_____
Unterschrift des Betreuers

TECHNISCHE UNIVERSITÄT WIEN

DIPLOMARBEIT

Mechanistic Platform Knowledge of Concomitant Sugar Uptake in *Escherichia coli* and Analysis thereof

Ausgeführt am Institut für
Verfahrenstechnik, Umwelttechnik und technische Biowissenschaften
der Technischen Universität Wien

unter der Anleitung von
Dipl.-Ing. Dr.nat.techn. Oliver Spadiut
und Dipl.-Ing. David Johannes Wurm

durch Johanna Hausjell
Marktgasse 49/15, 1090 Wien

_____
Datum

_____
Johanna Hausjell

Danksagung

Zunächst möchte ich mich an dieser Stelle bei allen bedanken, die mich bei der Durchführung und beim Verfassen dieser Diplomarbeit unterstützt und motiviert haben:

Mein besonderer Dank gilt Prof. Christoph Herwig, der mir die Durchführung dieser Masterarbeit in seiner Arbeitsgruppe ermöglicht und die notwendigen Rahmenbedingungen zur Verfügung gestellt hat.

Weiters möchte ich mich herzlich bei Dr. Oliver Spadiut vor allem für seine hervorragenden Instruktionen in Bezug auf Durchführungsstrategien und Versuchsplanung bedanken.

Ebenfalls herzlich bedanke ich mich bei Dipl. Ing. David Wurm für Anleitung, Hilfe und Erläuterungen im praktischen Arbeiten sowie bei der Datenauswertung.

Mein weiterer Dank gilt allen ArbeitskollegInnen, die mir während der Durchführung der Arbeit vor allem im Labor behilflich waren.

Zuletzt möchte ich mich noch ganz besonders bei meinen Eltern und meiner Schwester bedanken, die mir durch ihre Unterstützung mein Studium ermöglicht haben.

Abstract

When producing recombinant proteins *Escherichia coli* strain BL21(DE3) in combination with the T7-based pET-expression system is often the method of choice. Recently, we proposed a new induction strategy based on lactose, where the level of lactose uptake was shown to be strictly coupled to the abundance of glucose. Based thereon we developed a mechanistic model to determine the correlation of the concomitant sugar uptake and showed that establishment of the correlation can be used to gear productivity and product location through the level of lactose uptake.

In this study we substantiated applicability of the established correlation for concomitant glucose and lactose uptake as platform knowledge for BL21(DE3) cells with pET-plasmids: we show for the first time that the correlation of concomitant sugar uptake exhibits the same trend in *E. coli* BL21(DE3) cells expressing different products, only absolute values diverge. Furthermore, we developed a strategy for fast and accurate determination of this correlation. This study will help all scientists working with recombinant *E. coli* strains, as not only mechanistic platform knowledge for *E. coli* strains carrying the pET expression system is revealed, but also a strategy for fast determination of the mechanistic correlation is offered.

Zusammenfassung

Eines der am häufigsten angewendeten Expressionssysteme bei rekombinanter Proteinproduktion in *Escherichia coli* ist das T7-basierte pET-Expressionssystem. Kürzlich stellten wir im Rahmen einer Studie eine neue Induktionsstrategie basierend auf Laktose vor, in der sich zeigte, dass die Laktose-Aufnahme streng von der verfügbaren Menge an Glukose abhängig ist. Darauf basierend entwickelten wir ein mechanistisches Modell zur Bestimmung der Korrelation von gleichzeitiger Zucker-Aufnahme und zeigten, dass die Bestimmung dieser Korrelation die Möglichkeit bietet, Produktivität und Produktort über die Höhe der Laktose-Aufnahme zu steuern.

In dieser Studie untermauern wir die Anwendung der entdeckten Korrelation für gleichzeitige Glukose- und Laktose-Aufnahme als Plattform-Wissen für BL21(DE3) Zellen mit pET-Plasmiden: Zum ersten Mal zeigen wir, dass die Korrelation für gleichzeitige Zuckeraufnahme den gleichen Trend aufweist, in *E. coli* BL21(DE3) Zellen, die verschiedene Produkte exprimieren. Die absolut Werte der Laktose-Aufnahme sind jedoch unterschiedlich. Außerdem entwickelten wir eine Strategie für schnelle und verlässliche Bestimmung der Korrelation. Diese Studie kann allen Wissenschaftlern, die mit rekombinanter Proteinproduktion in *E. coli* Stämmen arbeiten, behilflich sein, da einerseits mechanistisches Plattform-Wissen für *E. coli* Stämme mit pET-Plasmiden präsentiert wird und andererseits eine Strategie zur schnellen Bestimmung der mechanistischen Korrelation vorgestellt wird.

Table of Contents

Introduction

1. Antibodies, Antibody Fragments and IgY

In 1986, the first recombinantly produced immunoglobulin, Orthoclone OKT, for human therapy was approved. Ever since, the demand for antibodies and antibody fragments is drastically increasing reaching levels of more than 1000 kg/year. To meet this high demand rapid and efficient production is a must [1,2].

Antibodies are used for the treatment of various diseases including cancer, viral infection or inflammation. Their popularity can be attributed to their outstandingly high selectivity and low toxicity [1]. In recent years especially engineered antibodies as well as antibody fragments are gaining ground [3].

Structure of Antibodies and Antibody Fragments

In general, all immunoglobulins exhibit the same basic structure: two heavy chains (V_H) on the scale of about 50-77 kDa are connected via one or more disulphide bonds. Furthermore, each of those heavy chains is joined to a light chain (V_L) of about 25 kDa [4]. Full length antibodies consist of the variable region (Fv) and the constant region (Fc). The constant region serves stabilisation and long serum half-lives by Fc-Receptor-mediated recycling. Additionally, the Fc is responsible for antibody-dependent cell-cytotoxicity and complement-dependent cytotoxicity [3,5]. Antibody fragments lack, depending on the type, either the entire Fc or the bigger part of it (Figure 1).

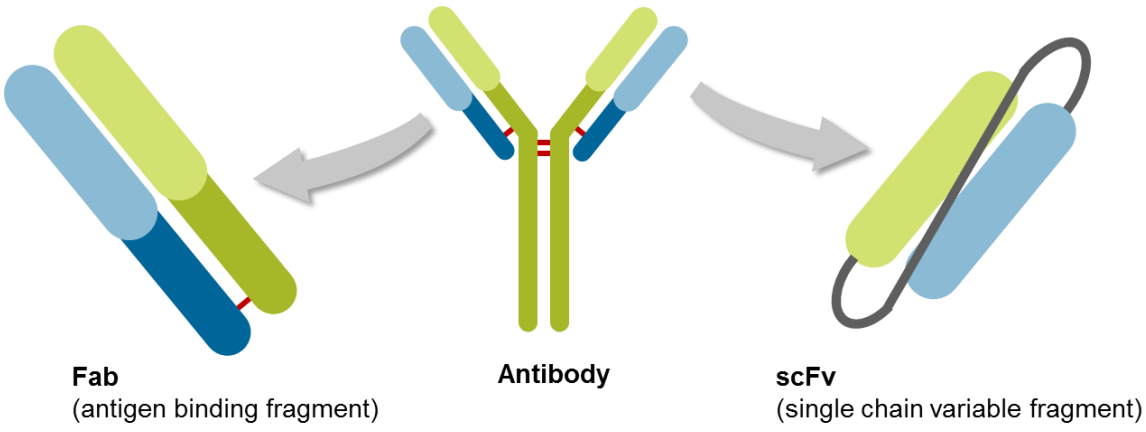


Figure 1 Left: Antigen binding fragment (Fab) consisting of only the antigen binding region of the full-length antibody. Middle: Full length antibody with heavy chains (green), light chains (blue), constant region (dark green and blue), variable region (light green and blue) and disulphide bonds (red). Right: Single chain variable fragment consisting of only the variable region of the full length antibody connected by a linker (grey).

The two most prominent types of antibody fragments are antigen binding fragments (Fabs) and single chain variable fragments (scFvs). The former consist of the antigen-binding region of the

antibody, while the latter comprise only the variable region of the heavy and the light chain connected by a linker [5]. In case of scFvs the length of the linker is crucial as it must not affect correct folding of the antigen binding domain [6].

Recombinant Production of Antibodies and Antibody Fragments

Full length antibodies exhibit post translational modifications including glycosylation patterns which avert fruitful expression in easily cultivatable organisms as *Escherichia coli*. Therefore, expression of full length antibodies is usually carried out in mammalian cell cultures, most popularly Chinese hamster ovary cells (CHOs) [7-9]. They can perform human like N-glycosylation, which impedes the hazard of immunogenic responses. However, cell-cultures entail many drawbacks such as complex and expensive media as well as shear sensitivity [7]. Furthermore, fermentations with batch and subsequent fed-batch phases, the way they are carried out in standard cultivation strategies, are problematic when using CHOs as gradually increasing amounts of lactate and ammonia are accumulated, which have inhibiting effects. In contrast the alternative, continuous culture, elevates the risk of contamination [10].

In the case of antibody fragments the smaller size as well as the lack of post translational modifications allows the facile expression in prokaryotic hosts such as *E. coli*, *Pichia pastoris* or *Saccharomyces cerevisiae* [10,11].

In 1988 the first antigen-binding fragment was successfully produced in *E. coli*. The major break-through was achieved by secreting the variable domains together into the oxidizing environment of the periplasm, where disulphide bonds could be formed [12]. However, periplasmic protein production entails difficulties as accumulation of too much protein may result in interference with normal cell functions and can thus lead to cell death [13]. Single chain variable fragments on the other hand do not have to be secreted into the periplasm as they consist solely of the variable region connected by a linker. This allows successful expression in the cytoplasm [12] as long as the linker does not contain disulphide bonds.

Manufacturing of Antibody Fragments

Antibody fragments can be manufactured comparatively simply and inexpensively. Initially antibodies were treated with pepsin or papain to cleave them proteolytically, and obtain the fragments [3]. For recombinant production an exemplary approach would be isolation of mRNA from hybridoma, followed by reverse transcription into cDNA and subsequent diverse PCR-techniques. This leads to libraries of various V_H and V_L genes. Consequent screening for suitable scFvs can be performed for instance by phage display [6].

Properties and Applications of Antibody Fragments

The reduced proportions of antibody fragments lead to outcompeting full-length antibodies in regard to tumour localization and faster blood clearance [3]. This is reflected in biopharmaceutical uses of antibody fragments in clinical or preclinical trials, which include numerous applications for cancer treatment including breast, ovarian and colorectal cancer [3]. These countless studies for cancer-treatment by antibody fragments is also a result of their excellent capacity in discriminating even low amounts of tumour-associated antigens [3]. However, at the same time the compact size of the fragments is a disadvantage when it comes to tumour retention. For circumventing this drawback often multivalent antibody fragments are used, which combine the major requirements when it comes to cancer treatment: rapid localisation of the tumour and high target retention followed by quick blood clearance [3].

Nevertheless, applications of scFvs and Fabs are not limited to biopharmaceuticals, but also include bioseperation techniques, purification and immune detection [10]. In bio-imaging, antibody fragments have been shown to be superior to full-length antibodies as their half-life is shorter resulting in better contrast [3].

IgY and Celiac Disease

Within this Thesis the expression of a tandem-scFv derived from IgY was conducted in *E. coli*. IgY is an antibody occurring in birds and reptiles and is related to mammalian IgG. In comparison to IgG it holds an additional constant domain and is believed to fulfil functions that are executed by IgG and IgE in mammals: Similar to IgG it is involved in complement fixation and opsonisation and in the same way as IgE it performs tissue sensitizing [14].

In recent years some research groups engaged in investigating the potential of anti-gliadin IgY for treatment of celiac disease [15,16]. Currently 1 % of the world´s population suffers from celiac disease. The chronic illness is caused by an abnormal immune response to dietary gluten and related prolamines, which mainly occur in wheat, barley and rye. Consumption of gluten, when suffering from celiac disease, can result in inflammation, villous atrophy, and crypt hyperplasia in the upper small intestine. So far the only way to circumvent those symptoms is a gluten free diet [15-17], which can be challenging for individuals suffering from the disease. This is why there is a need for treatment agents.

Gujral, Suh and Sunwoo created a chicken-derived anti-gliadin IgY by immunization with Sigma gliadin and showed that the obtained antibody could neutralize negative effects of toxic PT gliadin on Caco-2 cells [15]. Stadlmann generated a PT-binding scFv from RNA derived from splenic lymphocytes of immunized chicken suitable for recombinant expression [16]. An *E. coli*

strain carrying a plasmid for expressing a tandem version of this scFv was used in the course of this Thesis.

2. *Escherichia coli*

The organism *Escherichia coli* was first described in 1885 by the paediatrician Theodor Escherich. The gram negative, rod shaped, facultative anaerobe bacteria are part of the Enterobacter species and naturally reside in the intestine of mammals [18]. During the 20th century *E. coli* became one of the most well-studied bacteria, serving as a laboratory workhorse and model organism: In 1982, the first pharmaceutical using recombinant DNA was approved: human insulin produced in *E. coli* [19,20]. In 1997, *E. coli* was one of the first organisms with a fully sequenced genome, the project took 15 years [21]. In a long time period afterwards, *E. coli* was the preferred host for production of recombinant proteins, amongst others numerous biopharmaceuticals [22]. However, as antibodies and more complex molecules gained increasing popularity, mammalian cells took the place of microbes as they were more suited for expression of such [20,22]. Nevertheless, up to date about 30 % of approved therapeutic proteins are produced in *E. coli* [23]. In recent years the organism regained its attractiveness in particular, as antibody fragments, which can easily be produced by microbes, are gaining ground [10,11].

3. Recombinant Protein Expression in *Escherichia coli* - Advantages and Drawbacks

Advantages of using *E. coli* for recombinant protein production include extensive knowledge about its genome coming along with a variety of established methods for genetic modification, multiple different engineered strains as well as a stunning assortment of expression plasmids (which differ mainly in copy numbers, promoters, selection markers or affinity tags) [24].

Beyond that, *E. coli* is an unmatchedly fast growing organism. Given optimal conditions, the doubling time is about 20 min. Although, it should be mentioned that this number decreases when expressing recombinant protein as a metabolic burden is imparted on the bacteria [24]. Nevertheless, *E. coli* can be cultivated up to high cell densities (more than 100 g/L) on relatively inexpensive media. High cell density cultures entail advantages including increased cost effectiveness arising from lower production costs and reduced investment in equipment as less raw material is needed. All of this results in economical and rapid gain of large quantities of recombinant protein [25].

However, utilizing *E. coli* for recombinant protein production entails difficulties: The cells are, in contrast to higher organisms, not able to perform post translational modifications. Thus, for instance no glycosylation is realised and disulphide bond formation is not possible in the

cytoplasm [24,26]. The latter deficiency can be overcome by using signal peptides allowing protein-transport to the oxidizing environment of the periplasm or by using certain strains such as Origami, which carry mutations leading to enhanced disulphide bond formation in the cytoplasm [27,28].

A well-known occurrence of recombinant protein synthesis in *E. coli*, however, is the formation of insoluble, inactive protein aggregates, so called inclusion bodies [24,29]. Inclusion bodies are often a result of high translation rates, which exhaust the folding machinery of the organism, resulting in unfolded as well as misfolded proteins which tend to aggregate [28]. Besides this inclusion bodies are often formed if the expressed protein is not congenial to *E. coli* and requires, for instance, post translational modifications such as the afore mentioned disulphide bond formation [30].

Once inclusion body formation occurs, obtaining active proteins from aggregates requires the cumbersome and labour intensive process of solubilisation and subsequent refolding, which often results in low recovery yields [30-32]. Another drawback coming along with solubilisation and refolding is that strategies are often empiric and a matter of trial and error[30].

4. *Escherichia coli* strain BL21(DE3)

When using *E. coli* for recombinant protein production, BL21(DE3) is often the strain of choice [33-35]. The strain under the name *E. coli* B occurs for the first time in a publication dating back to 1942 by Delbrück and Luria [36]. During the time thereafter the B-line was mainly used for studies of phage sensitivity, restriction systems, mutagenic assays, and bacterial evolution [37]. Nowadays *E. coli* strain BL21(DE3) is particularly popular when expressing recombinant proteins. It lacks the Lon protease, responsible for degrading foreign protein, and the OmpT protease, an extracellular enzyme for degrading proteins that can be especially problematic in case of cell-disruption, which is often carried out in order to obtain the intracellularly expressed products [24].

E. coli BL21(DE3) in contrast to BL21 carries the DE3 prophage in its genome: The T7 polymerase is located downstream of the *lacUV5* promoter [38]. Thus the strain is optimally suited for recombinant protein production in combination with the T7-based pET-expression system.

5. T7-Expression System, Lac-Operon and Carbon Catabolite Repression

The T7-based pET-expression system takes advantage of the polymerase stemming from bacteriophage T7, which is extremely specific to the T7 late promoter sequence that does not naturally occur in the genome of *E. coli* [39]. *E. coli* BL21(DE3) carries a copy of the T7

polymerase in its genome under control of the *lacUV5* promoter. Thus upon induction T7 polymerase gets transcribed and translated. Subsequently it is able to initiate transcription from the T7 promoter, which is in this case located on the pET plasmid upstream of the target gene [40]. This system entails two major advantages: Firstly, the T7 promoter is considerably stronger compared to every known native *E. coli* promoter [41]. Secondly, this system allows stringent regulation of transcription by co-expression of T7 lysozyme, that is able to degrade basal levels of T7 polymerase [42].

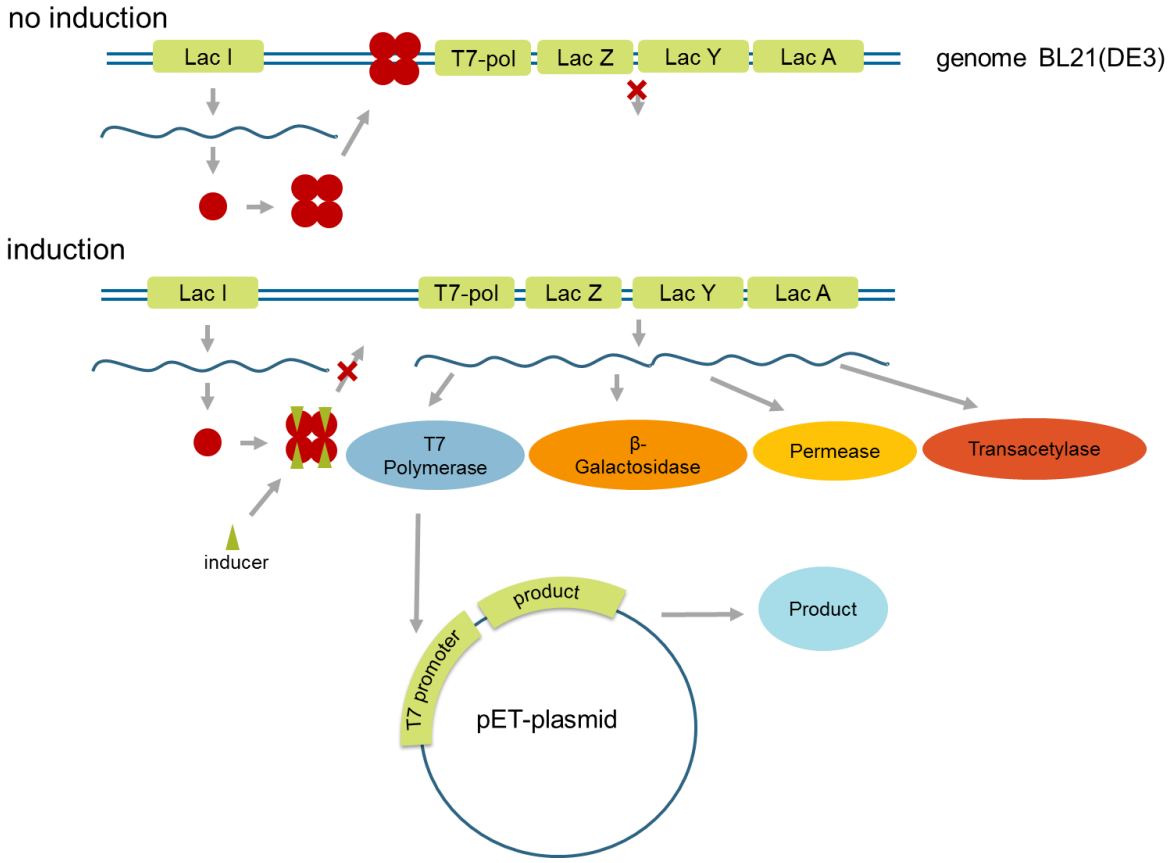


Figure 2: Scheme of the T7 based pET-expression system in combination with BL21(DE3) cells. Above: Under non-inducing conditions the lac repressor forms a tetramer and binds to the operator region. Transcription from the lac operon is not possible as the lac promoter is inaccessible to RNA-polymerase. Below: As inducer binds the lac repressor it can no longer attach to the operator region upstream of the lac operon. Genes located on the operon get transcribed. In BL21(DE3) cells the non-native gene encoding T7 polymerase is also transcribed. As the polymerase is translated transcription from the T7 promoter located on the pET plasmid upstream of the product can be initiated.

As shown above (Figure 2) transcription of the T7 polymerase is controlled by the lac operon. The wild type operon consists of three genes: *lacY* (β-Galactosidase), *lacZ* (β-Galactosid-Permease) and *lacA* (β-Galactosid-Transacetylase). In *E. coli* strain BL21(DE3) also transcription of the T7 polymerase is under its control. The expression is regulated by the lac promoter, which is subject to repression as well as positive transcription regulation: LacI, the lac repressor protein, is responsible for negatively regulating transcription. It binds as a tetramer to the operator region, resulting in a DNA loop which averts successful transcription as the

DNA is no longer accessible to RNA-polymerase [43]. In the presence of inducers such as lactose or synthetic analogues, most popularly Isopropyl-β-D-thiogalactopyranoside (IPTG) or alternatively thiomethyl- β -D-galactoside (TMG), those substances bind to LacI, which in turn can no longer attach to the operator region. Consequently, the promoter region is no longer coiled up which facilitates attachment of the RNA-polymerase [44,45].

Carbon catabolite repression influences transcription from the promoter by the mechanism of inducer exclusion: If glucose is present, Enzyme II A (EIIA) is dephosphorylated. The dephosphorylated form is able to inhibit lactose permease and thus also lactose transport into the cell. In succession LacI is unbound and able to attach to the operator region leading to a coiled up promoter region from which transcription cannot be initiated by RNA-polymerase [44-46].

Transcription from the operon is positively regulated by a complex of cyclic adenosine monophosphate receptor protein (Crp) and cyclic adenosine monophosphate (cAMP). The absence of glucose leads to a phosphorylation of EIIA, which activates adenylate cyclase and thus promotes formation of the Crp-cAMP complex. This complex then binds to the promoter region and presents the DNA to proteins such as RNA polymerase, facilitating attachment and subsequent transcription [45].

As mentioned, in BL21(DE3) the transcription of the T7 polymerase is regulated by the *lacUV5* promoter. This promoter is a mutated version of the wild type lac-promoter, which makes the promoter less susceptible to carbon catabolite repression as it allows transcription initiation in absence of the Crp-cAMP complex [47]. However, regulation of the operon is still dependent on Crp as it initiates, in conjunction with glucose, transcription of *ptsG*, the gene encoding Enzyme II BC (EIIBC). EIIBC is the phosphoryl acceptor for EIIA [45].

6. Soluble Protein vs. Inclusion Bodies: Strategies for Production Optimization

A well-known occurrence when using *E. coli* as host is the formation of insoluble and mostly inactive protein aggregates, so called inclusion bodies [30,31]. Expression of the product in inclusion bodies is often unwanted as active protein can be regained, but solubilization and refolding procedures are necessary, which are cumbersome and often lead to low recovery yields [30,31]. On the other hand, in certain cases inclusion body formation might as well be of use: the protein aggregates are, although inactive, highly pure, less likely to be affected by proteases and lead to higher initial yields [25,48].

If the product is expressed as soluble protein or in the form of inclusion bodies mainly effects cell physiology and subsequent downstream processing [30]. Process steps for inclusion body processes or processes designed for acquiring soluble protein respectively are shown in Figure 3.

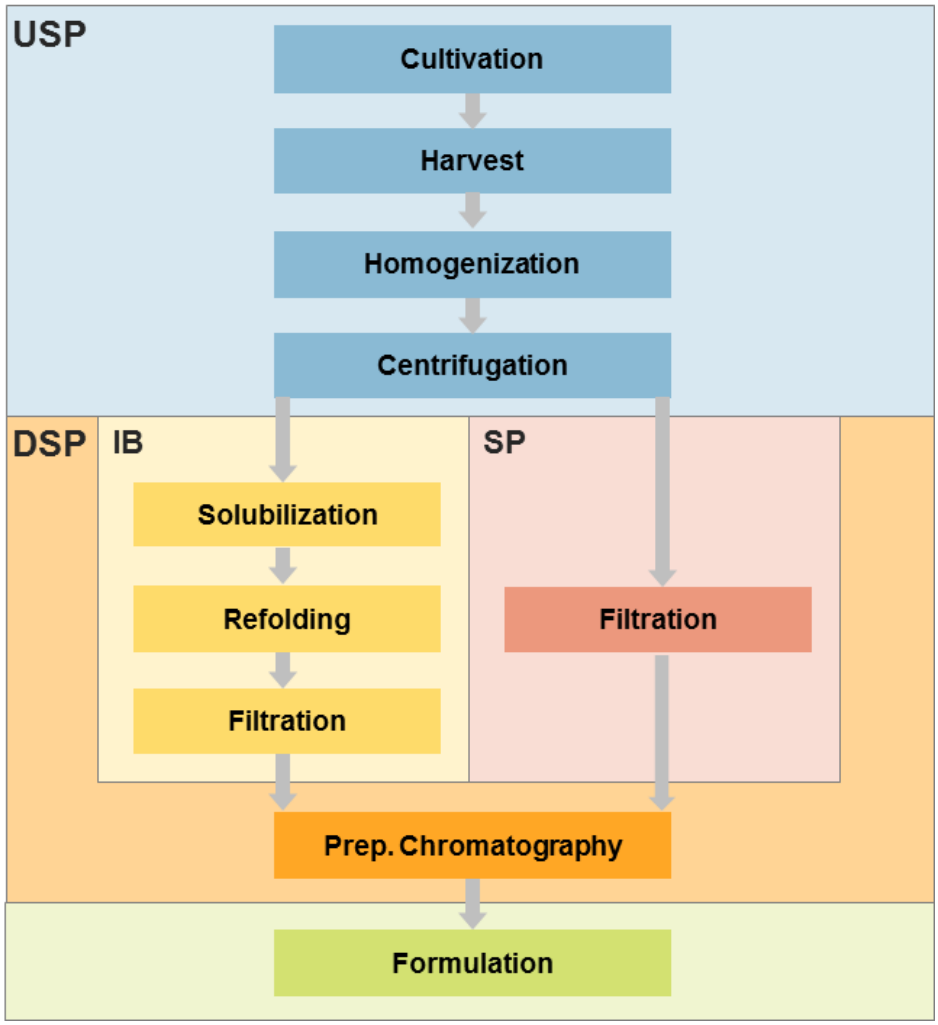


Figure 3 Schematic Process flow for upstream process (USP) and subsequent downstream process (DSP). Different DSP-steps when product is obtained in inclusion bodies (IB) or as soluble product (SP)

After disrupting the cells, soluble protein is separated from cell debris and inclusion bodies by centrifugation [49]. The consequent process steps depend on the product location.

When the product is expressed in inclusion bodies downstream processing requires solubilization and refolding [30-32]. At first inclusion bodies have to be solubilized, which is usually conducted by application of chaotropic reagents such as urea and guanidine hydrochloride or detergents such as SDS. Those reagents break up hydrogen bonds leading to fully denatured proteins. Additionally, reducing agents are added in order to maintain disulphide bonds in the reduced state. Reagents used include β-Mercaptoethanol or dithiothreitol. In order to avoid reoxidation by air, which is metal-catalysed often agents like

EDTA are added, which chelate such metals [31]. Subsequent refolding is conducted by removal of chaotropic reagents. On the small scale it is usually performed by dilution, which leads to reduced concentrations of chaotropic reagents and thus renaturation of the protein. Renaturation gets increasingly complicated with more disulphide bonds present as the risk of formation of incorrect bonds is raised. Additionally, aggregate formation is a potential jeopardy, which often occurs during refolding [31,32].

When the product is expressed as soluble protein, these steps can be omitted. However, when the yield of inclusion bodies is sufficiently higher compared to the yield of soluble protein in the upstream process, the more elaborate downstream processing is most likely worth the effort.

Inclusion body formation, or formation of soluble product on the other hand, can be influenced by either usage of genetic engineering or process-technological means [26,28,49,50].

Genetic optimization strategies

Genetic optimization can be performed by changing the expression vector, the construct or the host strain or by optimizing the sequence of the target protein [26].

Expression Vector and Construct

For *E. coli* a wide variety of different plasmids is available that mainly diverge in copy number, selection markers, promoters and fusion tags [24]. The latter two directly influence the expression of the target protein. Application of strong promoters such as the T7 promoter is more likely to result in inclusion body formation due to the elevated transcription rate [49]. On the other hand, it has been shown in numerous cases that fusing the protein to be expressed to a second one can increase solubility dramatically. For enhanced solubility, ideally a protein that is easily expressed by *E. coli* is inserted upstream of the target protein. This often comes along with the additional advantage of facilitating purification (for instance when glutathione-S-transferase (GST) or maltose binding protein (MBP) are used) [50].

Host Strain

Host strains can mainly influence protein expression regarding tRNA-concentration and disulphide bond formation [24,26]. *E. coli* cannot successfully form disulphide bonds in the reducing environment of the cytoplasm. This handicap can be overcome by using appropriate hosts such as Origami that carry mutations in the *gor* and *trxB* genes, which promote disulphide bond formation in the cytoplasm [26]. Another reason that the protein is not expressed properly is in many cases that the sequence of the protein is not optimized for the expression host [28]. The genetic code is redundant, meaning there is often more than just one triplet encoding for a

particular amino acid. Different organisms, however, tend to use certain triplets more often than others. Accordingly, the concentrations of accessible tRNAs diverge: for certain codons that are unfrequently used, only low amounts of tRNA are obtainable [51]. Thus, if the protein sequence is derived from a different organism, problems in expression might result. This can be overcome by supplying certain tRNAs. For instance, *E. coli* Rosetta strains have all rarely used tRNAs encoded [26,28].

Target Protein

Another approach that is often used when the protein sequence is not suitable for the host is codon optimization: The protein sequence is kept the same while codons, which are rarely used by the expression host are exchanged into more frequently used ones. This can be either done by gene synthesis or site directed mutagenesis [28]. Furthermore, regarding the sequence, expression of target protein can be facilitated, if the formation of secondary structures in the mRNA is avoided. Thus translation is eased leading to more soluble protein [26].

Co-expression of Chaperones

As inclusion body formation is often a result of exhaustion of the folding machinery of *E. coli*, co-overexpression of chaperones may lead to enhanced formation of soluble product. Suitable chaperones include the GroEL-GroES and DnaK-DnaJ-GrpE chaperone systems along with trigger factor [50].

Process Technological Optimization

Apart from genetic modifications varying expression conditions can have a tremendous impact on soluble protein expression [26].

Media Composition

Recombinant protein production requires nutrients for the organism. Those nutrients have to be provided by the media. However, during cultivation the media changes as substrates are taken up and products of various metabolic pathways accumulate in the reactor. These products can have inhibitory effects and impede cell division or efficient production of soluble protein [52]. Sometimes the addition of certain compounds to the media can facilitate yielding high amounts of soluble product. This includes the addition of required cofactors or prosthetic groups[50].

Process Parameters

It was found that the probability of obtaining soluble product increases at lower temperatures. Inclusion body formation can be described as an imbalance between transcription and translation: when too much DNA is transcribed, the folding machinery cannot keep up, which results in poorly folded proteins that tend to aggregate. However, at lower temperatures

transcription is slowed down as well as the process of aggregation, leading to more soluble product. In spite of this it has to be kept in mind that at lower temperatures the growth and protein synthesis rate decrease, resulting in prolonged induction times [28,53].

Another approach for decreasing the transcription rate and thus avoiding the consequent imbalance would be reducing inducer-concentration. It has been shown by Turner that on occasion cutting IPTG-concentrations in half (from 0.1 to 0.05 mM) can make all the difference [54].

7. Induction strategies: IPTG vs. Lactose

When employing the T7-based pET-expression system induction is conventionally performed by IPTG, a synthetic lactose analogue [55,56]. Using IPTG facilitates induction as the component is not metabolized by *E. coli*; thus one-point addition is sufficient and a constant amount of inducer is ensured coming along with stable induction [24]. Nevertheless, IPTG-induction entails many drawbacks: It has been shown to put a high metabolic burden on the organism [57,58] and is often associated with inclusion body formation [59-62]. Additionally, IPTG is an expensive compound [63,64], which is especially disadvantageous for large scale production of recombinant protein. Lactose on the other hand is considered a waste product and is thus comparatively inexpensive. More importantly it has also been shown to promote formation of soluble product and enhance productivity [65-69]. However, induction by lactose is scarcely used neither in academia nor industry. The cause of this is certainly the challenge resulting from metabolisable inducers as lactose: Continuous supply is indispensable for avoiding depletion and guaranteeing stable induction. Additionally, when inducing with lactose it is crucial to only add glucose in limiting amounts as otherwise lactose uptake is inhibited by carbon catabolite repression [46,70,71]. This makes controlled lactose induction unfeasible in batch or shake flask cultivations, which are often performed for small scale production of recombinant protein or during screening experiments. However, in conventional bioprocesses, which comprise the three phases of batch, fed-batch and induced fed-batch, this problem is inconsequential as during fed batch and induced fed-batch glucose is fed in limiting amounts. Problems might arise during prolonged cultivations as the biomass yield decreases and glucose accumulates. However, the main challenge of lactose induction lies in avoiding depletion on the one hand and immoderate concentrations and consequent osmotic stress on the other hand.

8. Sugar Uptake in *Escherichia coli* - Glucose dependent Lactose Metabolism and the Correlation of Concomitant Sugar Uptake

E. coli is, as many other organisms, geared towards usage of the most effortlessly metabolisable substrate, the one on which growth rate is the highest. Therefore, it is essential for the bacteria to regulate transcription of genes involved in the uptake of changing C-sources. Concomitant as well as subsequent uptake of carbohydrates is mainly regulated by the well-studied effects of carbon catabolite repression [70]. In the presence of glucose many proteins, responsible for taking up different sugars are not expressed. Thus, if two sugars, for instance glucose and lactose are present in excess, the phenomenon of diauxic growth is observed in *E. coli* [72]. Glucose is first metabolised resulting in rapid growth of the bacteria while lactose is not taken up due to inducer exclusion [45]. This phase is followed by a lag phase where growth of *E. coli* is nearly arrested, as the bacteria need all their energetic resources for expression of the lactose-metabolism related proteins. Thereafter the growth rate increases again, however not as steeply as before as now lactose is taken up, a carbohydrate on which the growth rate is slower compared to glucose [70,73,74].

Hence if glucose was present in excess, close to no lactose would be taken up by the cells and induction by the disaccharide would become an unaccomplishable task. Nevertheless, if glucose is fed in limiting amounts the cells do take up enough of it to initiate transcription from the lac operon [75]. How much of it is taken up, or the specific lactose uptake rate, $q_{s,lac}$ respectively, is only dependent on the specific glucose uptake rate $q_{s,glu}$ [69]. A schematic graph for the dependency is shown in Figure 4.

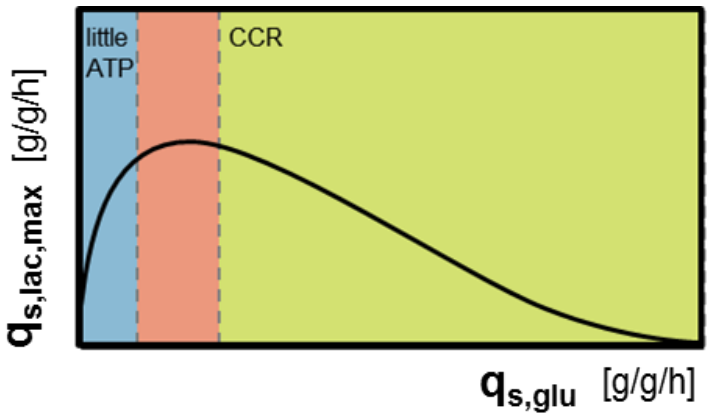


Figure 4 Schematic graph for the $q_{s,glu}/q_{s,lac,max}$ correlation: Little lactose uptake at low levels of $q_{s,glu}$ (blue) due to lack of ATP, followed by a steep increase in $q_{s,lac,max}$, and high levels (red) preceded by a comparatively shallow decrease (green) due to carbon catabolite repression.

At high amounts of glucose little lactose is taken up due to the mechanisms of carbon catabolite repression: when glucose is present EIIA is dephosphorylated and able to inhibit lactose

permease [46,70,71]. With decreasing levels of $q_{s,glu}$ more and more lactose is taken up per hour and per gram biomass. However surprisingly when no glucose is available again $q_{s,lac,max}$ is low. This is hypothesized to be a result of the ATP-related lactose transport into the cell (Figure 5).

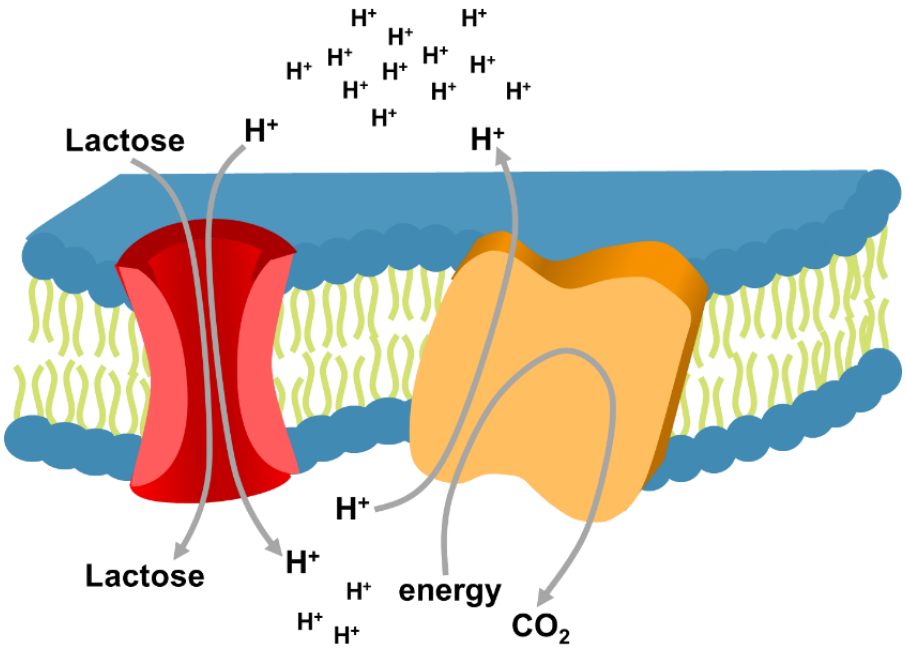


Figure 5 Scheme of lactose transport into the cell. The symporter lactose permease (red) simultaneously transports lactose and protons into the cell. This can only happen due to a proton gradient that is generated by the proton pump (orange), which is powered by ATP.

Lactose permease is a symporter, which contemporaneously transports lactose and protons into the cell. It is dependent on the proton pump which can create a pH gradient across the membrane: It ships H^+-ions out of the cell by consuming ATP [76-79]. When no glucose is available, the cells lack energy for the proton pump and hence no gradient for stimulation of lactose permease is established. Therefore, lactose permease cannot efficiently simultaneously carry lactose and protons into the cell.

Whenever efficient lactose induction is desired it is indispensable to be aware of the dependency of $q_{s,lac,max}$ on $q_{s,glu}$ as the specific lactose uptake rate influences productivity as well as product location, as has been shown in a previous study of our working group [69].

Regarding productivity, it has been shown that there is a trade-off between uptake of inducer and availability of ATP (Figure 6).

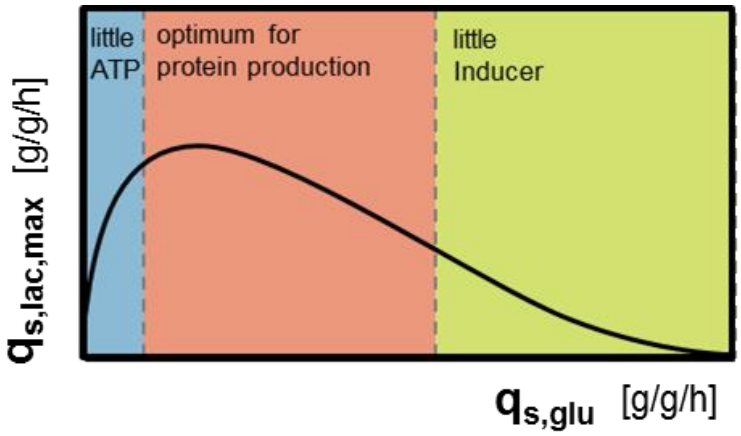


Figure 6 Schematic graph for the $q_{s,glu}/q_{s,lac,max}$-correlation: Little ATP at low levels of $q_{s,glu}$ (blue), followed by the optimum region for protein expression where levels of lactose uptake and thus inducer uptake are high and sufficient ATP is available (red), preceded by a region in which ATP-availability is high but little inducer is taken up due to carbon catabolite repression (green).

We have found that most product is expressed at $q_{s,glu}$ values around 0.2-0.3 g/g/h. At very high values of $q_{s,glu}$ less product is expressed, which is believed to be caused by the reduced lactose uptake. When only low amounts of inducer are present in the cell, transcription from the lac operon is diminished. Thus less T7 polymerase is expressed and consequently less target protein is transcribed from the pET-plasmid. On the other hand when $q_{s,glu}$ is extremely low, the cells suffer from starvation and the additional burden of overexpressing the product protein is not feasible.

Another study of our working group showed that performing induction at a controlled $q_{s,lac}$, at values beneath $q_{s,lac,max}$, impacts the ratio of soluble protein to inclusion bodies (Figure 7).

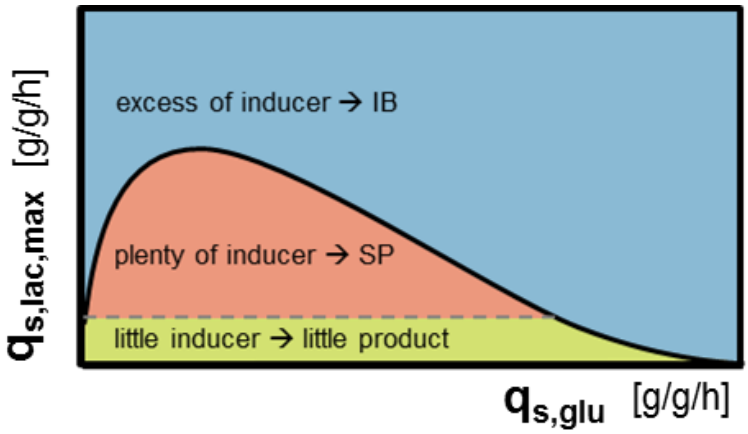


Figure 7 Schematic graph for the $q_{s,glu}/q_{s,lac,max}$-correlation: little product when $q_{s,lac}$ is low, as only low amounts of inducer are present (green). At moderate values of $q_{s,lac}$: enough inducer for effective induction, resulting in high amounts of soluble product (red). When lactose is in excess at $q_{s,lac,max}$: High amounts of inclusion bodies due to strong induction (blue).

Most inclusion bodies are formed at highest lactose uptake rates ($q_{s,lac,max}$). This is explicable by stronger induction, coming along with a high transcription rate. This favours the formation of inclusion bodies. However when $q_{s,lac}$ is decreased and lactose is only fed in limiting amounts,

fewer inclusion bodies are formed while more soluble product is obtained. This goes down to a certain point where too little inducer is present and therefore total product expression is diminished (manuscript in preparation).

However, before being able to influence $q_{s,lac}$ depending on the desired outcome, it is an essential prerequisite to establish the correlation of $q_{s,lac,max}$ in response to $q_{s,glu}$. Currently the curve is determined by performing 2 batch experiments as well as 2-3 fed-batch cultivations. The obtained data is subsequently fitted by a mechanistic model [69]. This approach involves a lot of experimental work, which is why there is a need for a more rapid way of strain characterisation in terms of lactose and glucose uptake. The development of this fast strategy was the subject of this Thesis.

9. Dynamic Process Conditions

It has been shown in many cases that dynamic process conditions can be of use for strain characterisation or bioprocess development. Dynamic conditions outcompete conventional ones such as continuous culture or repeated fed batch cultivations mainly in regard to experimental work. Instead of conducting numerous successional fermentations at different constant parameters for determining a certain response, often fewer cultivations with dynamic variation of those parameters can lead to the same outcome, requiring much less time and resources [80].

Alterable parameters for investigating desired responses during cultivations include physical parameters (such as temperature, dilution rate, aeration or stirring rate), chemical parameters (including pH, nutrient concentrations or osmolarity) and physiological parameters (mainly specific rates) [80].

Dynamic changes of process parameters can be carried out in different ways: It can be distinguished between shifts, ramps, pulses and oscillation (Figure 8).

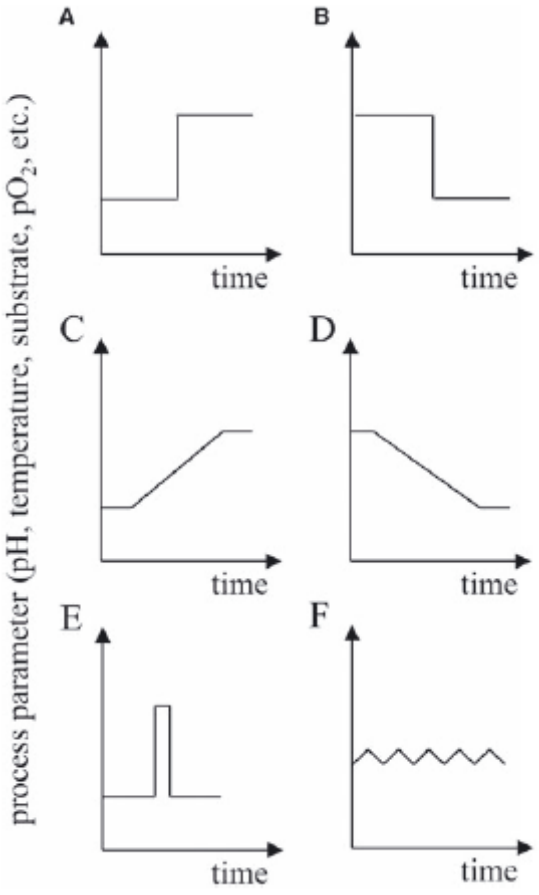


Figure 8 Overview of approaches for dynamic variation of process parameters: shift up (A), shift down (B), ramp up (C), ramp down (D), pulse (E), oscillation (F)[80]

When conducting shift experiments, a parameter is changed rapidly to a different level and is left constant on this level afterwards, which allows the systems to adapt to the changed parameter. Shifts can be of use when examining responses such as productivity to for instance a temperature- or pH- shift [80]. Sanden showed by a shift of the specific growth rate (μ), that productivity was double as high at a μ of 0.5 h^{-1} compared to a μ of 0.1 h^{-1} [81].

Another option when applying dynamic process conditions would be investigating the change of a response in dependence on the gradual variation of a parameter. In this case time-dependent effects have to be kept in mind. In order to obtain the same results as in static cultivations it is important not to change the parameters more rapidly than the system can adapt to the alteration: The time trajectory must not exceed the pace of physiological adaption. Most frequently substrate ramps are performed for investigation of productivity or growth [80].

When pulses are conducted a certain condition is changed rapidly instead of gradually. Pulses of different C-sources, for instance, allow to determine metabolite formation [80].

Oscillations describe conditions where certain parameters go up and down repeatedly. Often oscillating dO conditions were investigated in regard to productivity [80]. Oscillating process parameters have been used since the eighties, when, for instance, Larsson and Enfors investigated the effects of oxygen limitation on the respiration rate of *Penicillium chrysogenum* [82].

However, when applying dynamic process conditions, crucial prerequisites are on the one hand being able to accurately control the change in parameters and on the other hand to precisely determine the altered responses. Another drawback is the increased complexity in data analysis. In spite of that, if high resolution monitoring of process parameters is assured, dynamic approaches facilitate strain characterisation and bioprocess development tremendously [80].

10. Motivation and Scientific Questions

In recent years recombinant production of antibodies and antibody fragments is rapidly gaining ground, reaching levels of more than 1000 kg/year. This demonstrates the need for fast and efficient production [1,2]. Efficient recombinant protein production is achieved by monitoring processes, understanding the ongoing events and applying obtained knowledge for optimization. Within this Thesis it was attempted to gain further insight into the mechanistics of concomitant lactose and glucose uptake in *E. coli*. The experiments were conducted using an *E. coli* BL21DE3 strain expressing a tandem-scFv.

Lactose has an immense potential as alternative inducer whenever the T7-based pET-expression system is applied as it has been shown to benefit formation of soluble product and enhance productivity [65-69]. However, for successful lactose induction, being aware of the $q_{s,glu}/q_{s,lac,max}$-dependency is an indispensable prerequisite as the level of lactose uptake influences productivity and product location. The mechanistic correlation of concomitant sugar uptake, which was initially discovered for an *E. coli* BL21(DE3) strain expressing a scFv, was hypothesized to be applicable as a platform tool for induction optimization in various *E. coli* BL21(DE3) strains with pET-plasmids for expression of different products. Currently the correlation is determined by conducting 2-3 fed-batch as well as 2 batch experiments and subsequently fitting the curve by the developed mechanistic model. The objective of this study was to substantiate applicability of the correlation as platform tool and design and conduct a fast determination strategy by means of dynamic process conditions.

Within the scope of this Thesis two main scientific questions regarding strain characterisation in terms of concomitant lactose and glucose uptake have been addressed:

1. **Is the $q_{s,glu}/q_{s,lac,max}$-correlation platform knowledge for *E. coli* BL12(DE3) cells carrying pET plasmids?**

   Concomitant glucose- and lactose-uptake is investigated in *E. coli* BL12(DE3) cells carrying pET-plasmids for expressing four different products in the cytoplasm and periplasm. The obtained data points for each cultivation-series are fitted by the previously developed mechanistic model. Subsequently the resulting curves are compared to evaluate transferability of the discovered correlation.

2. **How does the strategy have to be designed to rapidly characterize the concomitant sugar uptake of a novel strain with little experimental workload?**

   In order to address this issue, the development of a rapid method for strain characterisation regarding the $q_{s,glu}/q_{s,lac,max}$-dependency is shown using a BL21(DE3) strain carrying a pET plasmid for the expression of a scFv. Cultivations are carried out at dynamic $q_{s,glu}$-ramps and in excess of lactose. Applied feeding strategies are optimized step by step tackling discovered challenges including hysteresis effects, adaption time for changing the metabolism to a different C-sources as well as maintenance of cell-fitness throughout the induction phase.

These two scientific questions will be answered elaborately within the following paper draft prepared for submission to *Scientific Reports*.

References Introduction

1 Lee, Y. J. & Jeong, K. J. Challenges to production of antibodies in bacteria and yeast. *J Biosci Bioeng* **120**, 483-490, doi:10.1016/j.jbiosc.2015.03.009 (2015).

2 Walsh, G. Biopharmaceutical benchmarks 2014. *Nat Biotechnol* **32**, 992-1000, doi:10.1038/nbt.3040 (2014).

3 Holliger, P. & Hudson, P. J. Engineered antibody fragments and the rise of single domains. *Nat Biotechnol* **23**, 1126-1136, doi:10.1038/nbt1142 (2005).

4 Maverakis, E. *et al.* Glycans In The Immune system and The Altered Glycan Theory of Autoimmunity: A Critical Review. *J Autoimmun*, 1-13, doi:10.1016/j.jaut.2014.12.002. (2015).

5 Nelson, A. L. Antibody fragments: hope and hype. *MAbs* **2**, 77-83 (2010).

6 Ahmad, Z. A. *et al.* scFv antibody: principles and clinical application. *Clin Dev Immunol*, doi:doi: 10.1155/2012/980250 (2012).

7 Li, F., Vijayasankaran, N., Shen, A. Y., Kiss, R. & Amanullah, A. Cell culture processes for monoclonal antibody production. *MAbs* **2**, 466-479, doi:10.4161/mabs.2.5.12720 (2010).

8 Birch, J. R. & Racher, A. J. Antibody production. *Adv Drug Deliv Rev* **58**, 671-685, doi:10.1016/j.addr.2005.12.006 (2006).

9 Rodrigues, M. E., Costa, A. R., Henriques, M., Azeredo, J. & Oliveira, R. Technological progresses in monoclonal antibody production systems. *Biotechnol Prog* **26**, 332-351, doi:10.1002/btpr.348 (2010).

10 Spadiut, O., Capone, S., Krainer, F., Glieder, A. & Herwig, C. Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol* **32**, 54-60, doi:10.1016/j.tibtech.2013.10.002 (2014).

11 Andersen, D. C. & Reilly, D. E. Production technologies for monoclonal antibodies and their fragments. *Curr Opin Biotechnol* **15**, 456-462, doi:10.1016/j.copbio.2004.08.002 (2004).

12 Hust, M. *et al.* Single chain Fab (scFab) fragment. *BMC Biotechnol* **7**, doi:10.1186/1472-6750-7-14 (2007).

13 Fakruddin, M., Mazumdar, R. M., Mannan, K. S. B., Chowdhury, A. & Hossain, M. N. Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant E. coli. *ISRN Biotechnol*, doi:10.5402/2013/590587 (2013).

14 Lundqvist, M. L., Middleton, D. L., Radford, C., Warr, G. W. & Magorb, K. E. Immunoglobulins of the non-galliform birds: Antibody expression and repertoire in the duck. *Dev Comp Immunol* **30**, 93-100, doi: 10.1016/j.dci.2005.06.019 (2006).

15 Naiyana, G., Suh, J. W. & Sunwoo, H. H. Effect of anti-gliadin IgY antibody on epithelial intestinal integrity and inflammatory response induced by gliadin. *BMC Immunol* **16**, doi:10.1186/s12865-015-0104-1 (2015).

16 Stadlmann, V. *et al.* Novel avian single-chain fragment variable (scFv) targets dietary gluten and related natural grain prolamins, toxic entities of celiac disease. *BMC Biotechnol* **15**, 109, doi:10.1186/s12896-015-0223-z (2015).

17 Comino, I., de Lourdes Moreno, M. & Sousa, C. Role of oats in celiac disease. *World J Gastroenterol* **21**, 11825-11831, doi: 10.3748/wjg.v21.i41.11825 (2015).

18 Mainil, J. Escherichia coli virulence factors. *Vet Immunol Immunopathol* **152**, 2-12, doi:10.1016/j.vetimm.2012.09.032. (2013).

19 Baeshen, N. A. *et al.* Cell factories for insulin production. *Microb Cell Fact* **13**, doi:10.1186/s12934-014-0141-0 (2014).

20 Swartz, J. R. Advances in Escherichia coli production of therapeutic proteins. *Curr Opin Biotechnol* **12**, 195-201, doi:10.1016/S0958-1669(00)00199-3 (2001).

21 Blattner, F. R. *et al.* The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453-1462, doi:10.1126/science.277.5331.1453 (1997).

22 Sanchez-Garcia, L. *et al.* Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microb Cell Fact* **15**, doi:10.1186/s12934-016-0437-3 (2016).

23 Baeshen, M. N. *et al.* Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives. *J Microbiol Biotechnol* **25**, 953-962, doi:10.4014/jmb.1412.12079 (2015).

24 Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **5**, 172, doi:10.3389/fmicb.2014.00172 (2014).

25 Choi, J. H., Keum, K. C. & Lee, S. Y. Production of recombinant proteins by high cell density culture of Escherichia coli. *Chem Eng Sci* **61**, 876-885, doi:10.1016/j.ces.2005.03.031 (2006).
26 Gopal, G. J. & Kumar, A. Strategies for the production of recombinant protein in Escherichia coli. *Protein Journal* **32**, 419-425, doi:10.1007/s10930-013-9502-5 (2013).
27 Makrides, S. C. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiol Rev* **60**, 512-538 (1996).
28 Francis, D. M. & Page, R. Strategies to optimize protein expression in E. coli. *Curr Protoc Protein Sci* **Chapter 5**, Unit 5 24 21-29, doi:10.1002/0471140864.ps0524s61 (2010).
29 Baneyx, F. Recombinant protein expression in Escherichia coli. *Curr Opin Biotechnol* **10**, 411-421, doi:Doi 10.1016/S0958-1669(99)00003-8 (1999).
30 Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M. & Panda, A. K. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microb Cell Fact* **14**, doi:10.1186/s12934-015-0222-8 (2015).
31 Singh, S. M. & Panda, A. K. Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng* **99**, 303-310, doi:10.1263/jbb.99.303 (2005).
32 Cabrita L. D. & P., B. S. Protein expression and refolding--a practical guide to getting the most out of inclusion bodies. *Biotechnol Annu Rev* **10**, 31-50, doi: 10.1016/S1387-2656(04)10002-1 (2004).
33 Huang, C.-J., Lin, H. & Yang, X. Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. *J Ind Microbiol Biotechnol* **39**, 383-399, doi:10.1007/s10295-011-1082-9 (2012).
34 Jia, B. & Jeon, C. O. High-throughput recombinant protein expression in Escherichia coli: current status and future perspectives. *Open Biol* **6**, doi:10.1098/rsob.160196 (2016).
35 Liu, M. *et al.* Metabolic engineering of Escherichia coli to improve recombinant protein production. *Appl Microbiol Biot* **99**, 10367-10377, doi:10.1007/s00253-015-6955-9. (2015).
36 Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S. & Kim, J. F. Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3). *J Mol Biol* **394**, 634-643, doi:10.1016/j.jmb.2009.09.022. (2009).
37 Han, M. J. Exploring the proteomic characteristics of the Escherichia coli B and K-12 strains in different cellular compartments. *J Biosci Bioeng* **122**, 1-9, doi:10.1016/j.jbiosc.2015.12.005. (2016).
38 Jeong, H., Kim, H. J. & Lee, S. J. Complete Genome Sequence of Escherichia coli Strain BL21. *Genome Announc* **3**, doi:10.1128/genomeA.00134-15. (2015).
39 Studier, F. W. & Moffatt, B. A. Use of Bacteriophage-T7 Rna-Polymerase to Direct Selective High-Level Expression of Cloned Genes. *J Mol Biol* **189**, 113-130, doi:Doi 10.1016/0022-2836(86)90385-2 (1986).
40 Baneyx, F. Recombinant protein expression in Escherichia coli. *Curr Opin Biotechnol* **10**, 10, doi:10.1016 (1999).
41 Tegel, H., Ottosson, J. & Hober, S. Enhancing the protein production levels in Escherichia coli with a strong promoter. *FEBS J* **278**, 729-739, doi:10.1111/j.1742-4658.2010.07991.x (2011).
42 Studier, F. W. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* **219**, 37-44 (1991).
43 Fulcrand, G. *et al.* DNA supercoiling, a critical signal regulating the basal expression of the lac operon in Escherichia coli. *Sci Rep* **6**, doi:10.1038/srep19243 (2016).
44 Deutscher, J. The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol* **11**, 87-93, doi:10.1016/j.mib.2008.02.007 (2008).
45 Stulke, J. & Hillen, W. Carbon catabolite repression in bacteria. *Curr Opin Microbiol* **2**, 195-201, doi:10.1016/S1369-5274(99)80034-4 (1999).
46 Warner, J. B. & S., L. J. CcpA-dependent carbon catabolite repression in bacteria. *Microbiol Mol Biol Rev* **67**, 475-490, doi:10.1128/MMBR.67.4.475-490.2003 (2003).
47 Maizels, N. M. The nucleotide sequence of the lactose messenger ribonucleic acid transcribed from the UV5 promoter mutant of Escherichia coli. *Proc Natl Acad Sci USA* **70**, 3585-3589 (1973).

48 Akbari, V., Sadeghi, H. M. M., Jafarian-Dehkordi, A., Chou, C. P. & Abedi, D. Optimization of a single-chain antibody fragment overexpression in Escherichia coli using response surface methodology. *Res Pharm Sci* **10**, 75-83 (2015).
49 Fahnert, B., Lilie, H. & Neubauer, P. Inclusion Bodies: Formation and Utilization. *Adv Biochem Eng Biotechnol* **89**, 93-142, doi:DOI 10.1007/b93995 (2004).
50 Sørensen, H. P. & Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microb Cell Fact* **4**, doi:10.1186/1475-2859-4-1 (2005).
51 dos Reis, M., Savva, R. & Wernisch, L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res* **32**, 5036-5044, doi:10.1093/nar/gkh834 (2004).
52 Khow, O. & Suntrarachun, S. Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pac J Trop Biomed* **2**, 159-162 (2012).
53 San-Miguel, T., Perez-Bermudez, P. & Gavidia, I. Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature. *Springerplus* **2**, 89, doi:10.1186/2193-1801-2-89 (2013).
54 Turner, P., Holst, O. & Karlsson, E. N. Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in Escherichia coli by using a soluble fusion-tag and by tuning of inducer concentration. *Protein Expr Purif* **39**, 54-60, doi:10.1016/j.pep.2004.09.012 (2005).
55 Durania, V., Sullivanb, B. J. & Maglierya, T. J. Simplifying protein expression with ligation-free, traceless and tag-switching plasmids. *Protein Expr Purif* **85**, 9-17, doi:10.1016/j.pep.2012.06.007 (2012).
56 Marbach, A. & Bettenbrock, K. lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J Biotechnol* **157**, 82-88, doi:10.1016/j.jbiotec.2011.10.009. (2012).
57 Dvorak, P. *et al.* Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway. *Microb Cell Fact* **14**, doi: 10.1186/s12934-015-0393-3 (2015).
58 Haddadin, F. T. & Harcum, S. W. Transcriptome profiles for high-cell-density recombinant and wild-type Escherichia coli. *Biotechnol Bioeng* **90**, 127-153, doi:10.1002/bit.20340 (2005).
59 Zhang, Z. *et al.* High-level production of membrane proteins in E. coli BL21(DE3) by omitting the inducer IPTG. *Microb Cell Fact* **14**, doi:10.1186/s12934-015-0328-z. (2015).
60 Sina, M., Farajzadeh, D. & Dastmalchi, S. Effects of Environmental Factors on Soluble Expression of a Humanized Anti-TNF-α scFv Antibody in Escherichia coli. *Adv Pharm Bull* **5**, 455-461, doi: 10.15171/apb.2015.062 (2015).
61 Changa, B. & Chaudharya, P. M. High level production and one-step purification of biologically active ectodysplasin A1 and A2 immunoadhesins using the baculovirus/insect cell expression system. *Protein Expr Purif* **37**, 162-169, doi:10.1016/j.pep.2004.04.026 (2004).
62 Xu, Y. *et al.* Characterization of the T7 promoter system for expressing penicillin acylase in Escherichia coli. *Appl Microbiol Biot* **72**, 529-536, doi: 10.1007/s00253-005-0293-2 (2006).
63 Gombert, A. K. & Kilikian, B. V. Recombinant gene expression in Escherichia coli cultivation using lactose as inducer. *J Biotechnol* **60**, 47-54, doi:Doi 10.1016/S0168-1656(97)00185-5 (1998).
64 Leea, S. K. & Keasling, J. D. Heterologous protein production in Escherichia coli using the propionate-inducible pPro system by conventional and auto-induction methods. *Protein Expr Purif* **61**, 197-203, doi:10.1016/j.pep.2008.06.008 (2008).
65 Bashir, H. *et al.* Simple procedure applying lactose induction and one-step purification for high-yield production of rhCIFN. *Biotechnol Appl Biochem*, doi:10.1002/bab.1426 (2015).
66 Fruchtl, M., Sakon, J. & Beitle, R. Expression of a collagen-binding domain fusion protein: effect of amino acid supplementation, inducer type, and culture conditions. *Biotechnol Prog* **31**, 503-509, doi:10.1002/btpr.2048 (2015).
67 Ma, X., Su, E., Zhu, Y., Deng, S. & Wei, D. High-level expression of glutaryl-7-aminocephalosporanic acid acylase from Pseudomonas diminuta NK703 in Escherichia coli by combined optimization strategies. *J Biotechnol* **168**, 607-615, doi:10.1016/j.jbiotec.2013.08.024 (2013).

68 Monteiro, R. A., Souza, E. M., Yates, M. G., Pedrosa, F. O. & Chubatsu, L. S. Use of lactose to induce expression of soluble NifA protein domains of Herbaspirillum seropedicae in Escherichia coli. *Can J Microbiol* **46**, 1087-1090 (2000).
69 Wurm, D. J. *et al.* The E. coli pET expression system revisited-mechanistic correlation between glucose and lactose uptake. *Appl Microbiol Biot* **100**, 8721-8729, doi:10.1007/s00253-016-7620-7 (2016).
70 Kremling, A., Geiselmann, J., Ropers, D. & de Jong, H. Understanding carbon catabolite repression in Escherichia coli using quantitative models. *Trends Microbiol* **23**, 99-109, doi:10.1016/j.tim.2014.11.002 (2015).
71 Brückner, R. & Titgemeyer, F. Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiol Lett* **209**, 141-148, doi:10.1111/j.1574-6968.2002.tb11123.x (2002).
72 Boulineau, S. *et al.* Single-Cell Dynamics Reveals Sustained Growth during Diauxic Shifts. *PLOS ONE* **8**, doi:10.1371/journal.pone.0061686 (2013).
73 Boulineau, S. *et al.* Single-cell dynamics reveals sustained growth during diauxic shifts. *PLOS ONE* **8**, doi:10.1371/journal.pone.0061686 (2013).
74 Loomis, W. F. J. & Magasanik, B. Glucose-lactose diauxie in Escherichia coli. *J Bacteriol* **93**, 1397-1401 (1967).
75 Mayer, S. *et al.* Lactose autoinduction with enzymatic glucose release: characterization of the cultivation system in bioreactor. *Protein Expr Purif* **94**, 67-72, doi:10.1016/j.pep.2013.10.024 (2014).
76 Johnson, J. L. & Brooker, R. J. Control of H+/Lactose Coupling by Ionic Interactions in the Lactose Permease of Escherichia coli. *J Membr Biol* **198**, 135-146, doi:10.1007/s00232-004-0667-x (2004).
77 West, I. C. & Mitchell, P. Stoicheiometry of lactose-H+ symport across the plasma membrane of Escherichia coli. *Biochem J* **132**, 587-592 (1973).
78 West, I. C. Lactose transport coupled to proton movements in escherichia coli. *Biochem Biophys Res Commun* **41**, 655-661, doi:10.1016/0006-291X(70)90063-X (1970).
79 Kaback, H. R. A chemiosmotic mechanism of symport. *Proc Natl Acad Sci USA* **112**, 1259-1264, doi:10.1073/pnas.1419325112 (2015).
80 Spadiut, O., Rittmann, S., Dietzsch, C. & Herwig, C. Dynamic process conditions in bioprocess development. *Eng Life Sci* **13**, 88-101, doi:10.1002/elsc.201200026 (2013).
81 Sandén, A. M. *et al.* Limiting factors in Escherichia coli fed-batch production of recombinant proteins. *Biotechnol Bioeng* **81**, 158-166, doi: 10.1002/bit.10457 (2003).
82 Larsson, G. & Enfors, S. O. Studies of insufficient mixing in bioreactors: Effects of limiting oxygen concentrations and short term oxygen starvation on Penicillium chrysogenum. *Bioprocess Eng* **3**, doi:doi:10.1007/BF00373475 (1988).

Paper Draft:

Mechanistic Platform Knowledge of Concomitant Sugar Uptake in *Escherichia coli* and Analysis thereof

David J. Wurm[1*], Johanna Hausjell[1*], Sophia Ulonska[1], Christoph Herwig[1,2], Oliver Spadiut[1,2]

[1]*Research Division Biochemical Engineering, Institute of Chemical Engineering, Vienna University of Technology, Vienna, Austria*

[2]*Christian Doppler Laboratory for Mechanistic and Physiological Methods for Improved Bioprocesses, Institute of Chemical Engineering, Vienna University of Technology, Vienna, Austria*

**These authors contributed equally to this work*

Correspondence and Requests for materials should be addressed to:

Oliver Spadiut, TU Wien, Institute of Chemical Engineering, Research Area Biochemical Engineering, Gumpendorfer Strasse 1a, 1060 Vienna, Austria. Tel: +43 1 58801 166473, Fax: +43 1 58801 166980, Email: oliver.spadiut@tuwien.ac

Abstract

When producing recombinant proteins *Escherichia coli* strain BL21(DE3) in combination with the T7-based pET-expression system is often the method of choice. In a recent study we introduced a mechanistic model describing the correlation of the specific glucose uptake rate ($q_{s,glu}$) and the corresponding maximum specific lactose uptake rate ($q_{s,lac,max}$) for a pET-based recombinant *E. coli* strain producing a single chain variable fragment (scFv) against celiac disease. We also showed the effect of $q_{s,lac,max}$ on productivity and product location underlining its importance for recombinant protein production processes.

In the present study we investigated the mechanistic $q_{s,glu}/q_{s,lac,max}$ correlation for four pET-based *E. coli* strains producing different recombinant products and, in fact, proved the mechanistic model to be platform knowledge for this kind of *E. coli* strains. However, we found that the respective model parameters strongly depended on the recombinant product. Driven by this observation we tested different dynamic bioprocess strategies to allow a faster investigation of this mechanistic correlation. In fact, we succeeded and finally propose an experimental strategy comprising only one batch cultivation, one fed-batch cultivation as well as one dynamic experiment, to reliably determine the mechanistic model for $q_{s,glu}/q_{s,lac,max}$ and get trustworthy model parameters for a pET-based recombinant *E. coli* strain.

Introduction

The bacterium *Escherichia coli* is one of the most widely used host organisms for recombinant protein production [1-3]. It features several advantages including extensive knowledge about its genome coming along with the availability of numerous established methods for genetic modification, multiple engineered strains as well as a dazzling array of expression plasmids [2,4]. Amongst those, the T7-based pET expression plasmids are frequently employed since the strong T7 promoter [5] allows exceptionally high yields of recombinant product [4,6]. The most common approach for inducing these pET-based *E. coli* strains is by IPTG [6,7]. IPTG is not metabolized by bacteria, which is why one-point addition is sufficient to guarantee induction [4]. However, IPTG is known to put a high metabolic burden on *E. coli* [8,9] and is often associated with the generation of misfolded protein aggregates, called inclusion bodies [10,11]. In contrast, the alternative inducer lactose has been shown to favour the production of soluble product and trigger enhanced productivity [12-15]. However, lactose is scarcely used in biochemical engineering since induction entails the challenge of continuous supply of the disaccharide as lactose gets metabolized by *E. coli*. Furthermore, cultivations have to be conducted at limiting amounts of glucose as otherwise lactose uptake is inhibited due to the well-known phenomenon of carbon catabolite repression (e.g. [16-18]).

To shed more light on the mechanistic correlation between the uptake of glucose and lactose, we recently performed a comprehensive study with a recombinant pET-based *E. coli* strain producing a single chain variable fragment (scFv) against celiac disease [15]. We succeeded in establishing a mechanistic model of the specific glucose uptake rate ($q_{s,glu}$) and the corresponding maximum specific lactose uptake rate ($q_{s,lac,max}$). Furthermore, we showed that $q_{s,lac,max}$ impacted productivity as well as product location and is thus a crucial parameter for recombinant protein production. Finally, we hypothesized that this mechanistic correlation might describe platform knowledge for *E. coli* strains carrying the pET expression system and proposed to conduct at least four bioreactor cultivations (batch and fed-batch experiments) to determine the mechanistic model for any pET-based *E. coli* strain [15].

In the present study we put our hypothesis to test and investigated the mechanistic correlation of $q_{s,glu}$ and $q_{s,lac,max}$ for four pET-based *E. coli* BL21(DE3) strains producing different recombinant proteins. We were able to prove the mechanistic model to be applicable for these strains and concluded that this model in fact describes platform knowledge for pET-based *E. coli* strains. However, we found that model parameters were strongly dependent on the recombinant product. This finding argues for the need of physiological characterization of each

recombinant pET-based *E. coli* strain in order to optimize recombinant protein production as well as to avoid sugar accumulation and resulting osmotic stress for the cells. Driven by this need we investigated different dynamic strategies to accelerate the establishment of the mechanistic model. In fact, we succeeded and propose a strategy comprising only three cultivations, including a batch, a standard fed-batch and a dynamic fed-batch cultivation, to determine the mechanistic $q_{s,glu}/q_{s,lac,max}$-model for pET-based *E. coli* strains. We believe that all scientists working with recombinant *E. coli* strains will benefit from this study as we do not only present mechanistic platform knowledge for *E. coli* strains carrying the pET expression system, but also offer a strategy for fast determination of this mechanistic correlation.

Results and Discussion

Is the Mechanistic $q_{s,glu}/q_{s,lac,max}$ Model Platform Knowledge for *E. coli*?

The main motivation for this study was to test if the previously generated mechanistic model of $q_{s,glu}$ and $q_{s,lac,max}$ for a recombinant *E. coli* BL21(DE3) strain producing a scFv with a pET expression system describes platform knowledge for pET-based *E. coli* strains. Thus, we investigated this mechanistic correlation for four different pET-based recombinant *E. coli* BL21(DE3) strains producing either 1) the model protein green fluorescent protein (GFP), 2) the plant enzyme horseradish peroxidase (HRP; e.g. [19]), 3) a scFv against celiac disease [20] and 4) a novel tandem construct of this scFv. Horseradish peroxidase was produced in the periplasm of *E. coli*, whereas the other three proteins were produced in the cytoplasm. As suggested previously [15], we performed at least four bioreactor cultivations (hereafter referred to as "static experiments") for each strain to determine the mechanistic correlation of $q_{s,glu}$ and $q_{s,lac,max}$. Then, we fitted the mechanistic model (Equation 3; [15]) to the data. In fact, we were able to demonstrate that the mechanistic model was applicable for all four strains (Figure 9). All curves followed the same trend: little lactose uptake at low $q_{s,glu}$, followed by a steep increase in $q_{s,lac,max}$, preceded by a comparatively shallow decrease at higher levels of $q_{s,glu}$. We explain this trend as follows: At low levels of $q_{s,glu}$ little lactose is taken up by the cells, as energy is required for the ATP-related lactose transport into the cell [21]. Towards higher $q_{s,glu}$, $q_{s,lac,max}$ increases but then gradually drops again due to the well-studied effects of carbon catabolite repression [15-18,22]. Due to the results depicted in Figure 9, we concluded that the mechanistic model (Equation 3) in fact describes platform knowledge for pET-based recombinant *E. coli* strains.

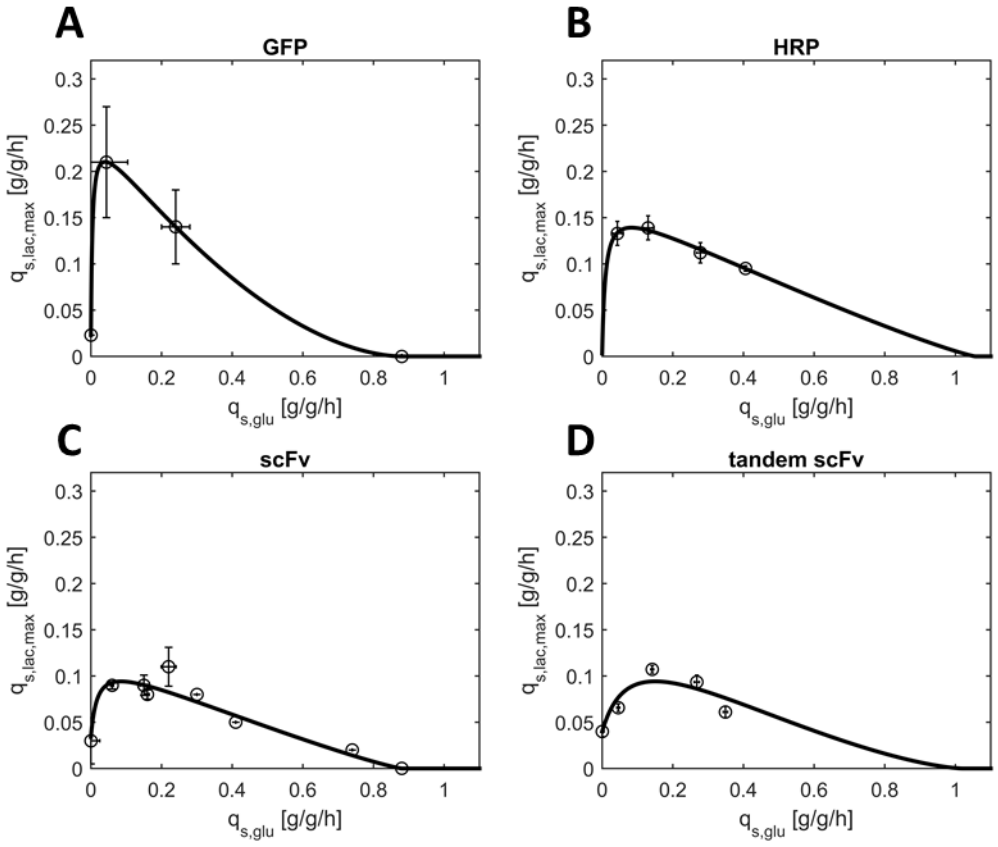


Figure 9 $q_{s,glu}/q_{s,lac,max}$-correlation for recombinant pET-based *E. coli* strains producing either A) GFP, B) HRP, C) scFv or D) tandem scFv. Data-points were obtained from batch and fed-batch cultivations with constant $q_{s,glu}$ and excess lactose ("static experiments") and subsequently fitted to the mechanistic model (Equation 3)

Even though all model parameters were within a physiologically meaningful range, we found striking differences between the four different strains (Table 1).

Table 1 Model parameters for recombinant pET-based *E. coli* strains producing GFP, HRP, scFv or tadem scFv.

rec. product	$q_{s,lac,max}$*	K_A	$q_{s,glu,crit}$	n	$q_{s,lac,noglu}$	NRMSE
	[g/g/h]	[g/g/h]	[g/g/h]	[-]	[g/g/h]	[%]
GFP	0.23	0.0042	0.88	1.77	0.023	0.04
HRP	0.17	0.0092	1.06	1.15	0.0032	5.14
scFv	0.09	0.019	0.88	1.16	0.034	9.72
tandem scFv	0.13	0.094	1.02	1.48	0.040	9.11

Thus, we performed an identifiability analysis to verify the model parameters. For the strains producing HRP, IGY and scFv the identifiability analysis was successful revealing that the parameters were trustworthy. For the strain producing the tandem scFv we had to include general knowledge derived from the other strains to identify the parameters making them

slightly more error prone. We believe that for that strain the static experiments were not ideally distributed over the whole $q_{s,glu}$ range. However, the normalized root mean square error (NRMSE) of all curves was below 10% attesting a good correlation between experimental data and the data points from the fitted curves. Since all these strains were pET-based BL21(DE3) chassis strains, we concluded that mechanistic model parameters mainly depended on the recombinant product. This fact strongly argues for physiological strain characterization of each recombinant *E. coli* strain prior to production processes, not only to optimize recombinant protein production, but also to avoid sugar accumulation and thus osmotic stress for the *E. coli* cells. In our previous study we proposed to conduct at least four bioreactor cultivations to determine the mechanistic $q_{s,glu}/q_{s.lac,max}$-correlation for any pET-based *E. coli* strain. Although this hypothesis obviously held true (Figure 8), we were eager to find another strategy allowing faster strain characterization and thus speed up bioprocess development.

Dynamics in *E. coli* Strain Characterization

We used the pET-based recombinant *E. coli* strain producing the tandem scFv against celiac disease to test different dynamic methods to possibly allow faster determination of the $q_{s,glu}/q_{s.lac,max}$-correlation. Applying dynamic process conditions to accelerate bioprocess development is a common approach in our working group [23-27]. The different dynamic strategies tested are schematically depicted in Figure 9 and Figure 10. In general, all experiments were conducted by employment of $q_{s,glu}$-ramps and lactose in excess.

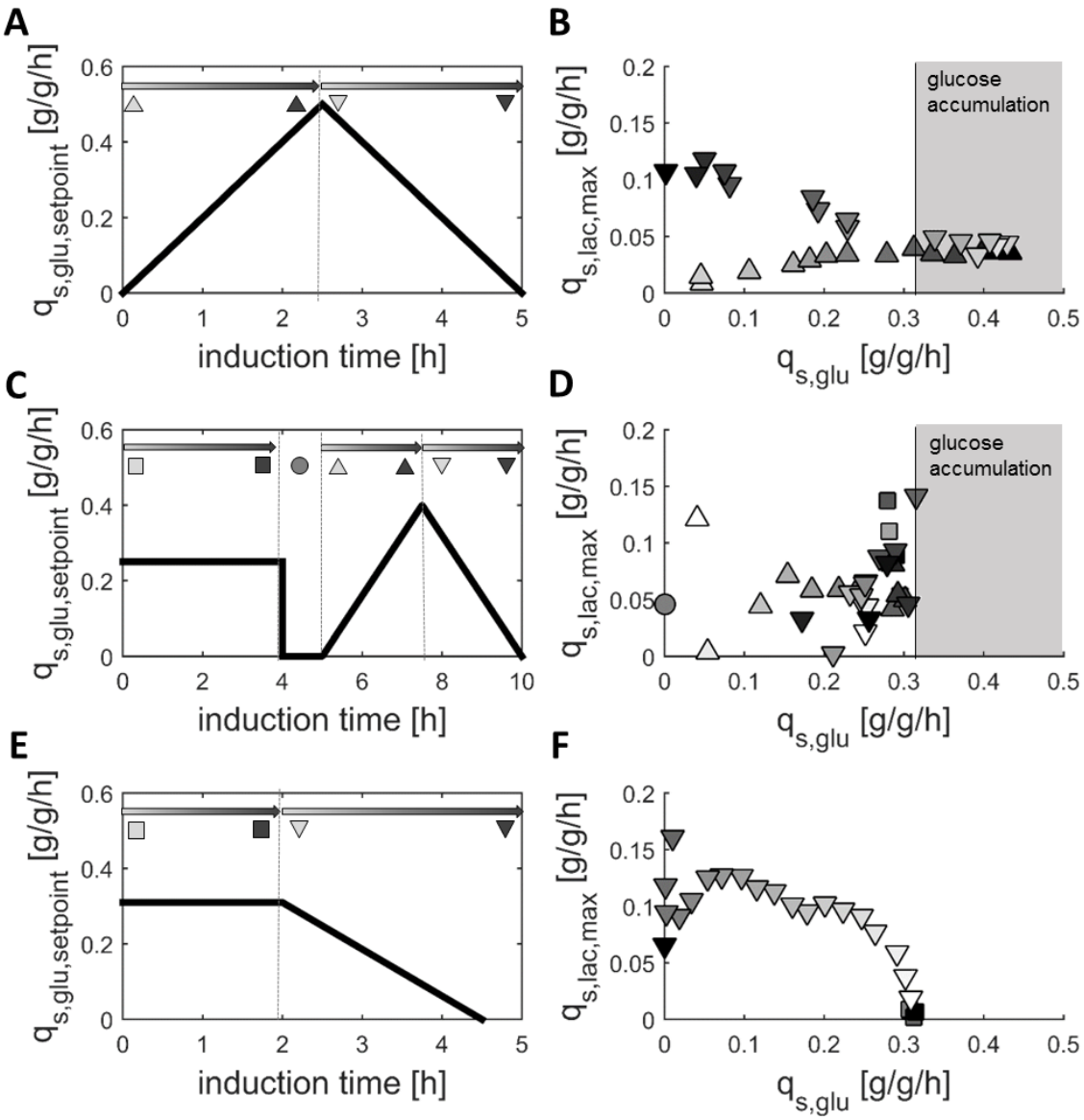


Figure 10 Dynamic bioprocess strategies (A, C, E) and respective resulting $q_{s,glu}/q_{s,lac,max}$-correlations (B, D, F). Different feeding phases are marked by different symbols. Time courses within phases go from light grey to dark grey, as indicated on the left. The shape and colour of the symbols of the $q_{s,glu}/q_{s,lac,max}$ data points on the right (B, D, F) correspond to the symbols on the left (A, C, E). Samples were taken every hour during adaptation, every 10 min during $q_{s,glu}$ ramps and every hour after the ramp.

Hysteresis of $q_{s,glu}$

In the first dynamic experiment $q_{s,glu}$ was increased from 0 g/g/h to 0.5 g/g/h and then again decreased to 0 g/g/h within 5 h resulting in a hysteresis of the specific uptake rate of glucose (Figure 10A). As shown in Figure 10B, $q_{s,lac,max}$ increased with increasing $q_{s,glu}$. However, the absolute values were far lower compared to the results in static experiments (Figure 9D). We hypothesized that *E. coli* needs time to adapt to lactose since enzymes required for uptake and metabolism of the disaccharide need to be expressed, a phenomenon which has been described for diauxic growth before [28,29]. Thus, we decided to include an adaptation phase to lactose before the $q_{s,glu}$ ramp.

Adaptation Followed by Hysteresis of $q_{s,glu}$

In the second dynamic experiment, we controlled $q_{s,glu}$ at 0.25 g/g/h for 4 h in the presence of lactose, followed by 1 h without glucose-feeding to investigate lactose uptake in the absence of glucose. Then we again performed a $q_{s,glu}$ hysteresis experiment (Figure 10C). However, as shown in Figure 10D the $q_{s,glu}/q_{s,lac,max}$-values did not follow the expected trend, but resulted in a quite chaotic cloud of data points. Viability measurements using FACS revealed fluctuating viability and up to 10 % dead *E. coli* cells during this dynamic cultivation (Supplementary Figure 1). We hypothesized that cell death and lysis led to fluctuating q_s-values during the experiment causing this unexpected result. We further believe that 1 h without glucose feed and the long overall induction time of 10 h were too stressful for the cells. However, we found that $q_{s,lac,max}$ values remained constant after approximately 2.0 h at $q_{s,glu}$ = 0.25 g/g/h indicating that the cells were fully adapted (Supplementary Figure 2). Furthermore, we observed glucose accumulation at $q_{s,glu}$ values higher than 0.32 g/g/h.

Adaptation Followed by $q_{s,glu}$ Ramp Down

Based on our observations that (1) adaption to lactose took approximately 2.0 h at $q_{s,glu}$ 0.25 g/g/h, (2) glucose accumulation was observed for fully adapted cells at $q_{s,glu}$ higher than 0.32 g/g/h, and (3) overall induction time should be kept short to maintain cell fitness, we designed the third dynamic experiment as follows: the adaption phase was conducted for 2 h at $q_{s,glu}$ of 0.31 g/g/h, before $q_{s,glu}$ was ramped down to 0 g/g/h within 2.5 h (Figure 10E). As shown in Figure 10F, the data followed the expected trend (Figure 9D) with two exceptions. First, $q_{s,lac,max}$ values at $q_{s,glu}$ of 0.31 g/g/h (light grey triangles in Figure 10F) were much lower compared to the values obtained in static experiments (Figure 9D). We concluded that 2 h of lactose presence at $q_{s,glu}$ of 0.31 g/g/h were not sufficient for full adaptation and that adaptation at $q_{s,glu}$ = 0.25 g/g/h was preferred. Furthermore, we observed quite high values for $q_{s,lac,max}$ at low $q_{s,glu}$ values. While in static experiments we determined a $q_{s,lac,max}$ of 0.04 g/g/h at $q_{s,glu}$ = 0 g/g/h (Figure 9D, Table 1), we found $q_{s,lac,max}$ values higher than 0.1 g/g/h in the dynamic experiment. However, we analyzed $q_{s,lac,max}$ values for a prolonged time and interestingly observed a constant decrease of $q_{s,lac,max}$ over time (Figure 10F). We believe that the *E. coli* cells still harboured a great amount of enzymes required for the transport and metabolism of lactose once they had been cultivated in the dynamic ramp experiment and that the presence of these enzymes was only slowly reduced resulting in the initially high $q_{s,lac,max}$ values. In contrast, when performing a static fed-batch experiment at $q_{s,glu}$ of 0 g/g/h, the cells actually do not have the required energy to produce a great amount of these enzymes resulting in a lower $q_{s,lac,max}$ (Figure 9D). We

obviously have to deal with great time effects when working with the pET-based *E. coli* systems and lactose induction.

Optimized Adaption Followed by two Ramp Experiments

Based on the conclusions drawn from the first three dynamic experiments, we finally tested a strategy comprising two ramp experiments. To guarantee fast adaptation, we adapted the cells at $q_{s,glu}$ = 0.25 g/g/h for 2.0 h. Once cells were fully adapted to lactose, indicated by a constant $q_{s,lac,max}$, $q_{s,glu}$ was either ramped up until glucose accumulation was observed (Figure 11A) or ramped down to $q_{s,glu}$ = 0 g/g/h at a rate of 0.14 g/g/h^2 to guarantee a total cultivation time of less than 5 h (Figure 11C).

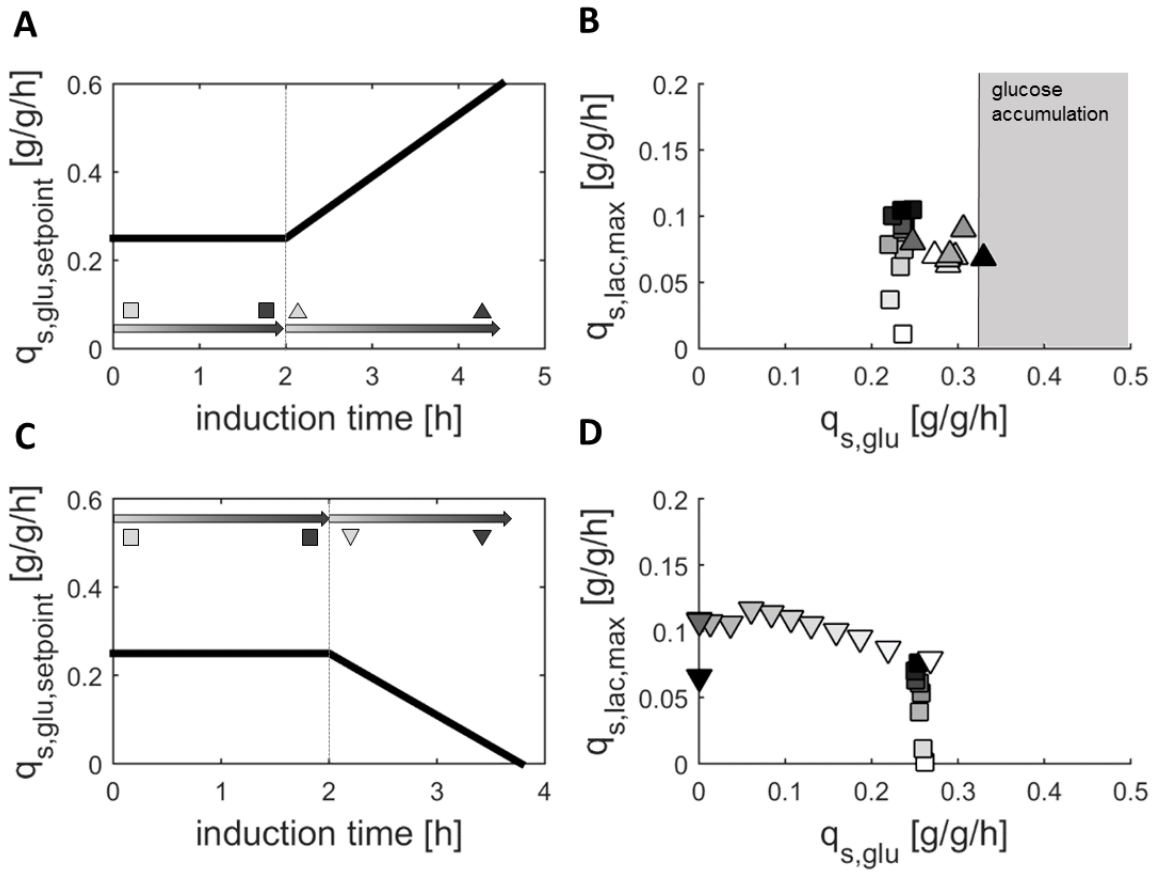


Figure 11 Optimized adaption followed by two dynamic experiments: ramp up (A, B) and ramp down (C, D). Samples were taken every 10 min during adaptation and every 30 min during $q_{s,glu}$ ramps. In the ramp down experiment another sample was taken after 4 hours at $q_{s,glu}$ = 0 g/g/h (black triangle in Figure 3D).

As shown in Figure 11B, cells were fully adapted after 2.0 h. We observed a decrease of $q_{s,lac,max}$ once we increased $q_{s,glu}$ which agreed to our previous results (Figure 9D). At $q_{s,glu}$ = 0.32 g/g/h we again observed glucose accumulation confirming our previous observations. When we decreased $q_{s,glu}$ from 0.25 g/g/h to 0 g/g/h we obtained $q_{s,lac,max}$ values which followed the expected trend, but again were higher compared to the values from static experiments (Figure

11D). We fitted the data of the two ramp experiments to the model and compared the model fit and the parameters to the results from static experiments (Figure 12; Table 2).

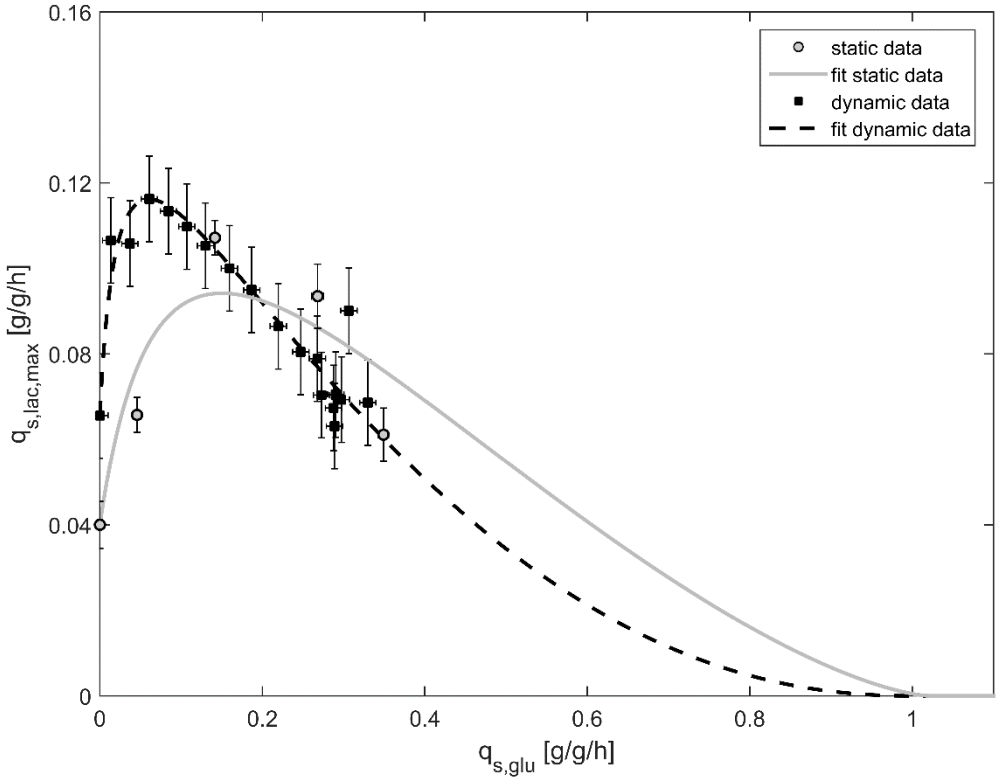
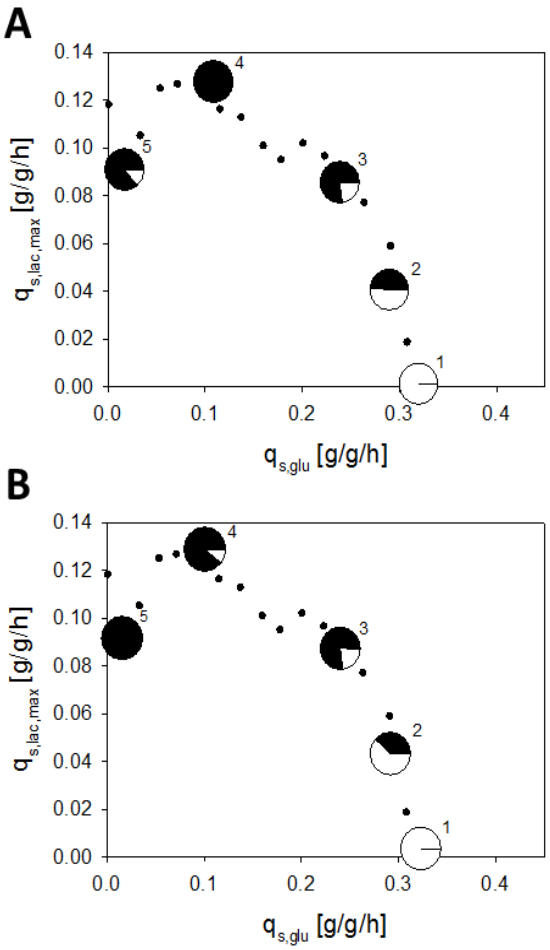


Figure 12 $q_{s,glu}/q_{s,lac,max}$-correlation derived from data points from static experiments (grey circles, solid line) and from dynamic ramp experiments (black squares, dashed line).

As shown in Figure 12, the main discrepancy between the two model fits can be found at low $q_{s,glu}$ values. However, this also impacted the shape of the curve at high $q_{s,glu}$ values. The NRMSE of the curve derived from ramp experiments was 22.1 % in respect to the static data points, compared to only 9.1 % for the curve derived from static experiments (Figure 12). Again, we hypothesized that there was still a high amount of enzymes for lactose uptake and metabolism available in the cells distorting the true $q_{s,glu}/q_{s,lac,max}$-correlation at low $q_{s,glu}$ setpoints. Apparently, the ramp speed was higher than the physiological adaptation of the cells. To put this hypothesis to test and further confirm our observations of the $q_{s,glu}/q_{s,lac,max}$-correlation by molecular biological data, we performed qPCR-analysis of β-Galactosidase (LacZ) and β-Galactosid-Permease (LacY) (Figure 13). The transcription of both genes was regulated by the well-studied lac operon which is why abundance of mRNA of either gene depended on the availability of the inducer, lactose [30].



ID	$q_{s,glu}$ [g/g/h]	$q_{s,lac,max}$ [g/g/h]	LacZ/CysG: $2^{\Delta\Delta Cp}$		
			Mean [-]	Std.dev.	Relative amount [%]
1	0.32	0	1.9	0.6	0.28
2	0.29	0.04	168.9	1.6	46.54
3	0.24	0.09	227.2	9.6	62.56
4	0.11	0.13	364.7	37.2	100.00
5	0.02	0.09	306.6	6.2	84.48

ID	$q_{s,glu}$ [g/g/h]	$q_{s,lac,max}$ [g/g/h]	LacY/CysG: $2^{\Delta\Delta Cp}$		
			Mean [-]	Std.dev.	Relative amount [%]
1	0.32	0	4.9	1.9	0.30
2	0.29	0.04	125.4	1.8	37.54
3	0.24	0.09	213.4	6.3	63.88
4	0.11	0.13	304.8	24.9	90.96
5	0.02	0.09	334.0	4.8	100.00

Figure 13 Time course of transcription of *lacZ* (A) and *lacY* (B). qPCR data were referenced to *CysG*. $2^{\Delta\Delta Cp}$ values were calculated by relating ΔCp to a reference sample which was taken before induction. Displayed percentages of transcription are in relation to the highest $2^{\Delta\Delta Cp}$ value obtained.

As shown in Figure 13, transcription levels of both genes strongly correlated with $q_{s,lac,max}$. Transcription levels gradually increased with decreasing $q_{s,glu}$ and the consequent increase in $q_{s,lac,max}$. The transcription level of *lacZ* was highest at $q_{s,glu}$ = 0.11 g/g/h and then decreased at $q_{s,glu}$ = 0 g/g/h, where also $q_{s,lac,max}$ decreased. However, the transcription level of *lacY* was highest at $q_{s,glu}$ = 0 g/g/h. We currently have no explanation for the discrepancy in the abundance of *lacY* and *lacZ* at this point, but we speculate that it might be measurement error in qPCR-analysis. However, comparing the transcript levels of both genes at $q_{s,glu}$ = 0 g/g/h and $q_{s,glu}$ = 0.24 g/g/h, where approximately the same $q_{s,lac,max}$ was reached and thus the same amount of inducer was present, we detected great discrepancies (Figure 13). The amount of mRNA for both genes was 20 – 40% higher at $q_{s,glu}$ = 0 g/g/h, supporting our hypothesis that the $q_{s,glu}$ ramp was faster than the adaptation capacity of the cells. Thus, we concluded that only static experiments reveal the true $q_{s,glu}/q_{s,lac,max}$ at low $q_{s,glu}$ levels.

Proposed Experimental Strategy to Determine the Mechanistic $q_{s,glu}/q_{s,lac,max}$ Correlation

In order to find a fast experimental strategy to determine the $q_{s,glu}/q_{s,lac,max}$-correlation for pET-based recombinant *E. coli* strains, we differently combined static and dynamic experiments and performed sensitivity analyses of the models to investigate the error of fit (Table 2).

Table 2 Comparison of parameters and NRMSE of models fitted to data from static and dynamic experiments as well as combinations thereof.

Datasets	$q_{s,lac,max}$*	K_A	$q_{s,glu,crit}$	n	$q_{s,lac,noglu}$	NRMSE
	[g/g/h]	[g/g/h]	[g/g/h]	[-]	[g/g/h]	[%]
Static data	0.13	0.094	1.02	1.48	0.040	9.11
Dynamic data (ramp up and down)	0.093	0.023	1.00	2.17	0.066	22.1
Dynamic data (ramp up)	0.072	0.392	1.00	0.83	0.066	17.9
Dynamic data (ramp up) & all static data	0.075	0.025	1.00	1.14	0.040	12.5
Dynamic data (ramp up) & two static data points (model based)	0.072	0.025	1.00	1.11	0.040	12.8

As shown in the values of NMRSE (Table 2), data derived from dynamic ramp experiments gave unsatisfactory model fits in regard to the static data-points (NMRSE = 22.1 % and 17.9 %, respectively). Combining the data from all static experiments and the ramp up experiment gave a satisfactory fit and a NMRSE of only 12.5 % (Table 2). However, since we wanted to develop a strategy comprising less experiments, we conducted a model based experimental design by sensitivity analysis [31]: we combined the data from the ramp up experiment and from two static experiments at $q_{s,glu}$ = 0 g/g/h and $q_{s,glu}$ =0.074 g/g/h, respectively, and fitted the model (Figure 14).

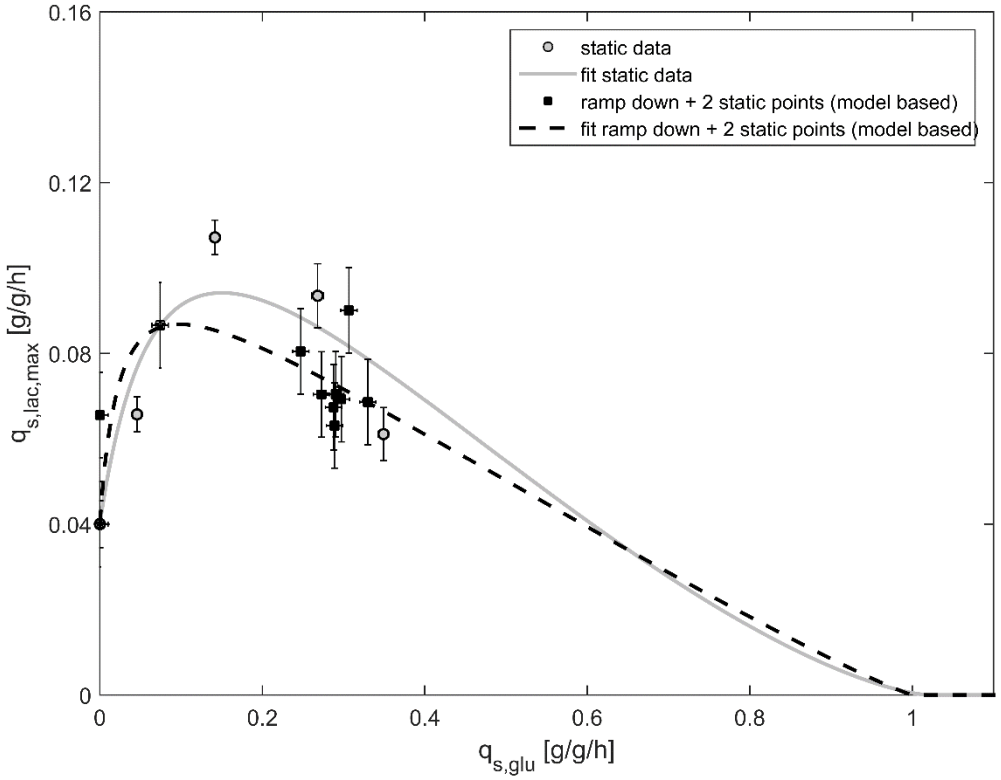


Figure 14 Comparison of the optimal fit for the static experiments and the combination of the dynamic ramp up and two static experiments derived from a model-based experimental design.

In fact, this combination of the dynamic ramp up and two static experiments gave a satisfactory fit and a NMRSE of only 12.8 % (Table 2). In terms of the deviating parameters K_A and $q_{s,lac,max}^*$ (Table 2), we had found that those parameters were hard to identify from static experiments and were therefore less trustworthy (*vide supra*). In contrast we were able to confirm parameter identifiability when we performed a practical identifiability analysis for the parameters derived from the combination of the dynamic and two static experiments. We concluded that the parameters derived from this experimental combination are in fact more trustworthy than the parameters derived from static experiments only. To prove the applicability of the model, we assumed different $q_{s,glu}$ values between 0.1 and 0.8 g/g/h and calculated the respective $q_{s,lac,max}$ values using the model derived from static experiments only as well as from the combination of the dynamic and two static experiments (Table 3).

Table 3 Comparison of $q_{s,lac,max}$ values at different $q_{s,glu}$ setpoints calculated from models derived from static experiments only as well as from the combination of the dynamic and two static experiments.

$q_{s,glu}$ [g/g/h]	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$q_{s,lac,max}$ [g/g/h] static fit = A	0.091	0.092	0.082	0.069	0.055	0.041	0.028	0.016
$q_{s,lac,max}$ [g/g/h] combination fit = B	0.087	0.081	0.072	0.061	0.050	0.039	0.029	0.018
percentual deviation = (A-B)/A	4.9%	12.2%	13.2%	11.6%	8.3%	3.4%	-3.7%	-13.8%

As shown in Table 3, the calculated values for $q_{s,lac,max}$ for both models were rather similar. A direct comparison of the two model fits gave a NRMSE of only 0.059 %, confirming the high similarity thereof.

Thus, we propose a strategy comprising only three experiments to determine the mechanistic $q_{s,glu}/q_{s,lac,max}$ correlation for a pET-based recombinant *E. coli* strain. Our strategy can be summarized as:

- Performing a batch cultivation on glucose, followed by a lactose pulse to determine the parameter $q_{s,lac,noglu}$ **(static experiment No. 1)** .
- Performing a fed-batch cultivation with lactose in excess (> 5 g/L) at a low $q_{s,glu}$ value of about 0.1 g/g/h, as the sensitivity analysis found data points in this region to be crucial for correct parameter estimation **(static experiment No. 2)**.
- Performing a **dynamic experiment**: Adapt *E. coli* to lactose at intermediate $q_{s,glu}$ of about 0.25 g/g/h for 2.0 h. However, since adaptation time might differ from strain to strain we recommend at-line HPLC measurements of sugar concentrations and biomass-estimation by OD_{600} every 30 min to reliably determine full adaptation to lactose (Supplementary Figure 3). After adaption increase $q_{s,glu}$ linearly at a rate of 0.14 $g/g/h^2$ to guarantee a total induction duration of less than 5 h in order to maintain cell fitness.
- Plotting the $q_{s,glu}/q_{s,lac,max}$ values and fitting them to the mechanistic equation to be able to determine the model parameters and thus the physiological limits of the respective *E. coli* strain.

Conclusions

In this study we were able to show that our previously generated mechanistic $q_{s,glu}/q_{s,lac,max}$ model in fact describes platform knowledge for pET-based recombinant *E. coli* strains. We found that model parameters were greatly affected by the recombinant product, pushing for physiological strain characterization of each *E. coli* strain to allow efficient recombinant protein production and to avoid sugar accumulation and the resulting osmotic stress for the cells. We compared data from different dynamic strategies to the data obtained from static experiments. Finally, we propose a strategy comprising one batch cultivation, one fed-batch cultivation as well as one dynamic experiment, to reliable determine the mechanistic model for $q_{s,glu}/q_{s,lac,max}$ and get trustworthy model parameters for a pET-based recombinant *E. coli* strain.

Material and Methods

Strains

All cultivations were conducted with the *E. coli* BL21(DE3) strain (Life technologies, Carlsbad, CA, USA). Green fluorescent protein (GFP) was expressed using a pET21a(+) plasmid. Periplasmic horseradish peroxidase (HRP) was expressed from a pET39(+) plasmid. For expression of both the recombinant scFv and the tandem-scFv a pET28a(+) plasmid was used.

Bioreactor Cultivations

Media

All fermentations were carried out in defined minimal medium according to DeLisa et al.[32]. Depending on the antibiotic resistance genes on the plasmid the medium was either supplemented with 0.1 g/L ampicillin or 0.02 g/L kanamycin. Feeds contained 250 g/L Glucose or 200 g/L Lactose, respectively.

Pre-culture

Pre-cultures were conducted by inoculating 500-mL of sterile DeLisa pre-culture medium in a 2500-mL High-Yield shake flask with frozen stocks (1.5 mL, −80 °C) and subsequent incubation in an Infors HR Multitron shaker (Infors, Bottmingen, Switzerland) at 37 °C and 230 rpm for 20 h. Bioreactors were inoculated using a tenth of the final batch volume.

Overall cultivation strategy

All cultivations comprised three phases (batch, non-induced fed-batch, induced fed-batch) including dynamic experiments during induction. Induction was performed by lactose which was applied by a pulse to reach concentrations of 20-25 g/L and then always kept higher than 5 g/L. For that purpose, at-line lactose measurements by HPLC were performed.

Static experiments

Experiments for static strain characterisation were performed in DASbox Mini Bioreactors (Eppendorf, Hamburg, Germany) with a working volume of 250 mL. The reactors were supplied with 2 vvm of a mixture of pressurized air and oxygen, the ratio was adjusted in a way to keep dissolved oxygen (dO) above 40 % during cultivation. dO was monitored using a fluorescence dissolved oxygen electrode Visiferm DO425 (Hamilton, Reno, NV, USA). The reactors were stirred constantly at 2,000 rpm. pH was monitored by a pH-Sensor EasyFerm Plus (Hamilton, Reno, NV, USA), and kept at 7.2. If necessary, it was adjusted by addition of NH_4OH (12.5 %). Base uptake was monitored via flowrates with the DASbox MP8 Multipumpmodul. A DASGIP GA gas analyzer (Eppendorf, Hamburg, Germany) was used for monitoring CO_2 and O_2 concentrations in the offgas. All process parameters were recorded and controlled by DASware control.

The batch phase was carried out at 35 °C with an initial glucose concentration of 20 g/L and yielded a biomass concentration of 8–9 g dry cell weight (DCW) per liter. When the CO_2 off-gas signal dropped, indicating the end of the batch phase or glucose depletion respectively, a fed-batch to generate biomass was conducted. Fed-batch phases were conducted at a $q_{s,glu}$ of 0.25 g/g/h. When the DCW reached 25 g/L, the temperature was set to 30°C and cultures were induced by a lactose pulse to reach a lactose concentration of 25 g/L in the bioreactor. The feed rate was adjusted to control $q_{s,glu}$ (Equation 1). DCW in the bioreactor was estimated by using a Soft-sensor-tool [33].

$$F = \frac{q_{s,glu} * X * \rho_F}{c_F}$$

Equation 1 Calculation of the feedrate

F	feedrate [g/h]
$q_{s,glu}$	specific glucose uptake rate [g/g/h]
X	absolute biomass [g]
ρ_F	feed density [g/L]
c_F	feed concentration [g/L]

Dynamic Experiments

For development of the dynamic strategy batch and fed-batch cultivations were done in a stainless steel Sartorius Biostat Cplus bioreactor (Satorius, Göttingen, Germany) with a working volume of 10 L. The reactor was stirred at 1,400 rpm. pH was monitored with an EasyFerm electrode (Hamilton, Reno, NV, USA) and was kept at 7.2 by addition of NH_4OH

(12.5 %) or HCl (18.75 %), respectively. Base and acid consumption were determined gravimetrically. For monitoring O_2 and CO_2 concentrations in the offgas a DASGIP GA gas analyzer (Eppendorf, Hamburg, Germany) was used. Aeration was performed with a mixture of pressurized air and pure oxygen at 1.5 vvm, varying the ratio of pressurized air to pure oxygen in a way that dissolved oxygen (dO) was kept above 40 % throughout all cultivations. dO was monitored with a fluorescence dissolved oxygen electrode Visiferm DO425 (Hamilton, Reno, NV, USA). Process parameters were recorded and controlled by the process information management system Lucullus (Biospectra, Schlieren, Switzerland).

The batch phase was conducted at 35 °C with an initial glucose concentration of 20 g/L and led to biomass concentrations of 8–9 g DCW per litre. After the end of the batch phase, a fed-batch to generate a biomass was carried out.

Fed-batches were conducted at a constant specific glucose uptake rate ($q_{s,glu}$) of 0.25 g/g/h. When the DCW reached 25 g/L, the temperature was lowered to 30°C and the culture was induced by lactose. During non-induced fed-batch as well as during induction with lactose, DCW was calculated assuming a constant biomass yield ($Y_{X/S}$ = 0.37 g/g, own unpublished data; Equation 2). The feed rate was adjusted to control $q_{s,glu}$ and was calculated according to Equation 1.

$$X = \frac{c_{X0} * V_R}{\rho_R} + Y_{\frac{X}{S}} * (0 - m_F) * \frac{c_F}{\rho_F}$$

Equation 2 Estimation of absolute biomass in reactor

c_{X0}	initial biomass concentration [g/L]
V_R	reactor volume [L]
ρ_R	density of fermentation broth [g/L]
$Y_{X/S}$	biomass-yield [g/g]
m_F	balance signal of feed balance [g]
ρ_F	feed density [g/L]
c_F	feed concentration [g/L]

Sampling and Analysis

Samples were taken at the beginning and end of the batch and the non-induced fed-batch. During induction, sampling was performed every 30 min to analyze DCW and OD_{600}. In addition, every 10 min supernatant for sugar analysis was collected via an in-line ceramic 0.2 µm filtration probe (IBA, Heiligenstadt, Germany). DCW was determined by centrifuging (4500g, 4 °C, 10 min) 1 mL cultivation broth, washing the obtained cell pellet with a 0.9 %

NaCl solution and subsequent drying at 105°C for 72 h. Optical density at 600 nm (OD_{600}) was determined using a Genesys 20 photometer (Thermo Scientific,Waltham, MA, USA). For staying within the linear range of the photometer (OD_{600} 0.1–0.8) samples were diluted with 0.9 % NaCl solution. A calibration correlation OD_{600} to DCW was established. Sugar concentrations were analysed via HPLC (Thermo Scientific,Waltham, MA, USA) on a Supelcogel column (Supelco Inc., Bellefonte, Pennsylvania, USA) with 0.1% H_3PO_4 as eluent at a constant flow of 0.5 ml/min. The method for sugar analysis lasted 15 min. Analysis of the chromatograms was performed using Chromeleon Software (Dionex, Sunnyvale, California, USA). In order to examine cell-death during bioreactor cultivations, fluorescence-activated cell sorting (FACS) was conducted via Cube 8 (Sysmex Partec, Görlitz, Germany) according to Langemann et al. [34].

Data-Analysis

Fitting the data to the mechanistic $q_{s,glu}/q_{s,lac,max}$-model was carried out according to our previous study [15]. In short, unknown parameters of Equation 3 were identified using the Nelder-Mead simplex method in MATLAB to minimize the objective function (Equation 4).

$$q_{s,lac} = q_{s,lac,max}{}^{*} \cdot \max\left(\left(1 - \frac{q_{s,glu}}{q_{s,glu,crit}}\right)^{n}, 0\right) \cdot \left(\frac{q_{s,glu}}{q_{s,glu} + K_A} + \frac{q_{s,lac,noglu}}{q_{s,lac,max}{}^{*}}\right)$$

Equation 3 Mechanistic $q_{s,glu}/q_{s,lac,max}$-model

$q_{s,lac}$	Specific lactose uptake rate [g/g/h]
$q_{s,lac,max}{}^{*}$	Maximum specific lactose uptake rate [g/g/h]
$q_{s,glu}$	Specific glucose uptake rate [g/g/h]
$q_{s,glu,crit}$	Critical specific glucose uptake rate up to which lactose is consumed [g/g/h]
$q_{s,lac,noglu}$	Specific lactose uptake rate at $q_{s,glu} = 0$ [g/g/h]
K_A	Affinity constant for the specific lactose uptake rate [g/g/h]
m, n	Type of inhibition (noncompetitive, uncompetitive, competitive)

$$S = \sum_{i=1}^{n} \left(\frac{q_{s,lac,max,meas,i} - q_{s,lac,max,model,i}}{\sigma_i}\right)^2$$

Equation 4 Objective function

S	objective function
$q_{s,lac,meas,i}$	i^{th} measurement of $q_{s,lac}$
$q_{s,lac,model,i}$	predicted $q_{s,lac}$ at timepoint of i^{th} measurement
σ_i	standard deviation of the i^{th} data point

For calculating $q_{s,lac,max}$-values biomass concentrations as well as lactose amounts were interpolated using a Savitzky-Golay-filter in MATLAB R2014b. A practical parameter identifiability analysis was performed by a method similar to Raue[35]: for each parameter a physiologically meaningful range [p_{min}, p_{max}] was defined and the parameter was held fix at various values inside this range. The objective function S (Equation 4) was then iteratively minimized for each parameter value with respect to the other parameters resulting in a trajectory S_p. If S_p has a minimum the parameter can be interpreted to be identifiable.

To suggest a model based experimental design to optimally estimate the mechanistic model parameters a local sensitivity analysis of the model parameters was conducted similar to Franceschini et al.[31]: the parameters were disturbed by ± 10%. Those $q_{s,glu}$ values where the deviation of the $q_{s,lac}$ values is maximal with respect to the original parameter values can be assumed to contain maximal information to estimate the parameter.

qPCR-Analysis

qPCR-Analysis was done as a commercial service offered by Microsynth AG (Balgach, Switzerland). Reference gene *CysG* was used. 2^(ΔΔCp)-values were calculated and normalized to the highest value found.

Acknowledgements

The authors thank SCIOTEC Diagnostic Technologies GmbH (Tulln, Austria) for providing the scFv and tandem-scFv strains and for fruitful collaboration.

Author Contributions

DJW and JH performed the experiments. SU assisted in data evaluation and model fitting. CH gave valuable scientific input. OS initiated and supervised the study. DJW, JH and OS wrote the manuscript.

Competing Financial Interests

The authors declare that they have no conflict of interest.

References Paperdraft

1 Huang, C.-J., Lin, H. & Yang, X. Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. *J Ind Microbiol Biotechnol* **39**, 383-399, doi:10.1007/s10295-011-1082-9 (2012).

2 Jia, B. & Jeon, C. O. High-throughput recombinant protein expression in Escherichia coli: current status and future perspectives. *Open Biol* **6**, doi:10.1098/rsob.160196 (2016).

3 Liu, M. *et al*. Metabolic engineering of Escherichia coli to improve recombinant protein production. *Appl Microbiol Biot* **99**, 10367-10377, doi:10.1007/s00253-015-6955-9. (2015).

4 Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **5**, 172, doi:10.3389/fmicb.2014.00172 (2014).

5 Tegel, H., Ottosson, J. & Hober, S. Enhancing the protein production levels in Escherichia coli with a strong promoter. *FEBS J* **278**, 729-739, doi:10.1111/j.1742-4658.2010.07991.x (2011).

6 Durania, V., Sullivanb, B. J. & Maglierya, T. J. Simplifying protein expression with ligation-free, traceless and tag-switching plasmids. *Protein Expr Purif* **85**, 9-17, doi:10.1016/j.pep.2012.06.007 (2012).

7 Marbach, A. & Bettenbrock, K. lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J Biotechnol* **157**, 82-88, doi:10.1016/j.jbiotec.2011.10.009. (2012).

8 Dvorak, P. *et al*. Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway. *Microb Cell Fact* **14**, doi: 10.1186/s12934-015-0393-3 (2015).

9 Haddadin, F. T. & Harcum, S. W. Transcriptome profiles for high-cell-density recombinant and wild-type Escherichia coli. *Biotechnol Bioeng* **90**, 127-153, doi:10.1002/bit.20340 (2005).

10 Zhang, Z. *et al*. High-level production of membrane proteins in E. coli BL21(DE3) by omitting the inducer IPTG. *Microb Cell Fact* **14**, doi:10.1186/s12934-015-0328-z. (2015).

11 Sina, M., Farajzadeh, D. & Dastmalchi, S. Effects of Environmental Factors on Soluble Expression of a Humanized Anti-TNF-α scFv Antibody in Escherichia coli. *Adv Pharm Bull* **5**, 455-461, doi: 10.15171/apb.2015.062 (2015).

12 Bashir, H. *et al*. Simple procedure applying lactose induction and one-step purification for high-yield production of rhCIFN. *Biotechnol Appl Biochem*, doi:10.1002/bab.1426 (2015).

13 Fruchtl, M., Sakon, J. & Beitle, R. Expression of a collagen-binding domain fusion protein: effect of amino acid supplementation, inducer type, and culture conditions. *Biotechnol Prog* **31**, 503-509, doi:10.1002/btpr.2048 (2015).

14 Ma, X., Su, E., Zhu, Y., Deng, S. & Wei, D. High-level expression of glutaryl-7-aminocephalosporanic acid acylase from Pseudomonas diminuta NK703 in Escherichia coli by combined optimization strategies. *J Biotechnol* **168**, 607-615, doi:10.1016/j.jbiotec.2013.08.024 (2013).

15 Wurm, D. J. *et al.* The E. coli pET expression system revisited-mechanistic correlation between glucose and lactose uptake. *Appl Microbiol Biot* **100**, 8721-8729, doi:10.1007/s00253-016-7620-7 (2016).
16 Kremling, A., Geiselmann, J., Ropers, D. & de Jong, H. Understanding carbon catabolite repression in Escherichia coli using quantitative models. *Trends Microbiol* **23**, 99-109, doi:10.1016/j.tim.2014.11.002 (2015).
17 Brückner, R. & Titgemeyer, F. Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiol Lett* **209**, 141-148, doi:10.1111/j.1574-6968.2002.tb11123.x (2002).
18 Warner, J. B. & S., L. J. CcpA-dependent carbon catabolite repression in bacteria. *Microbiol Mol Biol Rev* **67**, 475-490, doi:10.1128/MMBR.67.4.475-490.2003 (2003).
19 Spadiut, O. & Herwig, C. Production and purification of the multifunctional enzyme horseradish peroxidase. *Pharm Bioprocess* **1**, 283-295, doi:10.4155/pbp.13.23 (2013).
20 Stadlmann, V. *et al.* Novel avian single-chain fragment variable (scFv) targets dietary gluten and related natural grain prolamins, toxic entities of celiac disease. *BMC Biotechnol* **15**, 109, doi:10.1186/s12896-015-0223-z (2015).
21 Luo, Y., Zhang, T. & Wu, H. The transport and mediation mechanisms of the common sugars in Escherichia coli. *Biotechnol Adv* **32**, 905-919, doi:10.1016/j.biotechadv.2014.04.009 (2014).
22 Bettenbrock, K. *et al.* A quantitative approach to catabolite repression in Escherichia coli. *J Biol Chem* **281**, 2578-2584, doi:10.1074/jbc.M508090200 (2006).
23 Dietzsch, C., Spadiut, O. & Herwig, C. A fast approach to determine a fed batch feeding profile for recombinant Pichia pastoris strains. *Microb Cell Fact* **10**, doi:10.1186/1475-2859-10-85 (2011).
24 Zalai, D., Dietzsch, C., Herwig, C. & Spadiut, O. A dynamic fed batch strategy for a Pichia pastoris mixed feed system to increase process understanding. *Biotechnol Prog* **28**, 878-886, doi:10.1002/btpr.1551 (2012).
25 Spadiut, O., Rittmann, S., Dietzsch, C. & Herwig, C. Dynamic process conditions in bioprocess development. *Eng Life Sci* **13**, 88-101, doi:10.1002/elsc.201200026 (2013).
26 Jazini, M. & Herwig, C. Effect of post-induction substrate oscillation on recombinant alkaline phosphatase production expressed in Escherichia coli. *J Biosci Bioeng* **112**, 606-610, doi:10.1016/j.jbiosc.2011.08.013 (2011).
27 Dietzsch, C., Spadiut, O. & Herwig, C. A dynamic method based on the specific substrate uptake rate to set up a feeding strategy for Pichia pastoris. *Microb Cell Fact* **10**, doi:10.1186/1475-2859-10-14 (2011).
28 Boulineau, S. *et al.* Single-cell dynamics reveals sustained growth during diauxic shifts. *PLOS ONE* **8**, doi:10.1371/journal.pone.0061686 (2013).
29 Loomis, W. F. J. & Magasanik, B. Glucose-lactose diauxie in Escherichia coli. *J Bacteriol* **93**, 1397-1401 (1967).
30 Fulcrand, G. *et al.* DNA supercoiling, a critical signal regulating the basal expression of the lac operon in Escherichia coli. *Sci Rep* **6**, doi:10.1038/srep19243 (2016).
31 Franceschini, G. & Macchietto, S. Model-based design of experiments for parameter precision: State of the art. *Chem Eng Sci* **63**, 4846-4872, doi:10.1016/j.ces.2007.11.034 (2008).

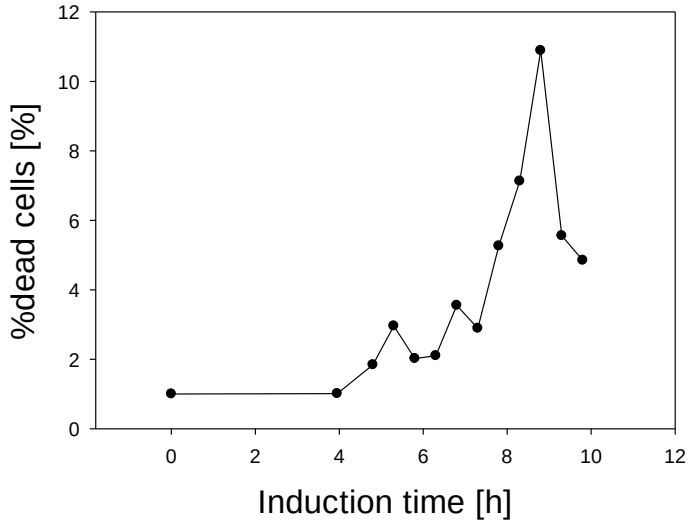
32 DeLisa, M. P., Li, J. C., Rao, G., Weigand, W. A. & Bentley, W. E. Monitoring GFP-operon fusion protein expression during high cell density cultivation of Escherichia coli using an on-line optical sensor. *Biotechnol Bioeng* **65**, 54-64, doi:doi:10.1002/(Sici)1097-0290 (1999).

33 Wechselberger, P., Sagmeister, P. & Herwig, C. Real-time estimation of biomass and specific growth rate in physiologically variable recombinant fed-batch processes. *Bioprocess Biosyst Eng* **36**, 1205-1218, doi:10.1007/s00449-012-0848-4 (2013).

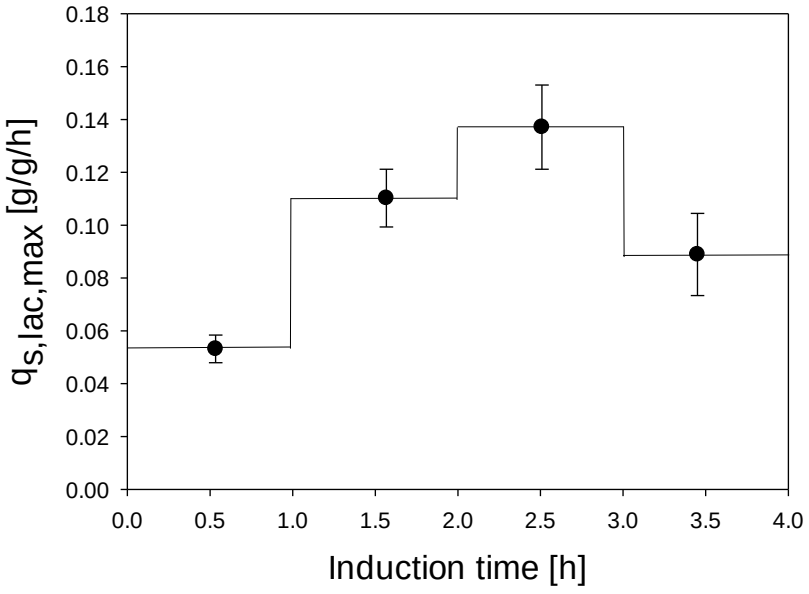
34 Langemann, T., Mayr, U. B., Meitz, A., Lubitz, W. & Herwig, C. Multi-parameter flow cytometry as a process analytical technology (PAT) approach for the assessment of bacterial ghost production. *Appl Microbiol Biot* **100**, 409-418, doi:10.1007/s00253-015-7089-9 (2016).

35 Raue, A. *et al.* Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**, 1923-1929, doi:10.1093/bioinformatics/btp358 (2009).

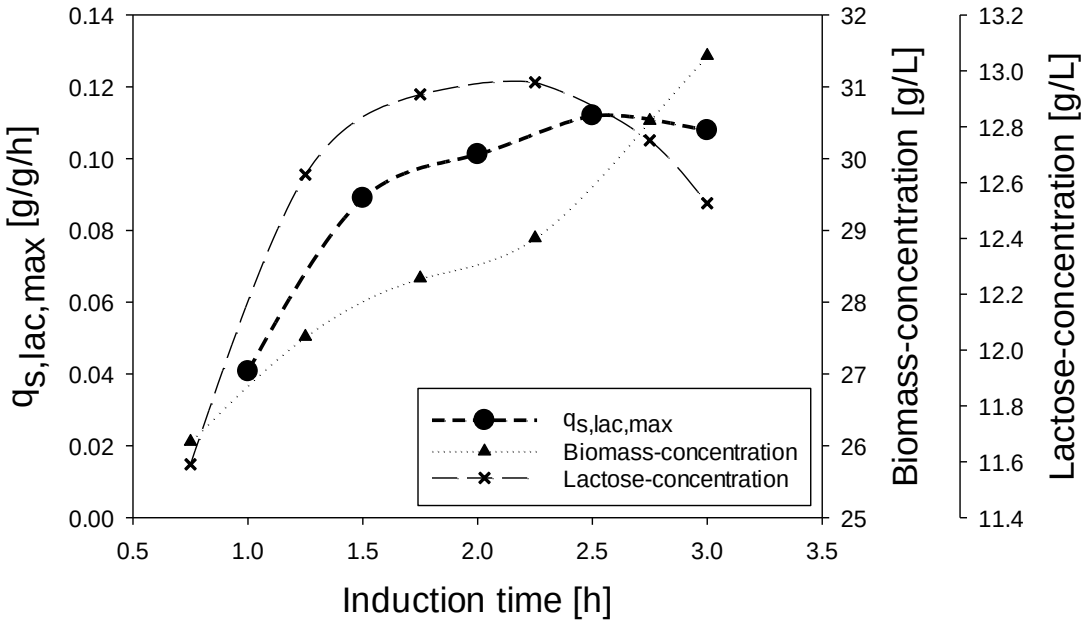
Supplementary Information



Supplementary Figure 1 Dead cells during cultivation measured by Flow cytometry.



Supplementary Figure 2 $q_{s,lac,max}$ values during first 4 h after induction for investigating duration of adaption phase. Errors were calculated by error propagation from standard deviations of DCW and sugar measurements which were carried out in triplicates.



Supplementary Figure 3 Obtained $q_{s,lac,max}$ values from at-line determination of biomass (via OD_{600}) and sugar analysis in the supernatant (via HPLC) showing a constant qs,lac and thus full adaption of the cells after 2 h.

Conclusions

The main findings of the conducted study provide answers to the previously raised questions:

1. **Is the $q_{s,glu}/q_{s,lac,max}$-correlation platform knowledge for *E. coli* BL12(DE3) cells carrying pET-plasmids?**

   Within the study we showed that the previously discovered $q_{s,glu}/q_{s,lac,max}$ dependency is transferable between *E. coli* BL21(DE3) cells carrying pET-plasmids for the production of different products. Four different proteins, including a periplasmic and three cytoplasmic products were investigated. This revealed that $q_{s,lac,max}$ in dependence on $q_{s,glu}$ always follows the same trend. However absolute values diverge. Thus it was shown that $q_{s,lac,max}$ is not only dependent on the abundance of glucose, but also on the product that is expressed.

   The $q_{s,glu}/q_{s,lac,max}$-correlation functions as platform knowledge for *E. coli* BL21(DE3) cells carrying pET-plasmids. The absolute values of the specific lactose uptake rate are reliant on the product to be expressed.

2. **How does the strategy have to be designed to rapidly characterize the concomitant sugar uptake of a novel strain with little experimental workload?**

   A method for rapid characterisation of the $q_{s,glu}/q_{s,lac,max}$-dependency was developed. It was suggested to perform the characterisation in three cultivations: a batch experiment, a conventional fed-batch cultivation and a dynamic fed-batch experiment:

   - The batch is carried out on glucose. After depletion a lactose pulse is applied to determine the lactose uptake rate in the absence of glucose.
   - The conventional fed-batch cultivation is carried out with lactose in excess at a constant $q_{s,glu}$ of around 0.1 g/g/h, as the results of the sensitivity analysis revealed this region to be essential for correct parameter determination in the model.
   - In the dynamic cultivation adaption is conducted at a $q_{s,glu}$ of 0.25 g/g/h until via at-line measurements of sugar- and biomass- concentration (OD_{600}) a constant $q_{s,lac,max}$ can be detected. Then $q_{s,glu}$ is increased linearly at a speed of 0.14 g/g/h² until glucose accumulation occurs. This strategy ensures short overall induction times for maintaining cell fitness and thus reliable data.
   - As a last point all obtained $q_{s,glu}/q_{s,lac,max}$-data is fitted by the mechanistic model.

   Applying this strategy allows attainment of the same information in only three instead of five cultivations, reducing the experimental workload severely.

Appendix: Development of an Upstream Process for the Production of a tandem-scFv in *Escherichia coli*

Apart from the developed dynamic method for strain characterisation also the upstream process for the production of a tandem scFv against celiac disease was developed within this Thesis. Different strategies aiming at boosting expression as soluble product or in the form of inclusion bodies were applied.

1. Materials and Methods

Cultivations

Cultivations were carried out as described in the paper draft.

Product Analysis

For analysis of product formation 5 ml of centrifuged fermentation broth were resuspended in lysis-buffer (100 mM Tris (12.14 g/L), 10 mM Na_2EDTA (3.7 g/L), pH 7.4) to reach a concentration of 5 g/L DCW. Cells were then disrupted by homogenization in an EmusiflexC3 Homogenizer (Avestin, Ottowa, ON, USA) at 1500 bar for 6 passages. A 15 ml Aliquot was centrifuged at 13000 g for 15 minutes at 4 °C. The supernatant was used for analysis of soluble protein. The remaining pellet containing inclusion bodies and cell debris was washed with 15 mL buffer A (50 mM Tris, 0.5 M NaCl, 0.02 % Tween 80 (w/v) pH 8) to dissolve cell debris. The suspension was centrifuged again (13000 g, 15 min, 4 °C). The obtained pellet was resuspended in 1.5 ml buffer B (50 mM Tris, 5 mM EDTA, pH 8).

Equal amounts of the suspension and Lämmli Buffer [1] were incubated at 95 °C for 10 minutes and centrifuged (14000 rpm, 21°C, 5 min.). Dilutions were performed in order to stay within concentrations of BSA-standards reaching from 0.1-0.5 g/L protein. Electrophoresis was conducted on polyacrylamide gels (Amersham ECL Gel 8-16%, GE Healthcare, Little Chalfont, UK) in 1x SDS buffer (3.03 g/L Tris, 7.2 g/L Glycine, 1.0 g/L SDS) in a horizontal chamber (Amersham ECL Gel Box, GE Healthcare, Little Chalfont, UK). Gels were pre-run at 160 V for 12 min. Electrophoresis was performed at 140 V for 80 min. Ladders from SeeBlue Plus2 Pre-stained Protein Standard (Thermo Fisher, Waltham, MA, USA) were used as size standard. Gels were stained with InstantBlue stain (Sigma Aldrich, St. Louis, MO, USA). Gel Doc XR system and ImageLab software (Bio-Rad, Hercules, CA, USA) were used for subsequent analysis including densitometric quantification.

2. Expression of Soluble Protein

In order to avoid cumbersome solubilisation and refolding procedures we wanted to produce the tandem-scFv in soluble form. Therefore, with the aim of obtaining soluble product, the established $q_{s,glu}/q_{s,lac,max}$ -correlation was applied and induction was carried out at a $q_{s,glu}$ of 0.25 g/g/h and limiting amounts of lactose: a $q_{s,lac}$ of 36 % of $q_{s,lac,max}$.

Induction by Lactose below $q_{s,lac,max}$

It had been shown in a previous study in our working group (manuscript in preparation) that most product was obtained when fermenting at values of $q_{s,glu}$ around 0.2 to 0.3 g/g/h, where the cells can take up enough inducer (as opposed to high $q_{s,glu}$, where only diminishing amounts of lactose are taken up) and where the cells are also provided with enough energy for bearing the burden of overexpressing the product (which is not given at low $q_{s,glu}$). $q_{s,lac}$,values for highest yields of soluble protein were shown to be below $q_{s,lac,max}$, where the folding machinery of *E. coli* can keep up with the transcription rate. However $q_{s,lac}$ should not be controlled at values that are too low as then not enough inducer is available leading to lessened transcription form the lac operon followed by decreased levels of T7 polymerase and consequently less product.

Goal and Strategy

Initially process parameters were geared towards high yielded production of soluble protein. Acting on the assumption that most soluble product can be obtained at $q_{s,glu}$ values around 0.2 to 0.3 g/g/h, and a $q_{s,lac}$ below $q_{s,lac,max}$, the first strategy for expressing soluble tandem scFv was cultivation at parameters found in Table 4 and specific sugar uptake rates as shown in Figure 15.

Table 4 Cultivation Parameters for production of soluble protein

Parameter	Level
Temperature	30°C
$q_{s,glu}$	0.25 g/g/h
μ	0.09 h^{-1}
Inducer	Lactose
Inducer-Amount	$q_{s,lac}$ = 0.034 g/g/h (36 % of $q_{s,lac,max}$)
pH	7.2
pO_2	> 40%
Induction time	8 h

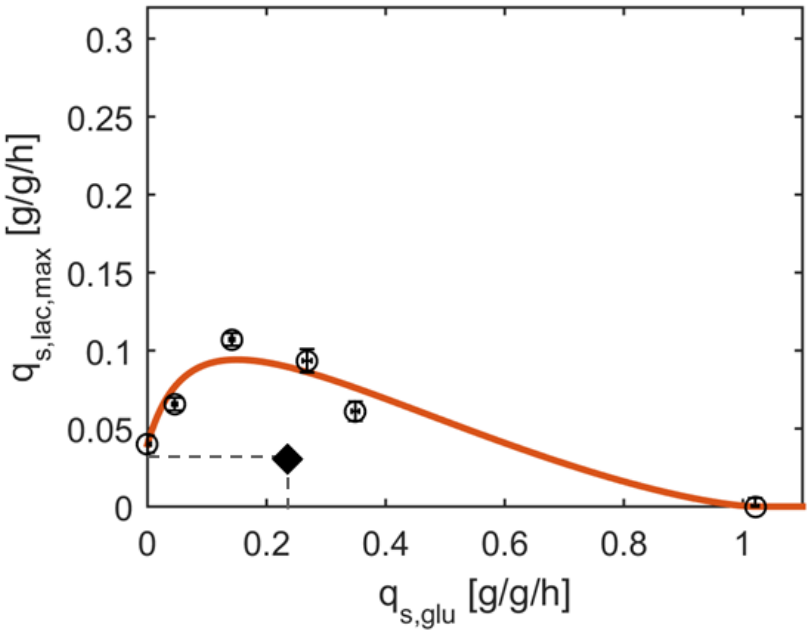


Figure 15 $q_{s,glu}/q_{s,lac,max}$ -correlation for *E. coli* BL21(DE3) cells carrying the plasmid for production of tandem-IgY. Diamond indicates specific uptake rates at which the cultivation was carried out.

Results and Discussion

The results of product analysis are shown Figure 16. The chosen parameters did not lead to formation of soluble protein, but high amounts of inclusion bodies.

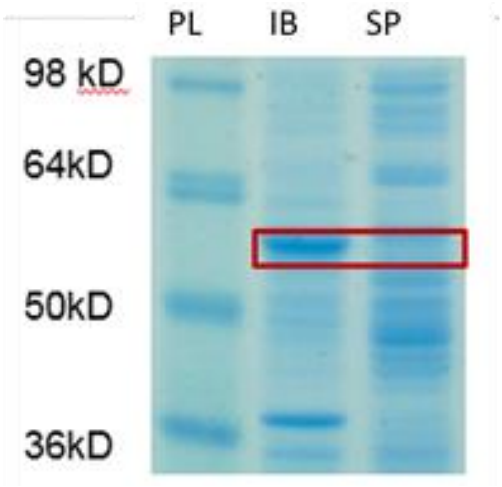


Figure 16 Obtained product from cultivation at the parameters in Table 4. Left lane: protein ladder (PL), middle lane: protein from inclusion bodies (IB), right lane: protein from the cytoplasm (SP). Product is circled red.

Even process parameters designed for yielding high amounts of soluble protein only led to formation of inclusion bodies. Most likely the scFv exhibits characteristics making soluble product formation unfeasible by only varying process parameters. The protein sequence was codon optimized for *E. coli* but it could not be precluded that the tandem-scFv exhibited disulphide bonds, which would make soluble expression in the reducing environment of the cytoplasm impractical [2]. Alternatively, if the protein showed high amounts of hydrophobic residues the chance of soluble expression is reduced [3]. Another cause for the elevated formation

of inclusion bodies might be that scFvs tend to aggregate. It has been found before that tandem-scFvs could only be expressed as inclusion bodies in *E. coli* and required refolding [4].

Conclusion

From the obtained results of the process designed for soluble protein expression it was deduced that expression of the tandem-scFv as soluble product was unlikely. Thus the further development of the upstream process focused on formation of high amounts of inclusion bodies.

3. Expression as Inclusion Bodies

As the task of expressing the tandem-scFv in soluble form seemed unfeasible, process parameters were redesigned to enhance formation of inclusion bodies. It was mainly focused on strong induction either by IPTG or by lactose in excess as this had been shown to favour formation of inclusion bodies. The temperature was not changed for facilitating comparability of cultivations.

Induction by 0.1 mM IPTG

In literature induction with IPTG is often associated with the formation of inclusion bodies [5-8]. Thus initially this inducer was chosen when redesigning the upstream process for inclusion body formation.

Goal and Strategy

The objective was to boost inclusion body formation by induction with IPTG. The parameters for cultivation are summarized in Table 5.

Table 5 Process parameters for enhanced inclusion body formation by induction with IPTG.

Parameter	Level
Temperature	30°C
$q_{s,glu}$	0.12 g/g/h
μ	0.04 h^{-1}
Inducer	IPTG
Inducer-Amount	0.1 mM
pH	7.2
pO_2	< 40 %
Induction time	8 h

Results and Discussion

Cultivating at parameters shown in Table 5 led to the results presented in Table 6.

Table 6 Results of cultivation at parameters in Table 5.

specific IB-yield [mg_{IB}/g_{DCW}]	r_p [mg_{IB}/h]	q_p [$mg_{IB}/g_{DCW}/h$]
48.1	208	7.1

IPTG-induction at 0.1 mM led to a specific inclusion body yield of 48.1 mg/g corresponding to an inclusion body formation rate of 7.05 g/g/h.

It is well known that process parameters have a strong impact on the formation of inclusion bodies, especially inducer concentration, cell density and temperature have been shown to influence product location [3,9]. Thus it was decided to investigate the effect of different process parameters in further experiments, to see if the inclusion body formation rate could be increased.

Conclusion

Induction with 0.1 mM IPTG resulted in a specific inclusion body formation rate of 7.05 g/g/h. It was decided to boost this inclusion body formation rate by varying different process parameters.

Induction by IPTG accompanied by pH and pO_2 ramps

If desired, processes can be geared towards acquisition of high inclusion body yields. Parameters known to positively influence inclusion body formation include high cell densities, strong induction (high inducer concentrations) and elevated temperaturesc[3,9,10]. Within this part of the Thesis it was attempted to enhance inclusion body formation by stressing the cells. This was conducted through dynamic variations of pH and pO_2, which was hypothesized to favour the formation of inclusion bodies.

Goal and Strategy

The objective of this part of the Thesis was to boost inclusion body formation by induction with IPTG while simultaneously stressing the cells by conducting pH and pO_2 ramps as it was postulated that this additional stress factor would increase inclusion body formation. pO_2 was regulated by accelerating or decelerating the stirrer. pH was regulated via base consumption or addition of acid respectively. The previous cultivation with constant pH and pO_2 served as control. The parameters for the cultivations are summarized in Table 7. Profiles of the conducted pH and pO_2 ramps are shown in Figure 17.

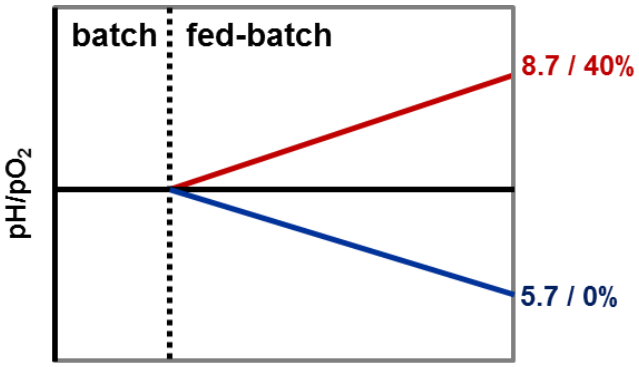


Figure 17 Profiles of dynamic variation of pH and pO_2 during cultivation.

Table 7 Process parameters for enhanced inclusion body formation by induction with IPTG while dynamically varying pH and pO_2.

Cultiv.	Temperature [°C]	$q_{s,glu}$ [g/g/h]	µ [h^{-1}]	Inducer	pO_2 [%]	pH [-]	Induction time [h]
A.1	30	0.12	0.04	IPTG, 0.1 mM	> 20 %	7.2	8
B.1		0.15			> 20 %	7.2 → 8.7	
C.1		0.14			> 20 %	7.2 → 5.7	
D.1		0.10			20 % → 40 %	7.2	
E.1		0.18			20 % → 0 %	7.2	

Results and Discussion

In total 5 cultivations are compared in the following, the previously conducted with constant pH and pO_2 and four where one parameter was either linearly increased or decreased.

In order to determine how the cells were affected by pH and pO_2 ramps viability was investigated throughout the induction phase. In Figure 18, the results of viability measurements by fluorescent activated cell sorting (FACS) are summarized.

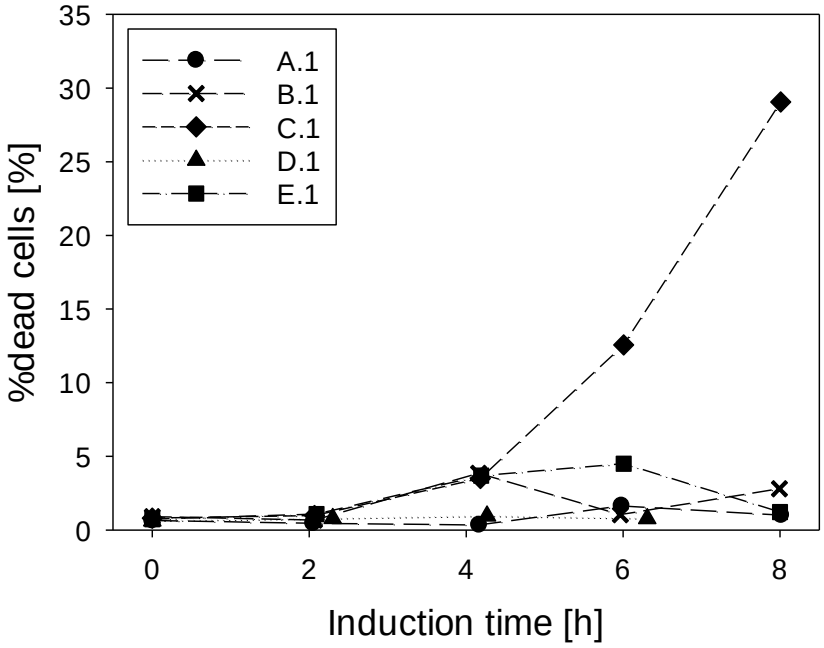


Figure 18 Results from FACS-measurements from cultivation at parameters in Table 7. Annotations according to Table 7. Left: Strongly increasing percentage of dead cells in the cultivation with rising pH.

The cells seemed to be under a lot of stress when the pH was increased to 8.7, which can be deduced from the comparatively high amount of dead cells (up to 30 %). Increasing or decreasing pO_2 as well as lowering the pH did not seem to impact the cells in regard to their viability.

The results regarding inclusion body formation in the cultivations are summarized in Table 8. The specific inclusion body formation rates are shown in Figure 19.

Table 8 Results of cultivations at parameters in Table 7. Annotation according to Table 7.

Cultivation	specific IB-yield [mg_{IB}/g_{DCW}]	r_p [mg_{IB}/h]	q_p [$mg_{IB}/g_{DCW}/h$]
A.1	48.1	208	7.1
B.1	27.4	100	4.1
C.1	-	-	-
D.1	25.7	73	3.1
E.1	31.0	123	5.3

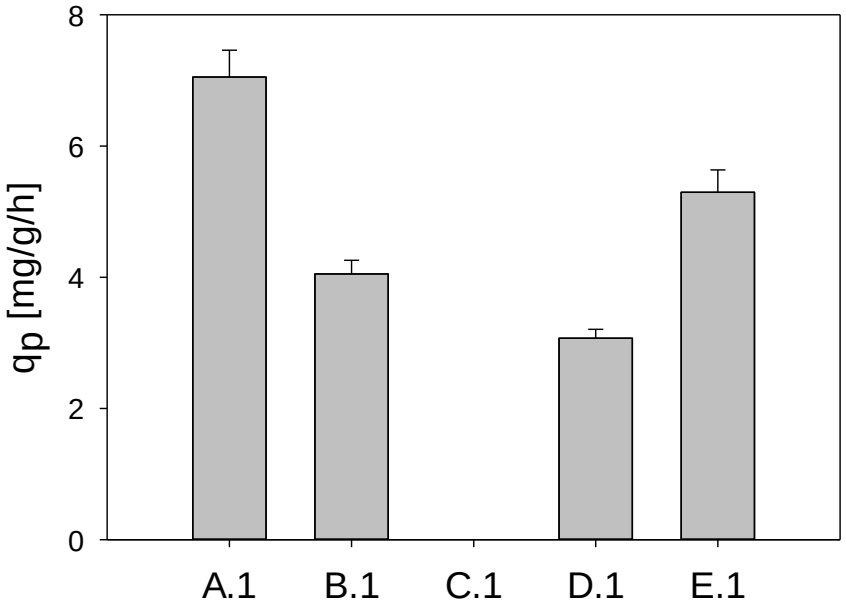


Figure 19 Specific inclusion body formation rates of cultivations at parameters in Table 7. Annotation according to Table 7. Highest specific inclusion body formation rate for cultivation at constant pH and pO_2 (A1). Lowest when increasing pH (C1).

As presented in Figure 19 stressing the cells did not lead to enhanced formation of inclusion bodies. Most product was detected in the first cultivation in which pH and pO_2 were kept constant (A.1). The lack of inclusion bodies when increasing the pH is most likely a result from the inviable cells (C.1). Product was below the limit of quantification on the gel, thus no specific inclusion body formation rate was calculated. However also increasing or decreasing pO_2 as well as decreasing the pH resulted in less product. This might be explained by the suboptimal process parameters.

Conclusion

Conducting pH and pO_2 ramps decreased productivity. The highest inclusion body formation rate was obtained when pH and pO_2 were kept constant. It was decided to next investigate the impact of parameters that are known to favour inclusion body formation: high growth rates and strong induction [3].

Induction by 1 mM IPTG

High growth rates and strong induction are described in literature to favour inclusion body formation [3,9]. Within this part of the Thesis it was aimed to boost inclusion body formation by inducing with high IPTG concentrations at elevated levels of $q_{s,glu}$ in a high cell-density cultivation.

Goal and Strategy

The objective of the following cultivation was to boost inclusion body formation by induction at high IPTG concentrations at a high $q_{s,glu}$ during induction phase in a high cell-density cultivation. The process parameters were set according to Table 9.

Table 9 Process parameters for enhanced inclusion body formation

Parameter	Level
Temperature	30°C
$q_{s,glu}$	0.3 g/g/h
μ	0.12 h^{-1}
Inducer	IPTG
Inducer-Amount	1 mM
pH	7.2
pO_2	> 20%
Induction time	5 h

Results and Discussion

Results obtained from cultivating at the parameters described in Table 9 resulted in inclusion body formation as summarized in Table 10.

Table 10 Results of cultivation at parameters according to Table 9.

specific IB-yield [mg_{IB}/g_{DCW}]	r_p [mg_{IB}/h]	q_p [$mg_{IB}/g_{DCW}/h$]
43.9	367	11.7

When comparing this result to the q_ps in Figure 19, inducing at the parameters according to Table 9 led to a higher inclusion body formation rate than the previous cultivation that was induced with lower amounts of IPTG and carried out at a lower $q_{s,glu}$. This result was expected as the chosen parameters were designed for enhanced inclusion body formation. An even higher q_p might be obtained at increased temperatures (e.g. 37 °C). This was not investigated within this Thesis as higher temperatures would aggravate comparability between cultivations. Another option to elevate inclusion body formation would be prolonged induction times, as a 5 h induction is comparatively short.

Conclusion

Inclusion body formation by induction with IPTG can be elevated when using higher amounts of the compound and by inducing at higher cell densities and higher $q_{s,glu}$.

Induction by Lactose at $q_{s,lac,max}$

As a last point we wanted to investigate changing the inducer and achieve induction by excess of lactose. It had previously been shown that cultivating at $q_{s,lac,max}$ resulted in higher inclusion body yields than when lactose was fed in limiting amounts (manuscript in preparation). This is most likely the case as strong induction leads to elevated transcription. Consequently, the folding machinery of the cells is exhausted and improperly folded proteins aggregate as inclusion bodies. It was taken advantage of this concept and process parameters were adjusted to yield high amounts of inclusion bodies [9,11].

Goal and Strategy

The objective was to yield high inclusion body amounts through inducing by excess of lactose. Formation of inclusion bodies was analysed at different levels of $q_{s,glu}$ (Figure 20). The process parameters were chosen as presented in Table 11.

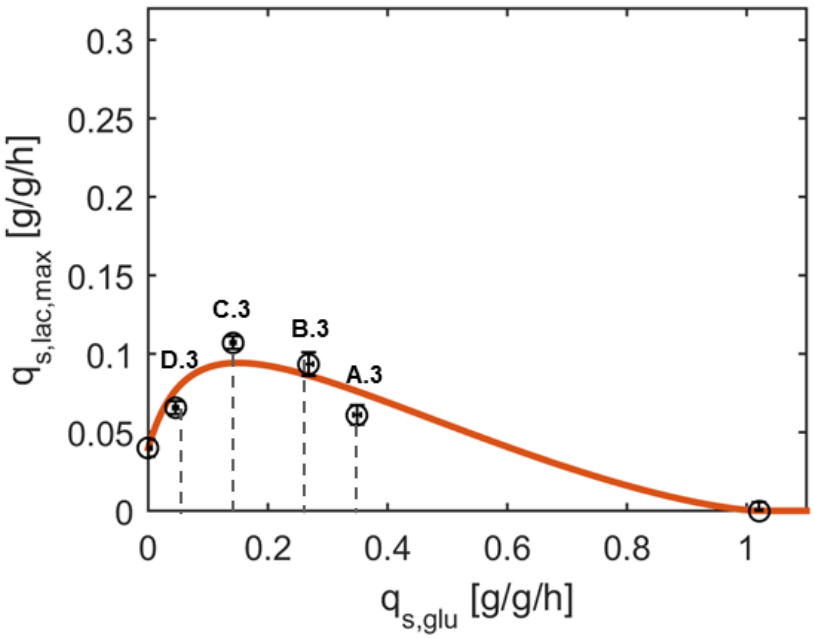


Figure 20 $q_{s,glu}/q_{s,lac,max}$-correlation for *E. coli* BL21(DE3) cells carrying the plasmid for production of tandem-IgY. A.3, B.3, C.3 and D.3 indicate points of specific sugar uptake rates at which cultivations were carried out.

Table 11 Cultivation parameters for investigating inclusion body formation at different $q_{s,glu}$ with lactose in excess.

Cultivation	Temperature [°C]	$q_{s,glu}$ [g/g/h]	μ [h^{-1}]	Inducer	pO_2 [%]	pH [-]	Induction time [h]
A.3	30	0.35	0.16	Lactose, $q_{s,lac,max}$	> 20	7.2	5
B.3		0.27	0.11				
C.3		0.14	0.06				
D.3		0.05	0.01				

Results and Discussion

As shown in Table 12 and Figure 21 the chosen parameters led to differing specific inclusion body formation rates (q_p).

Table 12 Results of cultivations at parameters in Table 11.

Cultivation	specific IB-yield [mg_{IB}/g_{DCW}]	r_p [mg_{IB}/h]	q_p [$mg_{IB}/g_{DCW}/h$]
A.3	298	1218	36.2
B.3	133	498	15.9
C.3	97.0	364	11.0
D.3	-	-	-

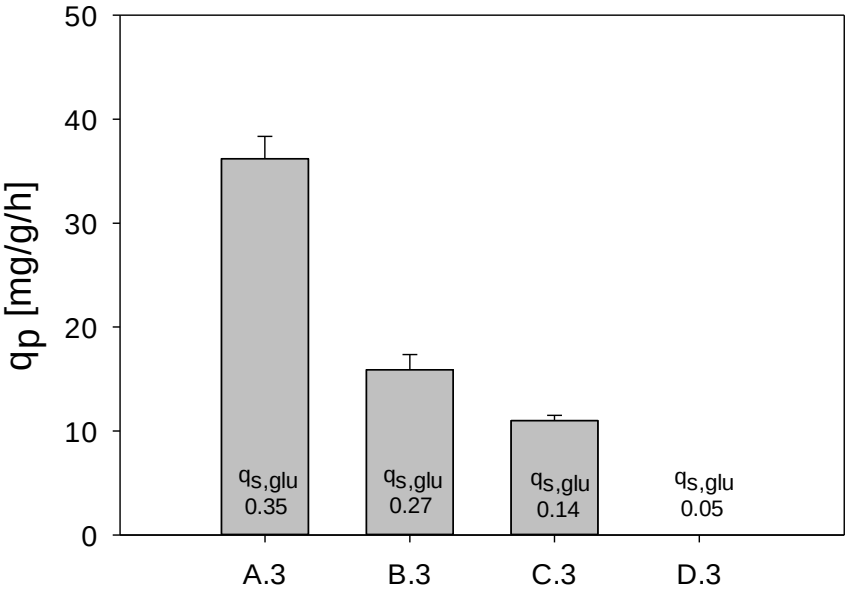


Figure 21 Specific inclusion body formation rates obtained from fermenting at process conditions listed in Table 11. Annotation according to Table 11. Specific inclusion body formation rate decreases with declining $q_{s,glu}$.

For cultivation D.3 q_p was not calculated as the product was below the limit of quantification on the SDS-gel. Regarding the other cultivations q_p increases with higher $q_{s,glu}$ reaching a value of 36.2 mg/g/h for the cultivation at $q_{s,glu}$=0.35 g/g/h. Most inducer (lactose) is taken up at a $q_{s,glu}$ of 0.27 g/g/h. At this value also transcription from the lac operon is maximized, which consequently leads to highest amounts of T7 polymerase. However, the expression of target protein is lower than when cultivating at $q_{s,glu}$ of 0.35 g/g/h. Consequently, this must be caused by events during translation or post-translation. It is hypothesized that the lowered yield at a $q_{s,glu}$ of 0.27 g/g/h is caused by a reduced availability of energy. At lower levels of $q_{s,glu}$ less glucose and thus less ATP is available to the cells. However, the burden of overexpressing

product comes along with high energy needs, which explains that most product occurs at a $q_{s,glu}$ of 0.35 g/g/h. Considering this an obvious conclusion would be to cultivate at even higher values of $q_{s,glu}$, however, for the strain and product used this is not possible as above values of 0.35 g/g/h glucose accumulation occurs.

Conclusion

When inducing with lactose in excess, highest specific inclusion body formation rates were obtained when fermenting at $q_{s,glu}$ of 0.35 g/g/h. Productivity seems to be increasing with higher $q_{s,glu}$, however above 0.35 g/g/h glucose accumulation occurs. Thus maximum productivity seems to occur at a $q_{s,glu}$ of 0.35 g/g/h.

4. Summary of the Upstream Development

Specific inclusion body formation rates obtained from cultivations at different process parameters are summarized in Table 13 and Figure 22. Space time yields are shown in Figure 23. For comparability reasons they were calculated for a biomass concentration of 60 g/L in all cultivations.

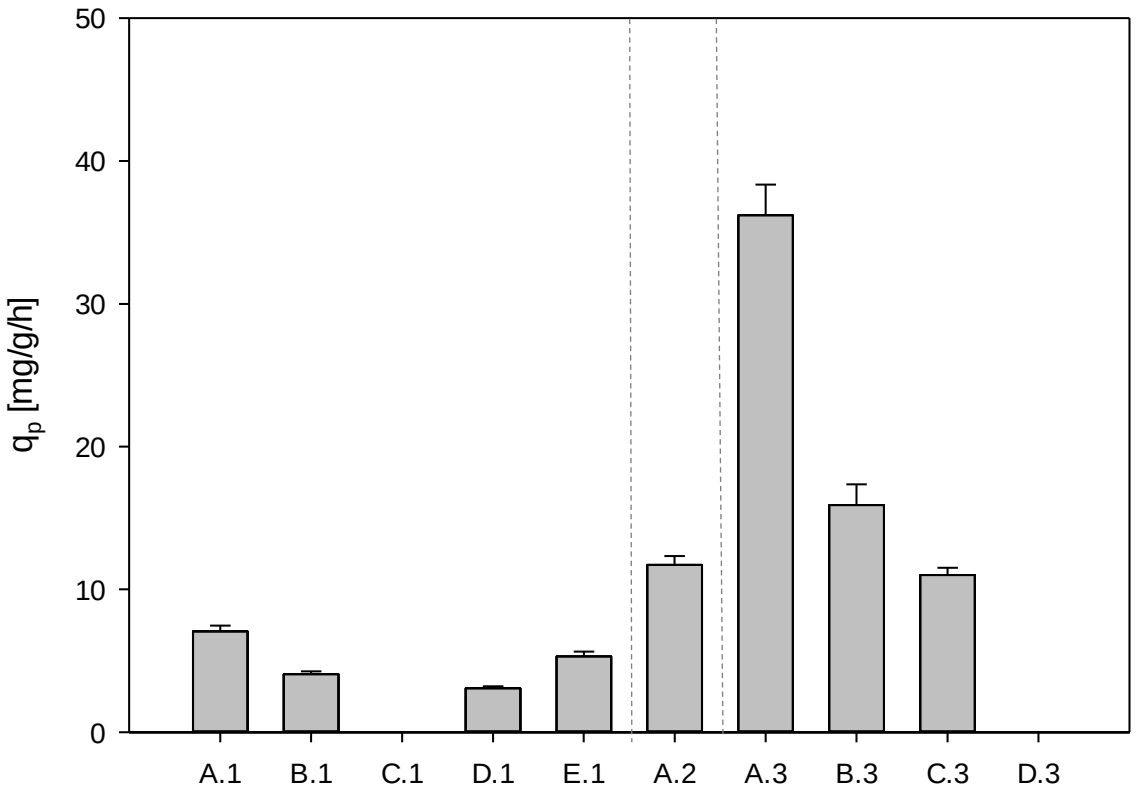


Figure 22 Specific inclusion body formation rates of all processes designed for enhanced inclusion body formation. Annotation according to Table 13. Induction by lactose in excess results in higher productivity than even high IPTG concentrations (A.2).

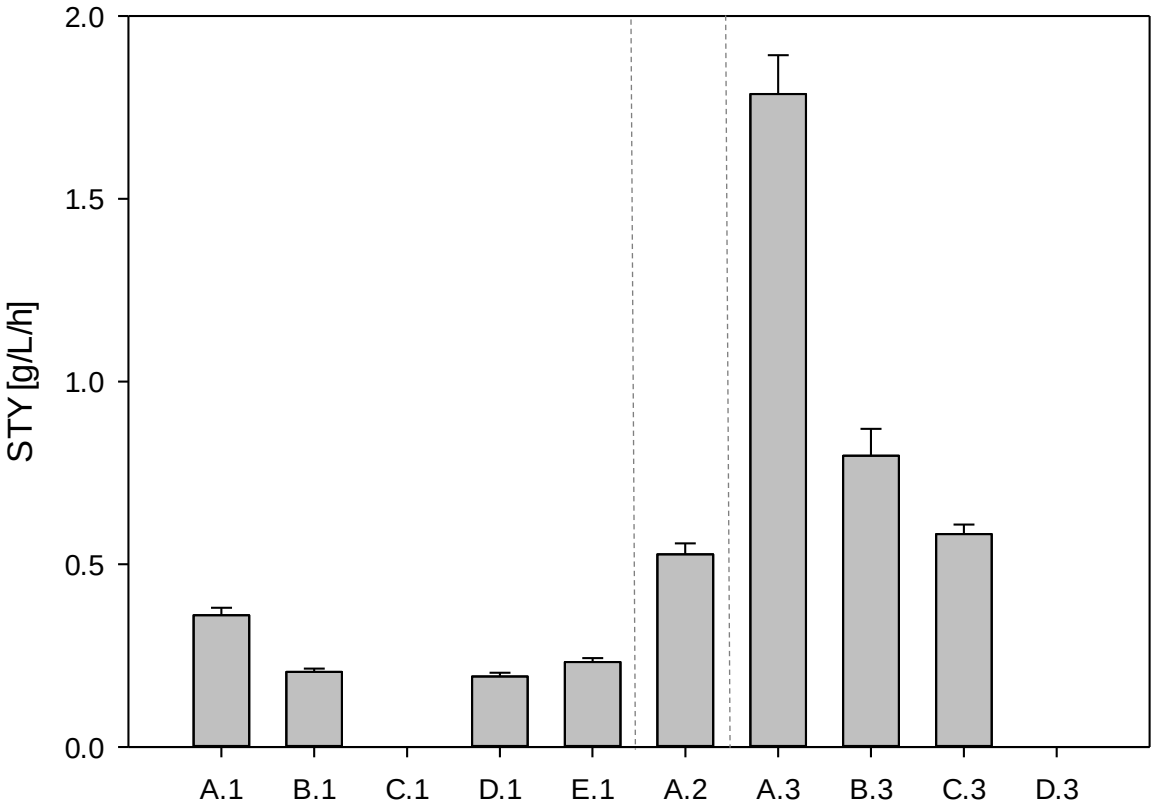


Figure 23 Space time yields for 60 g/L DCW of all processes designed for enhanced inclusion body formation. Annotation according to Table 13. Induction by lactose in excess results in higher space time yields than even high IPTG concentrations (A.2).

Table 13 Comparison of inclusion body formation of all processes designed for enhanced inclusion body formation.

ID	Inducer	$q_{s,glu}$ [g/g/h]	comm.	specific IB-yield [mg_{IB}/g_{DCW}]	r_p [mg_{IB}/h]	q_p [mg/g/h]	STY [g/L/h]
A.1	0.1mM IPTG	0.12		48.1	208	7.1	0.36
B.1	0.1mM IPTG	0.15	pH ↑	27.4	100	4.1	0.21
C.1	0.1mM IPTG	0.14	pH ↓	-	-	-	-
D.1	0.1mM IPTG	0.10	pO_2 ↓	25.7	73.3	3.1	0.19
E.1	0.1mM IPTG	0.18	pO_2 ↑	31.0	123	5.3	0.23
A.2	1 mM IPTG	0.3		43.92	367	11.7	0.53
A.3	$q_{s,lac,max}$	0.35		148.9	1218	36.2	1.79
B.3	$q_{s,lac,max}$	0.27		66.5	498	15.9	0.80
C.3	$q_{s,lac,max}$	0.14		48.5	363	11.0	0.58
D.3	$q_{s,lac,max}$	0.05		-	-	-	-

Induction with lactose results in higher inclusion body yields and higher specific and volumetric inclusion body formation rates. Also space-time-yields when inducing with lactose outcompete the ones when inducing with IPTG (even at 1 mM inducer). For sample C.1 and D.3 the concentration of the product on the gel was below the limit of quantification, thus no yields and rates were calculated.

For induction of the investigated strain expressing the tandem-scFv, lactose seems to be the superior inducer compared to IPTG. For the same induction time and comparable $q_{s,glu}$ set points (cultivations A.2 and A.3) the specific inclusion body formation rate is more than 3 times higher when inducing with lactose. Higher productivity when inducing with the disaccharide has been reported before, but mainly in regard to soluble product [12-15]. In this case it was demonstrated that even for obtaining inclusion bodies, induction by lactose seems to outcompete IPTG-induction. Approaches for further boosting inclusion body production would be cultivating at higher temperatures or fermenting at higher cell densities. The optimal approach for obtaining highest amounts of inclusion bodies for *E. coli* BL21(DE3) expressing the tandem-scFv, might be induction for 6 h at a $q_{s,glu}$ of 0.35 g/g/h with excess of lactose at 35°C and cell densities of 50 g/L.

References Appendix

1 Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).

2 Gopal, G. J. & Kumar, A. Strategies for the production of recombinant protein in Escherichia coli. *Protein Journal* **32**, 419-425, doi:10.1007/s10930-013-9502-5 (2013).

3 Fahnert, B., Lilie, H. & Neubauer, P. Inclusion Bodies: Formation and Utilization. *Adv Biochem Eng Biotechnol* **89**, 93-142, doi:DOI 10.1007/b93995 (2004).

4 Nelson, A. L. Antibody fragments: hope and hype. *MAbs* **2**, 77-83 (2010).

5 Zhang, Z. *et al.* High-level production of membrane proteins in E. coli BL21(DE3) by omitting the inducer IPTG. *Microb Cell Fact* **14**, doi:10.1186/s12934-015-0328-z. (2015).

6 Sina, M., Farajzadeh, D. & Dastmalchi, S. Effects of Environmental Factors on Soluble Expression of a Humanized Anti-TNF-α scFv Antibody in Escherichia coli. *Adv Pharm Bull* **5**, 455-461, doi: 10.15171/apb.2015.062 (2015).

7 Changa, B. & Chaudharya, P. M. High level production and one-step purification of biologically active ectodysplasin A1 and A2 immunoadhesins using the baculovirus/insect cell expression system. *Protein Expr Purif* **37**, 162-169, doi:10.1016/j.pep.2004.04.026 (2004).

8 Xu, Y. *et al.* Characterization of the T7 promoter system for expressing penicillin acylase in Escherichia coli. *Appl Microbiol Biot* **72**, 529-536, doi: 10.1007/s00253-005-0293-2 (2006).

9 Francis, D. M. & Page, R. Strategies to optimize protein expression in E. coli. *Curr Protoc Protein Sci* **Chapter 5**, Unit 5 24 21-29, doi:10.1002/0471140864.ps0524s61 (2010).

10 Choi, J. H., Keum, K. C. & Lee, S. Y. Production of recombinant proteins by high cell density culture of Escherichia coli. *Chem Eng Sci* **61**, 876-885, doi:10.1016/j.ces.2005.03.031 (2006).

11 San-Miguel, T., Perez-Bermudez, P. & Gavidia, I. Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature. *Springerplus* **2**, 89, doi:10.1186/2193-1801-2-89 (2013).

12 Bashir, H. *et al.* Simple procedure applying lactose induction and one-step purification for high-yield production of rhCIFN. *Biotechnol Appl Biochem*, doi:10.1002/bab.1426 (2015).

13 Fruchtl, M., Sakon, J. & Beitle, R. Expression of a collagen-binding domain fusion protein: effect of amino acid supplementation, inducer type, and culture conditions. *Biotechnol Prog* **31**, 503-509, doi:10.1002/btpr.2048 (2015).

14 Ma, X., Su, E., Zhu, Y., Deng, S. & Wei, D. High-level expression of glutaryl-7-aminocephalosporanic acid acylase from Pseudomonas diminuta NK703 in Escherichia coli by combined optimization strategies. *J Biotechnol* **168**, 607-615, doi:10.1016/j.jbiotec.2013.08.024 (2013).

15 Wurm, D. J. *et al.* The E. coli pET expression system revisited-mechanistic correlation between glucose and lactose uptake. *Appl Microbiol Biot* **100**, 8721-8729, doi:10.1007/s00253-016-7620-7 (2016).