Unterschrift des Betreuers

TECHNISCHE UNIVERSITÄT WIEN

DIPLOMARBEIT

The pET expression system revisited - how to boost soluble recombinant protein expression in *E. coli*

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

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

durch

Lukas Veiter

Stolzenthalergasse 16/3

1080 Wien

Wien, 21. Mai 2016

Lukas Veiter

Abstract

This Thesis deals with the production of an antibody fragment in *Escherichia coli* in preferably soluble form. Full length antibodies feature glycosylation and thus are typically produced in mammalian cells which usually involves low space time yields, risk of contamination and considerable cost for equipment and media. Since desired antigen-binding properties are located in the fragment antigen binding (Fab) region, recombinant expression of only a Fab or single chain fragment variable (scFv) instead of a whole antibody is promising and can be accomplished in *E. coli*.

The gut bacteria *E. coli* is one of the most widely used expression hosts for the production of recombinant proteins. It is superior to mammalian cells in several regards such as easier genetic manipulation and higher productivity in high cell density fermentations. The pET expression system features an inducible *lac* promotor which enables optimal cell growth in the Upstream Process (USP) until induction of gene expression, usually with Isopropyl β-D-1-thiogalactopyranoside (IPTG) as inducer. Induction by IPTG is used in most industrial fermentation processes as it cannot be metabolized by the cells and thus can be added in a simple one pulse addition. It however imposes a strong metabolic load onto the cells and is toxic at higher concentrations. Furthermore over-expression of recombinant protein in the cytoplasm results in inclusion body (IB) formation, containing the target protein in wrongly folded, non-functional form. Alternatively the *lac* operon's natural inducer lactose can induce expression of recombinant proteins equally effective.

Regarding IB formation it was found that solubility of recombinant protein can be increased by modification of fermentation parameters like decreasing the growth rate (μ). Also studies indicate that lactose induction of the *lac* operon results in enhanced solubility (SP) of recombinant protein.

The aim of this Thesis was to develop a mixed substrate feeding strategy for the production of a novel single chain fragment variable (scFv) using glucose and lactose in high density fermentation cultures and thereby increase the amount of soluble product as opposed to conventional IPTG induction. Advantages of this strategy include a higher fraction of soluble protein providing significant improvement in Downstream Process (DSP) complexity. Furthermore knowledge on the link between substrate uptake and product formation as IB or SP leads to tunability of protein production rate and form. Such a fermentation setup is challenging because lactose needs to be fed continuously since it serves both as substrate and inducer. Consequently concomitant substrate uptake must be taken into account to avoid over-feeding, as in our experience glucose presence is necessary for lactose uptake.

We found that lactose not only favours the recombinant production of soluble scFv when compared to IPTG, but furthermore that formation of soluble product can be tuned by the specific uptake rate of glucose during induction. On this basis a mechanistic correlation between specific uptake rates of lactose and glucose was determined in order to develop a corresponding model to be potentially valid for any other *E. coli* strain.

Zusammenfassung

Diese Diplomarbeit beschäftigt sich mit der Herstellung eines single-chain variable Antikörper-Fragments in *Escherichia coli* in möglichst löslicher Proteinform. Vollständige Antikörper verfügen über komplexe Glykosylierungsmuster welche eine Herstellung mittels Säugetierzellen notwendig machen, verbunden mit den bekannten Nachteilen eines solchen Verfahrens wie verhältnismäßig niedrige Ausbeuten, Kontaminationsrisiken und entsprechend hohe Kosten. Anstatt nun mit hohem Aufwand einen vollständigen Antikörper herzustellen, kann auch "nur" die Expression der Antigen bindenden Region eines Antikörpers (z.B.: fragment antigen binding = Fab bzw. single chain variable fragment = scFv) mittels *E. coli* Bakterien zielführend sein.

Bezüglich rekombinanter Proteinproduktion in *E. coli* wird oft über das pET Expressionssystem verfahren, welches über einen induzierbaren *lac* Promotor verfügt wodurch optimale Zelldichten gewährleistet sind bevor das System induziert wird. Als sogenannter "Inducer" wird im industriellen Maßstab Isopropyl ß-D-1-thiogalactopyranoside (IPTG) verwendet, da selbiges nicht metabolisiert werden kann ist ein einmaliger Zusatz praktischerweise ausreichend. Nachteile entstehen durch eine hohe metabolische Belastung durch IPTG, die wohl mitverantwortlich ist für die Entstehung von Einschlusskörperchen (Inclusion Bodies) bei der Überexpression rekombinanter Proteine. Solche Einschlusskörperchen beinhalten das Zielprotein in fehlgefalteter, unlöslicher Form und müssen unter erheblichem Aufwand denaturiert und neu gefaltet werden. Über niedrige Wachstumsraten (μ) während der Induktionsphase der Fermentation kann das Entstehen von Einschlusskörperchen weitgehend verhindert werden - stark auf Kosten der zeitlichen Ausbeuten.

Das pET Expressionssystem kann allerdings auch über Lactose, den natürlichen "Inducer" des *lac* operons induziert werden. Damit verbunden kann auch die Löslichkeit des Zielproteins entschieden verbessert werden, weil für die Zellen einiges an metabolischem Stress wegfällt. Für diese Diplomarbeit wurde deshalb ein sogenanntes "mixed-feed" System verwendet und angepasst um Glucose und Lactose gleichzeitig im Rahmen von Hochzelldichtefermentationen zu füttern. Damit ergibt sich nicht nur eine verbesserte Proteinlöslichkeit, sondern auch die Möglichkeit jene in die gewünschte Richtungen zu lenken. Lactose fungiert hier sowohl als Substrat als auch als "Inducer". Somit ist es notwendig, den Mechanismus der gemeinsamen Aufnahme von Glucose und Lactose zu kennen, um ein Überfüttern der Kultur zu vermeiden. Verkomplizierend kommt hinzu, dass Glucose in einem solchen System unbedingt benötigt wird, damit Lactose über einen aktiven Transportmechanismus überhaupt aufgenommen werden kann.

Unsere Erkenntnisse zeigen, dass die rekombinante Expression von löslichem Produkt durch den Einsatz von Lactose entschieden verbessert und kontrolliert werden kann. Auf dieser Basis wurde eine mechanistische Korrelation zwischen den Substrataufnahmeraten von Glucose und Lactose ermittelt um daraus ein simples Modell zu erstellen, welches als Ausgangspunkt für andere E. coli Stämme bzw. damit verbundene Produkte verwendet werden kann.

Danksagung

An dieser Stelle möchte ich zuallererst Prof. Christoph Herwig für die wohlwollende Aufnahme in sein Team und die damit verbundenen, ungemein günstigen Arbeitsbedingungen danken.

Besonders sei Priv.-Doz. Dr. Oliver Spadiut gedankt, der unseren Forschungsdrang in geordnete Bahnen gelenkt hat und mir dabei stets mit den richtigen Impulsen zur Seite stand, um meiner Arbeit den letzten Schliff zu geben.

Wissenschaftlicher Fortschritt verlangt nach frischen Ideen, und davon hat mein Betreuer Dipl.-Ing. David Wurm eine ganze Menge. Danke, denn ohne Deinen wachen Geist wären uns viele neue Erkenntnisse verwehrt geblieben.

Für die Unterstützung im Labor bin ich natürlich auch allen Kolleginnen und Kollegen zu Dank verpflichtet.

Abschließend möchte ich meine Familie nicht unerwähnt lassen, die mir weit über das Studium hinaus stets unter die Arme gegriffen hat.

Table of Contents

Introduction

Overview on antibodies and antibody fragments

Antibodies are immunoglobulins (Ig): glycoproteins which are formed in plasma cells as a response to immunogens which trigger an immune response. They recognise and subsequently bind the corresponding antibody generating (antigen) substance, thereby preventing an infection. The part of an antigen recognised by an antibody is called epitope and typically located on the surface of an antigen.

Regarding human therapeutics, antibodies have long been recognised for high selectivity, low toxicity and comparably long catabolic half-life when applied as serum (1). Since the first available recombinant IgG products in the 1980s, the pharmaceutical industry has steadily been expanding the market. Generally a trend to gradually replace chemically synthesized drugs can be observed. As of today, over 35 monoclonal antibodies or antibody fragments have been approved by the United States Food and Drug Administration, more importantly more than 200 related products are currently in clinical trials. As of 2015 the global antibody market comprises $80 billion and is constantly expanding, the most important sector are therapeutic and diagnostic applications (2).

Antibody structure

All Igs consist of at least four sub-units (3): two identical light chains comprising about 23 kDa and two identical heavy chains comprising 53 to 75 kDa (see Fig. 1). These sub-units associate via disulfide bonds and non-covalent interactions to form a Y-shaped molecule. There are five classes of Ig, differing in the structure of their heavy chain and sub-units. This Thesis was conducted with IgY, a functional equivalent to IgG found in chicken, which is effective against viruses and bacteria and is the most abundant type of Ig in human circulation. IgGs as wells as IgYs can be split into three 50 kDa fragments: two identical Fab fragments (fragment antigen binding) and one Fc (fragment crystallisable) fragment. Each Fab represents the region of an antibody which binds to antigens and consists of one complete light chain and the N-terminal part of the heavy chain (3). The Fc consists of the C-terminal parts of both heavy chains. Light chains mainly differ in the N-terminal parts, consequently these polypeptides are subdivided into variable (V_L) and constant regions (C_L). Heavy chains also consist of such variable (V_H) and constant regions (C_H) (3). Fc and Fab regions are connected via the hinge region.

While the Fab region defines the antibody's specificity, the Fc region causes the antibody's correct response to the antigen. This is achieved by binding to corresponding immune molecules such as Fc receptors, which subsequently lead to the desired physiological effects like cell lysis for pathogen elimination (4) through for instance macrophages. Thereby the complement system supports the process of clearing pathogens from the organism. Fig. 1 displays the general structure of an antibody.

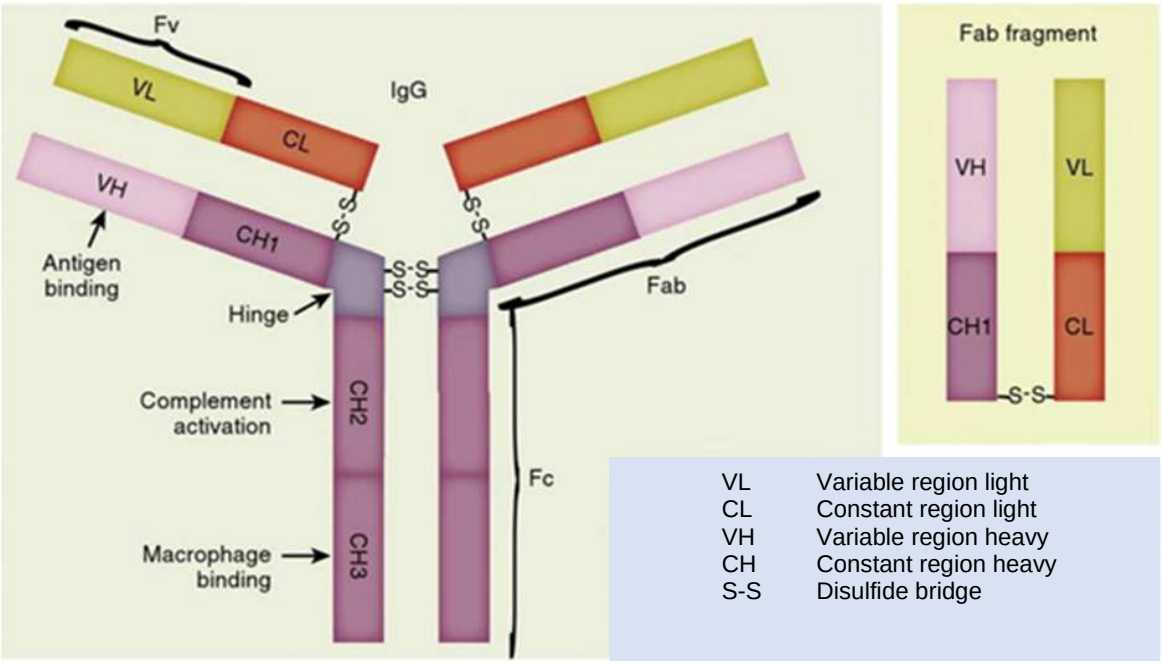


Fig. 1.: Antibody structure featuring Fab and Fc (5)

The antigen binding site is located on the tip of each Fab between the V_L and V_H - domains. Van-der-waals forces, hydrogen bonds, hydrophobic and ionic interactions are necessary to bind antigens. The dissociation constant is similar to enzyme - substrate bonds. Specificity and strength of an antibody-antigen complex is achieved by structural complementarity, consequently antibodies are highly useful reagents in medical diagnostics (6). An additional field of application is targeted cancer treatment via the detection of tumor cell antigens.

Full length native antibodies are glycosylated through post-translational modifications for stability and recognisability. While this does not have an effect on antigen detection, faulty glycosylation can cause unintended immune responses in patients after application, stability issues and a rapid clearance from the body.

Monoclonal antibodies vs. polyclonal antibodies

In antibody production a distinction between monoclonal antibodies (mAbs) and polyclonal antibodies (pAbs) has to be made. mAbs are derived from identical immune cells and all have affinity for the epitope on the same antigen, as opposed to pAbs which are able to recognise several epitopes on different antigens. pAbs are prone to some advantages (7), as high affinity and the recognition of multiple epitopes lead to a more robust detection of antigens. On the other hand there is significant batch-to-batch variability during production. Also several applications depend on the specificity of antibodies to provide reproducible results, thus mAbs are used in antibody assays to prevent cross-reaction with other proteins and enhance reproducibility.

mAbs are generally produced via the hybridoma technology: antibody generating B lymphocytes taken from mice are fused with immortal myeloma cancer cells resulting in semi-immortal hybridoma cells which produce mAbs. For this rather complex process considerable skills combined with high costs are required (7).

Alternatives to full-length antibodies

As full length antibodies feature glycosylation they are typically produced in mammalian cells. However the complicated technology involving mammalian cells results in long processing times, risk of viral contamination and considerable cost (8). Specifically glycan heterogeneity in the fragment crystallisable region (Fc) represents a major challenge (1). An alternative production strategy for full length antibodies involving microbes faces considerable obstacles as well-established bacterial hosts cannot perform glycosylation, while yeasts feature glycosylation machinery that is not compatible with mammalians.

However a full length antibody is not necessary for antigen detection, its Fab region is sufficient. For instance, in cancer treatment full length therapeutic antibodies are often not able to penetrate physical barriers of tumours (9), a possible solution for this problem are smaller fragments. Diagnostic tools like imaging often require the radioactively labelled antibody to be cleared from the body fast. Furthermore the blocking of certain signalling receptors for therapeutic reasons would only require a Fab (10). The simpler structure of the Fab enables recombinant protein production in microbial hosts (11). Consequently there are continuous efforts to manage Fab production in microbial hosts.

Naturally there are also drawbacks to Fabs which need to be addressed: The missing Fc-region reduces the half-life of Fabs significantly to days or hours as opposed to weeks and prolonged circulation of Fc-fusion proteins (via Neonatal Fc-mediated (FcRn) recycling) is not possible. This has led to recent developments which focus on Fcs featuring only one Fab "arm" (12), or generating so-called half-antibodies by reducing disulfide bonds.

Other alternatives are disulfide stabilized variable fragments (dsFv) and single chain variable fragments (scFv). For dsFvs mutations are introduced in both variable regions, disulfide bridges are formed via cysteine. scFvs are artificially manufactured antibody fragments which consist of variable domains of both the heavy and the light chain. Both domains are covalently linked by a linker sequence typically featuring glycine and serine (5). Apart from the missing Fc region, essential properties like selectivity and affinity are preserved. scFvs are mainly used in immunohistochemistry (13). The structure of dsFv- and scFv fragments is displayed in Fig. 2.

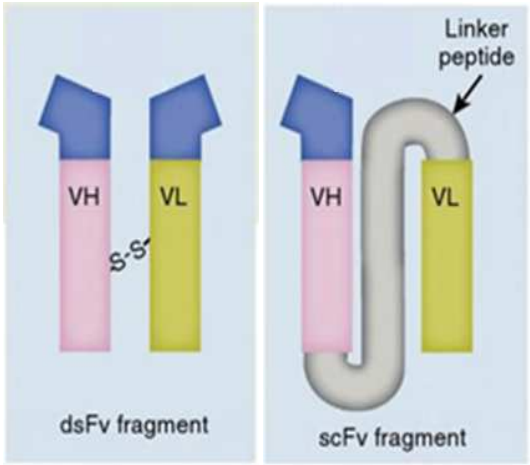


Fig. 2.: Structure of dsFv and scFv fragment (5)

Microbial host organisms for recombinant expression of mAbs and Fabs

For recombinant expression of mAbs and Fabs a series of microbial hosts comes to mind, the first and foremost distinction is to be made between prokaryotic and eukaryotic organisms (14). Bacteria are prokaryotes and as such do not possess membrane separated cell organelles, proteins are directly synthesised in the cytosol's reducing environment (15). The result can be formation of protein aggregates, so-called inclusion bodies. Inclusion bodies are generally insoluble and non-functional due to miss-folding, laborious re-folding steps are necessary to obtain functional protein: After a solubilisation step using reducing denaturants, the formation of correct disulfide bonds is achieved through removal of denaturant in presence of stabilizing additives (16).

As eukaryotic expression systems yeast can perform more complex processes involving post translational modifications like gylcosylation and the formation of disulfide bonds. Proteins can be secreted into the media via signal sequences. On the other hand yeasts also involve drawbacks: heterologous protein secretion can be tedious. Also there is the risk of proteolysis during high-cell density fermentations (17). The biggest drawback however involves hyperglycosylation, the high-mannose type glycan pattern of yeast is highly different from its human counterpart thus leading to immunogenicity (1).

Efforts involving *Saccharomyces cerevisiae*

The yeast *S. cerevisiae* is commonly used for production of recombinant proteins, mainly heavy chain antibody fragments (Hvv) - which comprise only two heavy chains - and Fabs. The first successful production of an scFv in *S. cerevisiae* was performed in 1998 employing folding assistance via chaperones and protein disulfide isomerase (PDI), a final yield of 20 mg/L was obtained (18).

Production process: Among its advantageous features are easy cultivation and well-known genetics and physiology. Many post-transitional modifications are performed by *S. cerevisiae* and expressed proteins are easily purified after secretion. Transformation is achieved via electroporation or removal of the cell wall resulting in protoplasts. Fermentation procedures typically make use of the special yeast metabolism which can shift from oxidative to fermentative. Glucose limited fed-batch conditions have proved to be the most successful fermentation technique because if carbon source is too abundant, toxic metabolites can be formed.

S. cerevisiae has an endogenous yeast episomal plasmid (Yep) plasmid to be used as a cloning vector containing the 2µ origin of replication which enables expression of genes without integration into the genome. Nevertheless genomic integration has lately been favoured because heterogenous expression and/or plasmid loss (19) can be avoided. In any case expression levels are rather low. To overcome this challenge self-replicating plasmids have been recently developed (20).

Challenges and recent advances: A major bottleneck for Fabs and mAbs production is miss-folding in the endoplasmatic reticulum. While this does not heavily affect the production of Hvv, it is problematic for single chain Fvs because of hydrophobic interaction in V_L and V_H chains (21). In yeast degradation of miss-folded protein follows endoplasmic reticulum-associated degradation pathways (ERAD), the result would be a loss of product (21).

Recent attempts to improve recombinant protein production in *S. cerevisiae* are based on enhancing the expression of corresponding proteins involved in secretion. Such approaches were effectively performed with the use of the mutant alpha mating factor 1 signal sequence, secretion of a scFv could be improved up to 16-fold over wild type. Additionally by using this approach the secretion of a full-length, correctly glycosylated IgG was improved 180-fold (22).

Efforts involving *Pichia pastoris*

Effective promoter regulation combined with the possibility of high cell densities during cultivation make *P. pastoris* the most favoured yeast expression host for antibody production (23). In 1995 a yield of 100 mg/L was achieved for the production of a scFv product (24). Current strategies to optimize secretion of scFvs involve the manipulation of chaperone and foldase levels in *P. pastoris* respectively over-expressing immunoglobulin binding protein (BIP) with methanol as substrate (25). In recent years, a number of studies involving *P. pastoris* were successfully carried out with the aim to further improve production yields of Fabs up to 450 mg/L (26).

Production process: Compared to *S. cerevisiae* ethanol formation is considerably reduced, as a result even in high cell density cultures ethanol concentrations hardly reach toxic levels. Transformation is generally achieved by integration into the genome. In *P. pastoris* the alcohol oxidase (AOXI) promoter used for foreign gene expression can be regulated efficiently, it is repressed in presence of glucose but induced when methanol is added. In addition it is a very strong promoter, during induction up to 30 % of total cell protein can be AOX. This promoter can be used for both scFvs and Fabs. Alternatively the weaker Glyceraldehydes-3-phosphate dehydrogenase (GAP) promoter is used to enhance folding time, thus ensuring more correctly folded product (1).

Challenges and recent advances: Whereas in *S. cerevisiae* homologous recombination (HR) predominates, non-homologous end joining (NHEJ) frequently occurs in *P. pastoris* leading to imprecise repairs and loss of nucleotides. To overcome this problem, the elimination of the necessary protein for NHEJ has been successfully performed resulting in an improved strain (26).

Efforts involving *Escherichia coli*

The main advantages of bacterial hosts in recombinant protein expression are well documented and involve fast growth, simple genetic manipulation, established cultivation systems combined with low costs (1). Over the years the majority of efforts concerning

antibody production have involved *E. coli* (see Table 1), the importance of other bacterial hosts is negligible.

Production process: To achieve recombinant protein expression a multitude of vectors is available, typical strong promoters are the bacteriophage T7, the *E. coli* lactose operon *(lac)* or the synthetic tryptophan operon (*trp*) promoter. Cultivation is usually performed in high cell density fermentations using glucose limited fed-batch conditions to prevent the formation of potentially inhibiting metabolites like acetate which is formed when glucose excess results in overflow-metabolism. Strategies to manage overflow-metabolism and connected adverse affects involve strain engineering as well as strain characterization. The latter is a reasonable alternative and possibility to refrain from unnecessary strain engineering - knowledge on strain physiological parameters (like specific substrate uptake rates) allows costume-made fermentation design (14).

Challenges and recent advances: Over-expression of recombinant protein is leading to inclusion body formation. Regarding inclusion bodies the main bottleneck is the development of functional and efficient refolding techniques. But inclusion bodies can also have advantages as product is incorporated in enriched form and protected against degradation via proteases. Also this way toxic effects of certain proteins on the host organism can be avoided (27).

There are also other issues: In the case of cytoplasmic expression it would be the limited ability to form disulfide bridges due to the cytoplasm's reducing environment. Possible adverse effects of foreign proteins due to faulty or missing post-translational modification are a challenge.

Because of *E. coli* 's inability to perform post-translational modifications, production of antibody related products has been predominantly limited to Fabs and scFvs. Since an oxidizing environment is required for disulfide bridge formation, efforts are being made to localize production of Fabs in the periplasm (1). The fact that most endogenous proteases are located in the cytoplasm further suggests this. First achievements were made in the late 1980s as active chimeric antibody fragments were successfully secreted (28). Expanding strategies focused on co-expressing periplasmic chaperones and foldases with the aim to obtain increased solubility and production yield. For secretion into the periplasm mostly the Sec-dependent pathway is used. This comes with the potential drawback of inclusion body formation as proteins are unfolded prior to secretion, which is promoting aggregation in the cytoplasm (1). An alternative pathway involves the signal recognition particle (SRP), here protein secretion occurs during translation, which makes inclusion body formation considerably less frequent. Combined with the co-expression of the SRP pathway related factor YiDC and using a defined nutrient feeding solution, a production yield of 90 mg/L scFv was possible (29). Proteases are also present in the periplasmic space. In order to avoid loss of product by proteolytic degradation, *E. coli* strains carrying mutations in genes encoding periplasmic proteases can be used (30).

A main disadvantage of the periplasm is its limited volume. As a result antibody production was also attempted in the cytoplasm employing *E. coli* mutants that enable an oxidative environment, for scFvs this was proven to be possible (31). Also existing cytoplasmic

chaperones (Gro EL/ES, Dna K/J, trigger factor) can be very helpful. Furthermore the cytoplasm is especially appropriate for intrabodies production. Intrabodies are Fabs that do not require disulfide bonds for their biological functionality (1), yields of 3 g/L for corresponding scFvs can be achieved (32).

To further lessen the effect of proteolysis and improve product recovery there is also the option to direct product into the extracellular space. A major challenge though is to overcome the cytoplasmic and the bacterial outer membrane, for this reason such a strategy has proven to be difficult and not efficient. In order to achieve extracellular production, pathways for secretion over both membranes can be used, specifically the hemolysin pathway (1). The hemolysin pathway can secret target protein directly from the cytoplasm into the extracellular space, such a method was published for scFvs in 2000 (33). Another study showed that the manipulation of fermentation characteristics enhances the extracellular recovery of a heavy chain single-domain antibody (V_HH fragment): Firstly the outer cell membrane permeability was increased by changing the cell growth rate resulting in 1.5-fold improvement of the ratio between released antibody and nucleic acid. Secondly cell permeability during fermentation was further increased by the use of polyethyleneimin (PEI) (34).

Though production of full-length antibodies in *E. coli* is not pursued extensively, there have been several attempts to circumvent the lack of a glycosylation system. Incorporating the glycosylation system of gram-negative bacterium *Campylobacter jejuni* into *E. coli* was performed in 2002 (35), its difference to corresponding mammalian glycosylation systems is substantial however. Consequently this approach has not been successful so far (1). Recently biologically active full length IgGs were obtained by engineering the cytoplasm of *E. coli* resulting in an oxidative environment, membrane translocation and gylcosylation were avoided altogether in the process. N-linked glycosylation in IgG is essential for in vivo function via binding to Fcg receptors, membrane receptors featuring specificity to corresponding Fcs. Fc mutations that enable IgGs to bind such receptors were identified and subsequently used to modify the underlying E. coli strains (36).

Summary on mAb and Fab expression in microbial hosts

Developments for mAb and Fab expression in microbial hosts are diverse, Table 1 is an attempt to summarise and chronicle some of the efforts.

Table 1.: Examples for successful expression of mAbs and Fabs in microbial hosts (1)

Host	Product	Notes	Product Yield	Year	Citation
S. cerevisiae	scFv	chaperons and PDI added	20 mg/L	1998	(18)
P. pastoris	scFv	first production of scFv in *P. pastoris*	100 mg/L	1995	(24)
	scFv	glycerol feeding	300 mg/L	2011	(37)
	scFv	manipulation of chaperone, foldase and BIP levels, methanol as substrate	~ 8000 mg/L	2007	(25)
	Fab	GAP promoter, overexpression of PDI	41 mg/L	2006	(38)
	Fab	recombinant humanized anti-HBsAg Fab	450 mg/L	2005	(26)
E. coli	IgG	first production of IgG in *E. coli*	150 mg/L	2002	(39)
	IgG	co-expression of the SRP pathway related factor YiDC	90 mg/L	2013	(29)

E.coli	IgG	highest yield for IgG, full length	~ 1000 mg/L	2010	(40)
	IgG	engineering of oxidative cytoplasm, avoid gylcosylation via Fc mutations	-	2015	(36)
	Fab	periplasmic protease manipulated strain	2500 mg/L	2004	(41)
	Fab	overexpression of disulfide interchange protein	3300 mg/L	2013	(32)
	v_HH	secretion into extracellular space via manipulation of fermentation conditions and cell permeability (PEI)	6000 mg/L	2015	(34)

This Thesis was performed in collaboration with an industrial project partner. Due to corresponding specifications all work was conducted with the expression host *E. coli* which therefore will be discussed in more detail in the following section.

The expression host *E. coli* and its induction

E. coli is one of the most widely used, genetically and physiologically characterized expression system for recombinant proteins. As an enteric gram-negative bacteria, rod-shaped with typically featuring flagellation in peritrichous form, its natural occurrence is the animal digestive tract which makes it a valid faecal indicator. Because of its rapid growth rate in defined media *E. coli* was chosen as a model organism early on, therefore a lot of cultivation protocols are available. Carbon sources and media costs are generally low, upscaling is simple compared to other expression hosts. Depending on the media a doubling time of 20-30 minutes can be observed in the exponential phase (42), cell densities up to 190 g/L dry cell weight can be achieved (43). The use of an inducible promotor enables optimal cell growth in the Upstream Process (USP) until induction of recombinant gene expression.

The T7 Expression system

The purpose of an expression system is the production of the protein of interest within the host cell. One of the most widely used systems is the pET vector featuring the T7 expression system (44). It features an easily controlled induction process and promoter specificity resulting in a target protein/cell protein ratio up to 50 % (45). The T7-RNA polymerase is the RNA-polymerase from the T7-bacteriophage. In combination with its own T7 promoter it is considered a strong expression system because of the specificity of the promoter which only binds to its corresponding T7 RNA polymerase. The vector contains the promoter sequence, the necessary ribosome binding site, the coding region for a protein of interest and a suitable promoter/terminator. In *E. coli* the DE3 Prophage is used to integrate the T7-polymerase into the genome, its expression is under the control of the *lac*UV5-promoter. This expression system can be induced with lactose and isopropylthiogalactoside (IPTG).

Regulation of gene expression - the *lac* operon

Under normal circumstances the *lac*-operon facilitates the response to lactose presence in the culture medium (46). For growth on lactose two proteins are necessary: lactose-permease transports lactose into the cells. There ß-galactosidase splits lactose into its mono-saccharides galactose and glucose. The physiological inducer of the *lac*-operon is 1,6-allolactose which is formed from lactose by ß-galactosidase. The genes on the *lac* operon Z,Y,A are transcribed in a row and code for ß-galactosidase (Z), lactose permease (Y) and thiogalactoside-transacetylase (A), the latter's physiological function is unknown as of now. Gene I codes for the *lac*-repressor which inhibits expression of the *lac* operon under lack of lactose.

The target sequence of the *lac*-repressor is the operator region of the *lac* operon. Without inducer present the *lac*-repressor binds specifically to three operator-sequences (O_1, O_2, O_3) and thereby prevents transcription. Furthermore the primary repressor binding site O_1 is protected from nucleases. O_1 is overlapping with the transcription starting point of the *lac*Z

gene, O_2 is located in the *lac*Z gene and O_3 is located at the end of the *lac*I gene. All three operator sequences together ensure maximum repression.

Fig. 3 provides an overview on the function and structure of the *lac* operon.

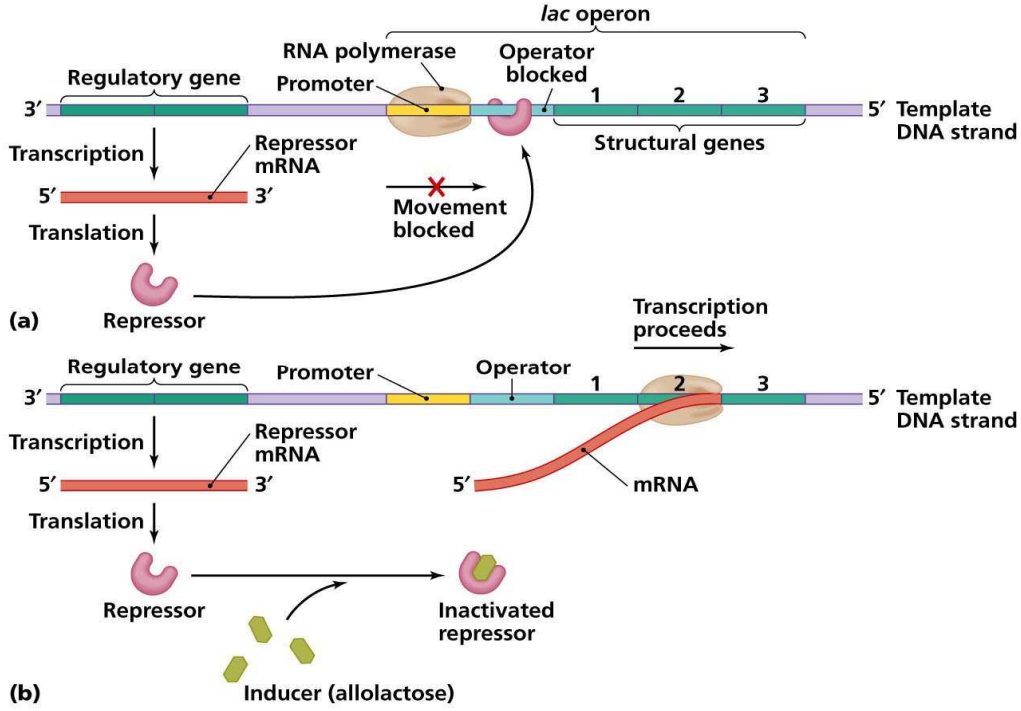


Fig. 3.: Function of the *lac* operon (47) featuring its essential components. For the sake of clarity all three operator sequences are displayed in a simplified manner as a single operator sequence. (a) displays a situation where the *lac* operon is repressed, after induction (b) the *lac* repressor dissociates from the DNA which subsequently enables transcription.

The *lac* repressor can be described as a homo-tetramer with four functional units (48): the N-terminal DNA-binding domain, a core consisting of two domains to bind the inducer, a linker featuring a short α-helix which links DNA-binding and core domain in hinge-like form and a C-terminal α-helix essential for its quaternary structure. When the inducer binds to the core, it does so at the linkage of the two domains comprising the core. This results in a change of conformation mediated through the hinge-like nature of the linker unit, consequently the repressor dissociates from the DNA because DNA-binding is not possible any longer.

Carbon catabolite repression

For *E. coli* metabolism glucose is the most important substrate. If enough glucose is present, gene expression is repressed for enzymes of the fermentative metabolism. This so-called carbon catabolite repression prevents the synthesis of unnecessary enzyme systems for energy supply. If glucose levels are down, catabolite repression is revoked in a mechanism involving cyclic adenosine monophosphate (cAMP) and the carbon activator protein (CAP), a dimer comprising two identical subunits. As soon as cAMP binds, the CAP undergoes a change in conformation, the resulting complex binds to the promoter region of the *lac*

operon and stimulates transcription - assuming the *lac* repressor has dissociated from the DNA before.

As a result two events are critical to ensure induction of the T7 expression system:

- The *lac* repressor must dissociate from the DNA after binding of the respective inducer (lactose or IPTG).
- Carbon catabolite repression must be suspended when glucose concentration becomes limiting. This happens through increased cAMP levels resulting in the transcription stimulating conformation of CAP.

Induction of the T7 expression system

A major advantage of the T7-based expression system is its high expression level when compared to other prokaryotic systems. On the other hand its main disadvantage is the accompanied formation of inclusion bodies in the cytoplasm. As a result a many studies exist with strategies to improve protein solubility, such as using low temperatures in the induction phase (49). Furthermore it was found that solubility of recombinant protein can be enhanced by modification of fermentation parameters like decreasing the growth rate (μ) by changing the feeding-profile during induction in fed-batch cultivations (50). A much broader, at least equally effective approach would be to change the inducer itself: The T7 expression system can be induced with IPTG or 1,6-allolactose which is derived from lactose via ß-galactosidase (see Fig.).

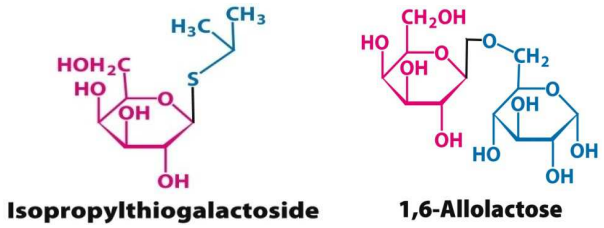


Fig. 4.: structure of inducers IPTG and allolactose

IPTG

As a structural equivalent to lactose, IPTG is considered state-of-the-art in production scale environments. Its sulfur bond cannot be hydrolyzed by ß-galactosidase, thus it cannot be metabolized by the cells and in batch fermentations inducer concentration remains stable throughout induction. Additionally it binds more irreversibly to the *lac* repressor than primary occurring allolactose, consequently even concentrations as low as 0.05 mM have been used successfully (51). IPTG can be added in a range of concentrations, however cost efficiency has to be observed carefully in industrial applications. Uptake of IPTG is closely related to the presence of lactose permease. In 2012 a study showed that a much higher inducer accumulation is found when transport occurs via lactose permease which signifies that IPTG uptake is similar to lactose uptake (52).

Negative aspects of IPTG induction include that it imposes a metabolic load onto the cells and is toxic at higher concentrations. As a result IPTG induction does not necessarily equal

high product yields. If *E. coli* is in fact overproducing product due to IPTG induction, excessive protein accumulation in the cytoplasm will lead to inclusion body formation, which in turn has a negative effect on cell viability respectively cell growth. Costs for IPTG are also rather high, especially in large-scale applications.

Lactose

While lactose is certainly not as prominently used as IPTG, there nevertheless have been studies which examine its inducer-potential. Some come to the conclusion that protein expression is equally effective (53). More importantly there is a clear indication that lactose induction enhances the solubility of recombinant protein (54) (55).

Rather low concentrations of lactose are sufficient for induction. The works of Pei et al. (56) and Mayer et al. (57) showed that a constant amount of 62 µM lactose is needed for induction at all growth rates. For this to happen a sufficient amount of lactose has to be transported into the cell: *Lac* permease classifies as a symporter. A proton gradient is used for the transportation of lactose into the cell. Such an active transport is ATP dependent, consequently glucose and lactose uptake are interrelated and both will also be used as substrate, with a notable preference for glucose.

Fig. 5 provides an overview glucose and lactose uptake.

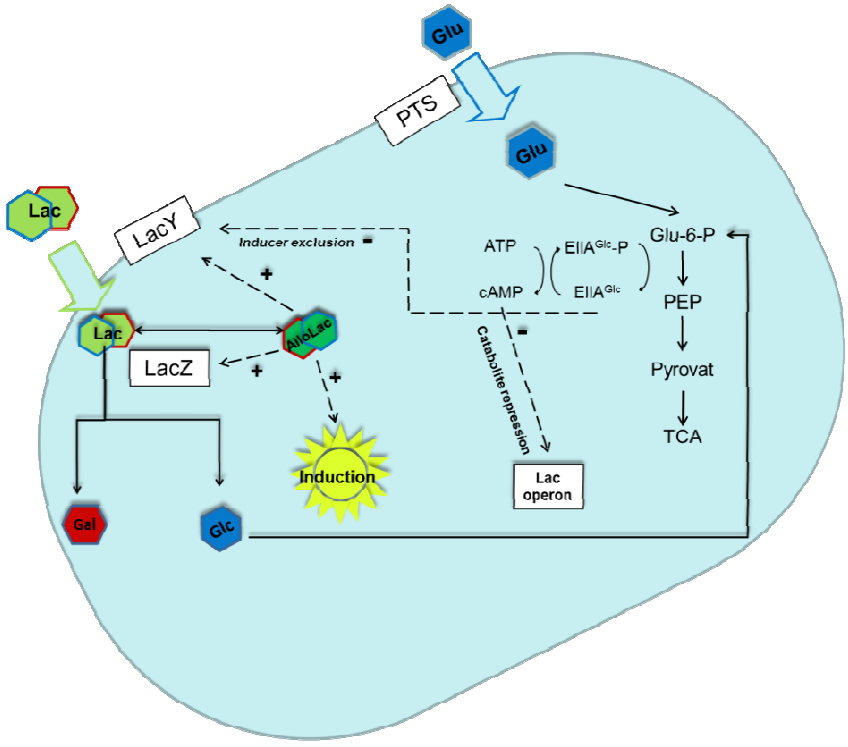


Fig. 5.: Schematics of lactose and glucose uptake - In the cell glucose gets phosphorylated to glucose-6-P whereas the enzyme II A (EIIA) domain of the glucose transporter $EIIA^{Glc}$-P gets converted to $EIIA^{Glc}$ which inhibits the uptake of non-phosphotransferase system (PTS) carbon sources. If glucose concentration is limited, the rephosphorylation of $EIIA^{Glc}$ to $EIIA^{Glc}$-P converts ATP to cAMP. cAMP binds to the catabolite activator protein (CAP). As a result the CAP undergoes a change in conformation, binds to the promoter region of the *lac* operon and thereby stimulates transcription of the *lac* operon's genes involving lactose permease.

There are several factors to be considered concerning lactose induction. First of all there is a notable delay from inducer application till induction which can be explained by the fact that ß-galactosidase has to form the actual inducing component allolactose from lactose, which subsequently must bind to the *lac* repressor. Combining this delayed response to lactose with the ability of *E. coli* to switch between carbon sources, there have been attempts to establish a so-called autoinduction of recombinant protein expression: If glucose and lactose are present in the medium, lactose is only taken up after glucose has reached limiting concentrations (57). As a result cells could grow on their preferred substrate glucose to suitable cell densities before induction happens by lactose uptake. The trouble with this is that cells enter a starvation response shortly after glucose depletion which results in a lag phase. Before *E. coli* can switch to lactose metabolism, enough *lac* permease has to be produced which leads to further delay. A promising strategy to avoid this is the use of glucose-limited conditions. It was previously found that adding lactose shortly before glucose depletion has positive effects on the cell's adaption time to the new substrate (58). Promising results could also be achieved when low amounts of glucose were continuously added throughout the induction phase via an exponential feeding profile (56). During such fed-batch conditions the cells take up glucose instantly, concentrations in the media remain virtually non-existent. Lactose is then added via pulses. While lactose uptake is definitely impaired by glucose presence, it is nevertheless sufficient for induction.

Other research has examined glucose and lactose fluxes during autoinduction using enzymatic glucose release: a medium containing lactose was provided in a bioreactor and a soluble glucose polymer was added. Results indicate that such autoinduction cultures are scalable for bioreactor cultivations (57). The autoinduction method allows for process simplification as no additional components have to be added during fermentation, also enhanced cell densities and product yields have been reported (57). However for industrial applications such a strategy would be difficult to control and reproduce.

An additional option would be to induce with lactose but use glycerol as primary carbon source, which is an already established method. While glycerol generally ensures high cell densities, there are challenges to recombinant protein expression as the uptake of lactose is heavily affected by aeration conditions. Therefore such fermentation processes have issues regarding reproducibility (59).

When induction is performed with lactose, one has to keep in mind that ß-galactosidase action increases the number of potential substrates as both glucose and galactose are formed. In 1980, a study showed that metabolites formed from lactose are mainly secreted into the medium: namely glucose, galactose and interestingly even allolactose. This was entirely unexpected, happening regardless of strain, growth conditions and product metabolism (60). It has been speculated that product secretion might be a storage device for products that are not immediately needed when substrate concentrations are abundant. Also there is the possibility that it is preferred by the cells from an energetic viewpoint. Secretion of glucose and galactose is performed by permeases dependent on a proton-gradient in the opposite direction of lactose transport, leading to the hypothesis that such a

transport is interrelated in both directions. Allolactose is probably secreted in order to prevent it from hydrolysis, thereby ensuring that inducer concentration stays constant (60).

Concluding it can be stated that lactose induction is more complex than the use of IPTG. It serves both as substrate and inducer, thus lactose uptake is highly dependent on the physiological state of the cells. Since a total switch between carbon sources is time consuming and stresses the cells unnecessarily, it would be worthwhile to further develop glucose-limited fed-batch processes with lactose induction - especially involving a mixed-feed solution. The main challenge here is to establish controlled fermentation conditions. As previously mentioned lactose and glucose uptake are interrelated. Therefore it is necessary to establish the correlation between both substrate uptake rates in order to develop a stable and reproducible bioprocess.

Goals and Scientific Questions of this Thesis

In this Thesis *E. coli* BL21(DE3) and the pET expression system were used for the production of a novel single chain fragment variable (scFv) (61). The goal of this study was to increase the amount of soluble product. It was hypothesized that this could be achieved by lactose induction as opposed to conventional induction via isopropylthiogalactoside (IPTG).

This task could be divided into three main scientific questions:

1) Does lactose favor the recombinant production of soluble scFv when compared to IPTG?

For this purpose several fermentations involving IPTG or lactose induction were conducted and analyzed. In the process the following parameters were varied: inducer type, inducer concentration and growth rate.

2) Can the formation of soluble product be influenced by the specific uptake rate of glucose during induction with lactose?

Different induction strategies for recombinant protein expression were tested including induction pulse-experiments and a mixed-feed (glucose and lactose) setup to investigate interdependent and concomitant substrate uptake.

3) Is it possible to determine a mechanistic correlation between specific uptake rates of lactose and glucose?

By determining the mechanistic correlation between both specific uptake rates it would be possible develop a model valid for basically any *E. coli* strain. Such a model also enables process control in a satisfactory manner as substrate uptake can be predicted while over-feeding can be avoided.

These scientific questions will be addressed in the following section consisting of our paper which was accepted for publication in the journal Applied Microbiology and Biotechnology. Further efforts beyond the scope of the paper regarding product quantification are summarized in the appendix, corresponding standard operating procedures (SOPs) are gathered as well.

References Introduction

1. Lee, Y. J., and K. Y. Jeong, Challenges to production of antibodies in bacteria and yeast. Journal of Bioscience and Bioengineering, Available online 23 April 2015.

2. Biocompare: 2015 Antibody Market Report, http://www.biocompare.com/Editorial-Articles/177815-2015-Antibody-Market-Report (11.03.2016).

3. Voet, D., Voet, J., Pratt, C.W., Lehrbuch der Biochemie. Wiley-VCH, 2010. p.230-231.

4. Woof, J.M. and D.R.Burton, Human antibody–Fc receptor interactions illuminated by crystal structures, 2004. 4: p.89 - 99.

5. Janeway, C.A., Travers, P., Walport, M. and M.J. Shlomchik, Immunobiology, 5th edition. New York: Garland Science, 2001.

6. Voet, D., Voet, J., Pratt, C.W., Lehrbuch der Biochemie. Wiley-VCH, 2010. p.234.

7. Abcam.com: A comparison between polyclonal and monoclonal antibodies, http://www.abcam.com/protocols/a-comparison-between-polyclonal-and-monoclonal (07.04.2016).

8. Khan K.H. Gene Expression in Mammalian Cells and its Applications. Adv Pharm Bull. 2013 Dec; 3(2): 257–263.

9. Christiansen, J., and Rajasekaran, A.K., Biological impediments to monoclonal antibody-based cancer immunotherapy. Mol. Cancer Ther., 2004. 3: p.1493-1501.

10. Absolute Antibody.com: Antibody Fragments, http://absoluteantibody.com/antibody-resources/antibody-engineering/antibody-fragments/ (10.09.2015).

11. Carter, P., et al., High Level Escherichia coli Expression and Production of a Bivalent Humanized Antibody Fragment. Nature Biotechnology, 1992. 10: p. 163 - 167.

12. Martens, T., et al., A Novel One-Armed Anti-c-Met Antibody Inhibits Glioblastoma Growth In vivo. Clin Cancer Res, 2006. 12: p.6144–6152.

13. Zhong, Y.W., et al., Preparation of human single chain Fv antibody against hepatitis C virus E2 protein and its identification in immunohistochemistry. World J Gastroenterol, 2002. 8(5): p.863-867.

14. Spadiut, O., et al., Microbials for the Production of Monoclonal Antibodies and Antibody Fragments. Trends in Biotechnology, 2014. 32.1: p.54–60.

15. Petrausch, U., Etablierung eines Expressionssystems in Pichia pastoris zur Herstellung von rekombinanten Fusionsantikörpern am Beispiel des A33rbGFP Fusionsantikörpers. Medizinische Fakultät Charité - Universitätsmedizin Berlin, 2004. Dissertation.

16. Yamaguchi, H. and M. Miyazaki, Refolding Techniques for Recovering Biologically Active Recombinant Proteins from Inclusion Bodies. Biomolecules, 2014. 4: p.235–251.

17. Frenzel, A., Hust, M., and T. Schirrmann, Expression of Recombinant Antibodies. Frontiers in Immunology, 2013. 4, p.217.

18. Shusta, E. V., Raines, R. T., Pluckthun, A., and Wittrup, K. D., Increasing the secretory capacity of Saccharomyces cerevisiae for production of single-chain antibody fragments. Nat. Biotechnol., 1998. 16, p.773-777.

19. Park, Y.N., et al., Application of the FLP/FRT system for conditional gene deletion in yeast Saccharomyces cerevisiae. Yeast, 2011. 28: p.673–681.

20. Maury, J., et al., Reconstruction of a bacterial isoprenoid biosynthetic pathway in Saccharomyces cerevisiae. FEBS Lett., 2008. 582: p.4032–4038.

21. Joosten, V.; et al., The production of antibody fragments and antibody fusion proteins by yeasts and filamentous fungi. Microbial Cell Factories, 2003. 2:1.

22. Rakestraw, J. A., et al., Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in Saccharomyces cerevisiae. Biotechnol. Bioeng., 2009. 103: p. 1192–1201.

23. Reinhart, C. and C. Krettler, Das Pichia pastoris-Expressionssystem. Spektrum Akademischer Verlag, 2012. p. 432.

24. Ridder, R., et al., Generation of rabbit monoclonal antibody fragments from a combinatorial phage display library and their production in the yeast Pichia pastoris. Biotechnology (NY), 1995. 13, p.255-260.

25. Damasceno, L. M., Anderson, K. A., Ritter, G., Cregg, J. M., Old, L. J., and C. A. Batt, Cooverexpression of chaperones for enhanced secretion of a single chain antibody fragment in Pichia pastoris. Appl. Microbiol. Biotechnol., 2007. 74, p.381-389.

26. Ahmad, M., et al., Protein Expression in Pichia Pastoris: Recent Achievements and Perspectives for Heterologous Protein Production. Applied Microbiology and Biotechnology, 2015. 98.12: p.5301-5317.

26. Ning, D., et al., Production of recombinant humanized anti-HBsAg Fab fragment from Pichia pastoris by fermentation. J. Biochem. Mol. Biol., 2005. 38, p.294-299.

27. Mayer, M., Expression und Faltung von Antikörpern im Cytoplasma von E. coli und Einfluss von Faltungshelfern auf die Antikörperfaltung in vitro. Technische Universität München, Dissertation, 2003. p.25.

28. Better, M., Chang, C. P., Robinson, R. R., and A.H. Horwitz, Escherichia coli secretion of an active chimeric antibody fragment. Science, 1988. 240, p.1041-1043.

29. Lee, Y. J. and K. J. Jeong, Enhanced production of antibody fragment via SRP pathway engineering in Escherichia coli. Biotechnol. Bioprocess Eng., 2013. 18, p.751-758.

30. Chen, C., et al.: High-level accumulation of a recombinant antibody fragment in the periplasm of Escherichia coli requires a triple-mutant (degP prc spr) host strain. Biotechnl. Bioeng., 2004. 85, p. 463-474.

31. Sonoda, H., Kumada, Y., Katsuda, T., and H. Yamaji, Functional expression of single-chain Fv antibody in the cytoplasm of Escherichia coli by thioredoxin fusion and co-xpression of molecular chaperones. Protein Expr. Purif., 2010. 70, p. 248-253.

32. Ellis, M. and D. P. Humphreys, Bacterial host strain expressing recombinant DsbC. WO 2011086141 A1, 2013.

33. Fernandez, L. A., Sola, I., Enjuanes, L., and V. de Lorenzo, Specific secretion of active single-chain Fv antibodies into the supernatants of Escherichia coli cultures by use of the hemolysin system. Appl. Environ. Microbiol., 2000. 66, p. 5024-5029.

34. Voulgaris, I., et al., Enhancing the selective extracellular location of a recombinant E. coli domain antibody by management of fermentation conditions. Applied Microbiology and Biotechnology, 2015. 99, p. 8441-8453.

35. Wacker, M., et al, N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. Science, 2002. 298, p.1790-1793.

36. Robinson, M.-P., et al., Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. Nat. Commun., 2015. 6, p.8072.

37. Khatri, N. K., Gocke, D., Trentmann, O., Neubauer, P., and F. Hoffmann, Single-chain antibody fragment production in Pichia pastoris: benefits of prolonged pre-induction glycerol feeding. Biotechnol. J., 2011. 6, p. 452-462.

38. Gasser, B., Maurer, M., Gach, J., Kunert, R., and D. Mattanovich, Engineering of Pichia pastoris for improved production of antibody fragments, Biotechnol. Bioeng., 2006. 94, p.353-361.

39. Simmons, L. C., et al., Expression of full-length immunoglobulins in Escherichia coli: rapid and efficient production of aglycosylated antibodies. J. Immunol. Methods, 2002. 263, p. 133-147.

40. Reilly, D. E. and D.G. Yansura, Production of monoclonal antibodies in E. coli. Biotechnology: pharmaceutical aspects. Current trends in monoclonal antibody development and manufacturing, Springer: AAPS Press, New York, 2010. 14, p. 295-308.

41. Chen, C., et al., High-level accumulation of a recombinant antibody fragment in the periplasm of Escherichia coli requires a triple-mutant (degP prc spr) host strain. Biotechnol. Bioeng., 2004. 85, p. 463-474.

42. Casali, N., Escherichia coli host strains. Methods in Molecular Biology, 2003. 235: p.27.

43. Shiloach, J. and R. Fass, Growing E. coli to high cell density--a historical perspective on method development. Biotechnol Adv, 2005. 23(5): p. 345-57.

44. Rosano, G.L., and E.A. Ceccarelli, Recombinant protein expression in Escherichia coli: advances and challenges. Front.Microbiol., 2014. 5,p.172.

45. Baneyx, F., Recombinant protein expression in Escherichia coli. Curr Opin Biotechnol, 1999. 10(5): p. 411-21.

46. Voet, D., Voet, J. and C.W. Pratt, Lehrbuch der Biochemie. Wiley-VCH, 2010. p.1115-1118.

47. Moran, L.A., Horton, R.A., Scrimgeour, G., and M. Perry, Principles of Biochemistry, 5/E. Pearson Prentice Hall, 2012.

48. Lewis, M., The lac repressor. Comptes Rendus Biologies, 2005. 328(6): p.521-548.

49. Schein, C.H., Production of soluble recombinant proteins in bacteria. Biotechnology, 1989. 7, p.1141–1149.

50. Sorensen, H.P. and K.K. Mortensen, Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. Microb Cell Fact., 2005. 4, p.1.

51. Donovan, R.S., Robinson, C.W., and B. R. Glick, Optimizing the expression of a monoclonal antibody fragment under the transcriptional control of the Escherichia coli lac promoter. Can. J. Microbiol., 2000. 46, p. 532–541.

52. Fernandez-Castane, A., Vine, C.E., Caminal, G., and J. Lopez-Santin, Evidencing the role of lactose permease in IPTG uptake by Escherichia coli in fed-batch high cell density cultures. J. Biotechnol., 2012. 157(3): p. 391-398.

53. Foor, F., Morin, N. and K.A. Bostian, Production of L-dihydroxyphenylalanine in Escherichia coli with the tyrosine phenol-lyase gene cloned from Erwinia herbicola. Appl Environ. Microbiol., 1993. 59: p.3070–3075.

54. Weng, Y.P., et al., Optimization of the overexpression of glutamate mutase S component under the control of T7 system by using lactose and IPTG as the inducers. Enzyme and Microbial Technology, 2006. 38(3-4): p.465-469.

55. Monteiro, R., et al., Use of lactose to induce expression of soluble NifA protein domains of Herbaspirillum seropedicae in Escherichia coli. Can J Microbiol., 2000. 46: p.1087–1090.

56. Pei, X., et al., Efficient production of a thermophilic 2-deoxyribose-5-phosphate aldolase in glucose- limited fed-batch cultivations of Escherichia coli by continuous lactose induction strategy. Appl. Biochem. Biotechnol., 2011. 165: p. 416–425.

57. Mayer, S., et al., Lactose autoinduction with enzymatic glucose release: Characterization of the cultivation system in bioreactor. Protein Expression and Purification, 2014. 94: p. 67–72.

58. Neubauer, P., Hofmann, K., Holst, B., Mattiasson, P., and P. Kruschke, Maximizing the expression of a recombinant gene in Escherichia coli by manipulation of induction time using lactose as inducer. Appl. Microbiol. Biotechnol., 1992. 36: p.739–744.

59. Ukkonen, K., et al., Use of slow glucose feeding as supporting carbon source in lactose autoinduction medium improves the robustness of protein expression at different aeration conditions. Protein Expression and Purification, 2013. 91(2): p. 147-154.

60. Huber, R.E., Lytton, J., and E.B. Fung, Efflux of ß-galactosidase Products from Escherichia coli. Journal of Bacteriology, 1980. 141(2): p.528-533.

61. Stadlmann, V., Harant, H., Korschineck, I., Hermann, M., Forster, F. and A. Missbichler, Novel avian single-chain fragment variable (scFv) targets dietary gluten and related natural grain prolamins, toxic entities of celiac disease. BMC Biotechnology, 2015. 15: p.109.

Published paper

The following manuscript was accepted for publication in Applied Microbiology and Biotechnology on May 7th, 2016.

The *E. coli* pET expression system revisited – mechanistic correlation between glucose and lactose uptake

David Wurm[1], Lukas Veiter[1], Sophia Ulonska[1], Britta Eggenreich[1,2], 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

[§]Corresponding author: Oliver Spadiut, Vienna University of Technology, 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.at

Acknowledgements

The authors acknowledge Sciotec Diagnostic Technologies for providing the used strain.

Abstract

Therapeutic monoclonal antibodies are mainly produced in mammalian cells to date. However, unglycosylated antibody fragments can also be produced in the bacterium *E. coli* which brings several advantages, like growth on cheap media and high productivity. One of the most popular *E. coli* strains for recombinant protein production is *E. coli* BL21(DE3) which is usually used in combination with the pET expression system. However, it is well known that induction by IPTG stresses the cells and can lead to the formation of insoluble inclusion bodies.

In this study, we revisited the pET expression system for the production of a novel antibody single chain variable fragment (scFv) with the goal of maximizing the amount of soluble product. Thus, we 1) investigated whether lactose favors the recombinant production of soluble scFv compared to IPTG, 2) investigated whether the formation of soluble product can be influenced by the specific glucose uptake rate ($q_{s,glu}$) during lactose induction, and 3) determined the mechanistic correlation between the specific lactose uptake rate ($q_{s,lac}$) and $q_{s,glu}$.

We found that lactose induction gave a much greater amount of soluble scFv compared to IPTG, even when the growth rate was increased. Furthermore we showed that the production of soluble protein could be tuned by varying $q_{s,glu}$ during lactose induction. Finally, we established a simple model describing the mechanistic correlation between $q_{s,lac}$ and $q_{s,glu}$ allowing tailored feeding and prevention of sugar accumulation. We believe that this mechanistic model might serve as platform knowledge for *E. coli*.

Keywords

Escherichia coli BL21(DE3), pET expression system, lactose induction, antibody fragment, soluble protein, mechanistic model

Introduction

Antibodies are used to treat a wide variety of human diseases. More than 35 monoclonal antibodies and antibody fragments have been commercialized and around 240 therapeutic monoclonal antibodies and antibody fragments are in clinical trials (1). Since more than 1,000 kg of these therapeutics are needed per year worldwide there is an urge for cheap and fast production (1, 2, 3, 4). Due to the requirement of post-translational modifications most therapeutic monoclonal antibodies and antibody fragments are produced in mammalian cells to date. However, there are many drawbacks such as glycan heterogeneity, low volumetric productivity, long cultivation times, expensive media and the potential risk of virus contamination (1, 6). Thus, the prokaryotic organism *Escherichia coli* has been investigated as alternative host for the production of unglycosylated antibody fragments, mainly single chain variable fragments (scFv), which are also suitable for antigen detection (1, 7, 8). *E. coli* can be cultivated on inexpensive media, has a high growth rate, high cell densities can be achieved, the genetics are very well characterized and an increasingly large number of cloning vectors and mutant host strains is available (e.g. (9, 10). The *E. coli* strain BL21(DE3) and its derivatives are by far the most used *E. coli* strains for recombinant protein production as they exhibit several biotechnological advantages compared to other *E. coli* strains, such as low acetate yield, high biomass yield and reduced expression of proteases (10, 11, 12). Usually the well-known pET expression system is used in combination with *E. coli* BL21(DE3) (13). The lac operon can be induced by allolactose and its molecular mimic Isopropyl β-D-1-thiogalactopyranoside (IPTG) (14). IPTG is a very strong inducer that is not metabolized by *E. coli*, which is why one point addition is sufficient. Thus, IPTG is usually used in industrial production processes with *E. coli* BL21(DE3). However, IPTG is known to put a high metabolic burden on the cells resulting in the formation of inactive aggregates of the recombinant target protein, known as inclusion bodies (IBs). Thus, lactose has been studied as alternative inducer. Lactose was found to be as effective as IPTG, to increase cell fitness, to reduce IB formation and to enhance the formation of soluble recombinant product (14, 15, 16, 17, 18, 19). However, lactose is metabolized by *E. coli* making stable induction more complicated as it has to be continuously supplied (20). In a previous study it was nicely shown that lactose metabolism strongly depends on the available amount of glucose (21). However, a potential mechanistic correlation between glucose and lactose uptake has not been investigated yet.

In this study we used *E. coli* BL21(DE3) and the pET expression system for the production of a novel scFv (22). We hypothesized that induction by lactose increases the amount of soluble product compared to IPTG. Thus, we 1) tested and compared IPTG and lactose as inducer, 2) investigated whether the formation of soluble product can be influenced by the specific uptake rate of glucose during induction with lactose, and 3) determined a mechanistic correlation between the specific uptake rates of lactose and glucose.

Materials and Methods

Strain

E. coli BL21(DE3) (Life technologies, Carlsbad, CA, USA) and the pET28a(+) expression system were used for production of the recombinant scFv which describes an engineered IgY fragment against PT-Gliadin useful for the treatment of celiac disease (22).

Bioreactor Cultivations

Media

A defined minimal medium according to DeLisa (23) supplemented with 0.02 g/L Kanamycin and different amounts of glucose and lactose (Table 1) were used for all cultivations.

Preculture

500 mL sterile DeLisa pre-culture medium were inoculated from frozen stocks (1.5 mL, -80°C) and incubated in a 2,500 mL High-Yield shake flask in a Infors HR Multitron shaker (Infors, Bottmingen, Switzerland) at 37°C and 230 rpm for 20 h. Then 4,500 mL DeLisa-batch medium in the bioreactor were inoculated with 500 mL of pre-culture.

Batch and Fed-batch cultivations

Batch and Fed-batch cultivations were done in a stainless steel Satorius Biostat Cplus bioreactor (Satorius, Göttingen, Germany) with a working volume of 10 liters. The bioreactor was aerated with a mixture of pressurized air and pure oxygen at 1.5 vvm and agitated constantly at 1,000 rpm. Dissolved oxygen (dO_2) was monitored with a fluorescence dissolved oxygen electrode Visiferm DO425 (Hamilton, Reno, NV, USA) and kept above 40 % throughout all cultivations by varying the ratio of pressurized air to pure oxygen. pH was monitored with an Easyferm electrode (Hamilton, Reno, NV, USA) and maintained constant at pH 7.2 by addition of NH_4OH (12.5 %). Base consumption was determined gravimetrically. CO_2 and O_2 concentrations in the off-gas were monitored by a DASGIP GA gas analyzer (Eppendorf, Hamburg, Germany). All process parameters were adjusted and logged by the process information management system Lucullus (Biospectra, Schlieren, Switzerland).

The batch phase was conducted at 35°C and yielded a biomass concentration of 8-9 g dry cell weight (DCW) per liter. After depletion of glucose, visible by a drop in the CO_2 off-gas signal, a fed-batch to generate a biomass was conducted. We fed at a constant specific glucose uptake rate ($q_{s,glu}$) of 0.2 g/g/h. When the final DCW reached 25 g/L the culture was induced by IPTG or lactose, respectively. During non-induced fed-batch and induction with IPTG, DCW in the bioreactor was estimated using a soft-sensor tool (24). During induction with lactose, DCW was calculated assuming a constant biomass yield ($Y_{X/S}$ = 0.37 g/g, own unpublished data). The feed rate was adjusted to maintain a constant $q_{s,glu}$ and was calculated according to Equation 1.

(1) $F = q_s * X * \frac{V}{W}$

F… feedrate [g/h]
q_s...specific substrate uptake rate [g/g/h]
X… DCW concentration [g/L]
V... reactor Volume [L]
W… amount of substrate per feed [g/g]

Cultivation strategy

In this study cultivations following a standard procedure comprising 3 phases (batch, non-induced fed-batch, induced fed-batch) as well as also dynamic experiments (pulses and shifts) were carried out. Applying dynamic process conditions to accelerate strain characterization and bioprocess development is a common approach in our working group (25, 26, 27, 28, 29). An overview of the different cultivations and their respective goals is shown in Supplementary Table S1. Induction was either performed by 0.5 mM IPTG or by lactose which was applied either as pulses or by continuous feeding. In these cultivations the lactose concentration in the media was kept in excess between 5 and 15 g/L.

Sampling

Samples were taken at the beginning and the end of the batch and the non-induced fed-batch. During induction sampling was performed at the beginning and the end of each shift/pulse and every hour during fed-batch cultivations. Specific product formation rates and final product yields are given for an induction phase of approximately 4 h for all cultivations. DCW was determined by centrifugation (4,500 g, 4°C, 10 min) of 5 mL cultivation broth, washing the obtained cell pellet with a 0.1 % NaCl solution and subsequent drying at 105°C for 48 h. Optical density at 600 nm (OD_{600}) was measured in the photometer Genesys 20 (Thermo Scientific, Waltham, MA, USA). Samples were diluted with deionized water to stay within the linear range of the photometer (OD_{600} 0.1-0.8). A linear correlation between OD_{600} and DCW was established to verify and, if necessary, correct the DCW estimation of the soft-sensor (Equation 2).

(2) $DCW = OD_{600} \cdot 0.445$

DCW... biomass dry cell weight [g/L]
OD_{600}… optical density at 600 nm

Substrate and metabolite quantification

Cell-free samples of the cultivation broth were analyzed for concentrations of substrates and metabolites by HPLC (Agilent Technologies, Santa Clara, CA, USA) with a Supelcogel C-610 H ion exchange column (Sigma-Aldrich, St. Louis, MO, USA) and a refractive index detector (Agilent Technologies, Santa Clara, CA, USA). The mobile phase was 0.1 % H_3PO_4 with a constant flow rate of 0.5 mL/min and the system was run isocratically at 30°C.

Product quantification

Cells were harvested (4,500 g, 4°C, 10 min), resuspended and diluted in Tris buffer (100 mM, 10 mM EDTA, pH 7.4) to a DCW concentration of 5 g/L and subsequently homogenized in an EmusiflexC3 Homgeniziser (Avestin, Ottowa, ON, USA) at 1,500 bar for 5 passages. After centrifugation (14,000 g, 4°C, 10 min) soluble protein (SP) was recovered in the supernatant and IBs in the pellet.

Cell debris from 0.5 mg DCW was resuspended in 1x Laemmli buffer and the supernatant was diluted with 2x Laemmli buffer before the samples were heated at 95°C for 10 minutes. Ten µL of each sample were loaded onto pre-cast SDS gels (8-16%) (GE Healthcare, Little Chalfont, United Kingdom). Gels were run in an Amersham ECL Gel Box, a horizontal mini-gel system (GE Healthcare, Little Chalfont, United Kingdom) for 90 min at 140V and stained with Coomassie Blue. On every gel three BSA standards (0.5 µg, 1.5 µg 3 µg per lane) were applied. The protein bands were evaluated densitometrically using the software Image Lab (Bio-Rad, Hercules, CA, USA). Calibrations always gave an R^2 above 0.96 and allowed the quantification of SP and IBs. This method has been used in over 700 publications to date and is known to give precise data (e.g. (30, 31, 32, 33).

Results

In this study we revisited the pET expression system in E. coli BL21(DE3) for the production of a novel scFv. We wanted to 1) prove that lactose favors the recombinant production of soluble scFv compared to IPTG, 2) investigate if the formation of soluble product can be influenced by $q_{s,glu}$ during lactose induction, and 3) determine a mechanistic correlation between $q_{s,lac}$ and $q_{s,glu}$.

Induction by IPTG vs. lactose

We tested and compared IPTG and lactose as inducers for the pET expression system. Slow uptake of glucose and thus a low specific growth rate (µ) was described to result in higher amount of SP (34). Thus, we wanted to investigate whether this strategy was also suitable for the expression of the novel scFv and conducted three fed-batch cultivations with different $q_{s,glu}$ during induction with 0.5 mM IPTG (Fig. 1a, Table 2).

At $q_{s,glu}$ of 0.45 g/g/h and 0.35 g/g/h only IBs were formed during IPTG induction. IB formation decreased when feeding at lower $q_{s,glu}$. At $q_{s,glu}$ of 0.14 g/g/h soluble scFv was produced. Apparently the amount of SP and IBs was strongly linked to $q_{s,glu}$ and thus µ. Furthermore, by decreasing the feeding rate from $q_{s,glu}$ 0.45 g/g/h to 0.14 g/g/h the acetate formation rate was reduced more than 10-fold from 3.11 mg/g/h to 0.24 mg/g/h. It is known that the formation of acetate has a negative impact on cell growth and recombinant protein production and should thus be kept at a minimum to ensure high product quality (35). Our findings that a low µ leads to higher production of SP as well as lower acetate formation are in good agreement with literature (36; 37; 38). Nevertheless more than half of the scFv was still found as insoluble and inactive IBs. A further decrease of $q_{s,glu}$ to potentially increase the amount of SP was not feasible as it would have led to slow growth and thus very long process times. Therefore the alternative inducer lactose which is described to enhance cell

fitness and to increase the amount of SP (39) was tested (Fig. 1b, Table 2). Even though $q_{s,glu}$ was 0.30 g/g/h during lactose induction and thus more than 2-fold higher than in the only experiment with IPTG which gave SP ($q_{s,glu}$ = 0.14 g/g/h) around 30-fold more soluble scFv was produced. Furthermore the acetate concentration was below the detection limit when lactose was used as inducer (Table 2).

Tunability of recombinant protein production by varying $q_{s,glu}$ during lactose induction

Furthermore we investigated the potential tunability of recombinant protein production by testing three different $q_{s,glu}$ during induction with lactose, which was always present in excess (Fig. 2).

Fig. 2 shows that both the formation of SP and IBs strongly depended on $q_{s,glu}$ during lactose induction. There was a clear optimum at $q_{s,glu}$ = 0.30 g/g/h. At $q_{s,gluc}$ of 0.06 g/g/h 14-fold less SP was produced and also IB formation was reduced 3-fold. We speculate that at this low $q_{s,glu}$ the cells basically only had enough energy for maintenance metabolism but not for recombinant protein production. However, we were rather surprised to see reduced production of both SP and IB also at the higher $q_{s,glu}$ of 0.74 g/g/h. Thus, we analyzed $q_{s,lac}$ at the respective $q_{s,glu}$. As shown in Table 3 $q_{s,lac}$ was strongly dependent on $q_{s,glu}$. At $q_{s,glu}$ of 0.74 g/g/h only 0.02 g/g/h lactose were metabolized. We speculate that this amount of inducer was too low to guarantee full induction, thus resulting in reduced productivity.

Mechanistic correlation between $q_{s,lac}$ and $q_{s,glu}$

Motivated by our observation of an apparent mechanistic correlation between $q_{s,lac}$ and $q_{s,glu}$ (Table 3), we performed several cultivations to shed more light on this physiological correlation. First we performed a batch cultivation with excess of both glucose and lactose. In this experiment we observed the well-known phenomenon of carbon catabolite repression (40) meaning that as long as glucose was present in excess no lactose was taken up (Supplementary Fig. S1). Only when glucose was depleted, lactose was metabolized at a very slow rate of 0.05 g/g/h, which is in agreement to literature (21).

On the contrary at limiting amounts of glucose but an excess of lactose, we observed a much higher $q_{s,lac}$ (Table 3). Several subsequent cultivations revealed that $q_{s,lac}$ was in fact a function of $q_{s,glu}$ (Fig. 3, Table 4).

As shown in Fig. 3, $q_{s,lac}$ increased with increasing $q_{s,glu}$ having a maximum at $q_{s,glu}$ of 0.2 - 0.25 g/g/h, before it decreased again. We used these findings to generate a simple mechanistic model describing $q_{s,lac}$ as a function of $q_{s,glu}$. Such a mechanistic model would greatly facilitate bioprocess development as only a few experiments would be required to determine the correlation between $q_{s,lac}$ and $q_{s,glu}$. Thus, lactose accumulation and resulting osmotic stress can be reduced (41) and tailored feeding and induction are possible. We hypothesized that the correlation between $q_{s,lac}$ and $q_{s,glu}$ was described by two phenomena, namely:

1) $q_{s,lac}$ depended Monod-like on $q_{s,glu}$ until a certain maximum was reached, before 2) $q_{s,lac}$ decreased at high $q_{s,glu}$ which was treated similarly to the phenomenon of substrate

inhibition (42). To describe this correlation we adapted the model proposed by Han and Levenspiel (43) (Equation 3).

$$(3)\ q_{s,lac} = q_{s,lac,max} \cdot \max\left(\left(1 - \frac{q_{s,glu}}{q_{s,glu,crit}}\right)^{n}, 0\right) \cdot \left(\frac{q_{s,glu}}{q_{s,glu} + K_A\left(1 - \frac{q_{s,glu}}{q_{s,glu,crit}}\right)^{m}} + \frac{q_{s,lac,noglu}}{q_{s,lac,max}}\right)$$

$q_{s,lac}$	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)

The unknown parameters in this model are the maximum specific lactose uptake rate ($q_{s,lac,max}$), the critical specific glucose uptake rate up to which lactose is consumed ($q_{s,glu,crit}$), the affinity constant (K_A), the two constants n and m indicating the type of inhibition (noncompetitive, uncompetitive, competitive; (43)) and the specific lactose uptake rate at zero glucose uptake ($q_{s,lac,noglu}$). For the parameter identification the Nelder-Mead simplex method in MATLAB (44) was used to minimize the objective function (Equation 4) which describes the distance between the experimental data and the predicted values by the model. For points, where no standard deviation was available, the mean of all standard deviations was taken.

$$(4)\ S = \sum_{i=1}^{n}\left(\frac{q_{s,lac,max,meas,i} - q_{s,lac,max,model,i}}{\sigma_i}\right)^2$$

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

Based on our observations we defined that $q_{s,lac}$ is greater than zero when no glucose is consumed. Furthermore, parameter values must have a mechanistic meaning, which is why we assumed $q_{s,lac}$, K_A and n to be positive and even constrained them further (e.g. $q_{s,lac} < q_{s,glu,crit}$, $K_A < 1$, $n > 0$, etc.). To analyze the model and to evaluate the impact of the model parameters we performed a sensitivity analysis, where we increased or decreased the parameters by 20 % (an example is shown in Supplementary Fig. S2).

The sensitivity analysis revealed the parameter m to have almost no impact on the curve. Furthermore parameter estimation revealed that this parameter is almost zero ($6.2*10^{-9}$). This is in accordance with our assumption of a noncompetitive inhibition which is described by m=0 (43). Thus, we set m to zero and simplified the model (Equation 5).

$$(5)\ q_{s,lac} = q_{s,lac,max} \cdot \max\left(\left(1 - \frac{q_{s,glu}}{q_{s,glu,crit}}\right)^{n}, 0\right) \cdot \left(\frac{q_{s,glu}}{q_{s,glu} + K_A} + \frac{q_{s,lac,noglu}}{q_{s,lac,max}}\right)$$

Additionally the sensitivity analysis showed that n, $q_{s,lac,max}$, K_A, $q_{s,lac,noglu}$, and $q_{s,glu,crit}$ had a significant impact on the curve. However, the impact of n was rather small and even a potential error of 20 % would lead to only small deviations of the curve (Supplementary Fig. S2). We used our experimental data for the recombinant E. coli strain producing the novel scFv and fitted the mechanistic model to the data using Equation 5. The curve with the estimated parameters fitted the data with a normalized root mean square error (NRMSE) of 10.3 % and a coefficient of variation (CV) of 18.6 %. Furthermore all parameters showed physiologically reasonable values. The resulting curve and the corresponding parameter values are shown in Fig. 4 and Table 5, respectively.

Discussion

In this study we used a pET expression system and an E. coli BL21(DE3) strain for the production of a novel antibody single chain variable fragment. Since we wanted to maximize the amount of soluble product and reduce the formation of IBs we 1) analyzed whether lactose favors the recombinant production of soluble scFv compared to IPTG, 2) investigated whether the formation of soluble product can be influenced by $q_{s,glu}$ during lactose induction, and 3) determined the mechanistic correlation between $q_{s,lac}$ and $q_{s,glu}$.

In fact, we showed that lactose allowed a much higher production of soluble product compared to IPTG even when μ was increased. This outcome is in good agreement with several recent studies that also showed the benefits of using lactose as inducer compared to IPTG for other recombinant products (15, 16, 19), but also for antibody fragments (39, 45). However, data on potential tunability of recombinant protein production by changing $q_{s,glu}$ and thus μ during lactose induction are scarce to date.

Thus, we investigated a potential tunability of recombinant protein production and found that the formation of soluble product could be tuned by varying $q_{s,glu}$ during lactose induction. Furthermore, we observed a mechanistic correlation between $q_{s,glu}$ and $q_{s,lac}$ in these experiments which motivated us to analyze this phenomenon in more detail.

In fact we determined a mechanistic correlation between $q_{s,lac}$ and $q_{s,glu}$ and established a simple model. The shape of this curve can be explained by different phenomena. At high $q_{s,glu}$, no lactose is taken up which can be explained by the phenomenon called carbon catabolite repression. When $q_{s,glu}$ decreases cAMP gets formed inside the cell. cAMP binds to the catabolite activator protein (CAP) which undergoes a change in conformation and binds to the promoter region of the lac operon. Consequently transcription of the lac operon's genes involving lactose permease and β- galactosidase is initiated and lactose can be taken up. However, in the absence of glucose there is hardly any lactose taken up. The transport of lactose into the cell is ATP related (46). Without glucose metabolism, basically no ATP is generated, which is why this transport can barely happen (21, 47, 48, 49).

The outcomes of this study can be used for strain characterization and fast bioprocess development. The values for the parameters $q_{s,lac,noglu}$, and $q_{s,glu,crit}$, which have a high impact on the curve, can be easily determined by simple batch cultivations. To determine the other parameters ($q_{s,lac,max}$, K_A and n) we recommend performing two to three experiments where lactose is provided in excess and $q_{s,glu}$ is adjusted to values below $q_{s,glu,crit}$. In the optimal case

one of these $q_{s,glu}$ values corresponds to the maximum of $q_{s,lac}$. However, if this is not the case, $q_{s,lac,max}$ is simply underestimated, leaving no risk for lactose accumulation. By performing these few experiments enough data can be gathered to establish the mechanistic model. The mechanistic model can then be used to interpolate unknown $q_{s,lac}$ values to $q_{s,glu}$ points and thus allows fed-batch fermentations at different $q_{s,glu}$ preventing unwanted lactose accumulation. Furthermore, $q_{s,lac}$ can be adjusted within the feasible range allowing to tune the production of soluble protein, as we showed for the novel scFv (Table 3). By this approach the production of soluble and active protein can be significantly increased.

Summarizing this work emphasizes the applicability of lactose, a cheap, nontoxic waste product, as inducer for the production of soluble recombinant proteins in E. coli. In future studies we will investigate if the mechanistic model established in this study describes platform knowledge applicable to different E. coli strains. We suggest that by performing two batch experiments and two to three fed-batches at different $q_{s,glu}$ and concomitant lactose excess, enough data are available to fit the model and obtain the mechanistic correlation for basically any E. coli strain. Furthermore we will test if induction by lactose triggers expression on a cellular level or if E. coli subpopulations are generated.

Ethical approval

The authors declare that they have no conflict of interest. This article does not contain any studies with human participants or animals performed by any of the authors.

Author Contributions

DJW, LV, CH and OS designed the study. DJW, LV and BE conducted the experiments and analyzed the data. DJW and SU developed the model. DJW and OS wrote the paper.

Figures

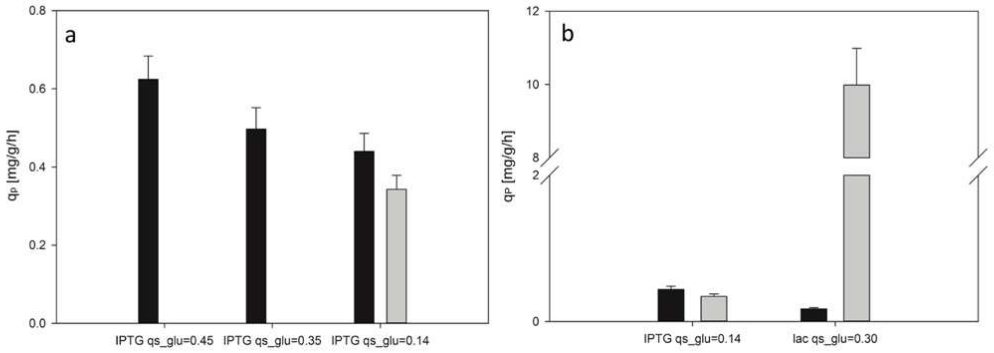


Fig. 1. a,b.: Specific product formation rate (q_P) at different $q_{s,glu}$ during IPTG induction. Black bars, specific production rate of inclusion bodies ($q_{P,IB}$); grey bars, specific production rate of soluble product ($q_{P,SP}$) b, Specific product formation rate (q_P) at different $q_{s,glu}$ during IPTG or lactose induction. Black bars, specific production rate of inclusion bodies ($q_{P,IB}$); grey bars, specific production rate of soluble product ($q_{P,SP}$)

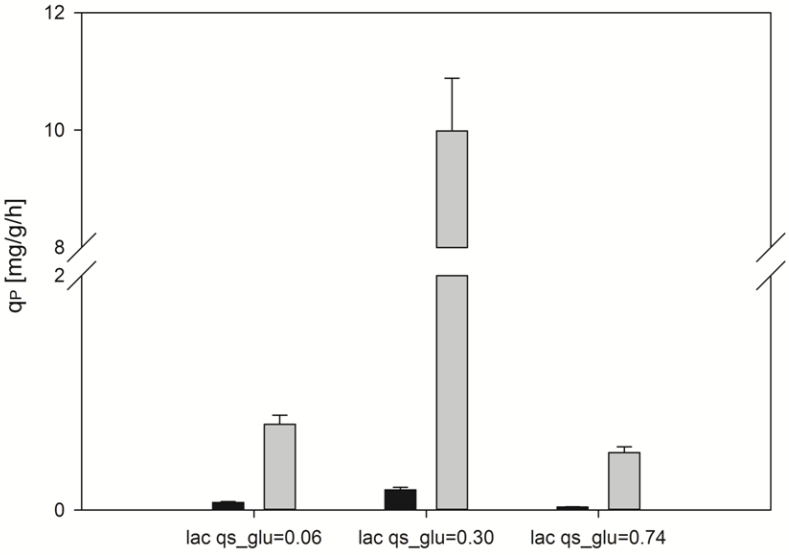


Fig. 2.: Specific product formation rate (q_P) at different $q_{s,glu}$ during lactose induction. Black bars, specific production rate of inclusion bodies ($q_{P,IB}$); grey bars, specific production rate of soluble product ($q_{P,SP}$)

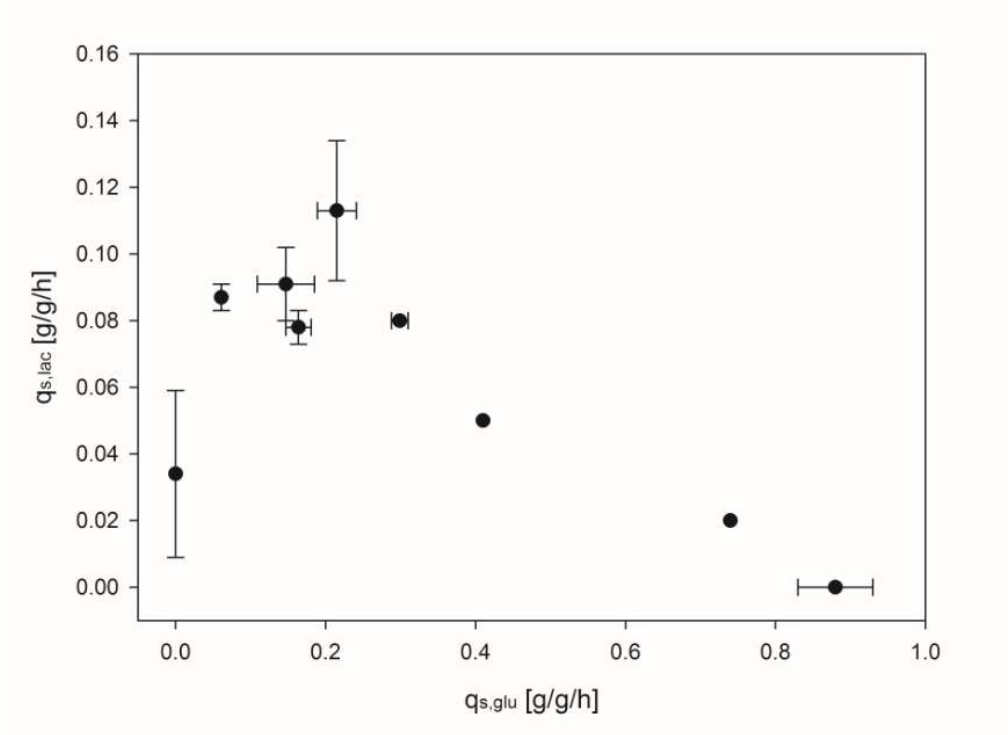


Fig. 3.: $q_{s,lac}$ as a function of $q_{s,glu}$ for the recombinant *E. coli* strain producing scFv with the pET expression system

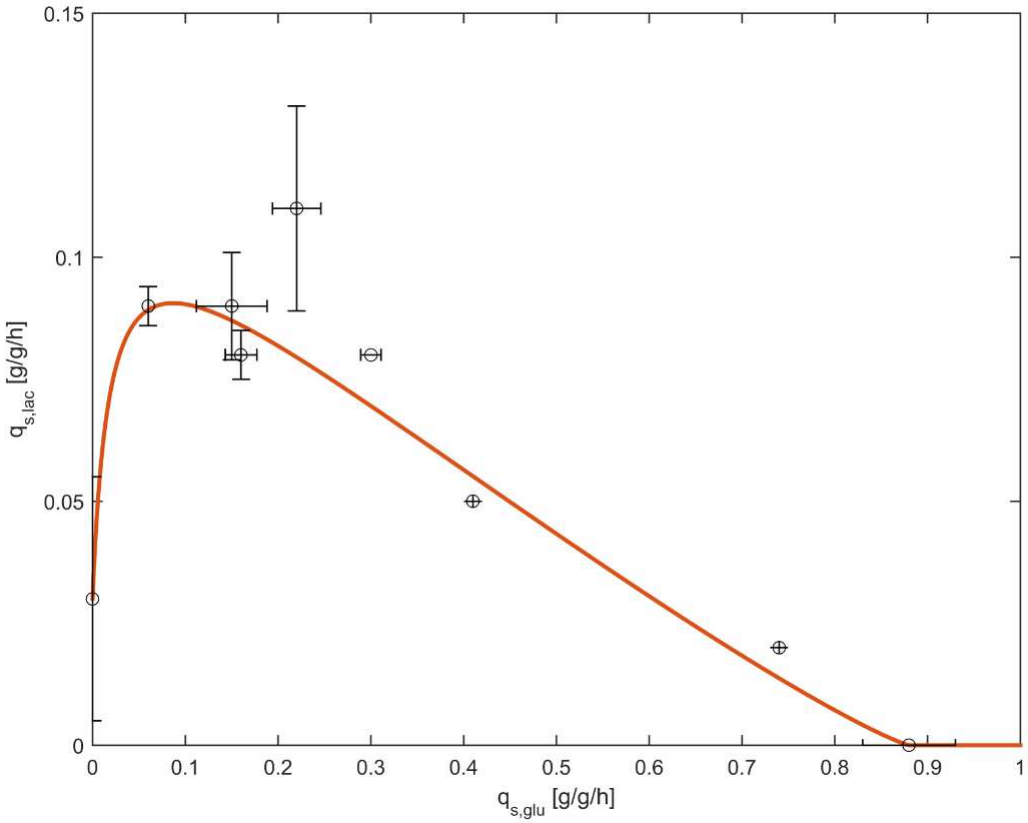


Fig. 4.: Optimal fit with parameters $q_{s,lac,max}$= 0.088 g/g/h, $q_{s,glu,crit}$ = 0.88 g/g/h, $q_{s,lac,noglu}$ = 0.034 g/g/h, K_A= 0.019 g/g/h, n=1.16

Tables and Table Captions

Table 1.: Sugar concentrations in different DeLisa media

Component	pre-culture	batch	feed glucose	feed lactose
$C_6H_{12}O_6 \cdot H_2O$ [g/L]	8.8	22.0	275	-
$C_{12}H_{22}O_{11}$ [g/L]	-	-	-	210

Table 2.: Strain physiological parameters of *E. coli* BL21(DE3) producing a scFv via a pET expression system during either IPTG or lactose induction

inducer	$q_{s,glu}$ [g/g/h]	μ [1/h]	$q_{P,IB}$ [mg/g/h]	Std. dev [mg/g/h]	$q_{P,SP}$ [mg/g/h]	Std. dev [mg/g/h]	$q_{P,acetate}$ [mg/g/h]	Std. dev [mg/g/h]	IB-titer [mg/g]	SP-titer [mg/g]
IPTG	0.45	0.16	0.62	0.059	0.00		3.11	0.006	2.50	0.00
IPTG	0.35	0.08	0.50	0.055	0.00		2.65	0.005	1.99	0.00
IPTG	0.14	0.03	0.44	0.046	0.34	0.036	0.24	0.001	1.76	1.37
lactose	0.30	0.10	0.17	0.021	9.99	0.899	0.00		0.69	39.55

Table 3.: Strain physiological parameters of *E. coli* BL21(DE3) producing a scFv via a pET expression system during lactose induction at different $q_{s,glu}$

inducer	$q_{s,glu}$ [g/g/h]	$q_{s,lac}$ [g/g/h]	$q_{P,IB}$ [mg/g/h]	Std. dev [mg/g/h]	$q_{P,SP}$ [mg/g/h]	Std. dev. [mg/g/h]	$q_{P,acetate}$ [mg/g/h]	IB-titer [mg/g]	SP-titer [mg/g]
lactose	0.06	0.09	0.06	0.007	0.73	0.077	0.00	0.26	2.92
	0.30	0.08	0.17	0.021	9.99	0.899	0.00	0.69	39.55
	0.74	0.02	0.03	0.003	0.49	0.049	0.00	0.12	1.96

Table 4.: Experimentally evaluated $q_{s,lac}$ at respective $q_{s,glu}$ for the recombinant *E. coli* strain producing scFv with the pET expression system

$q_{s,glu}$ [g/g/h]	Std. dev [g/g/h]	$q_{s,lac}$ [g/g/h]	Std. dev [g/g/h]
0.00	-	0.03	0.025
0.06	-	0.09	0.004
0.15	0.038	0.09	0.011
0.16	0.017	0.08	0.005
0.22	0.026	0.11	0.021
0.30	0.011	0.08	-
0.41	-	0.05	-
0.74	-	0.02	-
0.88	0.050	0.00	-

Table 5.: Parameters of optimal fit for the mechanistic correlation between $q_{s,lac}$ and $q_{s,glu}$ for the recombinant *E. coli* strain producing scFv with the pET expression system

$q_{s,lac,max}$ [g/g/h]	$q_{s,glu,crit}$ [g/g/h]	$q_{s,lac,noglu}$ [g/g/h]	K_A [g/g/h]	n [-]
0.088	0.88	0.034	0.019	1.16

References Paper

(1) Lee, Y. J., and K.J. Jeong, Challenges to production of antibodies in bacteria and yeast. Journal of Bioscience and Bioengeering, 2015. 120: p. 483-490. doi:10.1016/j.jbiosc.2015.03.009

(2) Elvin, J. G., Couston, R. G., and C.F. van der Walle, Therapeutic antibodies: market considerations, disease targets and bioprocessing. International Journal of Pharmaceutics, 2013. 440: p. 83-98. doi:10.1016/j.ijpharm.2011.12.039

(3) Rodrigues, M. E., Costa, A.R., Henriques, M., Azeredo, J., and R. Oliveira, Technological progresses in monoclonal antibody production systems. Biotechnology Progress, 2010. 26: p.332-351. doi:10.1002/btpr.348

(4) Walsh, G. Biopharmaceutical benchmarks. Nature Biotechnology, 2014. 32: p.992-1000. doi:10.1038/nbt.3040

(5) Liu, J. K., The history of monoclonal antibody development - Progress, remaining challenges and future innovations. Annals of Medicine and Surgery (London), 2014. 3: p.113-116. doi:10.1016/j.amsu.2014.09.001

(6) Khan, K.H., Gene expression in Mammalian cells and its applications. Advanced Pharmaceutical Bulletin, 2013. 3: p.257-263. doi:10.5681/apb.2013.042

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

(8) Wals, K., and H. Ovaa, Unnatural amino acid incorporation in *E. coli*: current and future applications in the design of therapeutic proteins. Frontiers in Chemistry, 2014. 2: p.15 doi:10.3389/fchem.2014.00015

(9) Baeshen, M. N., et al., Production of Biopharmaceuticals in *E. coli*: Current Scenario and Future Perspectives. Journal of Microbiology and Biotechnology, 2015. 25: p.953-962. doi:10.4014/jmb.1412.12079

(10) Rosano, G. L., and E. A. Ceccarelli, Recombinant protein expression in *Escherichia coli*: advances and challenges. Frontiers on Microbiology, 2014. 5: p.172. doi:10.3389/fmicb.2014.00172

(11) Choi, J. H., Keum, K. C., and S. Y. Lee, Production of recombinant proteins by high cell density culture of *Escherichia coli.* Chemical Engineering Science, 2006. 61: p. 876-885. doi:10.1016/j.ces.2005.03.031

(12) Ferrer-Miralles, N., Saccardo, P., Corchero, J. L., Xu, Z., and E. Garcia-Fruitos, General introduction: recombinant protein production and purification of insoluble proteins. Methods in Molecular Biology, 2015. 1258: p. 1-24. doi:10.1007/978-1-4939-2205-5

(13) Studier, F. W., and B. A. Moffatt, Use of Bacteriophage-T7 Rna-Polymerase to Direct Selective High-Level Expression of Cloned Genes. Journal of Molecular Biology, 1986. 189: p. 113-130. doi:10.1016/0022-2836(86)90385-2

(14) Neubauer, P., Hofmann, K., Holst, O., Mattiasson, B., and P. Kruschke, Maximizing the Expression of a Recombinant Gene in *Escherichia coli* by Manipulation of Induction Time Using Lactose as Inducer. Applied Microbiology and Biotechnology, 1992. 36: p. 739-744.

(15) Bashir, H., et al., Simple procedure applying lactose induction and one-step purification for high-yield production of rhCIFN. Biotechnology and Applied Biochemistry, 2015. doi:10.1002/bab.1426

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

(17) Gombert, A. K., and B. V. Kilikian, Recombinant gene expression in *Escherichia coli* cultivation using lactose as inducer. Journal of Biotechnology, 1998. 60: p. 47-54. doi:10.1016/S0168-1656(97)00185-5

(18) Pei, X.L., et al., Efficient production of a thermophilic 2-deoxyribose-5-phosphate aldolase in glucose-limited fed-batch cultivations of *Escherichia coli* by continuous lactose induction strategy. Applied Biochemistry and Biotechnology, 2011. 165: p. 416-425. doi:10.1007/s12010-011-9261-8

(19) Zou, C., Duan, X., and J. Wu, Enhanced extracellular production of recombinant *Bacillus deramificans* pullulanase in *Escherichia coli* through induction mode optimization and a glycine feeding strategy. Bioresources Technology, 2014. 172: p. 174-179.

(20) Striedner, G., Cserjan-Puschmann, M., Potschacher, F., and K. Bayer, Tuning the transcription rate of recombinant protein in strong *Escherichia coli* expression systems through repressor titration Biotechnology Progress, 2003. 19: p. 1427-1432. doi:10.1021/bp034050u

(21) Kremling, A., Geiselmann, J., Ropers, D., and H. de Jong, Understanding carbon catabolite repression in *Escherichia coli* using quantitative models. Trends in Microbiology, 2015. 23: p. 99-109. doi:10.1016/j.tim.2014.11.002

(22) Stadlmann, V., Harant, H., Korschineck, I., Hermann, M., Forster, F., and A. Missbichler, Novel avian single-chain fragment variable (scFv) targets dietary gluten and related natural grain prolamins, toxic entities of celiac disease. BMC Biotechnology, 2015. 15: p. 109. doi:10.1186/s12896-015-0223-z

(23) DeLisa, M.P., et al., Monitoring GFP-operon fusion protein expression during high cell density cultivation of *Escherichia coli* using an on-line optical sensor. Biotechnology and Bioengineering, 1999. 65: p. 54-64. doi:10.1002/(Sici)1097-0290

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

(25) Dietzsch, C., Spadiut, O., and C. Herwig C, A dynamic method based on the specific substrate uptake rate to set up a feeding strategy for *Pichia pastoris.* Microbial Cell Factories, 2011. 10: p. 14. doi:10.1186/1475-2859-10-14

(26) Dietzsch, C., Spadiut, O., and C. Herwig, A fast approach to determine a fed batch feeding profile for recombinant *Pichia pastoris* strains. Microbial Cell Factories, 2011. 10: p. 85. doi:10.1186/1475-2859-10-85

(27) Jazini, M., and C. Herwig, Effect of post-induction substrate oscillation on recombinant alkaline phosphatase production expressed in *Escherichia coli.* Journal of Bioscience and Bioengineering, 2011. 112: p. 606-610. doi:10.1016/j.jbiosc.2011.08.013

(28) Spadiut, O., Rittmann, S., Dietzsch, C., and C. Herwig, Dynamic process conditions in bioprocess development. Engineering in Life Science, 2013. 13: p. 88-101. doi:10.1002/elsc.201200026

(29) Zalai, D., Dietzsch, C., Herwig, C., and O. Spadiut, A dynamic fed batch strategy for a *Pichia pastoris* mixed feed system to increase process understanding. Biotechnology Progress, 2012. 28: p. 878-886. doi:10.1002/btpr.1551

(30) Matusica, D., et al., Inhibition of motor neuron death in vitro and in vivo by a p75 neurotrophin receptor intracellular domain fragment. Journal of Cell Science, 2016. 129: p. 517-530. doi:10.1242/jcs.173864

(31) Wahlang, B., et al., Polychlorinated Biphenyl-Xenobiotic Nuclear Receptor Interactions Regulate Energy Metabolism, Behavior, and Inflammation in Non-alcoholic-Steatohepatitis. Toxicology Science, 2016. 149: p. 396-410. doi:10.1093/toxsci/kfv250

(32) Wahlang, B., et al., Polychlorinated Biphenyl-Xenobiotic Nuclear Receptor Interactions Regulate Energy Metabolism, Behavior, and Inflammation in Non-alcoholic-Steatohepatitis. Toxicology Science, 2016. 149: p. 396-410. doi:10.1093/toxsci/kfv250

(33) Xu, B., et al., 2,3',4,4',5-Pentachlorobiphenyl Induces Inflammatory Responses in the Thyroid Through JNK and Aryl Hydrocarbon Receptor-Mediated Pathway. Toxicology Science, 2016. 149: p. 300-311. doi:10.1093/toxsci/kfv235

(34) Shin, C. S., Hong, M. S., Kim, D. Y., Shin, H. C., and J. Lee, Growth-associated synthesis of recombinant human glucagon and human growth hormone in high-cell-density cultures of *Escherichia coli*. Applied Microbiology and Biotechnology, 1998. 49: p. 364-370.

(35) De Mey, M., De Maeseneire, S., Soetaert, W., and E. Vandamme, Minimizing acetate formation in *E. coli* fermentations. Journal of Industrial Microbiology and Biotechnology, 2007. 34: p. 689-700. doi:10.1007/s10295-007-0244-2

(36) Hellmuth, K., Korz, D. J., Sanders, E. A., and W. D. Deckwer, Effect of growth rate on stability and gene expression of recombinant plasmids during continuous and high cell density cultivation of *Escherichia coli* TG1. Journal of Biotechnology, 1994 32: p. 289-298.

(37) Sanden, A. M., Bostrom, M., Markland, K., and G. Larsson, Solubility and proteolysis of the Zb-MalE and Zb-MalE31 proteins during overproduction in *Escherichia coli.* Biotechnology and Bioengineering, 2005. 90: p.239-247. doi:10.1002/bit.20433

(38) Shin, C. S., Hong, M. S., Kim, D. Y., Shin, H. C., and J. Lee, Growth-associated synthesis of recombinant human glucagon and human growth hormone in high-cell-density cultures of *Escherichia coli*. Applied Microbiology and Biotechnology, 1998. 49: p. 364-370.

(39) Donovan, R. S., Robinson, C., and B. Glick, Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of thelac promoter. Journal of Industrial Microbiology, 1996. 16: p.145-154.

(40) Stulke, J., and W. Hillen, Carbon catabolite repression in bacteria. Current Opinion in Microbiology, 1999. 2: p.195-201. doi:10.1016/S1369-5274(99)80034-4

(41) Shiloach, J., and R. Fass, Growing *E. coli* to high cell density - a historical perspective on method development. Biotechnology Advances, 2005. 23: p. 345-357. doi:10.1016/j.biotechadv.2005.04.004

(42) Sivakumar, A., Srinivasaraghavan, T., Swaminathan, T., and A. Baradarajan, Extended Monod Kinetics for Substrate Inhibited Systems. Bioprocess Engineering, 1994. 11: p.185-188. doi:10.1007/Bf00369628

(43) Han, K., and O. Levenspiel O, Extended monod kinetics for substrate, product, and cell inhibition. Biotechnology and Bioengineering, 1988. 32: p. 430-447. doi:10.1002/bit.260320404

(44) Lagarias, J. C., Reeds, J. A., Wright, M. H., and P. E. Wright, Convergence properties of the Nelder-Mead simplex method in low dimensions SIAM Journal on optimization, 1998. 9: p. 112-147.

(45) Sanchez-Arreola, P. B., et al., A baseline process for the production, recovery, and purification of bacterial influenza vaccine candidates. Biotechnology Progress, 2013. 29: p. 896-908. doi:10.1002/btpr.1749

(46) Voet. D., Voet, J. G., Pratt C. W., Beck-Sickinger, A., and U. Hahn, Lehrbuch der Biochemie. Wiley-VCH, Weinheim, 2010.

(47) Luo, Y., Zhang, T., and H. Wu, The transport and mediation mechanisms of the common sugars in *Escherichia coli*. Biotechnology Advances, 2014. 32: p. 905-919. doi:10.1016/j.biotechadv.2014.04.009

(48) Mayer, S., et al., Lactose autoinduction with enzymatic glucose release: characterization of the cultivation system in bioreactor. Protein Expression and Purification, 2014. 94: p. 67-72. doi:10.1016/j.pep.2013.10.024

(49) Siegal, M. L., Shifting sugars and shifting paradigms. PLoS Biology, 2015. 13: e1002068. doi:10.1371/journal.pbio.1002068

Conclusions

In this Thesis *E. coli* BL21(DE3) and the pET expression system were used for the production of a novel single chain fragment variable (scFv) with the emphasis on increasing the amount of soluble product via lactose induction as opposed to conventional induction via isopropylthiogalactoside (IPTG). In addition, the concomitant uptake of glucose and lactose as well as its effect on productivity and solubility was studied.

This task led to the aforementioned scientific questions which we were able to answer as follows:

1) Does lactose favour recombinant production of soluble scFv when compared to IPTG?

As previously hypothesized, lactose does significantly favour the recombinant production of soluble scFv when compared to IPTG. With IPTG induction soluble protein could only be detected at a $q_{s,glu}$ of 0.14 g/g/h (which amounts to a μ of 0.05 h^{-1}) resulting in a $q_{P,SP}$ of 0.34 mg/g/h. When using lactose we determined a maximum $q_{P,SP}$ of 9.99 mg/g/h which signifies an increase up to 29 fold, these results could be obtained at a higher $q_{s,glu}$ of 0.30 g/g/h. At the same time lactose induction also led to a decrease in IB production.

2) Can the formation of soluble product be influenced by the specific uptake rate of glucose during induction with lactose?

We found that the formation of soluble product can be tuned by varying $q_{s,glu}$ during lactose induction, this effect can also be observed when IPTG is used. The aforementioned maximum $q_{P,SP}$ of 9.99 mg/g/h was achieved at a $q_{s,glu}$ of 0.30 g/g/h, the existence of such an optimal $q_{s,glu}$ led to the realization that $q_{s,glu}$ affects inducer uptake, thus $q_{s,lac}$ and eventually $q_{P,SP}$. These findings led to scientific question number 3.

3) Is it possible to determine a mechanistic correlation between specific uptake rates of lactose and glucose?

As it was possible to determine the mechanistic correlation between $q_{s,glu}$ and $q_{s,lac}$, we were able to establish a simple model. This model is based on the assumption of two interconnected phenomena:

a) $q_{s,lac}$ is dependent on $q_{s,glu}$ in accordance with the Monod equation until a certain maximum is reached.
b) After reaching such a maximum, $q_{s,lac}$ decreases due to increasing substrate inhibition respectively carbon catabolite repression.

We suggest that by performing two batch experiments and up to three fed-batches at different $q_{s,glu}$ and concomitant lactose excess, enough data is available to obtain the corresponding mechanistic correlation for other *E. coli* strains. Such a model fit is performed using newly derived $q_{s,lac}$ /$q_{s,glu}$ points comparable to Fig. 3 as published in our paper.

Outlook

Refine conclusions on $q_{s,glu}/q_{s,lac}$ correlation and effects on SP/IB ratio in further experiments

While in this Thesis concomitant glucose/lactose uptake during lactose excess has been documented it is further necessary to study the behaviour under controlled $q_{s,lac}$ conditions. In doing so we expect further insights regarding tunability of expression rate and ratio of SP/IB. This can be achieved via additional fermentations under a mixed-feed setup using glucose and lactose:

- On the one hand the effect of slight variations in $q_{s,glu}$ and $q_{s,lac}$ on the determined maximum $q_{P,SP}$ is of prime interest. This could be done by lowering $q_{s,lac}$ during $q_{s,glu}$ conditions previously deemed optimal. The dotted arrow in Fig. 1 illustrates this.
- Additional insight can be obtained by performing dynamic experiments. For instance during lactose excess $q_{s,glu}$ could be dynamically varied in the course of one single fermentation rather than employing stepwise q_s shifts. Through such a dynamic variation it would be possible to further refine our knowledge on $q_{s,glu}/q_{s,lac}$ effects on maximum $q_{P,SP}$.

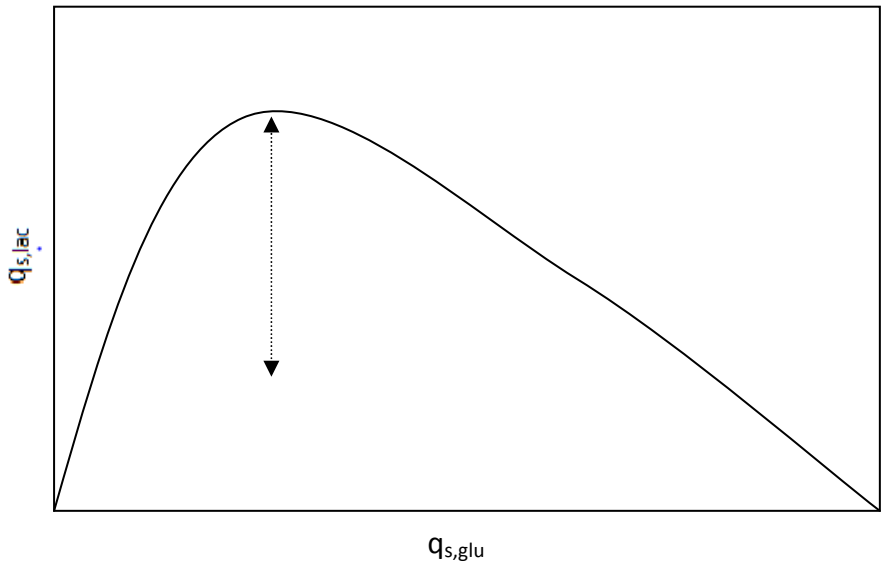


Fig. 1.: Experiments to further expand on $q_{s,glu}$ / $q_{s,lac}$ correlation - the dotted arrow illustrates a possible range for $q_{s,lac}$ variation.

Mixed feed glucose/lactose feeding strategy as platform knowledge

To broaden the scope of our mixed feed glucose/lactose feeding strategy it is necessary to implement it on other *E. coli* strains featuring different products. Partly this has already been performed in our group by Julian Quehenberger in his Master Thesis using an *E. coli* BL21(DE3) strain transformed with a pET 21a(+) plasmid harboring the T7*lac* promoter as well as the gene for the enhanced green fluorescent protein (eGFP), a cytoplasmic product. His conclusions can be summarized as follows: "The specific lactose uptake rate during concomitant glucose uptake of the strain *E. coli* BL21(DE3) depends on the expressed recombinant protein, but can be described with the same mechanistic model."(1)

In order to fully establish such a strategy as platform knowledge it is also necessary to study the applicability on a periplasmic product. Since such a protein must be able to pass the inner cell membrane from cytoplasm to periplasm, tuning of solubility and expression rate is essential to avoid accumulation issues. As the potential periplasmic product a Fab would come to mind (2).

Role of galactose

Especially when lactose pulses were added we observed a gradual accumulation of galactose in the fermentation broth. This means that galactose derived from the splitting of lactose subsequently must be transported out of the cell, as such it is only logical that such an accumulation is only detectable at latter stages of the fermentation process. In our experiments an *E. coli* BL21(DE3) strain was used which is not capable of utilizing galactose due to a mutation in the galactose metabolism. However it is capable of generating galactose via lactose metabolism which in turn can induce some *lac* operon based systems as previously reported in other studies (3).

Thus it is worthwhile to compare strains with either intact (gal+) or deficient (gal-) galactose metabolism capabilities. Especially the aforementioned accumulation of galactose could be prevented by the utilization of gal+ *E. coli* strains. Also glucose feeding could become redundant.

References Outlook

(1) Quehenberger, J., Expression tuning in *Escherichia coli* via a glucose/lactose mixed feed system. Vienna University of Technology, 2016. p. 64.

(2) Popplewell, A.G., Sehdev, M., Spitali, M. and A. N. C. Weir, Expression of Antibody Fragments by Periplasmic Secretion in *Escherichia coli.* Methods in Molecular Biology, 2005. 308: p. 17-30.

(3) Mattanovich, D., Kramer, W., et al., Rational design of an improved induction scheme for recombinant Escherichia coli. Biotechnology and bioengineering, 1998. 58(2-3): p. 296-298.

Appendix

Efforts beyond the scope of the paper

Quantification of IBs and SP via reversed-phase HPLC

In the scope of this Thesis it was attempted to develop a method for product quantification via reversed phase HPLC (RP-HPLC). RP-HPLC is potent tool for protein separation, characterization and quantification. Regarding antibody analytics several applications from full-length Ig molecules to Fabs and Fcs are possible (1). The basic principles of protein separation are well researched and a variety of columns featuring a wide range of column packing featuring different particle sizes and porosity is available.

As opposed to normal phase chromatography RP-HPLC utilizes a non-polar stationary phase combined with a polar mobile phase, interaction between analyte and column first and foremost involves the former's hydrophobic molecule components and an apolar silica based stationary phase incorporating covalently bonded alkyl chains of varying length. Polypeptides are adsorbed based upon differences in their hydrophobicity which is dependent on amino acid sequences and the resulting conformation. As polar mobile phase usually water and miscible organic solvents are used in various ratios, desorption happens dependent on variations in polarity and elution power.

Recovering the product - homogenisation and solubilisation procedure

Overview

The basic strategy for homogenisation and solubilisation of inclusion bodies (IB) for subsequent product quantification using RP-HPLC can be summarized as follows (see Fig. 1):

Re-suspension of cells		
Cells derived from cultivation broth are re-suspended in ice-cold lysis-buffer containing buffer Tris(hydroxymethyl)aminomethane (Tris) and Ethylenediaminetetraacetic acid (EDTA) added as protease inhibitor. Final biomass concentration is limited according to the capacity of the homogeniser (see SOP added beneath).		
↓		
Homogenisation		SP
Cells are homogenised using an EMULSIFLEX-C3 Homogeniser, sample is passed through a homogenising valve under pressure. Resulting supernatant contains product in soluble form (SP), recovered cell debris contains product as inclusion bodies (IB).	→	Supernatant containing SP is not processed any further.
↓		
Washing and re-suspension of IBs		
In order to remove lipids and membrane associated components cell debris is subjected to centrifugation and subsequent washing steps using Tris, NaCl, Polysorbate 80 and finally re-suspension in Tris/EDTA buffer.		
↓		
Re-solubilisation of IBs		
re-solubilisation of IBs from the re-suspended pellet is performed using guanidine hydrochloride (2), samples are then to be measured via RP-HPLC.		

Fig. 1.: Flow-chart detailing homogenisation of cells and solubilisation of IBs

As standard for HPLC quantification bovine serum albumin (BSA) was used, a cultivation sample taken before induction served as a reference for determining potential product peaks. The SP was measured without further processing steps. A detailed Standard Operating Procedure (SOP) is given below.

As reproducible results from preliminary HPLC runs could not be obtained from IB samples, it was decided to vary several parameters in the strategy. Samples from cultivations B_{Lac1} and FB_{IPTG1} were used in this matter as they encompass both induction strategies studied in this Thesis. Fig. 2 contains all parameters that were varied during method development.

Re-suspension of cells	Washing & re-suspension of IBs	Re-solubilisation of IBs
amount of fermentation broth to gain cells for re-suspension: *2 mL → 5 mL*	Volume of re-suspension buffer: *1 mL → 2.5 mL*	Reducing agent: *DTT, ß-mercaptoethanole* Solubilisation time: *30 → 120 min*

Fig. 2.: Parameters varied for IB homogenisation and solubilisation method development

Homogenisation

The original approach for homogenization required a pellet coming from 2 mL cultivation broth to be re-suspended to a maximum final biomass concentration of 5 g/L dry cell weight (DCW). Due to the homogenizer's dead volume of 13 ml the minimal required volume for homogenisation was fixed at 20 ml. As our fermentation samples mainly encompassed around 20 - 40 g/L DCW this led to a final cell concentration of roughly 0.04 - 0.08 g/20 mL (= 2 g/L) homogenisation buffer, a concentration which was deemed too low. Consequently 5 ml of cultivation broth were used in later stages, this resulted in a final cell concentration of 0.1 - 0.2 g/20 mL (= 5 g/L) homogenisation buffer.

Washing of IBs

Our approach for IBs subsequently included washing steps after which the pellet was to be re-suspended in Tris/EDTA buffer, the volume of this re-suspension buffer was adjusted from 2.5 ml to 1 ml if higher IB concentrations were desired for RP-HPLC quantification.

As mentioned SP was assumed to be present in the supernatant after homogenisation.

Solubilisation of IBs

A stock-solubilisation buffer containing 6M guanidine hydrochloride was prepared, dithiothreitol (DTT) was added as a reducing agent right before use. In subsequent experiments the solubilisation parameters were varied according to Table 1, ß-mercaptoethanol was introduced as a different reducing agent.

Table 1.: Solubilisation parameters

	Solubilisation buffer	Solubilisation time [min]
A	6 M guanidine hydrochloride, 50 mM Tris, 100 mM DTT, pH 8 (adjusted with HCl)	30/60/120
B	6 M guanidine hydrochloride, 50 mM Tris, 5 % v/v ß-mercaptoethanol, pH 8 (adjusted with HCl)	30/60/120

After solubilisation the samples were centrifuged and if any pellet formed, it was subjected to SDS - page analytics to determine whether product band was still present in the pellet,

therefore if solubilisation had been completed. For this purpose the samples were centrifuged, emerging pellets were resuspended in 10 µL of 1x Laemmli buffer and the supernatant was diluted in 2x Laemmli to a final volume of 50 µL. Then samples were heated at 95°C for 10 minutes , 10 µL of each sample were loaded onto SDS gels (8-16 %) (GE Healthcare, Little Chalfont, UK), gels were run for 90 min at 140 V and eventually stained with Coomassie Blue. Fig. 3 displays an examplary SDS - page gel from samples derived after a solubilisation time of 30 min with buffer A:

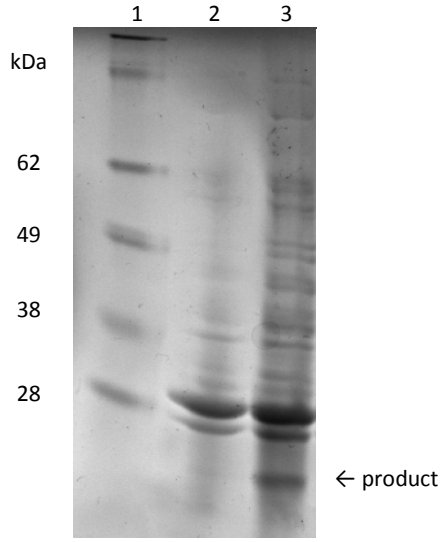


Fig. 3.: SDS-Page visualisation of incomplete solubilisation process from pellets derived after a solubilisation time of 30 min with buffer A. Lane 1, SeeBlue Plus2 Pre-stained protein standard, Lane 2, cultivation B_{Lac1} before induction, Lane 3, cultivation B_{Lac1} 6h after induction.

Clearly a solubilisation time of 30 min is not sufficient, however Fig. 4 demonstrates that the product band is barely visible after 60 and not visible after 120 min of solubilisation with buffer B. Therefore it was decided to stick with buffer B, utilizing a solubilisation time of 120 min to be sure.

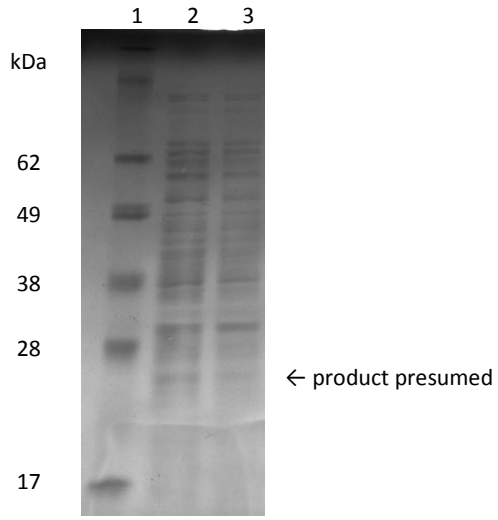


Fig. 4.: SDS-Page visualisation of solubilisation process from samples derived after a solubilisation time of 60/120 min with buffer B. Lane 1, SeeBlue Plus2 Pre-stained protein standard, Lane 2, cultivation B_{Lac1} 6h after induction/solubilisation time of 60 min, Lane 3, cultivation B_{Lac1} 6h after induction/solubilisation time 120 min.

Quantification of IB and SP via RP-HPLC

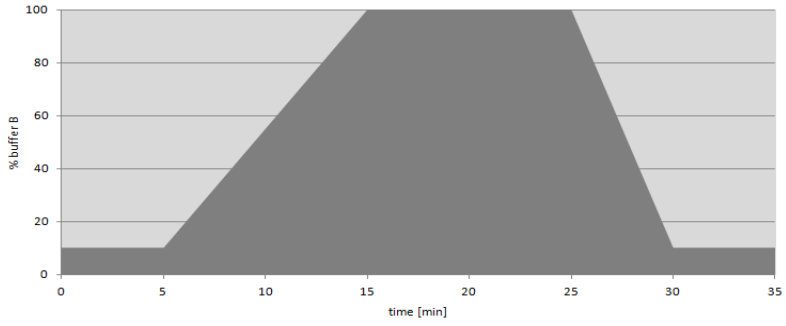
For RP-HPLC analysis a reversed-phase C8 column featuring a particle size of 5 µm and a 300 Å pore size was used. Larger particle sizes (15-30 µm) lead to peak width broadening thereby causing loss in resolution and are utilized in larger preparative work where an "overload" of the column is welcomed to enhance analyte throughput. Generally 5 µm particle sizes are preferred in analytical and small preparative separations. For polypeptides with a molecular weight higher than 2000 Da pore sizes of 300 Å are considered reasonable, as smaller pores would prevent the polypeptide from entering (3).

H_2O and ACN was used as mobile phase, additionally trifluoroacetatic acid (TFA) was added in order to avoid ionic interaction with the stationary phase, a method which is well-established in RP-HPLC (3).

First Trial

For preliminary runs the following HPLC settings were applied:

Table 2.: Preliminary HPLC settings

HPLC	UltiMate 3000; Thermo Fisher, Waltham, MA, USA
HPLC Column	EC 150/4.6 Nucleosil 300-5 C8; Macherey-Nagel, Düren, Germany column volume: 2.5 mL
Flow rate	2 mL/min
Buffers	Buffer A: MQ H_2O + 0.1 % v/v TFA Buffer B: ACN + 0.1 % v/v TFA
Column temperature	50 °C
Detection wavelength	280 nm
Injection volume	5 µL
Elution profile	

These HPLC settings led to unsatisfactory results. Though it was possible to determine retention times for peaks associated with the product, peak separation was not sufficient: in fact two overlapping peaks occurred which both displayed increasing intensity with longer cultivation induction times. Exemplary data from corresponding overlapped chromatograms can be found in Fig. 5, only IB samples are displayed.

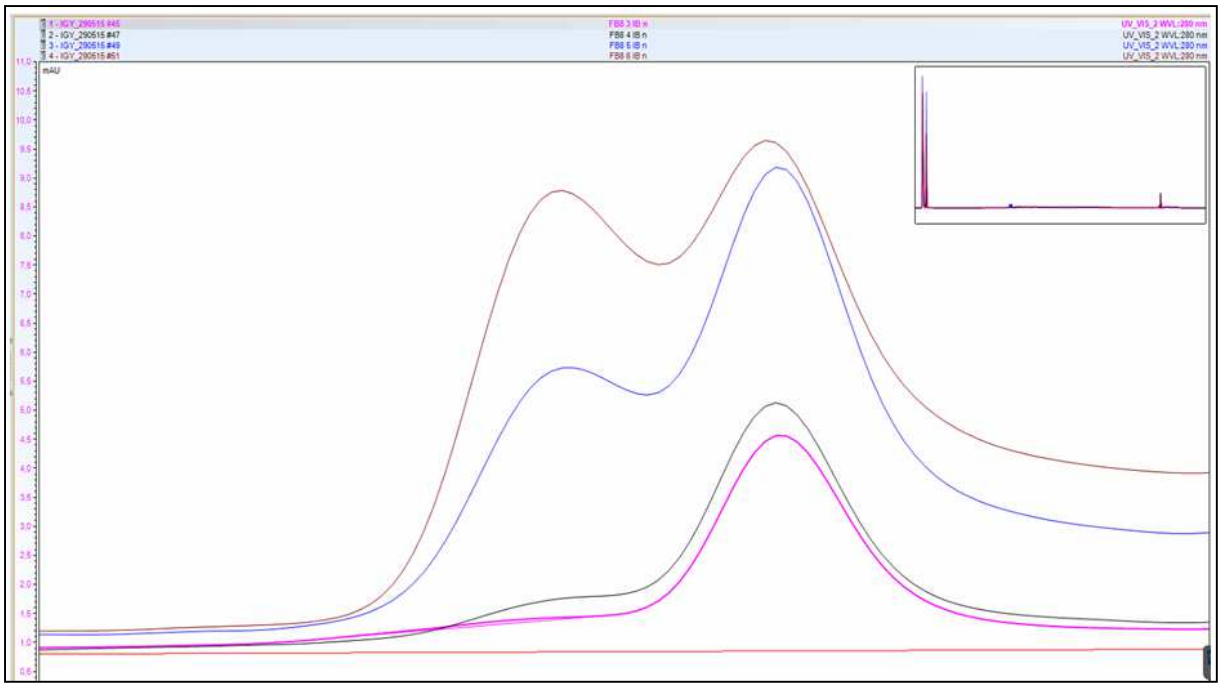


Fig. 5.: HPLC with settings from Tab. 2, excerpt from overlapped chromatograms - FB_{IPTG2} IBs, magenta/black: before induction, blue: 2.5 h after induction, brown: 4.5 h after induction

This first trial revealed problems regarding the peak separation performance, to adress these issues adjustments were made in a second trial:

Second Trial

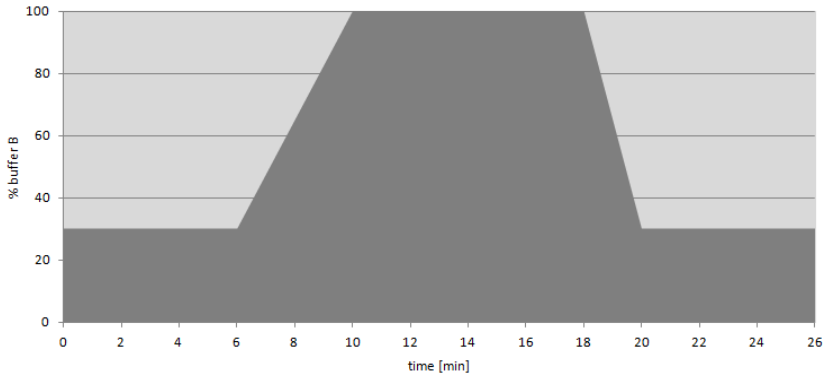
Preliminary runs utilized a gradient of 10 - 100 % of buffer B (ACN) in 8 column volumes (CV) (see Table 2), since the primary goal was to develop only a rough quantification method and due to time constraints, it was decided to first and foremost adjust the gradient. Also the column temperature was lowered to ensure product stability, additionally the injection volume was increased. After each sample one run solely injecting solubilisation buffer was done to make sure that residual protein was removed from the column. Changes to HPLC settings are summarised in Table 3.

Table 2.: Adjusted HPLC trial settings

Column temperature	30°C (lowered from 50 °C)
Injection volume	10 µL (from 5 µL)
Elution profile	Adjusted gradient: 30 - 100 % of buffer B (ACN) in 3,6 or 9 CVs blank run with solubilisation buffer performed after each sample run

An adjusted gradient utilizing 9 CVs improved peak separation, however this was not constructive for our purposes because it led to further issues as now two peaks had to be considered for quantification calculations. Therefore efforts were made to come up with only one quantifiable peak. As expected the use of 3 CVs led to a combination of both formerly separated peaks into one. Consequently the settings summarised in Table 4 were applied in the following HPLC runs, note that the method became considerable shorter (26 as opposed to 35 min). Corresponding chromatograms are displayed in Fig. 6 and Fig. 7.

Table 4.: Adjusted HPLC settings

HPLC	UltiMate 3000; Thermo Fisher, Waltham, MA, USA
HPLC Column	EC 150/4.6 Nucleosil 300-5 C8; Macherey-Nagel, Düren, Germany column volume: 2.5 mL
Flow rate	2 mL/min
Buffers	Buffer A: MQ H_2O + 0.1 % v/v TFA Buffer B: ACN + 0.1 % v/v TFA
Column temperature	30 °C
Detection wavelength	280 nm
Injection volume	10 µL
Elution profile	blank run with solubilisation buffer performed after each sample run 

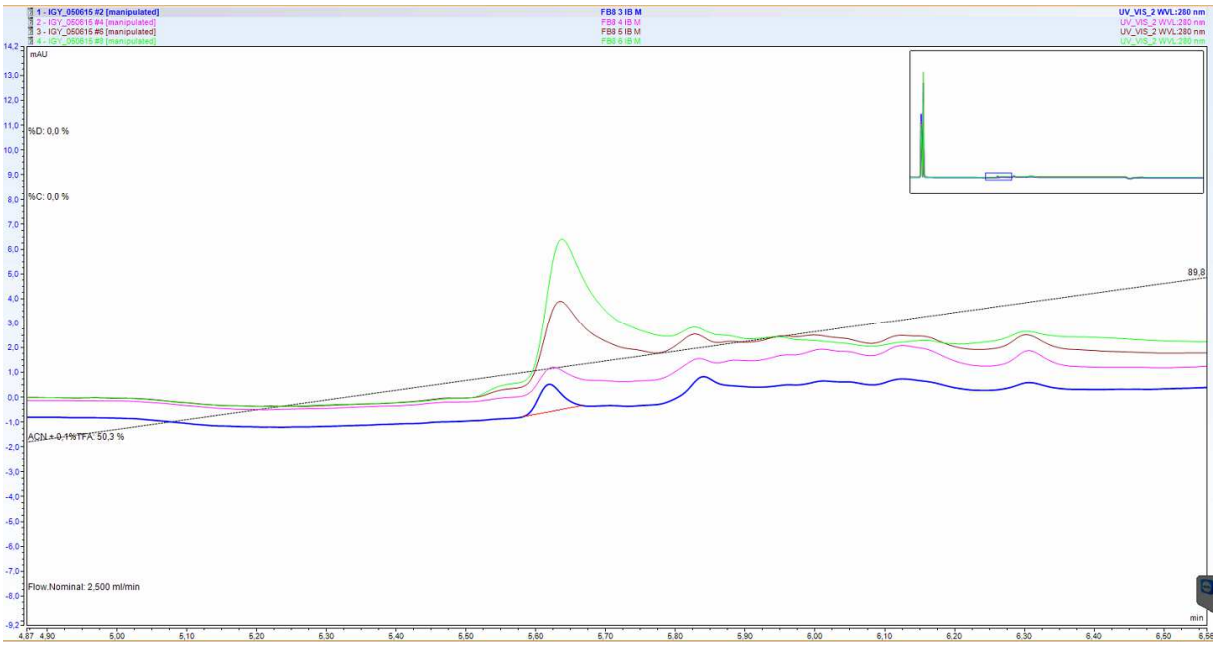


Fig. 6.: HPLC with adjusted settings (see Tab. 4), overlapped chromatograms - FB_{IPTG2} IBs, blue: before induction, magenta: 2.5 h after induction, brown: 4.5 h after induction, green: 6.5h after induction

Fig. 6 clearly displays an increasing, quantifiable peak area with extended induction time. However when chromatograms taken from samples before and after induction are overlapped, a drift respectively offset towards higher intensities is visible throughout the whole run, making the chromatograms problematic but not impossible for quantification purposes. As these samples are taken from a cultivation employing IPTG as inducer at a $q_{s,Glu}$ of 0.035 g/g/h, it is not surprising that the yield for SP seemed to be notably lower than for IBs (see section Paper Draft). Regarding SP the adjusted HPLC method did not deliver quantifiable results due to a poor signal/noise ratio (see Fig. 7).

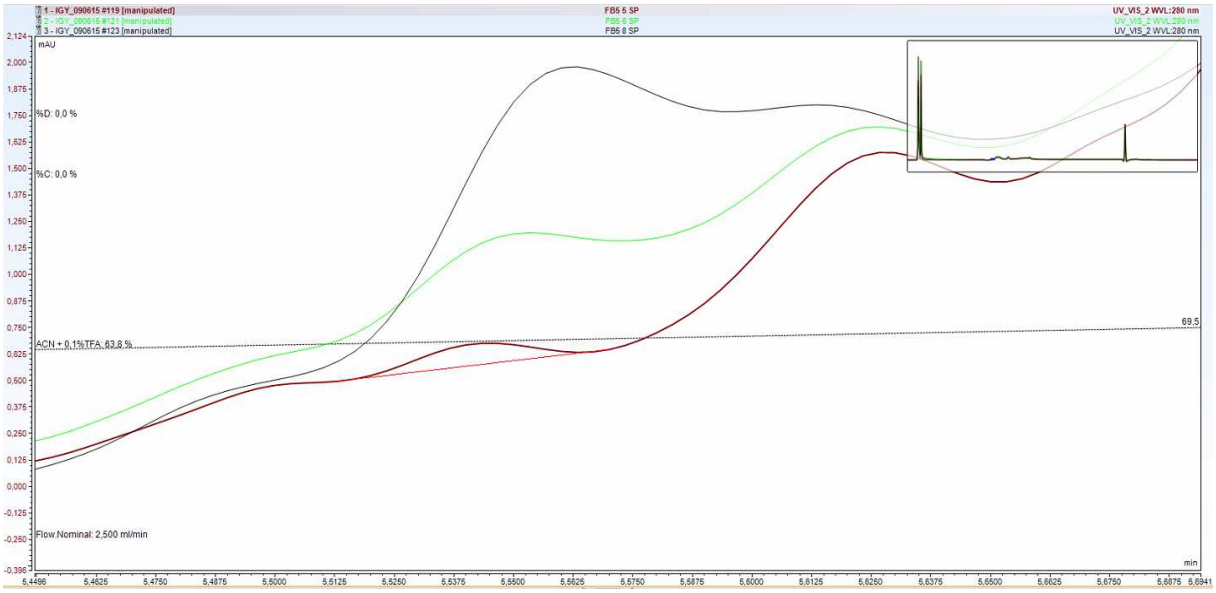


Fig. 7.: HPLC with adjusted settings (see Tab. 4), overlapped chromatograms - FB_{IPTG2} SP, brown: before induction, green: 2.5 h after induction, brown: 4.5 h after induction, black: 6.5 h after induction

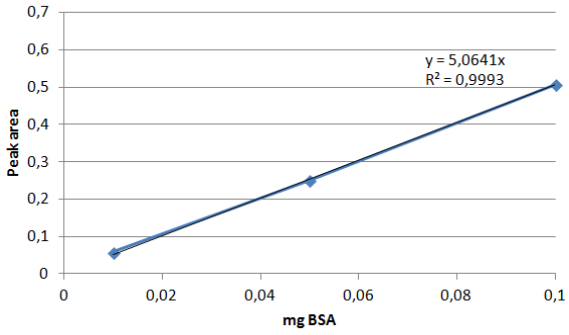
To determine the reliability of HPLC quantification results were examined for plausibility and furthermore cross-checked with already obtainable SDS-page quantifications.

Comparison with SDS-page quantification

Exemplarily the quantification of IB samples taken from FB_{IPTG2} will be demonstrated here:

As mentioned BSA standards were employed to generate a regression line which was applied to measured peak areas in order to quantify the results. Standard concentrations and corresponding data as well as comparisons with SDS-Page quantification can be found in Table 5. Generally HPLC quantification led to higher concentration results.

Table 5.: Exemplary IB quantification of FB_{IPTG2} via HPLC and SDS-Page. Sample 5: before induction, 6: 2.5 h after induction, 7: 4.5 h after induction, 8: 6.5 h after induction

Regression					
BSA [mg]	Peak area	Regression line including R-squared, coefficient of determination			
0.01	0.0583				
0.05	0.2501				
0.1	0.5072				
Note that the previously mentioned SOP for IB re-solubilisation (attached below) required 300 µL of IB sample to be mixed with 1200 µL of solubilisation buffer which ultimately led to a total amount of 9 mg CD in 1.5 mL of re-solubilised sample solution.					

FB_{IPTG2} IB - HPLC quantification					**SDS-Page quantification**
Sample	Peak area	Corrected area	mg IB / 9 mg CD	mg IB / g CD	mg IB / g CD
5	0.0088	0	0	0	0
6	0.0427	0.0339	0.00669	1.12	0.42
7	0.1421	0.1332	0.02630	5.52	1.99
8	0.2396	0.2308	0.04558	11.46	2.73

Inconsistencies occurred when the measurement of SP concentration via HPLC was attempted. As Table 6 demonstrates, SP seems to be present although it is not measureable with SDS-Page. Furthermore SP concentration is found to be decreasing at longer induction times, a trend which is highly unlikely and was never encountered in the scope of this Thesis. These mentioned issues concerning SP quantification also (and more significantly) occurred when cultivations induced with lactose were being analysed.

Table 6.: Exemplary SP quantification of FB_{IPTG2} via HPLC and SDS-Page. Sample 5: before induction, 6: 2.5 h after induction, 7: 4.5 h after induction, 8: 6.5 h after induction

FB_{IPTG2} SP - HPLC quantification					**SDS-Page quantification**
Sample	Peak area	Corrected area	mg SP	mg SP / g CD	mg SP / mg CD
5	0.0011	0	0	0	not detectable
6	0.0195	0.0195	0.00391	0.46	n.d.
7	0.0087	0.0087	0.00173	0.21	n.d.
8	0.0093	0.0093	0.00184	0.22	n.d.

Conclusions

Although the development of a robust RP-HPLC quantification strategy was leading to some promising - but only preliminary - results, it had to be abandoned for this Thesis due to time constraints. The reason for this can be found in some unrealistic results, poor reproducibility and significant discrepancies when compared to the already established method of SDS-Page quantification. Probably the parameters used were not entirely suitable for the product leading to stability issues.

Despite these aforementioned issues, this described method was successfully adapted and modified by Julian Quehenberger for another product in the scope of his Master Thesis (4) For this study an *E. coli* BL21(DE3) strain was used. It was transformed with a pET 21a(+) plasmid harbouring the T7*lac* promoter as well as the gene for the enhanced green fluorescent protein (eGFP) which served as the model protein.

Some speculations as to why the method did not work properly on our product and possible solutions can be found in Table 7. Personally I would recommend to refine the solubilisation process by further varying parameters like solubilisation time and composition of solubilisation buffer.

Table 7.: Possible issues related to product quantification using RP-HPLC

Possible issue	Potential action
Product stability Product is not stable in solubilisation buffer	Change composition of solubilisation buffer, use different reducing agent
Product stability Precipitation occurs when subjected to running buffer (ACN/H_2O) in the column, resulting carry-over effects falsify quantification results	Check if precipitation is observable when running buffer is added to sample
Product stability Product stability deteriorates with time	Make sure that time periods between solubilisation and measurement remain constant for each sample
Low titres IB titres are significantly lower than SP titres and there-fore IBS are not easily detectable	Further concentrate sample before solubilisation

Low sensitivity of UV - detector Absorption at wavelength of 280 nm is strictly due to tryptophan and/or tyrosine, however amino acid distribution can vary from sample to sample	Use other wavelengths (f.i.: 214 nm for peptide bonds)

The HPLC settings used and modified by Julian Quehenberger can be found in the attached SOP given below.

References

(1) Zhang, T., et al., HPLC for Characterization and Quality Control of Therapeutic Monoclonal Antibodies. LCGC North America, 2014. 32 (10).

(2) H.F. Cueto-Rojas, H. F., Interferon-α 2b quantification in inclusion bodies using Reversed Phase-Ultra Performance Liquid Chromatography (RP-UPLC). Journal of Chromatography B, 2010. 878 (13-14): p.1019-1023.

(3) Grace VYDAC, The Handbook of Analysis and Purification of Peptides and Proteins by Reversed-Phase HPLC. http://wolfson.huji.ac.il/purification/PDF/ReversePhase/VYDAChandbookRPC.pdf (11.04.2016)

(4) Quehenberger, J., Expression tuning in *Escherichia coli* via a glucose/lactose mixed feed system. Vienna University of Technology, 2016. p. 64.

SOP: Quantification of EGFP inclusion bodies and soluble protein via RP-HPLC

Materials

- Solubilized IB samples or SP samples
- Bovine serum albumin (BSA)
- Solubilization buffer prepared according to SOP “Homogenistation and Solubilisation of IBs”
- Acetonitrile HPLC grade (ACN)
- MQ water
- Trifluoro acetic acid HPLC grade (TFA)
- HPLC vials, caps and septa
- syringe filters (0.2 µm pore size), syringes and hollow needles

Equipment

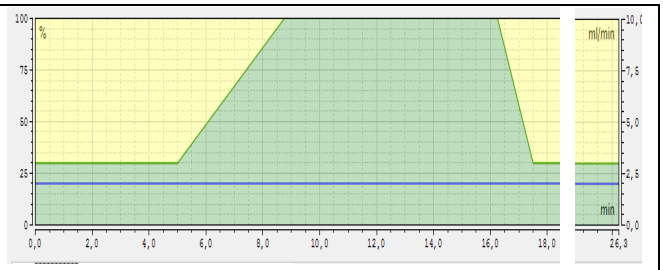
- 50 mL volumetric flask
- HPLC system
- RP-HPLC column EC 150/4.6 Nucleosil 300-5 C8 (Macherey-Nagel, Düren, Germany)

Procedure

- Preparation of Buffer A
    - MQ water + TFA (0.1%(v/v))
    - Sonicate for 15 minutes
- Preparation of Buffer B
    - ACN + TFA (0.1%(v/v))
    - Sonicate for 15 minutes
- Preparation of BSA standards
    - Dissolve 50 mg BSA in a 50 mL volumetric flask with deionized water
    - Prepare standards: 1.0 g/L, 0.75 g/L, 0.5 g/L, 0.4 g/L and 0.3 g/L by diluting with deionized water
    - Store at 4 °C
- Sample preparation
    - Dilute samples below 1.0 g/L EGFP if necessary
    - Samples must be filtered with syringe filters (0.2 µm)
    - Transfer into HPLC vials
    - Store at 4 °C
- HPLC settings
    - Flush the column for 30 min with 30% Buffer B
    - Flow rate: 2 mL/min
    - Column temperature: 30 °C
    - Upper pressure limit: 300 bar
    - Injection volume: 10 µL
    - Detection wavelength: 280 nm
    - After each sample and standard perform one blank run with solubilization buffer to remove residual protein off the column

- Elution profile

Time [min]	Buffer B [%]
0.00	30
5.00	30
8.75	100
16.25	100
17.50	30
26.25	30



- Retention times (in minutes)
    - BSA standard (can vary with supplier and production method): 6.95
    - EGFP (IBs): 7.10
    - EGFP (SP): 7.08
- Peak integration
    - Set the baseline manually according to the blank runs with solubilization buffer
    - Vertically split adjacent peaks
- Finishing the measurement
    - Flush the column for at least 30 minutes
    - Store the column in 40% Buffer A + 60% Buffer B
    - Remove the column and shut down the HPLC system according to the standard procedure

SOP: Homogenistation and Solubilisation of IBs by Wieland Reichelt, adapted by Lukas Veiter

TU VIENNA Technische Universität Wien Vienna University of Technology	**Standard Operating Procedure**	BioVT
	Title **Homogenistation and Solubilisation of IBs**	Page:83/86
Research Division ***Biochemical Engineering***	SOP Number: SOP Status: Effective/Editable	Date: 24.05.2015

Version	Number.0 for authorized versions; Number.1 for editions of a version, etc
Replaced version	
Author	Wieland Reichelt; Lukas Veiter
Date	
Signature	
Authorized by	
Date of authorization	
Signature	

Summary	Preparation solubilisation puffer

Materials

- Tris
- Guanidin Hydrochlorid **(98% for DSP/ 99,5 for HPLC analytics)**
- 2-Mercaptoethanol

Equipment

- Emulsiflex©-C3 HOMOGENIZER (read Operation of Emulsiflex©-C3 HOMOGENIZER before usage)
- Table top centrifuge
- Shaker
- pH meter

Procedure

Resuspension of Cell Pellets

i) Prepare lysis-buffer 100mM Tris (12,14 g/l), 10mM Na2EDTA (3,7 g/L) and adjust pH to 7,4 with HCl

ii) The pellet of 5 ml fermentation broth is resuspended on a scale in ice cold lysis buffer to a maximal final biomass concentration of 5 g/l (BDW). Due to the dead volume of 13 ml the minimal required volume for homogenisation is 20 ml.

Homogenisation and Washing of IBs

i) Prepare Buffer A: 50mM Tris, 0.5M NaCl, 0.02% Tween 80 (w/v) pH 8

ii) Prepare Buffer B: 50 mM Tris, 5mM EDTA, pH 8

iii) The cell suspension is homogenised at 1500 bar in 6 passages; it takes 13 ml 21 seconds for one passage (3min 30sec for 20ml)

iv) Rinse homogeniser after each sample with 30ml water.

v) Store sample on ice

vi) Take 15 mL of homogenate and centrifuge it at 13000 g for 15 minutes at 4 °C.

vii) Wash pellet with 15 mL buffer A and centrifuge again (13000 g, 15 min, 4°C)

viii) Resuspend pellet in __ ml buffer B, sample can now be stored at -20 °C

Solubilisation

i) Prepare solubilisation buffer 6M Guanidin Hydrochlorid (573,18g/l) and 50mM Tris (6,06 g/l) and adjust the pH to 8 (HCl) [(1)]

ii) Add 2-Mercaptoethanol (5 % volume) right before use.

iii) Take 300 µl of washed IB sample from step viii.

iv) Add 1200 µl of solubilisation buffer and vortex.

v) Put samples on shaker for 2h, vortex every 30 min.

vi) Centrifuge samples at 3000 rpm for 10 min.

vii) Samples are filled in HPLC vials through a syringe filter (0.2 µm)

Literature

- (1) 2010 H.F. Cueto-Rojas; „Interferon-α 2b quantification in inclusion bodies using Reversed Phase-Ultra Performance Liquid Chromatography (RP-UPLC)“

Acceptance of paper for publication

Von: em.amab.2.4b01f0.2667317d@editorialmanager.com
<em.amab.2.4b01f0.2667317d@editorialmanager.com> im Auftrag von Alexander Steinbüchel <em@editorialmanager.com>
Gesendet: Samstag, 7. Mai 2016 07:46
An: Spadiut Oliver
Betreff: AMAB: Your manuscript entitled The E. coli pET expression system revisited - mechanistic correlation between glucose and lactose uptake

Ref.: Ms. No. AMAB-D-16-00582R1

Dear Dr. Spadiut,

It is a pleasure to accept your manuscript in its current form for publication in Applied Microbiology and Biotechnology.

The article proofs will be sent to you about 3-4 weeks after receipt of this email.
You may check the progress of the publication process at My Springer under Article Tracking (http://www.springer.com/generic/registration?SGWID=0-40105-6-1004221-0).

Thank you for your fine contribution. On behalf of the Editors of the "Applied Microbiology and Biotechnology", we look forward to your continued contributions to the Journal.

Best regards,

Alexander Steinbüchel
Editor-in-Chief
Applied Microbiology and Biotechnology
steinbu@uni-muenster.de