universität wien

DIPLOMARBEIT

Plasmid free T7 based *E.coli* expression system: tandem gene integration

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag.rer.nat)

Verfasserin: Marianne Schimon
Matrikel-Nummer: 0306833
Studienrichtung/ Studienzweig: A490 Molekulare Biologie
Betreuer/Betreuerin: Prof. Karl Bayer (BOKU Wien), Prof. Angela Witte

Wien, am 3.12.2008

Danksagung

Mein herzlicher Dank gilt der Gruppe Proteinfermentation des Institutes für Angewandte Mikrobiologie der BOKU Wien für ihre freundschaftliche Arbeitsatmosphäre, meinen Betreuern Sabine Nemecek und Gerald Striedner für ihre fachliche Kompetenz und ihre hilfreichen Ratschläge und Herrn Prof. Karl Bayer für seinen Einsatz bei der Korrektur der Arbeit und der Analyse der Ergebnisse.
Mit dem Abschluss meines Studiums möchte ich mich aber noch bei den Menschen bedanken, die mich während meiner Studienzeit immer aufgebaut und vorangetrieben haben: meiner Familie und meinem Mann. Liebe Eltern ich danke euch für eure finanzielle und seelische Unterstützung, ohne die ich mein Studium nicht so schnell absolvieren hätte können. Lieber Rene, du bist die Inspiration meines Lebens, ich danke dir für die Kraft, die du mir täglich schenkst.

Contents

List of Tables

1 Introduction

1.1 Recombinant protein production

Recombinant protein production plays a central role in modern biotechnology. Every recombinant protein needs an optimized production process, which differs depending on the required features of the protein of interest and is affected by protein-specific properties. Research and development are mainly focused on the analysis of function and structure of the recombinant protein, whereas industrial protein production has to focus additionally on a cost-efficient process. Important parameters for protein production processes are therefore the growth rate of the host, the complexity and costs of the nutrition media, the strength of the gene expression system and the correct folding of the target protein. Correct protein folding is very important, because many eukaryotic proteins possess posttranslational modifications such as glycosylation, phosphorylation, acetylation and carboxylation, which influence the structure and function of the protein. In order to ensure the correct folding and therefore functionality of the protein, different expression hosts exist, ranging from prokaryotes to eukaryotes.

Prokaryotic production systems are characterized by a rapid growth on inexpensive media with frequently strong gene expression rates. Limited abilities of prokaryotes to modify proteins posttranslationally and folding problems of complex eukaryotic proteins reduce the applicability of prokaryotic expression hosts for production of certain recombinant proteins.

Yeast production systems using the eukaryotic host organism *Saccharomyces cerevisiae* and *Pichia pastoris* combine advantages of prokaryotic hosts, such as rapid growth and high production yields with characteristics of typical eukaryotes. Although glycosylation and modification systems exist in yeast for posttranslational modifications of recombinant proteins, certain limitations for correct modification of higher eukaryotic proteins can be encountered.

The desire of recent years for production of complex proteins, in particular antibodies and glycoproteins, has driven the development of insect or mammalian cell production, as an alternative to primitive eukaryotic protein production systems. The difficult handling, high costs, intensive production time, low growth rate, complex media and often low product yields characterize such higher eukaryotic production systems. In spite of the various different expression systems, the influence of the recombinant protein on the host strain can be hardly predicted, which makes an optimization for every single recombinant protein inevitable [12].

In reference to the production system used in this diploma thesis, the following introductional parts will focus on prokaryotic expression systems, mainly the production host *Escherichia coli*, for pharmaceutical applications.

1.2 *E. coli* as a cell factory

In this diploma thesis *E. coli* hosts were used for the integration of the recombinant genes in the chromosome. The gram negative bacterium *Escherichia coli* was first identified in the year 1885 by a German pediatrician namely Theodor Escherich. The originally named microorganism Bacterium coli commune, received the name *Escherichia coli*. The bacterium is widely distributed in the intestine tracts of warm-blooded animals and therefore also of humans and lives predominantly facultative anaerobe. It belongs to the family of Enterobacteriacea, beside other well-known pathogens such as Salmonella, Shigella and Yersinia [21].
The use of *E. coli* as laboratory strain started in the beginning of the 20th century. *E. coli* strain K-12 was isolated from the stool of a convalescent diphteria patient in the year 1922 and could be used as a laboratory strain in Stanford university, because of a significant stability under laboratory conditions [13]. After elimination of the pathogenic genes, *E. coli* could be used for laboratory experiments. In the year 1944 Gray and Tatum reported the isolation of X-ray-induced auxotrophic mutants of bacteria, including *E. coli* K-12 strain [24]. Studies on genetic recombination followed and since that time thousands of mutants of *E. coli* K-12 strain have been produced. *E. coli* emerged to become an important model organism for genetic studies [26]. The genome of *E. coli* was fully sequenced in the year 1997 and most of the 4288 genes are functionally characterized. The isolation of a gene was first realized in 1969 by Jonathan Beckwin, who isolated *E. coli* gene for the enzyme β-Galaktosidase. Driven by this achievement the genetic studies on *E. coli* as laboratory strain continued successfully [12].

Moreover *E. coli* developed to be a widely used organism for recombinant gene expression, because of the appearance of comprehensive knowledge about *E. coli* biochemistry and genetics over the last century. The production of recombinant proteins in microorganisms, is beside the production of microbial enzymes and the acquisition of microbial metabolites from industrial fermentation processes, an important field of the pharmaceutical and biotechnological industry.
In 1982 the first recombinant expressed protein for pharmaceutical purposes was approved. Eli Lilly's rDNA human insulin opened the market for pharmaceuticals produced by genetic manipulated organisms (GMO's). The low selling price and the large volume of production were feasible, because of a cost-efficient, highly effective and scalable process generated by the use of *E. coli* as host. For the most pharmaceutical purposes, the dominant role of *E. coli* as a production host did not change. The missing ability of *E. coli* to fold complex proteins in their proper conformation or to attach mammalian glycosylation chains is for many industrial products and applications not required. The large knowledge of *E. coli's* genetics and it's genetic flexibility permit continuous development, even though the field of in vivo production is an established technology [38].

1.3 Recombinant protein production of pharmaproteins

Recombinant proteins with high molecular masses have to be produced in pharmaceutical protein production processes. Proteins produced for pharma industry range from enzymes, such as proteases, hormones, such as insulin and erythropoetin (EPO), cytokines, such as interferons, to antibodies, inhibitors and vaccines. Because of their structural complexity in contrast to low molecular pharmaceutical substances, it is not possible to synthesize them chemically in a cost and work-efficient way.
The traditional way to overcome this problem was the isolation of pharmaproteins out of natural sources, such as mammalian or human tissues or human plasma. But the amount of such proteins is normally very low in the natural environment and therefore isolation is connected to a great effort. Furthermore the possible contamination of the proteins with pathogens harbours a great risk factor.
Production of pharmaproteins in genetically manipulated cells offered a way to receive high amounts of pathogen-free recombinant protein. The decision, which organism to use for successful production of the pharmaprotein, depends on the complexity of the protein. The functionality and the absolute human identical posttranslational modifications have to be verified, which makes high standards of quality control necessary for each production process [12].
As *Escherichia coli* is one of the main production hosts and was used in the subsequent work, the following sections focus on *E. coli* expression systems.

1.4 Recombinant protein expression systems

To design an efficient recombinant expression system it is necessary to keep in mind a number of central elements.

1.4.1 *E. coli* host strains

The host strain or genetic background for recombinant gene expression is highly important. It is necessary that host strains are deficient in the most harmful natural proteases, to avoid recombinant protein degradation. Genetic elements, which are relevant for the expression system have to exist. For different applications advantageous strains are available [34]. In this work BL21 and HMS174 host strains were used for the integration of the target cassette.

1.4.2 Transcriptional promoters

To produce sufficient levels of recombinant protein, a strong transcriptional promoter is necessary. Basal transcription in the absence of an inducer is minimized by the presence of a suitable repressor. The promoter induction can occur by addition of a chemical compound or by generation of a thermal difference, in case of temperature-sensitive promoters. The

most common used inducer is the sugar molecule isopropyl-beta-D-thiogalactopyranoside (IPTG) [14].
Because of the high utilization of *E. coli* lactose (lac) operon for many years, many promoters used to drive the transcription of heterologous proteins have been constructed from lac-derived regulatory elements. Beside the lac promoter there exists a close relative namely lacUV5, which is less sensitive to regulation by the cAMP-CRP (cAMP receptor protein). This reduced sensitivity can be enhanced by incorporation of 1% glucose in the cultivation medium to reduce cAMP levels. Both lac promoters are inducible by IPTG and are part for example in the most common used expression system : the T7 based expression system [34].

A commercially available form of this T7 based system is the pET system from Novagen. The pET expression system was first described by Studier and colleagues [37] in the year 1990 and since then more than 40 different pET plasmids were developed and are commercially available. The heterologous gene expression by this system requires a host strain lysogenized by a DE3 phage fragment, which encodes the T7 RNA polymerase (bacteriophage T7 gene 1), under the control of the IPTG inducible lacUV5 promoter.
Figure 1 shows the modified lac operon for the expression of T7 RNA polymerase gene 1 encoded in the DE3 phage region. T7 gene 1 is transcribed as the second gene in a bicistronic mRNA. The lac repressor (*lacI* gene product) binds to *lacO1*, and then interacts with pseudo-operators *lacO2* and *lacO3* to prevent transcription by *E. coli* RNA polymerase. The inducer IPTG binds to the repressor, reducing it's affinity for *lacO1* and thus enabling transcription to occur. When cAMP levels are sufficiently high (e.g., in the absence of glucose) the CAP/cAMP complex is formed and binds immediately upstream from the promoter to fully stimulate transcription. In the presence of glucose, CAP/cAMP is not formed and transcription is reduced. This is called the glucose effect, or catabolite repression [3].
A copy of the lacI gene is present on the *E. coli* genome and in many applications on the expression plasmid. The *lacI* gene is only weakly expressed and it is possible to receive a 10-fold enhanced repression by use of the promoter mutant $\mathrm{LacI^q}$.
The expression of T7 polymerase starts, when IPTG is added and binds to the lac promoter and triggers the release of the tetrameric lacI from the lac operator. The transcription of the target gene, which is under control of the T7/lac promoter starts subsequently after T7 polymerase expression. T7 polymerase transcribes five times faster than *E. coli* RNA polymerases, which is about 230 nucleotides per second in comparison to 50 nucleotides per second [34].
The T7 promoter is a 20 nucleotide sequence specific for the T7 polymerase and not recognized by the *E.coli* RNA polymerases. The strong expression system leads to large amounts of mRNA, accumulating the desired protein at levels of 30% of the total cell protein. Drawbacks of the system are the high amounts of mRNA which in turn cause ribosome destruction, a product formation rate, which is too high for host cell capacities and an increased plasmid replication, which might altogether even lead to cell death [14].

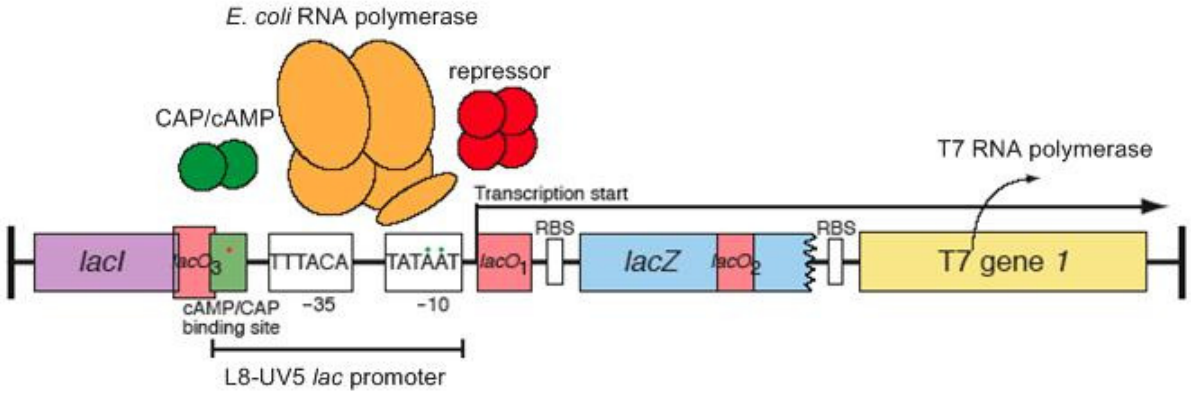

Figure 1: **DE3 phage fragment encodes for the T7 RNA polymerase gene 1 under the control of the lacUV5 promoter [3]**

A way to reduce the stress caused by the strong T7 system, is the tuning of the expression rate to a tolerable level for the cell. The increase of production yield caused by strong promoters is restricted by the short period of time through which recombinant gene expression can be maintained. By the limitation of the feed of inducer, a titration of the repressor molecule can be achieved, which leads to the establishment of a stable equilibrium between recombinant and cellular protein biosynthesis. With this extension of the production phase, the specific amount of recombinant protein can be increased [35]. Expression systems based on weak promoters are used to reduce the high metabolic stress on the host cell, caused by strong promoters as the T7 promoter. The araBAD promoter, which is inducible by the inexpensive sugar L-arabinose is an example of such a weak promoter. The regulation of the arabinose operon in *E. coli* is directed by the product of the *araC* gene. The AraC dimer can bind to three sites in the arabinose operon, O_1, O_2 and I. When arabinose is missing, the AraC dimer contacts the O_1 site in the promoter region and a DNA loop is formed, shown in Figure 2 picture 2.

Transcriptional repression of the araBAD promoter and the araC promoter occurs through AraC loop conformation forming, given in Figure 2 picture 2. Upon binding of arabinose the AraC dimer changes its conformation, binding to the O_2 site is replaced by binding to the I site at the araBAD promoter and transcription by RNA polymerase initiates, see Figure 2 picture 3. Binding of the AraC dimer to the O_1 and I sites is stimulated by cAMP receptor protein (CRP). The background expression from araBAD can be reduced by a glucose mediated catabolite repression [25]. Figure 2 picture 1 shows the linear arrangement of the regulatory gene *araC*, the operator sites and the *ara* operon, which encodes for three enzymes that are necessary to catalyze the metabolism of arabinose.

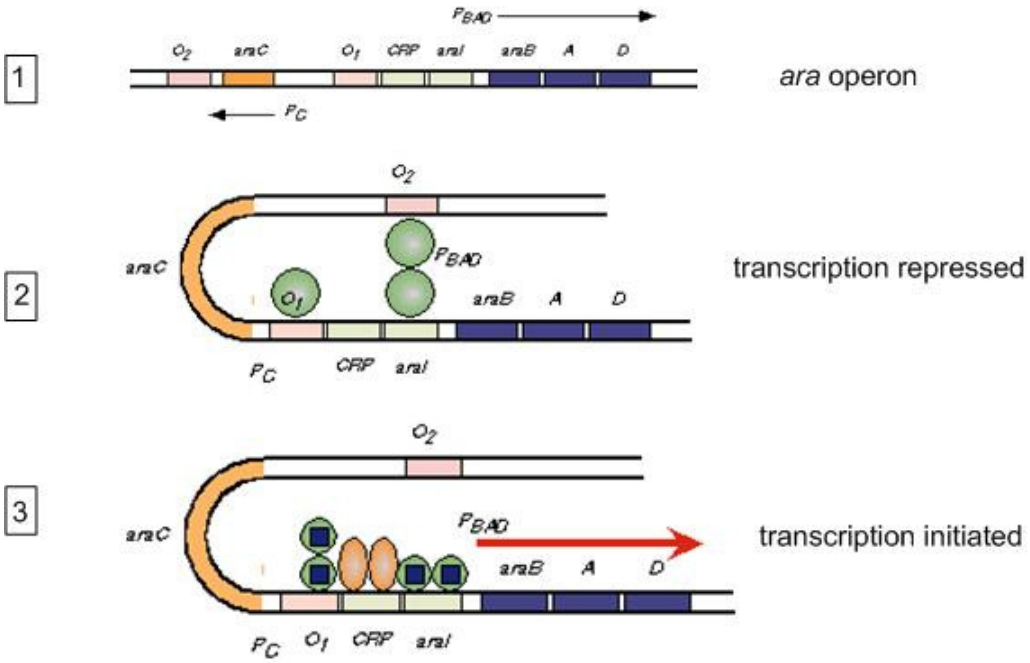

Figure 2: **araBAD promoter based expression system [4]**

1.4.3 mRNA stability

Another aspect to consider is the stability of mRNAs in *E. coli*. The levels of recombinant gene expression are mainly influenced by the efficiency of transcription, mRNA stability and the frequency of mRNA translation [14]. A way to control gene expression is the fast decay of mRNAs, which leads to rather unstable mRNAs in *E. coli*, with half-lives ranging between 30sec and 20min. The expression rate of a gene depends therefore directly on the stability of the inherent mRNA [34].

1.4.4 Translational issues

To initiate the translation of *E. coli* mRNAs from the translation initiation region (TIR) a Shine-Dalgarno (SD) sequence is necessary. The SD sequence is complementary to the 16S rRNA of the ribosome and is located 7 ± 2 nucleotides upstream from the initiation codon, which is most commonly AUG. Optimal translation initiation is obtained from mRNAs with the SD sequence 5'-UAAGGAGG-3'.
Ribosomal binding sites (RBS) include the SD sequence and the initiation codon and their secondary structure is important for translation initiation and efficiency is improved by high contents of adenine and thymine. In prokaryotes the transcription and translation are closely coupled and therefore engineering of the translation initiation region is a powerful tool for a promoter-independent gene expression modulation [14]. To prevent transcription

from irrelevant promoters on the plasmid or the genome, transcription terminators are placed downstream from the sequence encoding the target gene. Transcription termination stabilizes the mRNA by forming a stem loop at the three prime end and is preferably mediated by the stop codon UAA in *E. coli* [34].
For transcription termination T7 terminators were used in this work.

1.4.5 Inclusion body formation

The proteins used in this work are fully soluble proteins, but extensive recombinant protein production can lead to a phenomenon, called inclusion body formation.
The overproduction of heterologous proteins in the cytoplasm of *E. coli* can often lead to the aggregation of the proteins in amorph granules termed inclusion bodies. The underlying mechanism is an unbalanced equilibrium between in vivo protein aggregation and solubilization [34].
The structure of inclusion bodies and the mechanism of their formation are still far from resolved. These structurally complex aggregates often occur as a stress response, when recombinant protein is expressed at high rates. It is possible to observe inclusion body aggregates as refractile particles of up to $2\mu m^3$ by optical microscopy or as electron-dense aggregates by transmission electron microscopy [16].
To extract recombinant protein out of inclusion bodies, preparation procedures exist, which solubilize the purified aggregates in chaotropic reagents like urea and guanidium hydrochloride. To receive the recombinant protein in a native conformation from the solubilized inclusion bodies, in vitro refolding strategies are necessary. For every recombinant protein the refolding strategies have to be optimized [16].
Despite the time consuming refolding optimization for every protein [34], the production of recombinant protein in inclusion bodies has also advantages. The easier way of separation of the recombinant protein from the host cell proteins, the protection of degradation and often high yields of product protein can make production processes with inclusion body formation useful.

1.4.6 Plasmid based expression system

As this work investigates the production capacities of chromosomal based recombinant gene expression systems, it seems to be useful to overview plasmid based systems first.
Expression plasmids are designed in an order to achieve high gene dosage and therefore high production yields. Replicons, containing the origin of replication (ORI) and sometimes cis acting elements, are the ColE1 or p15A replicon. Plasmids replicate in a relaxed fashion and are present at 15-60 copies per cell ranging to a few hundred copies per cell for high copy plasmids [14].
The ColE1 replicon is derived from the pBR322 (copy number 15-20) or the pUC (copy number 500-700) family of plasmids, whereas the p15A replicon is derived from the

pACYC184 (copy number 10-12) plasmid. The advantages of these plasmids are, that they are stably replicated and maintained under selective conditions and are free of plasmid-free daughter cells [34].
The problem of plasmid loss occurs, when cells are cultivated at high densities or in continuous processes or when host cells carry very high copy number plasmids. A simple way to adress this problem is to supplement antibiotics to the growth medium and take advantage of the plasmid encoded antibiotic-resistance genes, to kill all plasmid-free cells. Drawbacks of this system are loss of selective pressure as a result of antibiotics degradation or contamination of the product or biomass by antibiotics, which may be unacceptable from a medical point of view [14].

Since the introduction of plasmids into *E. coli* induces a large number of changes, ranging from perturbation in the mechanism of DNA replication, transcription and translation, interaction with cellular membrane and alteration of carbon and energy metabolism, it is obvious that high copies of plasmids in order to produce high amounts of protein, may impose a high metabolic burden on the cells. The growth rate of the host is the most noticeable function affected, leading to an overgrowth of plasmid-free cells over plasmid-bearing cells in a mixed culture.
As soon as recombinant gene expression is induced, this negative effect on host cell metabolism is amplified. Energy consumption for maintaining a plasmid is only small, compared to the ATP amounts needed for overexpressing a gene [19]. Strong expression promoters, which are desired for high recombinant protein expression, can have adverse effects on the host cell metabolism, driving recombinant protein synthesis at the expense of host cell protein synthesis. To yield the optimum amount of recombinant protein, the expression rate has to be in relation to the synthesis capacity of the host cell [17], which is difficult for plasmid-encoded genes.

Through the excessive consumption of amino acids in protein synthesis of the recombinant product, levels of uncharged tRNAs can arise, which in turn interfere with the plasmid replication control mechanism. As a consequence the plasmid copy number (PCN) increases up to a point, where the metabolic capacity of the host is exhausted. Strong promoters as the T7 system are very susceptible for this phenomenon, even early during the production process [23].
The metabolic overload, caused by high plasmid numbers per cell, leads to a dramatic loss of cell viability and causes a significantly shortened period of product formation and a loss of overall product yield.

Because of missing alternatives plasmid based expression systems became, despite of the disadvantages, state of the art for production and isolation of recombinant proteins on a manufactoring scale.

1.4.7 Chromosomal integrated systems

Alternatives to the plasmid based recombinant gene expression are *E. coli* systems based on chromosomal integrated target genes. Although some disadvantages of plasmid systems, like plasmid instability or overload of the host metabolism by high copy numbers of plasmid, could be overcome by such chromosomal integrated systems, they are only sparsely distributed among industrial productions.
Progress to the development of alternatives for recombinant protein production was brought about by basic research on *E. coli*. Several methods for site directed chromosomal integration of DNA sequences, mainly for deletions or rearrangements of genomic DNA sequences, were established. To realize homologous recombination independently of bacterial recombinases, phage proteins, such as the λ phage proteins Redα and Redβ [29] or the rac prophage proteins RecE and RecT [39], were used.

Homologous recombination, important for repair of DNA double-stranded breaks and therefore for the maintenance of genomic integrity, is mainly initiated by the prokaryotic protein RecA in *E. coli*. By inactivation of RecA/ RecBCD pathway, alternative recombination pathways were explored. RecA/RecBCD mutants in *E. coli* can be suppressed by sbcA or sbcBC mutations. In sbcA strains, a cryptic Rac prophage operon is activated to express RecE and RecT. RecE is a 5'-3' exonuclease and RecT is a ssDNA-binding protein that promotes ssDNA annealing, strand transfer and strand invasion in vitro [30].
The genes of the λ Red system were among the earliest recombination genes to be described. It was found out that λ could recombine normally in recombination-deficient *recA*, *recB* and *recC* mutants of *E. coli*, which lead to the reasoning that λ must encode its own recombination function. Three genes were found out to be important for λ encoded recombination system. *Red*α or *exo* encodes a 5'-3' exonuclease, *red*β or β a ssDNA-binding protein with annealing and strand exchange activity and a third gene called *gam* or γ inhibits the host RecBCD recombinase.

The genes of the Red system are clustered in the P_L operon, one of two λ operons expressed early in the phages transcriptional program. The *exo* gene product, λ exonuclease processively degrades the 5'-ended strand of dsDNA, generating 3'-ended overhangs. β protein binds to ssDNA, promotes renaturation of complementary strands, and is capable of mediating strand annealing and exchange reactions in vitro. β protein forms complexes with λ exonuclease, and modulates both its nucleolytic and recombination-promoting activities. The *gam* gene product inhibits all of known RecBCD protein functions by binding [29].
Normally *E. coli* conjugational and transductional recombination proceed through the RecBCD pathway, in which the functions of *recB* and *recC* are essential. Blocking of the pathway by mutation in either gene can induce the RecE or RecF pathway by further mutation, leading to a restorage of recombination ability. λ strains lacking the Red system can engage in homologous recombination via any of these pathways, in contrast to λ

strains possessing Red system, which drive homologous recombination independent of bacterial recombination function.

Red-mediated recombination proceeds via two different mechanisms: invasion and annealing. In Figure 3 both mechanisms are shown. Recombination is initiated by dsDNA end, which is resected by β-modulated λ exonuclease, producing a 3'-overhang. If the only partner available for recombination with the dsDNA end is an unbroken circular homologous duplex, then recombination proceeds via strand invasion, and is dependent upon the bacterial RecA protein. If the partner is replicating, or has a double-stranded end at a genetically different location, then recombination can proceed by β protein-mediated annealing, which is RecA independent [31].

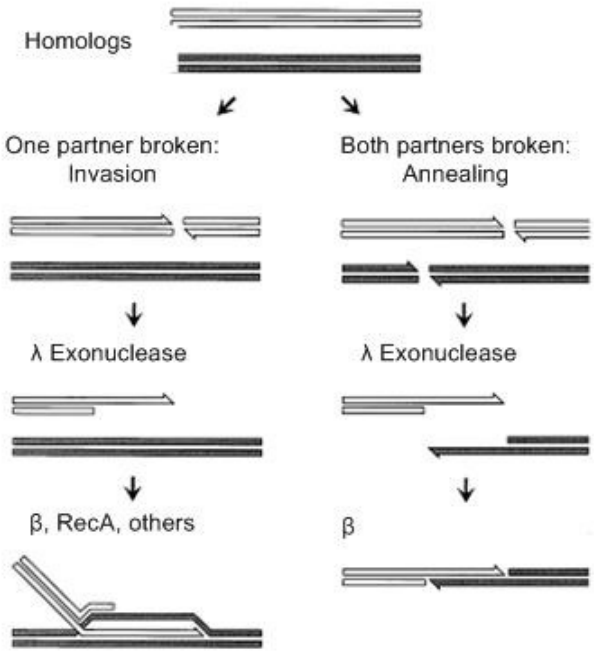

Figure 3: **Invasion and annealing mechanism of homologous recombination mediated by λ Red system [31]**

Expression of the genes *gam*, *bet* and *exo* at moderate levels under control of the *lac* promoter from a multicopy plasmid, was found to be non-toxic to *E. coli*. Recombination events between non-replicating homologs were found to be promoted at high levels. Furthermore Red system can be used for recombination of the bacterial chromosome and linear dsDNA introduced into the cell via electroporation [29]. Because of the presence of intracellular exonucleases, that degrade linear DNA in most bacteria strains, mutants lacking exonuclease V of the RecBCD recombination complex were used for integration of linear DNA into the chromosome [18].

Red system was shown to work in different bacterial strains for gene replacements, such as Salmonella or enteropathogenic and enterohemorrhagic *E. coli* [28] and was also efficiently used for DNA engineering in embryonic stem cells of mice [11].
These systems were mainly based on homologous recombination between DNA segments with identical sequences of several hundred bases. As a need for better genetic engineering, recombination of short homologous sequences was established. As demonstrated by Zhang and coworkers homologous recombination either based on the RecET or the Red system worked well for PCR-generated linear DNA species and plasmids with only 60 bases of homologous sequences. Drug resistance genes flanked by plasmid homologous sequences were generated by PCR and replaced after homologous recombination plasmid sequences, therefore generating a new plasmid [39].

Datsenko and Wanner used λ Red system for homologous recombination between PCR generated homology cassettes and the *E. coli* chromosome, sharing only 36 bases of homology. Red genes *gam*, *bet* and *exo* were placed on a low-copy plasmid under the control of the arabinose inducing promoter P_{araB}. With this approach 13 different *E. coli* disruption mutants were generated. *E. coli* genes were substituted by resistance genes flanked by FRT (FLP recognition target) sites, which were amplified with PCR primers that were homologous to regions adjacent to the gene to be inactivated. After selection of the mutants, they eliminated the resistance gene by using a helper plasmid expressing the FLP recombinase, which excised the directly repeated FRT sites flanking the resistance gene. The use of temperature-sensitive replicons on Red plasmid and FLP helper plasmid offered an easy way of curing bacterial cells from helper plasmids by a temperature-shift to 37°C [18].

The chromosomal integration procedure used in this work is based on the λ Red system and was developed based on the results achieved by Datsenko and Wanner [36]. The homologous recombination occurs between 300bp homologies on the Tn7 locus of *E. coli* host strains. The Tn7 site originates from a large transposable element with 14kb in size, which encodes resistance to the antibiotics trimethoprim, spectinomycin and spreptomycin. Unusually Tn7 transposes from episomal elements to a single site within the chromosome of *E. coli* at the high frequency of 10^{-2}/cell per generation. All transpositions of Tn7 to the *E. coli* chromosome which have been characterized, took place at a single site and in an unique orientation near minute 82 of the chromosome. As more detailed analysis show, lies the Tn7 unique site between the genes glmS and phoS, more precisely in the glmS transcriptional terminator [22].

1.5 Tandem gene genome integration

The integration of two identical product genes into the host chromosome, which was done in this work, was accompagnied by uncertainties based on the appearance of tandem

repeat instabilities.
Reports dealing with tandem repeat instability give reason to assume the occurrence of deletion or duplication events between tandem repeated genes [33]. Gene deletion of repeated sequences was reported to act RecA-dependently and RecA-independently, in both cases it is a problem of replication fork blockage during DNA replication [27]. Large repeats of several hundred base pairs were observed to delete at high frequencies and deletion rate decreases exponentially as the distance between the homologies increases [15].

1.6 Bioreactor cultivations

A main part of the underlying diploma thesis was the observation and analysis of the constructed tandem gene cassettes integrated in the host strain chromosome in the course of a shake flask and a bioreactor cultivation.
Bioreactor cultivations were run in a fed-batch mode. What is most important for bioreactor cultivations are tightly controlled conditions all over the cultivation broth. Maintaining of the process conditions for an optimal growth, is ensured by the observation of parameters, such as the temperature, the pH and the oxygen concentration by on-line probes and the subsequent regulation of the parameters. The determination of the growth rate of the bacterial cells during the cultivation, as well as the product yield and the product formation rate were performed to receive information about the functionality of the integrated cassettes and the host cell metabolism during recombinant protein production.

1.6.1 Growth of microorganisms

To describe the growth characteristics of microorganisms a batch cultivation is very useful. Because of the exponential growth of microorganisms, the biomass as a function of the time is presented in a semi-logarithmic way. A batch cultivation is based on a defined volume of media, which gets inoculated by a small amount of bacteria. The microorganisms start to proliferate and to consume substrate and at the same time metabolites get enriched. When the substrate is consumed, cell growth is reduced until it is stopped at a certain time point. The growth curve can be divided in certain phases. The "lag" phase starts immediately after inoculation and is characterized by no increase in biomass. The cells use this phase to adapt to the new environment. After finishing this process the exponential cell proliferation starts, which is presented in the semi-logarithmic graph as a linear ascent. In the end of this phase the ascent of biomass is reduced, because of a substrate limitation or metabolite inhibition. This is the phase of reduced growth. The last phase is the stationary phase, in which the biomass concentration remains constant. The velocity of cell growth is as high as the velocity of cell death [20].
Equation 1 shows the increase of biomass concentration (X) as a function of the time t for

the exponential phase and the phase of reduced growth. μ indicates the specific growth rate, which can be considered constant $\mu = \mu_{max}$ for the exponential phase.

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

The integration of equation 1 with the boundary conditions $X(t = 0) = X_0$ results in equation 2.

$$X = X_0 \cdot e^{\mu_{max} t} \tag{2}$$

The doubling time t_D for the exponential phase is defined as:

$$t_D = \frac{ln2}{\mu_{max}} \tag{3}$$

1.6.2 Batch and fed-batch

In batch cultivation substrates or nutrients are not added during the bacterial growth. The available substrates are given in the batch medium and substrate turnover results in biomass and product. The process is ended when cells stop to grow, which can be because of missing substrates or accumulated metabolites.
When the batch is used up, continuous feeding of medium starts. The addition of the feed can be followed by different strategies, for example by exponential algorithm to provide a constant growth rate [12].

1.6.3 Product formation

Knowledge of the kinetics of product formation is essential for process operation. On the one hand for microbial processes on the industrial scale, product yields are more important to know than the amount of biomass and on the other hand it is essential to know formation rates of inhibiting metabolites.
A correlation between the product formation rate and the growth kinetic is presented in equation 4 [20], when the substrate of the batch culture is consumed.

$$\frac{dP}{dt} = (Y_{PX}\mu)X \tag{4}$$

1.6.4 Product yield and turnover coefficients

Product yield and turnover coefficients define a ratio between the consumption of substrates and the production of biomass and other products, without knowing the exact cell metabolism. Table 1 shows the most important product yield and turnover coefficients [12].

Symbol	Definition
Y_{XS}	produced biomass in reference to consumed substrate
Y_{PS}	produced product in reference to consumed substrate
Y_{PX}	produced product in reference to produced biomass
Y_{XO}	produced product in reference to consumed oxygen
RQ	produced carbon dioxide in reference to consumed oxygen

Table 1: Typical product yield and turnover coefficients [12]

2 Aim of the work

The aim of the work is to determine the limits of recombinant protein expression in chromosomal encoding *E. coli* systems. Systems with an integrated single copy of the recombinant gene have already shown high productivity and stability, but the cellular capacities were not fully exploited.
In order to investigate the expression limits of chromosomal integrated hosts, gene constructs containing two copies of the product gene will be generated. GFP and SOD, two recombinant model proteins, will be integrated into *E. coli* (BL21 and K-12) hosts using homologous recombination based protocols.
The obtained gene constructs will be tested and characterized in shake flask and bioreactor cultivations. For evaluation of tandem gene systems, the obtained results will be compared with systems with an integrated single copy and with plasmid based systems.

3 Materials and Methods

3.1 Bacterial strains

As cloning host *E. coli* strain $DH5\alpha$ was used and *E. coli* recA(-) K12 strain HMS174(DE3) or *E. coli* B-strain BL21(DE3) provided by Novagen were used as expression hosts. Both strains are lysogens of λ DE3 prophage, and carry a chromosomal copy of the T7 RNA polymerase gene under control of the lacUV5 promoter. Therefore both strains are suitable for protein expression using T7 or T7 lac promoters.

3.2 Plasmids and linear cassettes

On all expression cassettes the desired gene is under control of the combined T7/lac promoter containing a 25bp lac operator sequence for a tight control of the strong T7 expression system. Termination is provided by the T7 terminator. Chloramphenicol resistance was amplified with it's own promoter from pLysS (Novagen). Chloramphenicol resistance gene encodes a chloramphenicol acetyltransferase (CAT).
Starting plasmids for cloning procedure of linear cassettes are pBKS(-) vectors carrying the Tn7::CAT:$T7^p$GFP inserted at NotI and KpnI restriction sites in the multiple cloning site. GFPmut3.1 starting plasmids are shown in Figure 4. hSOD plasmids were constructed in the same way, starting with the same starting plasmids, solely GFPmut3.1 replaced by hSOD.

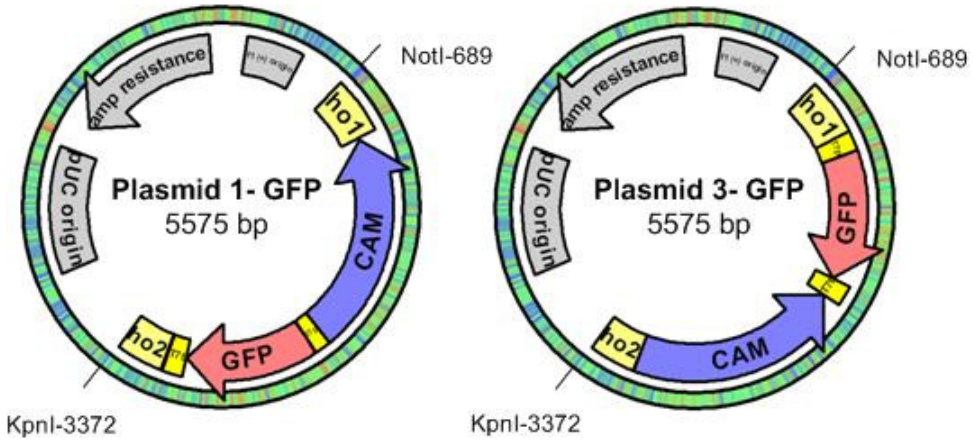

Figure 4: **Starting plasmids: pBKS(-) with linear cassettes Tn7::CAT:GFP or Tn7::GFP:CAT**

3.3 Model proteins

3.3.1 GFPmut3.1

The green fluorescent protein (GFP) is a 26,9 kDa protein, consisting of 238AA and was originally isolated from the jellyfish Aequorea victoria. It fluoresces green when it is exposed to blue light, which means that it has a major excitation peak at wavelength 395nm and a minor one at 475nm, whereas the emission peak lies at 509nm. The mutational form GFPmut3.1, which was used, is an FACS optimized variant with substitutions S65G and S72A. It has an improved folding and excitation at 488nm and is a highly soluble brightly fluorescing protein. The excitation and emission spectra of GFPmut3.1 differ from the spectra of the wild type form and fluorescence is described to be improved about 100-fold in comparison with wtGFP [32].

3.3.2 rhSOD

Human Superoxide (Cu/Zn) dismutase is a protein catalyzing the reduction of two free oxygen radicals to hydrogen peroxide. In the human body it functions as an important antioxidant defense in nearly all cells exposed to oxygen. It's molecular weight is 32kDa and it is a dimeric molecule which consists of two identical subunits (153AA each), containing a copper and a zinc atom in the active center. When it is expressed recombinantly in bacterial hosts, it is soluble and forms no inclusion bodies.

3.4 Cloning methods

The designated linear cassettes with two gene copies of GFPmut3.1 or hSOD were created by ligation of two PCR-products. All primers used were designed by the aid of SeqMan and pDraw32 and ordered by eurofins MWG GmbH. To create the tandem gene linear cassettes two PCR reactions were designed and after purification of the PCR products, both constructs were digested and ligated.
Length of the ligated constructs was verified by agarose gel electrophoresis and DNA was sequenced to exclude possible mutations. DNA sequencing samples were sent to AGOWA GmbH according to the instructions given for barcode sequencing [2].

3.4.1 PCR (Polymerase chain reaction)

PCR is the most common method to amplify specific DNA sequences. Starting point is usually a double-stranded DNA (dsDNA). The two specific primers (oligonucleotides specific to the 3' end of the target DNA segment), dNTPs and a thermostable polymerase (e.g. Novagen KOD-XL) are added in great excess. KOD XL polymerase was used to amplify long DNA sequences with a high fidelity. MBI taq polymerase (MBI Fermentas) was used

to screen colonies from agar plates or liquid cultures, as well as for Mini Preps, to determine the integration of the cassette of interest. All PCR reactions were run with Master Cycler Gradient (Eppendorf). Different PCR set-ups were performed to amplify DNA products using different templates, primers, DNA polymerases and temperature profiles. PCR set-ups for KOD-XL PCR and Taq polymerase PCR are given in appendix in tables 14, 15, 16 and 17. All primers were used (Appendix: table 13) at a concentration of 10pmol/μl; template DNA had an approximate concentration of 5ng for one reaction.

3.4.2 Restriction digestion

DNA restriction digests were used for preparing DNA fragments (PCR products) for combining or for ligation of a vector with an insert. Restriction endonucleases are enzymes, which are classified into three groups (type I, II and III) based on their subunit composition, their cleavage position, their sequence-specifity and cofactor-requirements.
Type I enzymes are complex, multisubunit and a combination of restriction and modification enzymes that cut DNA far from their recognition sequences. Because of their missing production of discrete restriction fragments or distinct gel-patterns they have little practical value. Type III enzymes are also a large combination of restriction and modification enzymes, but they cleave outside of their recognition sequence and lead rarely to complete digests and are therefore as useless for laboratory work as type I enzymes.
Type II enzymes recognize specifically 4-8bp sequences and cleave both DNA strands close to or within their recognition sequences. Because they produce discrete restriction fragments, they are the only class used in laboratory for DNA analysis and gene cloning. Many restriction enzymes cleave asymmetrically within the recognition sequence and provide "sticky ends". If the ends of the insert fragment and of the vector are compatible they can be easily ligated. Additionally restriction endonucleases can cleave DNA strands in a way to produce "blunt ends".

Used restriction enzymes and supplied 10x buffers were used from NEB and were stored at -20°C. Normally 2μl of enzyme were used for a 50μl reaction, but enzyme activities may differ and must be considered. Addition of 5μl of 10x buffer and 1μg of DNA for 50μl reaction volume completed the reaction and incubation was performed for 2h at 37°C and 300rpm on a heating block (Eppendorf). If necessary enzyme inactivation was maintained by incubation on 65° or 80°C for 20min, whereas temperature and time depended on the enzyme. Double digestions with two enzymes were performed in NEB buffer, which gives both enzymes highest activity [6].

3.4.3 Agarose gel electrophoresis

To determine the length and purity of DNA molecules gel electrophoresis methods are used. Because each nucleotide in a nucleic acid molecule carries a single negative charge, the nucleic acids move to the positive electrode in an electric field. To separate these

moving nucleic acids by their size, porous gels formed by dilute solutions of agarose, which is a polysaccharide isolated from seaweed, are used. These DNA separation methods are used for both analytical and preparative purposes. Because DNA bands are invisible on the gel, DNA has to be exposed to the dye ethidium bromide, which intercalates into nucleic acids and fluoresces under ultraviolet light [10].
For analytical gels, as well as for preparative ones, 1.5% agarose were dissolved in 1x TAE buffer (tris-acetate buffer at $pH = 8,0$ with EDTA that sequesters divalent cations) and for 120ml gel $3\mu l$ ethidium bromide were added. Analytical gels were run at 130V and 500mA and preparative gels at 90V and 500mA. As molecular marker O'GeneRuler^{TM} DNA ladder Mix (MBI fermentas), ladder sizes shown in appendix figure 21, was used.

3.4.4 DNA purification

Gel extraction and PCR purification

For extraction of appropriate DNA bands from an agarose gel or for the purification of PCR products after PCR reaction, Wizard SV Gel and PCR Clean-Up System (Promega, Cat. No: A9281) was used. With this kit DNA fragments of 100bp to 10kb can be extracted from agarose gels in Tris acetate (TAE) buffers. The purification of PCR products is commonly used for remove excess nucleotides and primers, to make PCR products ready for further use. The Clean-Up system is based on the ability of DNA to bind to silica membranes in the presence of chaotropic salts. The excised band after electrophoresis or the PCR product are dissolved in the presence of guanidine isothiocyanate. The isolation of DNA occurs by microcentrifugation to force the dissolved gel slice or PCR reaction through the membrane while simultaneously binding the DNA on the surface of the silica. After washing the DNA fragment or PCR product, the DNA is eluted in water [9].

Plasmid preparation

To isolate and purify plasmid DNA out of bacterial overnight cultures, the Wizard Plus SV Minipreps DNA purification System (Promega, Cat. No: A1330) was used. The system can isolate plasmids less than 20000bp in size from *Escherichia coli* hosts and delivers purified plasmids ready for further manipulation. Isolation of plasmids was performed from 1-10ml overnight cultures, cells were centrifuged at 4000rpm for 5min and the supernatant was discarded. Cells were resuspended in $250\mu l$ cell resuspension solution and mixed by invertion, not by vortexing to prevent shearing of chromosomal DNA. Afterwards $250\mu l$ cell lysis solution were added, inverted several times and incubated until the cell suspension cleared. In the next step $10\mu l$ alkaline protease solution was added and incubated for 5min after inverting several times. Alkaline protease inactivates endonucleases and other proteins released during the lysis of bacterial cells that can affect the quality of the isolated DNA. The reaction is stopped by adding $350\mu l$ of neutralization solution. The bacterial lysate is centrifuged for 10min at 14000rpm. Plasmid DNA purification occurs by binding of DNA

to silica membranes during microcentrifugation and washing several times in 95% ethanol until plasmid DNA is eluted in water [8].

3.4.5 Quantification of DNA

To determine DNA concentrations of PCR products, Mini Preps or gel extracted DNA, the NanoDrop ND-1000 Spectrophotometer was used. The NanoDrop ND-1000 is a full-spectrum (220-750nm) spectrophotometer that can measure $2\mu l$ samples of DNA. By surface tension the sample is hold in place and does not need to be placed in a cuvette. It is possible to measure nucleic acid samples up to 3700ng/μl (dsDNA) without dilution [5].

3.4.6 Ligation

For ligation of restriction enzyme digested PCR products or vectors and inserts T4 DNA ligase was used (NEB, Cat.No:M0202S). T4 DNA ligase catalyzes the formation of a phosphodiester bond between juxtaposed 5' phosphate and 3' hydroxyl termini in duplex DNA or RNA. The enzyme joins blunt end and cohesive end termini as well as repairs single stranded nicks in duplex DNA, RNA or DNA/RNA hybrids. Ligations were performed overnight at 16°C by adding $1\mu l$ T4 DNA ligase, $2\mu l$ T4 ligase buffer, the appropriate amounts of PCR products or vector and insert DNA and H_2O to $20\mu l$ reaction volume. For inactivation of T4 ligase incubation 10min at 65°C can be performed [7].

3.4.7 Isopropanol precipitation of DNA

To remove working buffers, e.g. after ligation, or for higher concentration of DNA isopropanol precipitation can be used. The precipitation is mediated by high concentrations of salt (sodium acetate) and the addition of isopropanol. 1/10 volume 3M sodium acetate and 0.6 fold volume of isopropanol were added, mixed and incubated 30min at -80°C. Precipitated DNA was centrifuged for 30min at 4°C at 13000rpm in a minicentrifuge and the supernatant was discarded completely. DNA was washed with 1ml 70% EtOH and centrifuged for another 5min at 4°C. Supernatant was discarded and the DNA pellet completely dried at room temperature (RT). DNA was dissolved in 10-20μl RO-H_2O and stored at -20°C for further use.

3.5 Transformation of DNA into bacterial cells

Transformation describes the genetic alteration of a cell resulting from the uptake and expression of foreign genetic material (DNA). To overcome the problem that only 1% of all bacterial species are naturally capable of taking up DNA, bacterial cells have to be treated to become competent. To make cell walls permeable for DNA electroporation can be performed. By application of an external field the electrical conductivity and permeability

of the cell wall is significantly increased, so that DNA can be taken up. To avoid the phenomenon of arcing, the salt content of the DNA solution and the bacterial suspension should be reduced to a minimum.

3.5.1 Preparation of electrocompetent cells

Cells to become electrocompetent were inoculated in 10ml LB (medium composition given in 7.1.3) with appropriate antibiotics and incubated overnight. On the next day cells were diluted 1:100 in fresh LB medium with antibiotics and cells were incubated until they reach an $OD_{600} = 0.6$. Cells were placed on ice and 50ml cell suspension were centrifuged for 5min at 4000rpm and 4°C (Beckman Coulter Avanti J-25 Centrifuge, rotor JA-10.500). The supernatant was discarded and the cell pellet resuspended in 1ml ice-cooled RO-H_2O. The cell suspension was transferred into 1,5ml Eppendorf tubes and centrifuged for 2min at 13000rpm at 4°C in a minicentrifuge. Supernatant was discarded and cells were washed with H_2O another three times in the same way as already described. In the final step $80\mu l$ ice-cooled H_2O was placed on the bacterial cell pellet and cells were left on ice until electroporation.

3.5.2 Electroporation

20-600ng of DNA were put to the $80\mu l$ ice-placed electrocompetent bacterial cells. Bacterial cells and DNA were put into an electroporation cuvette and put into the electroporation apparatus (Bio-Rad gene pulser), which was set to deliver an electrical pulse of $25\mu F$ capacitance, 2.5kV and 1000Ohm resistance. After electroporation bacteria were put into $900\mu l$ SOC medium (medium ingredients given in 7.3.1) and incubated for 1h at 37°C and 300rpm. $200\mu l$ and $800\mu l$ of the bacterial suspension were plated on LB agar plates containing the appropriate antibiotics. After drying the plates were incubated overnight on 37°C.

3.6 Genome integration methods

3.6.1 λ-Red recombination system

Chromosomal integration

PCR generated linear integration cassettes were checked by sequencing before they were used for genome integration. One colony of MG1655 bacteria was picked and inoculated in LB + Amp + 0.2% glucose and were grown overnight at 37°C. Bacteria were inoculated 1:100 in LB + 0.2% arabinose and in LB without arabinose as negative control on the next day and grown on 30°C until they reached an OD_{600} of 0.4 to 0.6. Bacteria were aliquoted in 50ml Falcons and centrifuged at 4°C and 10000rpm for 5min. Supernatant was discarded and cell pellet washed four times with 1ml ice-cooled H_2O. After the last washing step, the

cell pellet was resuspended in 80μl ice-cooled H_2O and linear DNA constructs were added at a concentration of 200, 400 and 600ng. Electroporation was performed as described in 3.5.2.
After 1h incubation at 37°C 500μl of the bacterial suspension was plated on LB agar plates and chloramphenicol and the other 500μl were incubated overnight at 37°C and 300rpm and plated on the next day. Positive colonies were screened by Taq-screening PCR, see for more detailed information subsection 3.4.1 and appendix tables 16 and 17, and sequenced to verify the right integration.
The genome integrated clones with the right sequence were cured from the pKD46 plasmid by growing colonies at 42°C and streaked out on LB-Amp plates to screen for ampicillin sensitive clones. If clones were still ampicillin resistant, clones were grown again on 42°C overnight until they became ampicillin sensitive.

Phage lysate preparation

MG1655 donor strain with the genome integrated cassette was grown overnight on LB containing 5mM $CaCl_2$ and then diluted 1:100 into 5ml LB medium containing 5mM $CaCl_2$. 0, 50, 100 and 200μl of a high titer P1 lysate was added to the culture and shaked at 37°C until lysis occured. Cultures have to be checked for lysis hourly. Then 100μl chloroform was added, mixed and cell debris were removed by centrifugation at 3000rpm for 10min. The supernatant (phage lysate) was stored at 4°C in a light safe location.

P1 vir transduction

300μl of an overnight culture of the acceptor strain, either rec(+) *E. coli* strain BL21(DE3) or the *E. coli* strain HMS174(DE3) containing a temperature sensitive helper plasmid (pTSA19recA), which was incubated in 5ml LB medium with 5mM $CaCl_2$, was mixed with 100μl phage lysate. Ratio of cells to phage particles should be 1:1 to 1:10. Controls are phage only and acceptor strain only reactions, to see if the phage lysate is contaminated with the donor strain.
After a static incubation of exactly 25min on 30°C (for HMS174(DE3) with pTSA19recA) or on 37°C (BL21(DE3)), 5ml LB medium containing 10mM NaCitrate were added. Cells were centrifuged at 3000rpm for 15min and the pellet was resuspended in 1ml LB medium with 10mM NaCitrate again. For allowing expression of antibiotic resistance genes, cells were shaken 1h at 30°C (37°C) and then spread on selective media. Curation of HMS174(DE3) strain from pTSA19recA helper plasmid was performed by incubation on selective plates at 42°C and screening for ampicillin sensitive clones. Positive clones were again checked by Taq screening PCR and sequencing.

3.6.2 Quick & Easy *E. coli* Gene Deletion Kit

HMS174(DE3) and BL21(DE3) *E. coli* strains were transformed with pRedET plasmid (with ampicillin or tetracycline resistance gene) according to the description given in 3.5.1 and 3.5.2. Figure 5 shows the vector map of pRedET plasmid.

After transformation cells were incubated for 70min at 30°C and 300rpm, with regard

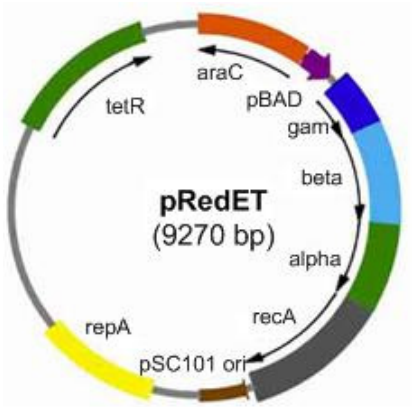

Figure 5: **pRedET plasmid of Quick & Easy *E. coli* Gene Deletion Kit [1]**

to the temperature sensitivity of the pRedET plasmid. Incubation overnight at 30°C followed on LB agar plates with ampicillin or tetracycline, depending on the resistance of the transformed plasmid. The received colonies were restroken on additional LB agar plates (+ antibiotics) and frozen in cryostocks at -80°C.
For genome integration a colony from the HMS174(DE3) pRedET-Amp or pRedET-Tet or BL21(DE3) pRedET-Amp or pRedET-Tet was picked and inoculated in LB + antibiotics at 30°C overnight. On the next day bacteria suspension was diluted 1:100 in LB medium and antibiotics and incubated until the cells reached an $OD_{600} = 0.3$. 20% L-arabinose solution was added to a final concentration of 0.4% L-arabinose. 5ml bacteria suspension were left without addition of L-arabinose for negative control of the experiment. Cells were incubated for 1h at 37°C to induce transcription of recombinational genes *bet*, *exo* and *gam*. 5ml aliquots of the bacteria suspension were centrifuged at 4°C and 10000rpm for 5min and cells were made competent as described in 3.5.1 and 3.5.2.

Figure 6 shows the mode of action of the rac prophage genes RecE and RecT, respectively the λ Red genes Redα and Redβ. 5'-3' exonculease RecE or Redα start to degrade the double-strand linear, leaving behind a single-stranded DNA. RecT or Redβ are single-strand binding proteins and bind immediately to the produced single strand. After joint molecule formation, DNA replication can occur.

Incubation after transformation was for 3h at 37°C and 300rpm and afterwards cells were plated on LB agar plates with chloramphenicol. Positive colonies occurred on the next day

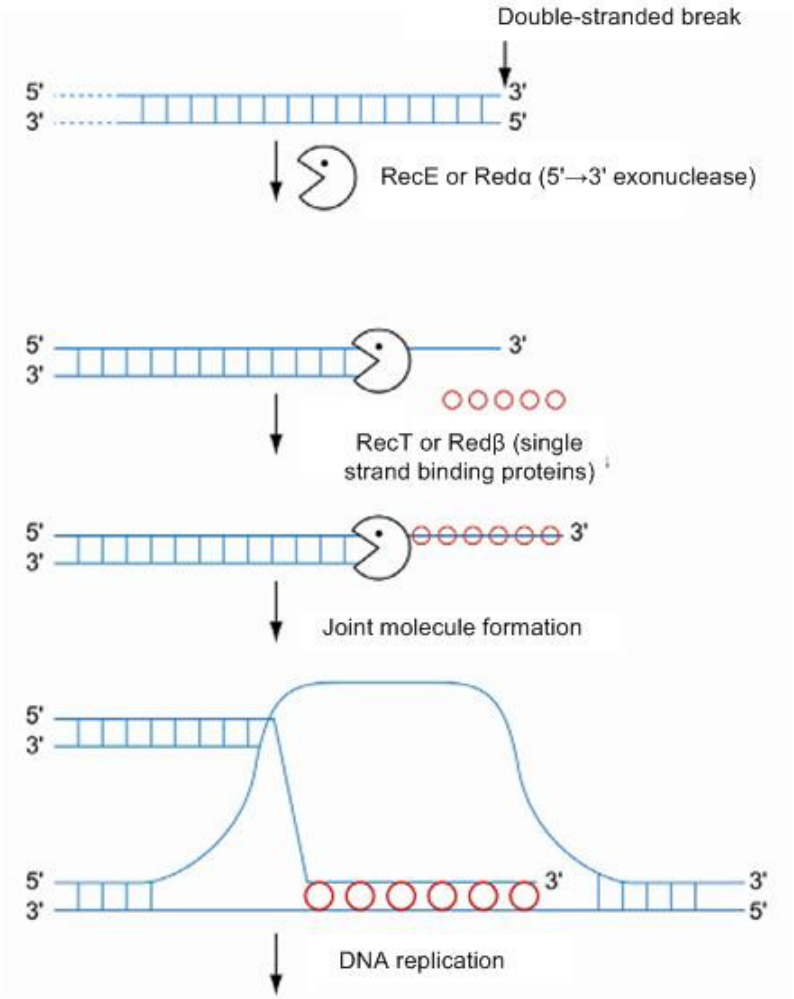

Figure 6: **Function of RecE/ Redα and RecT/ Redβ in homologous recombination [1]**

and were screened by Taq screening PCR (see subsection 3.4.1 and appendix tables 16 and 17). Positive colonies were sequenced and restroken on LB agar plates with chlorampenicol. Clones with accurate sequences were used for further experiments.

3.7 Preparation of cell banks

3.7.1 Research cell bank (RCB)

A single colony of a new strain was picked from a selective agar plate and cultivated in 3ml selective LB-medium. At $OD_{600} = 3$ one volume of this culture was well mixed with one volume sterile 87% glycerol and transferred into cryo-tubes for storage at -80°C.

3.7.2 Master cell bank (MCB)

The corresponding *E. coli* culture was cultivated an a solid selective medium. A single colony was picked and aseptically transferred to a flask filled with 40ml M9-ZB medium (see appenix table 3) with appropriate antibiotics. After overnight incubation, 80ml M9-ZB with appropriate antibiotics were inoculated with overnight culture at an $OD_{600} = 0.150$ and after 30min of incubation at 37°C and 180rpm, 40 ml of the cell culture were transferred to a new sterile 250ml flask (=reference control). A third 250ml shake flask which had been filled with 40ml sterile medium without antibiotics was incubated as a sterile control. OD_{600} of samples taken from the reference flask was measured during the cultivation. At $OD_{600} = 3$ 1ml of the WCB was taken for Koch testing. The remaining cell suspension was mixed with 1 volume of 87% glycerol and filled into cryo-vials for storage at -80°C. Koch plate testing was done as described in 3.9.2.

3.8 Cultivation in shake flasks

Shake flask experiments were started with an overnight culture (bacteria grown in 40ml semisynthetic medium) to inoculate 400ml semisynthetic medium (see for detail composition appendix table 4) in 2l shake flasks to start cell cultivation at $OD_{600} = 0.150$ at 37°C and 180rpm. At $OD_{600} = 2$ recombinant protein expression was induced by 1mM IPTG (IPTG stock solution: 0.3g in 3ml RO-H_2O, sterile filtered). Samples were taken shortly before induction, 30min, 60min, 120min, 180min and overnight after induction. For the determination of recombinant protein concentration (rhSOD, GFPmut3.1), 1ml sample was taken and centrifuged for 10min at 4°C at 13000rpm (Eppendorf Centrifuge 5414R). The supernatant was discarded and the pellet was dried and frozen at -20°C. OD_{600} and off-line fluorescence (for GFPmut3.1) were measured. From sample 180min post induction a Koch was performed to determine cells carrying the genome integrated cassette with the resistance gene.

3.9 Fed-batch cultivation in bioreactors

To ensure defined cultivation conditions and to minimize build-up of metabolic by-products (e.g. acetate) during the process, fed-batch cultivations were performed at a constant growth rate of 0.1 h^{-1}. Cells were grown in a 20l computer-controlled bioreactor (MBR; Wetzikon, CH) equipped with standard control units. For inoculation of the process, a deep-frozen (-80°C) working cell bank vial was thawed and 1ml ($OD_{600} = 1$) was transferred aseptically to the bioreactor. Batch and feed-media are given in appendix table 5.

3.9.1 Cultivation conditions

Temperature

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

pH

Calibration was performed with commercially available buffer solutions. pH was maintained at a set-point of 7.0 ± 0.05 by addition of 25% ammonia solution (ACROS Organics).

pO_2

Calibration of the Clark probe was performed after sterilisation at 800rpm and 37°C using N_2-gas or an O_2-simulator (Mettler Toledo) for 0% saturation and compressed air for 100% saturation. The dissolved oxygen level was maintained above 30% saturation by stirrer speed (800 to 1200rpm) and aeration rate control.

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 5 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 provided with an adequate amount of substrate. Feeding was initiated at the end of the batch when bacterial dry matter reached a value of 30g in 4l.

$$X_1 = X_0 \cdot e^{\mu_{max} t} \tag{5}$$

Equation 6 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.

$$S_0 = \frac{X_0}{Y_{X/S} \cdot c_S} \tag{6}$$

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 7.

$$\Delta S = S_0 \cdot (e^{\mu \cdot \Delta t} - 1) \tag{7}$$

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 8.

$$S_{target} = S_{start} - \Delta S \tag{8}$$

A PID control algorithm adjusted the pump speed accordingly this set point curve.

Addition of antifoam

To avoid foaming 1ml antifoam solution (Glanapon 2000, Bussetti, Vienna) was added per kg feed media .

Induction of recombinant protein production

Induction was performed using a conventional single pulse of IPTG directly into the bioreactor. To obtain a fully induced system, the supplied amount of IPTG was calculated to yield a concentration of 1μmol IPTG at the end of the process.

Cultivation media

A $KH_2PO_4/K_2HPO_4 * 3H_2O$ buffer provided the required buffer capacity and served also as P and K source. The other components were added to the media in relation to the gram bacterial dry matter to be produced, for detailed information see table 5 in the appendix. Yeast extract was added to the batch medium to accelerate initial growth. Feed medium was prepared according to the amount of biological dry matter to be produced during the feeding phase. Cu/Zn solution had to be added to provide intact rhSOD formation.

3.9.2 Off-line analysis

For off-line analysis 15-20ml of culture broth were taken aseptic out of the bioreactor and used for following offline analysis:

Optical density (OD_{600})

Optical density was measured at wavelength $\lambda = 600$ with a spectrophotometer (Amersham Biosciences Ultrospec 500 pro, Pharmacia Biotech Ultrospec 1000E). Dilution of samples was performed in RO-H_2O to reach linear range (0.1-0.6).

Cell dry weight (CDW)

Cell dry weight (CDW) was determined by centrifugation of 10ml of cell suspension for 10min at 5000rpm (Heraeus Labofuge 200) and washing the cells with 5 ml RO-H_2O (resuspension and centrifugation). Cells were resuspended and transferred to a pre-weighed beaker, which was dried for 24 h at 105 °C and reweighed. To calculate total CDW, the volume of cultivation broth was calculated by summarising initial batch volume, feed media and base added and subtracting the volume of the samples taken.

Samples for protein quantification

To obtain equal contents of CDW per sample for cell desintegration, volumes containing an estimated CDW of 1mg were transferred to an Eppendorf tube and centrifuged for 10min at 4°C. The supernatant was discarded and probes were stored at -20°C until cell desintegration, see 3.10.

Koch plating

In order to observe the cells containing the linear cassette during the process and to determine the stability of the cultivation, bacterial cells were cultivated and colony forming units (CFU) were counted after 24h. Cells were plated at 10^{-7}, 10^{-8} and 10^{-9} dilutions on nutrient broth agar plates and on plates containing 25mg/l chloramphenicol.

Off-line fluorescence

The fluorescence of cells was measured using the microplate reader SPECTRAmax GeminiXS and SOFTmax PRO software (Molecular Devices). 100μl of culture broth and 100μl of culture supernatant were measured in well scan mode of density 3x3 (9 points per well) in black, flat bottom plates with translucent bottom (Costar). Excitation wavelength was set to 488nm and emission wavelength to 530nm for GFPmut3.1. For GFPmut3.1 detection the best signal to noise ratio was obtained using a 515nm cut-off filter.
For calculation of the specific cellular fluorescence, the fluorescence of the supernatant had to be subtracted from the fluorescence of the culture broth and be divided through the BDM.

FACS analysis

Total Cell Number (TCN) and percentage of dead cells (DC) were determined using flow cytometric methods. All measurements were performed with a FACSCalibur flow cytometer (four-color system; Becton Dickinson), equipped with an air-cooled laser providing 15mW at 488nm and the standard filter setup. For acquisition and analysis of data, the CellQuest software package (version 4.1, Becton Dickinson) was used.

Forward light scatter (FSC), right-angle light scatter (side scatter; SSC), and fluorescence were collected as pulse height signals (four decades of a logarithmic scale). Side light scatter (SSC) served as the primary detection parameter (threshold value 125). In addition, software discriminators were set on side scatter light signal to further reduce electronic and small particle noise. Discrimination of dead cells was done via staining with propidium iodide, while total cell number was determined via ratiometric counting. Samples were spiked with a known amount of fluorescent counting beads (Becton Dickinson, USA) so that the absolute cell number could be back-calculated.
Viability staining was performed with a combination of 3 fluorescence dyes: Bis-oxonole (BOX), which is lipophilic and anionic and is accumulated if the membrane is depolarised; ethidiumbromide (EB), which binds to DNA and is slightly better membrane-permeable than PI and is actively excluded; propidiumjodide (PI), which binds to DNA and is almost not membrane-permeable, binds to DNA if the membrane is destroyed. A sample of the culture broth was taken and diluted if necessary to a concentration of approximately 108 to 109 cells/ml. From this dilution, 20μl were transferred to 400μl of fresh-made dye solution (PI 5μg/ml, EB 10μg/ml, BOX 10μg/ml, all diluted in PBS + 0.1 mM EDTA). Dye solutions were stable for approximately 12h. Directly before measurement, the FACS was rinsed for 15sec with the sample to ensure that no residues of previous samples were present.

3.9.3 On-line measurement

O_2/ CO_2

To measure the concentration of CO_2 (%) and O_2 (%) in the exhaust gas a gas analyser system using gas analyser modules from ABB Inc were used. CO_2 was measured by infrared spectroscopy (analyser type Uras), while O_2 was measured based on paramagnetic properties of oxygen (analyser type Magnos).

Base consumption

The consumption of base during the fermentation process was recorded by detecting weight loss of the base tank, which was put an a balance. Base volume was calculated using the specific density of 0.91kg/l for 25% ammonia solution.

3.9.4 Calculated values

Culture volume

To calculate the culture volume of the cultivation process, the initial batch volume, the feed media and the added base were summed up and the volume of the taken samples was subtracted.

Growth rate

Growth rate was calculated according to equation 9 or equation 10.

$$\mu = \frac{1}{X} \cdot \frac{dX}{dt} \tag{9}$$

$$\mu = \frac{ln\frac{X_1}{X_2}}{t_2 - t_1} \tag{10}$$

Specific product formation rate (qP)

The specific product formation rate is defined as formation rate of desired product related to the average biomass present. Equation 11 shows the calculation for qP ($[mg/gBDM*h]$) between two measurement points.

$$qP = \frac{1}{\frac{X_1+X_2}{2}} \cdot \frac{P_2 - P_1}{t_2 - t_1} \tag{11}$$

K_p

K_p is the formation rate of the desired product related to the formation rate of biomass in non-fully growth associated conditions (instead of qP). Equations 12 and 13 show the calculation of K_P ($[mg \cdot 100/gBDM]$) between two measurement points.

$$K_P = \frac{\frac{dP}{dt}}{\frac{dX}{dt}} \cdot 100 \tag{12}$$

$$K_P = \frac{\frac{P_2-P_1}{t_2-t_1}}{\frac{X_2-X_1}{t_2-t_1}} \cdot 100 \tag{13}$$

3.10 Analysis of product yield

3.10.1 SDS gel electrophoresis

To determine if inclusion bodies are formed during the cultivation process cell dry weight (CDW) samples are desintegrated and cell supernatant containing the soluble protein fraction and dissolved inclusion bodies were loaded on a SDS gel.
Cell desintegration was performed from a 1ml shake flask sample or 1mg CDW fed-batch cultivation sample. To cell pellets $200\mu l$ cell desintegration buffer (all used reagents are given in appendix tables 6, 7 and 8) was added and samples were vortexed until the

pellet was completely resuspended. For cell disruption $50\mu l$ of 2mg/ml lysozyme and $50\mu l$ benzonase were added and incubated for 10min at RT (shaking on a TurboMix). After addition of $100\mu l$ triton X-100 samples were incubated again for 10min at RT (shaking on a TurboMix). Centrifugation for 10min at 4°C and 13000rpm separates the cell debris from the dissolved proteins in the supernatant. Supernatant was transferred into a fresh Eppendorf tube and stored at 4°C until loading on the NuPAGE gel. The pellet contains beside cell debris also inclusion bodies and was washed two times with 1ml wash buffer (Tris/HCl, $pH = 8.2$, 100mM). Samples were vortexed until pellet was resuspended completely and centrifuged for 10min at 4°C and 13000rpm. The IB-pellet was resuspended in $400\mu l$ solvent buffer and incubated for 30min (shaking at TurboMix). After 10min centrifugation supernatant containing the dissolved inclusion body proteins could be used for SDS gel electrophoresis.

Used SDS gels were NuPAGE 4-12% BisTris gradient gels, as running buffer a MES (2-(N-morpholino) ethanesulfonic acid) buffer was used. NuPAGE LDS Sample buffer (4x) and NuPAGE Reducing Agent (10x) were used for sample loading. Molecular weight marker was Mark12 ready-to-use (22). Electrophoresis settings were 200V, 400mA and 30min. Gels were fixed in 50% ethanol, 10% acetic acid (100%) for 30min and stained in Coomassie Brilliant Blue R250 Solution containing 25% Ethanol and 12.5% acetic acid (100%) 10 min. Destaining solution (25% Ethanol and 12.5% acetic acid (100%)) was added for several hours or overnight, changing 1-2 times. Destaining reaction was stopped by transferring the gel to water and gels were scanned using Corel Photo-Paint 12 and pictures analysed by using ImageQuantTL to determine the ratio between dissolved proteins and proteins in inclusion bodies.

3.10.2 ELISA

To quantify the concentration of produced recombinant protein in the shaking flask experiments or fed-batch cultivations enzyme-linked immuno sorbent assays were performed.
For cell desintegration $200\mu l$ cell desintegration buffer (all reagents for cell desintegration are given in appendix tables 9, 10) was added to the cell pellet (1ml cell suspension respectively 1mg CDW) and pellet was resuspended completely. $50\mu l$ lysozyme was added and cells were incubated for 10min at 37°C, shaking at 350rpm. Then $750\mu l$ triton X-100 were added and incubated as previously. After 10min centrifugation at 4°C and 13000rpm the supernatant was transferred into a fresh Eppendorf tube and could be stored at -20°C until the ELISA. All ELISA stock solutions and buffers are given in detail in 11 and 12.

ELISA plates were coated with first antibody 2h at RT or overnight at 4°C. Primary SOD antibody (mAb IAM-SOD M05) or primary GFP antibody (mAb mouse α GFP Sigma art. nr. G6539) were diluted in coating buffer. Plates were washed three times with water with Columbus Washer (TECAN art.nr. I209 0049).
Samples and standard were diluted following experienced data to receive a concentration

of 50ng SOD/ml in the highest dilution row and 20ng GFP/ml. SOD-standard had a concentration of $2\mu g$/ml and was diluted for use to a concentration of 50ng/ml in dilution buffer. GFP standard was recombinant GFP (BD Bioscience Clontech art.nr.632373) with a concentration of 1mg/ml and diluted in dilution buffer to a concentration of 20ng/ml. GFP ELISA plates had to be blocked with dilution buffer for 2h at RT to block unspecific binding sites on the first antibody. $300\mu l$ samples and standards were added to row H of an dilution plate and were diluted in 1:2 steps from row G to A ($140\mu l$ dilution buffer already added in all wells).
$50\mu l$ of the diluted samples were transferred to the coated ELISA plates and incubation for 1h at RT. After washing for three times $50\mu l$ second antibody were added and incubated for 1h. Anti-SOD antibody is binded to alcalic phosphatase, which can catalyze a reaction with the staining solution. Secondary GFP antibody is linked to biotin (goat α GFP biotin AB Acris R1091B) and can therefore react with a streptavidin-horseradish peroxidase conjugate in the staining solution. SOD ELISA plates could be measured after 10-20min after adding $100\mu l$ staining solution per well on a multiple channel photometer (Sunrise Remote/ Touch Screen, TECAN) at 405nm and reference wavelength 620nm.
For GFP ELISA samples were incubated with $100\mu l$ streptavidin-HRP conjugate for 1h after three times washing after secondary antibody incubation. Adding of $100\mu l$ staining solution starts staining reaction, which can be stopped after 3-5 min and plates can be measured in the photometer. Data were analyzed with software Magellan.

3.10.3 Correlation of SDS page and ELISA data

SDS page analysis was performed to determine the formation of inclusion bodies in the course of cultivation. The SDS page gels were scanned and the scanned pictures were imported into the ImageQuantTL software. This software can count the pixels of the protein bands in the scanned gel photograph and realizes therefore a quantification of the protein bands. With the export of the counted values to Microsoft Excel, the ratio between the soluble protein fraction and the inclusion body fraction could be calculated.
ELISA analysis quantified the protein amounts in mg/g in the soluble protein fraction. With the combination of the ELISA data and the ratios determined by ImageQuantTL software of SDS gel quantification, total protein levels could be calculated.
Because SDS gel electrophoresis was not performed for all sampling times, the received distribution curve was interpolated. The mean values for all sampling times resulted in the calculation of total protein amounts for all samples.

4 Results

In order to determine the limits of recombinant protein expression in chromosomal encoding *E. coli* systems, gene constructs containing two copies of the product gene (GFPmut3.1 and hSOD) were generated. The obtained tandem gene constructs were integrated in *E. coli* production hosts and were tested in shake flask and bioreactor cultivations. Their performance was compared to integrated single copy and plasmid based systems, to evaluate the expression limits of chromosomal integrated hosts.

4.1 Creation of the expression systems

Plasmid based systems and chromosomal expression systems underlying this work are based on the high activity of the T7 RNA polymerase. The product gene (GFPmut3.1 or hSOD) of plasmids or single copy product gene chromosomal integrated strains are under control of the T7 promoter and the T7 terminator. Additionally they contain a chloramphenicol resistance gene with it's own promoter and terminator. The linear cassette of the single gene chromosomal integrated strain was starting point for the creation of the tandem gene linear cassette. Based on the available starting plasmids containing the linear cassette (CAT:$T7^p$GFP or $T7^p$GFP:CAT) two different cloning strategies were developed to create two different variants of the tandem gene cassettes.
Variant 1 was designed with a T7 promoter and T7 terminator for both copies of the product gene, interrupted by the gene of the chloramphenicol resistance. In the design of the variant 2, the two copies of the product genes share a T7 promoter and a T7 terminator and are only interrupted by a ribosomal binding site (rbs). The chloramphenicol resistance gene follows after the T7 terminator.
Cloned cassettes were integrated in the chromosomes of the production strains BL21(DE3) and HMS174(DE3) to obtain the tandem gene expression systems.

4.1.1 Cloning of linear cassette variant 1

The two identical product genes of linear cassette variant 1 are under control of their own T7 promoter and T7 terminator. The two genes are interrupted by the chloramphenicol resistance gene, which has it's own promoter and terminator.

Starting plasmids for the creation of the linear cassette variant 1 were pBKS(−):Tn7::CAT:$T7^p$GFP, containing a linear cassette encoding the chloramphenicol resistance gene and a single copy of the product gene, and pBKS(−):Tn7::$T7^p$GFP:CAT, containing the product gene and resistance gene in a reversed alignment.

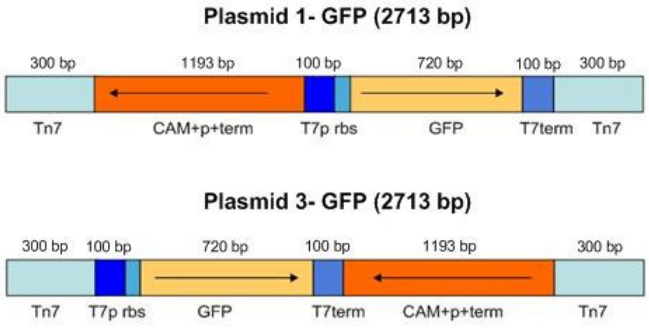

Figure 7: **Starting plasmids: pBKS(−) with linear cassettes Tn7::CAT:$T7^p$GFP or Tn7::$T7^p$GFP:CAT**

Figure 7 shows the linear cassettes of starting plasmids pBKS(−):Tn7::CAT:$T7^p$GFP and pBKS(−):Tn7::$T7^p$GFP:CAT, which were also called plasmid 1-GFP and plasmid 3-GFP. For linear cassettes containing hSOD product gene, the same starting plasmids were used containing the gene for hSOD instead of GFPmut3.1.

For the cloning strategy of the linear cassette variant 1, two PCR reactions were designed.

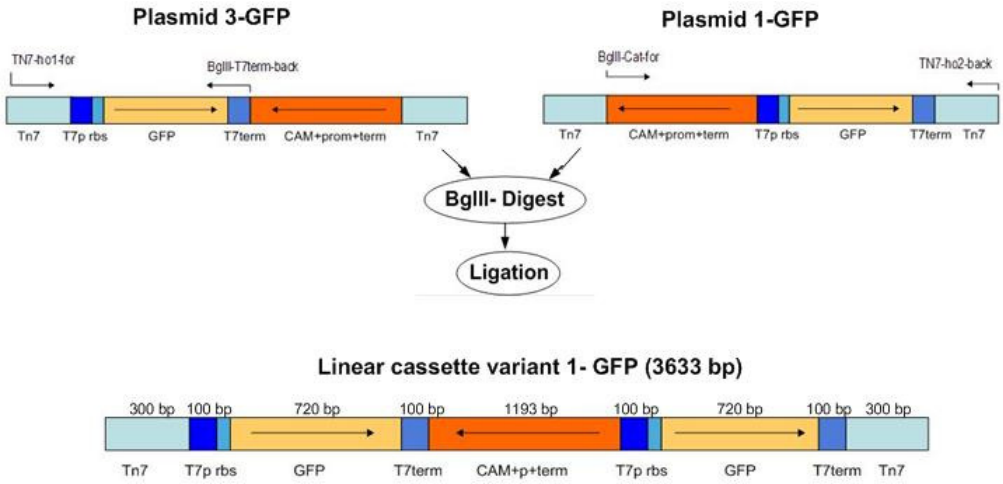

Figure 8: **Cloning procedure of linear cassette variant 1**

PCR of plasmid 1 taking primers BglII-CAT-for and Tn7-ho2-back resulted in a PCR product of 2413bp for GFPmut3.1 and 2161bp for hSOD and PCR of plasmid 3 taking primers Tn7-ho1-for and BglII-T7term-back delivered a PCR product of 1220bp for GFPmut3.1 and 968bp for hSOD. After purification of the PCR products a BglII digestion was performed and purified restriction fragments were ligated, resulting in the linear cassette of the acurate size. Figure 8 illustrates the single steps of the cloning procedure of linear cassette variant 1.

4.1.2 Cloning of linear cassette variant 2

The identical product genes in linear cassette variant 2 were arranged next to each other, only separated by a ribosomal binding site (rbs), sharing the T7 promoter and T7 terminator. The linear cassette variant 2 was created in a similar way as linear cassette 1, but in order to receive the target genes directly after each other, starting plasmid 3 was taken for the two PCR reactions. PCR reaction 1 was using primers Tn7-ho1-for and SpeI-GFP-back or SpeI-SOD-back to obtain PCR products of 1120bp length for GFPmut3.1 and 868bp for hSOD. PCR reaction 2 using SpeI-rbs-GFP-for or SpeI-rbs-SOD-for and Tn7-ho2-back resulted in PCR fragments of 2313bp length for GFPmut3.1 and 2061bp for hSOD. The subsequent SpeI digest was followed by a ligation, leading to the linear fragment of the desired size. Figure 9 shows the underlying steps of the cloning procedure of linear cassette variant-2.

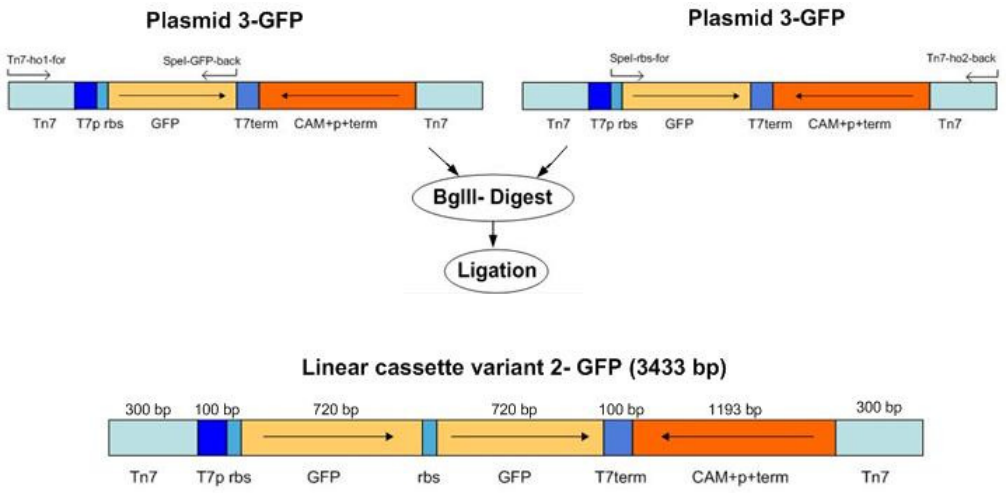

Figure 9: **Cloning procedure of linear cassette variant 2**

4.1.3 Results of the Cloning

The product yields after the individual PCR reactions lead to sufficient product for restriction digestions. After several trials it appeared to be more useful to clean up the PCR fragments and the digested PCR fragments only by a PCR clean-up kit and not by agarose gel electrophoresis to enhance the concentration of products. After ligation the long formed DNA of about 3500bp had to be amplified by PCR to receive sufficient amounts of the linear cassettes with the tandem genes. Optimization of the PCR reaction was necessary, because of encountered missing product formation after ligation. The rise of the annealing time and the increase of PCR cycles lead to DNA concentrations high enough to work with.

To verify the sequence of the ligated construct, DNA had to be sequenced. The sequencing of the linear cassettes was performed with Tn7-ho1-for and Tn7-ho2-back primers and T7 promoter and T7 terminator primers to ensure the sequence accuracy. Because of the possible emerging mutations during amplification by KOD-XL polymerase, only sequenced checked PCR products were used for the genome integration. Figure 10 shows the agarose gel photograph of the four different linear tandem gene cassettes generated by PCR (Var 1-GFP has a size of 3633bp, Var 1-SOD a size of 3129bp, Var 2-GFP a size of 3433bp and Var 2-SOD a size of 2929bp).

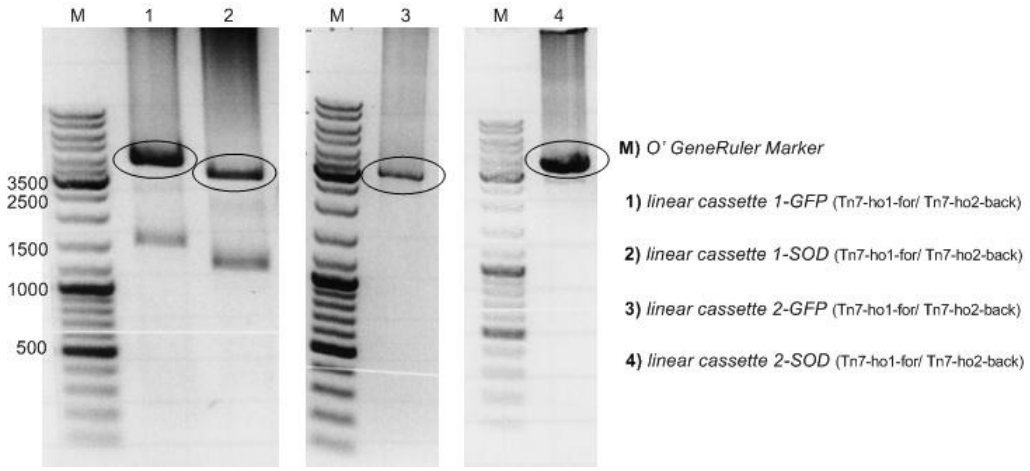

Figure 10: **PCR products of the linear cassettes var 1-GFP, var 2-GFP, var 1-SOD and var 2-SOD**

Cloning of the right sequenced linear tandem gene cassettes into pBKS(-) vectors, could be successfully performed for linear cassette variant 2-SOD. Mutations occurred in the other linear cassettes (var1-GFP, var2-GFP and var1-SOD) cloned into pBKS(-) vectors. A repetition was not performed, because of time reasons.

An important step to optimize the production of the linear cassettes by PCR would be the optimization of the PCR reaction itself. The utilization of different PCR polymerases, as the Pfu polymerase, the taq polymerase, the KOD-HiFi polymerase and the KOD-XL polymerase, showed greatest success with the KOD-XL polymerase. The increase of the PCR cycles from 30 to 35 cycles resulted in higher PCR product yields. The time-consuming experiments, which would be necessary to gain a further optimization stayed unattended at the expense of the integration of the linear cassettes into the host chromosome.

4.1.4 Chromosomal integration

λ Red recombination

Genome integration was started with transformation of linear cassette variant 2-GFP and SOD into Mg1655 *E. coli* bacterial cells. Mg1655 strain contains the Red helper plasmid pKD46, which is the plasmid with the λ Red genes *gam*, *bet* and *exo* under the control of P_{araB} promoter. pKD46 has a tl3 terminator and a temperature-sensitive origin, as described by Datsenko and Wanner [18]. For genome-integration 200, 400 and 600ng of linear DNA were used and positive clones received after transformation were analyzed by PCR screening.

Figure 11 shows the agarose gel of the PCR screening of Mg1655 clones. Primers binding at external sites outside of the Tn7 homologies were used to confirm the right integration of the linear cassette, ex1-for and ex2-back primers giving a PCR product of 3633bp. A PCR reaction with primer pairs (ho1-for and BsmBI-CAT-for resulting in a 2790bp fragment) and (ex1-for and BSMBI-CAT-for resulting in a 2890bp fragment) were used to verify the right length of the insert. Figure 11 shows 2 positive clones carrying a fragment of correct size. The bands of about 800bp in lane 1 and 4 result from dead cells on the plate taken for the PCR reaction, which have no linear cassette integrated. The size of 800bp is determined by the length of the homology regions which is about 300bp each and by the length of the external regions which is about 100bp each. After re-plating the clones on fresh plates containing chloramphenicol, this negative band disappears. DNA was sequenced after the cure of the pKD46 helper plasmid.

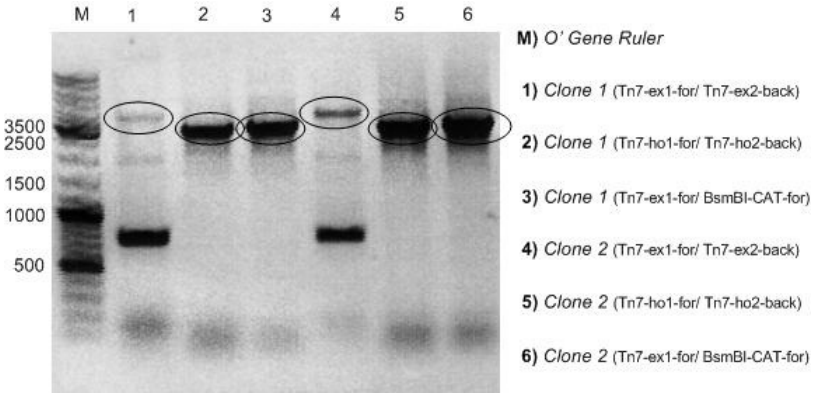

Figure 11: **Agarose gel of PCR screening of 2 clones picked after genome integration of linear cassette variant 2- GFP into Mg1655 chromosome**

Cure of the pKD46 plasmid is generated by the repeated streak-out on LB plates with chloramphenicol incubated at 42°C. The temperature-sensitive origin of replication (ORI) of the pKD46 plasmid avoided the amplification of the plasmid under such high temperatures. Although clones were received bearing an integrated linear cassette with the accurate

sequence, the integration strategy in detail described in materials and methods 3.6.1 was not pursued. The time-consuming steps of phage P1 transduction and receptor strain (HMS174(DE3) or BL21(DE3)) infection could be avoided by the use of the Quick & Easy *E. coli* Gene Deletion Kit (see reference [1] and subsection 3.6.2).

Quick & Easy *E. coli* Gene Deletion Kit

Chromosomal integration using the Quick & Easy *E. coli* Gene Deletion Kit facilitates the genome integration procedure by permitting a direct integration of the linear cassette into the production host. An intermediate integration into a cloning host followed by a P1 phage transduction is not necessary anymore, which helps to save time and working effort. The Quick & Easy *E. coli* Gene Deletion Kit depends mainly on the usage of the pRedET plasmid, which has the λ Red genes under the control of P_{araB} promoter and additionally the RecA gene. This plasmid allows a direct integration of the linear cassette into the host strain, for example HMS174(DE3), a recA-deficient strain.

HMS174(DE3) strain

Optimization of the Quick & Easy *E. coli* Gene Deletion Kit manual had to be performed for the use of HMS174(DE3) strain. Figure 12 shows the individual steps of the integration procedures. The pRedET plasmid carries the red α, β, γ genes of the λ phage together with the recA gene in a polycistronic operon under the control of the inducible P_{araB} promoter, which limits the recombination window therefore by the transient expression of the Red proteins.

Transformation of pRedET plasmids carrying ampicillin or tetracycline resistance were performed according to the Quick & Easy *E. coli* Gene Deletion Kit (see 3.6.2) in *E. coli* strains HMS174(DE3) and BL21(DE3). *E. coli* strains carrying the pRedET plasmid were stored at -80°C until genome integration was performed. To test the efficiency of the genome integration method, HMS174(DE3) cells were used. Because of the time and work-consuming procedure to receive enough amount of DNA for the long tandem gene PCR cassettes to be integrated, a shorter linear cassette was chosen to test the method.

Plasmid-1-GFP was used for the optimization of the Quick & Easy *E. coli* Gene Deletion Kit for HMS174(DE3) cells. Following the individual steps described in subsection 3.6.2 the integration of 500ng Tn7::CAT:$T7^p$GFP cassette in replicates, resulted in 60 and 150 clones on the two plates for pRedET-Amp plasmid and 400 and 500 clones for pRedET-Tet plasmid. To analyse the correct insertion of the linear cassette, a PCR screen was performed, which showed that all clones had the linear cassette (Tn7::CAT:$T7^p$GFP) integrated. Sequencing verified the observed results and showed that the Quick & Easy *E. coli* Gene Deletion Kit worked well for HMS174(DE3) cells.

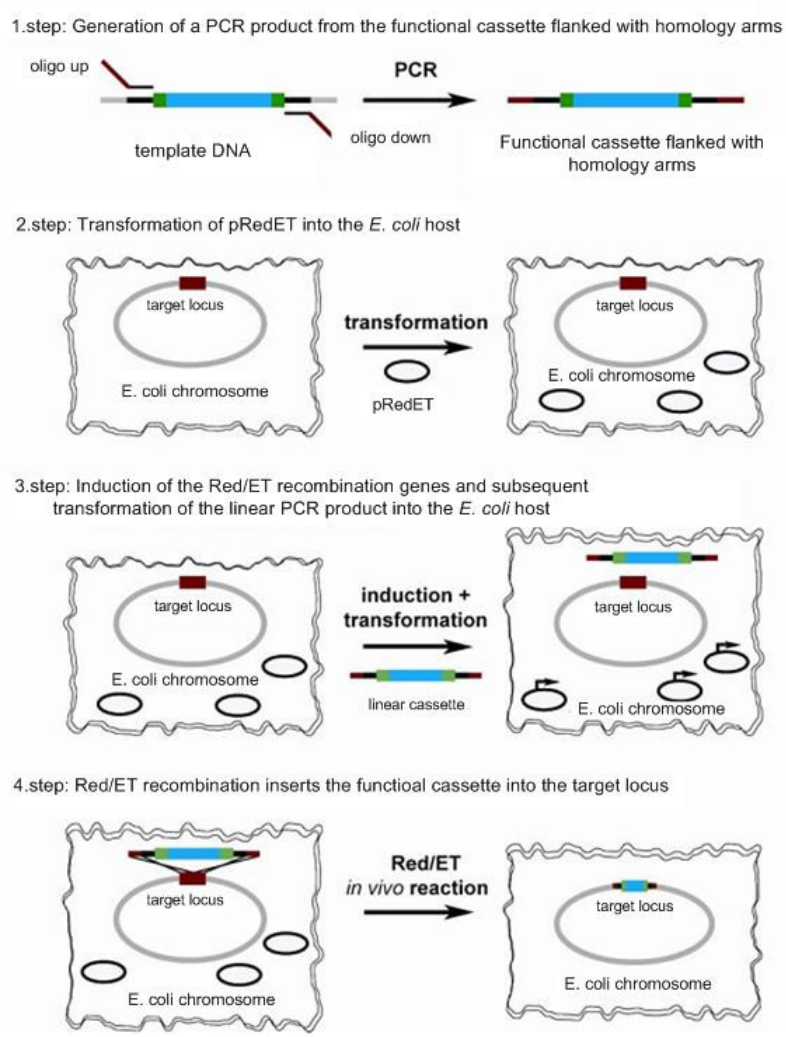

Figure 12: **Quick & Easy *E. coli* Gene Deletion Kit- individual steps**

Because of the positive results from the test integration with linear cassette Tn7::CAT:$T7^p$GFP, HMS74(DE3) strain with pRedET-Tet plasmid was chosen for the genome integration of the tandem gene cassettes. The preference for the pRedET-Tet plasmid was given, because of the high yield of clones after integration, which was 2-3 times higher as for ampicillin pRedET plasmid in the test integration. Another reason was the demand of the pharma industry to avoid the usage of ampicillin carrying plasmids in recombinant protein production processes. Although the pRedET plasmids are eliminated by the cultivation of the bacterial cells at 37°C, ampicillin resistance existed at distinct steps.

Integration of linear cassette variant 1-GFP and SOD and variant 2-GFP and SOD was performed in one experimental approach, using HMS174(DE3) pRedET-Tet strain. To

obtain an optimal amount of positive clones different concentrations of PCR product were used. 400 and 600ng DNA resulted in 20-200 clones on all plates. As a negative control in which expression of λ red genes was not induced by arabinose, linear cassette var2- GFP was used, because of the highest concentration of DNA.
PCR screening of positive clones could be performed on the day after tansformation. 7 clones from each variant were analysed by screening PCR using ex1-for and ex2-back primers. Figure 13 shows the agarose gel photograph of the PCR product of the external primer pair for one clone of each linear cassette. Clones shown are variant 1- GFP clone #7 (3833bp with external primers), variant 1- SOD clone #6 (3329bp with external primers), variant 2- GFP clone #4 (2633bp with external primers) and variant 2- SOD clone #5 (3129bp with external primers), which were sequenced, to verify the correctness of the integrated sequence. 800bp negative band can be seen for all clones. Sequencing showed that all four taken clones had a correct inserted tandem gene cassette.

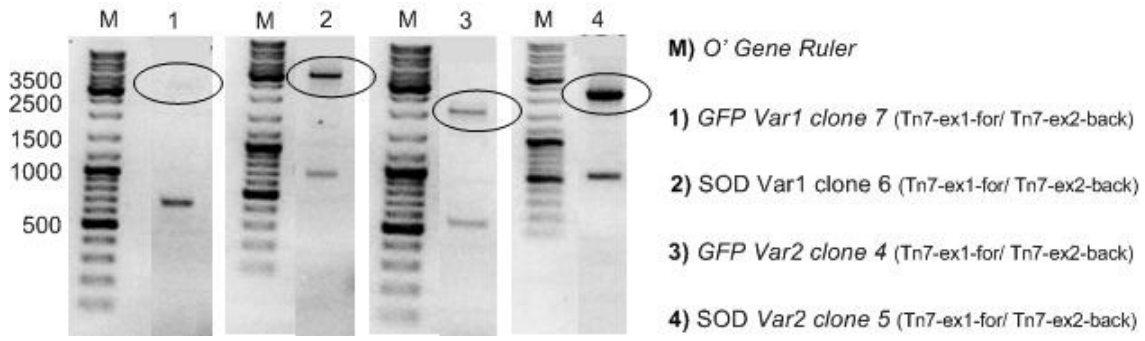

Figure 13: **Agarose gel of PCR screen of genome integrated tandem cassettes into host strain HMS174(DE3)**

BL21(DE3) strain

The successful production of HMS174(DE3) host cells carrying each of the four constructed cassettes lead to the desire to integrate the tandem gene cassettes also into the BL21(DE3) genome. BL21(DE3) genome integration using the Quick & Easy *E. coli* Gene Deletion Kit was repeated two times and did not result in any clones at all. The increased growth of BL21(DE3) cells after the shift from 30°C to 37°C brings about only a short incubation time at 37°C, which is necessary to start transcription of λ red genes, or a cell density, which brings cells out of the exponential growth phase. Both facts could be the reason for the inefficient genome integration. A large approach to test different cell densities, different incubation times at 37°C or different DNA concentrations would be necessary to establish the genome integration using the Quick & Easy *E. coli* Gene Deletion Kit in BL21(DE3), but was not realized in the course of this work, because of time reasons.

4.2 Testing of the constructs under cultivation conditions

4.2.1 Creation of cell banks

In order to obtain a reproducible inoculum in the experiments, it is necessary to create cell banks, in which cells were grown under equal conditions and are stored at equal densities. Overnight cultures of positive HMS174(DE3) clones were frozen in 1.5ml aliquots and stored as a research cell bank (RCB). Master cell banks (MCB) were prepared for HMS174(DE3):Tn7::GFP:CAT:GFP, HMS174(DE3):Tn7::SOD:CAT:SOD and HMS174(DE3):Tn7::SOD:SOD:CAT strains, according to the description given in 3.7.2. A master cell bank of HMS174(DE3):Tn7::GFP:GFP:CAT could not be prepared, because the cells did not grow in the M9-ZB medium used for master cell bank preparation.

4.2.2 Shake flask experiments

Shake flask cultivations were performed to test the functionality of the created linear cassettes, to determine the product yield of the recombinant protein and to analyse the influence of the product formation on the cellular growth. The cultivation medium was changed from LB to semisynthetic medium, to provide process conditions equal to bioreactor cultivations. Product formation of soluble recombinant protein was maintained by ELISA and inclusion body aggregated protein by SDS page analysis.
The behavior of the tandem gene strain was compared to a single copy integrated reference strain. The experiments were performed according to the description given in 3.8.

Functionality test of GFP constructs

To test the functionality of the GFP constructs a shake flasks experiment was performed with 40ml LB medium. GFP expression of HMS174(DE3):Tn7::GFP:CAT:GFP and HMS174(DE3):Tn7::GFP:GFP:CAT strain was started at an $OD_{600} = 1.2$. Samples were taken prior induction, 1h post induction and after 20h post induction.
GFP expression was found for both tandem GFP variants. Figure 14.1 shows the GFP production levels measured by offline fluorescence 20h post induction and indicates a higher production capacity for linear cassette variant 1- GFP. The fluorescence values are shown in dependence on OD_{600} values. After 20h of production the culture broth was stained green, because of GFP production, shown in Figure 14.3. The green staining could also be determined after centrifugation of the culture broth in the cell pellet, shown in Figure 14.2.

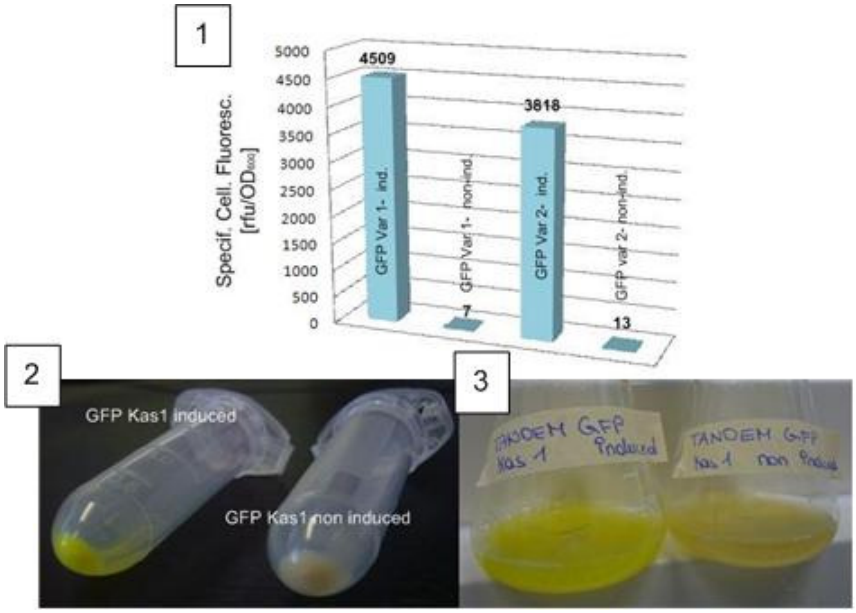

Figure 14: **GFP amounts after 20h post induction of 40ml shake flask experiment**

Shake flask cultivations of GFP constructs

Shake flask cultivation of HMS174(DE3):Tn7::GFP:CAT:GFP and the reference strain HMS174(DE3):Tn7::CAT:GFP showed that both strains have the same growth behavior and produce similar amounts of soluble GFP, detected by offline fluorescence.
In Figure 15 the diagramm 1 shows the courses of bacterial growth given by the OD_{600} values. The cellular fluorescence in relation to the OD_{600} is also given in the diagram and shows increased levels of soluble GFP for the tandem gene strain at 3h post induction, which was also confirmed by GFP-ELISA (Fig.15.3). Inclusion body formation was detected for samples 3h post induction. SDS gel presented in Figure 15.2 shows inclusion body formation for both strains, but a higher amount of inclusion bodies for the tandem gene strain.
Tandem gene strain has a higher amount of soluble protein, a strong increased amount of protein aggregated in inclusion bodies and therefore a higher total protein yield. The ratio of soluble protein to inclusion body aggregated protein of 73% to 27% in the single copy strain was shifted to 54% soluble protein to 46% inclusion body aggregated protein in the tandem gene strain. Total protein amounts in the single copy strain were 38mg/g recombinant protein and in the tandem gene strain 55mg/g recombinant protein at 3h post induction.

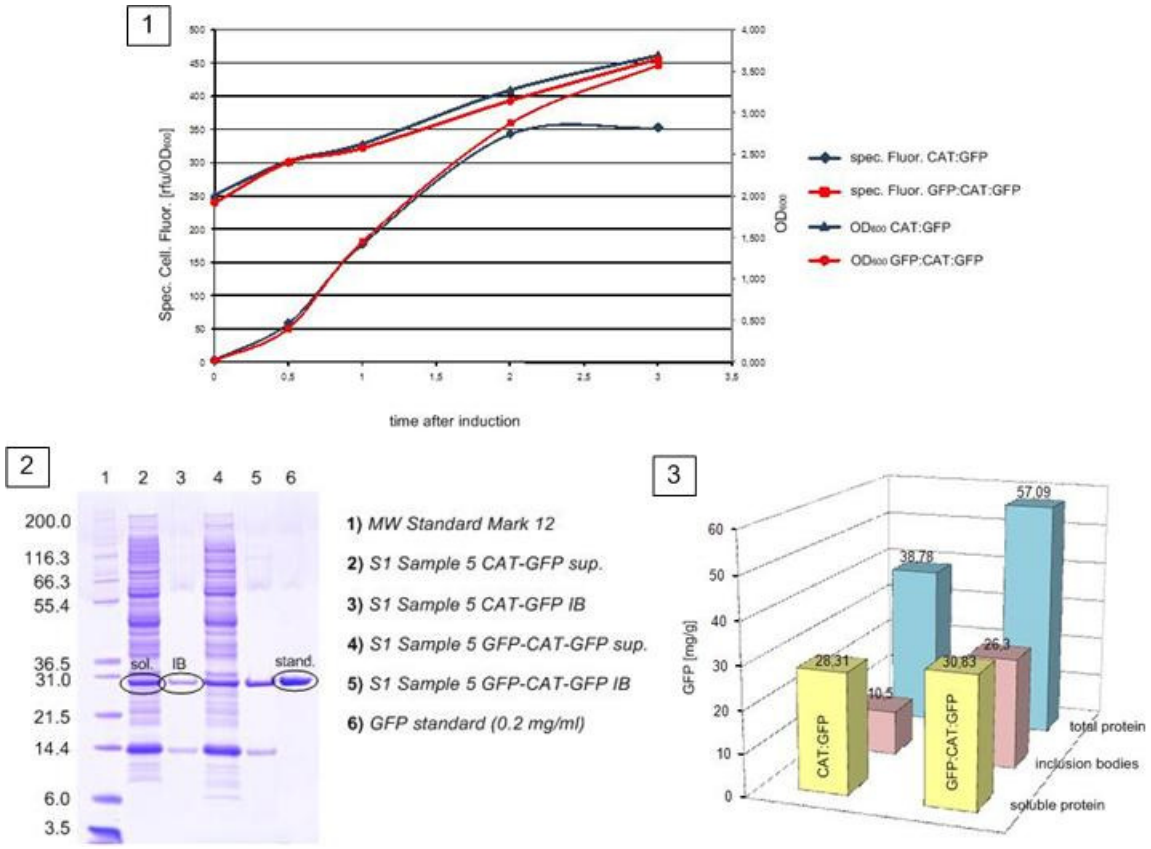

Figure 15: **Single copy strain HMS174(DE3):Tn7::CAT:GFP and tandem gene strain HMS174(DE3):Tn7::GFP:CAT:GFP in the course of a 400ml shake flask experiment. Growth behavior of the strains given in 15.1, SDS page analysis in 15.2 and total protein quantification in 15.3.**

Shake flask cultivations of SOD constructs

The comparison of tandem gene strains HMS174(DE3):Tn7::SOD:CAT:SOD and HMS174(DE3):Tn7::SOD:SOD:CAT with the reference strain HMS174(DE3):Tn7::CAT:SOD in a shake flask cultivation showed that recombinant protein productivity is increased for the tandem gene construct variant 1.
The comparison of the growth curves, given in Figure 16.1 by the measured OD_{600} values, shows that all three cultivated strains have similar growth rates.
SDS page and ELISA samples were taken 3h post induction. SDS page gel electrophoresis showed no inclusion body formation for all three strains (Fig.16.2). The total protein amount is equivalent to quantified ELISA protein concentrations, because of the absence of inclusion bodies. Total protein amounts are presented in Figure 16.3 and show that the tandem gene strain with the hSOD variant 1 integrated had the highest level of hSOD, the reference strain and the tandem gene strain with variant 2 produce similar levels of hSOD.

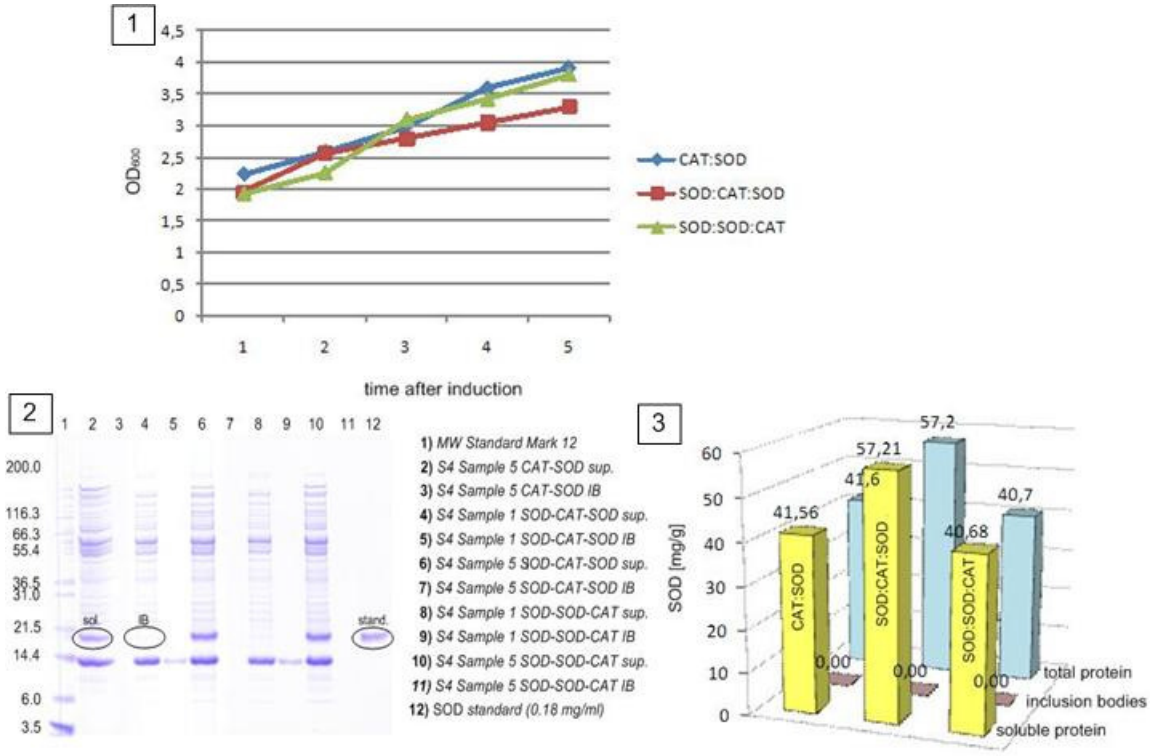

Figure 16: **Single copy strain HMS174(DE3):Tn7::CAT:SOD and the tandem gene strains HMS174(DE3):Tn7::SOD:CAT:SOD and HMS174(DE3):Tn7::SOD:SOD:CAT in the course of 400ml a shake flask experiments**

Conclusion of shake flask experiments

Shake flask experiments were useful to check the functionality of the new tandem gene integrated HMS174(DE3) strains. The functionality of the integrated cassettes could be demonstrated for all four constructs. It was shown that the linear cassette variant 1, containing two independently transcribed copies of the target gene, increases the level of produced target gene in comparison to the chromosomal integrated strain with one copy of the target gene. Inclusion body formation in the course of GFP construct cultivation was observed.

4.2.3 Bioreactor cultivation

Bioreactor cultivations were performed to test the production strains in stable and defined cultivation conditions. In the course of a bioreactor cultivation the product formation rates can be calculated, because of the high amount of defined points of sampling. Tandem gene strain HMS174(DE3):Tn7::SOD:CAT:SOD was taken for bioreactor cultivation, because of the detected high productivity of fully soluble recombinant protein in shake flask experiments. The Fed-batch cultivation was run with a carbon limited exponential feed and the

growth rate was set to $\mu = 0.1 * h^{-1}$ (detailed description given in 3.9).
The behavior of the tandem gene strain under bioreactor conditions, was compared to a single copy chromosomal system and a plasmid based system. HMS174(DE3):Tn7:CAT:SOD and HMS174(DE3) pet30s(hSOD) are the strains of the reference cultivations.

Cultivation behavior of the tandem gene strain

The comparison of the tandem gene strain cultivation with the reference cultivations of the single copy strain and the plasmid strain show that the tandem gene strain shows an intermediate behavior between the two other systems.
The plasmid based system is characterized by a high product formation rate of 35-40mg/gh starting immediately post induction, shown in Figure 17.1. This high recombinant protein formation, which is turned on by the induction of the strong T7 expression system, exceeds the cellular production capacity. The influence of the product formation on the growth rate is enormous, which can be derived from the deviation of the actual CDW from the calculated CDW. FACS analysis of the cells showed that cells stop to grow, but do not dye after induction.
Cell growth stops 2h post induction, but product formation is maintained for more than 8h. This means that product formation is decoupled from cellular growth. The resulting increase in recombinant protein amount depends therefore not on the increase of biomass, but on the metabolization of host cell proteins.
The specific content of recombinant protein reaches a maximal level of 200mg/g CDW, but recombinant protein production is stopped 9h post induction, because of the metabolic overload on the host cells..
In the experiment with the single copy strain (Fig.17.2) growth and product formation were maintained over the whole process. For the first 15h a constant qP of about 25mg/gh was calculated. In comparison to the stop of growth caused by the recombinant protein production from plasmids, the induction of recombinant protein expression in the chromosomal system has a relatively mild influence on the growth rate of the host cell. The specific recombinant protein production reaches a value of 180mg/g CDW at the end of the process and is growth-coupled until the end of the process. With this system host cell capacity is not exceeded, which allows constant production.
Recombinant protein production in tandem gene integrated hosts (Fig.17.3) showed, that two identical copies of the product gene, lead to a higher qP, but a decreased growth rate. The qP reached in this process was about 25-30mg/gh. The decrease of the qP after 8h post induction correlates with the increasing aberration of the actual CDW from the calculated CDW. The growth declines to half of the calculated growth rate, but prolongs until the end of the process. Specific recombinant protein production is higher for the tandem gene strain than for the single copy strain and reaches a maximal level of 240mg/g CDW. With the additional product gene copy, the limits of the host cell production capacities are reached. The additional protein synthesis influences the host cell growth rate, but does not cause a complete stop of growth as detected in plasmid systems.

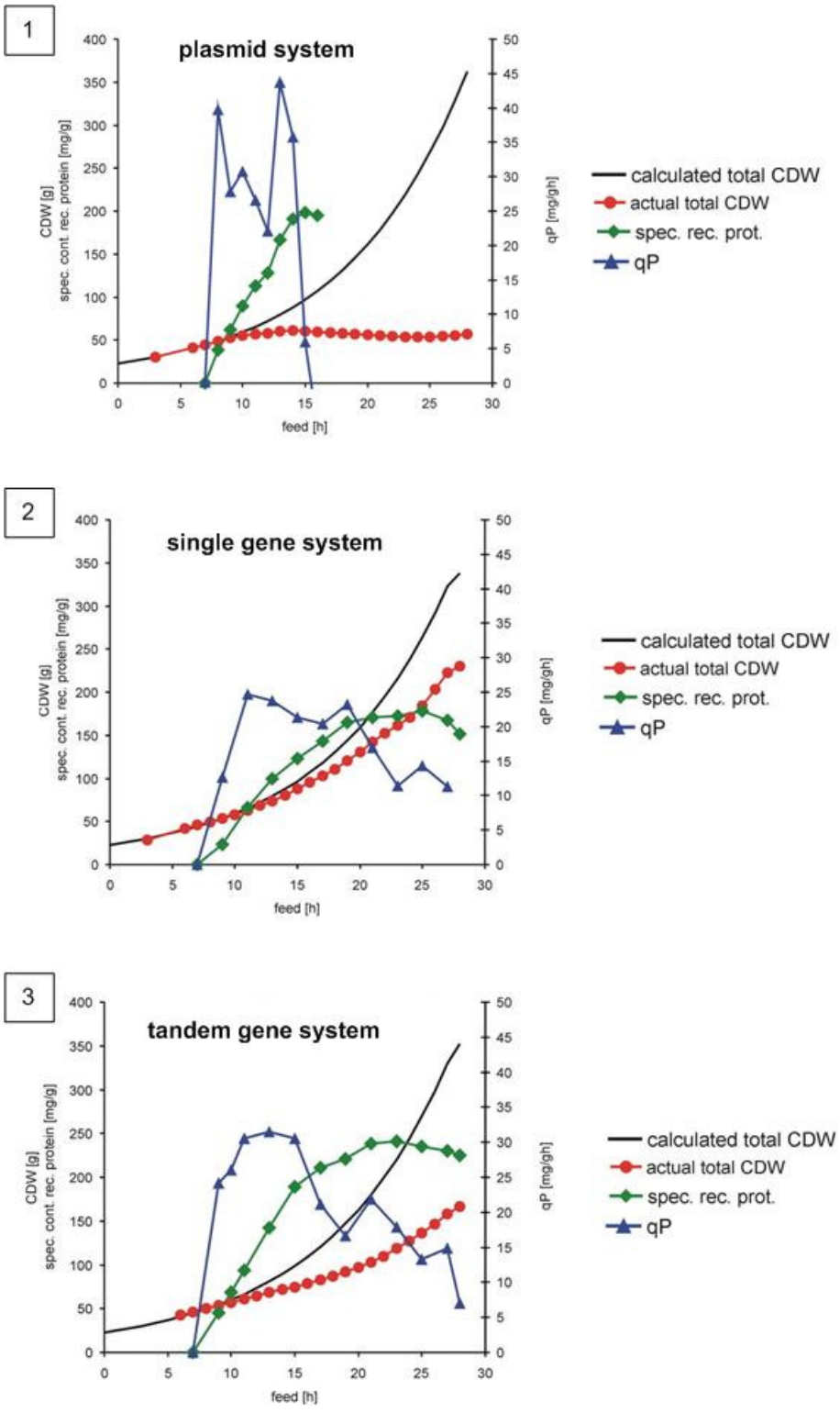

Figure 17: **Bioreactor cultivations of plasmid strain HMS174(DE3)pet30a(hSOD), single copy strain HMS174(DE3):Tn7::CAT:SOD and tandem gene strain HMS174(DE3):Tn7::SOD:CAT:SOD**

Specific recombinant protein production

The quantification of recombinant protein levels of the plasmid based system showed that the host cell produces about 200mg/g CDW recombinant protein at the end of the process, but about 60% of the produced recombinant protein is stored in inclusion bodies. Figure 18.1 shows total protein amounts in the course of cultivation and the distribution of the recombinant protein in soluble protein and inclusion body aggregated protein. In the first 2h post induction produced recombinant protein stays soluble and reaches a value of 60mg recombinant protein per g CDW. This soluble recombinant protein content remains the same throughout the process and starting with 2h post induction produced recombinant protein aggregates in inclusion bodies. Subsequent recombinant protein production only increases the levels of protein in inclusion bodies, but not in the soluble fraction. The ratios of inclusion body to soluble protein throughout the cultivation process are presented in the diagram 1 in Figure 20.

The single copy chromosomal integrated strain produces at the end of the process 180mg hSOD/g CDW, which stay soluble and do not aggregate in inclusion bodies. Figure 18.2 shows the specific recombinant protein amounts in the course of the cultivation. The total protein amounts are given by the soluble protein amounts, because no inclusion bodies were formed in the process.

What can be concluded from this cultivation is, that the product concentration of the recombinant protein is presumably not responsible for the formation of inclusion bodies, because the specific recombinant protein amounts are similar for the plasmid based and single copy system. The high product formation rate in the plasmid based systems seems to overload the host cell protein folding mechanisms and causes therefore the inclusion body formation. The single copy chromosomal system does not cause such an overload on the protein folding mechanisms of the host cell.

The recombinant protein levels in the tandem gene strain are higher than in the single copy strain, but inclusion bodies are formed in the course of cultivation. As Figure 18.3 shows, inclusion body formation starts very early in the process, almost 3h post induction. This phenomenon was observed by SDS page gel electrophoresis as shown in Figures 19.2 and 19.3. Since from the gel in Figure 19.2 IB formation between 3 to 7h post induction can be derived, a more detailed analysis was performed to determine the starting point of IB formation. Figure 19.3 shows that IB formation starts already 4h after induction. In contrast the single copy construct does not show IB formation (Fig. 19.1).

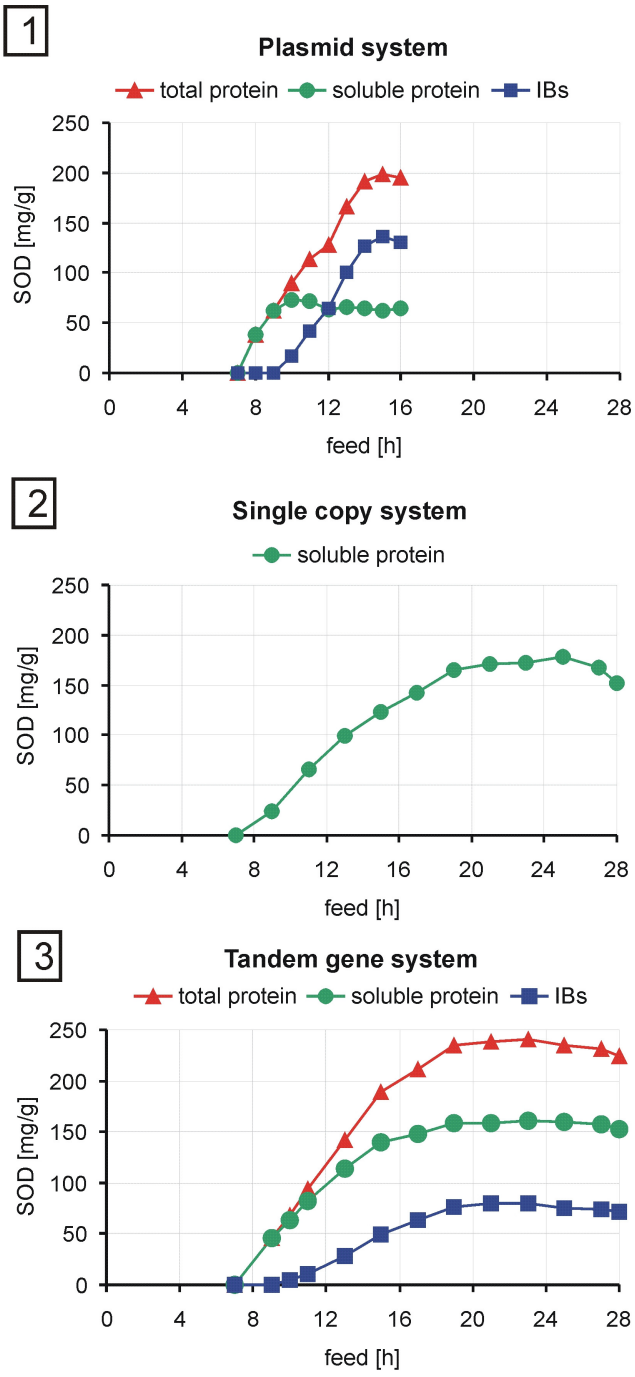

Figure 18: **Total protein amounts of plasmid strain HMS174(DE3)pet30a(hSOD), single copy strain HMS174(DE3):Tn7::CAT:SOD and tandem gene strain HMS174(DE3):Tn7::SOD:CAT:SOD**

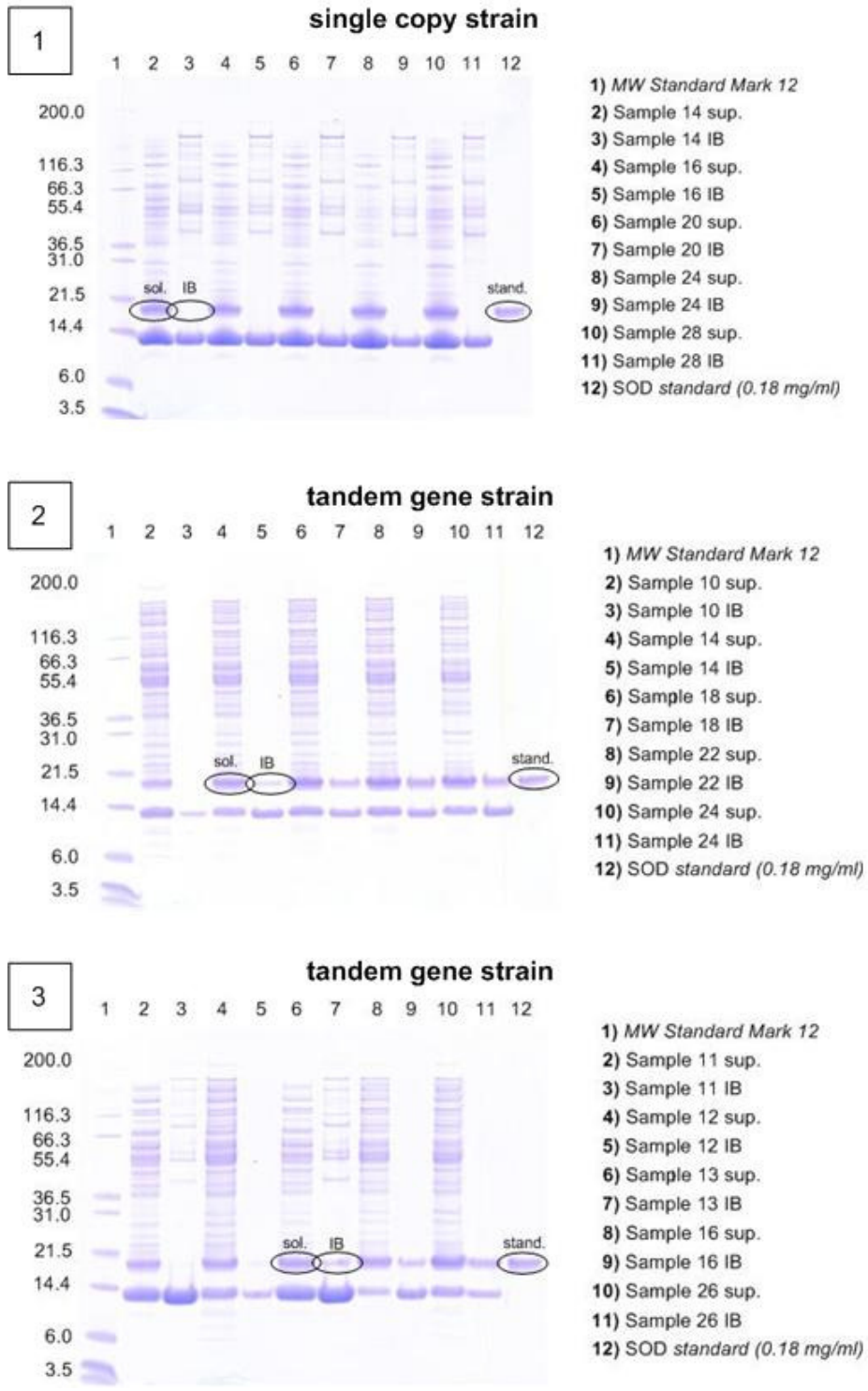

Figure 19: **Detection of inclusion body formation by SDS page gel electrophoresis of single copy strain HMS174(DE3):Tn7::CAT:SOD and tandem gene strain HMS174(DE3):Tn7::SOD:CAT:SOD in the course of a bioreactor cultivation (Sample number is equivalent to feed duration)**

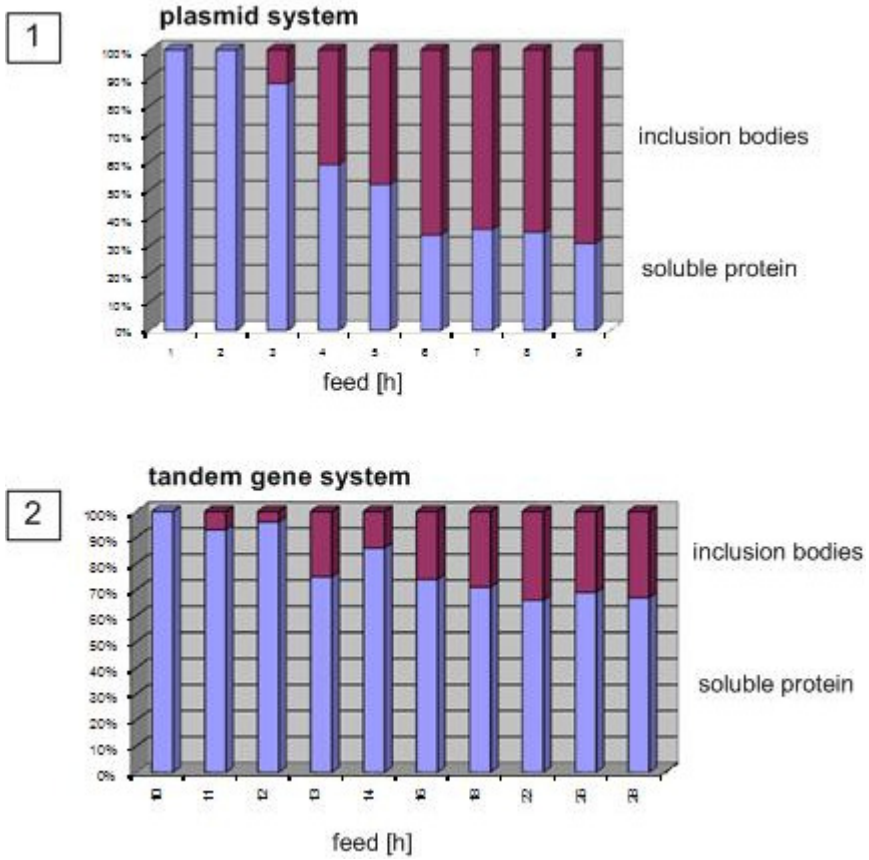

Figure 20: **Ratios of inclusion body formation and soluble protein of plasmid based system HMS174(DE3)pet30a(hSOD) and tandem gene system HMS174(DE3):Tn7::SOD:CAT:SOD**

The ratio of inclusion body to soluble protein over the course of cultivation for the tandem gene construct is given in Figure 20.2. The diagram shows the increase of inclusion body formation throughout the process with a maximum of 35% inclusion bodies. In comparison the plasmid based system (Fig.20.1) produces up to 60 % IBs.
The tandem gene integrated strain produces with a higher qP at the beginning of the process and attains a higher content of specific recombinant protein. The level of the soluble fraction is similar to the single copy construct, whereby the additional amount of recombinant protein accumulates as IBs. It seems that the maximum concentration of soluble hSOD is reached by the single copy strain and lies about 180mg recombinant protein per g CDW.
However, the integration of a second product gene in the host chromosome did not lead to the complete overload of the host metabolism, as it is seen with plasmid based systems. In order to avoid inclusion body formation, a tuning regime should be applied to control the transcription rate in relation to metabolic capacity of the host cell.

5 Conclusion

The aim of the work was the creation of a tandem gene chromosomal integrated expression system in order to determine the productivity limits of the chromosomal recombinant *E. coli* protein production systems.
The linear cassettes containing identical copies of the product gene and the subsequent integration of the cassettes into the chromosome of an *E. coli* production host was successfully performed. Tandem gene integrated hosts replicate stable and the system showed high stability during cultivations.

The recombinant protein production could be efficiently induced and cultivation experiments in shake flask or bioreactor showed superior product formation rates then the single copy strains. However tandem gene systems reduced growth, which can be referred to increased recombinant protein production. Nonetheless the overall yield of tandem gene systems is lower than single copy strains.
While cessation of growth, which is a problem of plasmid based systems, could not be observed in the tandem gene system, inclusion body formation was detected in plasmid and tandem gene systems, as well. The aggregation can presumably be addressed to the overload of the host cell folding machinery due to the higher qP resulting from the additional copy in the tandem gene strain.

Although the expression of the recombinant protein exceeded the capacity of the tandem gene strain, the determination of the accurate limits of the production capacity is difficult. The reasons therefore are the maintainance of recombinant protein production over a longer period of time in comparison to plasmid based systems. An important observation was that the chromosomal system did not collapse, as it is the case for plasmid based systems.

Considering the results of the presented experiments it comes clear that the expression of recombinant protein by tandem gene cassettes in combination with tuning strategies for controlling the transcription rate shows high potential for the development of an efficient production system. Moreover it can be strongly assumed that translation rather than transcription poses the bottleneck of recombinant gene expression.

References

[1] http://www.genebridges.com/gb/pdf/k006-ecoli-gene-deletion-kit-v2.0-2007.pdf.

[2] http://www.lgc.co.uk/pdf/barcode20flyer-web202008.pdf.

[3] http://www.microbiology.emory.edu/altman/f-protocols/f-tetramers/autoind-annot.html.

[4] http://www.mun.ca/biochem/courses/3107/topics/ara-operon.html.

[5] http://www.nanodrop.com/techsupport/nd-1000-users-manual.pdf.

[6] http://www.neb.com/nebecomm/tech-reference/restriction-enzymes/overview.asp.

[7] http://www.neb.com/nebecomm/techbulletinfiles/techbulletinm0202.pdf.

[8] http://www.promega.com/tbs/tb225/tb225.pdf.

[9] http://www.promega.com/tbs/tb308/tb308.pdf.

[10] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, a member of the Taylor&Francis group, 2002.

[11] P. O. Angrand, N. Daigle, F. van der Hoeven, H. R. Schöler, and A. F. Stewart. Simplified generation of targeting constructs using ET recombination. *Nucleic Acids Research*, 27, 1999.

[12] G. Antranikian. *Angewandte Mikrobiologie*. Springer Verlag, 2006.

[13] B.J. Bachmann. Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriology Reviews*, 36:525–557, 1972.

[14] F. Baneyx. Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*, 10:411–421, 1999.

[15] M. Bzymek and S. T. Lovett. Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *PNAS*, 98:8319–8325, 2001.

[16] M.M. Carrio and M. Villaverde. Protein aggregation as bacterial inclusion bodies is reversible. *FEBS letters*, 489:29–33, 2001.

[17] M. Cserjan-Puschmann, W. Kramer, E. Duerrschmid, G. Striedner, and K. Bayer. Metabolic approaches for the optimisation of recombinant fermentation process. *Applied Microbiology and Biotechnology*, 53:43–50, 1999.

[18] K. A. Datsenko and B. L. Wanner. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *PNAS*, 97:6640–6645, 2000.

[19] J.D. Diaz-Ricci and M.E. Hernandez. Plasmid-effects on Escherichia coli metabolism. *Critical Reviews in Biotechnology*, 20:79–108, 2000.

[20] P.M. Doran. *Bioprocess Engineering Principles*. Academic Press, San Diego, 2003.

[21] P. Feng, S. Weagant, and M. Grant. Enumeration of *Escherichia coli* and the coliform baceria. *Bacteriological Analytical Manual*, 8, 2002.

[22] N. J. Gay, V. L. J. Tybulewicz, and J. E. Walker. Insertion of transposon Tn7 into *Escherichia coli* glmS transcriptional terminator. *Biochemical Journal*, 234:111–117, 1986.

[23] R. Grabherr, E. Nilsson, G. Striedner, and K. Bayer. Stabilizing plasmid copy number to improve recombinant protein production. *Biotechnology and Bioengineering*, 77:142–147, 2002.

[24] C. H. Gray and E.L. Tatum. X-ray induced growth factor requirements in bacteria. *Proc. Nat. Acad. Sci. USA*, 30:404–410, 1944.

[25] L. M. Guzman, D. Belin, M. J. Carson, and J. Beckwith. Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. *Journal of Bacteriology*, 177:4121–4130, 1995.

[26] J. Lederberg and J.L. Tatum. Gene recombination in *Escherichia coli*. *Nature*, 158:558, 1946.

[27] A. S. Morag, C. J. Saveson, and S. T. Lovett. Expansion of DNA repeats in *Escherichia coli*: effects of recombination and replication functions. *Journal of Molecular Biology*, 289:21–27, 1999.

[28] K. C. Murphy and K. G. Campellone. Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic *Escherichia coli*. *BMC Molecular Biology*, 2003.

[29] K.C. Murphy. Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. *Journal of Bacteriology*, 180:2063–2071, 1998.

[30] J. P. P. Muyrers, J. Zhang, G. Testa, and F. A. Stewart. RecE/RecT and Redα/Redβ initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Development*, pages 1971–1982, 2000.

[31] A. R. Poteete. What makes the bacteriophage λ Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiology letters*, 201, 2001.

[32] H. Reischer, I. Schotola, G. Striedner, F. Pötschacher, and K. Bayer. Evaluation of the GFP signal and its amplitude for novel on-line monitoring strategies of recombinant fermentation process. *Journal of Biotechnology*, 108:115–125, 2004.

[33] C. J. Saveson and S. T. Lovett. Enhanced deletion formation by aberrant DNA replication in *Escherichia coli*. *Genetics*, 146:457–470, 1997.

[34] H.P. Sorensen and K.K. Mortensen. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology*, 115:113–128, 2005.

[35] G. Striedner, M. Cserjan-Puschmann, F. Pötschacher, and K. Bayer. Tuning the transcription rate of recombinant protein in strong *Escherichia coli* expression systems through repressor titration. *Biotechnology Progress*, 19:1427–1432, 2003.

[36] G. Striedner, I. Pfaffenzeller, M. Luchner, S. Nemecek, R. Grabherr, and K. Bayer. Plasmid-free T7 based *Escherichia coli* expression system. *in preparation*, 2008.

[37] F.W. Studier, A.H. Rosenberg, J.J. Dunn, and J.W. Dubendorff. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymology*, 185:60–89, 1990.

[38] J. R. Swartz. Advances in *Escherichia coli* production of therapeutic proteins. *Current Opinion in Biotechnology*, 12:194–201, 2001.

[39] Y. Zhang, F. Buchholz, J.P.P. Muyrers, and A.F. Stewart. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genetics*, 20:123–128, 1998.

6 Abbreviations

AA	amino acid
amp	ampicillin
bp	base pairs
cfu	colony forming units
cam	chloramphenicol
CAT	chlorampenicol trasnferase
DMSO	Dimethyl-Sulfoxide
DNA	desoxyribonucleic acid
E. coli	*Escherichia coli*
ELISA	Enzyme-Linked Immunosorbent Assay
GFP	green fluorescent protein
h	hour(s)
HPLC	High Performance Liquid Chromatography
IPTG	isopropyl β-D-thiogalactopyranoside
kD	kilo Dalton
MCB	master cell bank
min	minute(s)
mRNA	messenger ribonucleic acid
OD_{600}	optical debsity at wavelength of 600nm
PBS	phosphor buffered saline
PCR	polymerase chain reaction
qP	specific product formation rate
RCB	research cell bank
rfu	relative fluorescence unit [arbitrary unit]
hSOD	human superoxid (Cu/Zn) dismutase
RNA	ribonucleic acid
rpm	rotations per minute
sec	second(s)

7 Appendix

7.1 Cultivation of *E. coli*- media and buffers

All chemicals were purchased from Merck unless otherwise noted.

7.1.1 Stock solutions

Ampicillin stock solution

100mg/ml ampicillin dissolved in RO-H_2O, sterile filtered; working concentration 100μg/ml (1:1000)

Chloramphenicol stock solution

25mg/ml dissolved in 96% ethanol, sterile filtered; working concentration 25μg/ml (1:1000)

1M $MgSO_4$

24.6g $MgSO_4 * 7H_2O$ dissolved in 100ml RO-H_2O, autoclaved

0,9% sodium chloride solution

9g NaCl dissolved in 1l RO-H_2O, autoclaved

87% Glycerol

autoclaved

Stock solution of trace elements

The chemicals given in table 2 were dissolved in 5M HCl. The volume was adjusted to 1l. The solution was stored in a glass bottle at room temperature.

Order	Component	Amount per litre
1	$FeSO_4 * H_2O$	40.0g
2	$MnSO_4 * H_2O$	10.0g
3	$AlCl_3 * 6H_2O$	10.0g
4	$CoCl_2$	4.0g
5	$ZnSO_4 * 7H_2O$	2.0g
6	$Na_2MoO_2 * 2H_2O$	2.0g
7	$CuCl_2 * 2H_2O$	1.0g
8	H_3BO_3	0.5g

Table 2: Composition of a stock solution of trace elements

7.1.2 Solid media

Luria Bertani (LB) agar

1% NaCl
1% bacto-tryptone
0.5% yeast extract
1.5% agar-agar
The chemicals were resuspended in RO-H_2O and autoclaved. Before the antibiotic stock solution was added the LB agar was adjusted to 56°C. Petri dishes were cast in the laminar hood and stored at 4°C.

Nutrient agar (NA)

The nutrient agar was used for the determination of colony forming units.
10.0g nutrient broth
7g agar-agar
The chemicals were resuspended in 500ml RO-H_2O and autoclaved. Before the antibiotic stock solution was added the nutrient agar was adjusted to 56°C.

7.1.3 Liquid media

Luria Bertrani (LB) medium

1% NaCl
1% bacto-tryptone
0.5% yeast extract
The components were dissolved in RO-H_2O and autoclaved. Before an antibiotic stock

solution was added the medium had to be cooled. It was stored at 4°C.

M9-ZB medium

The components 1 to 4 were dissolved and autoclaved separately. After cooling down the solutions were combine. The medium was stored at 4°C.

Order	Component	Amount per litre
1	bacto-tryptone	10.0g
1	yeast-extract	5.0g
1	NaCl	5.0g
1	NH_4Cl	1.0g
2	KH_2PO_4	3.0g
2	Na_2HPO_4	6.0g
3	Glucose*H_2O	4.0g
4	$MgSO_4$ (1M)	1ml

Table 3: Components of M9-ZB medium

Semisynthetic medium for shake flask experiments

The components were dissolved separately according to the order given in table 4. The glucose solution was autoclaved separately and after cooling down added to the medium. The medium contains a phosphate buffer system. The phosphate salts (components 1) were added per litre of medium. The other components (2 to 9) were calculated according to the amount of biomass to be produced. For the correct folding of rhSOD, $CuCl_2 * 2H_2O$ and $ZnSO_4 * 7H_2O$ was added to the medium.

Order	Component	Amount per litre
1	KH_2PO_4	3.0g
1	$K_2HPO_4 * 3H_2O$	4.48g
2	bacto-tryptone	0.30g
2	yeast extract	0.15g
3	$Na_3 - citrate * 2H_2O$ (ACROS organics)	0.75g
3	$MgSO_4 * 7H_2O$	0.30g
4	$CaCl_2 * H_2O$	0.03g
5	Trace element stock solution	150μl
6	$CuCl_2 * 2H_2O$	12mg
7	$ZnSO_4 * 7H_2O$	9.6mg
8	$(NH_4)_2SO_4$	1.35g
8	NH_4Cl	1.11g
9	Glucose*H_2O	9.90g

Table 4: Components of semisynthetic medium for the production of 3g CDW/l

Semisynthetic and synthetic medium for fed-batch cultivations

The semisynthetic medium was used for batch cultivation in bioreactors; the synthetic medium was added during the fed-batch phase. The components were dissolved separately according to the order given in table 5. The glucose solution was autoclaved separately and after cooling down added to the medium aseptically. The medium contains a phosphate buffer system. The phosphate salts (components 1) were added per litre of medium. The other components were (2 to 9) were calculated according to the amount of biomass to be produced. For the correct folding of rhSOD, the medium contained $CuCl_2 * 2H_2O$ and $ZnSO_4 * 7H_2O$.

Order	Component	Semisynthetic (batch) medium	Synthetic (feed) medium
1	KH_2PO_4	3.0g	3.0g
1	$K_2HPO_4 * 3H_2O$	6.00g	6.00g
2	yeast extract	0.15g	-
3	$Na_3 - citrate * 2H_2O$ (ACROS organics)	0.25g	0.25g
3	$MgSO_4 * 7H_2O$	0.10g	0.10g
4	$CaCl_2 * H_2O$	0.02g	0.02g
5	$CuCl_2 * 2H_2O$	4.0g	4.0g
6	$ZnSO_4 * 7H_2O$	3.20g	3.20g
7	Trace element stock solution	50μl	50μl
8	Antifoam solution	0.5ml/kg	0.5ml/kg
9	Glucose*H_2O	3.00g	3.00g

Table 5: Components of semisynthetic medium (batch) and synthetic medium (feed) for fed-batch cultivations

7.2 Offline analysis

7.2.1 Protein quantification

Cell desintegration for NuPage gelelectrophoresis

Solutions	Concentration	Notes	Preparation
Tris/HCl (pH 8.2)	30mM	adjusted with 2N HCl	1.817g/500ml; 30ml 100mM Tris/HCl + 70ml H_2O
EDTA	0,5M	with KOH	25ml
$MgCl_2 * 6H_2O$	200mM	-	stock solution: 1M- (508,3mg/100ml)
Triton X-100	6%	in Tris/HCl (pH 8.2)	3g Triton + 47g Tris/HCl
β- mercaptoethanol	14.3M	-	-
Lysozyme	10mg/ml	aliquoted at 300μl	4mg/ml: 300μl + 450μl H_2O for IB-desintegration
Benzonase	50 units/ml	in benzonase buffer	-

Table 6: Stock solutions for cell desintegration

Contents	Volume
Tris/HCl (pH 8.2)	2.7ml
EDTA	150μl
$MgCl_2 * 6H_2O$	150μl
β- mercaptoethanol	6μl

Table 7: Cell desintegration buffer

Solutions	Concentration	Notes	Preparation
Tris/HCl (pH 8.2)	100mM	adjusted with 2N HCl	1.2144g/100ml
urea	8M	-	48.05g + 63.98g Tris/HCl
β- mercaptoethanol	100mM	added freshly	7μl/ml IB-solution buffer

Table 8: IB solution buffer

Cell desintegration for ELISA

Solutions	Preparation
Tris-buffer	20mM Tris-HCl, pH 8.2 (adjusted with 2N HCl); 1.2114g Tris/500ml
EDTA	0.5M EDTA (with KOH) pH 8.2
β- mercaptoethanol	14,3M
Lysozyme	10mg/ml
Triton X-100	0.5% Triton X-100 in Tris-buffer

Table 9: Stock solutions for cell desintegration of ELISA samples

Solution	Volume	final concentration
Tris-buffer	4.5ml	18mM
EDTA solution	250μl	5mM
β- mercaptoethanol	12.5μl	7mM

Table 10: Desintegration buffer for ELISA

GFP- ELISA- stock solutions and buffers

Solutions	Preparation
20mM Tris-HCl, $pH = 8.2$	242.28mg Tris base in 80 ml RO-H_2O adjusted to pH 8.2 with 2N HCl and filled up to 100ml.
0.5M EDTA, $pH = 8.0$	7.44g EDTA in 80ml RO-H_2O adjusted to pH 8.0 with 5M NaOH and filled up to 100ml.
10 mg/ml lysozyme	in RO-H_2O
0,5% Triton X-100	in 20mM Tris-HCl, $pH = 8.2$
Resuspension buffer	9ml of 20mM Tris-HCl ($pH = 8.2$), 500μl 0,5M EDTA ($pH = 8.2$) and 25μl mercaptoethanol freshly prepare each day
10 mg/ml OPD stock solution	stored at -20°C
Coating buffer (0,1N $NaHCO_3$-buffer, $pH = 9,6 - 9,8$)	8.4g $NaHCO_3$ and 4.2g Na_2CO_3 in 1l RO-H_2O. Buffer can be stored at 20°C up to one week.
Washing buffer (PBS, $pH = 7,2 - 7,4$)	1.15g $Na_2HPO_4 * 2H_2O$, 0,2g KH_2PO_4, 0.2g KCl and 8.0g NaCl in 1l RO-H_2O, 1ml Tween 20 is added cautiously to avoid foaming. The buffer may be stored at 4°C for 3 days.
Dilution buffer	1% bovine serum albumin in PBS, can be stored at 20°C for one day
Staining solution (Na_2HPO_4/ citric acid; $pH = 5.0$)	7.3g citric acid*H_2O and 11.86g Na_2HPO_4 in 1l RO-H_2O. The buffer can be stored at 4°C up to 4 weeks. A fresh solution of 10ml buffer combined with 100μl OPD stock solution and 3μl H_2O_2 is prepared for each analysis
Capture antibody (monoclonal mouse antibody anti GFP, Sigma)	diluted 1:1000 in coating buffer
Biotinylated secondary antibody (goat anti GFP- biotin, DPC-Biermann)	diluted 1:2000 in dilution buffer
recombinant GFP standard (1mg/ml)	diluted to 20ng/ml in dilution buffer
Avidin-HRP conjugate (Amersham Pharmacia)	diluted 1:1600 in dilution buffer
1.25 M H_2SO_4	35ml 95-97% H_2SO_4 made up to 1l with RO-H_2O

Table 11: Stock solutions and buffers for GFP- ELISA

hSOD- ELISA- stock solutions and buffers

Solutions	Preparation
20mM Tris-HCl, $pH = 8.2$	242.28mg Tris base in 80 ml RO-H_2O adjusted to pH 8.2 with 2N HCl and filled up to 100ml.
0.5M EDTA, $pH = 8.0$	7.44g EDTA in 80ml RO-H_2O adjusted to pH 8.0 with 5M NaOH and filled up to 100ml.
10 mg/ml lysozyme	in RO-H_2O
0.5% Triton X-100	in 20mM Tris-HCl, $pH = 8.2$
Resuspension buffer	9ml of 20mM Tris-HCl ($pH = 8.2$), 500μl 0.5M EDTA ($pH = 8.2$) and 25μl mercaptoethanol freshly prepare each day
10mg/ml OPD stock solution	stored at -20°C
Coating buffer (0,1N $NaHCO_3$-buffer, $pH = 9.6 - 9.8$)	8.4g $NaHCO_3$ and 4.2g Na_2CO_3 in 1l RO-H_2O. Buffer can be stored at 20°C up to one week.
Washing buffer (PBS, $pH = 7.2 - 7.4$)	1.15g $Na_2HPO_4 * 2H_2O$, 0.2g KH_2PO_4, 0.2g KCl and 8.0g NaCl in 1l RO-H_2O, 1ml Tween 20 is added cautiously to avoid foaming. The buffer may be stored at 4°C for 3 days.
Dilution buffer	1% bovine serum albumin in PBS, can be stored at 20°C for one day
PNPP (p-Nitrophylphosphat)(100mg/ml)	stock solution stored at -20°C
Staining solution ()coating buffer + PPNPP solution	A fresh solution of 10ml coating buffer combined with 100μl PNPP stock solution is prepared for each analysis.
Capture antibody (monoclonal mouse antibody IAM-SOD M05	diluted 1:300 in coating buffer
Second antibody enzyme-marked with alcaline phosphatase (Mouse mAb IAM-SOD 11H4x4AK)	diluted 1:1000 in dilution buffer
recombinant hSOD standard (2μg/ml)	diluted to 50ng/ml in dilution buffer

Table 12: Stock solutions and buffers for SOD- ELISA

7.3 Cloning methods- media and buffers

7.3.1 SOC medium

2% bacto-tryptone
0.5% yeast extract
10mM NaCl
3mM KCl
20mM $MgCl_2 * 6H_2O$
20mM Glucose*H_2O
10mM $MgSO_4 * 7H_2O$
The pH is adjusted to 7.0 with 5N NaOH.

Preparation: 20g tryptone, 5g yeast extract and 0.58g NaCl were dissolved in 450ml RO-H_2O. After addition of 0.22g KCl the mixture was stirred until the salt was dissolved. 2.03g $MgCl_2 * 6H_2O$ were added and the volume was made up to 975ml. Then the pH was adjusted to 7.0 with 5N NaOH. Finally the medium was autoclaved. An autoclaved 1M glucose stock solution (19,82g glucose*H_2O in 100ml water) and a sterile 2M $MgSO_4$ solution (49.3g $MgSO_4 * 7H_2O$ in 100ml water) were prepared, equilibrated to room temperature and pipetted into the nutrient medium, whereby 20ml glucose and 5ml $MgSO_4$-solution were added.

7.3.2 Buffers and stock solutions

50x TAE buffer

242g Tris base
57.1ml glacial acetic acid
100ml of 0.5M EDTA (pH 8.0)

Ethidium Bromide: Stock solution

10mg/ml, stored at 4°C in the dark

3M Na-acetate (pH 5.2)

40g Na-acetate*H_2O
The pH is adjusted to 5.2 with glacial acetic acid

0.5M EDTA (pH 8.0)

7.44g EDTA are dissolved in 80ml ROH_2O.
The pH is adjusted to 8.0 with 5M NaOH and filled up to 100ml.

HEPES buffer

1M stock solution: 23.83g HEPES in 100ml RO-H_2O
1mM HEPES buffer: 1:1000 dilution of the stock solution

7.3.3 Primer list

Name	Sequence
BglII-CAT-for	5'- GGA AGA TCT ACG GGG AGA GCC TGA GC- 3'
BglII-T7term-back	5'- GGA AGA TCT CTC GAG CAA AAA ACC CCT C- 3'
BsmBI-CAT-back	5'- AAT GAT GCG TCT CAC GAT GCC ATT GGG ATA TAT C-3'
BsmBI-CAT-for	5'- AAT GAT GCG TCT CCA TCG TAA AGA ACA TTT TGA GG- 3'
EcoRI-T7term-back	5'- AGT AGT GAA TTC CAA AAA ACC CCT CAA GAC C- 3'
GFP-SpeI-back	5'- GGA CTA GTC TTA TTT GTA TAG TTC ATC CAT GCC- 3'
KpnI-attTN7-back	5'- GTC GTC GGT ACC CAG AGG ACC CTC CTA AG -3'
KpnI-attTN7-for	5'- GTC GTC GGT ACC GTC TCC TGG GAG GAT TC -3'
SalI-CAT-for	5'- ACG CGT CGA CAC GGG GAG AGC CTG AGC -3'
SalI-T7term-back	5'- ACG CGT CGA CCT CGA GCA AAA AAC CCC TC -3'
SacI-TN7ho1-for	5'- GTC GTC GAG CTC GTT GCG ACG GTG GTA CG -3'
SpeI-rbs-GFP-for	5'- GGA CTA GTA ATA ATT TTG TTT AAC TTT AAG AAG GAG ATA TAC ATA TGG CTA GCA AAG GAG AAG- 3'
SpeI-rbs-SOD-for	5'- GGA CTA GTA ATA ATT TTG TTT AAC TTT AAG AAG GAG ATA TAC ATA TGG CGA CGA AGG CCG TGT G- 3'
SOD-SpeI-back	5'- GGA CTA GTT TAC TAT TGG GCG ATC CC- 3'
Tn7-ho1-for	5'- GTT GCG ACG GTG GTA CG- 3'
Tn7-ho2-back	5'- GGT ACC TGA AGA AGT TCG C- 3'
Tn7-ex1-for	5'- ACC GGC GCA GGG AAG G -3'
Tn7-ex2-back	5'- TGG CGC TAA TTG ATG CCG -3'
XhoI-TN7ho2-back	5'- GAC GAC CTC GAG GGT ACC TGA AGA AGT TCG C-3'

Table 13: Used primers

7.3.4 PCR set-ups

Contents	Volume (μl)
10x KOD-XL buffer	5
dNTPs	8
H_2O	33.7
primer 1(10pM)	1
primer 2(10pM)	1
DNA (2-20ng)	1
KOD-XL polymerase	0.3
reaction volume	50

Table 14: KOD-XL PCR Set-up

Temperature	Time	Cycles
94°C	1min	1x
94°C	30sec	30x
55°C	5sec	30x
72°C	3min15sec	30x
72°C	10min	1x

Table 15: KOD-XL PCR program

Contents	Volume (μl)
10x Thermopolymerase buffer	1,5
dNTPs	0.5
DMSO	0.5
H_2O	8.9
primer 1(10pM)	1
primer 2(10pM)	1
DNA (2-20ng)	0.5
KOD-XL polymerase	0.1
reaction volume	15

Table 16: Taq PCR Set-up

Temperature	Time	Cycles
94°C	2min	1x
94°C	15sec	30x
55°C	15sec	30x
72°C	3min15sec	30x
72°C	5min	1x

Table 17: Taq PCR program

7.3.5 Molecular weight markers

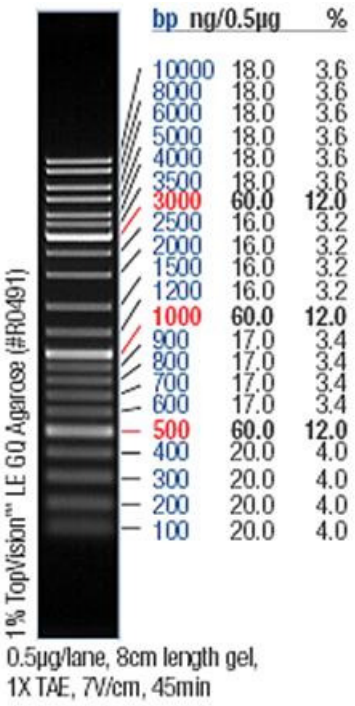

Figure 21: O' *GeneRuler*^{TM} 1 kb DNA Ladder, ready-to-use

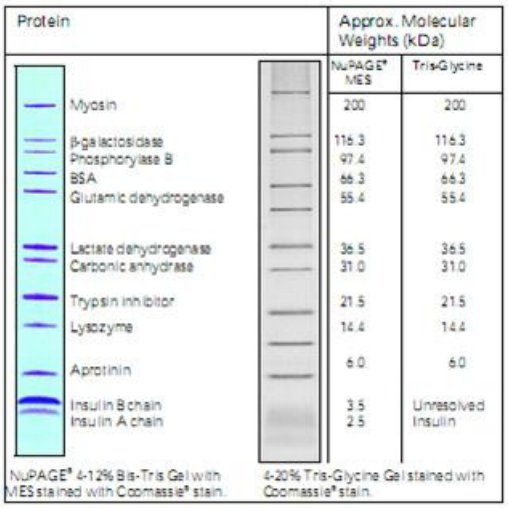

Figure 22: Mark 12^{TM} unstained standard

Abstract

Cultivation of microorganisms for the production of recombinant proteins is of central importance for many industrial fields. Many pharmaceutical applications are based nowadays on recombinant protein production. Easy and cost-efficient processes are in high demand, offering a great research area for process optimization and new process development.
E. coli production systems bearing a plasmid encoded heterologuous gene under a strong T7 expression promoter are widely distributed among industrial processes. Several phenomena encountered by T7 based plasmid systems, such as plasmid-loss during cultivation or break down of the system by metabolic overload, are unwelcome but accepted side-effects. Alternative expression systems are only sparsely available, giving plasmid based systems a predominant role in recombinant protein expression.
New concepts based on the integration of the target gene into the host chromosome offer features, which diminish system instabilities encountered by plasmid systems. Stable processes and high production yields plead for the establishment of chromosomal systems in industrial applications. Because of few research data and lack of experience with the production capacities of such chromosomal based systems, investigations on product formation abilities and limitations of hosts bearing chromosomal-integrated recombinant genes are of distinct importance.
Since recombinant gene expression by a single copy did not fully overstrain the host cell metabolism, two identical copies of the gene of interest should be inserted in this work. Comparisons of cultivation experiments between single and tandem integrated hosts, were performed to investigate the production capacity and limits of this system.
In addition the availability of genome integrated systems provides the prerequisites to investigate the metabolic load due to plasmid replication and plasmid encoded protein expression.

Kurzfassung

Rekombinante Protein Produktion ist ein zentrales Element der modernen Biotechnologie und findet Anwendung in den verschiedensten industriellen Sparten. Viele neue therapeutische Zulassungen in der Pharmaindustrie werden mittels rekombinanter Proteinproduktion hergestellt. Niedrige Produktionskosten und stabile Prozesse werden von Seiten der Industrie gewünscht, was die Forschung und Entwicklung im Bereich der Prozessoptimierung vorantreibt.
Plasmid codierte Expression von rekombinanten Genen unter starken Promotoren, wie dem T7 Promoter System, zählen zu den am häufigsten verwendeten Produktionssystemen in *E. coli*. Plasmidverlust oder Zusammenbruch des Produktionssystems durch Überbelastung des Wirtsmetabolismus sind nachteilige, aber akzeptierte Phänomene die durch das Fehlen von Alternativen zu erklären sind.
Neue Konzepte, die auf der Integration des Produktgens in das Wirtschromosom basieren, können die bei Plasmidsystemen auftretenden Instabilitäten und auch die mit Plasmidreplikation und Expression der Begleitproteine verbundene Stoffwechselbelastung reduzieren. Eine stabile Prozessführung und höhere Produktausbeuten als bei Plasmidsystemen sprechen für eine Etablierung der chromosomalen Systeme als industrielle Produktionsprozesse. Jedoch existieren noch wenig Forschungsdaten und Produktionserfahrungen. Die Erforschung der Produktionskapazitäten eines genomintegrierten Expressionssystems und die Festlegung der metabolischen Belastungsgrenzen sind von besonderem Interesse und Wichtigkeit.
Da die rekombinante Genexpression eines integrierten Gens den Wirtszellmetabolismus nicht vollständig auslastet, sollen im Zuge dieser Arbeit zwei idente Kopien des Produktgens ins Wirtschromosom integriert werden. Die Produktionsfähigkeit des Wirtsstamm sollte durch die Quantifizierung des produzierten Proteins ermittelt werden. Der Vergleich von Fermentationen des einfach genomintegrierten Stammes mit dem doppelt integrierten Stamm sollte mögliche Produktsteigerungen oder Prozesslimitierungen zeigen.
Weiters stellt die Verfügbarkeit von genomintegrierten Systemen die Voraussetzung für die Erforschung der metabolischen Belastung verursacht durch Plasmidreplikation und Plasmid codierte Proteinexpression dar.

Lebenslauf

Persönliche Daten

Name, Vorname	Schimon Marianne
Anschrift	Biragogasse 13/15; 3400 Klosterneuburg
email	marianne@schimon.biz
Telefonnummer	0699/ 10623142
Geburtsdatum	22.11.1984
Geburtsort	Klosterneuburg
Staatsbürgerschaft	Österreich
Familienstand	verheiratet seit 2007

Bildung, Werdegang

Volksschule	9/91 bis 7/95	Klosterneuburg
Bundesrealgymnasium	9/95 bis 7/03	Klosterneuburg
Universität Wien, Studium der Molekularen Biologie	9/03 bis 12/08	Wien
Europpass, Ferialpraktikum	7/03 und 9/03 - 7/04 und 9/04 - 7/05	Wien
VET MED Wien, Institut für Tierzucht und Genetik, Volontariatspraktikum	9/04 bis 1/05	Wien
Boehringer-Ingelheim Austria	9/05 ———— 2/06 und 3/06 7/06, 8/06, 9/06 2/07, 7/07, 9/07	Production of Hybridomas, Generation of Antibodies
Dept of Molecular Cell Biology; UNI Wien	4/07 und 5/07	Role of plactin in morphogenesis, signalling, disease
Dept. für Mikrobiologie und Immunbiologie; UNI Wien	11/07 und 12/07	B-raf knock out in mice
Boehringer- Ingelheim Austria	1/08 und 2/08	Identification of ligands of human G-protein coupled receptor 87
Dept. für Angewandte Mikrobiologie; BOKU Wien	4/08 bis 11/08	Diplomarbeit, Plasmid free T7 based *E. coli* expression systems