Design of substrate induced transcription for control of recombinant protein production in *Escherichia coli*

Maria Boström
M.Sc
Maria Boström (2004): Design of substrate induced transcription for control of recombinant protein production in *Escherichia coli*. The Swedish Centre for Bioprocess Technology, CBioPT, Department of Biotechnology, Royal Institute of Technology, KTH, Stockholm, Sweden

**ABSTRACT**

A suitable promoter for recombinant protein production should render several important characteristics. These characteristics were related to the two promoters in this work, the promoter of the lacUV5 operon, here called P<sub>lacUV5</sub> and the promoter of a maltose operon, here called, P<sub>malT</sub>. By the use of the catabolite repressed promoter P<sub>malT</sub> in glucose limited fed-batch production systems the growth and production phases could be separated. The activation of the promoter depends on the five MalT binding boxes and the four cAMP-CRP binding boxes. The affinity to these binding sites by MalT and cAMP-CRP, resulted in a possibility to modulate the transcription rate. The formation of cAMP, was important for the obtained production levels and was formed by two mechanisms, limited concentrations of glucose and at a switch between two carbon sources, here glucose to maltose. These two mechanisms were coupled to production techniques with the possibility to control the levels of cAMP, based on glucose/maltose limited fed-batch processes. Depending on the glucose feed a fed-batch process based on P<sub>malT</sub> resulted in a protein level of 6 to 10% of the total protein production, with β-galactosidase as reporter protein. The production rates obtained in these processes were controlled by the substrate feed rate, which controls the formation of cAMP and the formation of the endogenous inducer, maltotriose. Higher production rates were reached with glucose limited fed-batch systems with the addition of maltose but complicated the reproducibility of the process due to diauxic growth. However, the diauxic growth was controlled by the limited concentration of glucose and even avoided by the choice of a low limited glucose feed. A protein production of 30% of the total protein production was reached by the introduction of maltose and depended on the preceding substrate feed rate.

The hypothesis that the synthesis rate of the nascent polypeptide is an important factor controlling the inclusion body formation, proteolysis and translocation capacity was confirmed by the production of two model proteins Zb-MalE and the double mutant protein, Zb-MalE31, respectively. The synthesis rate was controlled by the promoters, P<sub>malT</sub> for production of Zb-MalE, and P<sub>lacUV5</sub> for production of Zb-MalE and Zb-MalE31. The proteolysis of Zb-MalE31 was decreased by a lower substrate feed rate compared to a high substrate feed rate in production systems based on P<sub>lacUV5</sub>. Also, the solubility of Zb-MalE31 was influenced by the substrate feed rate, where the highest solubility was obtained with a “low” feed rate. The levels of soluble Zb-MalE31 were lower compared to the soluble levels obtained with Zb-MalE and thus indicated the higher inclusion body formation with the double mutant protein. The inclusion body formation depended on the transcription rate, which was varied by different amounts of IPTG.

Soluble fractions of Zb-MalE were at all times accumulated in the cytoplasm during production based on P<sub>malT</sub> except during production at high substrate feed. This was due to a substantial degree of proteolysis as seen by the degradation pattern of the product. Zb-MalE was also subjected to inclusion body formation. Production in the periplasm resulted in accumulation of the full-length protein and furthermore this production system led to a cellular physiology where the stringent response could be avoided. Furthermore, the secretion could be used to abort the diauxic growth from production based on P<sub>malT</sub>. In production at high feed rate the accumulation of acetic acid, due to overflow metabolism, could be completely avoided.

**Key words:** *Escherichia coli*, substrate induced promoters, fed-batch processes, maltose operon, solubility, proteolysis, inclusion body formation
LIST OF PUBLICATIONS

This thesis is based on the following papers, which will be referred to by their Roman numerals:


INTRODUCTION

DEVELOPMENT OF BIOPROCESSES
The overall goal in bioprocesses is to reach high productivity, high yield and high quality, whereas the importance of each one of the parameter can vary between different kinds of processes. Since bioprocesses are dealing with living cells the challenge to optimise the yield and productivity, as well as increasing the product quality are very complex. Process techniques have to be developed to match the metabolic capacities of the living cells to reach these goals.

The development of the fed-batch technique in the end of 1800/ beginning of 1900 resulted in the possibility to reach higher cell densities and thus increase the productivity. The technology was used for the production of yeast and introduced with two patents (Hayduck 1919; Saks 1919). In yeast fermentation this technique increased the cell mass without the drawback of high ethanol production. The well-known process of Baker’s yeast is composed of living cells of Saccharomyces species. Baker’s yeast is a low-price product and the most important parameter for this process is the optimisation of the biomass yield coefficient. Two parameters are essential to control this process, strong aeration to suppress high ethanol formation and a strict control of the glucose concentration. The fed-batch technique enables the concentration of sugar to be kept at a rate limiting value, and thus controlling the specific growth rate to have a constant value lower than \( \mu_{\text{max}} \).

The production of antibiotics is formed through the secondary metabolism of moulds, i.e. the formation starts after the cell has entered the stationary phase. Most antibiotic production is thus catabolite repressed, which means that the carbon source is needed to be added in a limited concentration to induce the second metabolism. Furthermore, the production with moulds often results in limitations in oxygen transfer and cooling capacities, due to the high density and high viscosity that are obtained. Therefore the fed-batch technique is the option for a successful production technique where the above mentioned parameters can be controlled. Penicillin, as first discovered by Fleming 1929, is produced by the fed-batch technique. Other large production processes that are using the fed-batch technique are the production of citric acid and acetic acid.
The developments in genetics and microbiology lead to the improvement of bioprocesses in many different ways. The first example of an industrial use of the genetic engineering was the introduction of foreign DNA into a micro-organism. It became possible to produce larger amounts of proteins that previously only could be extracted from natural sources. Various production techniques were evaluated for the production of recombinant proteins. High cell density to reach high productivity was difficult to obtain with a traditional batch technique due to cooling and oxygen transfer limitations and resulted in the use of the fed-batch technique. In production of recombinant proteins that are growth rate dependent it is possible to increase the production yield with the fed-batch technique.

**Protein Products**

Biotechnological protein products are divided into at least three categories of industrially produced products (Sawers and Jarsch 1996). The first category includes technical enzymes, e.g. for detergents, and for food processing, which are characterised by a high quantity and low cost production. The other category involves the production of high quality proteins such as protein in diagnostics and human therapeutics, which are characterised by a high cost and a high quality production for example, insulin and erythropoietin.

![Figure 1. The relation between price and volume, for the production of biotechnical products (Schmid 2003).](image)

1. Erythropoietin, a growth factor produced in recombinant mammalian cells and used in chemotherapy
2. Insulin, a hormone that regulates the glucose levels in the blood, produced in recombinant bacteria and applied to treat diabetes
3. Tetracycline, an antibiotic, used in medicine and agriculture, produced in bacteria as well as yeast
4. Cephalosporins, antibiotics, see 3.
5. Aspartame, food additive, produced by chemical synthesis
6. Detergent protease, enzymes produced by recombinant bacteria
7. Citric acid, is used as a preservative in food and produced by fermentation of bacteria and glutamic acid is used as flavour enhancer in the food industry and produced by fermentation of bacteria
8. Ethanol is an industrial solvent and energy source and produced by the fermentation of yeast

Yield sensitive products require cheap unit operations, where the optimum is to obtain all substrate as product and are characterized by low cost high volume products. Quality sensitive products, on
the other hand, require a highly controlled process to optimise the product quality and are characterised by high cost products. Some of these industrially produced bio-products are exemplified in figure 1, where the relation between price and annual volume are shown.

**PROTEIN PRODUCTION TECHNIQUES**

Production of pharmaceuticals are divided into, pharma- and biopharmaceuticals, where the latter corresponds to 6% of the total production of pharmaceuticals. The development for the production of pharma/biopharma products, respectively, is different with respect to process and product goals. In pharma production, shown in the upper part of figure 2, multi variants of proteins in small scale are rapidly produced to serve as targets for new developed small molecules, drug leads. The requirements of the rapidly produced small scale proteins will differ in respect to proteins produced in large scale, as shown in the lower part of figure 2, where the goal is to develop a process for a single protein product in high amounts.

Figure 2. Two different goals to achieve in protein production. Upper part: Rapid production in small scale of proteins. Lower part: The research and developing of processes resulting in production of a large scale protein product as seen in the lower part of the figure, (adapted with the courtesy of professor P-A Nygren, Department of Biotechnology, KTH, Stockholm, Sweden).
The goal in small multiscale protein production is to obtain soluble protein products in high enough amounts for structural and functional analyses with the purpose of high throughput screening, HTS. The development of suitable production processes is therefore a very important part in the design of pharma production systems.

Since all proteins will have specific properties, the production of sufficient amounts of each one in an active and functional form will require a multiple set up of different production conditions. This multiple set up can be realised in a high throughput protein production system, HTPP. Important parameters in a HTPP system based on E.coli are described in a review (Stevens 2000). The system will include parallel production of proteins, expressed under a range of conditions according to the process environment and cellular/genetic properties. This will result in a number of successful as well as many unsuccessful process conditions. All this data will be analysed for a specific protein and the functional and nonfunctional expressing systems, respectively, will be known. This will increase the knowledge for further research of finding successful expression conditions. Naturally, this base of knowledge will increase with the number of tried expression conditions and will thus give the system a more efficient way to be used. Only a very specific and small part of these categories of proteins can be expected to be produced by already developed production systems.

The fed-batch process is a technique to reach high productivity, high yield and high quality. Two other modes of cultivation are batch, and continuous cultivation, which are described elsewhere (Enfors and Hägström 2000). The three modes of operations are shown in figure 3.

![Figure 3. Three operation modes used for cultivations.](image)

**Fed-batch process**

One of the main reasons to apply the fed-batch technique is the ability to obtain high cell densities, and thus reach higher productivity. The technique can in general be described as a batch process continuously fed with a nutrient limiting the cell growth. The substrate feed is often a sugar, but can
also be phosphate or nitrogen, with highest possible concentration to avoid the increase of culture volume. The limiting substrate allows a growth rate below $\mu_{\text{max}}$ and thereby the limiting parameters set by the reactor, oxygen and heat transfer, can be controlled. Suitable feed profiles of the substrate can be calculated from mass balance equations (Enfors and Häggström 2000; Yee and Blanch 1992). A fed-batch process can start with a batch phase to accumulate cell mass before the introduction of a constant feed of the growth limiting substrate, figure 4. The batch

![Graph](https://via.placeholder.com/150)

**Figure 4.** Biomass (closed circles) and specific growth rate, $\mu$ (open triangles in a fed-batch process in *E.coli*).

phase is most often abolished due to the formation of by-products. When applying a constant feed it will result in a linear increase of the biomass while the specific growth rate is declining, thus maintaining the cell mass productivity $X\cdot\mu$ approximately constant. If it is desired to control the cell growth early in a process that has not reached a high enough cell density, it might result in a very low constant feed to match the consumption rate at that time, which in turn leads to a low productivity. Thus, a further development of the fed-batch process is therefore to start with an exponentially increasing feed profile, which keeps the specific growth rate constant. An exponential feed profile will eventually lead to oxygen limitations due to the increase in cell densities, and the exponential feed rate has to switch into a constant feed rate. This process, figure 5, is widely used to obtain high productivity as in the Bakers yeast process. This technique is also commonly used to avoid acetate formation in *E.coli*. The linear increase of the biomass is proportional to the constant feed rate. If the feed into the reactor only contains the limiting component, a sugar, the culture will
eventually be limited on other medium components. To avoid this, these components could also be added into the reactor parallel to the substrate feed.

![Graph showing a fed-batch process with an initial batch phase, exponential growth at a constant growth rate, followed by a constant feed rate. The biomass (closed circles) accumulation and the specific growth rate (open triangles) are here shown.](image)

*Figure 5. A fed-batch process with an initial batch phase, an exponential growth at a specific growth rate followed by a constant feed rate. The biomass (closed circles) accumulation and the specific growth rate (open triangles) are here shown.*

Although the medium components do not become limiting, the specific growth rate and the specific energy substrate uptake rate decline during the constant substrate feed phase. Eventually, most of the substrate source is used for maintenance. Thus, there is a maximum cell quantity possible to obtain in a fed-batch culture if the maintenance reaction has priority to the available energy substrate as described elsewhere (Enfors and Häggström 2000). The scale-up in a fed-batch process is complicated due to the effect of concentration gradients created by the inlet feed. Since the concentration of the substrate inlet feed often is very high, it results in a delay to get it homogeneously distributed. The quasi-steady state concentration of limiting substrate depends not only of the feed rate but also of the saturation constant of the organism. Since the saturation constant, $K_s$, is in milligrams per litre there is obviously a significant difference between the inlet feed point and the desired substrate concentration in the reactor. This will lead to substrate gradients and results in lower cell yield and by product formation (Bylund et al. 1998).
PARAMETERS CONTROLLING THE PRODUCTIVITY AND QUALITY

In fed-batch processes the cell mass productivity \( \text{X} \times \mu \) is to be maintained to control limiting parameters such as oxygen and cooling capacities. Since the cell mass is expected to increase the specific growth rate \( \text{as} \) to decrease from this point on, meaning that the higher the cell mass the lower the specific growth rate. The production is eventually induced and promoters used are often based on the traditional production system where the first step is to optimise the cell mass and the second step is to induce the cells. However, production systems with strong promoters such as T7 and tac are only productive for short periods of time, 1-2 hours, whereafter the production ceases, (Dedhia N et al. 1997; Striedner G et al. 2002; Lin HY et al. 2001). This short production time in combination with induction at a low specific growth rate might make it difficult to reach high productivity. The longer the process is run the lower the specific growth rate. Many processes are run to specific growth rates of 0.05 \( \text{h}^{-1} \) or even below, before induction. This will result in an approach to starvation conditions as well as induction of carbon-induced stress, controlled by the alternative sigma factors (Hengge-Aronis 1996). Furthermore, at such low growth rates most carbon and energy has to be used for maintenance requirements. Production in this range would thus not in many cases be advised since the cells have severe problems in remaining viable. The control of the specific growth rate is thus crucial in the design of appropriate production protocols to reach high productivity. The total productivity of a process is defined by:

\[
Q_p = q_p \times X \times V \text{ [g/h],}
\]

where \( q_p \) is the specific production rate of the product (g product/g dry cells, h), X is the cell mass concentration (g/l) and V is the working volume (l). As seen from the equation, an increase in all three parameters separately, will lead to an increased total productivity. The production rate and the cell mass are both a function of the microenvironment in the bioreactor and a function of parameters in the “cell factory”, which here will be further described.

\( q_p \) as a function of the microenvironment in the bioreactor

The microenvironment in the bioreactor involves the culture medium, aeration, mixing as well as the addition of components into the bioreactor during cultivation, as substrate, pH-controlling agents and antifoam. To be able to control the specific production rate, \( q_p \), it is obviously very important to also control the specific growth rate. The control of the specific growth rate is possible by the introduction of different feed profiles that limits the cell growth. The feed profile effect on recombinant E.coli production have been studied with several different recombinant products as the alpha consensus interferon (Curless et al. 1990), malaria antigen (Zabriskie et al. 1987), chloramphenicol acetyl transferase, CAT (Ramirez and Bentley 1995) and human growth hormone/
human glucagons (Shin et al. 1998). In these studies the feed profile has a large impact on the product yield and the product quality according to inclusion body formation except for the production of malaria antigen (Zabriskie et al. 1987). It has furthermore been shown that a high productivity was inversely proportional to a high product quality, (Ryan et al 1996) and that the activity of the enzyme, β-galactosidase depends on the growth rate and reaches a maximal activity at a specific growth rate (Hellmuth et al 1994). Thus the control of the growth rate by different feed profiles are preferable to obtain optimal conditions for an acceptable quality and productivity.

**q_p as a function of parameters in the cell factory**

Important cellular parameters that influence the production rate are e.g. the host strain, expression vector, the efficiency and stability of the transcription and translation, codon usage, proteolytic stability, folding, localisation and characteristics of the individual protein. The “cell factory” with the different steps for protein production is shown in figure 6.

![Figure 6. The cell factory. Different mechanisms in the cell that influence the production rate of a protein followed by the purification. The substrate, S enters the cell and induces the promoter, P, and allow transcription, Tr, to start. The binding to the ribosomal binding site including the Shine Dalgarno, (S.D) results in the translation, Tn, and production of the nascent polypeptide. With the help from chaperons the polypeptide folds, F, and forms the product protein. The product is transported, Tp, to the medium, the periplasm or remains in the cytoplasm (adapted with the courtesy of Per-Åke Nygren, Department of Biotechnology, KTH, Stockholm, Sweden).](image)

**Hosts for production**

The choice of the host system depends on many factors such as the size, structure and stability of the gene product, and the requirements for post-translational modifications for biological activity. The necessary productivity, yield and quality as well as acceptable cost of the final product also have to be considered. The gram negative bacterium *E.coli* is the outmost well characterised host for production both in its genetics and metabolism and will here be further focused on. It is easy to grow to high cell densities and a number of expression vectors and mutant host strains are available
from which a wide variety of proteins have been successfully expressed (Baneyx 1999; Hockney 1994; Makrides 1996; Lee 1996).

The expression vector
The stability of the expression vector depends on the copy number, the structure of the vector, the specific growth rate, the strength of induction, growth temperature, the host cell and the culture medium.

There are two types of stability of the expression vector, segregational stability (plasmid loss resulting from defective partitioning during cell division), and structural stability (undesired plasmid modifications resulting from insertion, deletion, or rearrangement of DNA). The plasmid copy number hereafter called PCN, will influence the maintenance of the plasmid within the cells during replication and is determined by the origin of replication, “ori”, of the vector. There are several commercial vectors with low, 5-10, (pACYC), medium, 15-50, (pBR322), and high, about 100, (pUC19, pET) FCN. In terms of production yields it appears to be no advantage to use higher than medium PCN (Yansura and Henner 1990). If plasmid-borne genes are toxic it often reduces the cell growth, allowing cells with no or low PCN to dominate the cell culture (Friehs and Reardon 1993). Therefore, a number of alternative strategies have been developed to ensure that plasmid free cells will not overtake a culture (Baneyx 1999). The most common way to avoid plasmid free cells is to introduce a selective antibiotic resistance into the plasmid and supplement the antibiotic to the medium, and thereby kill the plasmid free cells. Ampicillin is commonly used for this purpose, but for human therapeutics production it is avoided since it can cause human allergic reactions. Furthermore, during cultivation the antibiotics can loose the selective pressure due to degradation and/or inactivation of the antibiotics. The transcription from strong promoters into the replication origin can destabilise plasmids owing to overproduction of the replication protein, REP, involved in the control of PCN and result in excessive plasmid replication. This leads to a high metabolic burden for the host cell that mainly derives from the protein synthesis itself. The PCN can be stabilised through sequence modification of the RNAs involved in the control of replication (Grabher et al. 2002).

The structure of an expression vector in *E. coli* requires several elements whose configuration must be carefully considered to ensure the desired levels of protein synthesis (Makrides 1996). The essential parts in an *E. coli* expression vector are shown in figure 7. The promoter is regulated by a
regulatory gene, which can be situated on the vector itself or on the chromosome. The promoter is commonly situated about 10 to 100 base pairs upstream from the ribosomal binding site. The promoter consists of a hexanucleotidic sequence about 35 base pairs upstream the transcriptional starting point (-35 region) separated by a short spacer from another hexanucleotidic sequence (-10 region). Additionally the vector often contains an antibiotic resistance gene and the origin of replication, ori.

**Parameters determining the final product quality**

**Folding and secretion**

It has been postulated that a high accumulation rate of the product protein leads to a large pressure on the posttranslational steps which are rate limiting. The folding intermediate, which is a loosely folded complex rapidly created from the nascent polypeptide (Hendrik and Hartl 1993), and subsequent steps is shown in figure 8. The subsequent steps leading to a correctly folded protein in

**Figure 8. Different pathways of the nascent polypeptide in the cell (adapted from Enfors S-O and Hägström L 2000).**

the cytoplasm or the periplasm are relatively slow processes due to the function of chaperones and folding catalysts as well as transporters of the secretory system. The unfolded protein is very sensitive to proteolysis and when produced in high amount the risk for aggregation to other unfolded proteins is increased which result in incusion bodies. Furthermore, if the amount of the foldases and/or chaperons are not sufficient to match the formation rate of partly folded protein, or
if they fail to recognise the heterologous protein, it will lead to a longer exposure time of the unfolded protein which, in turn, could result in misfolded protein and/or proteolysis. Methods to obtain properly folded protein from inclusion bodies require control of aggregation and might be successful (Lilie et al. 1998; de Bernardes et al. 1999). To find the optimal system for refolding is still relatively complex and the best results are found when a matrix of conditions affecting solubility and disulphide-bond formation and isomerisation are applied. The use of either soluble or immobilised folding aids can improve the results. It is still desirable, however, to obtain soluble, bioactive proteins, without the requirement of refolding in HTPP systems as earlier discussed.

To have a correctly folded in vivo protein, chaperons and foldases may play an important role. The molecular chaperons aid proper isomerisation and cellular targeting by transiently interacting with folding intermediates, in which the best characterised systems are GroEL/GroES and DnaK/DnaJ/GrpE. The E.coli cytoplasm is maintained in a reduced state that strongly disfavours the formation of stable disulfide bonds in proteins, therefore proteins that contain disulphide bonds are transported to the periplasm where the isomerisation of disulphide bonds can take place. However, there are some examples where the cytoplasm of mutant strains is sufficiently oxidising to enable cytoplasmic production of tPA and full-length vtPA (Bessette et al. 1999), and Fab antibody fragment (Levy et al. 2001). The disulphide bond formation requires the participation of two soluble periplasmic enzymes, DsbA and DsbC and two membrane bound enzymes DsbB and DsbD. Disulfide-bond containing proteins like the human growth hormone and a large number of single-chain Fv antibody fragments as well as full-lengths antibodies have been secreted and folded correctly in the periplasm (Simmons et al. 2002). The folding environment can be modified for example by the overexpression of DsbC which has shown to facilitate the folding of t-PA with 17 disulfide bonds in the periplasm (Qiu et al. 1998). In the case of the insulin-like growth factor (IGF-I) with three disulfide bonds the overexpression of DsbA and DsbC actually decreased the yield of folded product, but increased the accumulation of inclusion bodies (Joly et al. 1998). Overexpression of other helper proteins such as Skp and the periplasmic FkpA can help in the folding of secreted single-chain Fvs (Bothman and Pluckthun 2000).

However, the optimal environment for disulfide bonding and avoidance of proteases would be in the culture medium. The secretion of some recombinant proteins to the periplasm causes a destabilisation of the outer membrane, not fully understood (Sandkvist and Bagdasarian 1996), which becomes leaky and allows the protein to diffuse into the extracellular medium. Strategies of
potential transport systems for the introduction of proteins to the culture medium have been described in a review (Shokri et al. 2003) and are based on two mechanisms. The first is based on the active transport through the cytoplasmic membrane followed by the passive transport through the outer membrane by the interaction of external or internal destabilisation components. The second is based on the active transport over one or both of the membranes by the introduction of secretion mechanisms of pathogenic E.coli and other species.

**Proteolysis**
The proteolysis is a selective and highly regulated process that is involved in a variety of metabolic activities, such as the removal of abnormal and incorrectly folded proteins. E.coli contains a large number of proteases that are localised in the cytoplasm, in the periplasm and the inner and outer membranes (Maurizi 1992). The mechanisms for proteolytic degradation are not clearly understood but some rules according to the protein stability have been evaluated, such as the N-erd rule (Tobias et al. 1991). Strategies for minimising proteolysis as well as their limitations have been reviewed in detail (Makrides 1996; Murby et al. 1996). These strategies include targeting proteins to the periplasm or the culture medium, where the number of proteases are significantly less, the use of protease deficient host strains, growing the cells at a lowered temperature, construction of N-and C-terminal fusion proteins, fusing multiple tandem copies of the target gene, co-expressing molecular chaperons, replacing specific amino acid residues in order to eliminate protease cleavage sites, modifying the hydrophobicity of the target protein and optimising the cultivation conditions.

The accumulation of heterologous proteins in E.coli is sometimes relatively low due to a high degradation. This might be improved by the genetic fusion to a stable host protein. This combined fusion protein will be protected from degradation. The fused protein is normally situated in the N-or C-terminal of the foreign gene product. The first fusion protein system specifically aimed at increasing the solubility of a target protein was a thioredoxin, TRX, fused to the N-terminus of the target protein (La Vallie et al. 1993). Popular fusion tags are domains that include glutathione-S-transferase, GST (Smith and Johnson 1988), maltose binding protein, MalE, (Bedoule and Duplay 1988; di Guan et al. 1988), thioredoxin (La Vallie et al. 1993), and the Z-domain from protein A (Nilsson et al. 1987). MalE and GST were chosen as carriers because they enable the fusion protein to be affinity purified: MalE binds to amylose and GST binds to immobilised glutathione. These systems have resulted in successful overexpression of many heterologous proteins in E.coli but most often they do not posses maximal solubilisation characteristics. In a recent study 32 potentially interesting human proteins were selected with unknown structures to be expressed with six different
N-terminal fusion proteins and an N-terminal His tag (Hammarström et al. 2002). 85 % of these proteins were expressed in a soluble form. The methodology used and the expression vectors are suggested to be used for screening of genes for structural studies (Hammarström et al. 2002). In table 1 the advantages and disadvantages are summarised for the production of intracellular as well as secreted recombinant proteins.

<table>
<thead>
<tr>
<th>Product localisation</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secretion to the culture medium</td>
<td>Disulfide formation possible, Less protein degradation, Purification process simplified (no cell disruption needed), Improved folding</td>
<td>Transport to the medium not always possible, Protein diluted</td>
</tr>
<tr>
<td>Transport to the periplasm</td>
<td>Improved disulfide formation, Improved folding, Less protein degradation</td>
<td>Transport to the periplasm not always possible, No large scale procedure for selective release of periplasmic proteins available</td>
</tr>
<tr>
<td>Intracellular production as inclusion bodies</td>
<td>High protein yield, Inclusion bodies easy to isolate, Protection from proteases, Protein is inactive and can not harm host</td>
<td>Solubilisation and in vitro refolding necessary which usually give lower yields and higher costs, Refolding not always possible</td>
</tr>
<tr>
<td>Intracellular production as soluble product</td>
<td>No need for solubilisation and refolding, High protein yield</td>
<td>High levels of product can be harmful to the host, complex purification, proteolysis, disulfide formation usually not possible</td>
</tr>
</tbody>
</table>

Table 1. The advantages and disadvantages of different localisation of recombinant proteins.

**TRANSCRIPTIONAL CONTROL OF SUBSTRATE INDUCED PROMOTERS**

The control of promoters is performed by specific systems for transcription, which will further be described in the next chapter in connection to specific operons. Except for this mechanistic description transcription is furthermore influenced by environmental conditions requiring a specific cell metabolism. Important components directing the response to the environmental impact are for example: cAMP which will bind directly to the promoter and alternative σ factors (Hengge-Aronis 1996), binding to RNA polymerase, leading to enhanced transcription of the specific operons, and the stringent respons alarmone guanosine-3,5-tetr phosphate, ppGpp, by binding to RNA polymerase, resulting in transcription arrest. However, ppGpp also inhibit translation through inhibition of the tRNA synthesis leading to ribosome stalling.

Most of the promoters used for industrial recombinant protein production are characterised by a high product accumulation in a short period of time. Certain characteristics for promoters allowing high-level protein synthesis have been described (Goldstein and Doi 1995; Keasling 1999). A high production rate of the recombinant protein can also create the stringent response in the cell (Sandén et al. 2003; Cserjan-Puschman et al. 1999) due to an overload of the metabolic capacity of the host
cell and is created either by carbon starvation or a lack of charged tRNA (Chasel et al. 1996). This can result in a rapid increase of ppGpp.

The cellular routes for ppGpp metabolism are shown in figure 9. Synthesis of (p)pppGpp occurs in two ways. The first is when a ribosome is stalled due to insufficient amount of charged tRNA which result in a binding of uncharged tRNA to the ribosome. This leads to the activation of the ribosome activated protein, RelA which catalyses the production of (p)ppGpp. The second way to synthesise (p)pppGpp occurs during depletion of primary carbon and energy source and is independent of the relA gene product. Little is known about this second pathway but it includes a protein, SpoT, that catalyses as well as degrades (p)ppGpp and it is suggested to perform in two different conformations.

![Figure 9. The cellular routes for (p)pppGpp metabolism (Cashel et al. 1996).](image)

**Promoters, structure and function**

The structure of the promoter is important for the control of induction at different levels as will be exemplified in this text by different types of promoters. In general the regulation for transcription initiation can be either negative or positive, and the principles are illustrated in figure 10. In positive regulation a specific activator protein must be present either to permit RNA polymerase to initiate transcription or to increase the frequency of transcription initiation. The activity is controlled by a change in the metabolic status or the addition of a specific inducer molecule. The negative regulation means that a repressor, usually a protein, prevents the transcription by the binding to the operator region on the promoter. The derepression of the promoter can be controlled in two ways, either by addition of an inducer or removal of a co-repressor. The inducer binds to the repressor, which changes conformation and therefore no longer has the affinity to the promoter and thus liberates transcription. The tryptophan promoter is regulated by the presence or the depletion of
tryptophan, the corepressor. When tryptophan is present it prevents transcription by the binding of the repressor and when it is depleted it cancels the repression and thereby allows transcription.

**Figure 10.** The positive and negative regulation of promoters in E.coli. 1. Positive regulation, where RNA polymerase, RNAp, can interact with the promoter but can not form an open complex because the activator is in its inactivated form (A0). When the activator is exposed to a metabolic signal for instance, changes in oxygen, nitrogen, phosphorus or substrate concentration the activator changes the conformation and becomes active (A1). The activator can then bind to the operator, O, on the promoter, and allow transcription to start. 2. Negative regulation, controlled by induction. The inducer, I, binds to the repressor, R. The promoter gets derepressed and transcription can start. 3. Negative regulation controlled by derepression. The repressor, R, must interact with a small corepressor, C, to have affinity to the operator. Metabolism of the co-repressor activates the repressor and liberates the promoter (modified version Savers, 1996).

**The lactose operon**

The lactose operon is the most well understood of all bacterial promoters and has been extensively characterised at the molecular level (Reznikoff and Abelson 1980). The lactose operon, figure 11, encodes the genes necessary to express the enzymes involved in lactose degradation. The lacZ gene encodes β-galactosidase, which hydrolyses the α-1,4 bond between galactose and glucose in lactose, the lacY gene encodes the lactose permease, which spans the cell membrane to transport lactose or IPTG, isopropyl β-D-thiogalactoside, into the cell and the lacA gene encodes the thiogalactoside transacetylase, an enzyme with a detoxifying function (Andrews and Lin 1976). These genes are preceded by a region which is responsible for the negative regulation of the lactose metabolic genes, which encodes the gene lacI, expressing the repressor molecule, LacI. The gene encoding LacI can be chromosomal, inserted upstream the lac promoter, or on a separate plasmid either as lacI or the mutant lacI' (produces tenfold more repressor than lacI). LacI binds to the operator region of the operon to inhibit RNA polymerase to bind and initiate transcription. The lac
promoter is induced by lactose or the lactose analog, IPTG. Lactose is converted to the internal inducer allolactose in the cell, which binds to LacI and causes a conformational change so that LacI no longer can bind to the operator site. IPTG binds to LacI in a similar way as allolactose.

The lactose operon is also subject to positive control by the formation of the complex, cAMP-CRP, cyclic adenosine monophosphate-catabolite repressor protein. The phenomenon, where the complex cAMP-CRP is formed is called catabolite repression. Other examples of operons under the control of this global control mechanism are the arabinose and the maltose operons. As long as glucose is present it will be the preferred substrate for growth and no other sugar will be needed, thus there is no need to activate other operons. This phenomenon is called inducer exclusion. If there is an excess of glucose in the growth medium it will be transported into the cell by the action of the glucose transport system, the phosphoenolpyruvate-phosphotransferase system, PTS, and phosphorylated by the phosphate group donated by the transport system, figure 12. The PTS

![Diagram of lactose operon](image_url)

**Figure 11.** The lactose operon.

**Figure 12.** The phosphoenolpyruvate-phosphotransferase system, PTS (www.biologie.uni-hamburg.de/b-online/chimes/kanal/transport/pts/pts.htm).

transport system allows not only glucose transport but also transport of fructose, mannose, galactitol, mannitol, sorbitol, xylitol and N-acetylglucosamine. The phosphate group is transferred from phosphoenol pyruvate, PEP, through a series of intermediary proteins (Meadow et al. 1990; Postma et al. 1993). Some of the proteins are specific for an individual PTS sugar transport system.
and some are common in all PTS. One specific enzyme for the glucose PTS is the enzyme II, EI.
This enzyme contains three domains A, B and C. The domain C is the membrane channel, B is the
phosphorylation centre and A is the intermediate to donate the phosphate group. At low glucose
concentration in the growth medium or no glucose the phosphate group from EIIA is free to activate
adenylate cyclase, which stimulates the cAMP formation from adenosine triphosphate, ATP, figure
13, thus releases the catabolite repression.

![Diagram](image)

*Figure 13. Catabolite repression. Low levels of glucose result in a dephosphorylated form of EIIA (EI), which thereby
can activate the enzyme, adenylate cyclase to convert ATP to cAMP.*

As a result of the negative and positive regulations of the *lac* operon, there will be four basic states
of expression as illustrated in figure 14. With low glucose and lactose the adenylate cyclase will be
stimulated to cAMP formation. CRP will bind at its binding site together with cAMP. It will
promote the RNA polymerase to bind to the operator which leads to low transcription, since there
will be low amounts of lactose releasing the repressor LacI from the binding site. At high glucose
concentration the operon is catabolite repressed, i.e. low cAMP concentrations and no CRP bound
to the promoter, and the repressor LacI will inhibit the transcription of the operon. At high
concentrations of both glucose and lactose, the LacI will not be bound to the operator but the operon
will still be catabolite repressed. As long as glucose is present there is no need to metabolise lactose
because of the inducer exclusion. To reach the highest level of transcription, conditions of low
glucose and high lactose should be maintained. The operon will then be released from the catabolite
repression and the LacI repressor will be bound to lactose, thus inhibiting the binding to the
operator. RNA polymerase will be efficiently bound to the operator to start transcription.

A number of variants of the *lac* promoter have been reported. The *lacUV5* promoter is by the
mutation in the consensus region of the *lac* promoter not subjected to catabolite repression and
therefore not depending on the activation with cAMP. This has resulted in an increase in the
Design of substrate induced transcription for control of recombinant protein production

promoter strength (Reznikoff and Abelson 1980) and production systems with this promoter have reached levels as high as 30 % of the total protein production (Sandén et al. 2003). Two other derivatives of the lac promoter are the tac and trc promoters, which also allow expression levels of 15-30 % of the total protein production. The promoters are hybrids of the lac promoter, which both consist of the -35 region of the trp promoter and the -10 region of the lac promoter with the change in one base pair.

![Diagram showing the basic states of expression from the consequences of negative and/or positive regulation.]

IPTG is used as inducer in all these derivates and offers some advantages, especially in small scale bioreactors. Unlike, lactose, it is not metabolised by the cell, which ensures that the level of induction remains constant following the addition of IPTG to the growth medium. However, IPTG is not obviously the first choice due to its toxicity and expensiveness. Furthermore, it is shown that the addition of IPTG influences the specific growth rate of E.coli and induces the synthesis of heatshock proteins, DnaK, GroEL, GroES, (Kosinski et al. 1992). It is also very difficult to reproducibly control the induction strength since very small variations in IPTG leads to large variations in production level (Jensen Ruhdal and Hammer 1997). The problem to reproducibly control the induction and the fact that the lac promoter is relatively strong may cause an overoad of the metabolic capacity due to the uncontrolled induction. It has been shown that an overload of the metabolic capacity do not lead to maximal expression of a target protein (Glick 1995) and a metabolic overload causes a stress responses in the cell like the induction of the alamone, guanosine tetraphosphate, ppGpp (Sandén et al. 2003; Cserjan-Puschman et al. 1999). A study with
the tac promoter and chloramphenicol-acetyl-transferase, CAT, as reporter protein showed a severe reduction in growth rate caused by the increased levels of CAT (Bentley et al. 1991). However, the dependence of large variations in induction with IPTG changed qualitatively when a lacY mutation was introduced to the cells (Jensen Ruhdal et al. 1993; Jensen Ruhdal and Hammer 1997). In these lacY mutant cells, IPTG diffuses into the cell and allow for tuning the expression of the genes (Jensen Ruhdal and Hammer 1997). Furthermore, the most serious issue of IPTG toxicity can be circumvented by the use of lactose as inducer (Menzella et al. 2003).

The tryptophan operon
The trp operon consists of a set of structural genes that are expressed as a polycistronic mRNA which is translated into five enzymes responsible for the synthesis of tryptophan. The genes trpE through trpA, in figure 15, are the five enzymes that catalyse the synthesis of the amino acid tryptophan from chorismic acid.

![Tryptophan operon diagram](http://web.mit.edu/engbio/www/tpcc/tpceother.html)

**Figure 15. The regulation of the tryptophan operon**

In the absence of tryptophan, or at low concentrations, tryptophan will not bind to the aporepressor protein, encoded by trpR, located upstream the trp regulation region (not shown in the figure), which will inhibit the binding to the operator region and thus promote transcription. When
tryptophan is present, it binds to the aporepressor protein, which undergoes a conformational change and the aporepressor protein binds to the operator site and inhibits RNA polymerase from start transcription. The tryptophan operon has a second regulation of tryptophan control, i.e. except from the regulation with TrpR, that is called attenuation. This second regulation involves two components, tRNA, specifically tryptophanyl-tRNA\textsuperscript{Trp}, i.e. tRNA\textsuperscript{Trp} charged with tryptophan and the trpL gene. The trpL gene codes for a short 14 aa oligonucleotide, which contains two consecutive trp codons and therefore serves as a measure of the tryptophan supply in the cell. If the tryptophan concentration is high in the culture medium then the tRNA will be charged and the leader peptide will be translated without problem. In contrast, with a low concentration the tRNA will not be charged and translation will stall the trp codons.

![Tryptophan control by attenuation](www.mun.ca/biochem/courses/3107/Lectures/Topics/Trp_operon.html)

Figure 16. Tryptophan control by attenuation

This affects the transcription of the trp operon, since the trpL mRNA region can adopt a number of different conformations. It contains several self-complementary regions which can form a variety of stem-loop structures, figure 16. Since transcription and translation can occur simultaneously in bacteria the ribosome attaches to mRNA and influences the formation of secondary structures of the mRNA. In the case of trpL mRNA, when there is a high concentration of tryptophan, the ribosome follows right behind RNA polymerase until it is halted by a stop codon (UGA as seen in the bottom of region 1). This permits the formation of the stem-loops, as seen to the left in the figure, and will cause RNA polymerase to dissociate. On the contrary, a low concentration of tryptophan, results in no charged tRNA\textsuperscript{Trp} with tryptophan and the ribosome will stall waiting for a suitable tRNA to be brought. Because of the ribosome stalling, the region 2 and region 3 can base-pair as soon as both have been transcribed, as seen to the right. Attenuation as a means of regulating expression occurs.
in a number of other amino acid biosynthetic operons, such as the pheA, his, thr, leu and the ilv-operon. In all cases, the leader region is rich in codons for the particular amino acids that are synthesized by the enzymes encoded by the particular operon and it can form two alternative stem-loop structures, one of which is a transcription terminator.

The production with this promoter has resulted in levels up to 30 % of the total protein production and examples of proteins expressed with trp vectors include human proteins such as growth hormones, prolactin and interleukin (Yansura and Henner 1990). The vectors used are often low copy numbers plasmids, as a high copy number plasmid would risk saturation of the limited number of repressor molecules present in the bacterial cell. Although this promoter has been used to express a number of proteins, it obviously has some drawbacks, which in many cases has led to the use of other promoters. The induction ratio is poor, only about 50-fold (Bogosian and Somerville 1984; Hallewell and Emtage 1980), compared to the lac promoter that has a twenty times higher induction ratio, i.e. a 1000 fold. Thus, the leaky trp promoter would not be a choice when expressing toxic proteins. A second drawback is the induction condition, which is starvation of tryptophan. A minimal medium that lacks tryptophan is inoculated with up to 5% of an overnight culture in rich medium. After a period of about 2 hours the pool of tryptophan has been depleted, and the trp promoters have been fully induced. Indole-3-acrylic acid, IAA, is then added to the medium to speed up the synthesis of tryptophan needed to express the heterologous protein as well as the proteins needed for survival of the host cells. Since the expression levels obtained are rather high the proteins sometimes form inclusion bodies.

**The T7 promoter-driven systems**

The RNA polymerase of bacteriophage T7 is very selective for a relatively large and specific promoter sequence: the natural T7 promoters share a highly conserved sequence covering bp -17 to +6 relative to the start of the RNA chain (Dunn and Studier 1984). Furthermore, efficient termination signals for T7 RNA polymerase also appear to be rare, so that the enzyme should be capable of making complete transcripts from almost any DNA. Cloning of the gene for T7 RNA polymerase has provided a convenient source of T7 RNA polymerase, which can direct the selective production of large amounts of specific RNAs and proteins both in vivo and in vitro (Davanloo et al. 1984; Tabor and Richardson 1985). Several vectors for placing DNA under the control of T7 promoters have been described, primarily by commercial suppliers. In some cases, transcription by T7 polymerase is so active that transcription by the host RNA polymerase
apparently can not compete and almost all transcription in the cell rapidly becomes due to T7 RNA polymerase (Studier and Moffatt 1986).

The vector system based on the T7 promoter driven system, most widely used for cloning and expression of recombinant proteins is the pET system. The plasmids containing T7 promoter and the target gene are cloned in hosts not containing the T7 RNA polymerase gene to avoid plasmid instability due to the production of potentially toxic proteins. The vectors containing the target gene are transformed into a host containing a chromosomal copy of the T7 polymerase gene under the control of the lac promoter, mostly the derivative lacUV5 is used, and expression is induced by addition of IPTG. In figure 17 the control elements in a host and vector for T7 RNA polymerase levels are illustrated. The RNA polymerase expression is under the control of the lac promoter and the operator located at the host chromosome as seen to the right. The repressor gene, lacI, is located in the host chromosome and expresses the repressor molecule by the addition and induction by IPTG. Since the pET vector also contains the lac operator it will also be induced by the IPTG addition. Thus the expression of RNA polymerase and the transcription of the target gene will be operating almost simultaneously. The control with the lac promoter always allows low levels of transcription of RNA polymerase in an uninduced state. If the target protein has detrimental effects on the host cell, this low level of transcription will have to be avoided. By the use of hosts carrying either pLysS or pLysE, which encodes T7 lysozyme, a natural inhibitor of T7 RNA polymerase, the low level of RNA polymerase will be eliminated. pLysS produces low amounts of lysozyme while pLysE produces higher amounts of lysozyme, which enables the choice between a high or low level of induction control.

![Diagram showing the control elements in the pET expression system](http://wine1.sh.fiu.edu/bch5425/lec25lec25.htm)

Figure 17. The control elements in the pET expression system (http://wine1.sh.fiu.edu/bch5425/lec25lec25.htm)

Since, the induction with IPTG is both expensive and toxic the phage encoded cL repressor which encodes a temperature sensitive repressor has been used to express recombinant proteins. The repressor is functional at lower temperatures but denatures at temperatures higher then 37°C. The
induction is practically performed by a simple temperature shift. The introduction of the chromosomal copy of the T7 gene fused to the temperature regulated λP₁ and λP₉ tandem promoter has shown to give high production of the heterologous protein carbamoylase in the E. coli strain BL21 (Chao et al. 2002). The very high synthesis rate from the strong promoters as T7 and other phage promoters, creates a very rapid accumulation of the heterologous protein, which may cause aggregation forms of the proteins as inclusion bodies, since the capability of the system to correctly fold the proteins might be overloaded. However, there are examples where the high production rate can be tuned through repressor titration and thus better match the metabolic capacity of the host cell as shown for the pET expression system (Striedner et al. 2003).

The arabinose operon

The enzymes in the arabinose operon are encoded by the genes, araB, araA, araD, and are responsible for the degradation of the five carbon sugar, arabinose. The operon also contains a gene, araC, for a regulatory C protein, which have two operator sites, araO₁ and araO₂, and a promoter araI₁, figure 18.

![Figure 18. Map of the arabinose operon and its regulator genes.](image)

The formation of cAMP-CRP and expression of the protein C are required for efficient transcription. A distinctive feature of this operon is that the protein C also serves as a negative regulator. The direction of transcription of araC is opposite to that of araBAD and is controlled by the O₁ rather than the O₂ operator. The synthesis of the protein C is autoregulated, figure 19. With low levels of protein C and cAMP-CRP the mRNA for protein C is formed, in A, and with high levels of protein C it binds to the araO₁ gene and blocks the transcription. The binding of a second molecule of C to araO₂ and araI₁, makes a DNA loop formation, in B. The DNA loop accomplishes several things for the arabinose system. It sterically blocks access of RNA polymerase to the BAD promoter, thus holding the basal levels of BAD expression low. The loop also blocks the
access of RNA polymerase to the C promoter. The loop to $O_2$ also keeps AraC from occupying the $I_2$ half-site which could inappropriately activate the BAD genes. When the catabolite repression is relieved, thus forming cAMP which binds to CRP this DNA loop can not be formed. Arabinose binds to the protein C which undergoes a conformational change and bind to $araO_1$ and $araI_2, as shown in C. The presence of these positive factors enables the transcription of the BAD genes.

The tight control of the araBAD promoter makes it suitable for recombinant production. The promoter in multicopy plasmids has shown a relatively high induced/repressed levels of expression as shown with β-galactosidase (Cagnon et al. 1991) and alkaline phosphatase, AP, (Guzman et al. 1995), with the expression ratio, 240-700, and 200-1200, respectively. It is furthermore possible to regulate the promoter over a 1200 fold range and vectors have been constructed to give a 30 % yield of the total protein production (Cagnon et al. 1991). Another advantage is that the promoter can be turned off almost completely by the addition of fucose, which is a structural analog to arabinose, and this might be particular important with the production of toxic proteins. Moreover arabinose is
less expensive than IPTG. The promoter has the possibility to achieve graded levels of expression by varying the arabinose concentration, but an extensive heterogeneity in cell populations treated with subsaturating concentrations of the inducer has been observed (Siegele and Hu 1997). A host strain for an efficient uptake of arabinose by a constitutive synthesis of the arabinose transporters will be needed for the precise control of levels of protein accumulation with the araBAD promoter.

**The PHO regulon**

*E.coli* uses three kinds of phosphorus compounds for growth, inorganic phosphate, Pi, organophosphates and phosphonates, Pn. There are two transporters for the uptake of the preferred phosphorus, Pi. When Pi is in excess low affinity system, Pit is activated, and the high affinity transporter system, Pst and other genes for the use of alternative phosphorus sources are repressed. The genes expressed when Pi is limited belong to the phosphate regulon, hereafter called the PHO regulon, and include 31 or more genes arranged in eight separate operons. The expression of the different genes and their functions are described elsewhere (Wanner 1993).

The metabolism of phosphorus is complex with many controls acting on the PHO regulon. The assimilation of any phosphorus compound involves two basic steps, first the uptake of Pi or an alternative phosphorus compound and second the Pi or the alternative phosphorus compound has to be incorporated into ATP, the primary phosphorus donor in metabolism. Eventually, phosphorus is incorporated into essential components in membrane lipids, complex carbohydrates such as lipopolysaccharides, and nucleic acids. The PHO regulon is controlled by PhoB, the sensor, and PhoR, acting as a transcriptional activator, and is required in the phosphorylated form. The phosphorylation of PhoB is regulated by three controls. One is Pi-dependent and requires the PhoR sensor. The other two are Pi independent, in which one requires the sensor CreC and the other requires acetyl phosphate, and activate transcription in the absence of PhoR.

The Pi dependent control is a kind of transmembrane signal transduction. It is coupled to the Pi uptake. When Pi is in excess the Pst system and the PhoR protein is repressed and when Pi is limited the PHO regulon is activated by PhoR. The PhoR protein probably has a dual role in the Pi control, it can be either in an activated form or in a repressed form. During Pi limitation PhoR turns on genes of the PHO regulon by phosphorylating PhoB that in turn activates transcription by binding to promoters that share an 18-base consensus PHO box, the *phnC, phoA, phoB, phoH, pstS* and *ugpB* genes (Wanner 1993). When Pi is in excess, PhoR, Pst and PhoU together turn off the PHO regulon, presumably by dephosphorylating PhoB. In addition two Pi independent controls that
may be forms of cross regulation turn on the PHO regulon in the absence of PhoR. The sensor CreC, formerly called PhoM, phosphorylates PhoB in response to some (unknown) catabolite, while acetyl phosphate may be examples of underlying control mechanisms important for the general (global) control of cell growth and metabolism.

The promoter of the gene alkaline phosphatase, phoA, is induced by phosphate starvation and used in recombinant protein production. It has been shown that the induced state of the phoA promoter leads to a deletion of the recombinant plasmid (Lübke et al. 1995). This effect can however be reduced by growing the cells to a maximal cell density before induction (Lübke et al. 1995). A very successful use with the phoA promoter is the expression of antibodies in E.coli with a two :i:tron system with optimised light and heavy chain translation levels (Simmons et al. 2002).

Summary of inducible promoters

Strong promoters have been used widely with both advantages and disadvantages. Interferons for example are well suited for production with strong promoters. Human interferons have been produced with the strong lac and trp promoters of E.coli and also with the promoter of the recA gene in E.coli (Feinstein et al. 1983). The recA promoter is considered to be one of the strongest E.coli promoters as shown with the expression of β-lactamase and kanamycin phosphotransferase (Miki et al. 1981) respectively, when replacing the native promoter with the recA promoter.

However, the limitations of the established promoters and most common strong promoters lead to the research and development of new promoters. The most used substrate induced promoters for recombinant proteins is shown in table 2, where the most significant advantages and disadvantages are listed for each system. One limitation with a strong promoter is that the target protein is often unable to reach a native conformation and either partially or completely segregates to inclusion bodies. This problem has been partially solved by the co-expression of folding modulators or by fusion protein technology as discussed previously. An alternative approach is to use promoters that are induced by a temperature down-shift, as proper folding often is favoured in a low temperature (Baneyx 1999). The best characterised cold-shock promoter is that of the major E.coli cold shock protein CspA (Padhare et al. 1999; Mujacic et al. 1999). The promoter is rather well repressed at 37°C and above. The major disadvantage is that it is only activated in 1-2 hours after the shift in temperature. However, a temperature shift might not always be preferred since it may cause unwanted metabolic reactions in the host.
Starvation promoters as described previously, phoA and trp promoters, will have the difficulty to obtain graded levels of expression. The nutrient of which the cells is being starved will be depleted by the cell to a very low level in the medium, generally within a short period of time after imminent starvation is sensed. The cells will thereafter create a cascade response and the promoter will become fully induced.

<table>
<thead>
<tr>
<th>Name</th>
<th>Origin</th>
<th>Regulation principle</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_e$</td>
<td>Lactose operon</td>
<td>Repressed by LacI, LacI$, induction by lactose or IPTG, catabolite repressed</td>
<td>Strong promoter, rapid promoter activation, IPTG is not metabolised</td>
<td>IPTG is toxic and expensive, inclusion body formation, high basal levels</td>
</tr>
<tr>
<td>$P_{lac+IPT}$</td>
<td>Lactose operon</td>
<td>Repressed by LacI, LacI$, induction by lactose or IPTG</td>
<td>Not subjected to catabolite repression, IPTG is not metabolised</td>
<td>IPTG is toxic and expensive, inclusion body formation, high basal levels</td>
</tr>
<tr>
<td>$P_{tr}$</td>
<td>Lactose/Tryptophan operon</td>
<td>Repressed by LacI, LacI$, induction by lactose or IPTG, thermal</td>
<td>Allows accumulation of protein 15-30% of total cell protein production</td>
<td>IPTG is toxic and expensive, inclusion body formation,</td>
</tr>
<tr>
<td>$P_{tr}$</td>
<td>Lactose/Tryptophan operon</td>
<td>Repressed by LacI, LacI$, induction by lactose or IPTG, thermal</td>
<td>Allows accumulation of protein 15-30% of total cell protein production</td>
<td>IPTG is toxic and expensive, inclusion body formation,</td>
</tr>
<tr>
<td>$P_{tr}$</td>
<td>Tryptophane operon</td>
<td>Repressed by TrpR-trypthicolor complex, tryptophan starvation relieves repression, IAA mimics Trp starvation</td>
<td>Strong promoter, tightly regulated</td>
<td>IAA is expensive, defined synthetic medium is required, poor induction ratio, potential problems with mistranslation</td>
</tr>
<tr>
<td>$P_{re}$</td>
<td>Gene 10 promoter of phage T7</td>
<td>Thermal, IPTG(El+3 vectors), transcribed exclusively by phage T7 polymerase</td>
<td>Very strong promoter, exclusive transcription</td>
<td>The gene for T7 polymerase must be introduced into host cell and its synthesis controlled, owing to extremely high mRNA production, problems with aberrant translation can occur</td>
</tr>
<tr>
<td>$P_{ara}$</td>
<td>Arabinose operon</td>
<td>araC is corepressor, induced by L-arabinose</td>
<td>weaker than the lac promoter, tightly controlled, arabinose is less expensive compared to IPTG</td>
<td>heterogeneity in cell population,</td>
</tr>
<tr>
<td>$P_{pho}$</td>
<td>Phosphate operon</td>
<td>phoH (positive), phoB (negative), induced by phosphate starvation</td>
<td>Tight regulated</td>
<td>Need a tightly controlled phosphate limitation, can limit the duration of protein synthesis</td>
</tr>
</tbody>
</table>

*Table 2. Commonly used substrate induced promoters for recombinant protein production.*

The Maltose Regulon and metabolism

**Maltose uptake and metabolism**

The transport of maltose and higher maltodextrins, α-(1-4) maltodextrins from the medium to the cytoplasm occurs without chemical alteration of the substrate and against a considerable concentration gradient (Hengee and Boos 1983). The proteins involved in the uptake system are illustrated in figure 20. The λ receptor, the product of lamB, an integral membrane protein transports maltodextrins through the membrane to the periplasmic space. The periplasmic maltose binding protein, MalE, has a site with specific affinity to the λ receptor. Furthermore, it has a specific affinity towards maltodextrins, which is not limited by the length of the maltodextrins but the transport and growth of maltodextrins higher than six glucose residues is not possible (Wandersman et al. 1979). MalE is a monomeric and soluble protein. Upon binding to the maltodextrin, MalE undergoes a conformational change (Szmelman et al. 1976; Zukin 1979). The
actual membrane transporter from the periplasm to the cytoplasm contains a three protein complex with the composition, MalFGK₂. This ATP dependent transport with MalE, MalF, MalG and MalK involved is called the ABC cassette transport system. MalF and MalG are two integral membrane proteins and the cytoplasmic protein, MalK, is the energy-coupling protein binding to ATP.

The incoming maltodextrins of up to six glucose moieties are metabolised to glucose and glucose 1-phosphate by the combined action of three cytoplasmic enzymes, amylomaltase, MalQ, maltodextrin phosphorylase, MalP and maltodextrin glucosidase, MalZ as shown in figure 21. MalQ is a dextrinyl transferase that can transfer maltosyl and longer dextrinyl residues onto glucose, maltose and longer maltodextrins. Acting on maltotriose, the smallest substrate MalQ can recognise, it releases glucose from the reducing end, forms a maltosyl-enzyme complex, and transfers the maltosyl residue onto the nonreducing end of an acceptor, glucose, maltose or a longer maltodextrin. Thus, maltose is not a substrate of MalQ but only an acceptor in the transfer reaction catalysed by the enzyme. That means that for maltose degradation the cell has to be able to internally produce small amounts of maltodextrins as primers with the minimum size of maltotriose. MalP recognises maltopentose and longer maltodextrins and forms glucose-1-phosphate by sequential phosphorolysis of the nonreducing end of the maltodextrins. It is important that MalP does not attack maltotetraose and maltotriose, since those dextrans are required for full activity of MalQ. The reaction of MalP is reversible and the rate of phosphorolysis will be stimulated by the cytoplasmic phosphorus concentration. The phosphorus concentration will in turn depend on the external phosphorus concentration and the phosphorus containing organic compounds, as well as the state of the phoB-dependent PHO regulon (Xavier et al. 1995). MalZ hydrolyses maltotetraose and smaller maltodextrins to glucose and maltose, with the smallest substrate being maltotriose. In contrast to other glucosidases, MalZ preferentially removes glucose consecutively from the...
reducing end of the rialtodextrin chain (Tapio et al. 1991). The last enzyme of the maltose system, the product of the *malS* gene, MalS, is the periplasmic α-amylase. It preferentially cleaves maltohexose from the non-reducing end of longer maltodextrins, but it also recognises smaller maltodextrins down to maltotriose.

**Figure 21.** Maltose degradation of maltodextrins by the enzymes, MalQ, MalP and MalZ (Boos W and Shuman H 1998).

MalS is not essential for growth on maltose but allows slow growth on long dextrins (Freundlieb and Boos 1986). The resulting products from the action of, MalQ, MalP and MalZ on the maltodextrins are glucose and α-glucose 1-phosphate. Glucose is further phosphorylised by glucosidase, *glk*, to glucose-6-phosphate and α-glucose-1-phosphate is transformed by phosphoglucomutase, *pgm*, to glucose-6-phosphate. The glucose-6-phosphate then enters the glycolysis. Mutants unable to phosphorylate glucose due to the lack of glucokinase, Enzyme II^Glc^, encoded by *ptsG*, and enzyme II^Man^, encoded by *ptsM*, of the PTS system are unable to grow on maltose (Buhr et al. 1992).

**Gene level control**

MalT, the transcriptional activator, maltotriose and ATP are essential for the expression of the *mal* genes. The network regulation in this system occurs on two different levels (Schegel et al. 2002).
The first level concerns the gene regulation of the expression of MalT, and the second level concerns the control on protein level and affects the activity of MalT and will be discussed later. The mal genes and their functions have been listed and described in an excellent review (Boos and Shuman 1998). The mal genes are all under the regulation of MalT, the transcriptional activator, and under catabolite repression. The malT gene is not autoregulated by its own gene product, but it is subject to catabolite repression. Expression of malT is controlled by a global repressor protein, Mlc, encoded by the gene mlc. The repressor activity of Mlc is controlled by the transport status of the glucose-specific enzyme, EIICB of the PTS that causes sequestration of Mlc when glucose is transported. Mlc binds to the regulatory region of malT at a sequence with palindromic structure, the Mlc-box.

![Diagram](image)

**Figure 22.** The characteristic features of the mal promoters, two MalT boxes in direct repeat. Red boxes, open and closed corresponds to cAMP/CRP binding sites.

The characteristic feature of some of the MalT dependent promoters is the lack of the usual –35 region. The –10 region corresponds to those of constitutive promoters. Instead the MalT box is situated at –37.5 or at –38.5 upstream of the transcriptional start point (Bedouelle et al. 1982; Bedouelle 1983). Another characteristic feature found to be essential for MalT-dependent mal gene expression is two MalT boxes in direct repeat upstream of the –38 region (Raibaud et al. 1989; Vidal-Ingigliardi and Raibaud; Vidal-Ingigliardi et al. 1991). In figure 22, the MalT boxes and the cAMP-CRP binding sites are shown for some of the mal promoters. The promoter for malK and malE, respectively share the same MalT boxes and cAMP-CRP binding sites. The region has four binding sites for the complex cAMP-CRP and five binding sites for MalT (Vidal-Ingigliardi et al. 1991). The expression is suggested to require three out of four CRP binding sites and all five MalT binding sites to be fully activated.
Protein level control

Maltotriose, the endogenous inducer, activates MalT on a protein level, and is produced during growth on any maltodextrin but also under conditions of low levels of other carbon sources, such as glucose, lactose, galactose and trehalose (Decker et al. 1999). Thus, this indicates that maltotriose can be formed endogenously in the absence of maltodextrins in the medium. There are at least two suggested ways by which maltotriose can be formed endogenously: 1.)

By the degradation of glycogen and trehalose, and 2.) by the formation of free internal glucose as well as glucose-1-phosphate (Decker et al. 1993). By the action of three different enzymes on MalT, separately, it can lead to the deactivation of MalT. The first is the MalK protein, which acts as a repressor on MalT through the C-terminal end. The mechanism is a result of a direct MalK-MalT interaction that stabilises MalT in its inactivated form, figure 23. Furthermore MalK serves as the target for regulation of transport activity by binding of the unphosphorylated form of EIIBC of the PTS transport system, known as the inducer exclusion. The second is MalY, a pyridoxal phosphate containing enzyme exhibiting cystathionase activity. The repression of MalT by MalY does not require its cystathionase activity (Zdych et al. 1995). Purified MalY binds to monomeric MalT and inactivates it in an in vivo transcription assay using purified proteins. The interaction of MalY and MalT as well as the strength of inhibition is counteracted by the addition of maltotriose (Schreiber et al. 2000). Furthermore it is unclear what the natural substrate of MalY is, but the cluster of the malY and malX in an operon and the common regulation by MalI indicates that an
unknown sugar possibly a cystein derivative, might be the substrate of these gene products. The third protein that represses MalT is a cytoplasmic acetyl esterase, Aes. It is a negative effector of MalT that competes with maltotriose (Joly et al. 2002).
PRESENT INVESTIGATION

In the year 2001, 25 000 genes in the humane genome were reported. This has led to a start of a huge work of linking this information to the location and function of the corresponding proteins. This task includes the determination of the structure of the proteins and their function in the cell metabolism, how they interact and how this interaction changes with time, which might lead to the development of new drugs. A severe bottleneck for this development is the production of large quantities of proteins with unknown structure and function, in an active and functional form and in sufficient amounts for further studies. A significant part of the challenge lies in the expected but still unknown equally wide portfolio of production systems of these proteins. Only a small part of these proteins can be expected to be produced by already developed production systems.

The goal with this work is therefore to contribute to new production methods to produce soluble, active and correctly folded proteins. In this respect the present work has included methods to influence the inclusion body formation, proteolysis and translocation capacity. The hypothesis of this work is that the rate of synthesis of the nascent polypeptide is a main factor controlling the above mentioned parameters. The high accumulation rate leads to a large pressure on the posttranslational steps which are rate limiting. The folding intermediate, which is a loosely folded complex is created very fast from the nascent polypeptide (Henrick and Hartl 1993). The subsequent steps leading to a correctly folded protein in the cytoplasm or the periplasm are relatively slow processes due to the function of chaperones and folding catalysts as well as transporters of the secretory system. If the amount of these components is not sufficient it could result in misfolded protein and/or proteolysis as well. Furthermore, as long as the protein is unfolded there is a risk of aggregation to other unfolded proteins resulting in inclusion body formation.

It is reasonably to interfere with the synthesis rate of the nascent polypeptide as early as possible in the chain of cell operations, thus during the transcription, in order to reduce undesirable side effects and to minimise waste of precursors and energy.

The transcription rate is influenced by: 1) the promoter strength, 2) the substrate feed rate, 3) the gene dosage.
In this work we have chosen two substrate induced promoters, \( P_{\text{malK}} \) and \( P_{\text{lacU5}} \), respectively. Both promoters are induced during growth limiting conditions as in this work is performed in fed-batch cultivations. Different substrate feed rates in fed-batch cultivations are used to control the transcription from \( P_{\text{malK}} \). The transcription from \( P_{\text{lacU5}} \) is controlled by different amount of inducer, IPTG during fed-batch cultivations. The gene dosage in the respect of induction levels are controlled by different amounts of IPTG with \( P_{\text{lacU5}} \) and by different feed rates with \( P_{\text{malK}} \). A low copy number vector is chosen to avoid the effects of a high copy number vector.

The productivity in production systems based on \( P_{\text{malK}} \) was studied with the homologous model protein \( \beta \)-galactosidase (I, II).

To verify the hypothesis, the inclusion body formation and proteolysis were measured during glucose-limited fed-batch processes with two different model proteins (III, IV) based on both promoters. The model proteins used were the maltose binding protein, MalE, and a double mutant protein, MalE31. The latter is known to form inclusion bodies and would in this respect be more difficult to express to a soluble correctly folded protein.

\[ P_{\text{malK}} \]

This work has chosen to study a promoter from the catabolite repressed maltose regulon, the promoter of the \( \text{malK-lamB-malM} \) and \( \text{malE-F-G} \) operons, further on referred to as \( P_{\text{malK}} \). \( P_{\text{malK}} \) has shown to have qualities of a system characterised by alternative kinetic properties (I, II). The induction of \( P_{\text{malK}} \), and other promoters in the maltose regulon, except for the promoter of the \( \text{malT} \) gene, is based on the positive transcriptional activation by MalT and the binding to cAMP-CRP. The formation and activation of these two compounds rely on three mechanisms. These mechanisms lead to protein level activation of MalT by maltotriose, and/or formation of cAMP. The formation of cAMP leads to transcription of the \( \text{malT} \) gene. The three mechanisms are, as illustrated in figure 24, (1) low/high concentration of maltose which through the maltose metabolism, figure 21, forms maltotriose, (2) low concentration of glucose, which through the catabolite derepression elevates the levels cAMP, and (3) endogenous formation of glycogen and trehalose leading to the formation of maltotriose. The degradation of glycogen and trehalose is suggested to form free glucose and glucose 1-phosphate (Boos and Shuman 1998) from where maltotriose can be formed. Glycogen is accumulated with an excess of carbon source and degraded when the availability of carbon source drastically changes (II). The degradation of glycogen is most significant with the use of a low limited feed compared to a high limited feed on glucose with a preceding batch phase on glucose (II). The other important component, cAMP, is inversely proportional to the glucose concentration (II; Notley and Ferenci 1995). The levels of cAMP are also increased at a switch
between glucose and maltose, due to the catabolite derepression (I). There are thus two important routes from which high levels of cAMP can be reached.

**Figure 24.** Schematic summary of the mechanisms for induction of the catabolite repressed genes of the mal regulon excluding malT. This illustration excludes the induction due to other substrates other than glucose and maltose. K=MalK, T_i and T_e inactivated and activated forms of MalT, respectively.

Since the maltose regulon is specific for the uptake and transport of maltose, there is a strong reason to believe that external maltose should be the first choice as inducer of P_{malK}. However, batch cultivations on maltose resulted in low levels of production (II). In fact only lower levels were seen in batch cultivations on glucose where the maltose regulon is catabolite repressed. To understand the low production levels with excess of maltose the structure of the promoter was studied. The DNA sequence of P_{malK} is shown in figure 25, which includes five binding sites for MalT and four

\[
P_{malK}:
\begin{align*}
&\text{GAATTCGCGC AACTCCTTC CATCCTCCCT GCCCCCTACGC CCCACCGTCG} \\
&\text{CTTGTGTGA CGATGGGGG CGATATAAGC GAAATGATGAGATGCGCACA} \\
&\text{TAATATGCCC AGCATTTTGG CAGGCAACAT CAGAAATTTC CTGACATGAC} \\
&\text{CTCAGTTTAG TGACAGAAG CGGTGTTCTCT ATCCCTCCGGG CTCTCTCCCC} \\
&\text{ATAAAAAGGC CAGGGGTTGG AGGATTAAAG CCAATCTCTTG ATGACGCATA} \\
&\text{GCAGCCAT CATGAAAGTT GCATTGTGATG ACAGTTGTGTA CAAAAAGGGAG}
\end{align*}
\]

**Figure 25.** The DNA sequence of P_{malK}. The bold letters represents the MalT binding sites and the underlined letterst represent the cAMP-CRP binding sites. The –35 and –10 segments and the start site of transcription are indicated.

binding sites for cAMP-CRP as previously mentioned. The characteristics with two binding sites of MalT in direct repeat and the structural motif to fully activate the promoter as mentioned earlier is present in all of the promoters of the maltose regulon, although not always in the same location or in the same orientation with respect to the transcription start site. The different binding sites and the structural motif indicate the importance of the substances, activated MalT and cAMP-CRP, for
induction but also a possibility to achieve different levels of activation with varying amounts of MalT and/or cAMP-CRP. In conclusion, with an excess of maltose, maltotriose is formed through the maltose metabolism but can only activate the already present MalT molecules. No further formation of MalT occurs since there is no or very low levels of cAMP and transcription is not started.

\[ P_{\text{lacU5}} \]

\( P_{\text{lacU5}} \) is a derivative of \( P_{\text{lac}} \) and carries two point mutations in the CRP-binding site rendering the expression insensitive to catabolite repression (Arditti et al. 1973), figure 26, e.g. the promoter will not require cAMP to be fully activated and can thus be activated under excess of glucose by IPTG. IPTG binds to the repressor and cause a conformational change of the repressor which can no longer bind to the operator. The RNA polymerase is free to bind the operator and start transcription. A drawback with this promoter is the high basal levels occurring due to the independence of the catabolite repression.

\[
\begin{align*}
\text{P}_{\text{mut}}: & \\
\text{GCAAGCAGAT} & \text{TAATGAAGT} & \text{TACACACAC} & \text{ACTTATTAGG} & \text{CACCCAGGC} \\
\text{Mut.CRP-site} & & & & \\
\text{TTTACTTCT} & \text{ATGCTTCGG} & \text{CTGTATAA} & \text{GTGGAATT} & \text{GTGACGGA} \\
\text{-35} & & & \text{Lac-repressor binding site} & \text{-10} \\
\text{AACAATTCCA} & \text{GACCCGAAA} & \text{AGCTATGACC} & \text{S-D} & \text{MetThr}
\end{align*}
\]

\textit{Figure 26.} The DNA sequence of \( P_{\text{lacU5}} \). The two mutations in \( P_{\text{lacU5}} \) is in the CRP-site and the \(-10\) segment indicated by a larger size letter. SD is the Shine-Dalgarno sequence and MetThr is the start of transcription site.

**Expression system**

The expression vectors in work I, II, III and IV are different with respect to copy number, product protein and promoter, \( P_{\text{lacU5}} \) or \( P_{\text{malK}} \), respectively, as summarised in table 3. The first work (I) was based on a medium copy number plasmid derived from pBR322 (Casadaban et al. 1980) with the size of 10213 bp, coding for the \textit{lacZ}, \textit{lacY} and part of the \textit{lacA} gene, including both tetracycline and ampicillin resistance genes. The production levels obtained from this expression system were comparatively low and corresponded at highest to 1% of the total protein production. The reason of the low protein level can have many explanations, but could be the size of the plasmid, which cause a metabolic burden to the host cell. To optimise the production levels the size of the plasmid was reduced, which included depletion of the genes \textit{lacY} and \textit{lacA} and an elimination of the resistance gene for tetracycline. This resulted in a new plasmid (II) also derived from a pBR322 vector with the benefit of having a RNase III cleavage site to enable an unchanged 5’end for the \textit{lacZ} mRNA (Linn and Pierre 1990). This resulted in a considerable smaller expression vector of 8814 base pairs.
induction but also a possibility to achieve different levels of activation with varying amounts of MalT and/or cAMP-CRP. In conclusion, with an excess of maltose, maltotriose is formed through the maltose metabolism but can only activate the already present MalT molecules. No further formation of MalT occurs since there is no or very low levels of cAMP and transcription is not started.

**P_{lacU5}**

P_{lacU5} is a derivative of P_{lac} and carries two point mutations in the CRP-binding site rendering the expression insensitive to catabolite repression (Arditti et al. 1973), figure 26, e.g. the promoter will not require cAMP to be fully activated and can thus be activated under excess of glucose by IPTG. IPTG binds to the repressor and cause a conformational change of the repressor which can no longer bind to the operator. The RNA polymerase is free to bind the operator and start transcription. A drawback with this promoter is the high basal levels occurring due to the independence of the catabolite repression.

\[
\begin{align*}
P_{lacU5} & : \\
\text{GCAACGCAAT} & \text{TAATGTAAGT} \text{TAGCTCACTC} \text{ACTCAATTAGG} \text{CACCCAGGC} \\
\text{Mut.CRP-site} & \\
\text{TTTACTT} & \text{ATGCTCCGG} \text{CTCGTATAAAT} \text{GTGGAATT} \text{GTGACCGGA} \\
\text{-35} & \text{AACATTTCA} \text{CACGGAAAC} \text{AGCTATGACC} \\
\text{S-D} & \text{MetThr} \\
\text{Lac-repressor binding site} &
\end{align*}
\]

*Figure 26.* The DNA sequence of P_{lacU5}. The two mutations in P_{lacU5} is in the CRP-site and the –10 segment indicated by a larger size letter. SD is the Shine Dalgarno sequence and MetThr is the start of transcription site.

**Expression system**

The expression vectors in work I, II, III and IV are different with respect to copy number, product protein and promoter, P_{lacU5} or P_{malK}, respectively, as summarised in table 3. The first work (I) was based on a medium copy number plasmid derived from pBR322 (Casadaban et al. 1980) with the size of 10213 bp, coding for the lacZ, lacY and part of the lacA gene, including both tetracycline and ampicillin resistance genes. The production levels obtained from this expression system were comparatively low and corresponded at highest to 1 % of the total protein production. The reason of the low protein level can have many explanations, but could be the size of the plasmid, which cause a metabolic burden to the host cell. To optimise the production levels the size of the plasmid was reduced, which included depletion of the genes lacY and lacA and an elimination of the resistance gene for tetracycline. This resulted in a new plasmid (II) also derived from a pBR322 vector with the benefit of having a RNase III cleavage site to enable an unchanged 5’end for the lacZ mRNA (Linn and Pierre 1990). This resulted in a considerable smaller expression vector of 8814 base pairs.

36
mutant is less stable than the wild type (Betton and Hofnung 1996). The mutation is not supposed to influence the maltose binding site and the site where MalG binds due to its location.

Figure 28. The model fusion-protein Zb-MalE. Zb as shown in the figure is fused to the N-terminal of MalE. The localization of the mutation of Zb-MalE31 is marked by the rectangle at the top of the protein. The binding site for maltose is shown by the numbered amino acids and is located between the two domains where the maltose molecule is shown.

CONTROL OF GROWTH IN RELATION TO PRODUCTION

The three mechanisms for induction of P_{malK}, high/low concentrations maltose, low concentrations of glucose and endogenous degradation of glycogen and trehalose, are used to control growth and production. The three mechanisms were therefore coupled to fed-batch processes were the concentrations of glucose and maltose can be varied by the use of different feed profiles. The formation of cAMP, was achieved by glucose limited fed-batch and by glucose limited fed-batch cultivations followed by maltose addition, introducing a diauxic growth. Maltotriose formation due to glycogen degradation was achieved in glucose limited fed-batch cultivations. Maltotriose formation through the maltose metabolism was achieved by the addition of maltose after glucose limited fed-batch cultivations. The limiting glucose feed profiles used were, “very high”, “high”, “medium”, and “low” feed, corresponding to specific growth rates of 0.8/0.5/0.3/0.2 h⁻¹, respectively. A constant maltose feed or a maltose dose were after four generations in the glucose limited fed-batch introduced in two cultivations, thus creating a diauxic growth phase. The cultivations described here are illustrated in figure 29 and the mechanisms are indicated for each cultivation. When introducing a carbon source, which is not a PTS-sugar, a diauxic growth behaviour of the cells appear due to the time it takes to induce the genes involved in the uptake of the new substrate. The diauxic growth is depending on the limitation of the preceding glucose feed (Lendenmann and Egli 1995; I, II). Furthermore, the diauxic growth can be eliminated depending on the preceding feed of glucose (I, II). Figure 30 shows the cell mass accumulation during growth on maltose, where gram maltose/gram cells are the same in each case, with different feed rates of
Design of substrate induced transcription for control of recombinant protein production

glucose (II). Maltose accumulates during the diauxic growth and once the cell growth is resumed it results in a batch-like growth with the two highest preceding concentrations of glucose. With the “medium” preceding feed of glucose the diauxic growth is eliminated, which can be seen from the completely linear growth.

![Diagram of substrate induced transcription](image)

**Figure 29.** The glucose limited fed-batch cultivations performed in work II and III. The mechanisms to control induction are coupled to the processes and indicated in the figure.

![Graph of cell mass accumulation](image)

**Figure 30.** Cell mass accumulation during constant feed on maltose. The numbers indicated in the graph refer to the preceding feed on glucose, corresponding to specific growth rates of 0.3/0.5/0.8 h⁻¹, respectively (II).

The diauxic growth behaviour was also influenced by the secretion of product to the periplasm (III), figure 31. The production process was based on a “high” and “low” glucose limited fed-batch followed by addition of a maltose dose. The accumulation of cells in both the cytoplasmic and the secreted production, showed a similar exponential growth during the limited feeds of glucose. At addition of maltose the cell accumulation pattern differed between the systems.
The cell growth in the cytoplastic production system showed a significant diauxic growth while the cell growth in the periplasmic production system showed a reduced effect of the diauxic growth. The cell growth in the reference system, representing cells without plasmid, as also shown in the figure, showed a similar growth behaviour as for the cell growth in the periplasmic production. Thus, the burden to change the uptake system from glucose to maltose is influenced by the accumulation of product in the cytoplasm. This might further be explained by the metabolic response due to the switch in carbon source as will be discussed later.

![Graph showing cell growth](image)

**Figure 11.** The cell accumulation in five fed-batch processes during limited feeds on glucose, “high” (squares, corresponds to a specific growth rate of 0.5 h⁻¹) and “low” (circles, corresponds to a specific growth rate of 0.2 h⁻¹), respectively, followed by addition of maltose in excess as indicated by vertical lines. The closed symbols refer to cell growth with periplasmic production and the open symbols refer to cytoplastic production of Zb-MaIE. A reference cell accumulation pattern, cells grown without plasmid, is shown with “low” glucose feed and indicated by +.

Since $P_{malE}$ is controlled by catabolite repression, production systems based on this promoter can be divided into a growth phase and a production phase (I, II) in combination with the inducer exclusion of the PTS transport system. The lowest levels of production are obtained during catabolite repression, obtained in a batch phase on glucose. If the basal levels can be tolerated this will be an efficient means to separate growth from production although this system is substrate induced. If not, the basal levels mainly due to endogenous degradation of glycogen could probably be reduced or even avoided through abolishment of the batch phase, specifically when higher
glucose feed rate are used. This can be achieved by an immediate start of an exponential feed on glucose at the start of the production process.

The glucose controlled growth during “low” and “high” limited glucose feeds, respectively, results in different production levels (II). In figure 32A the production levels are plotted against the process time. The initial increase in production is faster with the more limited glucose feed. This is probably due to the more transient change in glucose availability, from batch on glucose to “low” feed on glucose compared to the “high” feed on glucose, where glycogen is degraded faster. However, the same level of production is reached during the first hour of the process. This might be due to that the same initial levels of maltotriose are formed during the preceding batch phase, which is the same in both systems. The maltotriose formation is the effect of degradation of glycogen (II).

The lower glucose availability results in higher cAMP levels (II). This is observed during the later part of the process which probably contributes to the higher production levels obtained with a “low” limited feed on glucose. In a process point of view the more rapid production will be with a high glucose feed as seen in figure 32A, but during a longer process time the low glucose feed production will result in higher production levels. Another way of studying these production processes is shown in figure 32B, where the production levels are plotted against developed generations. The production pattern does not seem particularly different during the first two developed generations, with the exception of the more rapid initial increase in production with the low glucose production. After four generations the production level reached is still higher with the more limited glucose feed as previously stated.

Although not shown here, the production levels can further be increased by the introduction of maltose. The production after addition of maltose shows a production lag due to the shift in carbon source, as also shown for the cell mass accumulation, figure 30, and is depending on the preceding feed of glucose. Once the production lag is overcome the levels can reach as much as 30% of the total protein production. The production in the maltose phase is due to the formation of cAMP and maltotriose which both are necessary to fully induce the promoter. Furthermore, production systems based on $P_{ malt }$ in glucose limited fed-batch cultivations can accumulate product for very long periods of time, i.e. well exceeding 30 hours or even more (I).
Figure 32. The production yield of β-galactosidase during two fed-batch processes on limited amounts of glucose. The numbers indicated in the graph correspond to a high, 0.5h⁻¹, and a low, 0.2h⁻¹, feed rate of glucose. A: the x-axis corresponds to developed generations and B: corresponds to process time (II).

**METABOLIC LOAD**

Production of recombinant proteins with strong promoter systems, most often result in a high metabolic load of the host cell “factory”, which is shown to result in stringent response. The stringent response can be measured by the increase of the cellular alarmone, ppGpp, as shown in figure 9. The levels of ppGpp, the cellular alarmone, figure 33, are measured during the glucose limited production process with addition of maltose as previously discussed with the production of cytoplasmic protein and protein secreted to the periplasm (II). At “high” feed rate on glucose, where production is shown to be lower (II), the levels of ppGpp are increased after addition of maltose. The level of ppGpp during the secreted accumulation of protein is 80 % of the ppGpp level seen when protein is accumulated in the cytoplasm. The reason for stress response is probably due to carbon starvation during the time of activation of the maltose uptake system.

At “low” feed rate, where production formation is shown to be higher (II), the increase in ppGpp occurs at the end of the glucose phase with the protein remaining in the cytoplasm and after the maltose addition when the protein is secreted to the periplasm. The increase in ppGpp with the protein in the cytoplasm does not seem to be coupled to the carbon starvation due to the change in uptake system. However, it could be a response due to limitation of carbon and energy source or lack of charged tRNAs. The increase in ppGpp, with the secreted protein at “low” glucose feed, occurs after the maltose addition and is significantly lower compared to the other processes. The stringent response in a “low” glucose limited fed-batch process seems to be coupled to the transport
of product, as no significant increase of the ppGpp levels is seen with the secreted protein. The transport to the periplasm might decrease the burden in the cytoplasm and might therefore result in a lower response to the lack of tRNAs.

Figure 33. Induction of the stringent response, shown as the accumulation of ppGpp, guanosine-3,5-tetraphosphate corresponding to the fed-batch processes in figure 31 (III).

**PROCESS DESIGN TO CONTROL PRODUCTIVITY**

As previously stated production from most industrial promoters have several drawbacks according to the possibility to reproducibly control the expression rate and to tune the production of protein. By the use of $P_{malK}$ in production processes based on glucose and maltose limited fed-batch cultivations, several features not characterised by the most frequently used promoters can be obtained. It was shown that production processes based on glucose and maltose can be established which are characterised by distinct and reproducible production profiles of glucose and maltose (II). Furthermore it was shown that the production profiles also depend on the feed rates of glucose and/or maltose addition, in fed-batch processes (II). The glucose limited fed-batch processes here performed, are conventional production systems also used with other promoters in other production processes, thus a high amount of knowledge is accumulated for this type of process to reach high productivity. The production obtained during two glucose limited fed-batch processes, with a “high” and a “low” feed on glucose (II), respectively, shows two different production curves. In the “low” feed process product is accumulated at a lower rate, but for a much longer production time.
which eventually leads to the establishment of a product level of 10% of the total protein which is 50% higher than with “high” feed. Higher production levels are reached with glucose limited production processes with the introduction of a maltose feed. Consequently also higher production rates are obtained after the lag time of the diauxic growth is overcome as previously discussed.

<table>
<thead>
<tr>
<th>PRODUCTION TECHNIQUE</th>
<th>PROMOTER</th>
<th>MECHANISM</th>
<th>qPmax [U/mg.h]</th>
<th>TIME TO qPmax [h]</th>
<th>DURATION [h]</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATCH</td>
<td>glucose</td>
<td>PmaltK</td>
<td>Catabolite repr.</td>
<td>1 &gt;3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>maltose</td>
<td>PmaltK</td>
<td>Maltotriose/mal</td>
<td>2 &gt;3</td>
<td>3</td>
</tr>
<tr>
<td>FEDBATCH:GLUCOSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;high&quot; feed</td>
<td>PmaltK</td>
<td>Maltotriose/gly</td>
<td>11 0.5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>phase I</td>
<td>PmaltK</td>
<td>Maltotriose/gly/unknown</td>
<td>13 &lt;0.5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>phase II</td>
<td>PmaltK</td>
<td>Maltotriose/gly/cAMP/cat.derepr</td>
<td>6 4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>phase III</td>
<td>PmaltK</td>
<td>Maltotriose/gly/cAMP/cat.derepr</td>
<td>3 8</td>
<td>&gt;7 5</td>
<td></td>
</tr>
<tr>
<td>FEDBATCH:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLUCOSE/MALTose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;med&quot; glucose feed</td>
<td>PmaltK</td>
<td>Maltotriose/mal/cAMP/cat.derepr</td>
<td>9 2 4</td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>&quot;high&quot; glucose feed</td>
<td>PmaltK</td>
<td>Maltotriose/mal/cAMP/cat.derepr</td>
<td>65 5 4</td>
<td>&gt;1 5</td>
<td></td>
</tr>
<tr>
<td>&quot;very high&quot; glucose</td>
<td>PmaltK</td>
<td>Maltotriose/mal/cAMP/cat.derepr</td>
<td>4 3 4</td>
<td>&gt;2 5</td>
<td></td>
</tr>
<tr>
<td>REFERENCE SYSTEM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fed-batch, high glucose</td>
<td>PlacIPT</td>
<td>Derepr. by IPTG</td>
<td>46 1.5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>fed-batch, low glucose</td>
<td>PlacIPT</td>
<td>Derepr. by IPTG</td>
<td>22 1</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. A summary of the production processes presented in this work, compared to a reference system 6 (Sandén et al. 2003). 1) Mechanism of formation of maltotriose through maltose metabolism, mal, or through glycolgen metabolism, gly. The formation of cAMP by catabolite derepression, cat. derepr. 2) The time at the maximal production rate, qPmax. 3) "med", "high" and "very high" feed of glucose followed by a constant feed of maltose, respectively. 4) The time from the induction by maltose, i.e. the time when the cell accumulation is resumed, referring to figure 5. 5) qPmax lasts through the process. 6) Induction at a specific growth rate of 0.5 h with 200μM IPTG. 7) Induction at a specific growth rate of 0.1 h with 200 μM IPTG.

Production processes based on the function of PmaltK, (II) have been summarised in table 4, according to productivity, induction mechanisms and duration of the maximal production rates. As seen in the table the production rates with the “low” feed might be distinguished by three different phases with a step-wise decrease in production rate, 13/6/3 U/mg.h, respectively. The reasons for the decrease is not fully understood, but the higher production level, as seen in figure 32, is believed to be the result of an increase in cAMP, that probably produced only when glycanogen levels are sufficiently low. To be able to further increase the production levels the exponential feed of glucose can be switched to a constant feed and thus result in a slow decrease in carbon/energy source, which might lead to constantly increased cAMP levels. This might result in a slow increase instead of a decreased production rate.
The introduction of maltose in the process results in the production of maltotriose in combination with cAMP formation at the event of catabolite derepression. In the table there is also a reference system based on \( P_{lacZ} \) production, with a \( lac^- \) host strain, on a low copy number plasmid and the same reporter protein for a comparison (Sandén et al 2003). This reference system is based on two identical fed-batch processes on glucose with the only difference in the time for induction, at “high” and at “low” feed, corresponding to a specific growth rate of 0.5h\(^{-1}\) and 0.1h\(^{-1}\), respectively. The systems based on \( P_{mfdK} \) are generally characterised by lower production rates compared to values obtained with \( P_{lacZ} \) as seen in the table. The only case where a higher production rate is seen is with the diauxic growth process, with a “high” preceding glucose feed, which result in an accumulation of 30% of the total protein reached as far as the process is run. The time to perform this production process depends on the diauxic phase and the preceding feed rate of glucose. Although this is a more complicated process to operate and control, overfeed leads to reversible loss of production (II). Furthermore, this production system can be designed so that growth never ceases.

**Transcriptional or translational effect?**

It has been concluded from the work in the reference system, table 4, that the translation is growth rate dependent due to different levels of ribosome accumulation at different growth rates. A high growth rate results in higher production of \( \beta\)-gal, due to the higher levels of ribosomes and the initial limitation in the synthesis machinery is set by translation (Sandén et al. 2003). The production based on \( P_{mfdK} \) results on the contrary in high production levels at low specific growth rates indicating that the ribosomal levels is not limiting during production at the low specific growth rates. Translation effects are thus not as clearly visible here as with reference system and production reached with \( P_{mfdK} \) are more likely explained by the metabolic events leading to a changed transcription and/or translation.

**PROCESS DESIGN TO CONTROL PRODUCT QUALITY**

Productivity is a very important parameter in recombinant protein production together with the equally important parameter, the product quality. Two quality aspects, inclusion body formation/solubility and proteolysis have been studied here. It was previously indicated that a high productivity was inversely proportional to a high product quality, due to post-translational modifications, as observed by the increased isoforms of the protein (unnamed) present the higher production rate (Ryan et al 1996). Furthermore, the activity of \( \beta\)-gal has shown to depend on the
growth rate and reached a maximal activity at a specific growth rate (Hellmuth et al 1994). This indicates that production systems where the production rates can be modulated are preferable to obtain optimal conditions for an acceptable quality and productivity. It is shown that the transcription rate can be tuned in the well known and strong pET based expression system through repressor titration and thus better match the metabolic capacity of the host cell (Striedner et al. 2003). The control of gene dosage, as discussed in the introduction of this chapter, will play an important role to obtain soluble proteins. This work presents production systems with the possibility to control the solubility of the model protein based on P$_{m&k}$, the weaker promoter and P$_{lacUV5}$, the stronger promoter, in a low copy number vector with different production rates following different amounts if inducer. The production processes used are based on “low”/”high” limited feed on glucose which previously have been established to be reproducible and stable according to productivity and cell growth (II). With the P$_{m&k}$ based processes maltose is furthermore added after four developed generations during limited growth on glucose.

**Inclusion body formation and proteolysis in production based on P$_{m&k}$**

It has previously been stated that reproducible production processes can be obtained in glucose limited fed-batch processes based on P$_{m&k}$ (II). In this work, the production process based on P$_{m&k}$ includes the production of the model protein Zb-MaIE (III) with and without the signal sequence of OmpA for secretion to the periplasm. The parameters that are suggested to control the quality of the model protein, are the substrate feed rates obtained by the feed profiles on glucose represented by a “low” and “high” feed and the secretion of the model protein to the periplasm. The production of Zb-MaIE resulted in very low detectable levels of full-length product because of a severe product degradation pattern. A Western blot is showing the degradation pattern obtained from the production of Zb-MaIE. The Western blot shown in figure 34 is blotted with antibodies against the Zb part of the fusion protein, but the same blot has also been blotted against MaIE antibodies for a comparison (not shown) and to be able to detect degradation products containing both parts, Zb and MaIE, respectively. The degradation products that are discussed here are seen with both antibodies and thus contains both the Zb and the MaIE part of the fusion protein. The full-length product was obtained when secreted to the periplasm both with “high” and “low” feed on glucose as seen in the figure, which is reflected by the milder stress response and reduced effect of diauxic growth as previously discussed. With the product remaining in the cytoplasm the full-length product is observed at “low” but not at “high” glucose feed. The major degradation products obtained in both systems are products with the approximated size of 33 and 38 kDa which have been detected in identical production processes based on P$_{lacUV5}$ (III, IV) not shown in the figure.
During protein production in E.coli there is always a degradation of proteins that belongs to the natural mechanism to regulate levels of native proteins and to eliminate damaged and abnormal proteins (Maurizi 1992), which is emphasised at low feed (Reeve et al. 1984). The in general low production where only a minority is the full-length product might be explained by a high and rapid degradation of the model protein. This high degradation seem to be due to the model protein since a similar degradation pattern is observed in identical production processes based on P\textsubscript{lacUV5} in the same low copy number expression vector (IV). One possibility to explain the low expression levels is by the limits from a weak promoter like P\textsubscript{malE} in combination with a low copy number plasmid (III).

\begin{center}
\includegraphics[width=0.5\textwidth]{figure34III.png}
\end{center}

\textit{Figure 34.(III) Western blot showing the product pattern obtained from production of the model protein Zb-MaIE. M=Molecular marker, kDa. R=Zb-MaIE purified on ion-exchange chromatography. g=four generations in glucose limited fed-batch. m=two hours after addition of maltose. Low=growth rate 0.2 h\textsuperscript{-1}. High=growth rate of 0.5 h\textsuperscript{-1}. Arrows to the right: The upper arrow indicate the full length protein, the middle arrow indicates the degradation product of approximately 38 kDa and the lower arrow indicate the degradation product of approximately 33 kDa. The Antibodies used is antiprotein A.}

\textbf{Inclusion body formation and proteolysis in production based on P\textsubscript{lacUV5}}

The solubility of Zb-MaIE and MalE31 can be controlled by different transcription/translation rates of the nascent polypeptide. The transcription rate is controlled by different amounts of IPTG and the translation is controlled by a “high” and a “low” feed rate of glucose in fed-batch process production processes (IV). By the different feed profiles used it was possible to control the productivity, as in production processes based on P\textsubscript{malE} (II). As suggested previously the control of production rate has a large impact on product quality as solubility and inclusion body formation, as was here shown by the addition of 4 different concentrations of IPTG, 5, 30, 100 and 300 μM. The results are shown in figure 35 and based on 15 fed-batch processes. As previously shown, in a diploma work at the department of Biochemistry and Biophysics, Stockholm University, the double mutant protein is more sensitive to proteolysis and thereof more difficult to express to high levels of soluble protein.
This is further concluded by the results obtained in this work (IV) where the double mutant protein results in lower levels of soluble product compared to Zb-MalE as seen in figure 35. In the figure there are some clear trends seen for both the production with Zb-MalE and Zb-MalE31. At “low” feed on glucose the highest levels of soluble protein are obtained. Inclusion bodies as in this case are preferred to be avoided are influenced by the transcription rate, a low amount of IPTG gives the lowest levels of inclusion bodies. These trends can be seen for both products.

![Figure 35](image)

**Figure 35 (IV).** The relation of soluble and insoluble product formation in 15 fed-batch processes. Zb-MalE corresponds to the eight processes to the left and Zb-MalE31 correspond to the eight processes to the right, as indicated in the figure. The y-axis shows a relative value in relation to a reference product of Zb-MalE (IV). The results are obtained from scanned Western blot membranes blotted against anti protein A antibodies. By the use of the software program Image Master the intensity of the bands could be measured, which are shown as bars in this figure.

During these fed-batch processes samples were furthermore taken to investigate the sensitivity to proteolysis. A sample from the bioreactor was withdrawn and diluted with culture medium to obtain an optical density of about 3-4 (OD/ml). A “zero” sample was taken before the protein synthesis was stopped with chloramphenicol and thereafter samples were taken every 5 minutes to study the degradation pattern of the full-length protein. Thus, this method will detect the proteolytic sensitivity of the already synthesised full-length protein. No degradation could be observed with Zb-MalE, (IV). However, the double mutant protein, Zb-MalE31 showed to be sensitive to proteolysis as seen in figure 36. The figure shows the degradation from samples taken from production at “high” and “low” feed with the different amounts of IPTG. The 100 % value of the y-axis corresponds to the initial sample taken without the treatment of chloramphenicol. From these data it was concluded that proteolysis seems to depend on the degree of limitation of glucose, where there is a faster degradation of samples taken from a “high” feed process compared to the “low”
Design of substrate induced transcription for control of recombinant protein production

feed process. The specific synthesis rate, from different IPTG amounts, does not seem to have a specific impact of the degradation rate as seen in the figure. With a “high” feed rate the degradation is linear during the 45 minutes of samples taken while with a “low” feed rate it takes 10-20 minutes before the degradation is showing a linear decrease.

In conclusion, the proteolysis is influenced by the substrate feed rate, where a “low” feed on glucose seem to decrease the proteolysis. The solubility of the model protein can be influenced by the substrate feed rate and by the transcription rate as obtained from the different amounts of IPTG.

Figure 36. Proteolysis of ZbMalE31 taken from a fed-batch process with a “high” and a “low” limited feed of glucose, respectively. The values of the y-axis is taken relative to the amount of product before the start of the proteolysis procedure (IV). The results are obtained from scanned Western blot membranes blotted against anti protein A antibodies. By the use of the software program, Image Master the intensity of the bands could be measured, which are shown as the dotted curves in this figure.
CONCLUSIONS

The hypothesis that the rate of synthesis of the nascent polypeptide is an important factor controlling inclusion body formation, proteolysis and translocation capacity was confirmed by the production of two model proteins Zb-MalE (III, IV) and the double mutant protein, Zb-MalE31, respectively, (IV). The two proteins represent here one easy produced product and one considerably more difficult product to produce, i.e. corresponding to the general characteristics of the still unidentified group of proteins of the sequenced genome of different organisms. The proteolysis of the difficult protein was decreased by production at a “low” substrate feed rate compared to a “high” substrate feed rate based on P\textsubscript{lac}U\textsubscript{53} (IV). Also, the solubility of the difficult protein was influenced by the substrate feed rate, where the highest solubility is obtained with a “low” feed rate, where both proteolysis and inclusion body formation are translationally controlled. However, the inclusion body formation can also be controlled by transcription rate when the lowest rate resulted in a much higher production of soluble protein. In conclusion, a low substrate feed in combination with a low gene dosage resulted in higher solubility and a lower proteolysis, even for the difficult protein.

In the present investigation a new promoter for recombinant protein production is suggested. However, promoters that are well suited for recombinant protein production are characterised by certain criteria to reach high product quality and productivity: 1) A promoter should allow low basal levels at the uninduced state. 2) The relation between inducer and production level should be clearly visible, meaning that the production level should not be sensitive to small variations of inducer addition. 3) The promoter should be able to recruit RNA polymerase during the limited growth conditions of industrial fed-batch cultivations. 4) The induction kinetics of a promoter is important and depends on how fast the promoter is activated. Promoters where the transcription rate is increased within hours instead of minutes are preferred to match subsequent post-translational steps. 5) The inducer should be inexpensive especially in large scale production. 6) A toxic inducer is to be avoided.

Production processes based on P\textsubscript{malK}, and P\textsubscript{lacU\v{S}}, have shown to fulfill some of these criteria. Despite that the promoter is substrate induced the production process based on P\textsubscript{malK} can be separated according to growth and production by the mechanisms of inducer exclusion and catabolite repression. The growth phase can be obtained during growth on excess of glucose and the
production phase can be obtained by the use of the glucose and/or maltose fed-batch technique. This growth and production separated process would be possible to obtain in any process based on a promoter controlled by catabolite repression, such as the lac and araBAD promoters. Production processes based on \( P_{\text{lacUFS}} \), which is not catabolite repressed, can also be separated in a growth and a production phase when a high enough concentration of repressor molecule is produced before the addition of IPTG, to avoid too high basal levels.

The basal protein levels when \( P_{\text{malK}} \) is catabolite repressed are obtained during a batch phase on glucose and were 2 U/mg dry weights with \( \beta \)-galactosidase as model protein, which corresponded to about 0.4% of the total protein production. The basal levels with a batch production on maltose corresponded to 2% of the total protein production. The basal levels are important when producing a toxic product. \( P_{\text{malK}} \) has qualities to be used in production systems when a too high production rate will harm or even kill the cells. To realise a process for a toxic product a process with an immediate start of a limited glucose feed is advised, which will reduce the initial rapid increase of production. A high glucose feed rate is further suggested that will reach a protein level of 6% of the total protein production.

Since \( P_{\text{malK}} \) is catabolite repressed it is activated and strong during limited growth, which is one of the criteria for a suitable promoter. Furthermore, since glucose and/or maltose are both inducers and carbon sources during production they will stand in proportion to the availability of carbon source. This will thus give better possibility for a balance of induction level, carbon/energy uptake and metabolism. The inducer is not toxic and it is not expensive since a mixture of glucose and maltodextrins can be used, due to the inducer exclusion. This mixture is furthermore cheaper than glucose and more transport stable. The substrate induced promoter, \( P_{\text{lacUFS}} \) has also shown to work well during limited growth conditions. A disadvantage is the use of IPTG, which is toxic to humans. The induction level is with a wild type host strain very sensitive to small variations of the added IPTG amount. In this work this is avoided by the use of a lac strains where IPTG diffuses into the cell instead of being actively transported.

\( P_{\text{lacUFS}} \) and \( P_{\text{malK}} \) show different kinetics during production, which can be explained by the mechanisms to activate the promoters, respectively. Except for cAMP, \( P_{\text{malK}} \) also needs activated MalT, while \( P_{\text{lacUFS}} \) needs the release of the repressor to start the transcription. These two different mechanisms result in different times to reach the maximal rate of production as shown in Table 4. In general the time to reach the maximal production rate is longer with production based on \( P_{\text{malK}} \).
is suggested to be important when producing protein sensitive to a fast production rate to avoid inclusion body formation and proteolysis. From the present data conclusions about the more favourable induction kinetics on product quality by production based on \( P_{milk} \) can not be drawn due to the too low production levels with the low copy number vector. A high copy number vector is advised for further work. However, the production of the secreted product to the periplasm resulted in a balanced metabolism not reached in other systems.

It is impossible to know only from the primary structure of a protein how a successful production system should be designed to produce an active and soluble protein. Therefore it would be advantageous to have a set of different production systems where the production conditions can be tested for each protein. In this work we have shown different production systems based on \( P_{milk} \) and \( P_{lacU75} \) respectively. These two promoters \( P_{milk} \) and \( P_{lacU75} \) represent a weak, catabolite repressed promoter, and a strong promoter, independent of the catabolite repression, respectively. The parameters to change the production conditions in production processes based on \( P_{milk} \) are two specific growth rates on glucose which can further be enhanced by maltose addition. The protein can be accumulated in the cytoplasm or in the periplasm, respectively. The parameters to change the production conditions in production processes based on \( P_{lacU75} \) are the specific growth rates and different amounts of IPTG. This leads to a set up of different combinations of production systems, which can rapidly be realised with multiparallel reactors. Moreover, a change in temperature would result in the double amount of production systems. With the designed production protocols, this multiparallel production system, results in different production conditions with variations in production rate, duration of production, translocation of product and temperature. This system leads to a randomisation in the selection of a process able to establish soluble protein in amounts enough for, e.g. structure and function determination. It is realised that this could also constitute a screening format for process development to pinpoint the best production system for a final protein product.
ABBREVIATIONS

ara  arabinose
ATP  adenosine monophosphate
cAMP cyclic adenosine monophosphate
CAT  chloramphenicol-acetyl-transferase
CRP  catabolite repressor protein
DNA  deoxyribonucleic acid
DOT  dissolved oxygen tension
E.coli  Escherichia coli
F  feed
GST  glutathione-S-transferase
HGH  human growth hormone
HTPP  high throughput protein production
HTS  high throughput screening
HUGO  human genome
IPTG  isopropyl-D-thiogalactoside
kDa  kilo Dalton
Leu  Leucine
LTPP  low throughput protein production
Lys  lysine
mal  maltose
MalE  maltose binding protein
MalE31  double mutant protein of MalE
mRNA  messenger RNA
ori  origin of replication
PCN  plasmid copy number
P_{lacUV5}  a derivative of the promoter of the lac operon
P_{MalR}  the promoter of the malK-lamb-matM and malE-F-G operons
PEP  phosphoenolpyruvate
Phe  phenylalanine
PHO  phosphate
Pi  organophosphates
Pn  phosphonates
ppGpp  guanosine-3,5-tetraphosphate
PTS  phosphoenolpyruvate-phosphotransferase system
q_c  specific production rate (U/mg.h)
Q_p  production rate (U/h)
RNA  ribonucleic acid
SD  Shine Dalgarno
tRNA  transfer RNA
TIR  translation initiation region
Trp  tryptophan
TRX  thioredoxin
Tyr  Tyrosine
V  volume (l)
X  cell dry weight (g/l)
Y_{ox}  yield of oxygen/substrate molecule (g/g)
Zb  Zbasic, the identification/purification tag
β-gal  β-galactosidase
μ  specific growth rate (h^{-1})
μ_{max}  maximal specific growth rate (h^{-1})
ACKNOWLEDGEMENTS

First of all I would like to thank all people that in some way has contributed to this work and I would especially like to thank the following:

Associate professor Gen Larsson, my supervisor, for being an inspiring and enthusiastic source with an incredible energy for new and exciting ideas. I made it thanks to you!
Thank you, professor Sven-Olof Enfors for reading my thesis and giving me good advice. Also for accepting me as a PhD student and being open for discussions.

Thanks to all senior researchers at our department, professor Lena Häggsström for very interesting courses and discussions, associate professor Andres Veide for reading through my thesis and giving me some good advice and always having time for discussions and being a party “höjare”.

I would like to thank Dr Sophia Hober for reading my thesis and for giving me a lot of support during the time of “never ending cloning and protein purification work”.

My Hedhammar for a lot of help with cloning and protein purification.

Thanks to the co-workers within the Centre of Bioprocess Technology and especially, Ioannis Tubulekas and Erik Holmgren at Biovitrum, for good advice and excellent help with construction of new expression vectors.

I would like to thank all my former colleagues and especially, Eric, for ALWAYS being available, Fredrik, for being helpful during my start up time as a PhD student, Atefeh for a good collaboration and remaining cool, no matter what.

A big hug to all colleagues at the department for making the environment nice and friendly and especially, Anna Maria for a good collaboration and for helping us through hard times of cloning, cultivations and for all chats, Katrin for a good collaboration with Western blots and ion-exchange and for keeping things in order. Thank you Ulrika for the jogging tours in Haga Parken, Eva for your organised mind, Ingrid for being helpful and positive, Helene for pep talk and cleaning my eye away from E.Coli, Erika, for the Yogi tea idea, it helped a lot when I was writing my thesis, Marie for many fruitful discussions, Bato for the “måndags kramar”.

I would furthermore like to thank Tommy having a technical solution on urgent problems, Gunilla for her positive spirit and for administrative help and “LaGunilla”, Marita for giving a smile and help with administrative stuff, Ela for the analysis from the multi cultivations and talks during these years.
Ulf for being an inspiring source, for always having fun ideas and for the help in photographing.

Many thanks to all my friends outside the world of Biotechnology, and especially,
Mikaela, for being there in “ur och skur”, you are a sun!
Petra for the help to chose the right education and for a true friendship,
Richard for lunches and peptalk.

Thousand thanks to my family, mam and dad for all the support and believing in me during difficult times and making me the person that I am today.
My brother **Mieke** for the never ending support and being there for pep talk.

My brother **Christoffer** being positive to life, no matter what, and bringing that feeling to me.

**Per Oscar** for being a part of my life! Thank you for believing in me when I did not! I am grateful for every minute I spend with you! You are the one!
REFERENCES


Design of substrate induced transcription for control of recombinant protein production


Harrison R G “Expression of soluble heterologous proteins via fusion with NasA protein.” in Novations 11: 4-7

Hayduck F (1919). "Verfahren der hefefabrikation ohne oder mit nur geringer alkoholerzeugung." German patent 300662.


Lendenmann U and Egli T (1995). "Is Escherichia coli growing in glucose-limited chemostat culture able to utilize other sugars without lag?" Microbiol141: 71-78

59
Design of substrate induced transcription for control of recombinant protein production


Design of substrate induced transcription for control of recombinant protein production


