screening purpose 2% agarose gels were used in order to separate fragments only differing in around 50 nucleotides of length. For separation of longer fragments (e.g shortened pET30a or fragments resulting from a control digest) a 1% agarose gel was utilized. The respective amount of agarose was melted in 1xTAE buffer (50xTAE stock solution: Tris base 242g, glacial acetic acid 57,1ml, 0.5M EDTA 100ml, pH=0,8), and 360ml of the agarose mix was supplied with 9µl of ethidium bromid stock solution (10mg/ml). Analytical gels were run at 120V and 500mA and preparative gels at 90V and 500mA.

3.4.3 DNA preparation

Standard Plasmid isolation

Isolation of plasmid DNA was done by using Wizard Plus SV Minipreps DNA purification System from Promega (Madison, WI, USA). For achieving cell material a single colony of an adequate streak out was inoculated into 7ml of LB medium (1% NaCl, 1% bacto-tryptone, 0,5% yeast extract) supplied with 7µl (1:1000) of a kanamycine stock solution (50mg/ml) and incubated at 37°C at 180rpm over night. Remaining cleaning steps were done according to the manufacturer's recommendations. At the end DNA was eluted in 30µl of sterile H_2O.

Extraction of DNA out of gels and PCR purification

For many cloning purposes it is necessary to separate a desired DNA fragment from by products by using agarose gel electrophoresis. After separation adequate bands were excised under observation with UV light. For recovering DNA, the gel slice was treated by using Wizard SV Gel and PCR Clean-Up System from Promega (Madison, WI, USA).

Sample treatment was done according to the manufacturer's recommendations. The same kit was also applied to clean up DNA fragments synthesised by PCR, in order to get rid of remaining reaction components, which may cause some problems in subsequent cloning steps. In both cases DNA was eluted in 20µl of aseptic H_2O.

- ***Plasmid isolation for in vitro transcription***

In vitro transcription has to be carried out using extremely pure plasmid preparations as template. For that purpose a standard isolation kit as described above turned out not to be adequate for the production of pure DNA solutions. Thus plasmids were isolated with the QIAfilter Plasmid Midi kit from QIAGEN (Venlo, LI, Netherlands). The kit enables the isolation of plasmids with sizes up to 150kb. As binding tips QIAGEN-tips 100 were utilized, possessing a maximal binding capacity of 100µg plasmid DNA, whereby real yield strongly depends on used host strain, plasmid copy number, culture volume and media composition. Sample preparation was done according to the manufacturer's recommendations.

As pre-culture a single colony was inoculated in 2ml LB medium supplied with 2µl of the kanamycine stock solution, resulting in an antibiotic working concentration of 50µg/ml. The tube was incubated for 8h at 37°C at 180rpm. Afterwards 200µl of the pre-culture was pipetted into 100ml LB medium supplied with kanamycine, resulting a dilution of 1:500 of the pre-culture. The culture flask was put into the incubator for over night incubation (37°C, 180rpm). At the end plasmid DNA was resolved in 100µl $T_{10}E_1$ (10mM Tris-HCl, 1mM EDTA; pH 7,5).

3.4.4 PCR

As PCR device the Eppendorf Mastercycler Gradient was used. Primers were ordered from Sigma Aldrich (St. Louis, MO, USA) in desalted or lyophilized form. In order to get a stock-solution, lyophilized primers were diluted in sterile water to gain a 100µM solution. For PCR reaction primers were employed in a concentration of 10pmol/µl.

- ***Screening PCR***

Due to select resistant clones carrying a correctly inserted cloning fragment, a PCR reaction was carried out. Either colonies were directly picked up from adequate plates and were directly added to the PCR mixture or plasmids were primarily isolated (see 3.4.3)

in order to use pure plasmid DNA (~1ng/µl) as PCR template. For screening PCR Taq polymerase from New England Biolabs (Ipswich, MA, USA) was utilized. Most of the used templates possessed cloned fragments with highly structured secondary elements, thus DMSO as denaturing reagent was used. For primer sequence see Appendix.

Contents	Volume
H_2O	22,25 µl
10xThermopolymerase buffer	3 µl
dNTPs mix (10mM each)	1 µl
DMSO	0,5 µl
primer T7DFor (10pmol/µl)	1 µl
primer T7DBack (10pmol/µl)	1 µl
Taq polymerase (5000U/ml)	0,25 µl
template DNA (~2ng/µl)	1 µl
reaction volume	30 µl

***Tab. 3.1.* Screening PCR reaction mix**

Temperature	Time	Cycles
98°C	2 min	1x
98°C	15 sec	25x
55°C	15 sec	25x
72°C	45 sec	25x
72°C	5 min	1x

Tab. 3.2. Screening PCR program

PCR for SOD amplification

The gene coding for hSOD was amplified with primers containing recognition sequences of restriction enzymes XbaI and BamHI at their 5' ends. The resulting fragment was used for cloning the hSOD gene into a original pET30a vector. As PCR template a former assembled pET30a containing an unfavourable cloned hSOD gen was utilized. KOD Hifi Polymerase purchased from Novagen (Merck KgaA, HE, Germany) was used. For primer sequence see Appendix.

Contents	Volume
H_2O	28,6 µl
10xBuffer#1	5 µl
dNTPs mix (2mM each)	5 µl
$MgCl_2$ (25mM)	2 µl
SOD_ BamHI_back (10pmol/µl)	4 µl
XbaI_rbs_kurz_for (10pmol/µl)	4 µl
KOD Hifi (2,5U/µl)	0,4 µl
template DNA (~2ng/µl)	1 µl
reaction volume	50 µl

Tab. 3.3. SOD amplification PCR reaction mix

Temperature	Time	Cycles
98°C	1 min	1x
98°C	15 sec	25x
49°C	2 sec	25x
72°C	10 sec	25x
4°C	hold	1x

Tab. 3.4. SOD amplification PCR program

Inverse PCR

To delete the intrinsic T7 termination signal on the original pET30a plasmid an inverse PCR was designed. Primers are constructed to bind to DNA templates in opposite direction according to their 3'end. Thus the DNA sequence present between the two primers binding sites gets deleted. Again KOD Hifi Polymerase (Merck KgaA, HE, Germany) was chosen for PCR reaction. For primer sequence see Appendix.

Contents	Volume
H_2O	28,6 µl
10xBuffer#2	5 µl
dNTPs mix (2mM each)	5 µl
$MgCl_2$ (25mM)	2 µl
primer 1_20 (10pmol/µl)	4 µl
primer_end (10pmol/µl)	4 µl
KOD Hifi (2,5U/µl)	0,4 µl
template DNA (~2ng/µl)	1 µl
reaction volume	50 µl

Tab. 3.5. Shortening PCR reaction mix

Temperature	Time	Cycles
94°C	1 min	1x
94°C	30 sec	30x
60°C	30 sec	30x
72°C	1 min	30x
4°C	hold	1x

Tab. 3.6. Shortening PCR program

3.4.5 Annealing of termination signals

Because of their highly secondary structured character, termination signals are tricky to ligate into a vector. Termination signals were ordered from Sigma Aldrich (St. Louis, MO, USA) as single stranded complementary nucleotide strands, possessing a phosphorylated 5' overhang according to the desired cloning site. The success of annealing those single stranded oligonucleotides strongly depends on the chosen temperature program and the employed nucleotide concentration. Annealing was carried out in Eppendorf Mastercycler Gradient and was verified through agarose gel electrophoresis. As starting material 5µl of the respective oligonucletide solutions (100pmol/µl) were pipetted together, put into the mastercycler and the temperature program shown in Tab. 3.7 was applied. Afterwards 1µl of the sample (total volume 10µl) was loaded on a 2% agarose gel.

Temperature	Time	Cycles
94°C	1 min	1x
-0,5°C	40 sec	60x
-0,5°C	20 sec	60x
4°C	hold	1x

Tab. 3.7. Annealing program to get dsDNA

3.4.6 Ligation of annealed termination signals.

Several dilutions (1:10, 1:100, and 1:200) from annealed oligos were prepared in sterile H_2O. Ligation was carried out with T4 DNA ligase from New England Biolabs at 16°C over night. The reaction mix comprised a varying volume of restricted and dephosphorylated vector DNA containing about 50ng DNA in total, 2µl of respective insert dilutions, 1µl of 10x ligase buffer and 0,5µl of T4 DNA ligase. The mixture was filled up with sterile H_2O in order to achieve a total reaction volume of 10µl. After incubation ligase got inactivated by incubation at 65°C for 10min. In order to get rid of salts within reaction mix, the ligation mixture was subjected to isopropanol precipitation (see 3.4.7).

3.4.7 Isopropanol precipitation

DNA of the ligation mixture was precipitated in the presence of isopropanol and sodium acetat in order to get rid of buffer components. For that purpose 6µl (0,6 fold volume) of isopropanol and 1µl (1/10 volume) 3M sodium acetat were added and shortly mixed by vortexing. Then the sample was centrifuged for 15min at 4°C at 13000rpm in a microcentrifuge and the supernatant was discarded by pipetting. Afterwards precipitated DNA was washed with 100 µl of 70% ethanol, centrifuged for another 10min under the same conditions. Supernatant was removed and the resulting pellet was dried by air for approximately 10min at 30°C on a heating block. Dried pellet got invisible, indicating that remaining ethanol was fully evaporated. Then 10µl of sterile H_2O was added and this pure DNA solution was used for following transformation (see 3.4.8.).

3.4.8 Transformation

Transformation of plasmid DNA was done through electroporation using a Biorad gene pulser (Micropulser) device. As host purchased electrocompetent DH5α (Novagen, Merck KgaA, HE, Germany) cells were used. 50µl of thawed electrocompetent cells were given into disposable cuvettes (wrapped with conductive aluminium foil), mixed with 10µl of precipitated DNA from ligation approach, and then an external electric field (1000Ω and 25µF) was applied. After the electric pulse cells were incubated at 37°C for 1h in 950µl of prepared SOC medium (2% bacto-tryptone, 0,5% yeast extract, 10mM NaCl, 3mM KCl, 20mM $MgCl_2$*$6H_2O$, 20mM glucose*H_2O, 10mM $MgSO_4$*$7H_2O$; pH=7) for cell recovery. Afterwards cell suspension was poured onto agar plates (1% NaCl, 1% bacto-tryptone, 0,5% yeast extract, and 1,5% agar-agar) supplied with adequate antibiotics. Plates were incubated at 37°C over night.

3.5 In vitro transcription

After construction of several pET30a derivatives possessing altered termination signals, their individual capability to terminate T7 polymerase mediated transcription was verified by using AmpliScribe™ T7 High Yield Transcription Kit from Epicentre Biotechnologies (Biozym Scientific GmbH, NI, Germany).

3.5.1 In vitro transcription assay

As template pure plasmid solutions purified with QIAfilter Plasmid Midi kit (see 3.4.3) were employed. Prior to use templates were linearised by restriction with SmaI (see Fig. 3.2). That restriction site is present in all templates and after digest a double stranded DNA with blunt ends is produced. The SmaI site is located about 1000bp downstream of the MCS of pET30a, representing the longest possible transcription product. Termination products and readthrough products differ in size of about 1000bp and were distinguishable after separation by electrophoresis. About 2µg of distinct templates DNA was digested with SmaI for 3h at 25°C. Subsequently the restriction enzyme was inactivated by incubation at 65°C for 20min and the reaction mixture was purified with Wizard SV Gel and PCR Clean-Up System (see 3.4.3). A volume containing about 200ng of linearised and purified plasmid DNA was used for subsequent *in vitro* transcription. The reaction was carried out as depicted in Tab. 3.8.

Contents	Volume	Final concentration
H_2O	6µl	-
200ng template DNA	2µl	10ng/µl
10x AmpliScribe T7 reaction buffer	2µl	1x
100mM ATP	1,5µl	7,5mM
100mM CTP	1,5µl	7,5mM
100mM GTP	1,5µl	7,5mM
100mM UTP	1,5µl	7,5mM
100mM DTT	2µl	10mM
AmpliScribe T7 enzyme solution	2µl	-
reaction volume	20µl	-

Tab. 3.8. In vitro transcription reaction mix

The whole transcription approach was incubated at 37°C for 2h-3h. Subsequently the optional Dnase I digest was carried out in order to get rid of template DNA. For that purpose 1µl of Rnas-Free Dnase I solution (1MBU/µl) was added to the *in vitro* transcription reaction mixture and digest was done for additional 15min at 37°C. One Molecular Biology Unit (MBU) of Dnase I is defined as amount of enzyme sufficient to digest 1µg of pUC19 DNA to oligodeoxynucleotides in 10 minutes at 37°C. Finally RNA was isolated as described in 3.5.2.

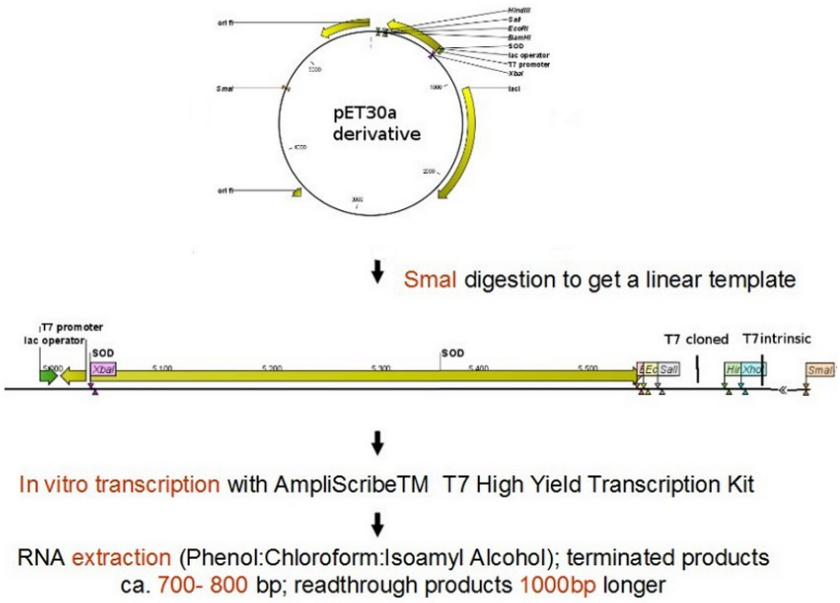


Fig. 3.2. Experimental design of *in vitro* transcription assays.

3.5.2 *Isolation of produced RNA*

Chloroform/Phenol extraction

To separate RNA from protein fraction a Chloroform/Phenol extraction was carried out. For a more practicable sample volume 280µl of H_2O were added to the *in vitro* transcription assay, resulting in a total volume of 300µl. Subsequently the same amount of a Chloroform/Phenol/Isoamylalcohol solution was added (Tris-buffered 50% phenol, 48% chloroform, 2% isoamylalcohol; pH4) to the reaction mixture, the tube was shortly put on a vortexer (5sec), incubated at room temperature (RT) for 5min and centrifuged at 13000rpm at RT for 2min. After centrifugation the RNA fraction should be present in the upper positioned aqueous phase, while proteins and DNA (already digested by Dnase I in that experiment) partition should exist in interphase and organic phase. 200µl of the aqueous phase were carefully transferred to a new Eppendorf tube without disturbing phase separation. Afterwards one volume (200µl) of water-saturated chloroform was added, again mixed for 5sec by vortexing, and centrifuged under the same conditions. For subsequent isopropanol preciptiation 100µl of the supernatant were transferred to a new Eppendorf tube.

- ***Modified Isopropanol precipitation***

The precipitation procedure is similar to the one described in 3.4.7, but is modified according to special requirements for RNA precipitation. 0,6 fold volume of isopropanol (99%) and 1/10 volume 3M sodium acetat were added and shortly mixed by vortexing. Afterwards the mixture was incubated on ice for 30min, centrifuged at 4°C at 13000rpm for 30min and the supernatant was decanted. For further cleaning 700µl 70% ethanol was added, the solution was again centrifuged at the same conditions but only for 15min and the supernatant was removed by pipetting. As observed by subsequent transcript analyses, RNA samples treated with 95% ethanol offer more precise peaks in subsequent gel electrophoresis analysis (see 3.5.3), a putative purification step was introduced. For that purpose 500µl of 95% ethanol were added to the pellet and centrifuged at 4°C at 13000rpm for additional 15min. After discarding the supernatant the RNA pellet was dried by air, resuspended in 30µl RNase free H_2O, incubated for 10min on ice and then frozen at -80°C.

3.5.3 Transcript analysis and calculation of TE

Purified RNA samples were analysed according to their size and quantity using the Bioanalyzer 2100 system provided from Agilent Technologies (Santa Clara, CA, USA). In principal the system relies on traditional electrophoresis but has been transferred to chip format, by etching micro channels into glass serving as supporting material. The chip format dramatically reduces separation time as well as sample and reagent consumption. In detail just 1µl comprising RNA of 25ng-250ng are sufficient for a single run. The system provides automated sizing and quantification information in a digital format.

For separation the micro channels get filled with a sieving polymer and fluorescence dye. After loading the samples a voltage is applied, thus charged biomolecules like nucleic acids migrate to positively charged pole. Because of a constant mass-to-charge ratio and the presence of a sieving polymer matrix, the molecules are separated by size. Smaller fragments are migrating faster than larger ones. Dye molecules intercalate into DNA or RNA strands and those complexes are detected by laser-induced fluorescence measurement. Data are translated into gel-like images (bands) or electropherograms (peaks). With the help of a ladder that contains components of known sizes, a standard curve of migration time versus fragments size is plotted. From the migration times measured for each fragment in the sample, the size is calculated. The ladder also contains components

of a distinct concentration, thus quantitation can be done by calculation of the ladder area and subsequent comparison with sample peak areas (http://www.chem.agilent.-com).

For separation and quantification of *in vitro* transcribed RNA the Agilent RNA 6000 Nano LabChip® kit was used. Sample preparation and analysis was done according to the manufacturer's recommendations. For all runs approximately 1µl containing 200ng of RNA were loaded on the chip. Each *in vitro* transcription assay was verified by loading at least 3 samples on one chip. Most of the RNA samples stemming from one transcription assay were also verified by loading on two different chips. In both cases standard deviation of calculated termination efficiency (TE) varied in a very small range and never exceeded ±0,4. In rare cases TE was also calculated from different *in vitro* transcription assays, whereby calculated TE showed higher standard deviation values of about ±1%. That low variety shows, that analyses carried out with the Bioanalyzer system are characterised by pronounced accuracy and reproducibility.

After separation of RNA transcripts according to their size the amount of the respective RNA fraction was assessed by calculation of area peak and subsequent comparison to ladder area. For assessing the termination efficiency the amount of RNA has to be transferred into molar ratios. Due to the known nucleotide composition of the gained transcripts, the molarity could be calculated according to equation (1), whereby A_n, U_n, C_n and G_n indicate the number of respective ribonucleotides.

$$MW = (A_n x\, 329.2) + (U_n x 306.2) + (C_n x\, 305.2) + (G_n x\, 345.2) + 159 \quad (1)$$

Termination efficiencies were calculated as the molar ratio between terminated transcript and the sum of terminated and read-through transcripts, and an average of at least three measurements was taken.

3.6 Preparation of cell banks

3.6.1 Research cell bank (RCB)

A single colony of a new strain was streaked out on a new LB agar plate (containing the appropriate antibiotics). The plate was incubated over night, and a single picked colony was cultivated in 3ml selective LB-medium. Culture was incubated over night or for several hours. Afterwards one volume of culture was mixed (vortex) with one volume of 87% sterile glycerol. About 1,2ml of resulting mixture was transferred to pre-cooled cryo-tubes and immediately frozen at -80°C.

3.6.2 Master cell bank (MCB)

A streak out of the corresponding strain was incubated over night. At the next day a single colony was inoculated into a sterile 250ml flask filled with 80ml of M9ZB medium (bacto-tryptone 10.0g, yeast-extract 5.0g, NaCl 5.0g, NH_4Cl 1.0g, KH_2PO_4 3.0g, Na_2HPO_4 6.0g, glucose*H_2O 4.0g and $MgSO_4$ (1M) 1ml; amounts for 1 litre of medium) and supplied with the proper amount of antibiotic. The flask was incubated for approximately 20min at 180rpm in order to ensure well mixing and a homogeneous solution. Afterwards the volume was equally divided into two 250ml flasks, each containing 40ml of the pre-culture. One of the flasks was used for the preparation of the proper MCB and the second was adducted for OD_{600} measurement. In addition a third 250ml flask solely comprising sterile M9ZB medium was used for sterile test. All three flasks were incubated at 37°C and 180rpm until measured OD_{600} in the reference flask reached a value of 3. Subsequently 1ml of the cell suspension of of the MCB flask was removed in order to accomplish Koch plating as described in 3.7.2. Remaining solution was mixed with the same volume of 87% glycerol and proportioned (~1ml) into pre-cooled cryo-vials. Vials were frozen at -80°C for storage.

3.7 Fermentation

At the beginning of recombinant protein production organism were normally grown in a process called batch cultivation. In that simple approach a vessel containing a distinct volume of growth medium gets inoculated with a small volume of starter culture. Cells are grown under aseptic conditions until a defined cell density is reached or exhausting substrates enforce stationary growth. Micro organisms get harvested and a set of purification processes ensure the receipt of desired product.

Nowadays batch cultivation is mostly replaced by fed batch cultivation, a system with several advantages. Primarily cells are also grown in batch mode, but as soon as process data (pO_2 measurement) indicate a decline in available substrates, a continuous feed of substrates starts. Feed phase is adjusted in accordance to exponential growth of bacterial cells, thus influx of feed medium exponentially increases with cultivation time. The rate of substrate addition strongly influence growth rate of bacteria, thus resembles a powerful tool to adjust growth rate (μ) to a desired value, making the whole process more predictable and reproducible. Beside the advantage of set up a distinct growth rate, proliferation of *E. coli* in fed batch mode also avoid the accumulation of disadvantageous by products like ethanol or acetat within medium, resulting in higher achievable cell densities. Because of the relative complex adjustment to achieve such a tight process control, fed batch cultivations are carried out in computer controlled bioreactors.

3.7.1 Cultivation conditions

All fermentations were run in fed batch mode. Adjusted growth rate, amount of inducer and medium feed rate were subjected to induction strategies (see Process design and induction of recombinant protein). As inoculum 1ml (OD_{600}) of a thawed master cell bank (MCB) vial was injected under aseptic conditions. Fermentation process was run under following process parameters and observed with indicated devices.

Bioreactor

The bioreactor used was a 20-L reactor (14 L net volume) from MBR Bioreactor AG (Wetzikon, Switzerland), equipped with standard control units (Siemens PS7, Intellution iFIX).

Temperature

Temperature was maintained at 37°C ± 0.5°C and measured with a Pt100 temperature probe.

pH

pH was maintained at pH 7.0 ± 0.05 by addition of 25% ammonia solution (ACROS Organics). For calibration commercially available buffer solutions were applied.

pO_2

Measurement was performed with a Clark probe. Calibration was done after sterilization of bioreactor at fermentation conditions (37° and 800rpm). For setting point for 0% saturation N_2-gas or an O_2-simulator (Mettler Toledo) were used. For assigning 100% saturation compressed air was streamed in (12l/min).

During fermentation pO_2 was constantly retained above 30% by regulating inlet air flow. When the maximum flow rate of 12l/min was reached, amount of solved oxygen was adjusted by changing stirrer speed (800 to 1200rpm).

Fed-batch cultivation

The substrate was added to the bioreactor following an exponential profile in order to keep the growth rate at the desired value. As equation (2) shows, the key parameter for the exponential feed regime was the amount of biomass present in the bioreactor after the batch phase X_0, since this biomass had to be supplied with an adequate amount of substrate.

$$X_1 = X_0 \cdot e^{\mu \cdot t} \tag{2}$$

Equation (3) shows the correlation between the amount of feed media required to produce the biomass present at the end of the batch phase (S_0), the yield coefficient ($Y_{X/S}$), the substrate concentration (c_s) and X_0.

$$S_0 = \frac{X_0}{Y_{X/S}} \cdot c_s \tag{3}$$

The calculation of the amount of media S required in a time interval t, when biomass X was produced at a constant μ, is defined in equation (4).

$$\Delta S = S_0 \cdot (e^{\mu \cdot \Delta t} - 1) \quad (4)$$

The media tank was placed on a balance to calculate S_{target} permanently after starting the exponential feed regime. Based on the initial weight of the media tank, an exponentially decreasing set point curve was calculated, shown in equation (5).

$$S_{target} = S_{start} - \Delta S \quad (5)$$

Feed control was achieved by increasing the pump speed according to the exponential growth algorithm (see equation (2)), with superimposed feedback control of weight loss of the substrate tank.

Addition of antifoam

Foaming was suppressed by addition of 0.5 ml antifoam (PPG 2000 Sigma Aldrich) per litre media.

Media composition

All chemicals were purchased from Merck unless otherwise noted. The minimal medium used for these cultivations contained 3 g KH_2PO_4 and 6 g K_2HPO_4*$3H_2O$ per litre. These concentrations provided the required buffer capacity and served as P- and K-source as well. The other components were added in relation to gram cell dry mass (CDM) to be produced: Na_3-citrate *$2H_2O$ (ACROS organics) 0.25g, $MgSO_4$*$7H_2O$ 0.10g, $CaCl_2$*$2H_2O$ 0.02g, trace element solution 50 μl and glucose*H_2O 3g. Additional to the trace element solution for cultivations with recombinant hSOD 4mg $CuCl_2$*$2H_2O$ and 3.2mg ZnSO4*$7H_2O$ per g CDM was added. Trace element solution: prepared in 5N HCl ($g*l^{-1}$): $FeSO_4$*$7H_2O$ 40.0, $MnSO_4$*H_2O 10.0, $AlCl_3$*6H2O 10.0, $CoCl_2$ (Fluka) 4.0, $ZnSO_4$*$7H_2O$ 2.0, Na_2MoO_2*$2H_2O$ 2.0, $CuCl_2$*$2H_2O$ 1.0, H_3BO_3 0.50.

The minimal medium for feeding phase was designed to achieve 386g of CDM (30g/L) in total. All components except glucose needed for the fed batch medium were mixed in a volume equivalent to about 5000g. Glucose was separately dissolved in a volume equivalent to 3500g. Both solutions were independently autoclaved and mixed together after cooling down. Nitrogen level was held by adding 25% ammonia solution for pH control.

After mixing all together the net weight was determined and served as input for the automated feed control algorithm. To accelerate initial growth of the population, the complex component yeast extract (0.15 g/g CDM) was added to the minimal medium to the batch medium, moreover 2.5 g/L ammonium chloride (NH_4Cl) and 2.1 g/L ammonium sulphate ($(NH_4)_2SO_4$) were added to avoid N-limitation in batch phase. The minimal batch medium was calculated to achieve 22,5g CDM (5,62g/L), and the components were dissolved in a volume equivalent to 3800g. Glucose was again separately dissolved in a volume equivalent to 300g. After autoclaving both solutions were aseptically combined.

Process design and induction of recombinant protein

All fermentations were carried out in accordance with the procedure depicted in Fig. 3.3. After end of batch phase, noticeable by a jumping up of pO_2 level, the exponential feed regime was started. Increasing the amount of added substrate was achieved by accelerative pump speed according to the exponential growth of the biomass. The amount and the composition of feed medium was laid out to permit growth of four generations. After starting the feed, cells were grown in not induced state in order to allow adaptation to fed batch conditions. After one doubling time (around 7h) cells got induced with IPTG. For induction two different strategies were implemented. In a first experimental design cells were fully induced by single addition of highly concentrated IPTG. The required amount of inducer was calculated to supply the expected CDM of 386g at the end of the process with an IPTG concentration of 20µmol/CDM. That classical pulse induction is achieved by adding the IPTG solution aseptically with a syringe provided with a sterile filter directly into the bioreactor.

The second induction strategy was conducted according to Striedner et al., 2003 (Striedner et al., 2003). Addition of limited amounts of inducer in a constant ratio to produced biomass allows a optimal exploitation of the cell's capacity to produce recombinant protein. IPTG concentration is maintained at a physiological tolerable level of 0,9µmol/g CDM. As in classical approach cells also get induced after one generation by external addition of IPTG by using a syringe. Afterwards IPTG is feeded according to the exponential regime. For that purpose an external inducer feed, similar to that arrangement conducted for the influx of feed medium, was implemented. A tank containing the inducer solution was put on a balance and IPTG was added according to the exponential growth of bacteria. Influx of inducer solution was controlled by adjusting pump speed and was controlled due to weight loss by an automated influx algorithm. An increase in biomass is encountered by an exponential feed of inducer, resulting in a constant ratio of

inducer to biomass over the whole process. Reduction of the concentration of IPTG reduces the transcription rate, thus transcription of foreign genes is diminished to a more tolerable level, ending up in a higher product yield.

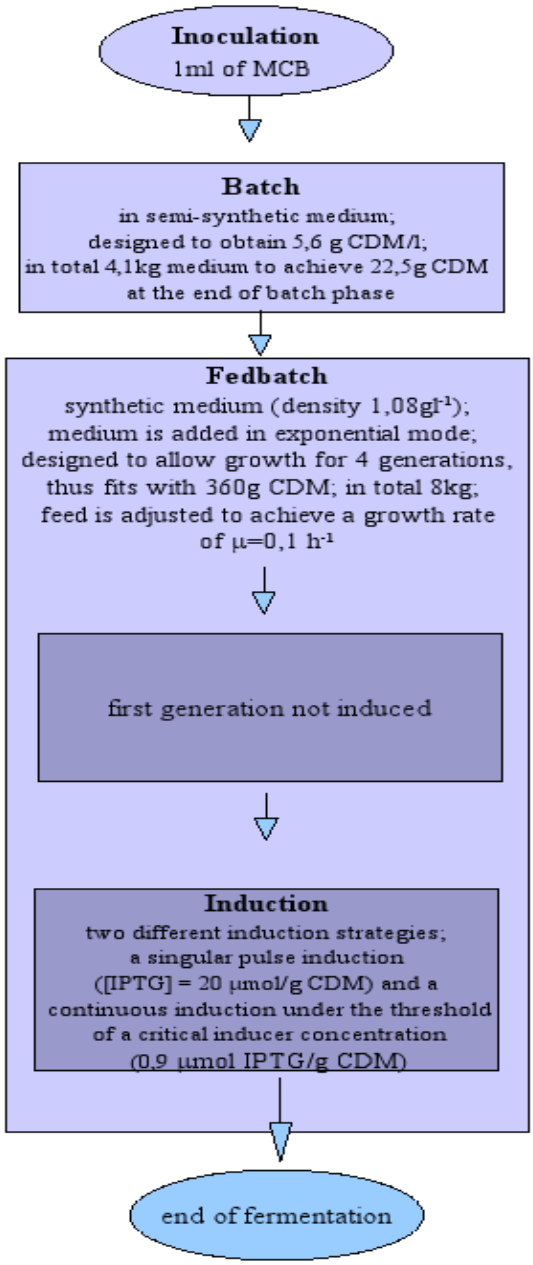


Fig. 3.3. Fermentation design for executed fed batch cultivations.

3.7.2 *Process monitoring (off-line)*

To achieve cell material for later off-line analysis, samples (15-20ml) were aseptically taken out of the bioreactor at distinct time points (see Fig. 4.1 and Fig. 4.5-Fig. 4.7) and used for following off-line analysis:

- ***Optical density OD_{600}***

To determine the turbidity of the fermentation broth samples were measured with a spectrophotometer (Amersham Biosciences Ultrospec 500 pro, Pharmacia Biotech Ultrospec 1000E) at wavelength λ= 600nm. To ensure a measurement within linear range (OD_{600} = 0,1-0,6), samples were diluted in RO-H_2O and water was also used as reference.

- ***Cell dry mass (CDM)***

10ml of cell material were aseptically taken out of the bioreactor and centrifuged for 10min at 5000rpm (Heraeus Laborfuge 200). Supernatant was discarded, cells were resuspended in 5ml RO-H_2O and centrifuged again. After removal of excess water, cells were suspended in RO-H_2O again and immediately transferred to a pre-weighed beaker. All beakers were dried for 24h at 105 °C and reweighed. For determination of total CDM in the bioreactor, the total broth volume at the time point at which samples were taken has to be calculated. The total volume in the bioreactor is given by summarizing the batch volume, the consumption of feed medium and the inflow of 25% ammonia solution. That value gets reduced by subtracting the volume of taken samples and is equal to the volume present in bioreactor at a distinct time point. Multiplying total volume with the determined CDM/ml displays the total CDM in the bioreactor at the moment of sample taking.

- ***Samples for hSOD quantification***

To achieve nearly the same amount of cells as starting material for cell disintegration, sample volumes (ml) containing about 1mg CDM were taken out of the reactor. Cells were transferred into a eppendorf tube and centrifuged for 10min at 4°C (abbr. 10 000g). The supernatant was decanted and the resulting cell pellet was stored at -20°C for further treatment (see 3.8)

- ***Determination of plasmid stability and amount of non-producing cells***

In order to determine plasmid stability, cells were plated on agar plates in the presence and the absence of kanamycin. Therefore 1ml of an adequate cell dilution series was pipetted on a petri dish and mixed with 55°C pre-warmed nutrient agar. In Accordance with its inventor that method is termed Koch plating.
Induced cells are highly stressed by producing recombinant protein, thus aren't able to proliferate in order to form colonies. To examine the number of antibiotic resistant, non producing cells, samples were also poured on agar plates containing kanamycin and IPTG. 1ml of sample got diluted in physiologically active saline salt solution (0,9%) to a final dilution of 10^{-9}. From each 10^{-7} to 10^{-9} dilution step 1ml was plated three times together with pre-warmed nutrient broth agar without kanamycin and also in the presence of kanamycin (100mg/l). From 10^{-3} to 10^{-5} dilutions 1ml was poured on nutrient broth agar plates containing kanamycin (100mg/l) and IPTG (200mg/l). Plates were incubated for 24h at 37°C in order to counter colony forming units (CFU).

3.7.3 Process monitoring (on-line)

- ***O_2/ CO_2***

The percentage of CO_2 and O_2 in exhaust gas was measured by using a system comprising a gas analyser module from ABB Inc. Measurement of CO_2 is based on absorption of light in the range of infra-red (infra-red spectrophotometry analyser type Uras). In contrast O_2 measurement is accomplished by the paramagnetic properties of oxygen (analyser type Magnos).

- ***Base consumption***

As described above pH was controlled by addition of 25% ammonia solution. The base tank was put on a balance and weight loss was detected during fermentation process. Considering the specific density (0.91kg/l) of 25% ammonia solution the consumed volume was calculated.

3.7.4 Calculation of fermentation key variables

Culture volume

The culture volume is given by summarizing the volume of initial batch medium, the influx of feed medium and consumed volume of 25% ammonia solution. That value gets reduced by the volume of taken samples, resulting in the overall volume of nutrient broth within fermenter at a distinct time point.

Growth rate

Growth rate is given as increase of biomass within time. Thus growth rate (μ) can be calculated according to equation (6).

$$\mu = \frac{1}{X} \cdot \frac{dX}{dt} \tag{6}$$

When measuring the biomass (X) at two distinct time points, growth rate could be calculated according to equation (7).

$$\mu = \frac{\ln \frac{X_1}{X_2}}{t1 - t2} \tag{7}$$

Specific product formation rate

The specific product formation rate is defined as alteration of product amount related to the present biomass (see equation (8)) with time.

$$qP = \frac{1}{X} \cdot \frac{dP}{dt} \tag{8}$$

When measuring product amount at two different time points, qP can be calculated in good approximation using equation (9).

$$qP = \frac{1}{\frac{X_1 + X_2}{2}} \cdot \frac{P_2 - P_1}{t_2 - t_1} \tag{9}$$

Determination of plasmid molecules per cell

In order to determine the number of plasmid molecules within a bacterial cell, the ratio of plasmid DNA (pDNA) to chromosomal DNA was calculated. pDNA was isolated by using a standard plasmid purification kit (Wizard Plus SV Minipreps DNA purification System from Promega, Cat. No: A1330). Because of the relative low growth rates (μ = 0,1) adjusted in the underlying fermentation processes, it can be assumed, that each bacterial cell just contains a single copy of its chromosome. Measurement of total DNA was done fluorimetrically with Höchst dye H33258 after disruption of cells with lysozyme and SDS. To determine loss of pDNA during purification, samples were treated by addition of an internal standard (pUC19) before purification. After purification the amount of plasmid DNA and its integrity were analysed through capillary electrophoresis (Agilent 2100 Bioanalyzer). The content of chromosomal DNA was determined by subtracting plasmid DNA/mg CDM from total DNA/mg CDM. The content of plasmid DNA in the sample was calculated according to equation (10). PCN was derived from the ratio of chromosomal DNA to plasmid DNA, and was evaluated according to equation (11).

$$Plasmid\ DNA\ within\ sample = \frac{Internal\ standard\ added}{Internal\ standard\ measured} \cdot Measured\ plasmid\ DNA \tag{10}$$

$$PCN = \frac{\sum Chromosomal\ base\ pairs \times Amount\ of\ plasmid\ DNA/mg\ CDM}{\sum Plasmid\ base\ pairs \times Amount\ of\ chromosomal\ DNA/mg\ CDM} \tag{11}$$

3.8 Quantification of recombinant protein (ELISA)

In order to quantify the amount of soluble hSOD in cytoplasm, an enzyme-linked immuno sorbent assay (ELISA) was performed. For that purpose hSOD gets immobilized by binding to primary SOD antibody, which in turn is fixed on the surface of an adequate micro-titer plate. After washing out excess reagents a second monoclonal mouse antibody is added. That secondary antibody recognizes the complex containing of primary antibody bound to hSOD and is linked to alkaline phosphatase, which is able to utilize the chromogenic substance p-Nitrophenylphosphat, resulting in formation of a yellowish product. The amount of coloured substance can be measured at 405nm with a photometer, thus the intensity of emitted light correlates with the amount of bound hSOD and enables the quantification of produced recombinant protein.

As starting material a cell volume ($3/OD_{600}$) corresponding to 1mg of CDM was harvested, centrifuged and the pellet was resuspended in 200µl of disintegration buffer (18mM Tris-HCI, pH 8.2; 5mM EDTA, 1mM $CuSO_4$, 0,1mM $ZnCl_2$, and 7mM β-Mercaptoethanol). After resuspension 50µl of a lysozyme (10 mg/ml) solution was added and incubated at 37°C at 350rpm for 10min. Then 750µl Triton X-100 (0.5% in 20 mM Tris buffer, pH 8.2) were pipetted and incubated at 37°C for 10 min with shaking at 350rpm. Afterwards the lysed material got centrifuged at 4°C (13000rpm) and the supernatant was transferred into a fresh Eppendorf tube and was stored at -20°C for further analysis (ELISA).

ELISA plates were coated with first antibody also called capturing antibody. Therefore antibody mouse anti-SOD mAb IAMSOD M05 (primary antibody) was diluted (200 µg/ml) in coating buffer (0.1 N $NaHCO_3$, pH 9.6±9.8). Diluted antibody solution was applied on ELISA plates, and plates were incubated for 2h at RT or overnight at 4°C. Then plates were washed three times with water using Columbus Washer (TECAN art.nr. I209 0049).

Coated ELISA plates were incubated with several dilutions of bacterial samples, starting from highest concentration of about 50ng hSOD/ml. For calibration a SOD standard with a concentration of 2µg/ml was used and also got diluted to a concentration of 50ng/ml in dilution buffer (1% bovine serum albumin in PBS). 50µl of diluted samples were added on pre-coated ELISA plates, incubated for 1h and washed three times with washing buffer (PBS, pH 7.2±7.4). After washing 50µl of mouse anti-SOD mAb IAMSOD A11H4 (secondary antibody) conjugated with alkaline phosphatase was added and incubated again for 1h. After incubation staining was induced by addition of 100µl staining solution (10ml coating buffer + 100µl PNPP stock solution) per well. Emitted light was measured 10-20min after staining induction at wavelength 405nm (reference wavelength 620nm) with a multiple channel photometer (Sunrise Remote/ Touch Screen, TECAN).

4 Results and Discussion

4.1 Investigation of a putative correlation between transcriptional termination and plasmid replication

Microarray data presented in 1.2.4 revealed different expression patterns of the two regulatory RNAs accomplishing the tight regulation of ColE1/pMB1 like origins of replication. After induction the activator of replication RNA II gets up-regulated nearly in the same extent as the gene encoding the recombinant protein. In contrast the expression level of the inhibitory RNA I just shows a moderate increase post induction. A putative explanation for that phenomenon could be the location of respective sequences on the plasmid. RNA II is encoded within oriV in the same orientation as potentially cloned genes coding for recombinant proteins. However we have to consider that there is a gap of about two thousand nucleotides between the entire MCS and the sequence encoding RNA II. Nevertheless it could be possible, that readthrough events, happening in an extent of 20% at the original T7 terminator, are able to prolong through the origin of replication.

Those read through events could dramatically increase the number of the 555bp long pre-RNA II molecules, normally synthesised by transcription induced at the natural RNA II promoter. In order to get functional RNA II primers for replication, the readthrough fragments, containing the pre-RNA II molecules at their 3' end, have to be recognized by RNAase H. Cleavage mediated through RNAse H would tremendously increase the number of available RNA II activators of replication, thus would lead to a loss of plasmid copy number control. In contrast RNA I is encoded in the opposite direction, hence isn't effected by readthrough events. The smooth rise of RNA I molecules in Fig. 1.9 could be explained by an increase of plasmid copy number, thus gene dosage of RNA I also increases and consecutive transcription of RNA I would result in a higher amount of that regulatory RNA.

To date the excessive plasmid replication observed after induction of recombinant protein (Cserjan-Puschmann et al., 1999) expression was explained by the accumulation of uncharged tRNAs (Grabherr et al., 2002). Especially the anticodon loops of some tRNAs share sequence homologies to the loops present in RNA I and RNA II. During excessive production of recombinant protein unloaded tRNAs accumulate, thus have the opportunity to interact with other molecules. Binding to RNA I would sequester the amount of

available inhibitor of replication, resulting in deregulation of plasmid copy number. By deleting the homology of RNA I loops by a simultaneous maintenance of secondary structure and melting temperature, a stabilisation of plasmid copy number during a fermentation process could be achieved.

Hence there are two different mechanism potentially explaining the rise of plasmid copy number post induction. In order to investigate the role of an insufficient transcriptional termination on deregulation of plasmid copy number control, a pET30a derivative lacking the intrinsic T7 terminator was produced. As starting material the original pET30 plasmid (Fig. 3.1) with the cloned (XbaI/BamHI) hSOD (see 3.4.4) as model protein was used. That plasmid was named Plasmid_I and functions as reference plasmid for all subsequent analyses. For deleting the T7 termination signal an inverse PCR (see 3.4.4) using primer 1_20 and primer_end (see Appendix) was applied. The resulting plasmid was named Plasmid_J and is 104bp shorter than Plasmid_I, which was used as template. After confirming the right deletion of the termination signal by sequencing the appropriate region, the shortened plasmid was transformed into *E. coli* HMS174(DE3) cells in order to investigate its behaviour during a fermentation process. The fermentation was carried out as described in 3.7. After one generation in feeding phase cells were fully induced by a high amount of inducer IPTG (20µmol/g CDM according to the calculated amount of CDM at the end of the process).

As depicted in Fig. 4.1 the absence of the transcriptional T7 terminator indeed showed some serious effects on host's growth behaviour, on product yield and on plasmid copy number control. Utilizing the wild-type plasmid, the maximal PCN comprises about 350 copies per cell and is reached at feed hour 14 and afterwards PCN gets stabilised at that value. In contrast the absence of a transcriptional terminator dramatically increases the number of plasmids within cells and at feed hour 15 the PCN reaches a value of nearly 600. In accordance to the shape of the PCN curve it could be assumed, that PCN doesn't reach the maximum at feed hour 15, rather PCN will show a further increase, but samples taken at later time points are missing. Nevertheless the data clearly reveal, that transcriptional termination indeed interfere with the control of plasmid replication. In the light of that new information the sole contribution of uncharged tRNAs on deregulation of PCN post induction can be doubted. To investigate the role of insufficient transcriptional termination on deregulation of plasmid copy number, the key objective was the creation of a stronger termination signal, showing a termination efficiency of nearly 100%.

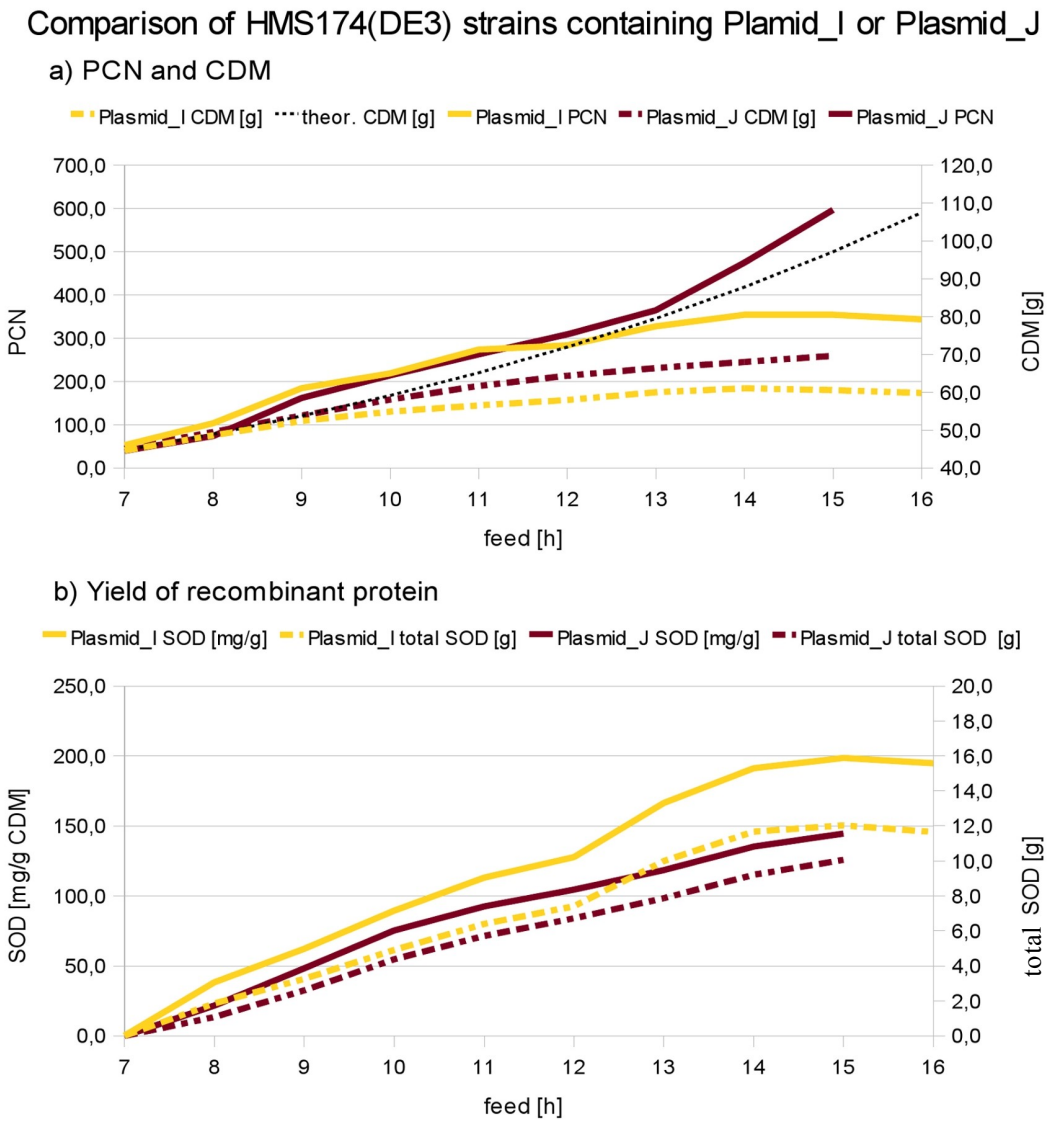


Fig. 4.1. Bioreactor cultivation of *E. coli* HMS174(DE3) strain carrying the indicated pET30a derivatives. Fermentation was carried out as described in 3.7., after one generation (7h) cells were fully induced by adding an inducer amount to achieve an IPTG concentration of 20µmol/g CDM according to the calculated amount of CDM at the end of the process. Only the first 16h after feed start are depicted, because after 16h no cell growth and no putative recombinant protein production could be detected. Data for Plasmid_J stop at feed hour 15, because the samples of feed hour 16 undergo wrong handling, thus couldn't be utilized for calculation. The black line resembles the theoretic cell growth when assuming an exponential growth without limitations.

Although cells carrying Plasmid_J show a tremendous rise of plasmids per cell (see Fig. 4.1), the putative burden to maintain the higher amount of plasmids didn't result in a lowering of growth rate. Rather the strain containing Plasmid_J reaches a higher total cell dry mass (CDM) at the end of the process, and deviation from theoretic biomass

growth occurs about one hour later (at 10h) compared to the strain harbouring Plasmid_I (at 9h). Thus cell growth phase gets enlarged for 1h compared to strains carrying the wild-type plasmid. When obtaining the product yield both the specific content of recombinant protein (hSOD/g CDM) as well as the total amount of hSOD (-17%) gets decreased, when the expression vector doesn't comprise a transcriptional terminator. Hence the lack of a termination signal leads to diminished product formation rates, resulting in a lowering of host's metabolic burden and therefore leading to an enlargement of growth phase, which is noticeable in an increase of total CDM. The metabolic burden resulting from plasmid replication in order to maintain the observed increase in PCN seems to be less important, showing little effect on cell growth.

4.2 Cloning of four different terminators

The aim of this work is to increase the termination efficiency at the transcriptional level in order to reduce the host's stress, originating from a leaky termination signal. If a more efficient termination signal leads to a stabilisation of PCN after induction of recombinant protein production, the number of available sequences encoding the recombinant protein will be drastically reduced, resulting in a decrease of the host's metabolic burden to produce recombinant proteins. It appears that the optimal hairpin for maximizing TE contains a stem of 8 or 9 mostly GC base pairs, a loop of 4-8 residues and at least a run of 6 uridines at the 3' end of the transcript and that within these structural limits, TE is moderately increased by increasing the stability of the hairpin (see 1.3). Lengthening or shortening the stem or the run of A residues (non-template strand) beyond these limits decreases (or abolishes) termination activity. Therefore a novel strategy for increasing TE is the combination of several well known terminators. Especially the presence of a class I and class II termination signal may exhibit an unexplored potential to enhance transcriptional termination. Most of the experiments in literature concerning the presence of several termination signals strongly indicate, that two termination signals don't automatically lead to higher TE. For example the cloning of the rrnC terminator sequence downstream *thr* attenuator leads to a nearly bisection of TE at the upstream located *thr* attenuator (Jeng et al., 1992). It was also shown that termination within the highly complex rrnB termination region predominantly occurs at the first T1 termination signal (Orosz et al., 1991).

Nevertheless the strategy of combining several well known termination signals was reinforced in order to increase TE. In a first experimental design four different terminators (see Fig. 4.3) were cloned (SalI/HindIII) upstream of the production plasmid pET30a

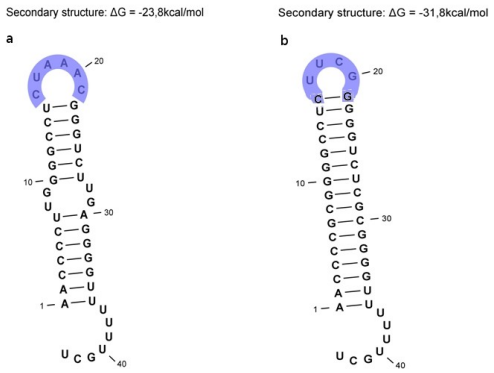


Fig. 4.2. The artificial terminator T7UUCG and the highly similar original T7 terminator. (a) Wild type T7 terminator contains a loop sequence of six nucleotides and an unpaired region comprising two unpaired G-U residues. (b) The artificial synthesized novel T7UUCG termination signal exhibit some modifications compared to the original T7 terminator, resulting in a higher overall ΔG value. The hexanucleotide loop is exchanged into the strong nucleation site UUCG. The unpaired region within stem structure is deleted by the presence of G-C base pairs instead of G-U.

(Plasmid_I) encoded T7 terminator. Three of those terminators are typical intrinsic terminators. The T7 terminator is the well known termination signal, also used in the original pET30a. T3 stems from bacteriophage T3 and its structure is similar to that of T7. The third terminator is almost identically to the original T7, but carries some mutations. The primarily hexaloop got exchanged to the extraordinary stable tetraloop UUCG (see 1.3.1.1e) and the two weak G-U base pairs within stem structure were exchanged to the more stable G-C base pairs (see Fig. 4.2).

The original T7 terminator contains more weakly bound G-U base pairs within stem structure, but some of them were described as important interaction sites with T7 polymerase (Schwartz et al., 2003). All these modifications led to an increased calculated ΔG value and could enhance TE. The last termination signal is a known class II termination signal, originating from vesicular stomatitis virus (VSV) (see 1.3.1.2). Altogether four novel plasmids, namely Plasmid_V, Plasmid_ K, Plasmid_L and Plasmid_ M (see Tab. 4.1) were generated, all containing a second termination signal in addition to the intrinsic plasmid encoded T7 termination signal. To obtain the TE of the four terminators when solely existent, the intrinsic T7 terminator of the original pET30a was deleted (Plasmid_J) and the four inserts were also cloned (SalI/HindIII) into the MCS of the resulting plasmid. Hence four putative plasmids (Plasmid_S, Plasmid_R, Plasmid_Q, and Plasmid_O) were produced (see Tab. 4.1) and those plasmids just contain a single termination signal at the 3' end of cloned hSOD gene.

All those plasmids in its linearised form were used as templates in an *in vitro transcription* assay as described in 3.5. Purified RNA fractions (see 3.5.2) were analysed with the Bioanalyzer 2100 system (see 3.5.3) in order to calculate termination efficiency. All consulted papers concerning the calculation of TE used radiolabeled α-^{32}P nucleotides for transcript synthesis and after separation by gel electrophoresis bands were detected by autoradiography in order to measure the intensity. To circumvent the use of radioactive

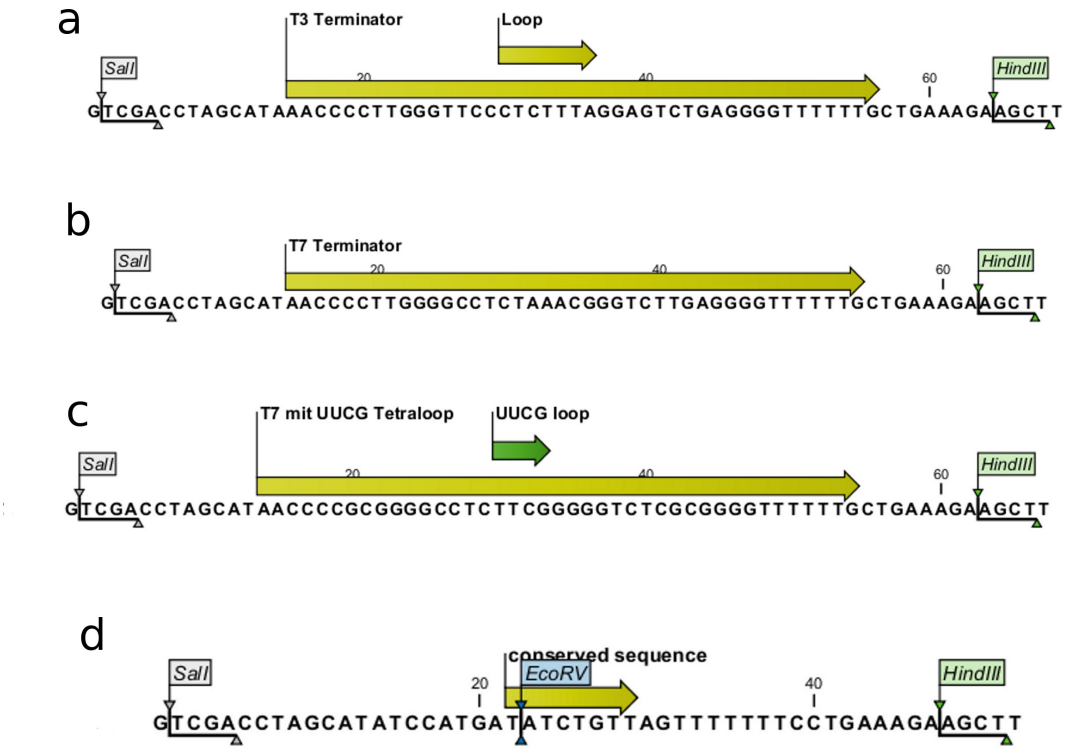


Fig. 4.3. The four termination signals which got cloned (restriction sites are indicated) into the multiple cloning site of pET30a (Plasmid_I). From up to down: the intrinsic terminator present in bacteriophage T3, the original T7 terminator, the modified T7 terminator with the altered tetraloop, and the class II terminator stemming from VSV.

substances, transcripts were analysed according to their respective length and quantity by using the Bioanalyzer system. To verify results gained from that system, Plamid_I with its original termination region containing the T7 terminator was used for *in vitro* transcription, and subsequent electrophoretic analysis was carried out by using the Bionalyzer system. It turned out that RNA purity is the most important factor influencing the calculated TE. For that purpose insufficiently pure plasmid preparations gained from Wizard Plus SV Minipreps (see 3.4.3) were replaced by more pure plasmid fractions resulting from utilizing QIAfilter Plasmid Midi kit from QIAGEN. As second action an additional purification step using 95% ethanol during RNA precipitation was implemented (see 3.5.2). Those two steps gained extremely distinct peaks when analysing the RNA transcripts with the Bioanalyzer system. Calculated TE based on those electropherograms revealed a TE of 79,45% ±0,35 at the T7 terminator present in Plasmid_I.

Hence that value fits well with a TE of 80% for T7 terminator mentioned in literature. The Bioanalyzer system seems to be an appropriate tool to measure TE and has the big advantage to avoid the handling with radioactive substances.

Analyses of the first eight generated plasmids revealed some very interesting and novel features. The individual calculated termination efficiencies are summarised in Tab. 4.1. The introduction of a second identical T7 termination signal as accomplished in Plasmid_V doesn't lead to a remarkable increase of TE. As also described in Jeng et al. the introduced termination signal just induce a bisection of termination events at the original terminator. The same result could be observed when the highly similar T3 terminator was cloned (Plasmid_K). Surprisingly the simultaneous presence of the artificial termination signal marked T7UUCG and the intrinsic T7 termination signal executed in Plasmid_L dramatically increases total termination efficiency to a value of 93,2%. Thus the adjusted modifications in order to enhance ΔG also triggers a higher TE. Nearly the same effect could be observed with Plasmid_M, which carries a combination of the two different types of Rho-independent termination signals.

As described above the four termination inserts were also cloned into the MCS of Plasmid_J in order to determine the individual TE. At the first view the results (Tab. 4.1), gained from *in vitro* transcription utilizing these plasmids as templates, are somewhat surprising. For instance the data reveal that the T7 terminator cloned into the MCS of the shortened plasmid (Plasmid_S) just exhibits a TE of 48,08%, while the same signal in the wild type plasmid (Plasmid_I) shows a calculated TE of nearly 80%. Hence the same terminator possess an extremely different capability to terminate transcription, depending on the actually surrounding sequence. That finding perfectly fits with the observation that the alteration of only few surrounding nucleotides can drastically influence TE of a distinct termination signal (see 1.3.1.1d). That dependence on encompassing sequences could also explain the difference in calculated TE when a distinct terminator is present in different cloning vectors. Structural investigation of the 104bp comprising deletion fragment was carried out by computer analyses and revealed putative hairpin structures (data not shown). Maybe those smaller hairpins aid in termination and are the reason for diminished observed TE in Plasmid_J derivatives. As seen in 4.1 all terminators cloned into the MCS of Plasmid_J exhibit a TE below 80%, thus aren't as efficient as the original T7 terminator.

Label	Used plasmid as starting material for cloning	Insert	Cloning site	TE at first terminator	TE at second terminator	TE at plasmid encoded T7	TE total (± Stdv)
S	pET30a with the deleted plasmid encoded T7 terminator	hSOD T7	XbaI/BamHI HindIII/SalI	48,08%	-	-	**48,08 %** ± 0,17
R		hSOD T7UUCG	XbaI/BamHI HindIII/SalI	67,37%	-	-	**67,37%** ± 0,34
Q		hSOD T3	XbaI/BamHI HindIII/SalI	39,45%	-	-	**39,45%** ± 0,15
O		hSOD VSV	XbaI/BamHI HindIII/SalI	75,87%	-	-	**75,87%** ± 0,32
V	pET30 with the plasmid encoded T7 terminator present	hSOD T7	XbaI/BamHI HindIII/SalI	42,68%		40,02%	**82,70%** ± 0,94
L		hSOD T7UUCG	XbaI/BamHI HindIII/SalI	58,28%	-	34,92%	**93,20%** ± 0,02
K		hSOD T3	XbaI/BamHI HindIII/SalI	30,42%	-	52,25%	**82,69%** ± 0,07
M		hSOD VSV	XbaI/BamHI HindIII/SalI	65,19%	-	27,15%	**92,30%** ± 0,04
W	pET30 with the plasmid encoded T7 terminator and the pausing signal CJ-T7 between this original terminator and the cloned one	hSOD T7 CJ-T7	XbaI/BamHI HindIII/SalI HindIII/NotI	52,65%	-	33,22%	**85,87%** ± 0,33
Z		hSOD T7UUCG CJ-T7	XbaI/BamHI HindIII/SalI HindIII/NotI	59,03%	-	28,42%	**87,45%** ± 0,09
T		hSOD T3 CJ-T7	XbaI/BamHI HindIII/SalI HindIII/NotI	36,85%	-	43,56%	**80,41%** ± 0,05
ptZENIT	pET30 with the plasmid encoded T7 terminator present and two putative terminators	hSOD T7UUCG rrnBT1	XbaI/BamHI HindIII/SalI HindIII/NotI	59,91%	32,64%	6,60%	**98,53%** ± 0,03
I	pET30 with the plasmid encoded T7 terminator present	hSOD	XbaI/BamHI			79,45%	**79,45% ±0,35**

Tab. 4.1. List of generated termination signals and their calculated termination efficiency (TE)

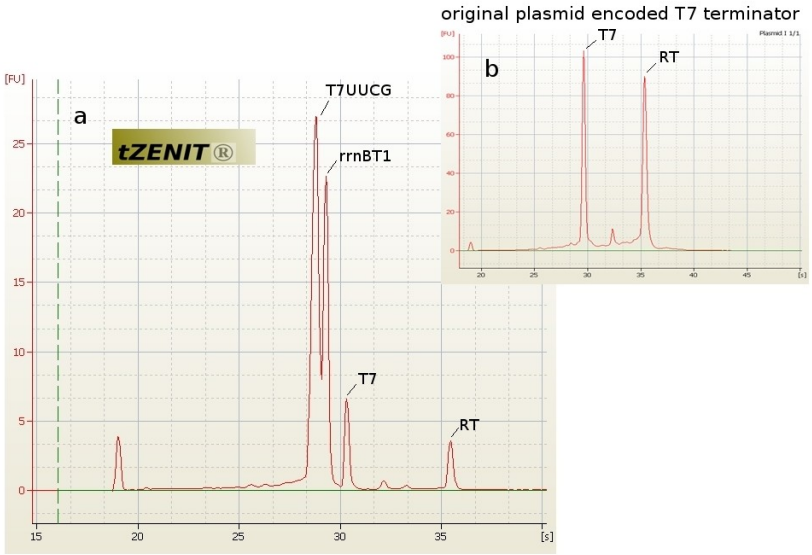
4.3 The introduction of a pausing signal

To further increase termination efficiency the pausing signal T7-CJ was cloned between the two intrinsic terminators of Plasmid_V, Plasmid_K and Plasmid_L. As described in 1.3.1.2 that pausing signal shares the consensus sequence of all known class II termination signals, but lacks an adjacent run of uridines at the 3' end. It was assumed that a sequence, which forces RNA polymerase to pause, may aid in a more efficient termination because of enlarging the time window within termination can occur. On the one hand the presence of a downstream located pausing signal could lead to a more efficient formation of hairpin structures, by enlarging the time within nucleation of newly synthesised hairpin sequences can occur. Thus hairpin formation gets favoured compared to competitive DNA-RNA hybrid formation. On the other hand a pausing signal in front of a terminator sequence may result in a deceleration of prolonging RNA polymerase, thus could enhance termination. As indicated in Tab. 4.1, the putative pausing signal indeed enhances TE at the first terminator, although total termination efficiency is just slightly increased in one case (Plasmid_W) and even decreased in the remaining two cases (Plasmid_Z and Plasmid_T). In regard to enhancing TE the cloning of the pausing signal T7-CJ wasn't successful.

4.4 Cloning of a ribosomal termination signal

Due to the data shown in Tab. 4.1, the cloning of another termination signal, the well known rrnBT1 terminator was carried out. Because of the highest measured TE of the plasmid containing the combination of T7UUCG and T7 this plasmid was taken for further cloning procedures. As described in 1.3.1.2b) terminator 1 (T1) of the enlarged termination region located at the 3' end of the rrnB gene contains both, a class I and class II termination signal within one secondary structure. Therefore the insertion of rrnBT1 provides the plasmid with a putative class II terminator, shown to successfully increase TE. In addition that ribosomal terminator will enhance secondary structures present at the 3' end, which was shown to protect mRNA from degradation by RNases (Baneyx, 1999). More stable mRNA molecules may have a benefit for recombinant protein expression and thus can result in higher recombinant protein amounts. The ribosomal terminator (sequence see Appendix) was cloned into the MCS of Plasmid_L directly following the primarily introduced T7UUCG signal. Restriction sites used for cloning were HindIII and NotI. After generating that plasmid, an *in vitro* transcription assay was carried out, and the transcripts were again separated by using the Bioanalyzer system. As seen in

Fig. 4.4 the putative termination signal again lead to a further increase of total TE. The calculated TE now shows a value of about 98,5%, and reveals this enlarged termination region as a highly efficient transcriptional terminator. Read through transcripts are dramatically reduced and because of its tremendous potential to stop transcription, this termination signal was named tZENIT, and the pET30a derivative carrying that signal was named ptZENIT.



c

Termination combination	First terminator	Second terminator	Third terminator	TE total
tZENIT®	59,91%	32,01%	6,60%	98,53%

Fig. 4.4. Bioanalyzer data achieved after *in vitro* transcription of a pET30a plasmid containing the tZENIT termination region. (a) pET30a plasmid possessing the enlarged termination region tZENIT, comprising the artificial T7UUCG, the rrnBT1 and the T7 terminator. Peaks caused by the appropriate terminator are indicated at peak tops. The amount of read through transcripts (RT) is dramatically reduced compared to Plasmid_I. (b) Bioanalyzer data of an *in vitro* transcription assay using the original pET30a plasmid with its T7 termination signal. Transcription terminates in an extent of 80%. (c) Calculated TEs of the plasmid containing the tZENIT region.

4.5 Fermentation of a strain carrying ptZENIT

As described in 4.4 the pET30a derivative ptZENIT is characterized by an extremely efficient transcriptional termination region. Under the presumption that insufficient transcriptional termination events have the capability to deregulate plasmid replication, such an enhanced termination should have a noticeable effect on plasmid copy number control. Therefore *E. coli* HMS174(DE3) cells were transformed with plasmid ptZENIT and that production strain was characterized in a bioreactor cultivation process. Fermentation was carried out according to 3.7. Recombinant protein production was induced after one generation in feed medium by injecting IPTG (20 µmol/g CDM), thus cells were fully induced.

When considering the results given in Fig. 4.5, it's obvious, that the PCN dramatically decreases from a maximum value of about 350 in reference strain (*E. coli* HMS174(DE3) with Plasmid_I) to a value of about 150 when harbouring ptZENIT. In both cases those maximal plasmid copy numbers are reached at about feed hour 14 and then are maintained over the whole process. When regarding remaining parameters like total CDM, specific product content and total hSOD amount, no deviations from the ones measured in the reference strain carrying the wild-type plasmid could be detected. The behaviour of *E. coli* HMS174(DE3) strains carrying Plasmid_I or ptZENIT are highly similar when excluding the PCN data. Nevertheless a further evidence for a putative correlation between transcriptional termination and plasmid replication was gained. Under fully induced fermentation conditions the metabolic burden stemming from recombinant protein production seems to be too high, hence reduced energy consumption for maintaining a lower plasmid copy number and the side effect of a reduced gene dosage appear to be of no consequence and don't lead to an enhanced growth of biomass.

To further characterize the consequences on host's viability when carrying ptZENIT as expression vector, a putative bioreactor cultivation was carried out. This time cells weren't fully induced, instead the IPTG concentration was maintained under the critical inducer concentration at 0,9µmol/CDM (see 3.7.1). As seen in Fig. 4.7 the more moderate process design drastically increases the viability of host's cells, indicated by a perpetuation of cell growth over the whole process. In comparison to a *E. coli* HMS174(DE3) strain carrying Plasmid_I and cultivated under same conditions, total CDM rises about +26%. That enhanced growth of bacteria cells indicates a lower stress level under production state when using ptZENT as expression vector. As already observed under standard cultivation conditions, again a reduction and stabilisation of PCN was ascertained. When considering the product yield, a remarkable increase of both specific hSOD

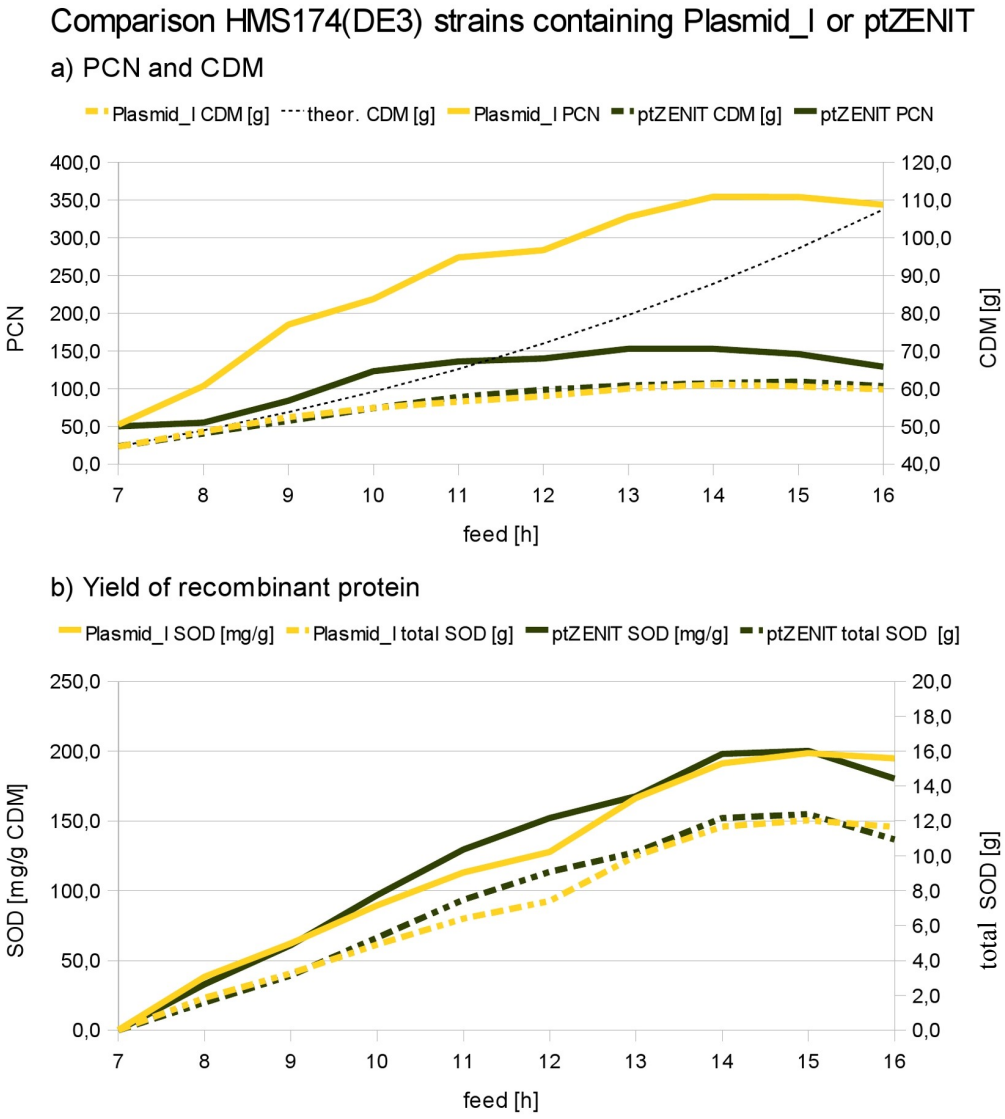


Fig. 4.5. Bioreactor cultivation of *E. coli* HMS174(DE3) strain carrying the indicated pET30a derivatives. Fermentation was carried out as described in 3.7., after one generation (7h) cells were fully induced by adding an inducer amount to achieve an IPTG concentration of 20µmol/g CDM according to the calculated amount of CDM at the end of the process. Only the first 16h after feed start are depicted, because after 16h no cell growth and no putative recombinant protein production could be detected.

content and total hSOD could be observed. Specific recombinant protein production rises from a nearly constant value of about 170mg/g CDM between feed hour 17h-28h to a value of about 240mg/g CDM in case of ptZENIT carrying cells at the same time period. Both the increase in produced biomass as well as the increase of specific hSOD formation escalates total hSOD yield to a value of about +64%. That means, that total

yield of hSOD increases from 25g to a total recovery of about 41g. Hence the improvement of the transcriptional T7 terminator signal present in the original pET30a plasmid was proved to stabilize plasmid copy number (see Fig. 4.5 and Fig. 4.6) and was also shown to possess the capability to positively influence recombinant protein yield.

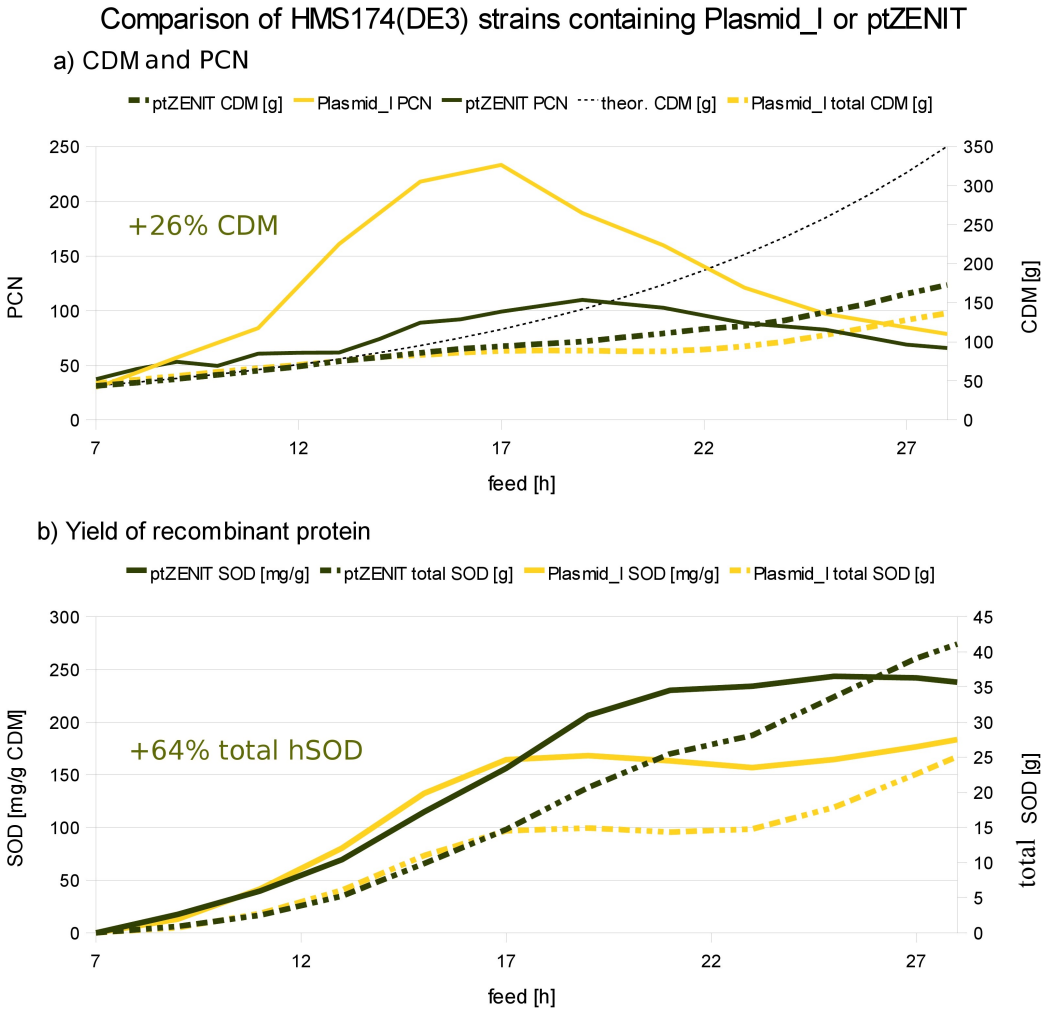


Fig. 4.6. Bioreactor cultivation of *E. coli* HMS174(DE3) strain carrying the indicated pET30a derivatives. Fermentation was carried out as described in 3.7. After one generation (7h) cells were induced by a continuous inducer feed according to the actually existent CDM (0,9µmol/g CDM). Cell growth and recombinant protein production was observed over the whole fermentation process. Cultivation was aborted after 28h fed batch phase.

The combination of modified inducer feed strategy in accordance to cell growth and the utilization of expression plasmid ptZENIT lead to an increase of total hSOD yield by a factor of nearly 3,5 compared to the standard process shown in Fig. 4.5. That example clearly shows, that we are still far away from a total understanding of occurrences happening during recombinant protein production. By adjusting such a basal issue like transcriptional termination in combination with an altered process control, the limits of plasmid encoded expression systems could be better exploited.

4.6 *Characterisation of a pET30a derivative containing no coding sequences within MCS*

Whether the excessive plasmid replication observed after inducing recombinant protein production issues from unloaded tRNAs (Grabherr et al., 2002) or is caused by read-through events into the origin of replication is still not distinguishable. An experiment, originally not carried out in order to answer that question, strongly indicates, that a leaky transcriptional terminator is the main actor causing deregulation of plasmid replication control. To quantify the stress solely originating from sustaining plasmids within a culture, the so called pMOCK plasmid was designed. That pET30a derivative lacks appropriate sequences within MCS which may be translated into protein fragments. As depicted in Fig. 3.1 even 'empty' pET30a plasmids, that means pET30a plasmids without any cloned gene sequences, contain numerous sequences coding for small peptides within MCS. Those sequences were originally designed to simplify recovery of recombinant protein. Hence the induction of *E. coli* HMS174(DE3) strains carrying the original pET30a plasmid may lead to a strong transcription of those 'tags', resulting in translation of small peptides. In that case the metabolic burden solely stemming from retaining the expression vector cannot be observed, because excessive translation can't be excluded.

The elimination of those tag sequences (bachelor thesis of Cornelia Lukasser, 2009) was done by applying inverse PCR similar to the procedure described in 3.4.4 for the generation of Plasmid_J. The resulting plasmid posses no putative sequences coding for peptides, but still harbours a functional ribosomal binding site (RBS), the strong T7 promoter with its *lac* operator sequence and the T7 terminator. That plasmid was called pMOCK and was transferred into *E. coli* HMS174(DE3) cells in order to observe its behaviour in a bioreactor cultivation. Again cells were fully induced (20µmol/g CDM) after one generation in fed batch phase. Results are shown in Fig. 4.7.

In regard to plasmid copy number the strain carrying pMOCK reveals an intermediate behaviour between the reference strain (*E. coli* HMS174(DE3) with Plasmid_I) and a *E. coli* HMS174(DE3) strain containing Plasmid_J, which lacks an appropriate termination signal. Even though the overwhelming production of recombinant proteins can be excluded and therefore also the presence of unloaded tRNA molecules, plasmid copy number is still increasing. PCN even exceeds the number observed in the reference strain and reaches a maximal value of about 470. Nevertheless cells carrying pMOCK seems to be exposed to a lower stress level, because CDM reaches an extraordinarily high value of about 250g in comparison to about 50g observed in reference strain. That fits with the already mentioned conclusion, that the induced stress level originating from recombinant protein production by far exceeds the one induced from plasmid replication.

That increase in PCN even in the absence of extensive expression of foreign gene sequences strongly indicates that indeed the excessive plasmid replication observed after addition of IPTG mainly originates from insufficient termination. Maybe both mechanisms cause deregulation of plasmid copy number control in an additive way. The wealth of new information concerning plasmid copy number control in such a strong expression system like the pET system may contribute to a better understanding of the whole process and finally may lead to a better exploitation of host's cells capacity for recombinant protein production in the future.

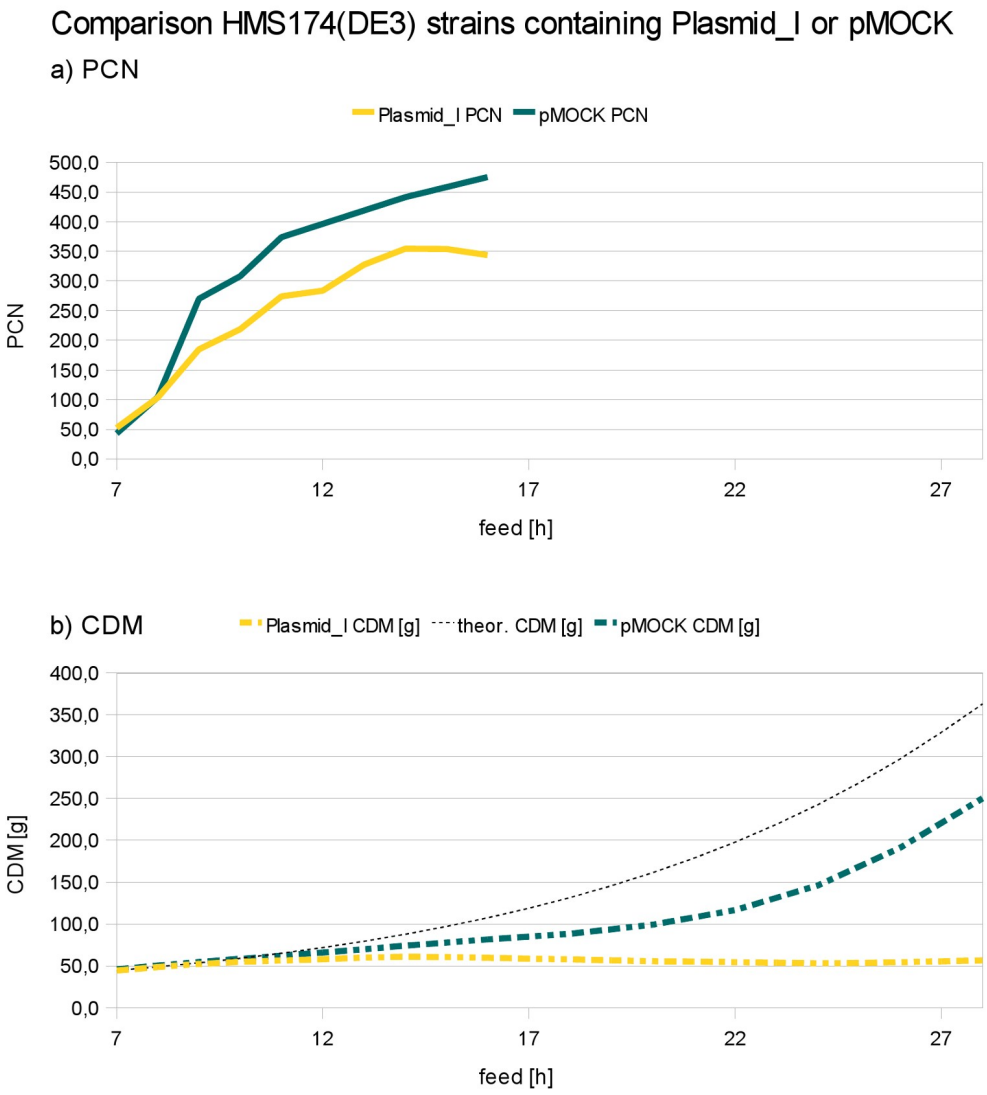


Fig. 4.7. Bioreactor cultivation of *E. coli* HMS174(DE3) strain carrying the indicated pET30a derivatives. Fermentation was carried out as described in 3.7. After one generation (7h) cells were induced by a continuous inducer feed according to the actually existent CDM (0,9µmol/g CDM). Cell growth and recombinant protein production was observed over the whole fermentation process. Cultivation was stopped after 28h fed batch phase.

5 Conclusion

The key objective of the underlying diploma thesis was the enhancement of transcriptional termination at the original T7 terminator sequence, present in all standard pET expression vectors, in order to investigate a putative correlation between read through events and an observed deregulation of plasmid replication control after induction of recombinant protein production. Because of the assumed limited gain to increase termination efficiency (TE) of the T7 terminator by sequence modifications, a combination of several termination signals was carried out. Especially the combination of the original T7 terminator with the class II termination signal VSV stemming from vesicular stomatitis virus and the combination of T7 termination signal with the artificially constructed T7UUCG revealed an obvious increase in termination efficiency (about +13% in both cases).

The cloning of the T7-CJ sequence, a signal already described to cause pausing of T7 polymerase, doesn't lead to a further increase of TE. Surprisingly the putative sequence even decreased overall termination in two cases. Finally a complex termination region comprising the artificial T7UUCG terminator, the T1 termination signal of the rrnB gene and the original T7 terminator was constructed and that configuration showed a tremendous ability to terminate T7 polymerase mediated transcription. In detail an overall TE of 98.53% was measured and because of its unique capability to stop transcription the region was named tZENIT and the resulting pET30 derivative was called ptZENIT.

Based on the presumption that readthrough events may influence plasmid replication control, series of bioreactor cultivations were carried out. To furnish a first evidence of a possible connection between an insufficient transcriptional termination and increased plasmid copy number after induction, a pET30a expression plasmid lacking any termination signal was created and characterized in a bioreactor cultivation process. In terms of plasmid copy number the analysis of the process revealed a dramatic increase of PCN, in more detail nearly a doubling of PCN compared to original pET30a plasmid was observed.

To aim at full exploitation of host cell's capacity to produce recombinant protein and to investigate its influence on PCN a strain carrying the ptZENIT expression plasmid with cloned hSOD gene was characterised by further bioreactor cultivations. To ensure constant bacterial growth rates and invariant environmental conditions, all cultivations were carried out in a bioreactor under well defined and controlled parameters, resulting in highly reproducible and comparable experimental results. Even by fully inducing

recombinant protein production (standard process design) a remarkable decrease of PCN and a stabilisation at a low level of maximal 150 copies per cell over the whole process could be obtained. Nevertheless decreased gene dosage resulting from lowering plasmid copy number doesn't influence growth behaviour, specific product formation or total hSOD yield. That clearly indicates, that under fully induced state host cells are still overstrained, even when the metabolic burden caused by plasmid replication gets decreased. For the sake of completeness the ptZENIT carrying strain was also analysed in another bioreactor cultivation with controlled inducer feed. Thereby inducer concentration was adjusted under a critical concentration and was maintained over the whole process. That process control accomplishes a more moderate gene expression, thus enables a better exploitation of host's capacity to produce recombinant proteins. The combination of a more efficient termination signal and the moderate inducer strength lead to an astonishing positive influence on PCN stabilization, bacterial growth and yield of recombinant protein. PCN was again stabilised at a lower level compared to reference strain. CDM increased about +26% and the combination of enhanced cell density with a higher specific product content per g CDM finally resulted in an increase of total hSOD of about +64%.

Those results clearly indicate, that we are still far away from a total understanding of biological processes induced by recombinant protein production and that there is still a margin to further improve expression systems in terms of protein yield and process stability.

Because of its inherent feature to improve such a basic biological process as cellular transcription, the tZENIT region may be applicable in other fields despite recombinant protein production. For example the predication of gene knock out analyses could be encouraged, by excluding possible effects at a transcriptional level as result of a read-through events into genome areas surrounding the integration site.

6 References

Baneyx F (1999) Recombinant protein expression in Escherichia coli. Curr Opin Biotechnol 10: 411-421

Benjamin Lewin, Genes IX, Inc. Ninth edition, Jones and Barlett Publishers 2008.

Brosius J (1984) Plasmid vectors for the selection of promoters. Gene **27:** 151-160

Brosius J, Dull TJ, Sleeter DD, Noller HF (1981) Gene organization and primary structure of a ribosomal RNA operon from Escherichia coli. J Mol Biol **148:** 107- 127

Carey J, Uhlenbeck OC (1983) Kinetic and thermodynamic characterization of the R17 coat protein-ribonucleic acid interaction. Biochemistry **22:** 2610-2615

Christiansen J (1988) The 9S RNA precursor of Escherichia coli 5S RNA has three structural domains: implications for processing. Nucleic Acids Res **16:** 7457-7476

Cserjan-Puschmann M, Kramer W, Duerrschmid E, Striedner G, Bayer K (1999) Metabolic approaches for the optimisation of recombinant fermentation processes. Appl Microbiol Biotechnol **53:** 43-50

d'Aubenton Carafa Y, Brody E, Thermes C (1990) Prediction of rho-independent Escherichia coli transcription terminators. A statistical analysis of their RNA stem-loop structures. J Mol Biol **216:** 835-858

Diaz GA, Rong M, McAllister WT, Durbin RK (1996) The stability of abortively cycling T7 RNA polymerase complexes depends upon template conformation. Biochemistry **35:** 10837-10843

Dong H, Nilsson L, Kurland CG (1995) Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. J Bacteriol **177:** 1497-1504

Dunn JJ, Studier FW (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J Mol Biol **166:** 477-535

Farnham PJ, Platt T (1982) Effects of DNA base analogs on transcription termination at the t ryptophan operon attenuator of EScherichia coli. Proc Natl Acad Sci U S A **79:** 998-1002

Friedman DI, Imperiale MJ, Adhya SL (1987) RNA 3' end formation in the control of gene expression. Annu Rev Genet **21:** 453-488

Gardner JF (1982) Initiation, pausing, and termination of transcription in the threonine operon regulatory region of Escherichia coli. J Biol Chem **257:** 3896-3904

Gill SC, Yager TD, von Hippel PH (1991) Escherichia coli sigma 70 and NusA proteins. II. Physical properties and self-association states. J Mol Biol **220:** 325-333

Grabherr R, Nilsson E, Striedner G, Bayer K (2002) Stabilizing plasmid copy number to improve recombinant protein production. Biotechnol Bioeng **77:** 142-147

Hartvig L, Christiansen J (1996) Intrinsic termination of T7 RNA polymerase mediated by either RNA or DNA. Embo J **15:** 4767-4774

He B, Kukarin A, Temiakov D, Chin-Bow ST, Lyakhov DL, Rong M, Durbin RK, Jeng ST, Gardner JF, Gumport RI (1992) Transcription termination *in vitro* by bacteriophage T7 RNA polymerase. The role of sequence elements within and surrounding a rho-independent transcription terminator. J Biol Chem **267:** 19306-19312

Kwon YS, Kang C (1999) Bipartite modular structure of intrinsic, RNA hairpin- independent termination signal for phage RNA polymerases. J Biol Chem **274:** 29149-29155

Ling ML, Risman SS, Klement JF, McGraw N, McAllister WT (1989) Abortive initiation by bacteriophage T3 and T7 RNA polymerases under conditions of limiting substrate. Nucleic Acids Res **17:** 1605-1618

Lyakhov DL, He B, Zhang X, Studier FW, Dunn JJ, McAllister WT (1997) Mutant bacteriophage T7 RNA polymerases with altered termination properties. J Mol Biol **269:** 28-40

Lyakhov DL, He B, Zhang X, Studier FW, Dunn JJ, McAllister WT (1998) Pausing and termination by bacteriophage T7 RNA polymerase. J Mol Biol **280:** 201-213

Lynn SP, Kasper LM, Gardner JF (1988) Contributions of RNA secondary structure and length of the thymidine tract to transcription termination at the thr operon attenuator. J Biol Chem **263:** 472-479

McAllister WT (1998) Characterization of an unusual, sequence-specific termination signal for T7 RNA polymerase. J Biol Chem **273:** 18802-18811

McDowell JC, Roberts JW, Jin DJ, Gross C (1994) Determination of intrinsic transcription termination efficiency by RNA polymerase elongation rate. Science **266:** 822-825

Mead DA, Szczesna-Skorupa E, Kemper B (1986) Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Protein Eng **1:** 67-74

Milligan JF, Uhlenbeck OC (1989) Synthesis of small RNAs using T7 RNA polymerase. Methods Enzymol **180:** 51-62

Miroux B, Walker JE (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

Muller DK, Martin CT, Coleman JE (1988) Processivity of proteolytically modified forms of T7 RNA polymerase. Biochemistry **27:** 5763-5771

Nudler E, Goldfarb A, Kashlev M (1994) Discontinuous mechanism of transcription elongation. Science **265:** 793-796

Orosz A, Boros I, Venetianer P (1991) Analysis of the complex transcription termination region of the Escherichia coli rrnB gene. Eur J Biochem **201:** 653-659

Reynolds R, Bermudez-Cruz RM, Chamberlin MJ (1992) Parameters affecting transcription termination by Escherichia coli RNA polymerase. I. Analysis of 13 rho-independent terminators. J Mol Biol **224:** 31-51

Reynolds R, Chamberlin MJ (1992) Parameters affecting transcription termination by Escherichia coli RNA. II. Construction and analysis of hybrid terminators. J Mol Biol **224:** 53-63

Rhodes G, Chamberlin MJ (1974) Ribonucleic acid chain elongation by Escherichia coli ribonucleic acid polymerase. I. Isolation of ternary complexes and the kinetics of elongation. J Biol Chem **249:** 6675-6683

Rolf Knippers, Molekulare Genetik, 8. Auflage, Georg Thieme Verlag 2001.

Sambrook J, Russel DW, Molecular Cloning: A Laboratory Manual, 3rd edn. Cold Spring Harbor 2001, ISBN: 0879695765

Schwartz A, Rahmouni AR, Boudvillain M (2003) The functional anatomy of an intrinsic transcription terminator. Embo J 22: 3385-3394

Sengupta D, Chakravarti D, Maitra U (1989) Relative efficiency of utilization of promoter and termination sites by bacteriophage T3 RNA polymerase. J Biol Chem **264:** 14246-14255

Sousa R, Patra D, Lafer EM (1992) Model for the mechanism of bacteriophage T7 RNAP transcription initiation and termination. J Mol Biol **224:** 319-334

Striedner G, Cserjan-Puschmann M, Potschacher F, Bayer K (2003) Tuning the transcription rate of recombinant protein in strong Escherichia coli expression systems through repressor titration. Biotechnol Prog **19:** 1427-1432

Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol **189:** 113-130

Telesnitsky A, Chamberlin MJ (1989) Terminator-distal sequences determine the *in vitro* efficiency of the early terminators of bacteriophages T3 and T7. Biochemistry **28:** 5210-5218

Telesnitsky AP, Chamberlin MJ (1989) Sequences linked to prokaryotic promoters can affect the efficiency of downstream termination sites. J Mol Biol **205:** 315-330

Tuerk C, Gauss P, Thermes C, Groebe DR, Gayle M, Guild N, Stormo G, d'Aubenton-Carafa Y, Uhlenbeck OC, Tinoco I, Jr., et al. (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. Proc Natl Acad Sci U S A **85:** 1364-1368

Uptain SM, Chamberlin MJ (1997) Escherichia coli RNA polymerase terminates transcription efficiently at rho-independent terminators on single-stranded DNA templates. Proc Natl Acad Sci U S A **94:** 13548-13553

von Hippel PH, Yager TD (1991) Transcript elongation and termination are competitive kinetic processes. Proc Natl Acad Sci U S A **88:** 2307-2311

Wanner BL, Kodaira R, Neidhardt FC (1977) Physiological regulation of a decontrolled lac operon. J Bacteriol **130:** 212-222

Wanner BL, Kodaira R, Neidhardt FC (1978) Regulation of lac operon expression: reappraisal of the theory of catabolite repression. J Bacteriol **136:** 947-954

Wilson KS, von Hippel PH (1994) Stability of Escherichia coli transcription complexes near an intrinsic terminator. J Mol Biol **244:** 36-51

Woese CR, Gutell R, Gupta R, Noller HF (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiol Rev **47:** 621-669

Yang MT, Scott HB, 2nd, Gardner JF (1995) Transcription termination at the thr attenuator. Evidence that the adenine residues upstream of the stem and loop structure are not required for termination. J Biol Chem **270:** 23330-23336

Zhang X, Studier FW (1995) Isolation of transcriptionally active mutants of T7 RNA polymerase that do not support phage growth. J Mol Biol **250:** 156-168

7 Abbreviations

AA	amino acid
ATP	adenosine triphosphate
bp	base pairs
cAMP	cyclic adenosine monophosphate
CAP	catabolite activator protein
cfu	colony forming unit
DMSO	dimethyl-Sulfoxide
DNA	deoxyribonucleic acid
E. coli	*Escherichia Coli*
ELISA	Enzyme-Linked Immunosorbent Assay
h	hours
hSOD	human superoxid dismutase
IPTG	isopropyl-β-D-thiogalactopyranoside
kan	kanamycine
kD	kilo Dalton
MCB	master cell bank
MCS	multiple cloning site
min	minute(s)
mRNA	messenger ribonucleic acid
OD_{600}	optical density at wavelength of 600nm
PBS	phosphor buffered saline
PCN	plasmid copy number
PCR	polymerse chain reaction
pDNA	plasmid deoxyribonucleic acid
qP	specific product formation rate
RBS	ribosome binding site
RNA	ribonucleic acid
rpm	ratations per minute
sec	second(s)
TE	termination efficiency
VSV	vesicular stomatitis virus

8 Appendix

8.1 List of used primers and inserts

Insert/primer	Sequence (5'→3')
T7DFor	-AGGGAAGAAAGCGAAAGGAG-
T7DBack	-AGATGACTTGGGCAAAGGTG-
SOD_ BamHI_back	-GTCGTCGGATCCTTACTATTGGGCGATCCC-
XbaI_rbs_kurz_for	-GTCGTCTCTAGAAATAATTTTGTTTAAC-
Primer1_20	-TTAGCAGCCGGATCTCAGTGGTGG-
Primer_end	-GGAGGAACTATATCCGGATTGGCG-
VSV1	-AGCTTCTTTCAGGAAAAAAACTAACAGATATCATGGATATGCTAGG-
VSV2	-TCGACCTAGCATATCCATGATATCTGTTAGTTTTTTTCCTGAAAGA-
T3_1	-AGCTTCTTTCAGCAAAAAACCCCTCAGACTCCTAAAGAGGGAACCCA AGG GGTTTATGCTAGG-
T3_2	-TCGACCTAGCATAAACCCCTTGGGTTCCCTCTTTAGGAGTCTGAGGGG TTTT TTGCTGAAAGA-
T7term_1	-AGCTTCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAA GGG GTTATGCTAGG-
T7term_2	-TCGACCTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTT TTT TGCTGAAAGA-
T7UUCG_1	- AGCTTCTTTCAGCAAAAAACCCCGCGAGACCCCCGAAGAGGCCCCGCG GG GTTATGCTAGG-
T7UUCG_2	-TCGACCTAGCATAACCCCGCGGGGCCTCTTCGGGGGTCTCGCGGGGTT TTT TGCTGAAAGA-
NotI_rrnBT1	- GGCCGCAGCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTTCGA CT GAGCCTTTCGTTTTATTTGA-
HindIII_rrnBT1	-AGCTTCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGT TTTA TCTGTTGTTTGTCGCTGC-
HindIII_T7-CJ	-AGCTTTGTGTCCCTATCTGTTACAGTCTCCTGC-
NotI_T7-CJ	-GGCCGCAGGAGACTGTAACAGATAGGGACACAA-

Tab. 8.1. List of used primers and inserts

8.2 ptZENIT sequence

```
LOCUS       ptZENIT                5863 bp    DNA     circular UNA 21-MAY-2002
COMMENT     http://www.informaxinc.com
FEATURES             Location/Qualifiers
     rep_origin      12..467
                     /vntifkey="33"
                     /label="f1 origin"
     CDS             complement(563..1375)
                     /vntifkey="4"
                     /label="kan sequence"
     rep_origin      2084
                     /vntifkey="33"
                     /label="ColE1 pBR322"
     CDS             complement(3518..4597)
                     /vntifkey="4"
                     /label="lac I"
     promoter        4988..5004
                     /vntifkey="30"
                     /label="T7 promoter"
     CDS             5007..5031
                     /vntifkey="4"
                     /label="lac operator"
     misc_feature    5035..5543
                     /note="originates from SOD XbaI/BgIII"
                     /label=Insert
     misc_feature    5039..5544
                     /note="originates from pET-30a(+)-2_XbaI_BglII_BamHI_XbaI"
                     /label=Insert
     CDS             5040..5543
                     /label=hSOD
     RBS             5062..5067
     modified_base   5081
                     /label="A zu G"
     modified_base   5084
                     /label="A zu G"
     misc_feature    5568..5624
                     /note="originates from tPHI UUCG rev_SalI_HindIII"
                     /label=Insert
     misc_feature    5568..5624
                     /note="originates from pET-30a(+)-2_XbaI_BglII__T7UUCG_SalI_HindIII_HindIII_SalI"
                     /label=Insert
     misc_feature    5568..5624
                     /note="originates from pET30a_hSOD_XbaI_BamHI_T7UUCG_SalI_HindII_SalI_HindIII"
                     /label=Insert
     terminator      5576..5616
                     /label="T7 mit UUCG Tetraloop"
     D-loop          5592..5595
                     /label="UUCG Tetraloop"
     terminator      5624..5698
                     /label=rrnBT1
     misc_feature    5629..5693
                     /note="originates from T1 rrnB_HindIII_NotI"
                     /label=Insert
     protein_bind    5707..5724
                     /vntifkey="31"
                     /label="His tag"
     terminator      5792..5838
                     /vntifkey="43"
                     /label="T7 terminator"
                     /label="T7 terminator intrinsisch"

   1 TGGCGAATGG GACGCGCCCT GTAGCGGCGC ATTAAGCGCG GCGGGTGTGG TGGTTACGCG
  61 CAGCGTGACC GCTACACTTG CCAGCGCCCT AGCGCCCGCT CCTTTCGCTT TCTTCCCTTC
 121 CTTTCTCGCC ACGTTCGCCG GCTTTCCCCG TCAAGCTCTA AATCGGGGGC TCCCTTTAGG
 181 GTTCCGATTT AGTGCTTTAC GGCACCTCGA CCCCAAAAAA CTTGATTAGG GTGATGGTTC
 241 ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT TTGACGTTGG AGTCCACGTT
 301 CTTTAATAGT GGACTCTTGT TCCAAACTGG AACAACACTC AACCCTATCT CGGTCTATTC
 361 TTTTGATTTA TAAGGGATTT TGCCGATTTC GGCCTATTGG TTAAAAAATG AGCTGATTTA
 421 ACAAAAATTT AACGCGAATT TTAACAAAAT ATTAACGTTT ACAATTTCAG GTGGCACTTT
 481 TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT CAAATATGTA
 541 TCCGCTCATG AATTAATTCT TAGAAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT
 601 TCATATCAGG ATTATCAATA CCATATTTTT GAAAAAGCCG TTTCTGTAAT GAAGGAGAAA
 661 ACTCACCGAG GCAGTTCCAT AGGATGGCAA GATCCTGGTA TCGGTCTGCG ATTCCGACTC
 721 GTCCAACATC AATACAACCT ATTAATTTCC CCTCGTCAAA AATAAGGTTA TCAAGTGAGA
 781 AATCACCATG AGTGACGACT GAATCCGGTG AGAATGGCAA AAGTTTATGC ATTTCTTTCC
 841 AGACTTGTTC AACAGGCCAG CCATTACGCT CGTCATCAAA ATCACTCGCA TCAACCAAAC
 901 CGTTATTCAT TCGTGATTGC GCCTGAGCGA GACGAAATAC GCGATCGCTG TTAAAAGGAC
 961 AATTACAAAC AGGAATCGAA TGCAACCGGC GCAGGAACAC TGCCAGCGCA TCAACAATAT
1021 TTTCACCTGA ATCAGGATAT TCTTCTAATA CCTGGAATGC TGTTTTCCCG GGGATCGCAG
1081 TGGTGAGTAA CCATGCATCA TCAGGAGTAC GGATAAAATG CTTGATGGTC GGAAGAGGCA
1141 TAAATTCCGT CAGCCAGTTT AGTCTGACCA TCTCATCTGT AACATCATTG GCAACGCTAC
1201 CTTTGCCATG TTTCAGAAAC AACTCTGGCG CATCGGGCTT CCCATACAAT CGATAGATTG
1261 TCGCACCTGA TTGCCCGACA TTATCGCGAG CCCATTTATA CCCATATAAA TCAGCATCCA
1321 TGTTGGAATT TAATCGCGGC CTAGAGCAAG ACGTTTCCCG TTGAATATGG CTCATAACAC
1381 CCCTTGTATT ACTGTTTATG TAAGCAGACA GTTTTATTGT TCATGACCAA AATCCCTTAA
1441 CGTGAGTTTT CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA
```

1501 GATCCTTTTT TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG
1561 GTGGTTTGTT TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC
1621 AGAGCGCAGA TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG
1681 AACTCTGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC
1741 AGTGGCGATA AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG
1801 CAGCGGTCGG GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC
1861 ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA
1921 AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT
1981 CCAGGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCGGGT TTCGCCACCT CTGACTTGAG
2041 CGTCGATTTT TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG
2101 GCCTTTTTAC GGTTCCTGGC CTTTTGCTGG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA
2161 TCCCCTGATT CTGTGGATAA CCGTATTACC GCCTTTGAGT GAGCTGATAC CGCTCGCCGC
2221 AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG CGGAAGAGCG CCTGATGCGG
2281 TATTTTCTCC TTACGCATCT GTGCGGTATT TCACACCGCA TATATGGTGC ACTCTCAGTA
2341 CAATCTGCTC TGATGCCGCA TAGTTAAGCC AGTATACACT CCGCTATCGC TACGTGACTG
2401 GGTCATGGCT GCGCCCCGAC ACCCGCCAAC ACCCGCTGAC GCGCCCTGAC GGGCTTGTCT
2461 GCTCCCGGCA TCCGCTTACA GACAAGCTGT GACCGTCTCC GGGAGCTGCA TGTGTCAGAG
2521 GTTTTCACCG TCATCACCGA AACGCGCGAG GCAGCTGCGG TAAAGCTCAT CAGCGTGGTC
2581 GTGAAGCGAT TCACAGATGT CTGCCTGTTC ATCCGCGTCC AGCTCGTTGA GTTTCTCCAG
2641 AAGCGTTAAT GTCTGGCTTC TGATAAAGCG GGCCATGTTA AGGGCGGTTT TTTCCTGTTT
2701 GGTCACTGAT GCCTCCGTGT AAGGGGGATT TCTGTTCATG GGGGTAATGA TACCGATGAA
2761 ACGAGAGAGG ATGCTCACGA TACGGGTTAC TGATGATGAA CATGCCCGGT TACTGGAACG
2821 TTGTGAGGGT AAACAACTGG CGGTATGGAT GCGGCGGGAC CAGAGAAAAA TCACTCAGGG
2881 TCAATGCCAG CGCTTCGTTA ATACAGATGT AGGTGTTCCA CAGGGTAGCC AGCAGCATCC
2941 TGCGATGCAG ATCCGGAACA TAATGGTGCA GGGCGCTGAC TTCCGCGTTT CCAGACTTTA
3001 CGAAACACGG AAACCGAAGA CCATTCATGT TGTTGCTCAG GTCGCAGACG TTTTGCAGCA
3061 GCAGTCGCTT CACGTTCGCT CGCGTATCGG TGATTCATTC TGCTAACCAG TAAGGCAACC
3121 CCGCCAGCCT AGCCGGGTCC TCAACGACAG GAGCACGATC ATGCGCACCC GTGGGGCCGC
3181 CATGCCGGCG ATAATGGCCT GCTTCTCGCC GAAACGTTTG GTGGCGGGAC CAGTGACGAA
3241 GGCTTGAGCG AGGGCGTGCA AGATTCCGAA TACCGCAAGC GACAGGCCGA TCATCGTCGC
3301 GCTCCAGCGA AAGCGGTCCT CGCCGAAAAT GACCCAGAGC GCTGCCGGCA CCTGTCCTAC
3361 GAGTTGCATG ATAAAGAAGA CAGTCATAAG TGCGGCGACG ATAGTCATGC CCCGCGCCCA
3421 CCGGAAGGAG CTGACTGGGT TGAAGGCTCT CAAGGGCATC GGTCGAGATC CCGGTGCCTA
3481 ATGAGTGAGC TAACTTACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA
3541 CCTGTCGTGC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT
3601 TGGGCGCCAG GGTGGTTTTT CTTTTCACCA GTGAGACGGG CAACAGCTGA TTGCCCTTCA
3661 CCGCCTGGCC CTGAGAGAGT TGCAGCAAGC GGTCCACGCT GGTTTGCCCC AGCAGGCGAA
3721 AATCCTGTTT GATGGTGGTT AACGGCGGGA TATAACATGA GCTGTCTTCG GTATCGTCGT
3781 ATCCCACTAC CGAGATGTCC GCACCAACGC GCAGCCCGGA CTCGGTAATG GCGCGCATTG
3841 CGCCCAGCGC CATCTGATCG TTGGCAACCA GCATCGCAGT GGGAACGATG CCCTCATTCA
3901 GCATTTGCAT GGTTTGTTGA AAACCGGACA TGGCACTCCA GTCGCCTTCC CGTTCCGCTA
3961 TCGGCTGAAT TTGATTGCGA GTGAGATATT TATGCCAGCC AGCCAGACGC AGACGCGCCG
4021 AGACAGAACT TAATGGGCCC GCTAACAGCG CGATTTGCTG GTGACCCAAT GCGACCAGAT
4081 GCTCCACGCC CAGTCGCGTA CCGTCTTCAT GGGAGAAAAT AATACTGTTG ATGGGTGTCT
4141 GGTCAGAGAC ATCAAGAAAT AACGCCGGAA CATTAGTGCA GGCAGCTTCC ACAGCAATGG
4201 CATCCTGGTC ATCCAGCGGA TAGTTAATGA TCAGCCCACT GACGCGTTGC GCGAGAAGAT
4261 TGTGCACCGC CGCTTTACAG GCTTCGACGC CGCTTCGTTC TACCATCGAC ACCACCACGC
4321 TGGCACCCAG TTGATCGGCG CGAGATTTAA TCGCCGCGAC AATTTGCGAC GGCGCGTGCA
4381 GGGCCAGACT GGAGGTGGCA ACGCCAATCA GCAACGACTG TTTGCCCGCC AGTTGTTGTG
4441 CCACGCGGTT GGGAATGTAA TTCAGCTCCG CCATCGCCGC TTCCACTTTT TCCCGCGTTT
4501 TCGCAGAAAC GTGGCTGGCC TGGTTCACCA CGCGGGAAAC GGTCTGATAA GAGACACCGG
4561 CATACTCTGC GACATCGTAT AACGTTACTG GTTTCACATT CACCACCCTG AATTGACTCT
4621 CTTCCGGGCG CTATCATGCC ATACCGCGAA AGGTTTTGCG CCATTCGATG GTGTCCGGGA
4681 TCTCGACGCT CTCCCTTATG CGACTCCTGC ATTAGGAAGC AGCCCAGTAG TAGGTTGAGG
4741 CCGTTGAGCA CCGCCGCCGC AAGGAATGGT GCATGCAAGG AGATGGCGCC CAACAGTCCC
4801 CCGGCCACGG GGCCTGCCAC CATACCCACG CCGAAACAAG CGCTCATGAG CCCGAAGTGG
4861 CGAGCCCGAT CTTCCCCATC GGTGATGTCG GCGATATAGG CGCCAGCAAC CGCACCTGTG
4921 GCGCCGGTGA TGCCGGCCAC GATGCGTCCG GCGTAGAGGA TCGAGATCGA TCTCGATCCC
4981 GCGAAATTAA TACGACTCAC TATAGGGGAA TTGTGAGCGG ATAACAATTC CCCTCTAGAA
5041 ATAATTTTGT TTAACTTTAA GAAGGAGATA TACATATGGC gACgAAGGCC GTGTGCGTGC
5101 TGAAGGGCGA CGGCCCAGTG CAGGGCATCA TCAATTTCGA GCAGAAGGAA AGTAATGGAC
5161 CAGTGAAGGT GTGGGGAAGC ATTAAAGGAC TGACTGAAGG CCTGCATGGA TTCCATGTTC
5221 ATGAGTTTGG AGATAATACG GCAGGCTGTA CCAGTGCAGG TCCTCACTTT AATCCTCTAT
5281 CCAGAAAACA CGGTGGGCCA AAGGATGAAG AGAGGCATGT TGGAGACTTG GGCAATGTGA
5341 CTGCTGACAA AGATGGTGTG GCCGATGTGT CTATTGAAGA TTCTGTGATC TCACTCTCAG
5401 GAGACCATTG CATCATTGGC CGCACACTGG TGGTCCATGA AAAAGCAGAT GACTTGGGCA
5461 AAGGTGGAAA TGAAGAAAGT ACAAAGACAG GAAACGCTGG AAGTCGTTTG GCTTGTGGTG
5521 TAATTGGGAT CGCCCAATAG TAAGGATCCG AATTCGAGCT CCGTCGACCT AGCATAACCC
5581 CGCGGGGCCT CTTCGGGGGT CTCGCGGGGT TTTTTGCTGA AAGAAGCTTC AAATAAAACG
5641 AAAGGCTCAG TCGAAAGACT GGGCCTTTCG TTTTATCTGT TGTTTGTCGC TGCGGCCGCA
5701 CTCGAGCACC ACCACCACCA CCACTGAGAT CCGGCTGCTA ACAAAGCCCG AAAGGAAGCT
5761 GAGTTGGCTG CTGCCACCGC TGAGCAATAA CTAGCATAAC CCCTTGGGGC CTCTAAACGG
5821 GTCTTGAGGG GTTTTTTGCT GAAAGGAGGA ACTATATCCG GAT

Fig. 8.1. ptZENIT sequence

Lebenslauf

Persönliche Daten

Name:	Wittwer Alexander
Adresse:	Baumgasse 16/13; 1030 Wien
Kontakt:	0650/5333083; witti20@gmx.at
Geboren am:	29. September 1981
Geburtsort:	Steyr (OÖ)
Staatsbürgerschaft:	Österreich
Familienstand:	Lebensgemeinschaft, erwarten unser erstes Kind

Ausbildung

Voraussichtlich 03/2010	Abschluss des Studiums mit Magistra der Naturwissenschaften (Mag.rer.nat.)
Seit 08/2008	Diplomarbeit an der Universität für Bodenkultur Wien (BOKU Wien) am Institut für angewandte Mikrobiologie (IAM) in der Arbeitsgruppe mikrobielle Fermentation (Prof. Karl Bayer) Diplomarbeitsthema: A novel transcriptional termination signal capable to terminate T7 polymerase mediated transcription more efficiently
Seit 04/2005	Diplomstudium Genetik – Mikrobiologie an der Universität Wien Studienschwerpunkte: Molekulare Mikrobiologie sowie Gen- und Biotechnologie
10/2002 - 04/2005	Diplomstudium Biologie (1.Diplomprüfung)
10/2001 - 01/2002	Studium für ein Semester an der Fachhochschule St. Pölten (Berufsfindungsphase) Studienzweig: Simulationsgestützte Nachrichtentechnik
09/2000 - 05/2001	Wehrdienst beim Pionierbataillon 3 in Melk (NÖ)
09/1992 - 07/2000	Bundesrealgymnasium Kirchdorf an der Krems (OÖ) mit abgeschlossener Reifeprüfung
09/1988 - 07/1992	Volksschule Pfarrkirchen bei Bad Hall
09/1987 - 07/1988	Vorschule Pfarrkirchen bei Bad Hall

Berufliche Praxis

03/2009 - 12/2009	Anstellung als Techniker an der BOKU Wien in der

Arbeitsgruppe von Prof. Karl Bayer
– Vollendung der Diplomarbeit bis zur Patentreife und Implementierung in einem Fermentationsprozess
– Einschulung von Bachelor-und MasterstudentInnen in das molekularbiologische Arbeiten
– Versuchsplanung und teilweise Umsetzung der Integration des T7-Polymerase Genes in das Genom eines Produktionsstammes

Sprachen

Deutsch – Muttersprache
Englisch – In Wort und Schrift
Portugiesisch – Grundkenntnisse

EDV

Microsoft Office Paket – sehr gute Kenntnisse
Open Office Paket – sehr gute Kenntnisse
Diverse Grafikprogramme (Photoshop, GIMP, etc.) - gute Kenntnisse
Diverse Anwender-Softwarepakete (CLC workbench, DNA Star, DNA Sequencer, etc.) bzw. web tools für molekularbiologische *in silco* Arbeiten