Universität für Bodenkultur Wien

Christian Doppler Forschungsgesellschaft

LONG-TERM STABILITY OF *ESCHERICHIA COLI* EXPRESSION SYSTEMS UNDER PRODUCTION CONDITIONS

Dissertation

for obtaining a doctorate degree

at the University of Natural and Life Sciences, Vienna

Department for Biotechnology

Institute of Bioprocess Science and Engineering

Head of Institute: Alois Jungbauer, Univ.Prof. DI. Dr.nat.techn

Advisor: Gerald Striedner, Assoc. Prof. Dipl.-Ing. Dr.nat.techn

Submitted by

Artur Schuller, Dipl.-Ing.

Vienna, November 2020

Danksagung

Zuallererst möchte ich mich bei meinem Betreuer Prof. Gerald Striedner bedanken. Er betreute mich bereits während meiner Masterarbeit bei Boehringer Ingelheim und gab mir anschließend die Möglichkeit eine Dissertation im neu gegründeten Christian Doppler Labor zu machen. Vielen Dank für die Unterstützung, wenn sie notwendig war aber auch für die Überlassung der freien Hand bei der Bearbeitung. Ich konnte mich sehr glücklich schätzen an einem so umfangreichen und interessanten Thema arbeiten zu können. Ebenso geht ein großer Dank an Dr. Monika Cserjan, welche mich zu Beginn meiner Arbeit sehr intensiv bei der Planung und Durchführung der Experimente unterstützt hat aber auch meine teilweisen sehr konfusen Manuskripte sehr geduldig korrigiert und mit zahlreichem Input optimiert hat. Ebenso geht ein Dank an mein Betreuer Team, Prof. Reingard Grabherr und Prof. Rainer Hahn, welche bei den zahlreichen Jour fixe Meetings zum Gelingen meiner Dissertation beigetragen haben.

Dann möchte ich mich bei der gesamten AG Striedner für die gute Zusammenarbeit und Hilfsbereitschaft bedanken. Nicht nur was die Arbeit anging, sondern auch abseits davon. Gerne denke ich an die Kaffeepausen auf unserer geliebten und überhaupt nicht dreckigen Couch im 2. Stock zurück oder an die zahlreichen fröhlichen Abende, sodass mittlerweile sehr viele Arbeitskollegen zu guten Freunden geworden sind. Besonders möchte ich mich bei meinem Wegbegleiter seit der Bachelor Zeit, Patrick Stargardt für die zahlreichen Empfehlungen zu Serien auf Netflix, günstigen Handyverträgen und supergünstigen Amazon Warehouse Deals bedanken.

Meinen Eltern bin ich für ihre bindungslose Unterstützung seit meinem ersten Studientag in Wien unendlich dankbar. Auch wenn ihr wahrscheinlich nie so ganz verstanden habt was genau ich da studiere, konnte ich mich immer, wenn es notwendig war auf eure Hilfe verlassen und ich kann euch nicht genug danken, dass ihr mir all das überhaupt ermöglicht habt. Ebenso möchte ich mich auch bei meinen Bruder Alex und seiner Frau Sabrina bedanken, für ihre Unterstützung, Interesse, ehrlichen Meinungen und sehr viel Ablenkung obwohl ich eigentlich hätte lernen sollen.

Mein größtes Dankeschön möchte ich an meine Frau Maria richten, welche mich nun schon seit mehr als 10 Jahren auf diesem Weg begleitet, mich unterstützt, aufmuntert und ermutigt diesen Weg zu gehen. Es ist wunderschön mit dir die Höhen und Tiefen unseres Lebens gemeinsam zu beschreiten. Danke!

Abstract

The genome integrated T7 expression system offers significant advantages in terms of productivity and quality, even if the gene of interest (GOI) is only present as a single copy per cell. In comparison to plasmid-based expression systems, this system has no plasmid-mediated metabolic load and the dosage of the GOI does not vary during the production process. Long-term cultivation with the T7 expression system is, however, only possible to a limited extent because mutations in the T7 RNA polymerase (RNAP) lead to a faster growing non-producing population. An investigation of the long-term effects of recombinant protein expression on the mutation pattern of *E. coli* is therefore not feasible.

In the course of this work, the BL21Q expression system was developed, which was successfully patented at the European Patent Office. It is based on the host RNAP in order to exclude mutations in the RNAP. The expression system was first tested for applicability with regard to genome-integrated recombinant protein production. Promoter / operator combinations were developed to achieve the desired properties, such as high expression rate, low basal gene expression and fine-tuning of gene expression by varying the inducer concentration. The newly developed expression system BL21Q was then examined for long-term stability under production conditions.

We performed repetitive fed-batch-like microbioreactor cultivation under production conditions. The easy-to-produce green fluorescent protein (GFP) and the difficult-to-produce Fab fragment served as model proteins. The BL21Q expression system remained stable in long-term cultivation during the production of the easy-to-produce protein GFP. For the production of proteins that are difficult to produce, mutations in the *lacI* gene of the BL21Q derivatives proved positive for long-term stability. Furthermore, mutations could be identified which play a role in the metabolism of different sugar sources. Our results showed that the adaptive evolution that was carried out with the genome-integrated *E. coli* BL21Q expression system in microbioreactor cultivations under industrially relevant production conditions is an efficient tool for the identification of mutation hotspots.

Zusammenfassung

Das genomintegrierte T7-Expressionssystem bietet erhebliche Vorteile hinsichtlich Produktivität und Qualität, selbst wenn das zu exprimierende Gen (*gene of interest*, GOI) nur als Einzelkopie in der Zelle vorliegt. Im Vergleich zu Expressionssystemen auf Plasmid Basis weist dieses System keine durch Plasmide vermittelte metabolische Belastung auf und die Dosierung des GOI variiert während des Produktionsprozesses nicht. Langzeitkultivierung mit dem T7-Expressionssystem sind allerdings nur eingeschränkt realisierbar da Mutationen in der T7-RNA-Polymerase (RNAP) zu einer schneller wachsenden nicht-produzierenden Population führen. Dadurch ist eine Untersuchung der Langzeit Auswirkungen der rekombinanten Proteinexpression auf das Mutationsmuster von *E. coli* nicht möglich.

Im Zuge dieser Arbeit wurde das BL21Q Expressionssystem entwickelt, welches erfolgreich beim Europäischen Patentamt patentiert werden konnte. Es basiert auf einem Promoter, welcher von der Wirts-eigenen RNA Polymerase erkannt wird, um Mutationen in der RNAP ausschließen zu können. Das Expressionssystem wurde zunächst auf die Anwendbarkeit hinsichtlich genomintegrierter rekombinanter Proteinproduktion getestet. Dafür wurden Promoter/ Operator Kombinationen entwickelt, um die gewünschten Eigenschaften, wie hohe Expressionsrate, geringe basale Genexpression und Feinabstimmung der Genexpression durch variierende Induktor Konzentration, zu erzielen. Das neu entwickelte Expressionssystem wurde anschließend hinsichtlich Langzeitstabilität unter Produktionsbedingungen untersucht.

Wir führten wiederholende fed-batch-ähnliche Mikrobioreaktorkultivierungen unter Produktionsbedingungen durch. Als Modellproteine dienten das einfach herzustellende Grün fluoreszierende Protein (GFP) und ein schwer zu produzierendes Fab-Fragment. Das auf Wirt-RNA-Polymerase basierende BL21Q-Expressionssystem blieb bei der Produktion des einfach herzustellenden Proteins GFP in Langzeitkultivierungen stabil. Für die Produktion schwer herzustellender Proteine erwiesen sich Mutationen im lacI-Gen der BL21Q-Derivate als positiv für die Langzeitstabilität. Des Weiteren konnten Mutationen identifiziert werden, welche eine Rolle im Metabolismus unterschiedlicher Zuckerquellen spielen. Unsere Ergebnisse zeigten, dass die adaptive Evolution, die mit dem genomintegrierten *E. coli* BL21Q Expressionssystemen in Mikrobioreaktorkultivierungen unter industriell relevanten Produktionsbedingungen durchgeführt wurde, ein effizientes Werkzeug zur Identifikation von Mutations-Hotspots ist.

Table of contents

1 Introduction

1.1 *E. coli* as production host for recombinant protein production

Since the first production of a human recombinant protein over 40 years ago (1), *E. coli* has established itself as a cell factory and is still being further developed. The advantages of *E. coli* are well known, such as the extremely rapid growth. Under optimal growth conditions, its doubling time is approximately 30 minutes (2). Furthermore, very high cell densities can be achieved. In theory it is possible to cultivate *E. coli* up to a cell density of approximately 175 g cell dry weight per liter (3). In addition, *E. coli* can grow in defined medium consisting of inexpensive components. Genetic manipulation is very simple and can be done in less than a few hours (4). There is an entire toolbox of various genetic tools to meet the requirements of the protein to be produced, and this is constantly being expanded.

Of the numerous different *E. coli* strains, four (K-12, B, C, and W) are regarded as model organism strains, with only K-12 and B strains being relevant for biotechnological use today. The K-12 strain *E. coli* MG1655 is considered the "wild-type" strain of numerous K-12 strains. This strain was sequenced by the Blattner laboratory because it most closely corresponds to the wild-type *E. coli* and could be maintained as a laboratory strain with minimal genetic manipulation. Only the bacteriophage lambda and the F plasmid were removed using ultraviolet light and acridine orange (5). These mutations are found in most K-12 strains used today. For this reason, MG1655 is also used as a model organism for *E. coli* genome, transcriptional regulation, transporters, and metabolic pathway databases. Strains originating from MG1655 include DH1 parent from DH5α and DH10β (6). These strains have mutations in recA1 and endA1. RecA1 prevents homologous recombination and endA1 the degradation of inserted plasmid DNA. Commercially available strains are already chemically competent and have been engineered for maximum transformation efficiency so that they are mainly used for the cloning of plasmid DNA and less for the production of recombinant proteins. Among the K-12 strains that are used to produce recombinant proteins, HMS174 is of particular biotechnological relevance. This strain also contains the recA mutation, which avoids the recombination of plasmid DNA, which has a positive effect on the plasmid stability during the production of recombinant proteins (7). BL21, arguably the most prominent representative of the B line, was described by Studier in 1986 when he made numerous modifications to the B line (8). BL21 cells are deficient in the Lon protease, which degrades both host cell proteins and foreign proteins (9). BL21 also lacks the outer membrane protease OmpT, which degrades extracellular proteins. Deletion of OmpT is particularly advantageous from a biotechnological point of view, since after cell lysis the recombinant

protein cannot be degraded by OmpT (10). Furthermore, plasmid loss is reduced by the hsdSB mutation, since the DNA methylation and degradation is interrupted (11). Due to the higher expression of the two sugar transporters MalE and GatD, the active sugar transport into the cell is higher compared to HMS174. BL21 thus achieves higher growth rates and lower acetate accumulation than K-12 strains if the cells are cultivated in large scale fermentation conditions (7).

1.2 Promoters: Strength, Basal Expression and Regulation

The first step in protein biosynthesis is the transcription and the selection of a suitable promoter is essential. The demands that are generally made on a promoter system in biotechnology are high expression of the GOI, low basal expression in the absence of the inducer and ideally the transcription should be tunable by means of varying inducer concentration.

A promoter is the DNA sequence to which the RNA polymerase (RNAP) binds and initiates the transcription of the messenger RNA (mRNA). In bacteria, the promoter consists of two short sequence elements which are approximately 10 (Pribnow Box) and 35 nucleotides upstream from the transcription start site. A consensus sequence of the two elements could be derived by sequence alignment of all *E. coli* promoters. The sequence at -10 has the consensus sequence TATAAT and at -35 TTGACA (12). Although these sequences are conserved on average across all promoters, there are only a few natural promoters that precisely contain these sequences. Artificial promoters, which contain the complete consensus sequence, showed reduced transcription frequencies (13). The optimal distance between -10 and -35 region is 17 bp and it was shown that the range can be very variable and is nevertheless recognized by the RNAP as a promoter sequence (14). For the transcription to take place in bacteria, sigma factors are required. Sigma factors are proteins that bind to the RNAP and initiate the specific binding of the RNAP to the promoter. The sigma factor together with the RNAP is known as RNA polymerase holoenzyme. *E. coli* has seven sigma factors and depending on genes and environmental conditions, a different sigma factor binds to the RNAP. The sigma factor 70 is the “housekeeping factor” and transcribes most of the genes of the growing *E. coli* cell (15). Depending on the environmental conditions, other sigma factors can become active, such as sigma 38 during starvation in the stationary phase or sigma 32 during heat shock. In this way, only genes and pathways are activated that are necessary for the respective situation (16).

The undisputedly best-known promoter is the lac promoter, which is part of the lac operon in *E. coli*. In the presence of lactose, the LacZYA genes are induced and lactose can be metabolized as a carbon source. Since its discovery in 1961 by Francois Jacob and Jacques Monod, the lac operon has been continuously developed for biotechnological use (17). A first step was the reduction of the so-called catabolite repression, which makes certain carbon sources of *E. coli* more preferred. If glucose and lactose are present in the medium at the same time, the lac promoter is not completely induced until glucose has been fully utilized (18). To make the lac operon inducible even in the presence of glucose, a mutation was introduced which led to the lacUV5 promoter. This property was caused by two changed base pairs in the -10 region of the lac promoter (14, 19). Synthetic hybrids that combine the strengths of other promoters with the advantages of the lac operon are available. The tac promoter consists of the -35 region of the tryptophan promoter and the -10 region of the lac promoter. As a result, the strength of the promoter (mRNA molecules produced per unit of time) could be increased by a factor of 10 (20).

The pET vector (Novagen), which is part of the T7 expression system, is one of the most popular expression systems for the production of recombinant proteins. It essentially consists of the T7-RNAP and the T7 promoter from the phage T7 (21). The T7 RNAP only recognizes its own promoter, does not require any additional sigma factors and is therefore a very good method of assimilating the host's metabolism and using it for the production of recombinant proteins. For biotechnological use, the T7 RNAP is either made available on a separate plasmid or, as usual, integrated into the bacterial chromosome using the prophage lambda DE3. The T7 RNAP is under the control of the lacUV5 promoter, which means that the system can be induced with lactose or its non-metabolizable analogue isopropyl β-D-1-thiogalactopyranoside (IPTG) even in the presence of glucose.

1.2.1 Regulation of Gene Expression and Basal Expression

A major disadvantage of lac-based expression systems, especially when high-copy number plasmids are used, is the unacceptable high gene expression in the absence of the inducer. This phenomenon is known as basal expression or system leakiness. For challenging proteins, even low basal expression can have negative effects on host metabolism or can even be toxic. Therefore, transforming such an expression system can be difficult. This is evident in a low frequency of transformants. The tightness of gene regulation is therefore an important quality feature of an expression system (22, 23). In the absence of lactose, the lac inhibitor LacI forms a homotetramer that binds to the lac operator site (*lacO*) and suppresses the transcription of the lacZYA operon (24). Conversely, if lactose or IPTG binds to LacI, it induces a conformational change in the protein structure and LacI can no longer bind to the

lacO site. This leaves the *lacO* site open for RNAP binding and the transcription can begin. The efficiency of repression essentially depends on two factors, the number of LacI molecules per cell and the binding affinity of the LacI to the *lacO* site. The binding affinity of the LacI to *LacO* results from the symmetry of the *lacO* site. The *lacO* sites are DNA sequences with inverted repeat symmetry (25). The higher the symmetry, the greater the LacI binding affinity for the operator sequence. An artificial, perfectly symmetrical *lacO* (sym-lacO) binds LacI with the greatest affinity (26). In contrast, the three wild-type operators *lacO1*, *lacO2* and *lacO3*, which showed approximate symmetry, showed lower affinities in the following descending order: *sym-lacO*> *lacO1*> *lacO2*> *lacO3* (27). In the native lac operon, LacI simultaneously binds to the primary operator *lacO1* as well as *lacO2* or *lacO3* via a DNA loop mechanism (28). *LacO2* is 401 bp downstream of *lacO1* and *lacO3* reads 92 bp upstream from *lacO1* (29). Because of their proximity, the DNA loop occurs mainly between *lacO1* and *lacO3*, and therefore these sites represent the major gene repression (27). As a result, the role of *lacO2* remains unclear. When LacI binds *lacO1* and *lacO3*, it also inhibits its own production because the 3' end of the *lacI* gene overlaps with *lacO3*. In the repressed state, *lacI* transcription leads to a truncated mRNA, which is rapidly degraded by the cell. Due to this autoregulation, the frequency of the LacI tetramer is ~ 0 molecules per cell in induced cells and ~15 molecules per cell in non-induced cells (30). The removal of *lacO3* to prevent autoregulation is not possible because the sequence is part of the protein structure and would therefore change the LacI protein. If the LacI concentration is to be increased to increase the tightness of the expression system, it can be achieved by introducing a mutated *lacI* gene, called $LacI^{Q}$ (31). As a result, the level of the LacI molecules rises to around 100 molecules per cell (30, 32)

1.2.2 Expression rate fine-tuning

Controlling the transcription rate of the expression system, also known as "tunability", serves to regulate protein production to a certain level. The optimal bioprocess is designed so that the maximum cell synthesis capacity can be maintained for a long time. This is particularly necessary if the protein of interest (POI) is to be produced as a soluble and correctly folded protein. In many cases, recombinant proteins can also have a toxic effect on the host if certain levels are reached and exceeded. Depending on the physical properties and metabolic requirements of the desired product, the transcription rate must match RNA stability, translation efficiency, protein folding, protein transport and all other system interactions.

Transcription rate tuning through titration with IPTG is not possible. IPTG, like lactose, is actively transported into the cell by the Lac permease LacY (33). Even small amounts of

IPTG or lactose induce the production of further LacY permease, which allows even more IPTG to penetrate the cell (34). The expression of Lac Permease LacY is heterogeneous and the number of active permeases in each cell of a population is highly variable. The final product titer can be varied due to the limited availability of the inducer but only if the entire population is considered. There is no single cell tunebaility, instead there are cells that are fully induced and which are weakly or not induced within a population (35). This is also known as the all-or-none induction phenomenon (36). With toxic or challenging proteins, this can lead to massive product loss, since the fully induced cells of this population can be outcompeted by faster growing non-producers (37). In the case of a two-stage, orthogonal expression system, such as the T7 expression system already described, this behavior is particularly evident. Small amounts of IPTG induce the production of T7 RNAP, which in turn initiates the transcription of the GOI. Once the T7 RNAP has been produced, the cell does not immediately degrade the T7 RNAP and the transcription of the GOI is initiated again.

To prevent this process and thus make a T7 system tunable, T7 lysozyme is co-produced with the commercially available strain Lemo21(DE3) (38). T7 lysozyme is produced in this expression system under the control of the tunable rhaPBAD promoter. T7 lysozyme binds to the T7 RNAP and thus prevents transcription initiation from the T7 promoter (39). By increasing the inducer L-rhamnose more T7 lysozyme is produced and thus less active T7 RNAP is available in the cell and the production of the POI decreases.

The Tuner™ (DE3) strain from Novagen takes a different approach. Tuner™(DE3) is a BL21 (DE3) derivative that has a mutation within the lac operon. As a result, the lac permease LacY is no longer produced, an induction cascade is prevented and the penetration of IPTG occurs evenly in every cell of the population. This means that a homogeneous level of the POI can be set in each cell using inducer titration (40).

1.3 Expression system: Plasmid-based versus genomically integrated

In biotechnology, plasmids are the most common vectors for recombinant DNA. Plasmids are small extrachromosomal, self-replicating DNA molecules within a cell. For a plasmid to be used as a vector for the GOI, it needs some basic properties. On the one hand, a multiple cloning site, which has a variety of restriction sites, so that the GOI can be cloned into the vector (41). Furthermore, selection markers, in most cases antibiotic resistance genes, so that positively transformed cells can be identified (42) and an origin of replication so that the vector can replicate independently of the chromosomal DNA (43). Commercially available vectors, such as the pET series, have a pMB1 ori, which is a ColE1 derivative with

approximately 15-60 copies per cell. A mutated variant of the pMB1 origin occurs in the pUC series (500-700 copies per cell) (44). One should now assume that a high copy number of the GOI also results in more recombinant protein. In the case of simple proteins to be produced, this is actually the case and recombinant protein yields of up to 50% of the total cell protein are possible (45, 46). However, proteins that are difficult to produce cause a high metabolic load, which limits cell growth, and the plasmid can be lost or mutate. This leads to the already mentioned phenomenon of the faster growing non-producing population, causing the entire production process to come to a standstill. (47, 48).

Another possibility is to integrate the GOI directly into the host's chromosome. By means of homologous recombination, a linear DNA fragment, which contains all required information (promoter / operator, GOI, terminator, resistance gene), is integrated into a non-coding part of the host genome (49). In comparison to the plasmid transformation, this process is labor-intensive and the transformation rates are very low, which means that more effort is required for the screening of positively transformed cells. The advantage of such a system is that host cells are freed from plasmid-related metabolic load and high gene dosage. In the case of a T7 based expression system, the high efficiency of the T7 RNAP compensates for the low gene dose and provides high expression rates of the GOI without extreme consequences for the host metabolism. This hardly affects the growth rate after induction of recombinant protein production, and cell growth and protein production can thus be maintained throughout the entire process. This consequently leads to higher biomasses and thus to higher product titers compared to plasmid-based expression systems. Another advantage of the reduced metabolic load is the quality of the protein produced. The capacities for protein translocation and folding are less utilized, which means that higher yields of the soluble and correctly folded protein can be obtained (50).

In summary, the production of a recombinant protein results in innumerable combinations of production strain, promoter system, type of gene regulation, resistance gene, the localization of the GOI, either genome integrated or plasmidic, all of which have an influence on quality and quantity. At this point it must be mentioned that even if an optimal combination of all these factors has been determined for a certain class of protein, these findings can only be transferred to another class of recombinant proteins to a limited extent. This requires a new screening and is still largely a trial and error approach. Therefore, a large number of screening experiments are required, which cannot be carried out in fully equipped lab-scale bioreactors.

1.4 Cultivation techniques of *E. coli* to produce recombinant proteins

1.4.1 Cultivation in Microbioreactors

Screenings of potentially new production strains are still mainly carried out in shake flasks or microtiter plates. As a result, numerous cultivations can be carried out in parallel, quickly, easily and, above all, inexpensively (51). However, process parameters such as dissolved oxygen (DO), pH and substrate uptake rate can only be monitored and controlled to a limited extent in such systems. In addition, cells grow in complex medium with excess glucose to guarantee rapid growth. This can lead to clones being selected from shake flask experiments, but perform poorly in an industrially relevant process in a bioreactor (52). Therefore, efficient clone screenings for the identification of possible *E. coli* production hosts should take place in a C-limited environment. As already described in the section “Promoters”, a lac-based expression system, for example, can only be completely induced in the absence or with limited availability of glucose. To facilitate scale-up, basic process parameters such as DO and pH should be monitored and, if necessary, regulated to a certain value. All of these conditions must nevertheless be able to be implemented with a high degree of parallelization and with the shortest possible set-up times (53).

There are now numerous manufacturers and devices for microbioreactors. RTS-8 plus bioreactor from Biosan (Biosan Medical-Biological Research & Technologies, Riga, Letvia) uses the patented Reverse-Spin® technology, in which a non-invasive, mechanically driven, innovative energy movement mixes the cell suspension in single-use Falcon tubes. The Ambr® 15 system from Sartorius takes a different approach. The working volume of the microbioreactors is 10-15 ml and have an integrated stirrer. They also contain a port for adding liquids and taking samples. With a sparge tube, aeration is very efficient, which means that this system comes very close to full-size bioreactors. The microbioreactor that was used in this work is the BioLector from m2p labs (m2p-labs GmbH, Baesweiler, Germany. The BioLector is a bench-top device, which consists of a temperature and humidity-controlled incubation chamber with an orbital shaker. It utilizes standard-sized microtiter plates, which enable automation and linking to other devices. An optical light cable underneath the plate enables the continuous monitoring of scattered light and fluorescence intensities of a bacterial culture and thus bacterial growth. Theoretically, biomass concentrations of up to 50 g/L can be measured by means of scattered light without the need to dilute the cell suspension (54). In addition, the protein formation of fluorescent proteins, such as GFP, can be measured directly by measuring fluorescence signals (55). Using immobilized fluorescent dyes, so-called optdodes, which are attached to the bottom of the microtiter plate, pH values and dissolved oxygen (DO) levels can be measured. In addition to

measuring important process parameters to characterize the cultivation, it is also necessary to simulate the cultivation strategies that are used in industrial production processes, as far as is technically possible. A very elegant strategy that enables carbon limitation without the use of external substrate feed is the enzymatic release of glucose from a dissolved polysaccharide. The activity of the enzyme, a glucoamylase, determines the glucose release and thus the growth rate of the cells (56).

1.4.2 Fed-batch cultivation

The biotechnological cultivation of bacteria in fed-batch mode is the most common method to produce recombinant proteins in *E. coli*. The substrate, usually the carbon source glucose, is fed to the reactor without the cell broth being removed from the reactor. This means that all products remain in the bioreactor until the end of cultivation (57). The advantage of this process is the control over the supply of the nutrients and thus the control over the growth rate and other process relevant factors. This process control enables very high cell densities of approximately 175 g cell dry weight per liter (3). Such high cell densities would only be possible in batch culture if the initial nutrient concentration were correspondingly high. However, an excessively high glucose concentration has an inhibiting effect on cell growth. If the glucose concentration is too high, the so-called crabtree effect occurs (58). Initially observed in the cultivation of yeasts, where too high a glucose concentration led to high ethanol production, the effect was also observed in *E. coli* and *Bacillus subtillis*, with the latter organisms increasingly producing organic acids such as acetic acid. Regardless of whether ethanol or organic acids, both have an inhibiting effect on cell growth and thus prevent a controlled bioprocess. The decisive factor for the production of recombinant proteins, however, is the catabolite repression, as already described in the section "Promoter".

To minimize all these effects there are various strategies for carrying out fed-batch cultivations. Usually a highly concentrated feed solution is used to prevent excessive dilution of the bioreactor. In principle, the feed solution can be fed to the reactor in two ways. From a technical point of view, the simplest way is to feed the feed solution constantly. The feed rate remains unchanged during the entire process and is not adapted to the increasing number of cells in the bioreactor. This results in a steadily decreasing cell growth rate. This process is well described and has advantages for recombinant protein production (59, 60). However, since cells grow exponentially under ideal conditions, the supply of the carbon source must also take place as such if a constant growth rate of the cells is to be achieved. In this way, very high cell densities are achieved in a short time. The exponential feed rate should be

chosen so that the cells grow below their maximum growth rate, so that the above-mentioned effects such as catabolite repression and crabtree effect do not occur and the cells are carbon-limited despite the exponential growth.

1.4.3 Chemostat cultivation

Regardless of whether the feed medium is supplied at a constant or exponential feed rate, the process is ended as soon as a certain biomass concentration or filling volume of the reactor has been reached. The cell broth is then processed further, and the bioreactor is cleaned, sterilized and prepared for the next fed batch. In terms of time, the actual fermentation and especially the production phase of the recombinant protein in such a batch-wise procedure is relatively short. That is why continuous production is becoming more and more interesting, since it enables the greatest possible space time yield to be achieved with the optimal use of the installed assets (61). In such chemostat cultivations, the cells are kept in a steady-state environment by supplying fresh feed medium to the reactor and simultaneously removing the cell suspension to the same extent (62). Technically speaking, such a process is quite simple and was first described in the 1950s (63). The growth rate can be set similar to that of fed-batch cultivation, the feed rate or dilution rate. The advantages of this process control are obvious; the volume of the reactor always remains constant, the average residence time of a producing cell and the product is always the same, which means that stable volumetric productivity and thus high space time yields can be achieved (64).

Regarding quality control, continuous production presents some difficulties. Compared to batch process control, where product pools can be subjected to quality control clearly separated from each other (65), this is not possible with continuous process control. One approach would be to split a continuous process into different batch numbers, which artificially creates single batch processes (66). Nevertheless, due to the steadily increasing interest in continuous processes and the increasingly advanced process technology, the regulatory authorities have also developed further. The American Food and Drug Administration (FDA) has not only established guidelines for the continuous production of recombinant proteins but is also encouraging the production of biopharmaceutically relevant proteins using a continuous manufacturing approach.

For this reason, continuous processes to produce recombinant proteins have already been implemented, but only for processes which use mammalian cell cultures. *E. coli* processes in continuous process control have still not been implemented due to process instability.

1.5 Mutations and Escape Variants Caused by Metabolic Burden

As already described, the production of recombinant proteins exhibits a metabolic load on the host organism. Especially if the POI is difficult to be produced or shows toxic effects on cell physiology. When *E. coli* is cultured under laboratory conditions, the mutation rate is 1×10^{-3} per generation (67). However, the mutation rate is highly stress-dependent (68). In cells that no longer produce the recombinant protein due to plasmid loss or a mutation, the metabolic load is reduced. Non-producing cells therefore have a growth advantage. As a result, they overgrow producing cells, which reduces the overall product yield (48, 69). Because of this, longer production times or even continuous production is in *E. coli* difficult to achieve.

To avoid these obstacles, it is therefore necessary to reduce the metabolic load and thus the genetic escape (70). This is possible on the one hand at the genetic as well as on the bioprocess level. At the bioprocess level, the metabolic load can be reduced by decoupling growth and production. By cascading two reactors, only biomass is generated in the first bioreactor without the addition of the inducer. The recombinant protein production then takes place in the second reactor, which contains the inducer. This arrangement ensures that the system is always supplied with fresh, uninduced cells, thus avoiding the aforementioned phenomena such as plasmid loss and genetic instability caused by recombinant protein production (71).

At the cellular level, host genes, which are responsible for genetic instability, can be knocked out. The removal rate of insertion elements (IS), deletion of the recombinase gene recA or error prone DNA polymerases has already reduced the mutation rate and thus increased the stability of the host organism (72). Interventions in the genome of the host organism in order to change individual genes may seem meaningful from the point of view of a biotechnologist, but such changes can affect proteins and metabolic processes, which in turn can have a negative influence on the fitness and growth of the cell. This rational approach, where genes and proteins are influenced in a targeted manner, is often a trial on error approach and is therefore time consuming and very often unsuccessful. Furthermore, this approach depends on the POI to be produced and cannot always be transferred directly to other classes of POIs.

A completely different approach is directed evolution. It is a method that imitates the process of natural selection and directs the genome and thus the proteome of a host organism in a certain direction (73). A population of cells is exposed to mutagenesis in repeated cycles, resulting in a variety of genetic variants within a population. Then, those individuals with the desired properties are selected and propagated, which creates a new population with new

properties. This can be done in vivo with living organisms or in vitro. A directed evolution approach can circumvent the complexity of the biogenesis process and its adverse effects on the host cell (74-76). One of the best-known examples of such an approach is the experiment by John E. Walker and Bruno Miroux (74). In a selection-based approach, BL21(DE3) cells expressing a toxic membrane protein were streaked onto agar plates containing the IPTG inducer. Those cells that survived the recombinant protein production and thus adapted to the production of toxic membrane proteins were selected and characterized more precisely. *E. coli* C41(DE3) is a derivative from this experiment, which shows a mutation in the lacUV5 promoter, whereby the expression of the T7 RNAP was reduced. *E. coli* C43(DE3), a derivative that originated from another direct evolution round from C41, has a mutation in the LacI gene, which reduces induction by IPTG, which in turn reduces the expression of the GOI. C41(DE3) and C43(DE3), commonly known as the Walker strains, are currently widely used to produce membrane and toxic proteins (76, 77)

Nevertheless, these strains use a plasmid-based expression system and are subject to the aforementioned obstacle to plasmid loss and would therefore not be suitable for long-term cultivation. Another disadvantage is the use of the T7 expression system. In a study by Striedner et al. long-term stability of genome-integrated T7-based expression systems has been investigated. Chemostat cultivations with a dilution rate of 0.1 /h showed a decrease in productivity after about 14 generations (50). The reason for this was a single point mutation within the T7 RNAP, which resulted in a complete loss of productivity.

In order to obtain a fundamental understanding of the interaction between the host, the recombinant protein and the long-term response of the recombinant gene expression, it is necessary to investigate a genome-integrated expression system that is not dependent on the T7 RNAP and T7 promoter. An expression system based on the host RNAP could be suitable for this approach since a mutation in the host RNAP can be excluded because the cell needs it to maintain the metabolism.

2 Objective and definition of the research topic

The aim of this thesis is to further develop and characterize genome-integrated expression systems and to investigate the long-term response of recombinant gene expression on the mutation patterns in *E. coli*. Currently, only the strong IPTG inducible T7 promoter with different model proteins has been tested. The use of alternative, inducible promoter systems specific for the *E. coli* host RNAP in a genome integrated expression system enables the potential to reduce the stress level that is triggered by the production of recombinant proteins and thus to enable long-term cultivation. Furthermore, a mutation in the host RNAP can be excluded, as this would not provide the cell a growth advantage. This would be the first time that the long-term response to recombinant gene expression could be examined at the genome level.

The scientific question is therefore, are genome-integrated expression systems that depend on the host RNAP suitable to characterize system stability and host cell response to long-term production and to identify mutation patterns and mutational hot spots?

To answer the scientific question / hypothesis, the objectives defined for this thesis are:

- Replace the T7 promoter with the strong IPTG inducible host RNAP specific promoters $T5_{N25}$ and $T7_{A1}$.
- Characterization of both promoters in terms of productivity, basal expression and tunability.
- Investigation of long-term stability under production conditions by adding IPTG in repeated cycles of cultivation in the BioLector under carbon-limited production conditions.
- Detailed investigation of the underlying mutations in different clones using whole genome sequencing.
- Cytosolic GFP serves as an easy-to-produce model protein while the periplasmic Fab fragment serves as a challenging model protein.
- Confirmation of the results from the BioLector experiments in bench-top scale chemostat cultivations under conditions similar to production processes.

The results will allow a detailed investigation to what extent and where recombinant protein production triggers mutations in the host organism. Mutation hot spots could be examined in this way and, in the best case, mutations could be identified which have a positive effect on the GOI expression in continuous bioprocesses.

3 Results and discussion

The research work has been published in two scientific articles.

Publication I

Schuller A., Cserjan-Puschmann M., Tauer C., Jarmer J., Wagenknecht M., Reinisch D., Grabherr D., Striedner G., **Escherichia coli σ^{70} promoters allow expression rate control at the cellular level in genome-integrated expression systems**. Microbial Cell Factories, Volume 19, Issue 1, March 2020, Page 58

In this study the protein expression potential was investigated of two modified phage-derived promoters, T5 and A1, both of which are recognized by the sigma^{70} *E. coli* RNAP was investigated. The promoter sequences were modified so that they contain either one, two or three *lacO* sites. As a result, seven promoter/operator combinations were designed that control the model protein GFPmut3.1. The BL21 (DE3) T7 expression system served as a reference.

The T5 and A1 promoter operator combinations were cultivated under fed-batch like conditions in the micro-titer BioLector device. The recombinant protein expression was induced by adding IPTG.

In this study, we were able to show that in genome integrated expression systems, the regulatory elements of the lac operon must be well balanced to regulate sigma^{70} promoters. The number and position of the *lacO* sites have a direct influence on the promoter strength and basal expression.

The promoter strength of the genome-integrated host RNAP dependent expression systems were determined relative to the genome integrated T7 expression system. With the A1 expression system, the relative promoter strength was approx. 30% of the T7 system regardless of whether one or two *lacO* sites regulate the promoter. The relative promoter strength of the T5 promoter was 20% if the promoter was controlled by one or two *lacO* sites and only about 10% if it was regulated by three *lacO* sites. The reason for the low promoter strength when using three *lacO* sites can be seen in the perfectly symmetrical *lacO* directly in the initially transcribed sequence (ITS) of a promoter (78).

Promoters which only contain one *lacO* site showed a particularly high promoter strength, but also a higher basal expression. In these expression systems, the *lacO* site was placed directly within the promoter sequence between the -10 and the -35 sequence (27). If LacI binds to the *lacO* in the non-induced state, the RNAP can no longer bind and thus prevents transcription. However, we were able to show that in genome-integrated expression systems

that only have the native level of LacI concentration (~ 10 molecules / cell (30)), the repression was not sufficient, despite the absence of the inducer.

In promoter / operator combinations that contained two *lacO* sites showed a very high binding affinity for LacI, whereby the basal expression could be significantly reduced without having a negative influence on the strength of the promoter. However, these expression systems exhibited a few hours after induction a complete stop of recombinant protein production, when cells were partially induced. To explain this unusual phenomenon, we hypothesized that the complete stop of productivity in partially induced cells might be due to LacI autoregulation:

If the binding constant of the LacI to the *lacO* sites of the GOI is greater than to the native lac operon, the first LacI molecules that were not inactivated by IPTG will preferentially bind to the GOI rather than to *lacO1* / *lacO3* of the lac operon. As a result, LacI autoregulation cannot not take place and more and more LacI molecules are produced until there is a complete stop in production of the POI.

We were able to support this hypothesis by comparing the effect of LacI autoregulation using the example of a two-*lacO* regulated expression system and a wild-type BL21 without GOI. Using western blot analysis, we estimated the LacI content of non-induced, partially-induced and fully induced cells and we were able to show that the LacI concentration in an expression system regulated by two *lacO* sites was significantly higher than in the native BL21 wild-type strain. This indicated that LacI autoregulation could not take place.

This negative effect was avoided by combining the positive properties of a one-*lacO* regulated expression system with an increased number of LacI molecules, thus reducing the basal expression. For this, the promoter sequence of the native *lacI* promoter was replaced by that of the $lacI^Q$ promoter, which resulted in an increase in the LacI concentration by a factor of 10. This new genome-integrated expression system BL21Q <1lacOA1>, showed high expression rates, very low basal expression and additional tunability at the cellular level, which makes it perfectly suitable for the production of challenging proteins, as it does not contain plasmid based high metabolic load and is free of phage related components.

This expression system was submitted for a patent together with the company partner Boehringer Ingelheim RCV GmbH & Co KG.

Publication II

Schuller A., Cserjan-Puschmann M., Köppl C., Grabherr R., Wagenknecht M., Schiavinato M., Dohm J. C., Himmelbauer H., Striedner G., **Adaptive evolution in producing microtiter cultivations generates genetically stable Escherichia coli production hosts for continuous bioprocessing.** Biotechnology Journal, 2020

The research interest of Publication II was to investigate the long-term response of the expression system characterized in Publication I.

In this study, we examined the genetic stability of *E. coli* expression systems under long-term production conditions. The scientific question was: How metabolic load, triggered by recombinant protein production influences the mutation characteristics of *E. coli*?

To answer this question, we have carried out repetitive fed-batch like microbioreactor cultivations under production conditions. The easy-to-produce protein GFP and the challenging protein antigen-binding fragment (Fab) served as model proteins. As production strains, the conventional BL21(DE3) strain with the T7 expression system was compared with the BL21Q<1lacOA1> strain described in Publication I.

In comparative whole genome sequencing analyzes, we were able to identify mutations that allowed the cells to grow unhindered, despite recombinant protein production. The host RNA polymerase-dependent BL21Q expression system remained genetically stable during the production of the easy-to-produce protein GFP over a period of 45 generations. During the production of the challenging protein Fab, we were able to identify mutations in the *lacI* gene that had a positive impact on long-term stability. These mutations were located in LacI's inducer binding pocket, which in turn affects the binding affinity of IPTG. This means that despite a sufficient amount of the inducer IPTG, the cells can no longer be completely induced and some of the LacI molecules can always bind to the *lacO* sites and thus prevent transcription of the GOI. As a result, the cells tuned themselves to maximum tolerable productivity. Those adapted cells were able to maintain a high growth rate despite recombinant protein production in chemostat cultivations, and mutated non-producing cells no longer had any growth advantage and therefore could not overgrow the producing cells. This led to very high process stability, which enabled the adapted cells to produce the challenging protein Fab in a chemostat culture over a period of 17 generations.

However, the high process stability could only be achieved by reducing specific productivity. Mutants that had shown process stability over several generations only produced about half of the recombinant protein compared to the non-mutated parent strain when it was cultivated in the fed batch. This could be compensated by continuous production in chemostat. The space time yield (mg/L/h) of the mutant was 43% higher than that of the parent in fed-batch

cultivation, since production lasted for 132 hours continuously. The necessary downtimes such as cleaning and sterilization and reactor setup, as is necessary with fed-batch cultivations, are significantly reduced in chemostat cultivations, which means that the bioreactor can be used more efficiently.

As already expected, the genetic stability of the T7-based BL21 (DE3) expression system was not sufficient to maintain constant product levels in long-term cultivation. Mutations that led to non-producers were localized either in the T7 RNAP gene and / or in the T7 promoter. This could be shown for the easy-to-produce protein GFP as well as for the challenging protein Fab.

4 Summary and Conclusions

In summary, it can be stated that the objectives of this doctoral thesis were met:

- A host-RNAP-dependent genome-integrated expression system was developed, which has a sufficiently high promoter strength, low basal expression and tunability on a cellular level.
- Long-term stability studies under production conditions could be carried out. In microscale BioLector cultivations as well as in small-scale chemostat cultivations.
- Using whole genome sequencing, detailed investigations of the underlying mutations were determined.
- We were able to show that a T7 expression system is neither suitable for the production of challenging proteins nor for long-term stability studies, since mutations in the T7 RNAP or in the T7 promoter always lead to non-producers.
- With the BL21Q <1lacOA1> expression system, we were able to identify mutations that have a positive impact on long-term stability.

We were able to show that in the applied adaptive evolution setup mutations that increase long-term stability always come along with a reduction in the productivity of the POI to the physiologically tolerable level. We suspect that the growth rate of producing cells is high enough that they can no longer be overgrown in a chemostat culture by mutated non-producers.

To fully exploit the potential of the genetically stable *E. coli* mutants, it would be necessary to couple production and cell viability. This has already been successfully implemented in so-called product-addicted expression systems. These systems use essential gene circuits, which are only expressed under production conditions. The cells only express the essential genes when the desired product is produced which means that viability is linked to productivity. In this way, a directed evolution approach would be possible, whereby mutants could arise which adapt to a high expression of the recombinant protein. However, this has only been successful in a few cases and is very much dependent on the desired product (79).

A more straightforward approach would be to optimize cultivation conditions. So far, the mutants described in this work have only been tested under fixed conditions for stability and productivity. A variation of the dilution rate, the production temperature and the media composition could increase the yield of the recombinant protein production and would offer further potential for successful continuous production (80-82). These optimizations could be

used for both expression systems T7 and host RNAP-depend, since the conventional genome-integrated T7 expression system also provides a very high space time yields for easy-to-produce proteins over several generations. Other approaches such as complex 2-stage cascade reactor systems could thus be avoided.

Since the interest in continuous protein production in *E. coli* is great, it is only a question of time before the advantages of fed-batch cultivation (high specific productivity), with the advantages of chemostat cultivation (high space-time yield) are combined. With this work we were able to contribute to a better understanding of the underlying mutation patterns and provided a strong basis for further research.

5 Table of abbreviations

DO	Dissolved oxygen
Fab	Antigen-binding fragment
FDA	Food and Drug Administration
GFP	Green fluorescent protein
GOI	Gene of interest
IPTG	Isopropyl β-d-1-thiogalactopyranoside
IST	Initially transcribed sequence
mRNA	messenger ribonucleic acid
POI	Protein of interest
RNAP	Ribonucleic acid polymerase

6 References

1. Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F, et al. Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. Science. 1977;198(4321):1056-63.

2. Plank LD, Harvey JD. Generation time statistics of Escherichia coli B measured by synchronous culture techniques. J Gen Microbiol. 1979;115(1):69-77.

3. Markl H, Zenneck C, Dubach AC, Ogbonna JC. Cultivation of Escherichia coli to high cell densities in a dialysis reactor. Applied microbiology and biotechnology. 1993;39(1):48-52.

4. Pope B, Kent HM. High efficiency 5 min transformation of Escherichia coli. Nucleic Acids Res. 1996;24(3):536-7.

5. Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of Escherichia coli K-12. Science. 1997;277(5331):1453-62.

6. Durfee T, Nelson R, Baldwin S, Plunkett G, Burland V, Mau B, et al. The Complete Genome Sequence of *Escherichia coli* DH10B: Insights into the Biology of a Laboratory Workhorse. Journal of Bacteriology. 2008;190(7):2597.

7. Marisch K, Bayer K, Scharl T, Mairhofer J, Krempl PM, Hummel K, et al. A comparative analysis of industrial Escherichia coli K-12 and B strains in high-glucose batch cultivations on process-, transcriptome- and proteome level. PLoS One. 2013;8(8):e70516.

8. Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol. 1986;189(1):113-30.

9. Gottesman S. Proteases and their targets in *Escherichia coli*. Annual Review of Genetics. 1996;30(1):465-506.

10. Grodberg J, Dunn JJ. ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. Journal of bacteriology. 1988;170(3):1245-53.

11. Sain B, Murray NE. The hsd (host specificity) genes of E. coli K12. Molecular and General Genetics MGG. 1980;180(1):35-46.

12. Eckstein F. LDMJ. Nucleic Acids and Molecular Biology 4: Springer-Verlag Berlin Heidelberg; 1990. XI, 291 p.

13. Oliphant AR, Struhl K. Defining the consensus sequences of E.coli promoter elements by random selection. Nucleic acids research. 1988;16(15):7673-83.

14. Lanzer M, Bujard H. Promoters largely determine the efficiency of repressor action. Proc Natl Acad Sci U S A. 1988;85(23):8973-7.

15. Kang JG, Hahn MY, Ishihama A, Roe JH. Identification of sigma factors for growth phase-related promoter selectivity of RNA polymerases from Streptomyces coelicolor A3(2). Nucleic acids research. 1997;25(13):2566-73.

16. Gruber TM, Gross CA. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. Annual review of microbiology. 2003;57(1):441-66.

17. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology. 1961;3(3):318-56.

18. Wanner BL, Kodaira R, Neidhardt FC. Regulation of lac operon expression: reappraisal of the theory of catabolite repression. Journal of bacteriology. 1978;136(3):947-54.

19. Pribnow D. Bacteriophage T7 early promoters: Nucleotide sequences of two RNA polymerase binding sites. Journal of Molecular Biology. 1975;99(3):419-43.

20. de Boer HA, Comstock LJ, Vasser M. The tac promoter: a functional hybrid derived from the trp and lac promoters. P Natl Acad Sci USA. 1983;80(1):21-5.

21. Martin CT, Esposito EA, Theis K, Gong P. Structure and Function in Promoter Escape by T7 RNA Polymerase. Progress in Nucleic Acid Research and Molecular Biology. 80: Academic Press; 2005. p. 323-47.

22. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Applied microbiology and biotechnology. 2006;72(2):211-22.

23. Briand L, Marcion G, Kriznik A, Heydel JM, Artur Y, Garrido C, et al. A self-inducible heterologous protein expression system in Escherichia coli. Scientific reports. 2016;6(1):33037.

24. Barkley MD, Riggs AD, Jobe A, Burgeois S. Interaction of effecting ligands with lac repressor and repressor-operator complex. Biochemistry. 1975;14(8):1700-12.

25. Oehler S, Eismann ER, Kramer H, Muller-Hill B. The three operators of the lac operon cooperate in repression. EMBO J. 1990;9(4):973-9.

26. Sadler JR, Sasmor H, Betz JL. A perfectly symmetric lac operator binds the lac repressor very tightly. Proc Natl Acad Sci U S A. 1983;80(22):6785-9.

27. Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, Müller-Hill B. Quality and position of the three lac operators of E. coli define efficiency of repression. EMBO J. 1994;13(14):3348-55.

28. Mossing MC, Record MT, Jr. Upstream operators enhance repression of the lac promoter. Science. 1986;233(4766):889-92.

29. Reznikoff WS, Winter RB, Hurley CK. The location of the repressor binding sites in the lac operon. Proc Natl Acad Sci U S A. 1974;71(6):2314-8.

30. Semsey S, Jauffred L, Csiszovszki Z, Erdossy J, Steger V, Hansen S, et al. The effect of LacI autoregulation on the performance of the lactose utilization system in Escherichia coli. Nucleic Acids Res. 2013;41(13):6381-90.

31. Glascock CB, Weickert MJ. Using chromosomal lacIQ1 to control expression of genes on high-copy-number plasmids in Escherichia coli. Gene. 1998;223(1-2):221-31.

32. Muller-Hill B, Crapo L, Gilbert W. Mutants that make more lac repressor. Proc Natl Acad Sci U S A. 1968;59(4):1259-64.

33. Fernandez-Castane A, Vine CE, Caminal G, Lopez-Santin J. Evidencing the role of lactose permease in IPTG uptake by Escherichia coli in fed-batch high cell density cultures. Journal of biotechnology. 2012;157(3):391-8.

34. Hansen LH, Knudsen S, Sørensen SJ. The Effect of the lacY Gene on the Induction of IPTG Inducible Promoters, Studied in Escherichia coli and Pseudomonas fluorescens. Current Microbiology. 1998;36(6):341-7.

35. Siegele DA, Hu JC. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. Proceedings of the National Academy of Sciences. 1997;94(15):8168-72.

36. Afroz T, Biliouris K, Kaznessis Y, Beisel CL. Bacterial sugar utilization gives rise to distinct single-cell behaviours. Mol Microbiol. 2014;93(6):1093-103.

37. Doherty AJ, Connolly BA, Worrall AF. Overproduction of the toxic protein, bovine pancreatic DNasel, in Escherichia coli using a tightly controlled T7-promoter-based vector. Gene. 1993;136(1):337-40.

38. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, et al. Tuning Escherichia coli for membrane protein overexpression. Proc Natl Acad Sci U S A. 2008;105(38):14371-6.

39. Stano NM, Patel SS. T7 lysozyme represses T7 RNA polymerase transcription by destabilizing the open complex during initiation. The Journal of biological chemistry. 2004;279(16):16136-43.

40. Khlebnikov A, Keasling JD. Effect of lacYExpression on Homogeneity of Induction from the Ptac and Ptrc Promoters by Natural and Synthetic Inducers. Biotechnology Progress. 2002;18(3):672-4.

41. Thomson JM, Parrott WA. pMECA: A Cloning Plasmid with 44 Unique Restriction Sites that Allows Selection of Recombinants Based on Colony Size. BioTechniques. 1998;24(6):922-8.

42. Andrews CA, Lesley SA. Selection Strategy for Site-Directed Mutagenesis Based on Altered β-Lactamase Specificity. BioTechniques. 1998;24(6):972-80.

43. Sakai H, Komano T. DNA Replication of IncQ Broad-host-range Plasmids in Gram-negative Bacteria. Bioscience, Biotechnology, and Biochemistry. 1996;60(3):377-82.

44. Minton NP. Improved plasmid vectors for the isolation of translational lac gene fusions. Gene. 1984;31(1):269-73.

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

46. Graumann K, Premstaller A. Manufacturing of recombinant therapeutic proteins in microbial systems. Biotechnology Journal. 2006;1(2):164-86.

47. Bentley WE, Mirjalili N, Andersen DC, Davis RH, Kompala DS, William E. Bentley Ib. Plasmid-encoded protein: The principal factor in the “metabolic burden” associated with recombinant bacteria. Biotechnology and bioengineering. 2009;102(5):1283-97.

48. Glick BR. Metabolic load and heterologous gene expression. Biotechnology advances. 1995;13(2):247-61.

49. Sharan SK, Thomason LC, Kuznetsov SG, Court DL. Recombineering: a homologous recombination-based method of genetic engineering. Nat Protoc. 2009;4(2):206-23.

50. Striedner G, Pfaffenzeller I, Markus L, Nemecek S, Grabherr R, Bayer K. Plasmid-free T7-based Escherichia coli expression systems. Biotechnology and bioengineering. 2010;105(4):786-94.

51. Vincentelli Rea. High-throughput protein expression screening and purification in Escherichia coli. Methods. 2011;55:65-72.

52. Šiurkus J, Panula-Perälä J, Horn U, Kraft M, Rimšeliene R, Neubauer P. Novel approach of high cell density recombinant bioprocess development: Optimisation and scale-up from microlitre to pilot scales while maintaining the fed-batch cultivation mode of E. coli cultures. Microbial cell factories. 2010;9(1):35.

53. Bareither R, Pollard D. A review of advanced small-scale parallel bioreactor technology for accelerated process development: Current state and future need. Biotechnology Progress. 2011;27(1):2-14.

54. Samorski M, Müller-Newen G, Büchs J. Quasi-continuous combined scattered light and fluorescence measurements: A novel measurement technique for shaken microtiter plates. Biotechnology and bioengineering. 2005;92(1):61-8.

55. Kensy F, Zang E, Faulhammer C, Tan RK, Buchs J. Validation of a high-throughput fermentation system based on online monitoring of biomass and fluorescence in continuously shaken microtiter plates. Microbial cell factories. 2009;8:31.

56. Hemmerich J, Wenk P, Lütkepohl T, Kensy F. Fed-batch cultivation in baffled shake flasks. Genetic Engineering and Biotechnology News. 2011;31(14).

57. Yamanè T. SS. Fed-batch techniques in microbial processes. In: Bioprocess Parameter Control. Advances in Biochemical Engineering/Biotechnology. 1984;30.

58. Crabtree HG. Observations on the carbohydrate metabolism of tumours. Biochemical Journal. 1929;23(3):536-45.

59. Norsyahida A, Rahmah N, Ahmad RMY. Effects of feeding and induction strategy on the production of BmR1 antigen in recombinant E. coli. Letters in Applied Microbiology. 2009;49(5):544-50.

60. Nakagawa S, Oda H, Anazawa H. High Cell Density Cultivation and High Recombinant Protein Production of Escherichia coli Strain Expressing Uricase. Bioscience, Biotechnology, and Biochemistry. 1995;59(12):2263-7.

61. Walther J, Godawat R, Hwang C, Abe Y, Sinclair A, Konstantinov K. The business impact of an integrated continuous biomanufacturing platform for recombinant protein production. Journal of biotechnology. 2015;213:3-12.

62. Ziv N, Brandt NJ, Gresham D. The Use of Chemostats in Microbial Systems Biology. JoVE. 2013(80):e50168.

63. Novick A, Szilard L. Description of the Chemostat. Science. 1950;112(2920):715-6.

64. Peebo K, Neubauer P. Application of Continuous Culture Methods to Recombinant Protein Production in Microorganisms. Microorganisms. 2018;6(3).

65. Lee SL, O'Connor TF, Yang X, Cruz CN, Chatterjee S, Madurawe RD, et al. Modernizing Pharmaceutical Manufacturing: from Batch to Continuous Production. Journal of Pharmaceutical Innovation. 2015;10(3):191-9.

66. Jungbauer A. Continuous downstream processing of biopharmaceuticals. Trends in biotechnology. 2013;31(8):479-92.

67. Lee H, Popodi E, Tang H, Foster PL. Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing. Proc Natl Acad Sci U S A. 2012;109(41):E2774-83.

68. Maharjan R, Ferenci T. Mutational signatures indicative of environmental stress in bacteria. Mol Biol Evol. 2015;32(2):380-91.

69. De Gelder L, Ponciano JM, Joyce P, Top EM. Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. Microbiology. 2007;153(2):452-63.

70. Rugbjerg P, Sommer MOA. Overcoming genetic heterogeneity in industrial fermentations. Nat Biotechnol. 2019;37(8):869-76.

71. Schmideder A, Weuster-Botz D. High-performance recombinant protein production with Escherichia coli in continuously operated cascades of stirred-tank reactors. J Ind Microbiol Biotechnol. 2017;44(7):1021-9.

72. Csorgo B, Feher T, Timar E, Blattner FR, Posfai G. Low-mutation-rate, reduced-genome Escherichia coli: an improved host for faithful maintenance of engineered genetic constructs. Microbial cell factories. 2012;11:11.

73. Lutz S. Beyond directed evolution—semi-rational protein engineering and design. Current Opinion in Biotechnology. 2010;21(6):734-43.

74. Miroux B, Walker JE. Over-production of Proteins inEscherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. Journal of Molecular Biology. 1996;260(3):289-98.

75. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, et al. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature. 2009;461(7268):1243-7.

76. Kwon SK, Kim SK, Lee DH, Kim JF. Comparative genomics and experimental evolution of Escherichia coli BL21(DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction. Scientific reports. 2015;5:16076.

77. Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). Protein Expression and Purification. 2004;37(1):203-6.

78. Hsu LM, Cobb IM, Ozmore JR, Khoo M, Nahm G, Xia L, et al. Initial transcribed sequence mutations specifically affect promoter escape properties. Biochemistry. 2006;45(29):8841-54.

79. Rugbjerg P, Sarup-Lytzen K, Nagy M, Sommer MOA. Synthetic addiction extends the productive life time of engineered Escherichia coli populations. Proceedings of the National Academy of Sciences. 2018;115(10):2347.

80. Turner C, Gregory ME, Turner MK. A study of the effect of specific growth rate and acetate on recombinant protein production of Escherichia coli JM107. Biotechnology Letters. 1994;16(9):891-6.

81. Sandén AM, Prytz I, Tubulekas I, Förberg C, Le H, Hektor A, et al. Limiting factors in Escherichia coli fed-batch production of recombinant proteins. Biotechnology and bioengineering. 2003;81(2):158-66.

82. Gutiérrez-González M, Farías C, Tello S, Pérez-Etcheverry D, Romero A, Zúñiga R, et al. Optimization of culture conditions for the expression of three different insoluble proteins in Escherichia coli. Scientific reports. 2019;9(1):16850.

7 List of Publications

I Schuller A., Cserjan-Puschmann M., Tauer C., Jarmer J., Wagenknecht M., Reinisch D., Grabherr D., Striedner G., ***Escherichia coli* σ^{70} promoters allow expression rate control at the cellular level in genome-integrated expression systems**. Microbial Cell Factories, Volume 19, Issue 1, March 2020, Page 58

II Schuller A., Cserjan-Puschmann M., Köppl C., Grabherr R., Wagenknecht M., Schiavinato M., Dohm J. C., Himmelbauer H., Striedner G., **Adaptive evolution in producing microtiter cultivations generates genetically stable *Escherichia coli* production hosts for continuous bioprocessing.** Biotechnology Journal, 2020

Contribution to the publications:

In publication I, Artur Schuller performed the silico design of the promoter/operator combination, the subsequent cloning into the vectors and the integration of the expression cassettes into the genome of *E. coli.* He carried out the small-scale screenings in the BioLector and developed an in-house method for determining the LacI concentration per cell and developed the hypothesis of LacI autoregulation in genome-integrated expressions systems and wrote the manuscript.

In publication II, Artur Schuller generated through repeated cycles of small-scale fed-batch cultivations, mutants that were able to grow unhindered despite recombinant protein production. He analyzed the results of the whole genome sequencing and carried out chemostat cultivation with selected mutants. Finally, Artur Schuller wrote the manuscript.

Publication I

Microbial Cell Factories

RESEARCH Open Access

Escherichia coli σ^{70} promoters allow expression rate control at the cellular level in genome-integrated expression systems

Artur Schuller[1], Monika Cserjan-Puschmann[1*], Christopher Tauer[1], Johanna Jarmer[2], Martin Wagenknecht[2], Daniela Reinisch[2], Reingard Grabherr[1] and Gerald Striedner[1]

Abstract

Background: The genome-integrated T7 expression system offers significant advantages, in terms of productivity and product quality, even when expressing the gene of interest (GOI) from a single copy. Compared to plasmid-based expression systems, this system does not incur a plasmid-mediated metabolic load, and it does not vary the dosage of the GOI during the production process. However, long-term production with T7 expression system leads to a rapidly growing non-producing population, because the T7 RNA polymerase (RNAP) is prone to mutations. The present study aimed to investigate whether two σ^{70} promoters, which were recognized by the *Escherichia coli* host RNAP, might be suitable in genome-integrated expression systems. We applied a promoter engineering strategy that allowed control of expressing the model protein, GFP, by introducing *lac* operators (*lacO*) into the constitutive T5 and A1 promoter sequences.

Results: We showed that, in genome-integrated *E. coli* expression systems that used σ^{70} promoters, the number of *lacO* sites must be well balanced. Promoters containing three and two *lacO* sites exhibited low basal expression, but resulted in a complete stop in recombinant protein production in partially induced cultures. In contrast, expression systems regulated by a single *lacO* site and the *lac* repressor element, $lacI^{Q}$, on the same chromosome caused very low basal expression, were highly efficient in recombinant protein production, and enables fine-tuning of gene expression levels on a cellular level.

Conclusions: Based on our results, we hypothesized that this phenomenon was associated with the autoregulation of the *lac* repressor protein, LacI. We reasoned that the affinity of LacI for the *lacO* sites of the GOI must be lower than the affinity of LacI to the *lacO* sites of the endogenous *lac* operon; otherwise, LacI autoregulation could not take place, and the lack of LacI autoregulation would lead to a disturbance in *lac* repressor-mediated regulation of transcription. By exploiting the mechanism of LacI autoregulation, we created a novel *E. coli* expression system for use in recombinant protein production, synthetic biology, and metabolic engineering applications.

Keywords: Recombinant protein expression, *Escherichia coli*, LacI autoregulation, Tunable expression, σ^{70} promoters, Genome-integrated expression systems

Background

In industrial recombinant protein production processes, regulation of the gene of interest (GOI) is an important prerequisite. Transcription rates are controlled by the interaction between a promoter and the RNA polymerase (RNAP). This interaction must be understood

*Correspondence: monika.cserjan@boku.ac.at
[1] Christian Doppler Laboratory for Production of Next-level Biopharmaceuticals in E. coli, Department of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria
Full list of author information is available at the end of the article

and externally regulated to provide process control, and thereby, the optimization of product yield and quality. In particular, challenging proteins of interest, like antibody fragments, membrane proteins, or toxic proteins, require low basal expression in non-induced states and a reduced transcriptional activity after recombinant protein induction [1–3]. The final yield of challenging proteins is not only directly determined by the strength of the promoter system but also by further processing steps, such as translation, folding, translocation into the periplasm, and proper disulfide bond formation. The most prominent and well-studied genetic regulatory mechanism is the *lac* operon of *Escherichia coli* [4]. In wild-type *E. coli*, the *lac* repressor protein (LacI) evolved to sense the presence of lactose. In the absence of lactose, LacI forms a homo-tetramer that binds to the *lac* operator site (*lacO*) and represses the transcription of the *lacZYA* operon [5]. Conversely, when lactose or isopropyl β-D-1-thiogalactopyranoside (IPTG, a non-metabolizable structural mimic of allolactose) binds to LacI, it induces a conformational change in the protein structure, and LacI can no longer bind to *lacO* site. This leaves the *lacO* site open to RNAP binding, and thus, transcription can start. The *lacO* sites are DNA sequences with an inverted repeat symmetry [6]. The higher the symmetry, the greater the LacI binding affinity of the operator sequence. An artificial, perfectly symmetric *lacO* (sym-*lacO*) was found to bind LacI with the greatest affinity [7]. In contrast, three wild-type operators *lacO1*, *lacO2*, and *lacO3*, which exhibited approximate symmetry, showed lower affinities, in the following descending order: sym-*lacO* > *lacO1* > *lacO2* > *lacO3* [8]. LacI binds simultaneously to both the primary operator, *lacO1*, and to either *lacO2* or *lacO3* through a DNA-looping mechanism [9]. *LacO2* is located 401 bp downstream of *lacO1*, and *lacO3* lies 92 bp upstream of *lacO1* [10]. Due to their close proximity, the DNA-looping mainly occurs between *lacO1* and *lacO3*, and thus, these sites provide the main gene repression [8]. Consequently, the role of *lacO2* remains unclear. Furthermore, when LacI binds *lacO1* and *lacO3*, it inhibits its own production, because the 3′ end of the *lacI* gene overlaps with *lacO3*. In the repressed state, *lacI* transcription results in a truncated mRNA, which is rapidly degraded by the cell. Due to this autoregulation, the abundance of the LacI tetramer is ~ 40 molecules per cell in induced cells and ~ 15 molecules per cell in non-induced cells [11].

One application of the *lac* regulatory mechanism is known as the pET system, which is currently the most widely used *E. coli* expression system for recombinant protein production [12, 13]. The pET system is based on the specific interaction between the phage-derived, T7-specific RNAP and the strong T7 promoter for the GOI. The recombinase functions of bacteriophage lambda were used for site-directed insertion of the T7 RNAP gene into the *E. coli* chromosome. Expression of the T7 RNAP is controlled by the lacUV5 promoter, a variant of the *lac* promoter that is insensitive to catabolic repression [14]. The addition of IPTG induces the expression of the T7 RNAP at high levels, which in turn, transcribes the target gene under the control of the T7 promoter [13]. This orthogonal expression system offers very high product titers for recombinant proteins that, consequently, can be efficiently produced in *E. coli*. However, the extraordinary strength of the T7 expression system, particularly when combined with high-copy-number plasmids, exerts an extreme metabolic load on the host cells. When the GOI codes for a challenging protein, the stress and metabolic burden often lead to reduced yield, shortened production periods, and even cell death [15, 16].

Plasmid-mediated stress, due to high gene dosages and the expression of antibiotic resistance genes, can be overcome by integrating the GOI into the host chromosome [17, 18]. The high efficiency of the T7 RNAP compensates for low gene dosages and provides high rates of recombinant gene expression [15]. Nevertheless, the high expression rates also cause stress to the cell, which results in reduced growth rates. In a previous study [18] we showed that during continuous production, the genome-integrated T7 expression system became instable approximately 70 h past induction. The reason for this could be found in a mutated T7 RNAP, which led to a faster growing non-producing population [unpublished data]. These non-producing cells grew more rapidly and prevailed over the producing population; this resulted in a massive loss in product yield.

We expected that expression systems that are coupled to the host metabolism would have increased genetic stability, because transcription relies on constitutive phage-derived promoters that are recognized by the σ^{70} *E. coli* RNAP, rather than relying on transcription machinery that is orthogonal to *E. coli*. The pQE vectors from Qiagen (Hilden, Germany) provide two *lacO* sites that control the $T5_{N25}$ promoter. The pJexpress 401-406 (T5) vectors from ATUM (Newark, NJ, USA) contain two wild-type *lacO* sites and one symmetric *lacO* site to avoid basal expression. The *E. coli* pAVEway™ expression system from Fujifilm Diosynth Biotechnologies (Hillerød, Denmark) employs two symmetrical *lacO* sites to control the expression of the $T7_{A3}$ promoter. However, all these expression systems are plasmid-based, and thus, they are subject to the obstacles mentioned above, like high gene copy number, plasmid replication, and process instability caused by plasmid loss.

The present study aimed to generate inducible promoters that were recognized by the σ^{70} *E. coli* RNAP and were originally derived from two constitutive phage promoters, T5 ($T5_{N25}$) [19–21] and A1 ($T7_{A1}$) [22]. We aimed to investigate their potential transcription efficiency, basal expression rate, and transcription rate control, in genome-integrated expression systems. For transcription rate control, we introduced one [21], two [23], or three *lacO* sites [7], into the promoter sequences. We integrated these into *E. coli* strains with wild-type *lacI* and $lacI^Q$ promoters. The $lacI^Q$ promoter is a variant with a single C → T change within the −35 promoter motif. This mutation causes a tenfold increase in LacI expression [24]. The resulting promoter/operator combinations were investigated to determine expression strength, tunability, and basal expression of the cytosolic model protein, GFPmut3.1 [25]. We also evaluated cell growth in plasmid-based and genome-integrated *E. coli* BL21 expression systems. We reasoned that the addition of *lac* operators on the chromosome in the genome-integrated expression systems might influence the endogenous *lac* operon activity. Therefore, we also measured the LacI levels in selected strains. The production clones were compared in micro-titer fermentations, under fed-batch-like conditions [26], over a production period of 12 h, and they were benchmarked with the T7 RNAP-dependent T7 promoter expression systems.

Results and discussion

In this study, we investigated the protein expression potential of two modified phage-derived promoters, T5 and A1, which were recognized by the σ^{70} *E. coli* RNAP. The promoter sequences were modified to contain one, two, or three *lacO* sites. We created seven promoter/operator constructs combined with the open reading frame of the model protein, GFPmut3.1 (Fig. 1).

Productivity of σ^{70} dependent promoter/operator combinations

The T7 expression system is known to provide high expression rates, even from a single target gene copy, when integrated into the *E. coli* chromosome. First, we wanted to check whether the same productivity could be reached with σ^{70} *E. coli* RNAP-dependent promoters in the same experimental set-up. Therefore, we compared the genome-integrated (indicated with pointed

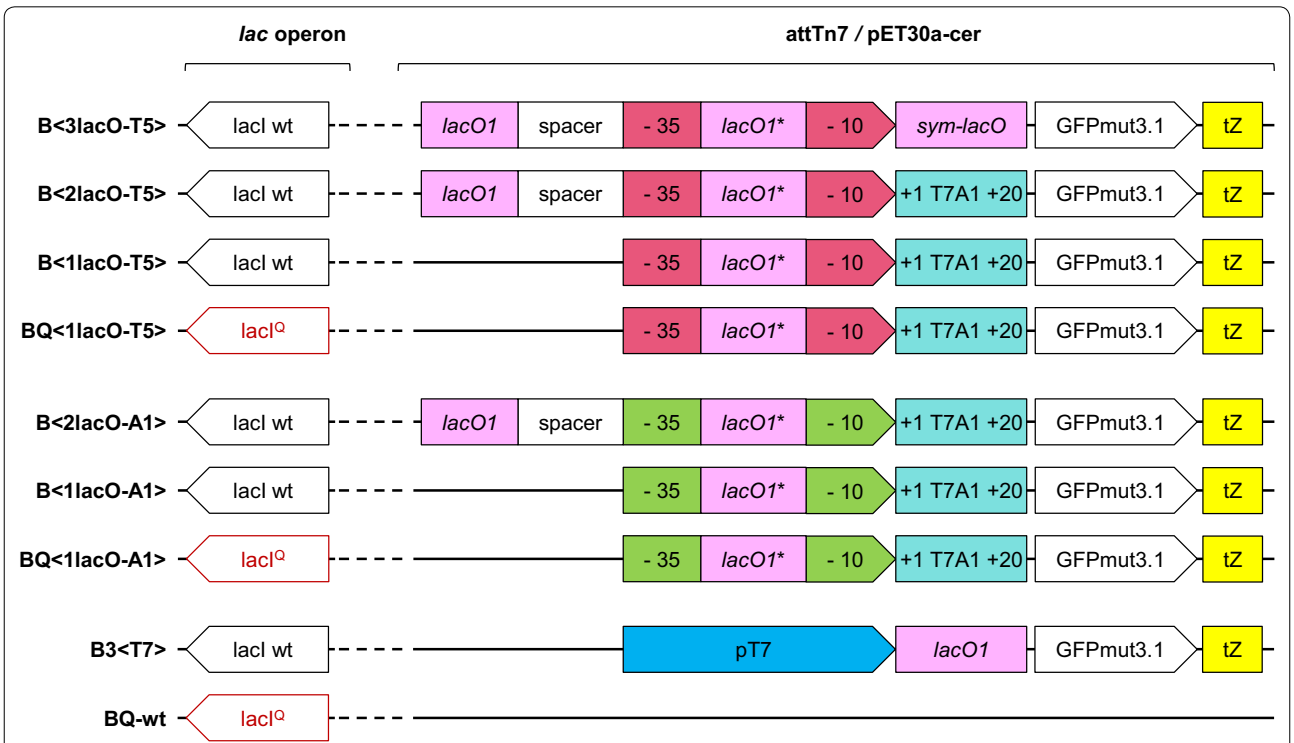


Fig. 1 Schematic of GFPmut3.1 expression cartridges controlled by seven different promoter/operator combinations. The cartridges were integrated into the attTN7 site (indicated with<pointed brackets>) of the *E. coli* BL21 chromosome, or they were cloned into the pET30a-cer vector (indicated with round brackets (), but not shown in this figure). In two promoter/operator combinations, the wild-type *lacI* promoter (B, black) was exchanged with the $lacI^Q$ promoter (BQ, red). *LacO1** is a 2-bp truncated version of wild-type *lacO1*. Sym-*lacO* is the perfectly symmetric *lacO*. The native, initially transcribed sequence of the A1 promoter, is labeled +1 T7A1 +20. Transcription is terminated by tZENIT (tZ) [27]. The BL21(DE3) T7 expression system (B3<T7>) is used as a reference. The BQ-wt carried the wild-type sequence, with the $lacI^Q$ promoter

brackets: <>) and plasmid-based (indicated with round brackets) T5 and A1 promoter/operator combination expression systems to the T7 expression system. The cells were grown in fed-batch-like conditions, in micro-titer fermentations, over a period of 22 h. Expression of GFPmut3.1 was induced with 0.5 mM IPTG after 10 h.

In all promoter/operator combinations, the cells maintained growth in the micro-titer fermentations. The average growth rate was $\mu = 0.05$/h, during the 12-h production period. We directly compared average growth rates between the T7 and the σ^{70} promoters (Additional file 1: Figure S1, Additional file 2: Figure S2).

On-line fluorescence measurements of the plasmid-based expression systems (Fig. 2b) showed that all promoter/operator combinations, except B(3lacO-T5), expressed comparable amounts of GFPmut3.1. In contrast, with the genome-integrated expression systems (Fig. 2a), we observed quite distinct differences between the different promoter/operator combinations. The A1 expression systems produced 1.5-fold GFPmut3.1 yields compared to the T5 expression systems. These results were consistent with previously published data [20, 21, 28]. In the genome-integrated T7 expression system, induction of GFPmut3.1 expression led to 145 rfu and a specific soluble GFPmut3.1 concentration of ~ 135 mg/g cell dry matter (CDM). The same experiment with the A1 expression systems yielded almost 50 rfu and a GFPmut3.1 concentration of 37 mg/g CDM. A comparison of protein solubility in the plasmid-based and genome-integrated systems indicated that a large proportion of insoluble GFPmut3.1 was produced in the plasmid-based expression systems. Conversely, over 90% of the recombinant protein was soluble in the genome-integrated expression systems (Additional file 3: Figure S3).

The reduced productivity observed with the plasmid-based B(3lacO-T5) and the genome-integrated B<3lacO-T5> might have been due to the presence of the perfectly symmetric *lac* operator (sym-*lacO*) [7], which replaced the initially transcribed sequence (ITS). This symmetric *lacO* could influence promoter escape, and therefore, productivity [29]. This effect was less evident with the plasmid-based 3(lacO-T5) expression system, where the high plasmid copy number compensated for the reduced promoter activity. However, in the genome-integrated expression system, the promoter activity was quite low; therefore, we discarded the 3*lacO* version with the A1 promoter. For the one and two *lacO* promoter/operator combinations, we replaced sym-*lacO* with the native ITS of the A1 promoter (+1 *T7A1* +20). This resulted in a 1.4-fold increase in productivity, in the case of the T5 promoter.

Basal expression in σ^{70} dependent expression systems

For challenging proteins, even low basal expression can have adverse effects on the host metabolism, or it may even be toxic to the host cell. Hence, in those cases, equipping the host with an expression construct, either plasmid-based or genome-integrated, can be rather difficult. This difficulty is typically represented by the low frequency of transformants or integrants, respectively. Thus, the tightness of gene regulation is an important quality criterion for expression systems.

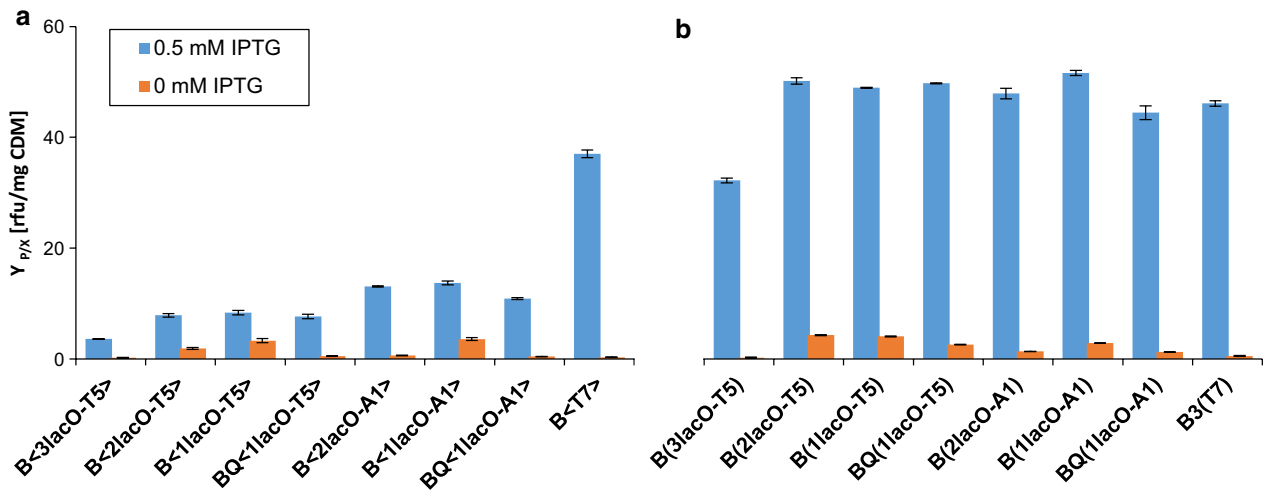


Fig. 2 Promoter activities of different promoter/operator combinations, under non-induced (0 mM IPTG) and induced (0.5 mM IPTG) conditions. The specific fluorescence of the reporter protein, GFPmut3.1 ($Y_{P/X}$), is given in relative fluorescence units per mg of cell dry matter [rfu/mg CDM]. This value was used to characterize **a** genome-integrated expression systems and **b** plasmid-based expression systems. Error bars indicate standard error of the mean (n = 3). Expression system names are defined in Fig. 1

In the plasmid-based systems, promoters that were controlled by one *lac* operator (1*lacO*) showed the highest basal expression, at a level of ~ 4 rfu/mg CDM, particularly under carbon-limited conditions (Fig. 2b). The addition of a second *lacO* (2*lacO*) or an increase in LacI production, by introducing the $lacI^Q$ promoter, reduced the basal expression of the A1 promoter to 1 rfu/mg CDM. In constructs with the T5 promoter, only the inclusion of three *lac* operators (3*lacO*) reduced the basal expression to almost 0 rfu/mg CDM. In contrast to the plasmid-based expression systems, all genome-integrated systems showed that the promoter/operator combination significantly impacted the system leakiness (Fig. 2a). Both an increase in the number of LacI molecules and the addition of a second *lacO* site reduced the basal expression of A1 expression systems from 4 rfu/mg CDM to nearly no significant background expression. Importantly, productivity was not affected. Although both promoters contained *lacO* sites at the identical position, only an increased level of LacI molecules or three *lacO* sites could sufficiently reduce basal expression in the T5 expression systems. Similar findings were obtained by Lanzer and Bujard [21]. They concluded that the promoter strength was not correlated with effective repression. The host RNAP recognized the A1 promoter only half as efficiently as the T5 promoter [28]. When one *lacO* site was located within the promoter sequence, between the −10 and −35 promoter elements, the host RNAP and LacI competed with each other for their respective binding sites, and this competition determined how efficiently promoter activity was controlled by the repressor. The RNAP and T5 promoter form a complex at one of the highest complex-formation rates known in nature [28]. Thus, controlling this promoter requires either a high repressor binding affinity in the operators or a high concentration of repressor molecules.

Control of recombinant gene expression rate

The control of the transcription rate, also referred to as "tunability", is used to fine-tune protein production. This fine-tuning is highly relevant in bioprocessing. Optimal bioprocesses are designed to maximally exploit cell synthesizing capacities for long periods to yield correctly folded, processed proteins. Depending on the physical properties and metabolic requirements of the desired product, transcription rates must be adapted to RNA stability, translation efficiency, protein folding, protein transport, and all other interactions in the system.

To evaluate the tunability of the promoter/operator combinations described herein, we tested a series of fed-batch-like microtiter cultivations at varying IPTG levels and benchmarked protein production to the genome-integrated T7 expression system. The range of IPTG concentrations for fully and partially induction with IPTG was determined in a preliminary experiment. The strains B<3lacO-T5> and B3<T7> were induced with following IPTG concentrations: 1.0, 0.5, 0.1, 0.05, 0.01, 0.005 mM IPTG (Additional file 4: Figure S4). Based on these results, we decided on the concentrations 0.005, 0.01 and 0.5 mM IPTG. On-line fluorescence measurements and end-point flow cytometry analyses were used to characterize the different promoter/operator combinations.

Expression systems controlled by one *lacO* site for gene regulation exhibited the highest basal expression and the least pronounced gradation of GFPmut3.1 expression at increasing inducer concentrations (Fig. 3c, f). Although promoters controlled by two *lacO* sites showed sufficiently low basal expression, they also produced less protein at the lower inducer concentrations (Fig. 3b, e). The promoter/operator combinations controlled by 3lacO-T5 and 2lacO-A1 led to a complete stop (plateau) of recombinant GFPmut3.1 production after a certain time, independent of the inducer concentration (Fig. 3a, e). We did not observe this behavior in promoter/operator combinations with only one *lacO* site. The combination of promoters controlled by one *lacO* site and $lacI^Q$ repressor (Fig. 3d, g) and the T7 expression system (Fig. 3h) resulted in the desired system properties, including tunability and low system leakiness.

T7 expression systems exhibit an all-or-none induction phenomenon, where reduced expression in partially induced cultures results from the formation of subpopulations of fully induced and non-induced cells [30]. Therefore, we investigated transcription rate tuning at the cellular level with flow cytometry analyses of all genome-integrated promoter/operator combinations (Fig. 4). We confirmed that the all-or-none phenomenon occurred in genome-integrated T7 expression systems. In fact, we observed a mixture of fully, partially, and non-induced cells, particularly at very low inducer concentrations (Fig. 4h, red line). In the B<2lacO-A1> expression system, flow cytometry analyses revealed that these expression systems stopped GFPmut3.1 production, although the cells continued to grow (Additional file 1: Figure S1, Additional file 2: Figure S2). This result indicated that there were two distinct subpopulations of producing and non-producing cells. We also observed this behavior in B<3lacO-T5> (Fig. 3a, e). But the BQ<1lacO-A1> system showed different behavior. There, the induction of the *gfpmut3.1* gene resulted in a homogenous population at any given IPTG concentration (Fig. 3g). Consequently, this expression system provided proof that the expression rate was controlled on a cellular level.

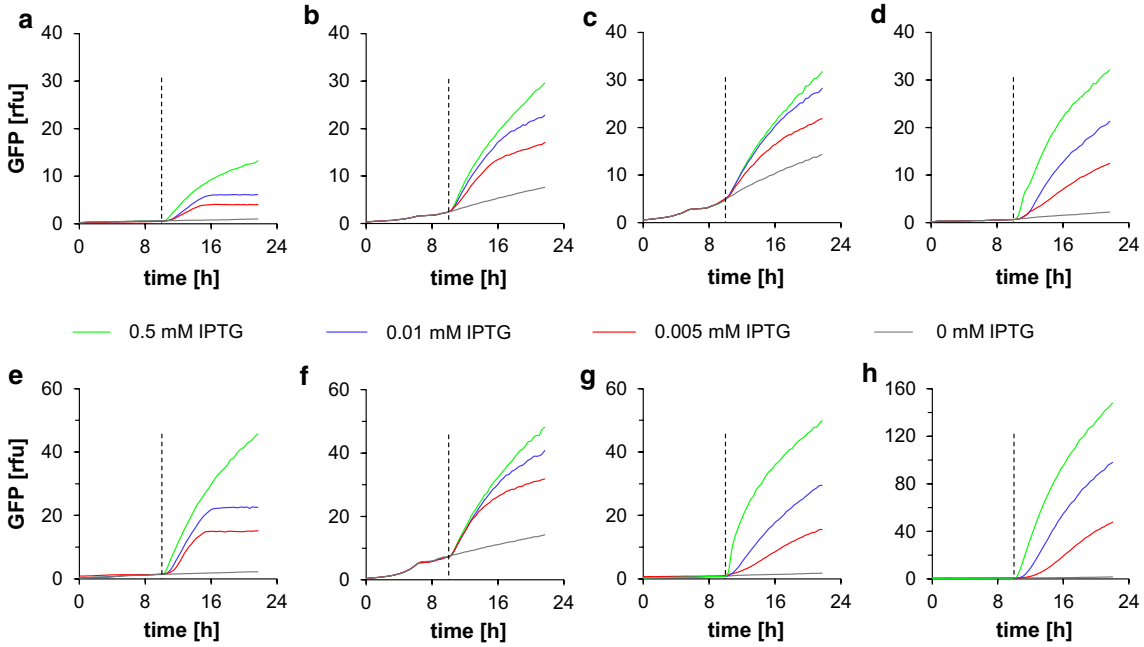


Fig. 3 Influence of *lac* operators on expression rate control, shown by the change in on-line GFPmut3.1 fluorescence in fed-batch-like microtiter cultivations. The dashed vertical lines indicate the time of induction. Induction was performed with 0 (gray, not induced), 0.005 (red), 0.01 (blue), or 0.5 mM (green) IPTG. A–D: The T5 promoter is controlled by: **a** three *lacO*, **b** two *lacO*, **c** one *lacO*, and **d** one *lacO*/*lacI*^{Q} sequences. **e**–**g** The A1 promoter is controlled by **e** two *lacO*, **f** one *lacO*, and **g** one *lacO*/*lacI*^{Q} sequences. **h** The T7 expression system is used as a reference. The Y-axis scale is adjusted to the respective expression rates. The mean relative GFP fluorescence intensity (rfu) represents triplicate samples

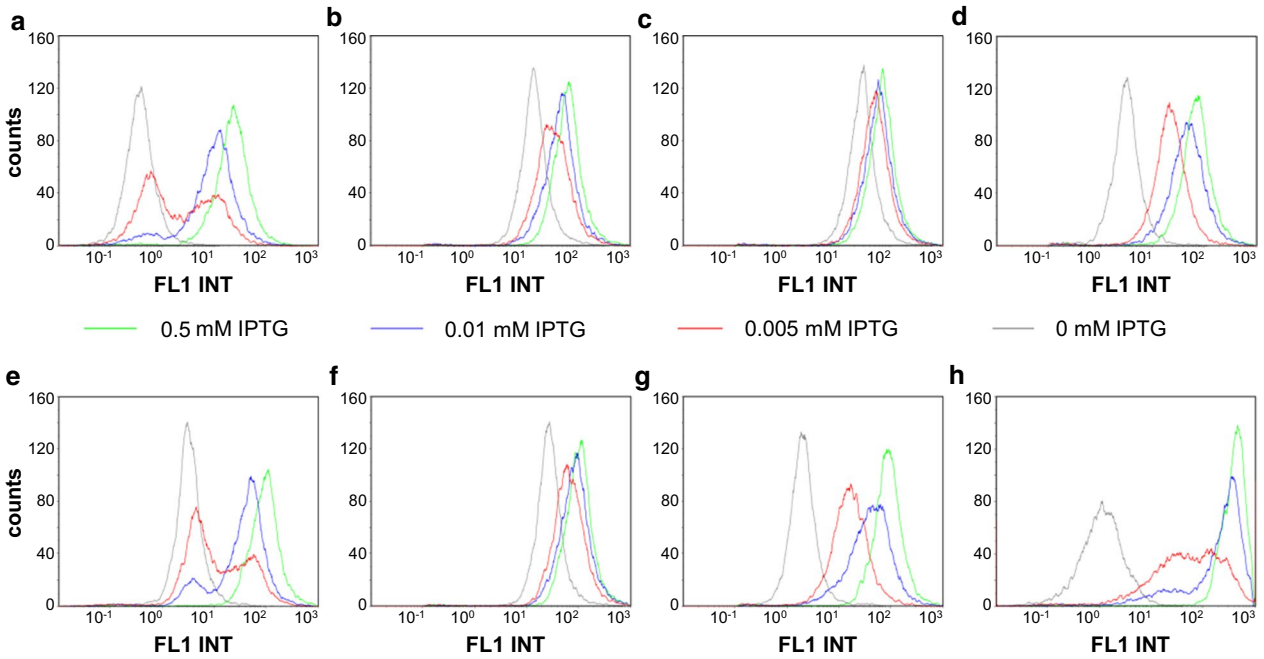


Fig. 4 Flow cytometry analysis of single-cell expression of GFPmut3.1. Induction was performed with 0 (gray, not induced), 0.005 (red), 0.01 (blue), or 0.5 mM (green) IPTG. A–D: T5 promoter controlled by: **a** three *lacO*, **b** two *lacO*, **c** one *lacO*, or **d** one *lacO*/*lacI*^{Q} sequences. **e**–**g** A1 promoter controlled by: **e** two *lacO*, **f** one *lacO*, or **g** one *lacO*/*lacI*^{Q} sequences. **h** The T7 expression system is used as a reference

Influence of LacI autoregulation on expression rate control

We assumed that the complete stop in productivity, observed when the B<3laco-T5> and B<2lacO-A1> systems were partially induced, was associated with the autoregulation of the *lac* repressor. The native *lac* operon is regulated by three *lacO* sites (Fig. 5a). The LacI molecule simultaneously binds to two sites, either *lacO1* and *lacO3* or *lacO1* and *lacO2* [6]. The *lacO3* sequence overlaps with the 3'end of the *lacI* gene. When LacI binds to *lacO1* and *lacO3*, it causes the DNA to form a loop. This results in truncated *lacI* mRNA molecules, which are degraded by the cell. This autoregulation of LacI production resulted in a constant level of ~ 10 LacI molecules per cell in the absence of an inducer [11, 31, 32].

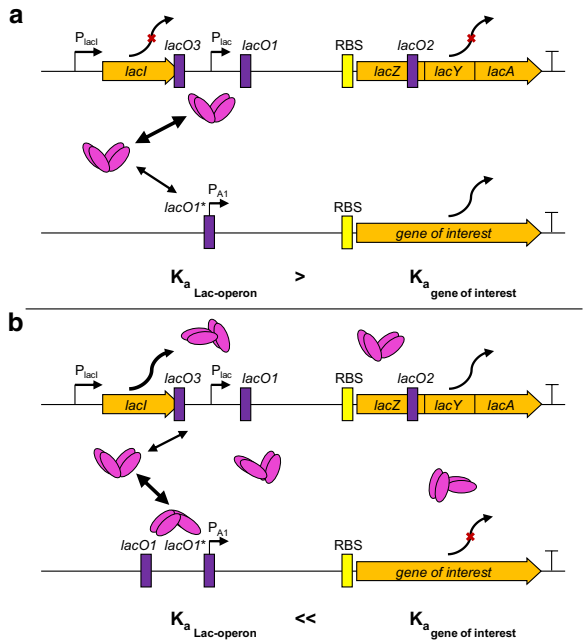


Fig. 5 Schematic of *lac* operators in the native *lac* operon (top panels) and its regulation of the gene of interest (bottom panels). LacI (pink tetramer) production effects are shown, when the promoter for the gene of interest was regulated by **a** one *lac* operator or **b** two *lac* operators, respectively. K_a = association constant; red cross = stopped production

We hypothesized that, when the binding constant (K_a) of LacI to the *lacO* sites of the GOI was greater than the binding constant to the *lacO* sites of the *lac* operon, the first LacI molecules, which are not inactivated by IPTG, will preferentially bind to the *lacO* site of the GOI, instead of the *lacO3*/*lacO1* within the *lac* operon. Hence, autoregulation of LacI would not intervene, and LacI molecules would continue to be produced. This would cause the whole system to become overregulated, which would result in a complete stop in production (Fig. 5b).

To test this hypothesis, we compared the effect of autoregulation on LacI in B<2lacO-A1> and BL21 wild-type cells (BL21-wt). We estimated the LacI content of non-induced, partially-induced, and fully-induced cells with western blot analyses. The band intensities were quantified and normalized by the cell number (Fig. 6).

In fully induced (0.5 mM IPTG) BL21-wt cells, the number of LacI molecules was 3.3-fold greater than the number observed in non-induced BL21-wt cells. Partial induction with 0.01 mM IPTG only led to a 0.7-fold increase. The 3.3-fold change in fully induced BL21-wt cells was consistent with previous results from Semsey et al. In that study, they measured an average of 15 LacI molecules per cell in the absence of inducer and ~ 40 molecules per cell in fully induced cells [11].

In B<2lacO-A1> cells, LacI numbers in non-induced and partially induced conditions were clearly higher than the numbers observed in uninduced BL21-wt cells. LacI yields were 2.4-fold greater in the absence of inducer and 3.2-fold greater in partially induced cells, relative

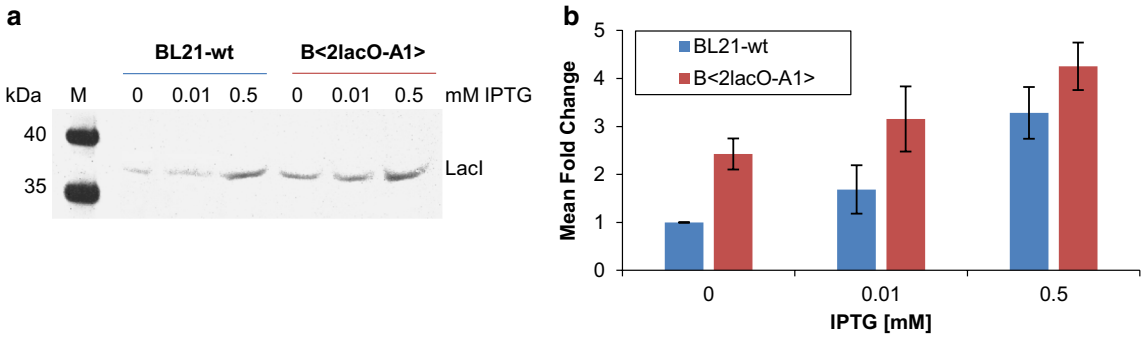


Fig. 6 Influence of additional *lacO* sites on cellular LacI concentrations. Proteins of ~ 1.2×10^7 cells were separated with SDS-PAGE and analyzed on western blots, probed with an anti-LacI antibody. **a** Western blot of BL21 wild-type cells and B<2lacO-A1> cells, which were grown without IPTG, 0.01 mM IPTG, or 0.5 mM IPTG. (M) PageRuler™ Plus Prestained Protein Ladder. **b** Fold changes of band intensities determined in panel **a** are shown relative to the levels observed in 0 mM IPTG BL21-wt cells. Error bars indicate the standard error of the mean (n = 3)

to uninduced BL21-wt cells. In fully induced cells, LacI yields were 4.3-fold greater than those observed in uninduced BL21-wt cells, which was similar to the yield in fully induced BL21-wt cells.

Although the addition of 0.01 mM IPTG resulted in almost half-maximal GFPmut3.1 expression in B<2lacO-A1> cells (Fig. 3e), it had little or no influence on the LacI levels. This suggested that LacI continued to bind to *lacO1*/*lacO3* in the lac operon; hence, it could maintain autoregulation under these conditions. In the fully induced state, the LacI concentrations are almost the same with a fourfold increase regardless of whether it is the BL21-wt or the B<2lacO-A1> expression system. LacI therefore no longer binds to its operators and thus the expression of LacI itself is no longer inhibited. The small fold change of 4 results from the weak constitutive LacI promoter, which provides about one new mRNA per cell generation [33]. Thus, the high LacI levels in non-induced and partially induced B<2lacO-A1> cells clearly supported our hypothesis that LacI autoregulation impacted the expression rate control in genome-integrated *E. coli* production strains (Fig. 5).

The effect of LacI autoregulation was only observed in genome-integrated, host RNAP-dependent expression systems, which were controlled by two or three *lacO* sites. In contrast, this effect was not observed in plasmid-based, host RNAP-dependent expression systems or in the conventional T7 expression system. This discrepancy might be explained by differences in the balance between *lacO* sites and LacI concentrations. The T7 expression system harbors a second *lacI* gene sequence within its DE3 lysogen, which would, theoretically, double the LacI concentration per cell. The plasmid-based expression systems used in this study were based on the pET plasmid system, which encodes a second *lacI* gene sequence. In turn, depending on the plasmid copy number, that resulted in an extra 15–20 *lacI* gene sequences [34]. However, the effect of LacI autoregulation on partially induced cells was also observed in plasmid-based expression systems, like the *E. coli* pAVEway™ expression system, from Fujifilm Diosynth Biotechnologies (Hillerød, Denmark). In the pAVEway™ expression system, transcription control was enabled by two perfectly symmetric *lac* operators, one positioned upstream and one downstream of the T7A3 promoter. The high affinity of LacI to the symmetric *lacO* sites, combined with the ability to form a DNA loop, resulted in very low basal expression, but also, a complete stop in productivity in partially induced cultures.

Considering the autoregulation of *lac* repressor synthesis, we identified BQ<1lacO-A1> as the σ^{70} promoter/operator combination that fulfilled the desired properties. It showed a high expression rate, negligible basal expression, and true tunability of the expression rate on a cellular level, even at low inducer concentrations, without a complete stop in productivity.

Conclusion

The regulation of transcription in *E. coli* has recently received considerable attention, because it is the first step in the process of recombinant protein production [35–38]. Transcription control of the GOI allows a cell to divide up its resources between cellular and recombinant proteins in a physiologically balanced manner. Tight and tunable transcription control of the GOI is essential for successful bioprocesses. We showed that, in genome-integrated expression systems, the regulatory elements of the *lac* operon must be well balanced to control σ^{70} promoters. Three *lacO* sites reduced the basal expression, but also reduced the recombinant protein production rate. The perfectly symmetric *lacO* in the ITS hampered the escape of RNAP from the promoter. As shown by Hsu et al. [29], the wild-type ITS of A1 is enriched in purines, and it displayed one of the best promoter escape efficiencies. Promoters that contained only one *lacO* site exhibited a considerably higher promoter strength, but also higher leakiness. In promoter/operator combinations that contained two *lacO* sites, when the two *lacO1* sites were located within a distance of 62 bp, they exhibited very strong binding affinity with the repressor molecule, which prevented LacI autoregulation. These conditions resulted in a complete stop in productivity in partially induced cells. However, we did not consider that all promoters with two *lacO* sites were unsuitable, in general. The binding affinity can be reduced by using less symmetric *lacO* sites, like *lacO3* or *lacO2*, or by varying the distance between the *lacO* sites [8, 23]. The combination of one *lacO1* site and the $lacI^Q$ promoter (which increased LacI levels) resulted in high GFPmut3.1 expression rates, low basal expression, and true tunability on a cellular level. Thus, we concluded that this novel genome-integrated, host RNAP-dependent expression system would be advantageous for the production of challenging proteins, because it obviates the plasmid-mediated metabolic load, and it confers true tunability on a cellular level.

Methods

Strains and culture conditions

Escherichia coli K-12 NEB5-α [*fhuA2Δ(argF-lacZ)U169 phoA gln V44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*] (New England Biolabs [NEB], Ipswitch, MA, USA) was used for all cloning procedures. Linear DNA cartridges were integrated into the bacterial chromosome at the attTN7 site of *E. coli* BL21 [*fhuA2 [lon] ompT gal [dcm] ΔhsdS*] (NEB). For reference

experiments, the same strains were transformed with the respective plasmids, except that they carried the sequence for the soluble protein, GFPmut3.1, which was used as a recombinant model protein [25].

The strains were cultivated in the BioLector microfermentation system, in 48-well Flowerplates® (m2p-labs, Baesweiler, Germany), as described by Török et al. [39]. We used a synthetic Feed in Time (FIT), fed-batch medium, with 1 g/L glucose and 16.5 g/L dextran as carbon sources (m2p-labs GmbH, Baesweiler, Germany). Additionally, the medium contained (g/L): 27.40 MOPS, 6.54 $(NH_4)_2SO_4$, 1.96 K_2HPO_4, 1.96 trisodium citrate·$2H_2O$, 1.31 Na_2SO_4, 0.65 NH_4Cl, 0.33 $MgSO_4 \cdot 7H_2O$, and 0.0065 Thiamin·HCl.

The trace element solution contained (mg/L): 0.36 $ZnSO_4 \cdot 7H_2O$, 0.33 $CuSO_4 \cdot 5H_20$, 0.20 $MnSO_4 \cdot H_2O$, 27.30 $FeCl_3 \cdot 6H_2O$, 21.84 Titriplex III, 0.36 $CoCl_2 \cdot 6H_2O$, and 1.31 $CaCl_2 \cdot 2H_2O$. Immediately prior to inoculation, 0.6% (v/v) glucose releasing enzyme mix (EnzMix) was added. Expression levels were monitored at an excitation wavelength of 488 nm and an emission wavelength of 520 nm. The signals are expressed in relative fluorescence units [rfu]. The cycle time for all parameters was 20 min. The initial cell density was equivalent to 0.3 optical density at 600 nm (OD_{600}). For inoculation, a deep-frozen (− 80 °C) working cell bank ($OD_{600} = 3.5$) was thawed, and the biomass was harvested by centrifugation (7500 rpm, 5 min). Cells were washed with 500 µL of the corresponding medium to remove residual glycerol. Next, cells were centrifuged, and the pellets were resuspended in the total cultivation medium. All cultivations were prepared in three replicates at 30 °C for 22 h. Recombinant gene expression was induced with 0.005 mM, 0.01 mM, or 0.5 mM IPTG at 10 h after the start of cultivation.

Construction and characterization of promoter/operator combinations

Basic cloning methods, like restriction endonuclease digestions, agarose gel electrophoresis, plasmid engineering, and transformation of *E. coli* plasmids, were carried out according to Sambrook et al. [40]. For the integration of the $lacI^Q$ promoter into *E. coli* BL21 (NEB), we constructed the plasmid, pETAmp-lacIq. This plasmid contained the ampicillin resistance gene (Amp), flanked by FRT sites [41], and the *lacI* gene controlled by the $lacI^Q$ promoter [33]. The pBR322 ori and the *lacI* gene were amplified from pET30a with the overhang PCR technique to add a C → T mutation within the *lacI* promoter. The linear $lacI^Q$ DNA cartridge for genome-integration was amplified with the Q5® High-Fidelity DNA Polymerase (NEB), according to the manufacturer's instructions. Integration into the bacterial chromosome occurred at the *lac* operon site of *E. coli* BL21, which carries the pSIM5 plasmid, as described by Sharan et al. [42]. This strain was designated $BL21^Q$. The sequences of the $T7_{A1}$ and the $T5_{N25}$ promoters were adopted from Lanzer and Bujard [21] (designated as $P_{A1/04}$ and $P_{N25/04}$, respectively). These promoters contained a 2-bp truncated *lacO1* sequence, inserted between the − 10 and − 35 region, upstream of the promoter. These promoters were purchased as gBlocks® Gene Fragments (Integrated DNA Technologies, IA/USA), which contained a 5′ spacer sequence from pET30a and the restriction sites, SphI (5′) and XbaI (3′); these were subsequently cloned into the pET30a-cer-tZENIT-GFPmut3.1 backbone. The tZENIT terminator was described elsewhere [27]. A second *lacO1* sequence, 62 bp upstream of the first *lacO1* sequence, was added via the overhang PCR technique. The 3lacO-T5 promoter/operator combination was adopted from the pJexpress 401–406 (T5) vector from ATUM (Newark, NJ, USA). Linear DNA cartridges were integrated into the bacterial chromosome at the attTN7 site of *E. coli* BL21 or *E. coli* $BL21^Q$.

GFPmut3.1 off-line expression analysis and quantification

In addition to on-line measurements of recombinant GFPmut3.1, expressed in rfus, we performed absolute quantifications with ELISA, according to Reischer et al. [43]. Inclusion body formation was analyzed with SDS-PAGE, as previously described [44] and fractions of soluble and insoluble protein were estimated with ImageQuant TL software (GE Healthcare, Chicago, IL, USA).

Flow cytometry

A Gallios flow cytometer (Beckman Coulter, Brea, CA, USA) was used to determine the fraction of GFPmut3.1-producing cells. Cells were harvested 12 h after induction, then diluted 1:2025 in PBS. GFPmut3.1 fluorescence was excited with an OPSL Sapphire Laser at 488 nm, and the subsequent emission was measured with the FL1 Channel (505–545). Data were recorded for 15,000 cells per sample at ~ 300 events/sec. Analyses were performed with Kaluza analysis software (Beckman Coulter).

Analysis of LacI with western blots

Cell extracts were prepared with ~ 1.2×10^7 BL21-wt and B<2lacO-A1> cells, respectively, and proteins were separated with SDS-PAGE, as previously described. After separation, the proteins were blotted with the iBlot® Dry Blotting System, according to the manufacture's instructions (Invitrogen™/Thermo Fisher Scientific, Waltham, MA, USA). Subsequently, proteins were blocked for 4 h at room temperature with 3% nonfat dry milk in PBST (1x PBS Dulbecco and 0.05% Tween 20). The blots were then incubated with primary antibody (1:1000 anti-LacI Antibody, clone

9A5; Sigma-Aldrich/Merck, St. Louis, MO, USA) for 1 h at room temperature. Blots were then incubated with alkaline phosphatase-conjugated secondary antibody (1:2000 Anti-Mouse IgG, whole molecule, Sigma A5153; Sigma-Aldrich) for 1 h at room temperature. Blots were developed with SigmaFAST™ BCIP®/NPT tablets (Sigma-Aldrich) according to the manufacturer's instructions. Band intensities were quantified with ImageQuant TL software (GE Healthcare, Chicago, IL, USA).

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12934-020-01311-6.

Additional file 1: Figure S1. Growth characteristics of genome-integrated expression systems with different promoter/operator combinations. Cells were grown in enzymatic glucose release media in micro-titer fermentations over a period of 22 h. The dashed vertical lines indicate the time of induction with 0.5 mM IPTG. (A, B) Biomass trends (CDM) and (C, D) growth rates (μ) are shown for (A, C) induced and (B, D) non-induced cells. The mean values of triplicates are shown. The promoter/operators are defined in Fig. 1.

Additional file 2: Figure S2. Growth characteristics of plasmid-based expression systems with different promoter/operator combinations. Cells were grown in enzymatic glucose release media in micro-titer fermentations over a period of 22 h. The dashed vertical lines indicate the time of induction with 0.5 mM IPTG. (A, B) Biomass trends (CDM) and (C, D) growth rates (μ) are shown for (A, C) induced and (B, D) non-induced cells. The mean values of triplicates are shown. The promoter/operators are defined in Fig. 1.

Additional file 3: Figure S3. Solubility analysis of GFPmut3.1. SDS-PAGE images show soluble (S) and insoluble (I) fractions of proteins produced under the indicated *lacO*-promoter combinations in genome-integrated (indicated with pointed brackets <>) and plasmid-based (indicated with round brackets ()) expression systems.

Additional file 4: Figure S4. Determination of IPTG concentrations for full and partial induction. The dashed vertical lines indicate the time of induction. Induction was performed with 0 (gray, not induced), 0.005 (red), 0.01 (blue), 0.05 (orange), 0.1 (violet), 0.5 (green) or 1.0 (black) mM IPTG. (A) B<3lacO-T5>. (B) B3<T7>. The mean relative GFP fluorescence intensity (rfu) represents triplicate samples.

Authors' contributions

AS, MC, RG, and GS designed the experiments and drafted the manuscript. AS and CT performed experiments. AS, MC, JJ, MW, DR, RG, and GS analyzed data. AS and RG wrote the paper. All authors read and approved the final manuscript.

Funding

We gratefully acknowledge the financial support received from the Austrian Federal Ministry of Science, Research and Economy; from the National Foundation of Research, Technology, and Development; and from Boehringer Ingelheim RCV GmbH & Co KG.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its Additional files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Boehringer Ingelheim RCV GmbH & Co KG filed a patent covering the use and application of the expression technology employed in this study.

Author details

[1] Christian Doppler Laboratory for Production of Next-level Biopharmaceuticals in E. coli, Department of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 18, 1190 Vienna, Austria. [2] Boehringer Ingelheim RCV GmbH & Co KG, Dr.-Boehringer-Gasse 5-11, 1120 Vienna, Austria.

Received: 12 December 2019 Accepted: 17 February 2020
Published online: 05 March 2020

References

1. Angius F, Ilioaia O, Amrani A, Suisse A, Rosset L, Legrand A, et al. A novel regulation mechanism of the T7 RNA polymerase based expression system improves overproduction and folding of membrane proteins. Sci Rep. 2018;8(1):8572.
2. Chia-Chang Hsu ORTTaTWO. Periplasmic expression in and release of Fab fragments from *Escherichia coli* using stress minimization. J Chem Technol Biotechnol. 2015;91(3):815–22.
3. Saida F, Uzan M, Odaert B, Bontems F. Expression of highly toxic genes in *E. coli*: special strategies and genetic tools. Curr Protein Pept Sci. 2006;7(1):47–56.
4. Riggs AD, Bourgeois S. On the assay, isolation and characterization of the lac repressor. J Mol Biol. 1968;34(2):361–4.
5. Barkley MD, Riggs AD, Jobe A, Burgeois S. Interaction of effecting ligands with lac repressor and repressor-operator complex. Biochemistry. 1975;14(8):1700–12.
6. Oehler S, Eismann ER, Kramer H, Muller-Hill B. The three operators of the lac operon cooperate in repression. EMBO J. 1990;9(4):973–9.
7. Sadler JR, Sasmor H, Betz JL. A perfectly symmetric lac operator binds the lac repressor very tightly. Proc Natl Acad Sci USA. 1983;80(22):6785–9.
8. Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, Müller-Hill B. Quality and position of the three lac operators of *E. coli* define efficiency of repression. EMBO J. 1994;13(14):3348–55.
9. Mossing MC, Record MT Jr. Upstream operators enhance repression of the lac promoter. Science. 1986;233(4766):889–92.
10. Reznikoff WS, Winter RB, Hurley CK. The location of the repressor binding sites in the lac operon. Proc Natl Acad Sci USA. 1974;71(6):2314–8.
11. Semsey S, Jauffred L, Csiszovszki Z, Erdossy J, Steger V, Hansen S, et al. The effect of LacI autoregulation on the performance of the lactose utilization system in *Escherichia coli*. Nucleic Acids Res. 2013;41(13):6381–90.
12. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol. 2014;5:172.
13. Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol. 1986;189(1):113–30.
14. Silverstone AE, Arditti RR, Magasanik B. Catabolite-insensitive revertants of lac promoter mutants. Proc Natl Acad Sci USA. 1970;66(3):773–9.
15. Mairhofer J, Scharl T, Marisch K, Cserjan-Puschmann M, Striedner G. Comparative transcription profiling and in-depth characterization of plasmid-based and plasmid-free *Escherichia coli* expression systems under production conditions. Appl Environ Microbiol. 2013;79(12):3802–12.
16. Glick BR. Metabolic load and heterologous gene expression. Biotechnol Adv. 1995;13(2):247–61.
17. Marchand I, Nicholson AW, Dreyfus M. High-level autoenhanced expression of a single-copy gene in *Escherichia coli*: overproduction of bacteriophage T7 protein kinase directed by T7 late genetic elements. Gene. 2001;262(1–2):231–8.
18. Striedner G, Pfaffenzeller I, Markus L, Nemecek S, Grabherr R, Bayer K. Plasmid-free T7-based *Escherichia coli* expression systems. Biotechnol Bioeng. 2010;105(4):786–94.
19. Gentz R, Bujard H. Promoters recognized by *Escherichia coli* RNA polymerase selected by function: highly efficient promoters from bacteriophage T5. J Bacteriol. 1985;164(1):70–7.

20. Brunner M, Bujard H. Promoter recognition and promoter strength in the *Escherichia coli* system. EMBO J. 1987;6(10):3139–44.
21. Lanzer M, Bujard H. Promoters largely determine the efficiency of repressor action. Proc Natl Acad Sci USA. 1988;85(23):8973–7.
22. Kadesch TR, Rosenberg S, Chamberlin MJ. Binding of *Escherichia coli* RNA polymerase holoenzyme to bacteriophage T7 DNA. Measurements of binding at bacteriophage T7 promoter A1 using a template competition assay. J Mol Biol. 1982;155(1):1–29.
23. Camsund D, Heidorn T, Lindblad P. Design and analysis of LacI-repressed promoters and DNA-looping in a cyanobacterium. Journal of biological engineering. 2014;8(1):4.
24. Glascock CB, Weickert MJ. Using chromosomal lacIQ1 to control expression of genes on high-copy-number plasmids in *Escherichia coli*. Gene. 1998;223(1–2):221–31.
25. Cormack BP, Valdivia RH, Falkow S. FACS-optimized mutants of the green fluorescent protein (GFP). Gene. 1996;173(1 Spec No):33–8.
26. Kensy F, Zang E, Faulhammer C, Tan RK, Buchs J. Validation of a high-throughput fermentation system based on online monitoring of biomass and fluorescence in continuously shaken microtiter plates. Microb Cell Fact. 2009;8:31.
27. Mairhofer J, Wittwer A, Cserjan-Puschmann M, Striedner G. Preventing T7 RNA polymerase read-through transcription-A synthetic termination signal capable of improving bioprocess stability. ACS Synth Biol. 2015;4(3):265–73.
28. Deuschle U, Kammerer W, Gentz R, Bujard H. Promoters of *Escherichia coli*: a hierarchy of in vivo strength indicates alternate structures. EMBO J. 1986;5(11):2987–94.
29. Hsu LM, Cobb IM, Ozmore JR, Khoo M, Nahm G, Xia L, et al. Initial transcribed sequence mutations specifically affect promoter escape properties. Biochemistry. 2006;45(29):8841–54.
30. Afroz T, Biliouris K, Kaznessis Y, Beisel CL. Bacterial sugar utilization gives rise to distinct single-cell behaviours. Mol Microbiol. 2014;93(6):1093–103.
31. Gilbert W, Muller-Hill B. Isolation of the lac repressor. Proc Natl Acad Sci USA. 1966;56(6):1891–8.
32. Garcia HG, Phillips R. Quantitative dissection of the simple repression input-output function. Proc Natl Acad Sci USA. 2011;108(29):12173–8.
33. Muller-Hill B, Crapo L, Gilbert W. Mutants that make more lac repressor. Proc Natl Acad Sci USA. 1968;59(4):1259–64.
34. Lin-Chao S, Bremer H. Effect of the bacterial growth rate on replication control of plasmid pBR322 in *Escherichia coli*. Mol Gen Genet MGG. 1986;203(1):143–9.
35. Brewster RC, Jones DL, Phillips R. Tuning promoter strength through RNA polymerase binding site design in *Escherichia coli*. PLoS Comput Biol. 2012;8(12):e1002811.
36. De Mey M, Maertens J, Boogmans S, Soetaert WK, Vandamme EJ, Cunin R, et al. Promoter knock-in: a novel rational method for the fine tuning of genes. BMC Biotechnol. 2010;10:26.
37. Kim SK, Lee DH, Kim OC, Kim JF, Yoon SH. Tunable control of an *Escherichia coli* expression system for the overproduction of membrane proteins by titrated expression of a mutant lac repressor. ACS Synth Biol. 2017;6(9):1766–73.
38. Marschall L, Sagmeister P, Herwig C. Tunable recombinant protein expression in *E. coli*: promoter systems and genetic constraints. Appl Microbiol Biotechnol. 2017;101(2):501–12.
39. Toeroek C, Cserjan-Puschmann M, Bayer K, Striedner G. Fed-batch like cultivation in a micro-bioreactor: screening conditions relevant for *Escherichia coli* based production processes. SpringerPlus. 2015;4:490.
40. Green JFSaMR. Molecular cloning: a laboratory manual. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2012.
41. Andrews BJ, Proteau GA, Beatty LG, Sadowski PD. The FLP recombinase of the 2 micron circle DNA of yeast: interaction with its target sequences. Cell. 1985;40(4):795–803.
42. Sharan SK, Thomason LC, Kuznetsov SG, Court DL. Recombineering: a homologous recombination-based method of genetic engineering. Nat Protoc. 2009;4(2):206–23.
43. Reischer H, Schotola I, Striedner G, Potschacher F, Bayer K. Evaluation of the GFP signal and its aptitude for novel on-line monitoring strategies of recombinant fermentation processes. J Biotechnol. 2004;108(2):115–25.
44. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970;227(5259):680–5.

Publisher's Note

Publication I

Supporting Information

Supplementary Figures

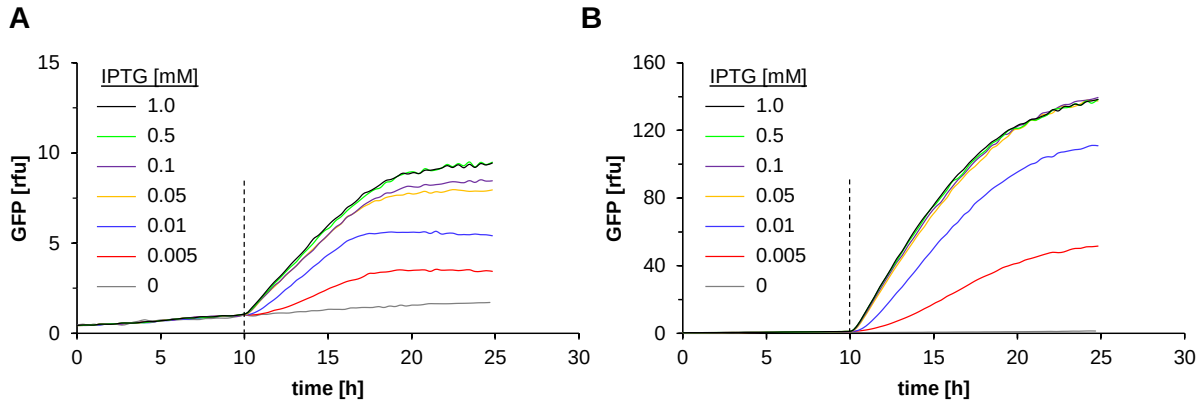


Figure S1 Determination of IPTG concentrations for full and partial induction. *The dashed vertical lines indicate the time of induction. Induction was performed with 0 (gray, not induced), 0.005 (red), 0.01 (blue), 0.05 (orange), 0.1 (violet), 0.5 (green) or 1.0 (black) mM IPTG. (A) B<3lacO-T5>. (B) B3<T7>. The mean relative GFP fluorescence intensity (rfu) represents triplicate samples.*

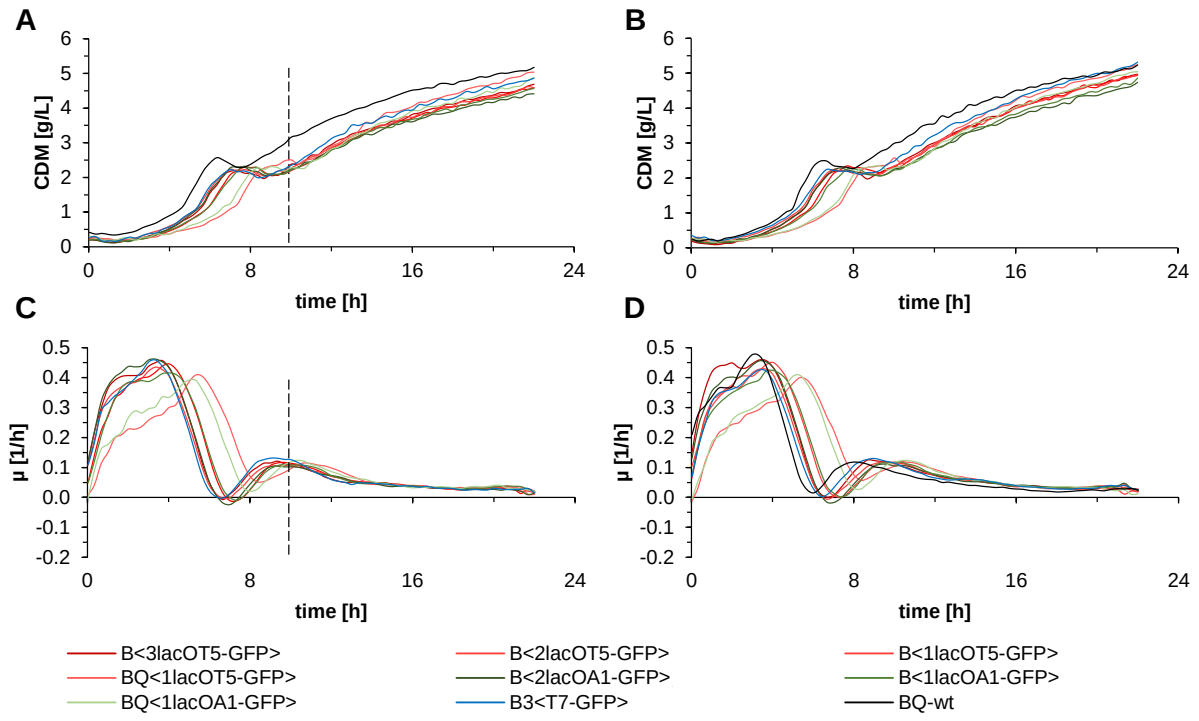


Figure S2 Growth characteristics of genome-integrated expression systems with different promoter/operator combinations. Cells were grown in enzymatic glucose release media in micro-titer fermentations over a period of 22 h. The dashed vertical lines indicate the time of induction with 0.5 mM IPTG. (A, B) Biomass trends (CDM) and (C, D) growth rates (μ) are shown for (A, C) induced and (B, D) non-induced cells. The mean values of triplicates are shown. The promoter/operators are defined in Figure 1.

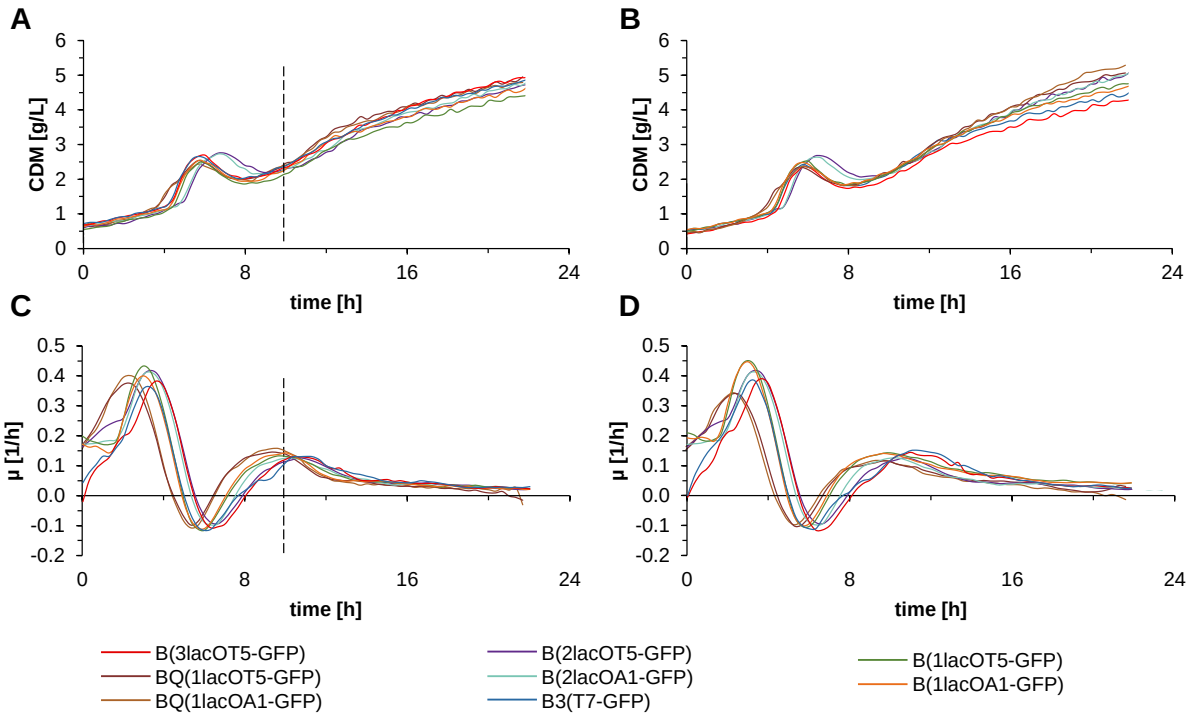


Figure S3 Growth characteristics of plasmid-based expression systems with different promoter/operator combinations. Cells were grown in enzymatic glucose release media in micro-titer fermentations over a period of 22 h. The dashed vertical lines indicate the time of induction with 0.5 mM IPTG. (A, B) Biomass trends (CDM) and (C, D) growth rates (μ) are shown for (A, C) induced and (B, D) non-induced cells. The mean values of triplicates are shown. The promoter/operators are defined in Figure 1.

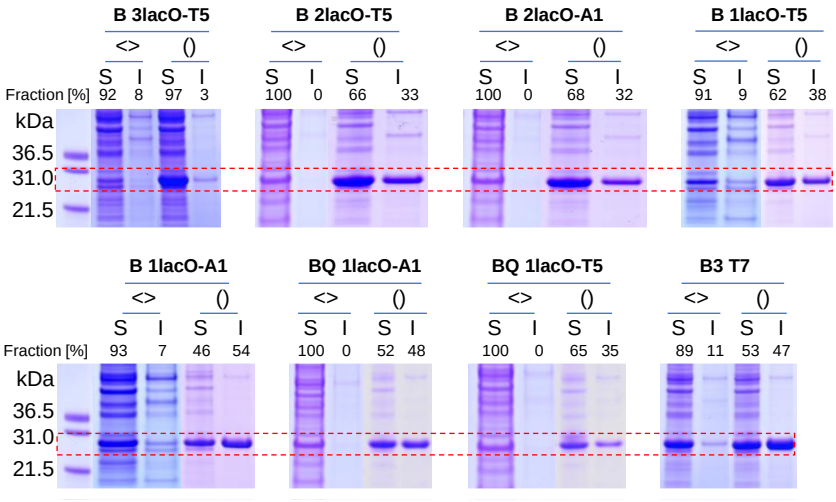


Figure S4 Solubility analysis of GFPmut3.1. SDS-PAGE images show soluble (S) and insoluble (I) fractions of proteins produced under the indicated *lacO*-promoter combinations in genome-integrated (indicated with pointed brackets <>) and plasmid-based (indicated with round brackets ()) expression systems.

Publication II

Biotechnology Journal
www.biotechnology-journal.com

Adaptive Evolution in Producing Microtiter Cultivations Generates Genetically Stable *Escherichia coli* Production Hosts for Continuous Bioprocessing

Artur Schuller, Monika Cserjan-Puschmann, Christoph Köppl, Reingard Grabherr, Martin Wagenknecht, Matteo Schiavinato, Juliane C. Dohm, Heinz Himmelbauer, and Gerald Striedner*

The production of recombinant proteins usually reduces cell fitness and the growth rate of producing cells. The growth disadvantage favors faster-growing non-producer mutants. Therefore, continuous bioprocessing is hardly feasible in *Escherichia coli* due to the high escape rate. The stability of *E. coli* expression systems under long-term production conditions and how metabolic load triggered by recombinant gene expression influences the characteristics of mutations are investigated. Iterated fed-batch-like microbioreactor cultivations are conducted under production conditions. The easy-to-produce green fluorescent protein (GFP) and a challenging antigen-binding fragment (Fab) are used as model proteins, and BL21(DE3) and BL21^{Q} strains as expression hosts. In comparative whole-genome sequencing analyses, mutations that allowed cells to grow unhindered despite recombinant protein production are identified. A T7 RNA polymerase expression system is only conditionally suitable for long-term cultivation under production conditions. Mutations leading to non-producers occur in either the T7 RNA polymerase gene or the T7 promoter. The host RNA polymerase-based BL21^{Q} expression system remains stable in the production of GFP in long-term cultivations. For the production of Fab, mutations in lacI of the BL21^{Q} derivatives have positive effects on long-term stability. The results indicate that adaptive evolution carried out with genome-integrated *E. coli* expression systems in microtiter cultivations under industrial-relevant production conditions is an efficient strain development tool for production hosts.

1. Introduction

Fed-batch bioprocessing is the most common cultivation method in the industrial microbial production of biopharmaceuticals. This batch-wise process essentially includes the repetitive steps of media preparation and reactor setup, fermentation, and subsequent cleaning in place (CIP) and sterilization in place (SIP). In terms of time, the actual fermentation, and particularly the production phase of the recombinant protein, is relatively short. As a result, continuous production becomes more and more interesting due to the greatest possible space–time yields and optimal use of the installed assets.[1] In chemostat cultivations, cells are maintained in a steady-state growth environment by adding fresh medium to the reactor at a constant flow. Simultaneously, the cell suspension, and thus the recombinant protein, is removed at the same rate.[2] The growth rate (μ) can be specified depending on the dilution rate (D). Through this process, stable volumetric productivity and high space–time yield can be achieved.[3] In contrast to fed-batch fermentation, the average residence time of a producing cell is

A. Schuller, Dr. M. Cserjan-Puschmann, C. Köppl, Prof. R. Grabherr, Prof. G. Striedner
Christian Doppler Laboratory for Production of Next-Level Biopharmaceuticals in E. coli, Department of Biotechnology
University of Natural Resources and Life Sciences
Muthgasse 18, Vienna A-1190, Austria
E-mail: monika.cserjan@boku.ac.at

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/biot.202000376

DOI: 10.1002/biot.202000376

Dr. M. Wagenknecht
Boehringer Ingelheim RCV GmbH & Co KG
Dr.-Boehringer-Gasse 5-11 Vienna A-1120, Austria
Dr. M. Schiavinato, Prof. J. C. Dohm, Prof. H. Himmelbauer
Department of Biotechnology
Institute of Computational Biology
University of Natural Resources and Life Sciences
Muthgasse 18 Vienna A-1190, Austria

always the same, which can be advantageous in terms of product quality. Examples of microbial continuous processes for the production of recombinant proteins have already been described in the literature.[4–6]

For industrial microbial production of recombinant proteins, the *Escherichia coli* strain BL21(DE3) is often used due to low acetate formation and high production rates, resulting from the integrated T7 RNA polymerase (RNAP).[7,8] In combination with a pET series plasmid, which harbors the gene of interest (GOI) under control of the T7 promoter, extraordinarily high expression rates can be achieved after induction with the non-metabolizable lactose analog isopropyl β-d-1-thiogalactopyranoside (IPTG).[9]

Genetic heterogeneity caused by high physiological burden and toxicity can be problematic at all industrial scales, especially for challenging proteins.[10–12] In bacterial production processes, challenging proteins impose adverse effects on host physiology, even at low concentrations. Escape variants, which have a growth advantage due to mutations or plasmid loss, can overgrow high-performing producer cells, reducing the overall product yield.[13–15] Therefore, longer production phases, or even continuous production mode, are hardly feasible in such *E. coli* expression systems.

To obtain stable, high-yield, and predictable *E. coli* production hosts, engineered producer strains must focus on reducing metabolic load and genetic escape. Attempts to reduce process instability caused by metabolic burden have been made on both the genetic and bioprocessing levels. The metabolic burden and increased selection pressure can be reduced by decoupling growth and production in cascading chemostat cultivations with two bioreactors.[6] Genetic escape can be reduced by removing insertion elements, and by deleting recA or error-prone DNA polymerase genes.[16] Another promising strategy uses the evolution approach combined with fluorescence-activated cell sorting to select cells with a lower plasmid mutation rate.[17]

In previous studies, we showed that genomic integration of the GOI under the control of the strong T7 promoter reduces the metabolic burden because the plasmid-mediated metabolic load is eliminated, and strong expression has been shown from even a single copy of the GOI. However, after approximately seven doublings under production conditions, mutations in the T7RNAP gene lead to a faster-growing non-producing cell population. This phenomenon can be excluded in systems using the host RNAP-specific A1 promoter because the full functionality of the host RNAP is required for cell growth.[18]

In combination with an adaptive evolution approach, this would allow the characterization of mutated production hosts, which have adapted themselves to the burden triggered by recombinant protein production and could potentially enable continuous protein production. Moreover, an adaptive evolution approach can circumvent the complexity of the process of biogenesis and its adverse effects on the host cell.[19–21]

For example, in the study by Walker et al.,[22] derivatives of BL21(DE3) were adapted by evolution to produce cell membrane proteins that are toxic to host cells. These strains are currently widely used for the production of a variety of membrane proteins and toxic proteins.

In the present study, we performed iterated carbon-limited fed-batch-like microbioreactor cultivations under production conditions. The goal was to investigate whether and how long-term burden triggered by the production of recombinant proteins influences the characteristics of mutations occurring in different genome-integrated *E. coli* production systems.[23,24] We compared the host RNAP-dependent BL21^{Q} A1 expression system (BQ<A1>)[18] with the T7-based BL21(DE3) expression system (B3<T7>). To study mutation characteristics based on the protein of interest (POI), we used the easy-to-produce protein GFPmut3.1 and the challenging protein Fab fragment dFTN2 as model proteins. To investigate the underlying mutations in the different clones in more detail, we performed comparative whole-genome sequencing analyses. We also performed long-term chemostat cultivations in laboratory-scale bioreactors with the abovementioned clones and additional robust production strains obtained by the adaptive evolution approach.

2. Experimental Section

2.1. Construction of Expression Systems

2.1.1. Strains

E. coli K-12 NEB5-α (fhuA2Δ(argF-lacZ)U169 phoA gln V44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17) was obtained from New England Biolabs (NEB, Ipswich, USA) and used for all cloning procedures. For recombinant protein expression, linear DNA cartridges controlled by the T7 promoter were integrated into the bacterial chromosome at the attTn7 site of *E. coli* BL21(DE3) (fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5) (NEB). Hereafter, this strain is referred to as B3<T7-gene>. Linear DNA cartridges controlled by the A1 promoter were integrated into the bacterial chromosome at the attTn7 site of *E. coli* BL21 (fhuA2 [lon] ompT gal [dcm] ΔhsdS) (NEB) containing the lacI^{Q} promoter as described by Glascock and Weickert.[25] Briefly, the pETAmp-lacIq plasmid was constructed for the integration of the lacI^{Q} promoter in *E. coli* BL21. This plasmid contains the ampicillin resistance gene flanked by FRT sites and the lacI gene controlled by the lacI^{Q} promoter. The pBR322 ori and lacI were amplified from pET30a using the overhang PCR technique to introduce a C → T mutation in the lacI promoter. The linear lacI^{Q} DNA cartridge for genome integration was amplified using Q5 high-fidelity DNA polymerase (NEB) according to the manufacturer's manual. Integration into the bacterial chromosome occurred at the lac operon site of *E. coli* BL21 carrying the pSIM5 plasmid.[26] This strain was designated as BL21^{Q} and referred to hereafter as BQ <A1-gene>.

The cytosolic protein GFPmut3.1 was used as a recombinant "easy-to-produce" model protein.[27] An antigen-binding fragment (Fab) against TNF-α (FTN2) with the DsbA signal sequence (dFTN2) was used as a recombinant "challenging" model protein.[28]

2.1.2. Plasmids and Integration of Expression Cassettes into the E. coli Chromosome

Plasmids harboring the integration cassettes of either GFPmut3.1 or dFTN2 were constructed and integrated into the *E. coli* chromosome as described by Fink et al.[28]

2.2. Microbioreactor Cultivations

The strains were cultivated in the BioLector microfermentation system in 48-well flowerplates (m2p-labs, Baesweiler, Germany) as described by Török et al.[29] Synthetic feed in time (FIT) fed-batch medium containing 1 g L^{-1} glucose and 16.5 g L^{-1} dextran as carbon sources (m2p-labs) was used with the following additions (g L^{-1}): 27.40 MOPS, 6.54 $(NH_4)_2SO_4$, 1.96 K_2HPO_4, 1.96 trisodium citrate·$2H_2O$, 1.31 Na_2SO_4, 0.65 NH_4Cl, 0.33 $MgSO_4 \cdot 7H_2O$, and 0.0065 thiamin·HCl.

The trace element solution contained (mg L^{-1}) 0.36 $ZnSO_4 \cdot 7H_2O$, 0.33 $CuSO_4 \cdot 5H_2O$, 0.20 $MnSO_4 \cdot H_2O$, 27.30 $FeCl_3 \cdot 6H_2O$, 21.84 Titriplex III, 0.36 $CoCl_2 \cdot 6H_2O$, and 1.31 $CaCl_2 \cdot 2H_2O$. Immediately before inoculation, 0.6% v/v of the glucose releasing enzyme mix (EnzMix) was added. The GFPmut3.1 expression level was monitored at an excitation of 488 nm and an emission of 520 nm. The signal is given in relative fluorescence units (rfu). The cell dry matter (CDM, given in grams per liter) was calculated from light scatter signals using calibration settings obtained by linear regression analysis. The cycle time for all parameters was 20 min.

The initial cell density was equivalent to an optical density of OD_{600} = 0.3. For the inoculation of passage 1, a deep-frozen (–80 °C) working cell bank (WCB; OD_{600} = 3.5) was thawed and biomass harvested by centrifugation (7500 rpm, 5 min). Cells were washed with 500 µL of the corresponding medium to remove residual glycerol and centrifuged. Pellets were then resuspended in the total cultivation medium. All subsequent passages were inoculated with induced cells from the previous passage, but without adding batch glucose again, keeping the cells in carbon-limited conditions. All cultivations were performed at 30 °C. Recombinant gene expression was induced with IPTG at a final concentration of 0.5×10^{-3} M.

2.3. Fed-Batch Cultivations

For fed-batch fermentation, cells were grown in a 1.5 L (1.2 L working volume, 0.4 L minimal volume) DASGIP Parallel Bioreactor System (Eppendorf AG, Hamburg, DE) equipped with standard control units. The pH was maintained at 7.0 ± 0.05 by the addition of 12.5% ammonia solution (Thermo Fisher Scientific, Waltham, USA). The temperature was maintained at 37 ± 0.5 °C during the batch phase and decreased to 30 ± 0.5 °C during the feed phase. The dissolved oxygen (O_2) level was stabilized at >30% saturation by controlling the stirrer speed and aeration rate. Foaming was suppressed by the addition of antifoam suspension (Glanapon, 2000, Bussetti, AT). For inoculation, a deep-frozen (–80 °C) WCB vial was thawed and 1 mL transferred aseptically to a 250 mL preculture shake flask containing 25 mL M9ZB[30] for cultivation for at least 8 h at 37 °C. Subsequently, a volume equivalent to 25 OD_{600} units ($25/OD_{600}$ = volume in milliliter) was transferred aseptically to the bioreactors.

Feeding was initiated when the culture, grown to 10 g L^{-1} CDM in 0.6 L batch medium, entered the stationary phase. A fed-batch regimen with exponential carbon-limited substrate feed was used to provide a constant growth rate of $\mu = 0.1$ h^{-1} over 19 h or 2.74 doublings. The substrate feed was controlled by increasing the pump speed according to the exponential growth algorithm, $x = x_0 e^{\mu t}$, with superimposed feedback control of weight loss in the substrate bottle. The CDM yield coefficient on glucose was 0.33 g g^{-1}. Therefore, the feed medium provided 66 g L^{-1} glucose and sufficient components to yield a final CDM concentration of 30 g L^{-1} in 1.2 L. The expression system was induced by adding IPTG to the reactor to yield a concentration of 10 µmol g^{-1} CDM. The minimal medium was prepared as previously described.[15]

2.4. Chemostat Cultivations

Chemostat cultivations were run at a density of 30 g L^{-1} CDM and a dilution rate of $D = 0.1$ h^{-1} in a working volume of 0.67 L. The starting batch process was followed by a fed-batch phase. The batch volume was set to 400 mL and the batch medium allowed production of 4.72 g CDM; the feed volume was set to 270 mL with a medium designed to produce another 15.4 g of CDM, which corresponds to a final CDM concentration of 30 g L^{-1}. In the next step, the bioreactor was shifted to chemostat mode with a dilution rate of 67 mL h^{-1}. The medium was used as a continuous feed to the reactor to provide nutrients to maintain a CDM concentration of 30 g L^{-1}. The temperature was decreased to 30 °C and recombinant protein production induced with 10 µmol IPTG g^{-1} CDM. Volume was kept constant via an immersion tube adjusted to the right height of the liquid surface, ensuring that the bleed pump was working at a higher rate than the feed pump.

2.5. Analysis of Recombinant Protein Expression

2.5.1. Flow Cytometry

A CytoFLEX S flow cytometer (Beckman Coulter, Brea, USA) was used to monitor the subpopulations of GFPmut3.1-producing cells. Cells were diluted 1/2025 in PBS. Excitation of GFPmut3.1 fluorescence was performed at 488 nm, with subsequent emission measured using the FL1 Channel (525BP40-A). Data were recorded for 15 000 cells per sample at ≈300 events/s and were analyzed by Kaluza analysis software (Beckman Coulter).

2.5.2. Offline Measurement of GFP Fluorescence

Offline fluorescence measurements were made using a Tecan analyzer infinite 200Pro. For GFPmut3.1, the excitation wavelength was 485 nm, and the emission wavelength of 520 nm. Calibration with the in-house purified target protein was used for quantification.

2.5.3. Cell Lysis

Samples from microbioreactor experiments were taken at the end of each passage. Samples from fed-batch cultivations were taken every 2 h and samples from chemostat cultivations every 24 h. The sample volume for intracellular protein quantification was calculated as $3.5/OD_{600}$ = volume in milliliter, which corresponds to ≈1 mg CDM. Cell lysis for quantification of intracellular recombinant proteins was performed as described elsewhere.[28]

2.5.4. Quantification of Soluble Recombinant Protein

Recombinant GFPmut3.1 and FTN2 were quantified by enzyme-linked immunosorbent assay as described previously.[28,31]

2.5.5. Fab Expression Pattern on Dot Blot

The FTN2 samples from microbioreactor cultivations were analyzed initially in a high-throughput manner using dot blot to obtain a simple yes-or-no-answer. For that purpose, 200 µL of cell suspension was centrifuged in a 48-deep-well plate at 1200 rpm for 10 min. The supernatant was removed and pelleted cells resuspended in 400 µL TE buffer (pH 8). Resuspended cells were incubated for 1 h at 60 °C with shaking, and then subsequently centrifuged as described above. The supernatant (3 µL) was pipetted on a nitrocellulose membrane. Blocking, incubation with antibodies, and membrane development were performed as described below for Western blotting.

2.5.6. Fab Expression Pattern via Western Blot

Soluble Fab expression, inclusion body (IB) formation, and basal expression levels were analyzed by Western blot as described previously.[28]

2.5.7. Paired-End Library Preparation and Whole-Genome Sequencing

Genomic DNA extraction was achieved by standard phenol/chloroform extraction as described previously.[32] The quality of the genomic DNA was checked on a 0.6% standard agarose gel stained with ethidium bromide using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher). Quantification was performed using a Qubit 3.0 Fluorometer and a dsDNA BR assay kit (Thermo Fisher). Paired-end libraries for Illumina sequencing were prepared using a TruSeq Nano DNA Low Throughput Library Prep kit (Cat# 20015964, Illumina, San Diego, CA, USA) according to the supplier's instructions. Briefly, library construction began with the fragmentation of 200 ng of genomic DNA to a peak fragment size of 550 bp using a Covaris M220 instrument (Covaris, Woburn, MA, USA) with the following settings: 45 s at 20% duty cycle, intensity 50, temperature 20 °C, and 200 cycles per burst. The DNA fragments were then purified using SPB beads included in the TruSeq kit, followed by end-repair, A-tailing, and ligation of Illumina adapters to the ends of the fragments. After SPB purification, the library was PCR-amplified using the following cycling conditions: initial denaturation at 95 °C for 3 min, followed by eight cycles at 98 °C for 20 s, 60 °C for 15 s, and 72 °C for 30 s, and a final extension at 72 °C for 5 min. After PCR clean-up, 1 µL of the library was used for validation in a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). Library quantification was accomplished on a Qubit 3.0 Fluorometer using a dsDNA BR assay kit. The library was then sequenced at the VBCF NGS Core facility (Vienna, Austria) on an Illumina HiSeq 2500 sequencing instrument using v4 sequencing chemistry and a 2 × 125 nt paired-end sequencing protocol.

Raw genomic short reads from all wild-type and mutant strains were quality trimmed with Trimmomatic v0.35.[33] (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 AVGQUAL:20 MINLEN:50). Four reference sequences were generated that represented the versions of the T7 and A1 wild-type strains: B3<T7-GFP> wt, B3<T7-Fab> wt, BQ<A1-GFP> wt, and BQ<A1-Fab> wt. Variants between the T7 wild-type strains and the publicly available BL21(DE3) genome (NCBI reference sequence: NC_012892.2) were extracted using their quality-trimmed reads. The same was repeated for A1 wild-type strains using the BL21 genome.[34] To extract the variants, breseq[35] was used (limit fold-coverage 150, minimum mapping quality 20, maximum read mismatches 15, no junction prediction, require match-fraction 0.5). The four updated references were obtained introducing the found variants with gdtools APPLY provided with breseq.

Using the quality-trimmed reads of the mutant strains, one genome sequence was assembled for each mutant. The peak insert size of each sequencing library was determined by mapping up to 10 000 read pairs with HISAT2[36] on the appropriate reference genome sequence. Quality-trimmed reads were then down-sampled to coverage of 150× using the seqtk sample (https://github.com/lh3/seqtk). Down-sampled reads were assembled using SOAP-denovo[37] (–M 3 –L 100 –K 75 –k 25 –d 5 –D 5). The peak insert size determined for each library was used to calibrate the genome assembler. The specified config file parameters were reverse_seq = 0, asm_flags = 3, max_rd_len = 125, rd_len_cutoff = 125, rank = 1, map_len = 30. Assembly metrics (contig N50 length and contig N90 length) were assessed using Biopython.[38]

Single nucleotide polymorphisms (SNPs) and short insertion/deletion (indel) variants up to 20 bp in length between mutant and wild-type strains were extracted. Short reads from each mutant were mapped onto their corresponding wild-type reference using breseq.[35] The program was run in polymorphism mode (polymorphism prediction, polymorphism frequency cutoff 0.05) using the same parameters as for the generation of mutated references. True variants were then selected by mapping the contigs of the assembled genomes of the mutants; only variants with both read and contig mapping were retained. The clonality of each variant was inferred from the "AF" field in the VCF files.

The presence of candidate genome rearrangements was screened using the assembled genome sequences for the mutants. Contigs were mapped onto the corresponding wild-type reference sequence with nucmer[39] (mum, breaklen 10, mincluster 500, delta, diagfactor 0.05, maxgap 30000, minmatch 50). The produced mapping records were then passed to the show-diff tool, available with nucmer, which highlights potential rearrangements and their type.

3. Results and Discussion

3.1. Adaptive Evolution and Isolation of Protein Production Strains

We applied repeated fed-batch-like cultivations in a 48-well microbioreactor system to characterize mutation patterns triggered by recombinant protein production under long-term

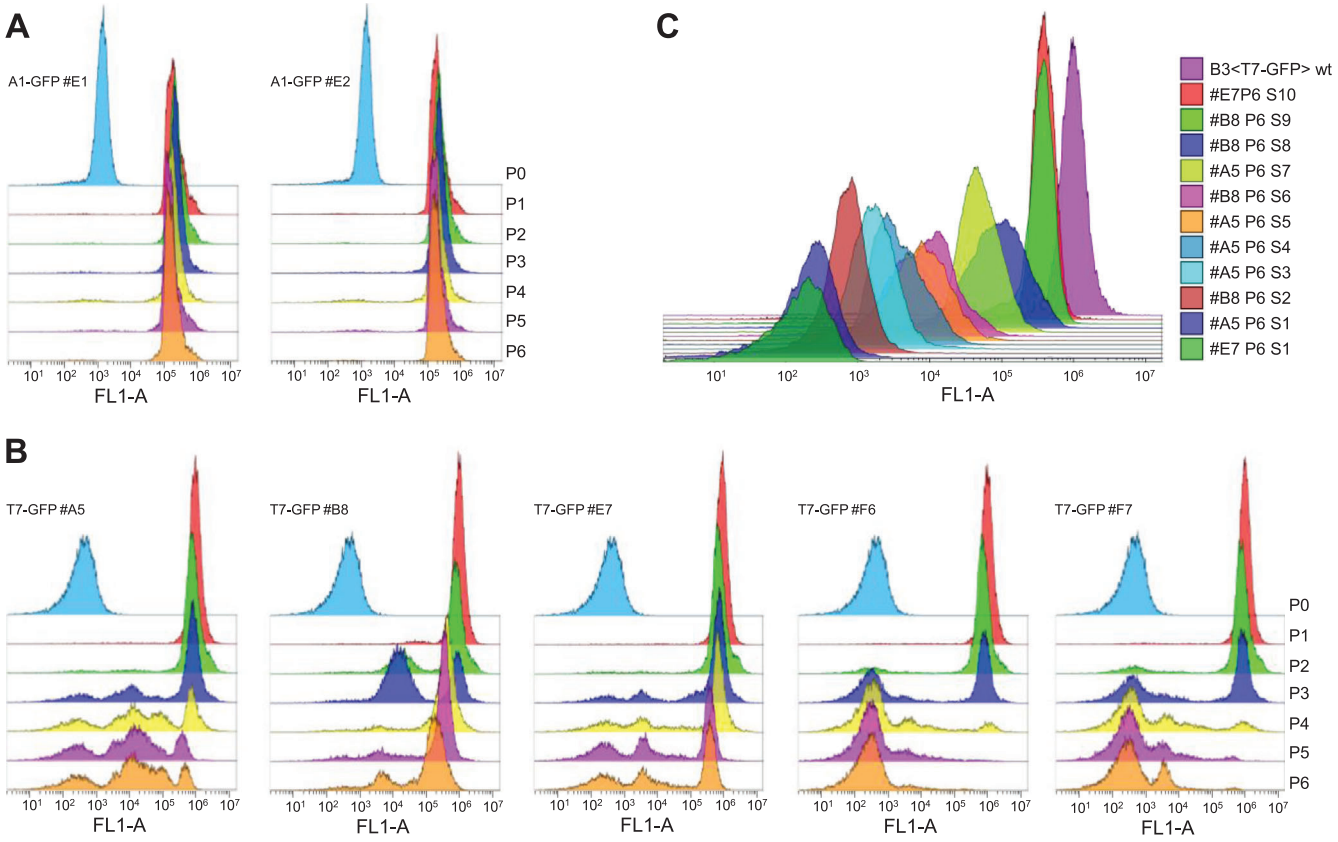


Figure 1. Analysis of single-cell expression using GFP as a model protein. A,B) Flow cytometry throughout six passages. Passage 0 represents the uninduced control. A) Derivatives of BQ<A1-GFP>. B) Derivatives of B3<T7-GFP>. C) Derivatives of B3<T7-GFP> separated and categorized according to the strength of their expression. S1: weak producing to S10: strong producing subclones.

cultivations. Twenty-four wells were inoculated with the host RNAP-dependent A1 expression system (BQ<A1>) and 24 with the T7 RNAP-dependent BL21(DE3) expression system (B3<T7>), in which both strains have the GOI integrated into the chromosome. GFP-producing cells were induced with 10 µmol IPTG g^{-1} CDM at the beginning of each passage, which corresponds to the full induction of recombinant protein production. Fab-producing cells were induced after 6 h of the first passage and cultivated in a medium that already contained 10 µmol IPTG g^{-1} CDM in all other passages. Cells were passaged several times until no difference in growth behavior was observed. We performed the experiments with two different model proteins, the easy-to-produce protein GFPmut3.1, and challenging protein dFTN2. For GFP-producing strains, the fluorescence of GFP was used to distinguish between producing and non-producing clones throughout six passages and a total of 42 generations. Fab-producing cells were cultivated throughout three passages (21 generations) and dot blot analysis performed to detect producing clones.

3.1.1. Evolution of GFP-Producing E. coli Strains

Induction of cells producing GFP under the control of the A1 promoter, BQ<A1-GFP>, showed no reduction in cell growth (Figure S1, Supporting Information). The cells were able to reach the calculated end biomass of 6 g L^{-1} CDM in each of the six passages. Similarly, the productivity remained unchanged and was always ≈25 rfu g^{-1} CDM at the end of each passage (Figure S2, Supporting Information). Flow cytometry revealed continuous homogeneous populations of all 24 measured cultivations, as depicted for clones #E1 and #E2 in **Figure 1**B.

Unlike cells producing GFP under the control of the strong T7 promoter (B3<T7-GFP>), induction of the GOI at the beginning of the cultivation led to a considerable reduction in cell growth. In passage 1, cells reached the final biomass of only 1.5 g L^{-1} CDM. However, all 24 T7 strains were able to reach the calculated end biomass of 6 g L^{-1} already in passage 2. This was accompanied by a strong reduction in productivity. As shown in Figure 1A, the first subpopulations of non-producers and weak producers appeared in passage 2 and continued to outgrow producing strains throughout six passages, as shown by derivatives #F6 and #F7. However, in derivatives #B8, #E7, and #F6, we found mixed populations of non-producers, weak producers, and strong producing cells, which could be maintained over the whole period of six passages.

To separate single cells from this mixture, cell suspension after passage 6 was streaked out on agar plates and isolated colonies were analyzed by flow cytometry. We selected 10 clones that formed homogeneous populations covering expression

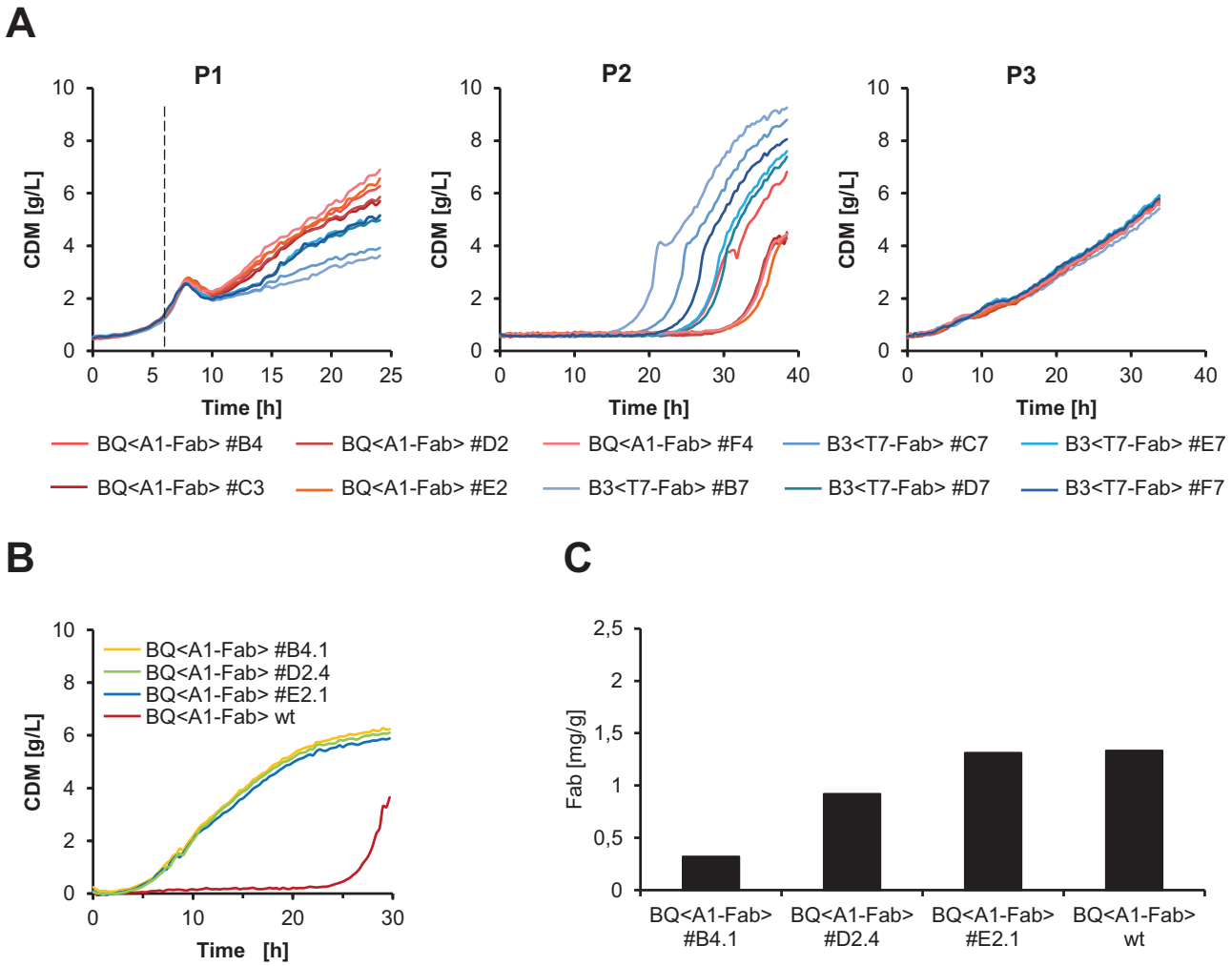


Figure 2. Growth characteristics and product formation of Fab-producing strains. A) Growth behavior of BQ<A1-Fab> (red lines) and B3<T7-Fab> (blue lines) throughout the three passages. The dashed vertical line indicates the time point of induction. B) Cells were fractionated on agar plates and re-tested for growth and productivity. C) The specific concentration of Fab is given in milligrams Fab per gram of cell dry matter (mg g^{-1} CDM).

levels from low (subclone 1, S1) to very high (subclone 10, S10) (Figure 1C).

3.1.2. Evolution of Fab-Producing E. coli Strains

Fab-producing strains were cultivated in the same way as described for the GFP-producing strains, except cells were kept induced for a total of three passages. All Fab-producing clones initially exhibited a long lag phase in passage 2 triggered by the recombinant protein production. After about 20 h, the cells were able to recover, which was evident from a strong increase in the growth rate. In passage 3, the cells finally grew exactly according to the kinetics of the glucose-releasing amylase, indicating unrestricted growth. (**Figure 2**A). Dot blot analysis revealed that none of the 24 T7 clones produced Fab anymore, whereas three A1 clones exhibited the desired properties of fast cell growth and an ability to produce the challenging protein. To ensure homogeneous populations, the three A1 clones were separated on agar plates. Four colonies were picked from each plate to inoculate further cultivation in media containing IPTG to confirm their productivity. As shown in Figure 2B,C, we were able to isolate three subclones (#B4.1, #D2.4, and #E2.1) that produced Fab without reducing cell growth, but in different quantities; under these conditions clone #E2.1 produced comparable amounts of Fab as the non-mutated BQ<A1-Fab> wt strain, whereas the other two subclones exhibited reduced productivity.

3.2. Comparative Genomic Analysis of *E. coli* BQ<A1> and B3<T7> Derivatives

To search for genetic factors involved in the ability to maintain a high growth rate under production conditions, we sequenced the genomes of 11 B3<T7-GFP> and 2 BQ<A1-GFP> derivatives (**Table 1**) and 3 B3<T7-Fab> and 4 BQ<A1-Fab> derivatives (**Table 2**). We found one to three mutations (SNPs or indels) per genome, corresponding to the mutation rate for *E. coli* of ≈1 ×

Table 1. Mutations in GFP-producing derivatives of B3<T7-GFP> and BQ<A1-GFP> isolated after fractionation.

Strain	Clone	Position	Mutation	Region	Gene	Function
B3<T7-GFP>	E7P6 S1	3800855	Δ13 bp	Intergenic	–73 from *gfpmut3.1*	T7 promoter
	A5P6 S1	743064	Δ10 872 bp	Coding	–	Part of λ DE3
	B8P6 S2	43017	GGT → GTT (G98V)	Coding	*caiB*	Type III coA transferase
		80353	T → G	Intergenic	–73 from *setA*	Sugar transporter
		3800868	+TA	Intergenic	–72 from *gfpmut3.1*	T7 promoter
	A5P6 S3	752558	GGA → TGA (G716*)	Coding	T7 RNAP gene	Enterobacteria phage T7 RNAP
	A5P6 S4	753056	TTC → CTC (F882L)	Coding	T7 RNAP gene	Enterobacteria phage T7 RNAP
	A5P6 S5	2407211	CTG → CCG (L256P)	Coding	*ptsI*	Bacterial phosphotransferase system
	B8P6 S6	43017	GGT → GTT (G98V)	Coding	*caiB*	Type III coA transferase
		3800859	+C	Intergenic	–81 from *gfpmut3.1*	T7 promoter
	A5P6 S7	751679	CGC → TGC (R423C)	Coding	T7 RNAP gene	Enterobacteria phage T7 RNAP
		3412013	$(T)_{5\rightarrow4}$	Coding	*malP*	Maltodextrin phosphorylase monomer
	B8P6 S8	43017	GGT → GTT (G98V)	Coding	*caiB*	Type III coA transferase
		2669647	Δ2 bp	Coding	*srlR*	GutR transcriptional repressor
	B8P6 S9	43017	GGT → GTT (G98V)	Coding	*caiB*	Type III coA transferase
		335936	C → T	Intergenic	–104 from *lacZ*	Upstream region of pLac
		3406745	C → T	Intergenic	–168 from *gntT*	*gntT* operator site
	E7P7 S10	335936	C → T	Intergenic	–104 from *lacZ*	Upstream region of *pLac*
BQ<A1-GFP>	E1P6	337099	G → A	Intergenic	–186 from *lacI*	$pLacI^Q$
		3375909	C → T	Intergenic	–168 from *gntT*	*gntT* operator site
	E2P6	337099	G → A	Intergenic	–186 from *lacI*	$pLacI^Q$
		3375909	C → T	Intergenic	–168 from *gntT*	*gntT* operator site

10^{-3} per generation for cells cultured under standard laboratory conditions.[40] Notably, the spontaneous mutation rate is stress-dependent[41] and mutants that result in an increased growth rate under production conditions are more likely to be selected.[42,43] We applied read mapping and a de novo assembly approach to examine SNPs and short indels, as well as to detect larger deletions or subpopulations of genomes.

3.2.1. Genetic Escape Variants of GFP-Producing Derivatives

Out of the 24 BQ<A1-GFP> derivatives isolated after adaptive evolution, the genomes of two randomly selected derivatives were sequenced, since all of them exhibited the same phenotype after six passages. Interestingly, both derivatives exhibited the same mutation within the $lacI^Q$ promoter (Table 1). Next to the introduced C→T substitution[25] within the –35 region of the LacI promoter, we found a G→A transition. However, this mutation does not influence the basal expression or tunability of the A1 expression system (Figure S3, Supporting Information).

Of the 24 B3<T7-GFP> derivatives, we sequenced the genomes of the 11 isolated subclones (Figure 1C). The mutations found in the B3<T7-GFP> derivatives can be divided into three clusters: mutations within the T7 promoter or T7 RNAP gene, mutations within genes involved in the metabolism and transport of various sugars, and mutations within the lac operon. Among the non-producing and weakly producing strains (subclones S1–S4, Figure 1C), mutations within the T7 RNAP gene or the T7 promoter were responsible for the observed phenotype. Derivative #E7P6 S1 had a 13 bp deletion within the T7 promoter encompassing almost the complete T7 promoter sequence. In #A5P6 S1, 10 872 bp of the DE3 lysogen were deleted, harboring the complete sequence of the T7 RNAP gene. In general, no mutational hotspot within the T7 RNAP gene was seen. In each sequence of the B3<T7-GFP> derivatives, which had a mutation in the T7 RNAP, we found a different SNP or indel that caused a loss of function of the T7 RNAP. However, this genetic instability of the T7 RNAP gene and the T7 promoter can be exploited to improve protein production, as is the case with the commercially available C41 strain.[19] Schlegel et al. investigated the underlying mechanism in more detail and concluded that the positive properties of C41 are due to a mutation of the lacUV5 promoter, which controls the expression of the T7 RNAP. A reduction in the T7 RNAP resulted in a reduced total recombinant protein production and thus allowed the production of toxic proteins.[44] This led to the development of the Lemo21 (DE3) expression system.[45,46]

As a second cluster, we identified mutations within genes involved in the metabolism and transport of various sugars. One mutation found in the A1 and T7 expression system was a C→T mutation within the gntT operator site.[47,48] The influence of the gluconate transporter on lactose-inducible gene regulation and its influence on growth during production cannot be fully explained because the medium used for cultivation did not contain any gluconate. Derivative #B8P6 S2 exhibits a T → G transversion

within the promoter sequence of setA. SetA is an efflux pump capable of transporting a range of sugars and sugar analogs. Cells overexpressing SetA exhibit decreased accumulation of lactose and IPTG.[49] We speculate that this mutation increases the SetA expression rate, preventing the intracellular accumulation of IPTG, and reducing the induction of the GOI.

Derivative #A5P6 S5 has a T→C mutation within ptsI, leading to an L256P substitution. PtsI is a cytoplasmic protein that serves as the gateway for the phosphoenolpyruvate/sugar phosphotransferase system (PTS^{sugar}). *E. coli* mutants with < 1% residual PtsI activity are unable to consume the PTS sugars glucose, fructose, mannose, mannitol, sorbitol, *N*-acetylglucosamine, and *N*-acetylmannosamine, and the non-PTS sugars glycerol, melibiose, maltose, and lactose.[50] The L256P substitution may reduce PtsI activity to a level at which glucose can still be used as a substrate, but the intake of IPTG is reduced, which also reduces the induction of the GOI.

In addition to a mutation within the T7 RNAP gene, derivative #A5P6 S7 also exhibited a mutation within malP. The $(T)_{5\rightarrow4}$ mutation causes a frameshift and probably complete loss of function. MalP is involved in maltose metabolism and has a high affinity for short, linear α-1,4 linked oligoglucosides.[51] A mutation within srlR was found in derivative #B8P6 S8. This gene encodes the glucitol repressor GutR and is a DNA-binding transcription factor that represses the gut operon involved in the transport and utilization of glucitol.[52]

A mutation in the type III CoA transferase gene, caiB,[53] was found in four derivatives. It is unlikely that the mutation in caiB influences recombinant protein production or growth behavior, because all B8 derivatives, which can be classified as both strong and weak producers, exhibit this mutation.

Mutations in the third cluster comprise the lac operon itself. Two mutations were found in the pLac promoter in derivatives belonging to strongly producing subclones S9 and S10. The positive properties of mutations in pLac on lac-regulated recombinant gene expression are already known;[19] expression of the lac operon is reduced, including reduced expression of LacY, the sugar transport protein. Consequently, less IPTG enters the cell, which weakens the induction of the GOI.[54] The burden is reduced, explaining why the selected derivatives no longer have a disadvantage in growth. We speculate that this phenomenon also applies to the mutations observed in other genes, such as setA, ptsI, malP, srlR, and gntT, which are involved in the metabolism of various sugars, even though they are not known to be directly related to IPTG or glucose transport into the cell. However, a negative influence on recombinant protein production can only be attributed in the case of ptsI, as this mutation does not occur in combination with a mutation in the T7 promoter or the T7 RNAP gene. Unfortunately, this is not the case with mutations in setA and malP. However, the mutations in srlR and gntT have a positive influence on the growth rate under conditions of recombinant protein production.

3.2.2. Genetic Escape Variants of Fab-Producing Derivatives

In the case of Fab-producing clones, we found only one mutation per sequenced genome, which is attributable to the shorter cultivation time and lower generation number.[40] Also, the production of a challenging protein exerts high selection pressure, which means that single mutations leading to a faster-growing population are selected much faster. As mentioned above, we could not find a B3<T7-Fab> derivative that was able to produce Fab. The three sequenced genomes exemplified that this common phenotype was always due to a mutation in the T7 RNAP gene (Table 2). In contrast, we were able to isolate three BQ<A1-Fab> derivatives capable of producing Fab without limiting growth. Besides, we sequenced the BQ<A1-Fab> non-producer derivative C3. In the latter case, the reason for the lack of production is a mutation in the –35 region of the A1 promoter. In vivo, the promoter function correlates with the degree of sequence homology of –35 and –10 with the consensus sequence of all prokaryotic promoters.[55,56] The σ^{70} factor of the host RNAP forms a specific connection with the helix-turn-helix DNA binding motif of the –35 region. Thus, a mutation in this region could prevent the sigma factor from binding to the promoter. The other three derivatives, B4.1, D2.4, and E2.1, have a mutation in the lac repressor LacI, indicating the reduced productivity and increased growth rate. The reason for the low productivity is the amino acid exchange in LacI, which was also described to occur in Walker strain C43.[19,22] The mutations Q207H, Q146C, and D108E in #B4.1, #D2.1, and #E2.1, respectively, are located in the inducer binding pocket of LacI and, thus, influence the binding affinity for IPTG and/or allolactose[21] (Figure S4, Supporting Information). As a result, cells have tuned themselves to the physiologically tolerated productivity. This seems to be a

Table 2. Mutations in Fab-producing derivatives of B3<T7-Fab> and BQ<A1-Fab> isolated after fractionation.

Strain	Clone	Position	Mutation	Region	Gene	Function
B3<T7-Fab>	C7	752679	CGC→CTC (R756L)	Coding	T7 RNAP gene	Enterobacteria phage T7 RNAP
	D7	752266	Δ11 bp	Coding	T7 RNAP gene	Enterobacteria phage T7 RNAP
	F7	752331	GGG→GAG (G640E)	Coding	T7 RNAP gene	Enterobacteria phage T7 RNAP
BQ<A1-Fab>	B4.1	336293	CAG→CAT (Q207H)	Coding	*lacI*	Lac repressor
	D2.4	336478	GGC→TGC (G146C)	Coding	*lacI*	Lac repressor
	E2.1	336590	GAT→GAG (D108E)	Coding	*lacI*	Lac repressor
	C3	3771324	A → G	Intergenic	–96 from *gfpmut3.1*	pA1

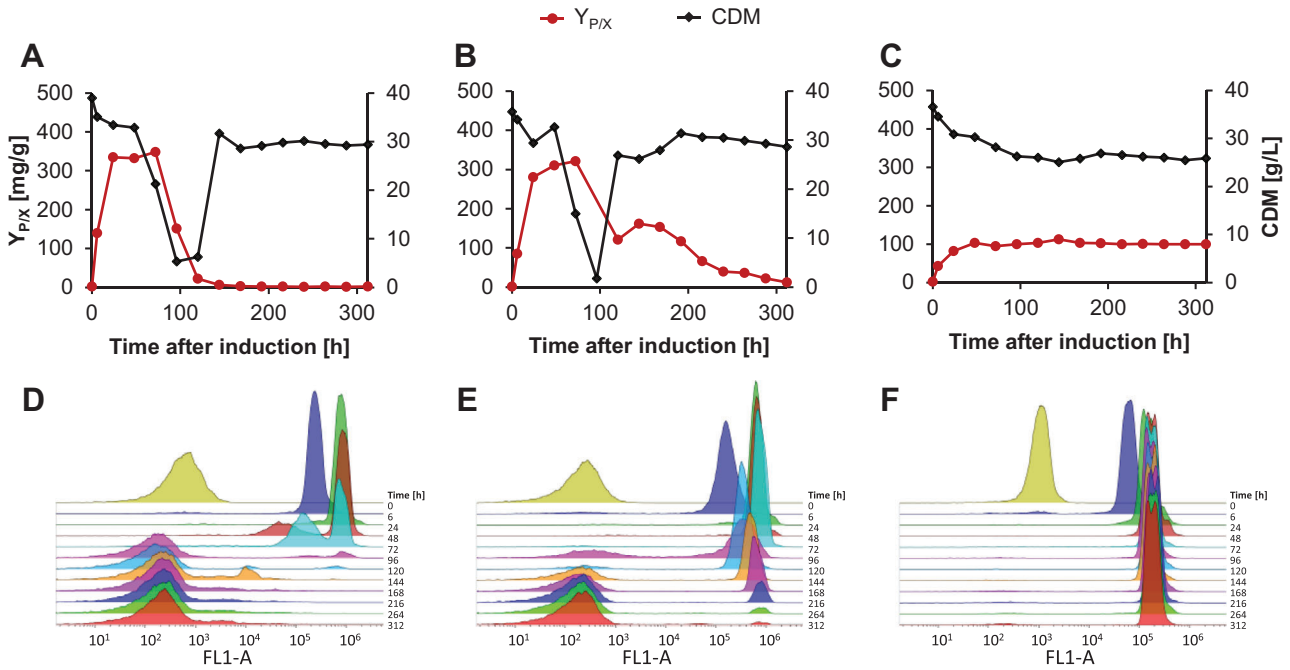


Figure 3. Process characteristics of GFP-producing strains in chemostat culture. A–C) Courses of the specific content of GFP ($Y_{P/X}$) and biomass. D–F) Single-cell expression analysis. A,D) B3<T7-GFP> wt, B,E) B3<T7-GFP> #E7P6 S10, and C,F) BQ<A1-GFP> wt.

universal phenomenon for lac-regulated gene expression. Under completely different experimental conditions, a similar result was achieved in Walker strain C43. Miroux and Walker produced mutants by spreading cells on IPTG-containing agar. Although the cells were cultured without process control, and despite the presence of the DE3 lysogen, the same type of variations developed in two independent experiments. However, Miroux and Walker's experiments did not focus on long-term stability, but on the productivity of toxic membrane proteins. Based on our results, we conclude that T7 RNAP-dependent expression systems acquire at least one mutation in the T7 RNAP gene and/or T7 promoter in long-term cultivation, regardless of the protein to be produced.

3.3. Long-Term Stability of Select Derivatives in Chemostat Cultivations

The derivatives described above were generated under carbon-limited fed-batch-like production conditions in the microbioreactor. This raised the question about how to select derivatives that can grow unhindered despite recombinant protein production behave in long-term cultivations and whether they can continue to maintain productivity over several generations.

To address this question, we performed comparative lab-scale chemostat cultivations. In the case of the easy-to-produce protein GFP, the best mutant B3<T7-GFP> #E7P6 S10, which had a mutation in the pLac promoter, was compared to the non-mutated wild-type strains BQ<A1-GFP> wt and B3<T7-GFP> wt in long-term chemostat cultivations.

For B3<T7-GFP> wt, induction with IPTG resulted in a very high specific content of 348 mg GFP g^{-1} CDM (**Figure 3**A), which was higher than the product titer from a fed-batch culture with the same final biomass of 30 g L^{-1} (**Figure 4**A and **Table 3**). This productivity could be maintained for 72 h after induction. An extreme decrease was observed in the biomass and product titer. This phenomenon was demonstrated in previous studies in which we also observed a decrease in the product titer 60 h after induction.[15] After ≈120 h, the biomass recovered and increased to the intended 30 g L^{-1} CDM, but without product formation. The specific GFP content remained at 0 mg g^{-1} until the end of the fermentation. As shown in Figure 3D, non-producers had completely asserted themselves, and the results of the microtiter experiments were reproduced. Similar behavior was observed with the derivative #E7P6 S10 (Figure 3B), with an extreme decrease in the biomass to almost 0 g L^{-1} after 72 h in chemostat mode. After 120 h, the cells recovered to the intended biomass of 30 g L^{-1} CDM. Unlike its wild-type ancestor, the specific GFP content did not decrease to zero but decreased steadily until the end of the fermentation. The peak at a fluorescence intensity of 10^2 in the FL1-A channel (Figure 3E) suggests that a mixed population emerged in which non-producers have gradually overgrown the producing cells, causing the production to decrease throughout the cultivation.

By determining the T7 RNAP gene sequence via PCR amplification of the corresponding chromosomal region and Sanger sequencing, we identified the insertion element insD-3 within the T7 RNAP gene of B3<T7-GFP> wt and a mutation within the lacUV5 promoter in derivative #E7P6 S10, which may have resulted in the complete loss of GFPmut3.1 productivity.

In contrast, no extreme decrease in biomass was observed for the host RNAP-dependent expression system BQ<A1-GFP> wt, indicating extraordinarily high stability (Figure 3C). Over the entire cultivation period of 312 h and 45 generations, the specific

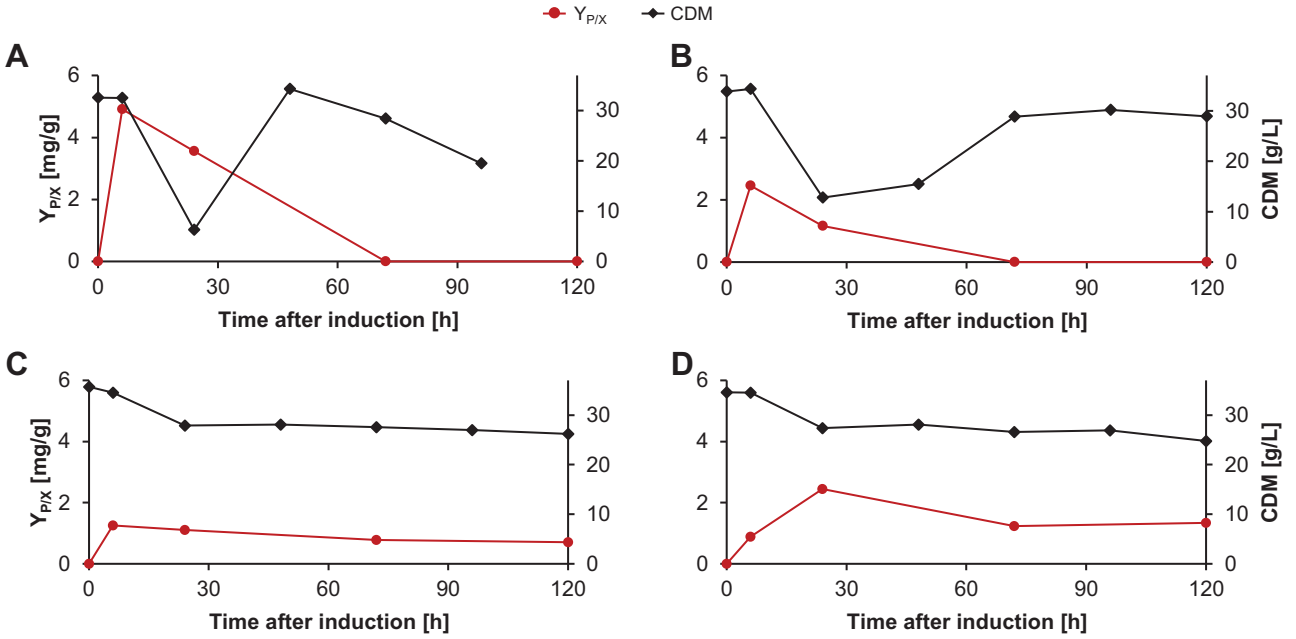


Figure 4. Process characteristics of Fab-producing strains in chemostat culture. Courses of the specific content of Fab ($Y_{P/X}$) and biomass. A) B3<T7-Fab> wt, B) BQ<A1-Fab> wt, C) BQ<A1-Fab> #D2.4, and D) BQ<A1-Fab> #E2.1.

Table 3. Comparison of productivity between fed-batch and chemostat cultivations (final biomass 30 g L^{-1} CDM).

Strain	Cultivation mode	Maximum volumetric yield[a] [g L^{-1}]	Process time[b] [h]	STY[c] [mg L^{-1} h^{-1}]
B3<T7-GFP> wt	Fed-batch	8.3 ± 0.6	29	286.2
B3<T7-GFP> wt	Chemostat	10.9	60[d]	847.1
BQ<A1-GFP> wt	Fed-batch	1.8 ± 0.2	29	62.8
BQ<A1-GFP> wt	Chemostat	2.6	324	284.1
B3<T7-GFP> #E7P6 S10	Chemostat	10.2	60[d]	642.2
B3<T7-Fab> wt	Fed-batch	0.052 ± 0.006	29	1.8
B3<T7-Fab> wt	Chemostat	0.164	18	3.4
BQ<A1-Fab> wt	Fed-batch	0.087 ± 0.008	29	3.0
BQ<A1-Fab> wt	Chemostat	0.085	18	1.8
BQ<A1-Fab> #E2.1	Chemostat	0.035	132	4.3
BQ<A1-Fab> #D2.4	Chemostat	0.019	132	2.3

[a] For fed-batch cultivations (Figure S5, Supporting Information), the mean ± standard deviation for individual determinations is given ($n = 3$) [b] Process time is the sum of reactor setup and the actual fermentation [c] Space time yield (STY) calculated according to Equation (1) [d] Chemostat cultivations in which the cultivation period was limited to the phase before population collapse.

product content remained constant at ≈100 mg GFP g^{-1} CDM (Figure 3C). Flow cytometry confirmed the process stability. A homogenous population was confirmed at each time point (Figure 3F). The lower productivity of this expression system in combination with GFP represents such a low burden on the host's physiology that there was no population collapse. Potential mutants did not have a growth advantage, which means that the population of producing cells could persist for a period of 45 generations.

Long-term chemostat cultivation with Fab was performed with two BQ<A1-Fab> variants, #D2.4 and #E2.1 (Figure 4C,D). The wild-type non-mutated strains B3<T7-Fab> wt and BQ<A1-Fab> wt served as reference (Figure 4A,B).

Unlike GFP, the wild-type strain BQ<A1-Fab> wt exhibited the same behavior as the strong T7 expression system B3<T7-Fab> wt. In both wild-type strains, a decrease in the biomass was already observed 6 h after induction and concomitant reduction of the product. The maximum amount of product was 4.9 and 2.5 mg g^{-1} for B3 <T7-Fab> wt and BQ <A1-Fab> wt, respectively. This corresponds to the maximum measured product titers from fed-batch cultivations (Figure 4C,D and Table 3). Based on the results obtained for GFP, this behavior was to be expected with the T7 expression system. Interestingly, the phenomenon of the population collapse was observed using the BQ<A1> expression system. As mentioned above, the physiological burden on the host depends not only on the expression system but also on the

specific recombinant protein to be produced. In the case of Fab, the expression of even the weaker promoter in BQ<A1-Fab> wt was too strong. Sanger sequencing of the A1 promoter again revealed an A → G substitution within the –35 region. In contrast, no notable decrease in biomass was observed in the two A1 derivatives. Derivative #D2.4 was able to consistently produce ≈1.0 mg Fab g^{-1} CDM over a total of 120 h, which corresponds to 17.3 generations. Derivative #E2.1 produced even more at 1.6 mg g^{-1}. Nevertheless, long-term stability was accompanied by a reduction in productivity. Compared to the conventional fed-batch processing, in which BQ<A1-Fab> wt produced 2.7 mg g^{-1} (Figure 4D), the product titers of the corresponding derivatives were lower, approximately half in the case of the best producing mutant #E2.1.

If long-term stability is only possible through reduced productivity, the question of process efficiency arises. To compare fed-batch and continuous production, we used the space-time-yield (STY; Equation (1)) as the evaluation criterion

$$\mathrm{STY} = \frac{\dfrac{\text{Total protein (mg)}}{\text{Reactor volume (L)}}}{\text{Process time (h)}} \quad (1)$$

For calculation of the processing time for both process types, we added a downtime of 12 h for media preparation, reactor setup, CIP, and SIP to the cultivation time. In chemostat cultivations with population collapse, we only considered the cultivation time before the decrease in cell density, which was the case with B3 <T7-GFP> after 48 h.

As can be concluded from the productivity comparisons (Table 3), the fed-batch process with the T7 expression system and the model protein GFP were inferior to the chemostat process in terms of volumetric yield and STY. The reason for this can be attributed to the duration of the induction. In fed-batch cultivation, the production phase takes 19 h. Chemostat cultivation can be extended to 48 h because the volume in the reactor is kept constant.

As already mentioned, a longer chemostat process would not be possible under these circumstances because the process ultimately becomes unstable. However, short-term chemostat cultivation of 48 h could be interesting for easy-to-produce proteins. With 874 mg L^{-1} h^{-1}, the STY of B3<T7-GFP>wt in the chemostat tripled compared to the fed-batch cultivation. In the case of the host RNAP-dependent A1 expression system, the STY of 62.8 mg L^{-1} h^{-1} in fed-batch cultivation was relatively low compared to the STY of 286 mg L^{-1} h^{-1} of the conventional T7 expression system. In addition to the short production time of 19 h, the A1 expression system was inherently weaker. However, the low titer can be compensated by continuous production. Due to the very long stable production time of 312 h, the STY of the chemostat cultivation more than quadrupled compared to the fed-batch cultivation. Nevertheless, for the production of GFP, the conventional T7 expression system is still superior overall, in both fed-batch and chemostat modes, especially when considering short-term chemostat cultivation of the T7 wild-type strain (Table 3).

The situation is different with Fab. The extraordinary strength of the T7 expression system is a disadvantage for Fab production. In neither fed-batch nor chemostat mode, the T7 expression system was able to outperform the host RNAP-dependent A1 expression systems. Due to the mutation in lacI, the volumetric yield of BQ<A1-Fab>wt was reduced from 0.087 to 0.035 g L^{-1}, as seen with derivative #E2.1. However, long-term cultivation without population collapse was possible. Due to the long production time of 120 h, the STY of the BQ<A1-Fab>wt was increased from 3.00 g L^{-1} h^{-1} in fed-batch mode to 4.33 g L^{-1} h^{-1} in chemostat mode for derivative #E2.1 when comparing the two strains to one another.

These results indicate that, in continuous *E. coli* bioprocesses, the choice of expression system depends on the recombinant protein to be produced. In the case of an easy-to-produce protein, such as GFP, the conventional T7 expression system is still the expression system of choice. The duration of the chemostat cultivation is limited by the inevitable population collapse, but this can be compensated by the extraordinary strength of the T7 system.

However, the moderately strong host RNAP-dependent A1 system is more suitable for the production of a challenging protein. As a result of the large physiological burden that is triggered, a reduction in the expression rate is advantageous. Thus, the low titer can not only be compensated but even increased based on the STY. Furthermore, there is still potential to increase yields through various cultivation conditions. So far, all chemostat experiments described herein have been carried out at a dilution rate of $D = 0.1$ 1 h^{-1} and 30 °C. Variations in the growth rate and temperature could further increase the yields also in continuous production, as has already been shown for fed-batch cultivations.[57–59]

As already mentioned, the time in which the actual fermentation occurs and recombinant protein is produced in fed-batch cultivation is relatively low. Also, the cost of CIP and SIP must be included. Although these process steps must also be carried out in long-term chemostat cultivation, the phase in which the actual production occurs is inevitably prolonged, which means that the CIP and SIP frequency can be reduced.

Nevertheless, only mutants that have adapted to long-term stability under production conditions can be cultivated in long-term chemostat mode. In wild-type strains, the population collapses, and adaptive rescue during long-term chemostat cultivation does not allow long-term production. This phenomenon has been observed frequently in other *E. coli* chemostat cultivations and prevents the direct use of non-adapted strains for this purpose.[5,60,61]

4. Concluding Remarks

Identification of genetically stable *E. coli* mutants using high-throughput serial fed-batch microtiter cultivations was successfully implemented. Here, we reported a selection-based approach based on the growth rate under production conditions.[23,24] As expected, the genetic stability of a T7 expression system is not sufficient to maintain constant product formation levels during long-term cultivation under production conditions. Mutations leading to non-producers are located in the T7 RNAP gene and/or the T7 promoter. The host RNAP-based A1 expression system, which has moderate expression strength, remained stable in the production of the easy-to-produce protein GFP during long-term cultivation. For the production of challenging proteins (e.g., Fabs), which trigger a more severe burden on cell physiology, mutations in lacI of BQ<A1-Fab> derivatives reduce the

expression levels but have positive effects on long-term stability. We could not find any mutations with positive effects on protein expression, such as mutations in folding helpers, the Sec translocon, or proteases that could serve as modification targets in a rational approach. This finding was not unexpected, as no selection pressure was applied to the production of the recombinant protein.

In addition to obvious mutations in the T7 RNAP gene and/or T7 promoter that led to reduced burden and higher growth rates, we were able to find mutations in the metabolism and transport mechanism of various sugars. These mutations led most to decreased inducer uptake and, thus, reduced induction. However, three mutants were found with unrestricted growth and stable Fab expression at a reduced, but physiologically acceptable, level. Hence, for continuous recombinant protein production in *E. coli*, new production clones can be generated much faster and more efficiently than with a rational approach. The production host regulates itself to the physiologically tolerable level of recombinant protein production, which can differ from protein to protein and is not predictable. Rational methods such as inductor concentration variation, promoter engineering, or interventions in the regulation of recombinant protein production require labor-intensive and time-consuming experimental designs and cannot guarantee success.

Regarding industrial regulations, the U.S. FDA regulatory body encourages the biological industries to use continuous manufacturing approaches for the production of new products. This aspect applies to the entire bioprocess, including downstream processing. Thus far, continuous downstream processing has been more relevant to mammalian perfusion processes.[12] However, with the production strains described in this study, *E. coli* processes could also be relevant.

In conclusion, we have shown that, through an adaptive evolution approach, a high-throughput screening process comes very close to industrial production processes, and we were able to find derivatives that have a positive effect on long-term stability. We postulate that long-term stability studies with *E. coli* can only be carried out with genome-integrated expression systems. As plasmid loss is no longer a problem in this regard, we were able to characterize how metabolic load triggered by recombinant protein production influences the characteristics of mutations in *E. coli*. Thus, adaptive evolution in microtiter cultivations could be an efficient strain development method in addition to a rational approach. Although the specific product titers are reduced, they can be compensated by continuous production.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors thank Philipp Peters for data analysis at an early stage of the project and Nancy Stralis-Pavese for the preparation of the sequencing libraries. Some calculations were performed at the Vienna Scientific Cluster (VSC). Illumina sequencing was performed at the Vienna BioCenter Core Facilities (VBCF) Next Generation Sequencing Facility. The financial support by the Austrian Federal Ministry for Digital and Economic Affairs, the National Foundation for Research, Technology, and Development, and the Christian Doppler Research Association is gratefully acknowledged. Furthermore, the authors gratefully acknowledge the financial support received from Boehringer Ingelheim RCV GmbH & Co KG. The funding agencies had no influence on the conduct of this research. The authors thank the colleagues of their company partner BI RCV for the scientific input and their continuous scientific support.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

A.S., M.C.-P., R.G., J.C.D., H.H., and G.S. designed the research; A.S., C.K., and M.S. performed the research; A.S., M.C.-P., R.G., M.W., J.C.D., M.S., H.H., and G.S. analyzed the data; A.S. and G.S. wrote the article, with contributions from M.S., J.C.D., M.C.-P., and M.W. All authors approved the final version of the manuscript.

Data Availability Statement

The data that support the findings of this study are available in the Supporting Information of this article.

Keywords

adaptive evolution, challenging proteins, chemostat cultivation, *E. coli*, Fab fragments, genetic stability, long-term stability

Received: August 10, 2020
Revised: October 6, 2020
Published online:

[1] J. Walther, R. Godawat, C. Hwang, Y. Abe, A. Sinclair, K. Konstantinov, *J. Biotechnol.* **2015**, *213*, 3.
[2] N. Ziv, N. J. Brandt, D. Gresham, *J. Visualized Exp.* **2013**, *80*, e50168.
[3] K. Peebo, P. Neubauer, *Microorganisms* **2018**, *6*, 56.
[4] S. J. Yeom, D. H. Lee, Y. J. Kim, J. Lee, K. K. Kwon, G. H. Han, H. Kim, H. S. Kim, S. G. Lee, *PLoS One* **2016**, *11*, e0166890.
[5] J. Kopp, A.-M. Kolkmann, P. G. Veleenturf, O. Spadiut, C. Herwig, C. Slouka, *Front. Bioeng. Biotechnol.* **2019**, *7*, 297.
[6] A. Schmideder, D. Weuster-Botz, *J. Ind. Microbiol. Biotechnol.* **2017**, *44*, 1021.
[7] F. W. Studier, B. A. Moffatt, *J. Mol. Biol.* **1986**, *189*, 113.
[8] S. B. Noronha, H. J. C. Yeh, T. F. Spande, J. Shiloach, *Biotechnol. Bioeng.* **2000**, *68*, 316.
[9] G. L. Rosano, E. A. Ceccarelli, *Front. Microbiol.* **2014**, *5*, 172.
[10] O. Zelder, B. Hauer, *Curr. Opin. Microbiol.* **2000**, *3*, 248.
[11] P. Rugbjerg, N. Myling-Petersen, A. Porse, K. Sarup-Lytzen, M. O. A. Sommer, *Nat. Commun.* **2018**, *9*, 787.
[12] P. Rugbjerg, M. O. A. Sommer, *Nat. Biotechnol.* **2019**, *37*, 869.
[13] B. R. Glick, *Biotechnol. Adv.* **1995**, *13*, 247.
[14] L. De Gelder, J. M. Ponciano, P. Joyce, E. M. Top, *Microbiology* **2007**, *153*, 452.
[15] G. Striedner, I. Pfaffenzeller, L. Markus, S. Nemecek, R. Grabherr, K. Bayer, *Biotechnol. Bioeng.* **2010**, *105*, 786.
[16] B. Csorgo, T. Feher, E. Timar, F. R. Blattner, G. Posfai, *Microb. Cell Fact.* **2012**, *11*, 11.

[17] D. E. Deatherage, D. Leon, A. E. Rodriguez, S. K. Omar, J. E. Barrick, *Nucleic Acids Res.* **2018**, *46*, 9236.
[18] A. Schuller, M. Cserjan-Puschmann, C. Tauer, J. Jarmer, M. Wagenknecht, D. Reinisch, R. Grabherr, G. Striedner, *Microb. Cell Fact.* **2020**, *19*, 58.
[19] B. Miroux, J. E. Walker, *J. Mol. Biol.* **1996**, *260*, 289.
[20] J. E. Barrick, D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, J. F. Kim, *Nature* **2009**, *461*, 1243.
[21] S. K. Kwon, S. K. Kim, D. H. Lee, J. F. Kim, *Sci. Rep.* **2015**, *5*, 16076.
[22] L. Dumon-Seignovert, G. Cariot, L. Vuillard, *Protein Expression Purif.* **2004**, *37*, 203.
[23] R. P. Maharjan, T. Ferenci, *PLoS Biol.* **2017**, *15*, e2001477.
[24] R. P. Maharjan, T. Ferenci, *Environ. Microbiol. Rep.* **2018**, *10*, 626.
[25] C. B. Glascock, M. J. Weickert, *Gene* **1998**, *223*, 221.
[26] S. K. Sharan, L. C. Thomason, S. G. Kuznetsov, D. L. Court, *Nat. Protoc.* **2009**, *4*, 206.
[27] B. P. Cormack, R. H. Valdivia, S. Falkow, *Gene* **1996**, *173*, 33.
[28] M. Fink, S. Vazulka, E. Egger, J. Jarmer, R. Grabherr, M. Cserjan-Puschmann, G. Striedner, *Biotechnol. J.* **2019**, *14*, 1800637.
[29] C. Toeroek, M. Cserjan-Puschmann, K. Bayer, G. Striedner, *SpringerPlus* **2015**, *4*, 490.
[30] F. W. Studier, *Protein Expression Purif.* **2005**, *41*, 207.
[31] H. Reischer, I. Schotola, G. Striedner, F. Potschacher, K. Bayer, *J. Biotechnol.* **2004**, *108*, 115.
[32] F. He, *Bio-Protoc.* **2011**, *1*, e97.
[33] A. M. Bolger, M. Lohse, B. Usadel, *Bioinformatics* **2014**, *30*, 2114.
[34] H. Jeong, H. J. Kim, S. J. Lee, *Genome Announce.* **2015**, *3*, e00134.
[35] D. E. Deatherage, J. E. Barrick, *Methods Mol. Biol.* **2014**, *1151*, 165.
[36] D. Kim, B. Langmead, S. L. Salzberg, *Nat. Methods* **2015**, *12*, 357.
[37] R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W. Cheung, S. M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T. W. Lam, J. Wang, *Gigascience* **2012**, *1*, 18.
[38] P. J. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M. J. de Hoon, *Bioinformatics* **2009**, *25*, 1422.
[39] S. Kurtz, A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L. Salzberg, *Genome Biol.* **2004**, *5*, R12.
[40] H. Lee, E. Popodi, H. Tang, P. L. Foster, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E2774.
[41] R. Maharjan, T. Ferenci, *Mol. Biol. Evol.* **2015**, *32*, 380.
[42] P. L. Foster, *BioEssays* **2000**, *22*, 1067.
[43] D. Choe, J. H. Lee, M. Yoo, S. Hwang, B. H. Sung, S. Cho, B. Palsson, S. C. Kim, B.-K. Cho, *Nat. Commun.* **2019**, *10*, 935.
[44] S. Schlegel, P. Genevaux, J.-W. de Gier, *Cell Rep.* **2015**, *10*, 1758.
[45] S. Schlegel, E. Rujas, A. J. Ytterberg, R. A. Zubarev, J. Luirink, J.-W. de Gier, *Microb. Cell Fact.* **2013**, *12*, 24.
[46] S. Schlegel, J. Löfblom, C. Lee, A. Hjelm, M. Klepsch, M. Strous, D. Drew, D. J. Slotboom, J.-W. de Gier, *J. Mol. Biol.* **2012**, *423*, 648.
[47] A. Porco, N. Peekhaus, C. Bausch, S. Tong, T. Isturiz, T. Conway, *J. Bacteriol.* **1997**, *179*, 1584.
[48] N. Peekhaus, T. Conway, *J. Bacteriol.* **1998**, *180*, 1777.
[49] J. Y. Liu, P. F. Miller, M. Gosink, E. R. Olson, *Mol. Microbiol.* **1999**, *31*, 1845.
[50] R. D. Simoni, S. Roseman, M. H. Saier Jr., *J. Biol. Chem.* **1976**, *251*, 6584.
[51] R. Schinzel, B. Nidetzky, *FEMS Microbiol. Lett.* **1999**, *171*, 73.
[52] M. Yamada, M. H. Saier, *J. Mol. Biol.* **1988**, *203*, 569.
[53] T. Elssner, C. Engemann, K. Baumgart, H. P. Kleber, *Biochemistry* **2001**, *40*, 11140.
[54] A. Fernandez-Castane, C. E. Vine, G. Caminal, J. Lopez-Santin, *J. Biotechnol.* **2012**, *157*, 391.
[55] D. A. Siegele, J. C. Hu, W. A. Walter, C. A. Gross, *J. Mol. Biol.* **1989**, *206*, 591.
[56] T. Gardella, H. Moyle, M. M. Susskind, *J. Mol. Biol.* **1989**, *206*, 579.
[57] C. Turner, M. E. Gregory, M. K. Turner, *Biotechnol. Lett.* **1994**, *16*, 891.
[58] A. M. Sandén, I. Prytz, I. Tubulekas, C. Förberg, H. Le, A. Hektor, P. Neubauer, Z. Pragai, C. Harwood, A. Ward, A. Picon, J. T. de Mattos, P. Postma, A. Farewell, T. Nyström, S. Reeh, S. Pedersen, G. Larsson, *Biotechnol. Bioeng.* **2003**, *81*, 158.
[59] M. Gutiérrez-González, C. Farías, S. Tello, D. Pérez-Etcheverry, A. Romero, R. Zúñiga, C. H. Ribeiro, C. Lorenzo-Ferreiro, M. C. Molina, *Sci. Rep.* **2019**, *9*, 16850.
[60] J. M. Couto, A. McGarrity, J. Russell, W. T. Sloan, *Microb. Cell Fact.* **2018**, *17*, 8.
[61] N. Rai, L. Huynh, M. Kim, I. Tagkopoulos, *Biotechnol. Bioeng.* **2019**, *116*, 693.

Publication II

Supporting Information

Supporting Information

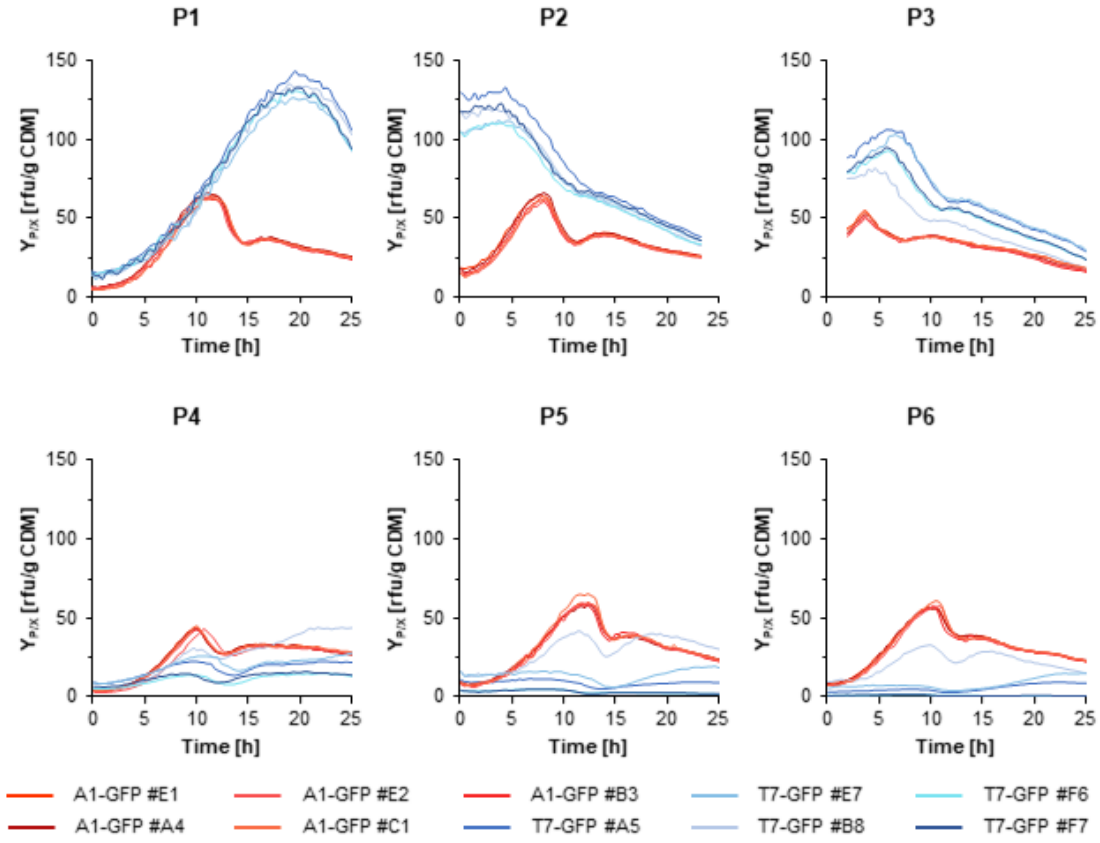


Figure S1. Specific fluorescence of the easy-to-produce protein GFPmut3.1 over the course of 6 passages. The specific productivity of BQ<A1-GFP> derivatives (reddish lines) and B3<T7-GFP> (bluish lines) is given in relative fluorescence units per g of cell dry matter [rfu/g CDM].

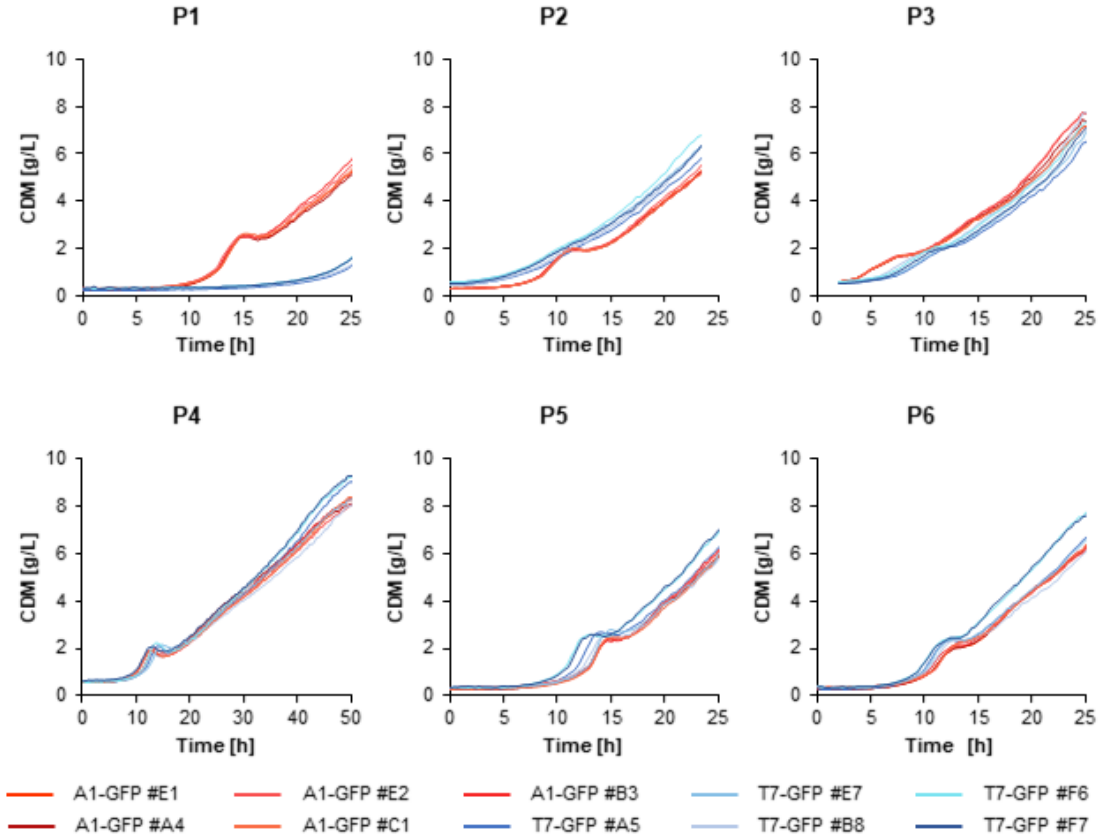


Figure S2. Growth characteristics of GFPmut3.1 producing cells over the course of 6 passages. Biomass trends of BQ<A1-GFP> derivatives (reddish lines) and B3<T7-GFP> (bluish lines) are given in g/L cell dry matter (CDM).

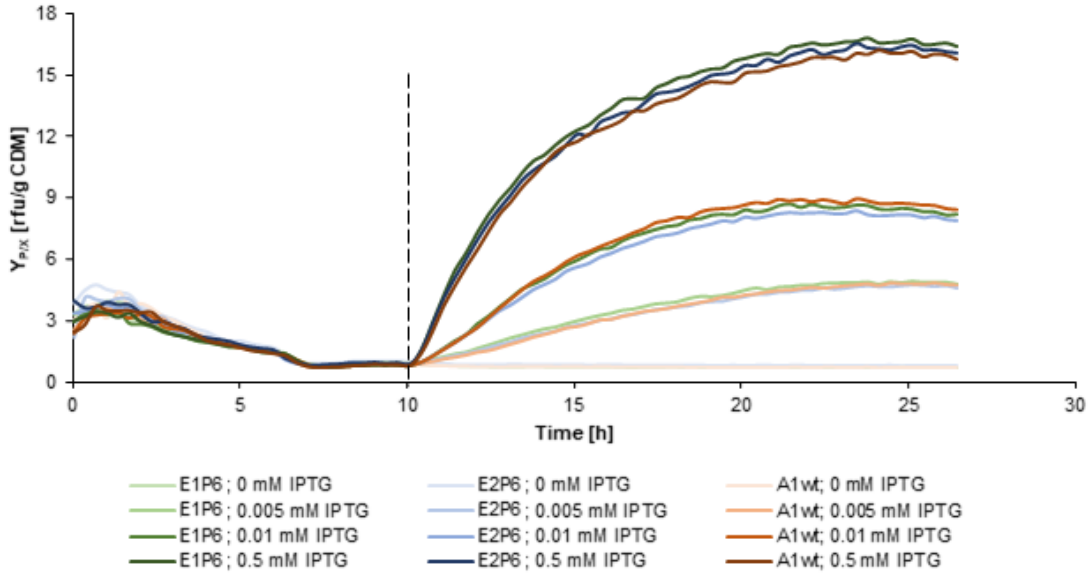


Figure S3. Influence of mutations in the lacI^{Q} promoter of BQ<A1-GFP> derivatives E1 (green) and E2 (blue) on expression rate control. Derivatives were compared to their

unmated ancestor BQ<A1-GFP> wt (orange) The dashed vertical line indicates the time of induction. Induction was performed with 0, 0.005, 0.01, or 0.5 mM IPTG

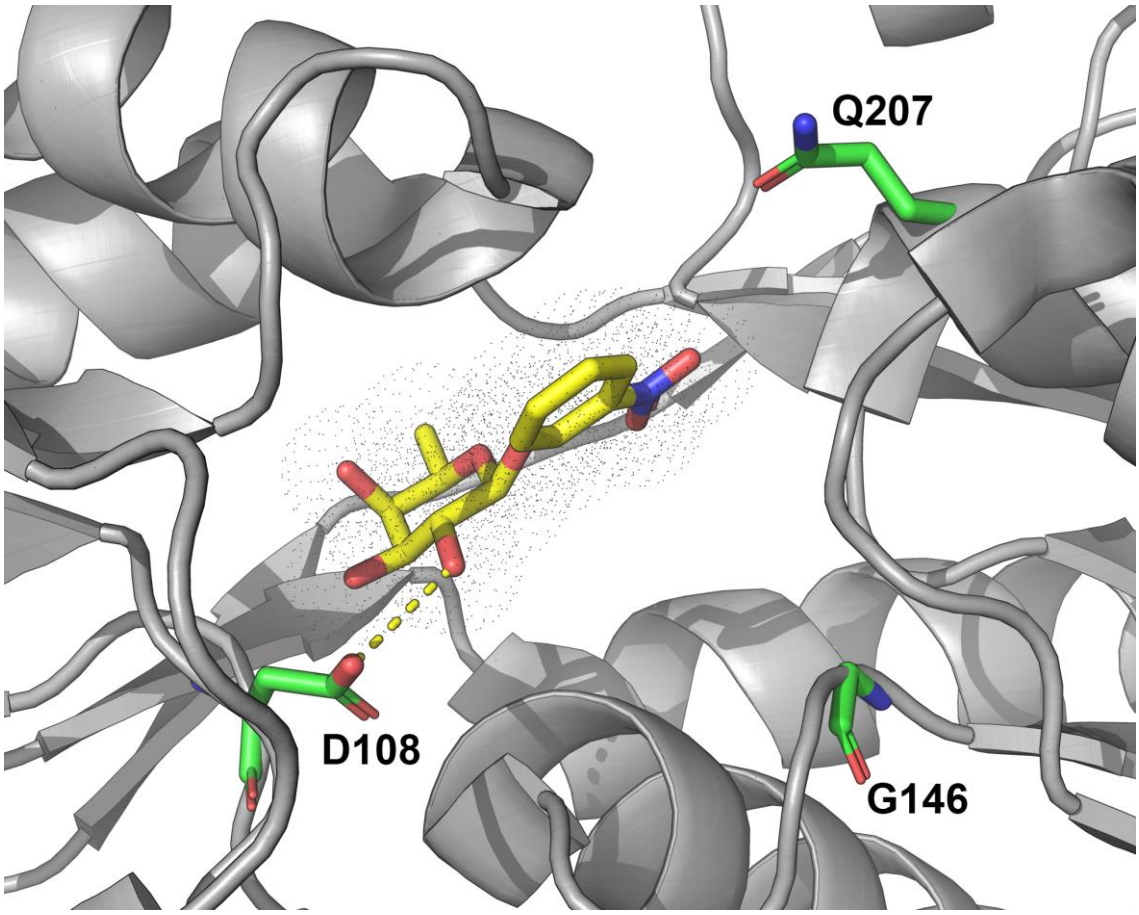


Figure S4. Structural analysis of the Q207, G146 and D108 amino acids in the LacI protein. The model shows the spatial proximity of the three mutated amino acids (green colored carbon) to the inducer IPTG (yellow colored carbon). The model was illustrated with PyMOL.

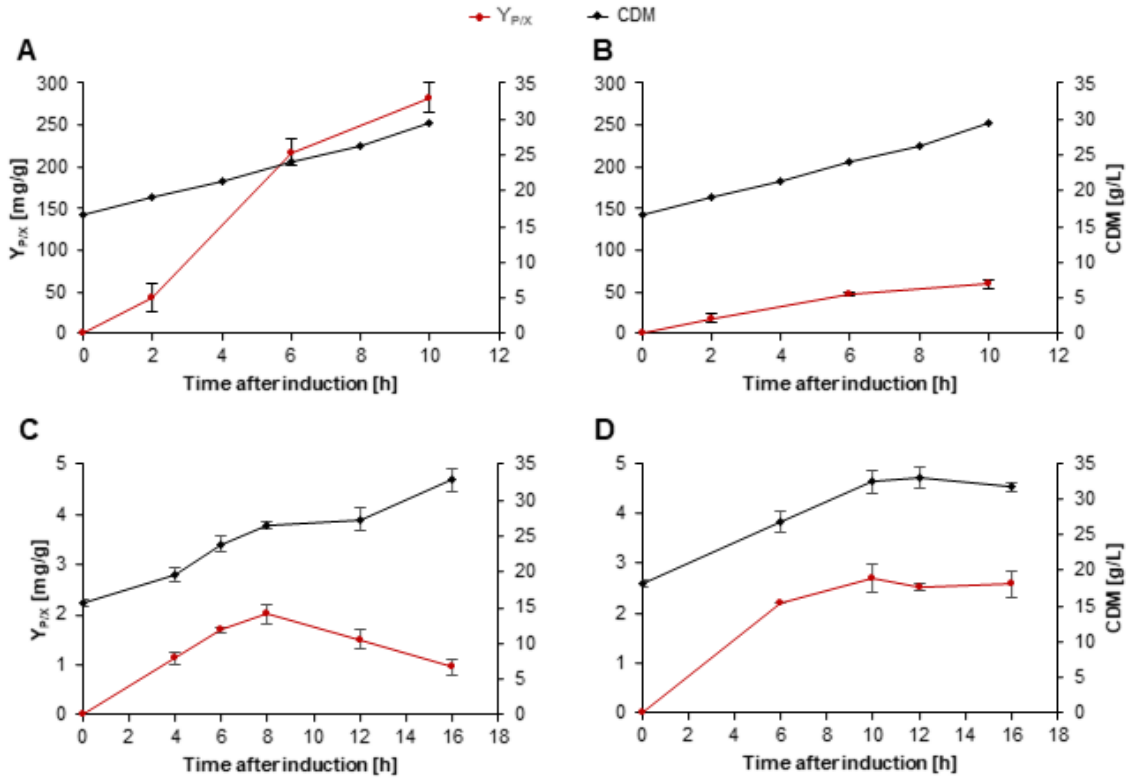


Figure S5. Process characteristics of GFP and Fab producing strains in fed-batch culture operated at μ = 0.1 /h and an intended final biomass of 30 g/L. (A-B) Courses of the specific content of GFP content ($Y_{P/X}$) and biomass (g/L). (C-D) Courses of the specific content of Fab content ($Y_{P/X}$) and biomass (g/L). (A) B3<T7-GFP> wt, (B) BQ<A1-GFP> wt, (C) B3<T7-Fab> wt, (D) BQ<A1-Fab> wt. The mean values and standard deviations for individual determinations are shown (n=3).

8 Appendix

Patent WO2020053285A1

Inducible expression system for plasmid-free production of a protein of interest

Inventor: Artur Schuller, Monika Cserjan, Reingard Grabherr, Gerald Striedner

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 March 2020 (19.03.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/053285 A1

(51) International Patent Classification:
C12N 15/63 (2006.01) *Cl 2N 15/70* (2006.01)

(21) International Application Number:
PCT/EP20 19/074239

(22) International Filing Date:
11 September 2019 (11.09.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
18193655.0 11 September 2018 (11.09.2018) EP

(71) Applicant: **BOEHRINGER INGELHEIM RCV GMBH & CO KG** [AT/AT]; Dr. Boehringer Gasse 5-11, 1121 Vi¬enna (AT).

(72) Inventors: **SCHULLER, Artur;** Beingasse 30/7, 1150 Vi¬enna (AT). **CSERJAN, Monika;** UniversitatfurBodenkultur Wien, Department ftir Biotechnologie, Muthgasse 18, 1190 Vienna (AT). **GRABHERR, Reingard;** Universitat fur Bodenkultur Wien, Department ftir Biotechnologie, Muthgasse 18, 1190 Vienna (AT). **STRIEDNER, Gerald;** Haberlandtgasse 83/1, 1220 Vienna (AT).

(74) Agent: **LOIDL, Manuela** et al.; Donau-City-Strasse 11, 1220 Vienna (AT).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, Cl, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

WO 2020/053285 A1

(54) Title: INDUCIBLE EXPRESSION SYSTEM FOR PLASMID-FREE PRODUCTION OF A PROTEIN OF INTEREST

(57) Abstract: A genome-based expression system for production of a protein of interest (POI) in a prokaryotic host, comprising at least an RNA polymerase (RNAP) gene, a gene encoding a POI, comprising a coding sequence, a promoter operably linked to said coding sequence, wherein said promoter is recognized by the RNAP expressed from the RNAP gene, and at least one lac operator (lacO) within the sequence of said promoter; and a lacI gene encoding a lac repressor protein (LacI) comprising a coding sequence, a lacI promoter operably linked to the lacI coding sequence, wherein the lacI promoter is a wild-type lacI promoter or a lacI promoter which increases LacI expression; wherein the expression rate of the POI is regulated by an inducer binding LacI.

INDUCIBLE EXPRESSION SYSTEM FOR PLASMID-FREE PRODUCTION OF A PROTEIN OF INTEREST

FIELD OF THE INVENTION

The invention relates to the field of plasmid-free inducible systems for expression of a protein of interest in a prokaryotic host. It further relates to methods of using such systems for the production of a protein of interest in a prokaryotic host.

BACKGROUND OF THE INVENTION

In industrial protein production processes, gene regulation is an important prerequisite. Transcription rates are controlled by the interaction of a promoter and the RNA polymerase (RNAP). Understanding and external regulation of this interaction is necessary to provide process control and optimization of product yield and quality. A reduced promoter strength can be beneficial, especially for challenging proteins, like antibody fragments, membrane proteins or toxic proteins (1-3). The final product yield of soluble and proper folded proteins is often not directly determined by the strength of the promoter system but by further processing of the peptide chains, like translocation into the periplasm and proper disulfide bond formation.

The most prominent and well-studied genetic regulatory mechanism is the lac operon (4). In wild-type *E. coli, the* /ac-inhibitor (Lad) forms a homo-tetramer that binds to the /ac-operator sequences (lacO) and represses the transcription of the *lacZYA* operon (5). In the presence of lactose or the non-metabolizable isopropyl β-D-l-thiogalactopyranoside (IPTG), Lad changes in structure and can no longer bind to the lac-operator, resulting in induction of transcription. The lac-operator sites are DNA sequences with inverted repeat symmetry (6).

The higher the symmetry, the greater the binding affinity of Lad to the operator sequence. An artificial perfectly symmetric lacO (sym-lacO) was found to bind Lad with the greatest affinity (7), whereas the three wild-type operators Iac01 , Iac02 and Iac03 exhibiting an approximate symmetry showed lower affinities, resulting in the following order: sym-lacO> Iac01 > Iac02 > Iac03 (8). Lad binds simultaneously to both, the primary operator Iac01 and to either Iac02 or Iac03 through a DNA-looping mechanism (9). Lac02 is located 401 bp downstream of Iac01 , whereas Iac03 lies only 92 bp upstream of Iac01 (10). The role of Iac02 is still not clear, because the main contribution

to repression comes from the DNA-looping of Iac01 and Iac03 due to their closer proximity (8). Furthermore, when Iac01 and Iac03 are bound by Lad, the production of Lad itself is prevented. The 3' end of the *lad* gene overlaps with Iac03. In a repressed state, transcription of *lacI* results in a truncated mRNA, which is rapidly degraded by the cell. Due to this autoregulation, the concentration of the Lad tetramer is ~ 40 molecules in induced cells and ~ 15 molecules in non-induced cells (11).

Several mutants of the Lad repressor protein and the pLacI promoter exist. Penumetcha et al. tested various combinations of repressor and promoter mutants in an effort to discover a system with reduced leakiness in transcription. They report that use of the wild-type Lad repressor protein in combination with the pLacI^{Q1} Promoter gives high levels of induction and low levels of leaky transcription (34).

Oehler et al. tested the effect of systematic destruction of all three *lac* operators of the chromosomal *lac* operon of *Escherichia coli* on repression by Lac repressor and report that the three operators of the *lac* operon cooperate in repression (35).

The tetrameric Lac repressor can bind simultaneously to two *lac* operators on the same DNA molecule, thereby including the formation of a DNA loop. Muller et al. report that repression increases significantly with decreasing inter-operator DNA length (36).

The effects of placing a *lac* operator at different positions relative to a promoter for bacteriophage T7 RNA polymerase have been tested. Transcription can be strongly repressed by *lac* repressor bound to an operator 15 base-pairs downstream from the RNA start (37).

W02003/050240A2 discloses an expression system for producing a target protein in a host cell comprising a homologously integrated gene encoding T7 RNA polymerase, and a non-integrated gene encoding a target protein.

One of the first applications of the lac regulatory mechanism was the pET system, which today is the most widely used *E. coli* expression system for recombinant protein production (12, 13). This system is based on the specific interaction of the T7-phage derived T7 RNAP with the strong T7 promoter. The recombinase functions of bacteriophage lambda were used for site-directed insertion of the T7 RNA polymerase gene into the *E. coli* genome. Expression of the T7 RNAP is controlled by the lacUV5 promoter, a variant of the lactose promoter that is insensitive to catabolic repression. Addition of IPTG, induces the expression of the T7 RNAP at high levels, which in turn transcribes the target gene which is under control of the T7 promoter. This orthogonal expression system offers very high product titres for recombinant proteins that can

efficiently be produced in *E. coli.* However, the extraordinary strength of the T7 expression system, especially if combined with high-copy number plasmids exerts an extreme metabolic load on the host cells. When the gene of interest codes for challenging proteins, stress and metabolic burden often lead to reduced yield, shortened production periods and even cell death (14, 15).

Plasmid-mediated stress effects, such as high gene dosage and transcription of antibiotic resistance genes, can be overcome by integration of the gene of interest (GOI), i.e. the gene encoding the protein of interest, into the host's genome (16, 17).

W02008/142028A1 discloses a method for producing a protein of interest, wherein the DNA encoding the protein of interest is integrated into a bacterial cell's genome at a pre-selected site.

Striedner et al. disclose a plasmid-free T7 based *Escherichia coli* expression system, wherein the target gene is site-specifically integrated into the genome of the host *(17).*

Genome integrated T7-based expression systems offer significant advantages. Compared to plasmid-based expression systems there is no plasmid mediated metabolic load and no variation in gene dosage during the production process. However, the T7 RNA polymerase (RNAP) is prone to mutations under long-term production conditions. This was demonstrated by Striedner et al. (*17)* in chemostat cultivations, where mutations in the T7 RNAP led to faster growing of a non-producing cell population and thus, to a massive loss in product yield.

There is thus a clear need in the field for improved inducible expression systems which result in improved expression rates, low basal expression and true tunability of expression rates on a cellular level, even at low inductor concentrations.

SUMMARY OF THE INVENTION

It is the objective of the present invention to provide an improved inducible system with improved control of expression rate of a protein of interest and very low basal expression for plasmid-free production of a protein of interest.

The problem is solved by the present invention.

According to the invention, there is provided a genome-based expression system for production of a protein of interest in a prokaryotic host, comprising at least

a) an RNA polymerase (RNAP) gene,

b) a gene for expression of a protein of interest, comprising

- a coding sequence encoding the protein of interest,
- a promoter operably linked to said coding sequence, wherein said promoter is recognized by the RNAP expressed from a), and
- at least one lac operator (lacO) within the sequence of said promoter; and

c) a lad gene for expression of a lac repressor protein (Lad) comprising

- a lad coding sequence,
- a lad promoter operably linked to the lad coding sequence, wherein the lad promoter is selected from the group consisting of wild-type lad and a lad promoter which increases lad expression;

wherein the expression rate of the protein of interest is regulated by an inducer binding LacI.

According to a specific embodiment, there is provided a genome-based expression system for production of a protein of interest in a prokaryotic host, comprising at least

a) an RNA polymerase (RNAP) gene,

b) a gene for expression of a protein of interest, comprising

- a coding sequence encoding the protein of interest,
- a promoter operably linked to said coding sequence, wherein said promoter is recognized by the RNAP expressed from a), and
- a lac operator (lacO), preferably Iac01 , within the sequence of said promoter; and

c) a lacI gene for expression of a lac repressor protein (LacI) comprising

- a lacI coding sequence
- a lacI promoter operably linked to the lacI coding sequence, wherein the lacI promoter is a lacI promoter which increases expression of lacI, preferably it is the $lacI^Q$ promoter;

wherein the expression rate of the protein of interest is regulated by an inducer binding LacI.

Specifically, the gene for expression of a protein of interest contains one lacO within the sequence of the promoter operably linked to the coding sequence, and the lacI promoter is a promoter which increases LacI expression.

According to a further specific embodiment, there is provided genome-based expression system for production of a protein of interest in a prokaryotic host, comprising at least

a) an RNA polymerase (RNAP) gene,

b) a gene for expression of a protein of interest, comprising

- a coding sequence encoding the protein of interest,

- a promoter operably linked to said coding sequence, wherein said promoter is recognized by the RNAP expressed from a), and

- at least two lac operators (lacOs) that are at least 92bp, specifically 94bp, apart, wherein one lacO is within the sequence of the promoter and the other lacO is upstream of the promoter; and

c) a lad gene for expression of a lac repressor protein (Lad) comprising

- a lad coding sequence

- a lad promoter operably linked to the lad coding sequence, wherein the lad promoter is the wild-type lad promoter;

wherein the expression rate of the protein of interest is regulated by an inducer binding LacI.

According to an alternative embodiment, there is provided an inducible system for plasmid-free production of a protein of interest in a prokaryotic host, comprising at least

a) an RNA polymerase (RNAP) gene in the chromosome of the host,

b) a gene for expression of a protein of interest comprising

- a coding sequence encoding the protein of interest,

- a promoter operably linked to said coding sequence, wherein said promoter is recognized by the RNAP expressed from a), and

- at least one lac operator (lacO) within the sequence of said promoter; and

c) a lacI gene for expression of a lac repressor protein (lacI) comprising

- a lacI coding sequence,

- a lacI promoter operably linked to the lacI coding sequence, wherein the lacI promoter is selected from the group consisting of wild-type lacI and a lacI promoter which increases expression of lacI;

wherein the affinity of lacI to the one or more lacO / lacOs of b) is lower than the affinity of lacI to the lac operators lac01 and lac03 of the endogenous lac operon of the host and wherein the expression rate of the protein of interest is regulated by an inducer binding LacI.

According to further embodiment, there is provided an inducible system for plasmid-free production of a protein of interest in a prokaryotic host, comprising at least

a) an RNA polymerase (RNAP) gene in the chromosome of the host,

b) a gene for expression of a protein of interest comprising

- a coding sequence encoding the protein of interest,
- a promoter operably linked to said coding sequence, wherein said promoter is recognized by the RNAP expressed from a), and
- a lac operator (lacO), preferably Iac01 , within the sequence of said promoter; and

c) a lad gene for expression of a lac repressor protein (lad) comprising

- a lad coding sequence
- a lad promoter operably linked to the lad coding sequence, wherein the lad promoter is a lad promoter which increases expression of lad, preferably it is the $lacI^Q$ promoter;

wherein the affinity of lad to the one lacO of b) is lower than the affinity of lad to the lac operators Iac01 and Iac03 of the endogenous lac operon of the host and wherein the expression rate of the protein of interest is regulated by an inducer binding LacI.

According to a further specific embodiment of the invention, there is provided an inducible system for plasmid-free production of a protein of interest in a prokaryotic host, comprising at least

a) an RNA polymerase (RNAP) gene in the chromosome of the host,

b) a gene for expression of a protein of interest comprising

- a coding sequence encoding the protein of interest,
- a promoter operably linked to said coding sequence, wherein said promoter is recognized by the RNAP expressed from a), and
- at least two lac operators (lacOs) that are at least 92bp apart, wherein one lacO is within the sequence of the promoter and the other lacO is upstream of the promoter; and

c) a lacI gene for expression of a lac repressor protein (lacI) comprising

- a lacI coding sequence
- a lacI promoter operably linked to the lacI coding sequence, wherein the lacI promoter is the wild-type lacI promoter;

wherein the affinity of lacI to the at least two lacOs of b) is lower than the affinity of lacI to the lac operators Iac01 and Iac03 of the endogenous lac operon of the host and wherein the expression rate of the protein of interest is regulated by an inducer binding LacI.

Specifically, the prokaryotic host is *Escherichia coli* (*E.coli*). Specifically, the host is *E.coli* of the strain BL21 or K-12.

Specifically, the RNAP is an RNAP homologous to the host, specifically σ^{70} *E.coli* RNA polymerase.

Specifically, the promoter operably linked to the coding sequence encoding the protein of interest is selected from the group consisting of T5, T5N25, T7A1 , T7A2, T7A3, lac, lacUV5, tac or trc or functional variants thereof with at least 20, 30, 40, 50, 60, 70, 80 or 90% sequence identity to T5, T7A1 , T7A2, T7A3, lac, lacUV5, tac or trc.

According to a preferred embodiment of the inducible system described herein, the lad promoter is a promoter which increases expression of Lad compared to the wild type host, which is the $lacI^Q$ promoter (SEQ ID NO:1). Specifically, the gene encoding the protein of interest includes only one lacO, preferably lacOI , and the lad promoter is $lacI^Q$ (SEQ ID NO:1).

Preferably, the gene encoding the protein of interest comprises at least one lacO selected from the group consisting of lacOI , Iac02 or Iac03 and any combination thereof. Specifically, the gene encoding the protein of interest comprises two lacOs, preferably lacOI and lacOI or lacOI and Iac02 or lacOI and Iac03.

Specifically, the at least one lac operator comprised in the gene encoding the protein of interest is a lacOI (SEQ ID NO:3), Iac02 (SEQ ID NO:4) or Iac03 (SEQ ID NO:5).

Specifically, the at least one lac operator is a functional variant of lacOI , Iac02 or Iac03 with at least 65% sequence identity or a perfectly symmetric lacO. Specifically, the lac operator is a functional variant of lacOI , Iac02 or Iac03 with at least 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91 or 95% sequence identity to wild-type lacOI , Iac02 or Iac03. According to an alternative, a functional variant of lacOI , Iac02 or Iac03 comprises 1, 2, 3, 4 or 5 point mutations or deletions of 1, 2, 3, 4 or 5 base pairs (bps).

Specifically, said promoter operably linked to the coding sequence encoding the protein of interest comprises an initial transcribed sequence (ITS), preferably a native T7A1 initial transcribed sequence (SEQ ID NO:2).

According to the system provided herein, the expression rate of the protein of interest is regulated by an inducer binding LacI. Specifically, Lad binds to the at least one lacO thereby repressing transcription of the gene encoding the protein of interest. Specifically, upon addition of an inducer capable of binding LacI interaction of LacI with

the at least one lacO is prevented, resulting in induction of transcription of the gene encoding the protein of interest.

Specifically, the inducer is selected from the group consisting of isopropylthiogalactoside (IPTG), lactose, methyl^-D-thiogalactoside, phenyl^-D-galactose and ortho-Nitrophenyl^-galactoside (ONPG).

Specifically, the promoter operably linked to the coding sequence expressing the protein of interest comprises an initial transcribed sequence, preferably the native T7A1 initial transcribed sequence. Specifically, the initial transcribed sequence is not limited to the ITS of T7A1 and can be any ITS known to a person skilled in the art.

According to a specific embodiment of the inducible system provided herein, the gene for expression of a protein of interest contains one lac01 operator within the sequence of the promoter operably linked to the native T7A1 initial transcribed sequence (SEQ ID NO:2) and to the coding sequence, and wherein the Lad promoter is a $lacI^Q$ promoter.

According to a further specific embodiment of the inducible system provided herein, the gene of interest contains two lac operators which are at least about 92 or 94 basepairs (bps) apart, preferably at least about 103, 105, 114, 116, 125, 127, 136, 138, or 149 bps apart, wherein one lac operator is located within the sequence of the promoter operably linked to the coding sequence and the second lac operator is upstream of the promoter.

Specifically, the gene encoding the protein of interest is a heterologous gene. Specifically, said gene that is heterologous to the prokaryotic host is a recombinant gene that is introduced into the host.

According to a further specific embodiment, the gene encoding the protein of interest is a homologous gene. Specifically, said gene that is homologous to the prokaryotic host, comprises a coding sequence, encoding the protein of interest, a promoter operably linked to said coding sequence, wherein said promoter is recognized by an RNAP that is expressed from a gene in the chromosome of the host, and at least one lac operator (lacO) within the sequence of said promoter.

Specifically, said gene that is homologous to the prokaryotic host is a recombinant gene that is introduced into the host. According to yet a further specific embodiment, said gene that is homologous to the prokaryotic host is modified by replacement of the promoter endogenous to said gene with a promoter described herein. Replacement can also mean the integration of the promoter described herein so that it is operably linked

to the endogenous homologous gene / polypeptide in the chromosome / genome of the host cell wherein the naturally occurring promoter of the endogenous homologous gene / polypeptide is inactivated by at least one point mutation within the naturally occurring promoter. Specifically, the promoter endogenous to said gene is replaced with a promoter described herein comprising at least one lacO within the sequence of the promoter, preferably at least two lacOs, wherein one lacO is within the sequence of the promoter and a second lacO is upstream of the promoter. Specifically, the affinity of lad to the one or more lacO / lacOs of the promoter replacing the endogenous promoter of the gene encoding the protein of interest is lower than the affinity of Lad to the Iac01 and Iac03 of the endogenous lac operon.

Specifically, the promoter operably linked to the coding sequence of the gene for expression of a protein of interest, is a recombinant promoter. Specifically, said promoter is not the wildtype *lac* promoter, it can, however, be a variant of the *lac* promoter. In the case, where the promoter described herein is a variant of the *lac* promoter, it comprises at least one lacO within its sequence, specifically it comprises at least one lacO within the sequence between the -10 and -35 promoter elements.

Further provided herein is a method of plasmid-free production of a protein of interest in a prokaryotic host, using the inducible system described herein, comprising the steps of

a) cultivating the host cells and inducing expression of the gene of interest by addition of an inducer,

b) harvesting the protein of interest, and

c) isolating and purifying the protein of interest and optionally

d) modifying the protein of interest and

e) formulating the protein of interest.

According to a specific embodiment of the system described herein, the gene for producing the protein of interest and/or the *lad* gene for producing a lac repressor protein are comprised in at least one expression cassette. Preferably, said expression cassette is used to integrate the gene for producing the protein of interest and/or the *lacI* gene for producing a lac repressor protein into the chromosome of the prokaryotic host.

Also provided herein is an expression cassette comprising at least one heterologous gene configured to produce at least one heterologous protein of interest, the gene of interest including

a) one or more coding sequences encoding the one or more proteins of interest,

b) a promoter operably linked to the coding sequence, and

c) at least one lac operator (lacO) operably linked to said promoter.

Specifically, the affinity of Lad to the at least one lacO comprised in the expression cassette is lower than the affinity of Lad to the lac operators lac01 and lac03 of the lac operon of a host cell. Preferably, said lac operon is the lac operon endogenous to the host cell.

According to a specific embodiment of the expression cassette provided herein, the heterologous gene configured to produce at least one heterologous protein of interest includes two lac operators, which are at least 92 or 94 bp apart, wherein one lac operator is located within the sequence of the promoter and the second lac operator is upstream of the promoter. Preferably, said two lac operators are at least about 92 to 134 bps apart, preferably they are at least about 103, 105, 114, 116, 125 or 136 or 138 or 149 bps apart. Specifically, said two lac operators are 92, 94, 103, 105, 114, 116, 125,1 36, 138 or 149 bps apart.

According to a specific embodiment of the expression cassette provided herein, the heterologous gene configured to produce at least one heterologous protein of interest comprises a lacOI operator within the sequence of the promoter operably linked to the coding sequence and a native T7A1 initial transcribed sequence (SEQ ID NO:2). Specifically, said expression cassette further comprises a heterologous lad promoter, which is the $lacI^Q$ promoter (SEQ ID NO:1).

Further provided herein is a method of plasmid-free production of a protein of interest in a prokaryotic host on a manufacturing scale, using the expression cassette described herein, comprising the steps of

a) integrating the expression cassette into the chromosome of the prokaryotic host,

b) cultivating the host cells and inducing expression of the gene of interest by addition of an inducer,

c) harvesting protein of interest, and

d) isolating and purifying the protein of interest. and optionally

e) modifying the protein of interest and

f) formulating the protein of interest.

According to a specific embodiment of the method and the system provided herein, the prokaryotic host contains the expression cassette integrated at an attachment site, preferably the attTn7, lacZ, recA, tufa or attnB site.

FIGURES

Figure 1: Scheme of integration cartridges. Expression of GFPmut3.1 is controlled by seven different promoter/operator combinations. The T7 expression system is used as reference. The cartridges were cloned into pET30a-cer vector (designated with round brackets) or were integrated into the attTN7 site (designated with squared brackets) of the BL21 genome (B) resp. BL21 ^{Q} (as described in Example 1) (BQ). In two promoter/operator combinations the wild-type lad promoter (lad wt) was exchanged by the lacI ^{Q} promoter (lacI ^{Q}). LacOI * is a 2 bp truncated version of wild-type Iac01 . Sym-lacO is the perfectly symmetric lac operator. +1 T7A1 +20 is the native ITS of the T7A1 promoter. Transcription is terminated by tZENIT (tZ). GFPmut3.1 is the coding sequence for expression of the GFPmut3.1 protein. Iac01 is the wild type Iac01 . -35 and -10 are the -35 and -10 promoter regions of the respective promoters, A 1 and T5.

Figure 2: Promoter activities of different promoter/operator combinations under uninduced (0 mM IPTG) and induced (0.5 mM IPTG) conditions. The fluorescence of reporter GFPmut3.1 (y-axix) was used to characterize genome-integrated expression systems (A) and plasmid-based expression systems (B). The integration cartridges cloned into pET30a-cer vector are designated with round brackets, those integrated integrated into the attTN7 site of the BL21 genome (B) resp. BL21Q (as described in Example 1) (BQ) are designated with squared brackets.

Figure 3: Influence of lac-operators on GFP expression and tuneability of expression of GFP expressed by the course of GFP on-line fluorescence (y-axis) in fedbatch-like microtiter cultivation. The dashed vertical lines indicate time of induction. A - D: **T5N25** promoter controlled by three lacO (B<3lacO-T5>) (A), two lacO (B<2lacO-T5>) (B), one lacO (B<1 lacO-T5>) (C) and one lacO / lacI ^{Q} promoter (BQ<1 lacO-T5>) (D). E -G: **T7AI** promoter controlled by two lacO (B<2lacO-A1 >) (E), one lacO (B<1 laqO-A 1>) (F) and one lacO / lacI ^{Q} promoter (BQ<1 lacO-A1) (G). The T7 expression system is used as reference (FI).

Figure 4: Control of recombinant gene expression with different levels of inducer. Flow cytometry analysis of GFPmut3.1 expression in B<2lacO-A1>, BQ<1 lacO-A1> and B3<T7>.

Figure 5: Scheme of lac-operator binding sites on native lac-operon (top) and gene of interest (bottom). Promoters for the gene of interest are regulated by one lac-operator (A) or two lac-operators that are 62bp apart (B).

Figure 6: SEQ ID NOs referred to herein.

Figure 7: Influence of recombinant expression rate control on Lad concentration. (A) BL21 wild-type cells (lanes 1-3) and B<2lacO-A1> (lanes 4-6) were grown without IPTG (lanes 1 and 4), 0.01 mM IPTG (lanes 2 and 5) and 0.5 mM IPTG (lanes 3 and 6). Proteins of ~ 1.2 x 10^7 cells were separated by SDS-PAGE and analyzed by western blotting, using an anti-LacI antibody. (B) Fold changes are shown relative to 0 mM IPTG BL21-wt. Error bars indicate standard error of the mean (n = 2).

Figure 8: Process characteristics and product formation kinetic of B3<T7-dFTN2> during the carbon-limited exponential fed-batch cultivation. Cultivations were conducted in a 1.5 L DASGIP® parallel bioreactor system with a final volume of 1.2 L. The dashed vertical lines indicate time of induction.

Figure 9: Process characteristics and product formation kinetic of BQ<A1-dFTN2> during the carbon-limited exponential fed-batch cultivation. Cultivations were conducted in a 1.5 L DASGIP® parallel bioreactor system with a final volume of 1.2 L. The dashed vertical lines indicate time of induction.

DETAILED DESCRIPTION

Unless indicated or defined otherwise, all terms used herein have their usual meaning in the art, which will be clear to the skilled person. Reference is for example made to the standard handbooks, such as Sambrook et al, "Molecular Cloning: A Laboratory Manual" (2nd Ed.), Vols. 1-3, Cold Spring Harbor Laboratory Press (1989); Lewin, "Genes IV", Oxford University Press, New York, (1990), and Janeway et al, "Immunobiology" (5th Ed., or more recent editions), Garland Science, New York, 2001.

The terms “comprise”, “contain”, “have” and “include” as used herein can be used synonymously and shall be understood as an open definition, allowing further members or parts or elements. “Consisting” is considered as a closest definition without further elements of the consisting definition feature. Thus “comprising” is broader and contains the “consisting” definition.

The term “about” as used herein refers to the same value or a value differing by +/-5 % of the given value.

Genome integrated, i.e. plasmid-free, expression systems offer significant advantages. Compared to plasmid-based expression systems there is no plasmid mediated metabolic load and no variation in gene dosage during the production process. However, the current state of the art T7-based expression system employing the strong T7 promoter dependent on the T7 RNA polymerase which is under the control of an inducible promoter, still suffers from considerable drawbacks. The strength of the T7 expression system exerts an extreme metabolic load on the host cells. When the gene of interest codes for challenging proteins, the stress and metabolic burden often lead to reduced yield, shortened production periods and even cell death. Moreover, the T7 expression system is leaky, because it shows significant basal expression, and the T7 RNA polymerase is prone to mutations under long-term production conditions.

The plasmid-free inducible expression system provided herein has the profound advantage that the rate of expression is tunable on a single cell level, it exhibits very low basal expression and it is highly efficient in recombinant protein production. Moreover, it provides true control of expression rate, negligible basal expression and a high expression rate even at low inductor concentrations, which is particularly beneficial for production of challenging proteins.

The terms **“plasmid-free”** or **“genome-based”** as used herein, refer to an expression system of a protein of interest in a prokaryotic host, wherein the gene for the expression of the protein of interest is located in the genome of the host. Specifically, said gene is an endogenous homologous gene which is located on the chromosome of the prokaryotic host, or is a recombinant heterologous or homologous gene that is integrated into the chromosome of the prokaryotic host.

According to a specific embodiment, a gene for expression of a protein of interest and optionally a lad gene for expression of a lac repressor protein or a recombinant lad promoter are integrated into the genome of the host using one or more expression cassette(s) comprising said genes.

Specifically, further recombinant heterologous or homologous genes, such as genes encoding an RNA polymerase or genes encoding helper proteins are introduced into the prokaryotic host. Said further recombinant heterologous or homologous genes may be introduced into the chromosome of the host or may be present in the host cell on a plasmid.

The terms **“expression cassette”,** or simply **“cassette”,** synonymously used with “expression cartridge” or simply “cartridge”, refer to a linear or circular DNA construct to be integrated into the prokaryotic genome, such as the bacterial genome. As a result of integration, the expression host cell has an integrated expression cassette. Preferably, the cassette is a linear DNA construct comprising essentially a promoter, a gene of interest, immediately upstream of the gene of interest a Shine-Dalgarno (SD) sequence, also termed ribosome binding site (RBS) and two terminally flanking regions which are homologous to a genomic region and which enable homologous recombination. In addition, the cassette may contain other sequences such as for example sequences coding for antibiotic selection markers, prototrophic selection markers or fluorescent markers, markers coding for a metabolic gene, genes which improve protein expression or two flippase recognition target sites (FRT) which enable the removal of certain sequences (e.g. antibiotic resistance genes) after integration.

The expression cassette is synthesized and amplified by methods known in the art, in the case of linear cassettes, usually by standard polymerase chain reaction, PCR. Since linear cassettes are usually easier to construct, they are preferred for obtaining the expression host cells used in the system and method provided herein. Moreover, the use of a linear expression cassette provides the advantage that the genomic integration site can be freely chosen by the respective design of the flanking homologous regions of the cassette. Thereby, integration of the linear expression cassette allows for greater variability with regard to the genomic region.

Expression vectors comprise the expression cassette described herein and in addition optionally comprises flanking regions homologous to the genome integration site, a number of restriction enzyme cleavage sites, an initial transcribed sequence (ITS) and a transcription terminator, and optionally one or more selectable markers (e.g., an amino acid synthesis gene or a gene conferring resistance to antibiotics such as ampicillin, kanamycin, chloramphenicol or streptomycin), which components are operably linked together. A common type of vector is a “plasmid”, which generally is a self-contained molecule of double-stranded DNA that can readily accept additional (foreign) DNA and which can readily be introduced into a suitable host cell. Specifically, the term “vector” or “plasmid” refers to a vehicle by which a DNA or RNA sequence (e.g. a foreign gene) can be introduced into a host cell, so as to transform the host and promote expression (e.g. transcription and translation) of the introduced sequence.

As used herein, the term **"prokaryotic host"** refers to any bacterial host, in particular it refers to bacterial host cells. In principle, there are no limitations regarding the choice of bacterial host cells, except for certain specific requirements detailed below. The bacterial host cells may be eubacteria (gram-positive or gram-negative) or archaebacteria, as long as they allow genetic manipulation for insertion of a gene of interest, advantageously for site-specific integration. Preferably, the bacterial host cells allow cultivation on a manufacturing scale. Preferably, the host cell has the property to allow cultivation to high cell densities. Examples for bacterial host cells that have been shown to be suitable for recombinant industrial protein production are *Escherichia coli, Bacillus subtilis, Pseudomonas fluorescens* as well as variations thereof and *Lactococcus lactis* strains. Preferably, the host cells are *E. coli* cells.

A requirement to the host cell is that it comprises an RNA polymerase that can bind to the promoter controlling the gene encoding the protein of interest.

In certain embodiments, the host cell carries, in its genome, a marker gene in view of selection.

In view of site-specific gene insertion, another requirement to the host cell is that it contains at least one genomic region (either a coding or any non-coding functional or non-functional region or a region with unknown function) that is known by its sequence and that can be disrupted or otherwise manipulated to allow insertion of a heterologous sequence, without being detrimental to the cell.

With regard to the integration locus, the expression system used in the invention allows for a wide variability. In principle, any locus with known sequence may be chosen, with the proviso that the function of the sequence is either dispensable or, if essential, can be complemented (as e.g. in the case of an auxotrophy).

Integration of the gene of interest into the bacterial genome can be achieved by conventional methods, e.g. by using linear cartridges that contain flanking sequences homologous to a specific site on the chromosome, as described for the attTn7-site, e.g. in (30). Moreover, the use of a linear expression cartridge provides the advantage that the genomic integration site can be freely chosen by the respective design of the flanking homologous regions of the cartridge. Thereby, integration of the linear expression cartridge allows for greater variability with regard to the genomic region. In a preferred embodiment, integration of a linear cartridge is at an attachment site like the attB site or the attTn7 site, which are well-proven integration sites. Examples, without limitation, of other integration methods useful in the present invention are e.g. those based on Red/ET

recombination, e.g. described in (31). Alternatively, an expression cassette can first be integrated into the genome of an intermediate donor host cell, from which it can then be transferred to the host cell by transduction by the P1 phage, e.g. described in (32). The integration method used herein is not limited to the above-mentioned examples; rather any integration method known in the art can be used.

The integration methods for obtaining the expression host cell are not limited to integration of one gene of interest at one site in the genome; they allow for variability with regard to both the integration site and the expression cassettes. By way of example, more than one gene of interest may be inserted, i.e. two or more identical or different sequences under the control of identical or different promoters can be integrated into one or more different loci on the genome. By way of example, it allows expression of two different proteins that form a heterodimeric complex. Heterodimeric proteins consist of two individually expressed protein Subunits, e.g. the heavy and the light chain of a monoclonal antibody or an antibody fragment.

Although the invention allows plasmid-free production of a protein of interest, it does not exclude that in the expression host cell a plasmid may be present that carries sequences to be expressed other than the gene of interest, e.g. helper proteins and/or recombination proteins. Preferably, care should be taken that in such embodiments the advantages of the invention should not be overruled by the presence of the plasmid, i.e. the plasmid should be present at a low copy number and should not exert a metabolic burden onto the cell.

Integration of one or more recombinant genes into the genome results in a discrete and pre-defined number of genes of interest per cell. In the embodiment of the invention that inserts one copy of the gene, this number is usually one (except in the case that a cell contains more than one chromosome or genome, as it occurs transiently during cell division), as compared to plasmid-based expression which is accompanied by copy numbers up to several hundred. In the expression system used in the method of the present invention, by relieving the host metabolism from plasmid replication, an increased fraction of the cells synthesis capacity is utilized for recombinant protein production.

A particular advantage is that the inducible expression system described herein has no limitations with regard to the level of induction. This means that the system cannot be “over-induced as it often occurs in plasmid-based systems, or systems employing strong promoters such as the T7 expression system. Since the genome-based

expression system allows exact control of protein expression, it is particularly advantageous in combination with expression targeting pathways that depend or rely on well-controlled expression. In a preferred embodiment, the method of the invention includes secretion (excretion) of the protein of interest from the bacterial cytoplasm into the periplasm and/or culture medium. The advantage of this embodiment is an optimized and sustained protein secretion rate, resulting in a higher titer of secreted protein as compared to prior art secretion systems. Specifically, this can be achieved by fusing a signal peptide N-terminal to the protein of interest / a nucleotide sequence encoding a signal peptide, which leads the protein of interest to the transporters of the host, causes translocation into the periplasma of the host and is cleaved by the signal peptidase of the host. Any signal peptide known in the art can be used such as but not limited to the ompA-, pelB, malE-, phoA-, dsbA-, lysC-, IoIB-, pyrL- leader peptides.

As used herein, the term **"RNA polymerase (RNAP) gene"** refers to a gene expressing an RNAP, which gene is comprised in the genome, e.g. in a plasmid, or chromosome of the prokaryotic host. Preferably, said gene expresses an RNAP that is endogenous to the prokaryotic host.

In bacteria, the same enzyme catalyzes the synthesis of mRNA and non-coding RNA (ncRNA). RNAP is a large molecule; the core enzyme has five subunits (-400 kDa). In order to bind promoters, RNAP core associates with the transcription initiation factor sigma (σ) to form RNA polymerase holoenzyme. Sigma reduces the affinity of RNAP for nonspecific DNA while increasing specificity for promoters, allowing transcription to initiate at correct sites. The complete holoenzyme therefore has 6 subunits (-450 kDa). The core enzyme is responsible for binding to template DNA to synthesize RNA, which is complemented by a σ factor to form a holoenzyme that recognizes the promoter sequence to begin promoter-specific transcription.

According to a preferred embodiment, the prokaryotic host cells of the system described herein are *E.coli* cells and the RNAP is an RNAP that is endogenous to *E.coli,* most preferably it is σ^{70} ***E.coli*** **RNA polymerase.** The σ subunit of bacterial RNA polymerase (RNAP) is required for promoter-specific transcription initiation. In the case of *E. coli* and other gram-negative rod-shaped bacteria, the "housekeeping" or "primary" sigma factor is σ^{70}. Every cell has a "housekeeping" sigma factor that keeps essential genes and pathways operating. When complexed with the RNAP core enzyme (subunit structure $\alpha_2\beta\beta'\omega$), different σ factors specify the recognition of different classes of promoters. Genes recognized by σ^{70} all contain similar promoter consensus sequences

consisting of two parts. The primary σ factor in *Escherichia coli,* σ^{70}, typically directs transcription initiation from promoters defined by two conserved hexameric DNA sequence elements, termed the -10 and -35 elements for their relationship to the transcription start site (position +1). Relative to the DNA base corresponding to the start of the RNA transcript, the consensus promoter sequences are characteristically centered at 10 and 35 nucleotides before the start of transcription (-10 and -35).

The term **"expression"** is understood in the following way. Nucleic acid molecules containing a desired coding sequence of an expression product such as *e.g.,* a recombinant protein as described herein, and control sequences such as *e.g.,* a promoter in operable linkage, may be used for expression purposes. Hosts transformed or transfected with these sequences are capable of producing the encoded proteins. In order to effect transformation, the expression system may be included in a vector; however, most preferably the relevant DNA is integrated into the host chromosome.

The term **"gene"** as used herein refers to a DNA sequence that comprises at least promoter DNA, optionally including operator DNA, and coding DNA which encodes a particular amino acid sequence for a particular polypeptide or protein. Promoter DNA is a DNA sequence which initiates, regulates, or otherwise mediates or controls the expression of the coding DNA. Promoter DNA and coding DNA may be from the same gene or from different genes, and may be from the same or different organisms.

The term **"recombinant"** as used herein shall mean "being prepared by or the result of genetic engineering". A recombinant host specifically comprises a recombinant expression vector or cloning vector, or it has been genetically engineered to contain a recombinant nucleic acid sequence, in particular employing nucleotide sequence foreign to the host. A recombinant protein is produced by expressing a respective recombinant nucleic acid in a host.

With regard to the **protein of interest (POI),** there are no limitations. More specifically, the protein may either be a polypeptide not naturally occurring in the host cell, i.e. a heterologous protein, or else may be native to the host cell, i.e. a homologous protein to the host cell, but is produced, for example, upon integration by recombinant techniques of one or more copies of the nucleic acid sequence encoding the homologous POI into the genome or chromosome of the host cell, or by recombinant modification of the promoter sequence controlling the expression of the gene encoding the POI. The POI can be a monomer, dimer or multimer, it can be a homomer or heteromer.

Examples for proteins that can be produced by the method of the invention are, without limitation, enzymes, regulatory proteins, receptors, peptides, e.g. peptide hormones, cytokines, membrane or transport proteins. The proteins of interest may also be antigens as used for vaccination, vaccines, antigen-binding proteins, immune stimulatory proteins, allergens, full-length antibodies or antibody fragments or derivatives. Antibody derivatives may be for example single chain variable fragments (scFv), Fab fragments or single domain antibodies.

The **DNA** molecule encoding the protein of interest is also termed **"gene of interest".** Specifically, the gene of interest includes the **DNA** sequence encoding the protein of interest, a promoter operably linked to the coding sequence and at least one lac operator within the sequence of the promoter.

Further, the gene of interest encoding the POI can be a naturally existing DNA sequence or a non-natural DNA sequence. One or more gene of interests can be under the control of one promoter as described herein. Alternatively, each gene of interest is under one promoter. The gene of interests may all be on the same expression cassette or on multiple expression cassettes. The POI can be modified in any way. Non-limiting examples for modifications can be insertion or deletion of post-translational modification sites, insertion or deletion of targeting signals (e.g.: leader peptides), fusion to tags, proteins or protein fragments facilitating purification or detection, mutations affecting changes in stability or changes in solubility or any other modification known in the art. In certain embodiments of the invention the recombinant protein is a biopharmaceutical product, which can be any protein suitable for therapeutic or prophylactic purposes in mammals.

The term **"promoter"** as used herein refers to an expression control element that permits binding of RNA polymerase and the initiation of transcription. Specifically, the promoter operably linked to the gene of interest as described herein, comprises at least one lac operator within its sequence. Specifically, said at least one lac operator is situated between the -10 and -35 elements, which elements are preferably located 10 and 35 nucleotides before the start of transcription (-10 and -35), as exemplified in Figure 1.

The lac promoter is the promoter of the lac operon, which controls transcription of the three lac genes, *lacZ, lacY* and *lacA.* The wildtype lac promoter does not comprise a lac operator within its sequence, as it does not comprise a lacO between the -10 and -35 promoter elements. Preferably, in the inducible expression system described herein,

the lac promoter is the endogenous lac promoter comprising the endogenous lac operators. According to a specific embodiment, one or more lac operators of the endogenous lac promoter are genetically modified to increase their binding affinity to the lac repressor molecule LacI. Specifically, they are genetically modified so that their affinity to the lac repressor molecule LacI is greater than the affinity of the lac operators of the promoter operably linked to the gene of interest.

The lacI promoter as used herein, is the promoter operably linked to the coding sequence of the *lacI* gene. Specifically, the inducible system described herein, includes the wild-type lacI promoter or a genetically modified lacI promoter which increases expression of LacI, such as the exemplary $lacI^Q$ promoter described herein. Specifically, the lacI promoter is a constitutive promoter. Specifically, any constitutive promoter stronger than the native lacI promoter can be used as lacI promoter according to the present invention. Specifically, any promoter stronger than the native lacI promoter can be used as lacI promoter according to the system provided herein, such as but not limited to T5, T7A1 , T7A2, T7A3, T7, dnaK/J, spac, bla, nptII, cat promoters.

The promoter operably linked to the gene encoding the protein of interest as described herein, can be any inducible promoter that is recognized by an RNAP encoded by an RNAP gene comprised in the chromosome of the host.

According to certain embodiments of the invention, the gene of interest may be under the control of the lac, lacUV5, tac or the trc promoter, the lac or the lacUV5 promoter, the T5 promoters (Gentz and Bujard, 1985), such as the T5N25, or the T7 promoters (Hawley and McClure, 1983), such as T7 C or T7 D or the T7A promoters, such as T7A1 , T7A2 or T7A3 promoters (all inducible by lactose or its analogue IPTG), or other promoters suitable for recombinant protein expression, which all use *E. coli* RNA polymerase. The sequences of such promoters are well known in the art, such as e.g. those described by Gentz and Bujard, 1985 (33) or Hawley and McClure, 1983 (38). Specifically, the sequences of said promoters are modified to comprise at least one lacO within their sequence, as described herein.

According to a specific embodiment, the promoter described herein, which is in operable linkage to the sequence encoding the protein of interest, comprises a lacO within its sequence. In bacteria, the sequence of a promoter typically contains two short sequence elements, which, in wild type promoters, are typically approximately 10 and 35 nucleotides upstream of the transcription start site. These sequences are conserved among many bacterial strains. For example, the sequence at -10 nucleotides (also called

the -10 element) typically has the consensus sequence TATAAT (SEQ ID NO:34), and the sequence at -35 (also called the -35 element) has the consensus sequence TTGACA (SEQ ID NO:35). The above consensus sequences, while conserved on average, are not found intact in all promoters. On average, only 3 to 4 of the 6 base pairs in each consensus sequence are found in any given promoter. Few natural promoters have been identified to date that possess intact consensus sequences at both the -10 and -35 elements. Specifically, artificial promoters with complete conservation of the -10 and -35 elements transcribe at lower frequencies than those with a few mismatches with the consensus.

Specifically, the promoter described herein comprises at least one lacO between the -10 and -35 elements.

The term **"inducer",** synonymously used with "inductor", refers the factor capable of leading to the induction of transcription through direct or indirect regulation of promoter activity. Specifically, as used herein, inducer is any factor that is capable of binding the lac repressor molecule and inhibiting its interaction with the promoter operably linked to the gene of interest. Preferably, the inducer used herein is isopropylthiogalactoside (IPTG), lactose, methyl^-D-thiogalactoside, phenyl^-D-galactose or ortho-nitrophenyl-β-galactoside (ONPG).

There is no limitation as regards the mode by which induction of protein expression is performed. By way of example, the inductor can be added as a singular or multiple bolus or by continuous feeding, the latter being also known as "inductor feed(ing)". There are no limitations as regards the time point at which the induction takes place. The inductor may be added at the beginning of the cultivation or at the point of starting continuous nutrient feeding or after (beyond) the start of feeding. Inductor feeding may be accomplished by either having the inductor contained in the culture medium or by separately feeding it. The advantage of inductor feeding is that it allows to control inductor dosage, i.e. it allows to maintain the dosage of a defined or constant amount of inductor per constant number of genes of interest in the production system. For instance, inductor feeding allows an inductor dosage which is proportional to the biomass, resulting in a constant ratio of inductor to biomass. Biomass units on which the inductor dosage can be based, may be for instance cell dry weight (CDW), wet cell weight (WCW), optical density, total cell number (TCN; cells per volume) or colony forming units (CFU per volume) or on-line monitored signals which are proportional to the biomass (e.g. fluorescence, turbidity, light scatter, dielectric capacity, carbon dioxide

concentration in the exhaust gas etc.)· Essentially, the method of the invention allows the precise dosage of inductor per any parameter or signal which is proportional to biomass, irrespective of whether the signal is measured off-line or online. Since the number of genes of interest is defined and constant per biomass unit (one or more genes per cell), the consequence of this induction mode is a constant dosage of inductor per gene of interest. As a further advantage, the exact and optimum dosage of the amount of inductor relative to the amount of biomass can be experimentally determined and optimized.

It may not be necessary to determine the actual biomass level by analytical methods. For instance, it may be sufficient to add the inductor in an amount that is based on previous cultivations (historical biomass data). In another embodiment, it may be preferable to add the amount of inductor per one biomass unit as theoretically calculated or predicted. For instance, it is well known for feeding-based cultivations (like fed-batch or continuous) that one unit of the growth-limiting component in the feed medium, usually the carbon source, will result in a certain amount of biomass.

Preferably, the inducer is used at a concentration ranging from 0.005mM to 1mM, even more preferably from 0.01 mM to 0.5mM. Specifically, the concentration of IPTG is in the range of 1-100 pmol/g CDW.

As provided herein, the host used in the inducible expression system described herein comprises a lac operon, preferably a wild-type lac operon, and a lad gene.

As referred to herein, the **endogenous lac operon** contains three genes: *lacZ, lacY,* and *lacA.* These genes are transcribed as a single mRNA, under control of one promoter. In addition to the three genes, the lac operon comprises the **lac promoter** and the **lac operators** lacOI , Iac02 and Iac03. The lac promoter is the binding site for the RNA polymerase. The lac operator is the negative regulatory site bound by the lac repressor protein. The operator overlaps with the promoter, and when the lac repressor protein is bound, RNA polymerase cannot bind to the promoter and start transcription. According to a specific embodiment, the endogenous lac operon is modified to increase the binding affinity of Lad to at least one of the lac operators lacOI , Iac02 or Iac03. Specifically, at least one of the lac operators lacOI , Iac02 or Iac03 is modified, i.e. the endogenous lac operon comprises a functional variant of lacOI , Iac02 and/or Iac03 with increased affinity for Lacl.

As used herein, the term **“lacl gene”** refers to a gene for expression of the lac repressor protein, also called lac inhibitor (Lacl), or any functional variant thereof with at

least 30% sequence identity to *lad* (SEQ ID NO:26). Specifically, said gene comprises a *lacI* coding sequence, a lad promoter operably linked to the *lad* coding sequence, wherein the lad promoter is selected from the group consisting of the wild-type lad promoter and a lad promoter which increases expression of *lad.* Specifically, the *lad* gene expresses Lad or a functionally active variant thereof comprising at least 40, 50, 60, 70, 80 or 90% sequence identity to Lad (SEQ ID NO:27). Specifically, the lad promoter which increases expression of Lad is a strong promoter, which increases expression of Lad by at least 1.5, 2, 2.5 or 5-fold, preferably 10-fold or more. Specifically, it increases the expression of Lad by at least 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold or even 100-fold. An exemplary embodiment of the inducible system provided herein comprises the $lacI^Q$ promoter as the lad promoter which increases expression of lacI. The $lacI^Q$ promoter includes a point mutation, a single C→T change, in the promoter region upstream of the native *lad* gene, resulting in a 10-fold increase in mRNA transcription. The promoter for the *lad* coding sequence may include the native *lad* initiation codon or any variants thereof. The *lad* gene is preferably incorporated into the host's chromosomal DNA or contained on a single-copy vector.

In wild-type *E. coli,* the lac repressor protein forms a homo-tetramer that binds to the /ac-operator sequences (lacO) and represses the transcription of the *lacZYA* operon. In the presence of lactose or the non-metabolizable isopropyl β-D-l-thiogalactopyranoside (IPTG), LacI changes its structure and can no longer bind to the lac-operator, resulting in induction of transcription. The lac-operator sites are DNA sequences with inverted repeat symmetry.

The higher the symmetry, the greater the binding affinity of LacI to the operator sequence. An artificial perfectly symmetric lacO (sym-lacO) was found to bind LacI with the greatest affinity, whereas the three wild-type operators lacOI , Iac02 and Iac03 exhibiting an approximate symmetry showed lower affinities, resulting in the following order with respect to the affinity to LacI: sym-lacO > lacOI > Iac02 > Iac03. LacI binds simultaneously to both, the primary operator lacOI and to either Iac02 or Iac03 through a DNA-looping mechanism. Lac02 is located 401 bp downstream of lacOI , whereas Iac03 lies only 92 bp upstream of lacOI . The main contribution to repression comes from the DNA-looping of lacOI and Iac03 due to their closer proximity. Furthermore, when lacOI and Iac03 are bound by LacI, the production of LacI itself is prevented. The 3' end of the *lad* gene overlaps with Iac03. In a repressed state, transcription of *lad* results in a truncated mRNA, which is rapidly degraded by the cell. Due to this

autoregulation, the concentration of the Lad tetramer is ~40 molecules in induced cells and ~15 molecules in non-induced cells.

Sequences of lac operators are well known in the art. Exemplary lac operator sequences are provided by SEQ ID NO:3-5.

Suitable variants of the nucleic acid or polypeptide sequences, specifically Iac01 , Iac02 and Iac03, disclosed herein are functional variants having the same type of activity (without regard to the degree of the activity) as the nucleic acid or polypeptide to which the sequence corresponds. Such activities may be tested according to the assays described in the Examples below and according to methods known in the art.

The term **"functional variant"** or functionally active variant also includes naturally occurring allelic variants, as well as mutants or any other non-naturally occurring variants. As is known in the art, an allelic variant is an alternate form of a nucleic acid or peptide that is characterized as having a substitution, deletion, or addition of one or nucleotides or more amino acids that does essentially not alter the biological function of the nucleic acid or polypeptide.

Functional variants may be obtained by sequence alterations in the polypeptide or the nucleotide sequence, e.g. by one or more point mutations, wherein the sequence alterations retains or improves a function of the unaltered polypeptide or the nucleotide sequence, when used in combination of the invention. Such sequence alterations can include, but are not limited to, (conservative) substitutions, additions, deletions, mutations and insertions.

A point mutation is particularly understood as the engineering of a poly-nucleotide that results in the expression of an amino acid sequence that differs from the non-engineered amino acid sequence in the substitution or exchange, deletion or 5 insertion of one or more single (non-consecutive) or doublets of amino acids for different amino acids.

An exemplary functional variant of the Iac01 operator is a 2 base-pair truncated version of wild-type Iac01 , which comprises a deletion of 2bp at its 5' end, lacO* (SEQ ID NO:6).

Transcription rate control, also referred to as fine-tuning of protein production or "tunability" is highly relevant in bioprocessing. Bioprocesses are designed to maximally exploit the cells' synthesizing capacity during a maximal long period, yielding properly folded and processed protein. But, strong expression systems, such as e.g. the T7 expression system, are known to exhibit an "all-or-none" behavior, where the reduced

expression level in partially induced cultures is the result of the formation of subpopulations of fully induced and non-induced cells. Such problem is solved by the inducible expression system described herein which allows tunability, specifically single-cell tunability. In the inducible expression system described herein, the affinity of Lad to the at least one lacO of the promoter operably linked to the gene of interest is lower than the affinity of Lad to the lac operators lac01 and lac03 of the endogenous lac operon of the host. If the binding constant (K_a) of Lad to the at least one lacO at the gene of interest (GOI) is higher than the binding constant to the lacO at the lac-operon, the first Lad molecules, which are not inactivated by IPTG will preferentially bind to the lacO binding sites of the GOI instead of the lac03/lac01 on the lac-operon. Hence, autoregulation of Lad does not intervene and more Lad molecules are being produced leading to an overregulation of the system which results in a complete stop of transcription of the gene of interest in this cell. In particular, at low inducer concentrations, such a system leads to at least two distinct sub-populations, of POI producing and non-producing cells, as such expression systems stop their productivity, but still continue to grow.

In the inducible expression system described herein, however, the binding constant (K_a) of Lad to the at least one lacO at the gene of interest (GOI) is lower than the binding constant to the lacO at the lac-operon. Therefore, Lad preferentially binds to the operators of the endogenous lac operon, preventing transcription of the three lacZ, lacY and lacA genes and also preventing further production of Lad through the autoregulation of Lad, resulting in a homogenous population at any given inducer concentration.

As used herein, the term **"affinity"** or "binding affinity" refers to strength of association between a ligand and a receptor as defined by the dissociation and/or the association constant. Dissociation constant (Kd) is the rate constant of dissociation at equilibrium, defined as the ratio kofr/kon, wherein koffis the rate constant of dissociation of the ligand from the receptor and k_{o_n} is the rate constant of association of the ligand to the receptor. The Association constant (K_a) is the opposite of Kd. When K_a is high, Kd is low, and the ligand has a high affinity for the receptor (fewer molecules are required to bind 50% of the receptors).

Usually a binder is considered a high affinity binder with a dissociation constant of at least Kd<10^{-7} M, in some cases higher affinities are required such as, e.g. Kd<10^{-8} M, preferably Kd<10^{-9} M, even more preferred is Kd<10^{-10} M.

In the inducible expression system described herein, the binding affinity of Lad to the one or more lacO/lacOs of the gene of interest is lower than the affinity of Lad to the lac operators Iac01 and Iac03 of the endogenous lac operon. Specifically, lad binds to the lac operators Iac01 and Iac03 with a Kd of at least $Kd<10^{-7}$ M, preferably $Kd<10^{-8}$ M, preferably $Kd<10^{-9}$ M, even more preferred is $Kd<10^{-10}$ M. Specifically, Lad binds to the one or more lacO/lacOs of the gene of interest with a Kd that is increased by at least 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 or 100% or more. Consequently, Lad binds to the one or more lacO/lacOs of the gene of interest with a K_a that is about 5, 10, 15, 20, 30, 40, 50, 60, 70, 80 or 90% lower than the K_a of Lad to the lacOI and Iac03 of the endogenous lac operon.

Specifically, binding affinity is determined by an affinity ELISA assay. In certain embodiments binding affinity is determined by a BIAcore, ForteBio or MSD assay. In certain embodiments binding affinity is determined by a kinetic method. In certain embodiments binding affinity is determined by an equilibrium/solution method. Those skilled in the art can determine appropriate parameters to determine binding affinity of a ligand to a certain molecule. The binding affinity can be routinely determined by one skilled in the art.

“Sequence identity” or “percent (%) amino acid sequence identity” as described herein is defined as the percentage of nucleotides or amino acid residues in a candidate sequence that are identical with the nucleotides or amino acid residues in the specific nucleotide or polypeptide sequence to be compared (the “parent sequence”), after aligning the sequence and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

The term **"operably linked"** as used herein refers to the association of nucleotide sequences on a single nucleic acid molecule, i.e. the vector, in a way such that the function of one or more nucleotide sequences is affected by at least one other nucleotide sequence present on said nucleic acid molecule. For example, a promoter is operably linked with a coding sequence encoding the protein of interest, when it is capable of effecting the expression of that coding sequence. Specifically, such nucleic acids operably linked to each other may be immediately linked, *i.e.* without further elements or nucleic acid sequences in between or may be indirectly linked with spacer sequences or

other sequences in between. Specifically, in the context of a lac operator being operably linked to a promoter refers to the ability of the lac operator to regulate the ability of the promoter to control expression of the coding sequence under specific conditions. Such as the ability of the lac operator to inhibit promoter-dependent expression of the gene of interest when lac repressor protein is bound thereto.

The term **"heterologous"** as used herein with respect to a nucleotide or amino acid sequence or protein, refers to a compound which is either foreign, *i.e.* "exogenous", such as not found in nature, to a given host cell; or that is naturally found in a given host cell, e.g., is "endogenous", however, in the context of a heterologous construct, *e.g.,* employing a heterologous nucleic acid, thus "not naturally-occurring". The heterologous nucleotide sequence as found endogenously may also be produced in an unnatural, *e.g.,* greater than expected or greater than naturally found, amount in the cell. The heterologous nucleotide sequence, or a nucleic acid comprising the heterologous nucleotide sequence, possibly differs in sequence from the endogenous nucleotide sequence but encodes the same protein as found endogenously. Specifically, heterologous nucleotide sequences are those not found in the same relationship to a host cell in nature *(i.e.,* "not natively associated"). Any recombinant or artificial nucleotide sequence is understood to be heterologous. An example of a heterologous polynucleotide or nucleic acid molecule comprises a nucleotide sequence not natively associated with a promoter, *e.g.,* to obtain a hybrid promoter, or operably linked to a coding sequence, as described herein. As a result, a hybrid or chimeric polynucleotide may be obtained. A further example of a heterologous compound is a POI encoding polynucleotide or gene operably linked to a transcriptional control element, e.g., a promoter, to which an endogenous, naturally-occurring POI coding sequence is not normally operably linked.

The invention furthermore comprises the following items:

1. A genome-based expression system for production of a protein of interest (POI) in a prokaryotic host, comprising at least

a) an RNA polymerase (RNAP) gene,

b) a gene encoding a POI, comprising

- a coding sequence,

- a promoter operably linked to said coding sequence, wherein said promoter is recognized by the RNAP expressed from a), and

- at least one lac operator (lacO) within the sequence of said promoter; and

c) a *lacI* gene for expression of a lac repressor protein (Lad) comprising

- a *lacI* coding sequence,

- a lad promoter operably linked to the *lad* coding sequence, wherein the lad promoter is a wild-type lad promoter or a lad promoter which increases Lad expression;

wherein the expression rate of the protein of interest is regulated by an inducer binding LacI.

2. The genome-based expression system of item 1, wherein the gene encoding a POI contains (i) one lacO within the sequence of the promoter or (ii) one lacO within the sequence of the promoter and one lacO upstream of the first lacO.

3. The genome-based expression system of item 1 or 2, wherein the gene encoding a POI contains one lacO within the sequence of the promoter, and the lacI promoter is a promoter which increases LacI expression.

4. The genome-based expression system of any one of items 1 to 3, wherein the gene encoding a POI contains one lacO within the sequence of the promoter and one lacO upstream of the first lacO, and the lacI promoter is a promoter which increases LacI expression.

5. The genome-based expression system of any one of items 1 to 4, wherein the prokaryotic host is *Escherichia coli (E.coli).*

6. The genome-based expression system of any one of items 1 to 5, wherein the host is *E.coli* of the strain BL21 or K-12.

7. The genome-based expression system of any one of items 1 to 6, wherein the RNAP is a heterologous or homologous RNAP, preferably the RNAP is an RNAP homologous to the host, specifically it is an *E.coli* RNA polymerase, preferably the σ^{70} *E.coli* RNA polymerase.

8. The genome-based expression system of any one of items 1 to 7, wherein the promoter in b) of item 1 is selected from the group consisting of T5, T5N25, T7A1, T7A2, T7A3, lac, lacUV5, tac or trc.

9. The genome-based expression system of any one of items 1 to 8, wherein the lacI promoter is the lacI promoter which increases LacI expression, which is the $lacI^{Q}$ promoter (SEQ ID NO:1).

10. The genome-based expression system of any one of items 1 to 9, wherein the lac operator is a Iac01 (SEQ ID NO:3), Iac02 (SEQ ID NO:4) or Iac03 (SEQ ID NO:5).

11. The genome-based expression system of item 10, wherein the lac operator is a functional variant of lacOI , Iac02 or Iac03 with at least 65% sequence identity or a perfectly symmetric lacO.

12. The genome-based expression system of any one of items 1 to 11, wherein said promoter operably linked to the coding sequence encoding the protein of interest comprises an initial transcribed sequence (ITS), preferably a native T7A1 initial transcribed sequence (SEQ ID NO:2).

13. The genome-based expression system of any one of items 1 to 12, wherein the inducer is selected from the group consisting of isopropylthiogalactoside (IPTG), lactose, methyl^-D-thiogalactoside, phenyl^-D-galactose and ortho-Nitrophenyl- β-galactoside (ONPG).

14. The genome-based expression system of any one of items 1 to 13, wherein the gene for expression of a protein of interest contains one lacOI operator within the sequence of the promoter operably linked to the coding sequence and the native T7A1 initial transcribed sequence (SEQ ID NO:2), and wherein the lad promoter is a $lacI^Q$ promoter.

15. The genome-based expression system of any one of items 1 to 14, wherein the gene of interest contains two lac operators which are at least 92 or 94 basepairs (bps) apart, preferably 103, 105, 114, 116, 125, 127, 134, 136, 138 or 149 bps apart, wherein one lac operator is located within the sequence of the promoter operably linked to the coding sequence and the second lac operator is upstream of the promoter.

16. The genome-based expression system of any one of items 1 to 15, wherein the gene encoding the protein of interest is a heterologous gene.

17. The system of any one of items 1 to 16, wherein at least one lac operator of the lac operon of the prokaryotic host is genetically modified to increase its binding affinity to the lac repressor molecule Lacl.

18. A method of plasmid-free production of a protein of interest in a prokaryotic host, using the genome-based expression system of any one of items 1 to 17, comprising the steps of

a) inducing expression of the gene encoding the POI by addition of an inducer,

b) harvesting the POI,

c) isolating and purifying the POI, and optionally

d) modifying, and

e) formulating the POI.

19. An expression cassette comprising at least one heterologous gene configured to produce at least one heterologous POI, including

a) one or more coding sequences encoding the one or more POI,

b) a promoter operably linked to the one or more coding sequences, and

c) at least one lac operator (lacO) within the sequence of said promoter;

wherein the affinity of Lad to lacO of c) is lower than the affinity of Lad to the lac operators lac01 and lac03 of the endogenous lac operon of a host cell.

20. The expression cassette of item 19, wherein the heterologous gene configured to produce at least one heterologous protein of interest includes two lac operators, which are at least 92 or 94 bp apart, wherein one lac operator is located within the sequence of the promoter and the second lac operator is upstream of the promoter.

21. The expression cassette of item 19 or 20, further comprising a heterologous lad promoter, which is the $lacI^Q$ promoter (SEQ ID NO:1).

22. The expression cassette of any one of items 19 to 21, wherein the heterologous gene configured to produce at least one heterologous POI comprises a lacOI operator within the sequence of the promoter operably linked to the coding sequence and a native T7A1 initial transcribed sequence (SEQ ID NO:2).

23. A method of plasmid-free production of a protein of interest in a prokaryotic host on a manufacturing scale, using the expression cassette of any one of items 19 to 22, comprising the steps of

a. integrating the expression cassette into the chromosome of the prokaryotic host,

b. inducing expression of the gene encoding the POI by addition of an inducer,

c. harvesting the POI,

d. isolating and purifying the POI, and optionally

e. modifying, and

f. formulating the POI.

24. An inducible system for plasmid-free production of a protein of interest (POI) in a prokaryotic host, comprising at least

a) an RNA polymerase (RNAP) gene in the chromosome of the host,

b) a gene encoding a POI comprising

- a coding sequence,
- a promoter operably linked to said coding sequence, wherein said promoter is recognized by the RNAP expressed from a), and
- at least one lac operator (lacO) within the sequence of said promoter; and

c) a *lad* gene encoding a lac repressor protein (Lad) comprising

- a *lacI* coding sequence,
- a lad promoter operably linked to the *lad* coding sequence, wherein the lad promoter is a wild-type lad promoter or a lad promoter which increases Lad expression;

wherein the affinity of Lad to the one or more lacO / lacOs of b) is lower than the affinity of lad to the lac operators Iac01 and Iac03 of the endogenous lac operon of the host and wherein the expression rate of the POI is regulated by an inducer binding LacI.

25. The system of item 24, wherein at least one lac operator of the lac operon of the prokaryotic host is genetically modified to increase its binding affinity to the lac repressor molecule LacI.

The examples described herein are illustrative of the present invention and are not intended to be limitations thereon. Different embodiments of the present invention have been described according to the present invention. Many modifications and variations may be made to the techniques described and illustrated herein without departing from the spirit and scope of the invention. Accordingly, it should be understood that the examples are illustrative only and are not limiting upon the scope of the invention.

EXAMPLES

Example 1: Overview and Materials and Methods used in the Examples herein.

Aim of this work was to investigate the feasibility of the two constitutive phage-derived promoters **T5N25** and **T7A1** , recognized by the σ^{70} *E. coli* RNAP in terms of transcription efficiency, basal expression rates and tuning capacity. The promoter sequences were modified to contain either one, two or three lacO binding sites (SEQ ID NO:28-33). The seven promoter/operator combinations that were tested with the model

protein GFPmut3.1 are shown in Figure 1. Expression strength, tunability, basal expression and cell growth were investigated in plasmid-based and plasmid-free BL21 expression systems. The resulting set of production clones was cultivated and compared under fed-batch like conditions in micro-titer fermentations.

Strains and culture conditions. *Escherichia coli* K-12 NEE35-a [*fhuA2A(argF-lacZ)U169 phoA gin V44 Φ80 A(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*\ was obtained from New England Biolabs (MA, USA) and used for all cloning procedures. Linear DNA cartridges were integrated into the bacterial chromosome at the attTN7 site of *Escherichia coli* BL21 *[fhuA2 [lon] ompT gal [dcm] AhsdS]* (New England Biolabs, MA, USA). For reference experiments, the same strains were transformed with the respective plasmids. The soluble protein GFPmut3.1 was used as recombinant model protein (19).

Basic cloning methods like restriction endonuclease (REN) digest, agarose gel electrophoresis (AGE), ligation and transformation of *E. coli* plasmids were carried out according to Sambrook et al. (24).

For cloning purposes, cells were routinely grown in M9ZB-medium, recovered in SOC-medium and plated on M9ZB-agar. The following antibiotic concentrations were used: ampicillin (Amp) 100 µg/ml or 30 pg/ml, kanamycin (Kan) 50 µg/ml or 30 µg/ml and chloramphenicol (Cm) 20 pg/ml or 10 pg/ml for plasmid-based and plasmid-free expression systems, respectively.

Culture Conditions

The strains were cultured in the BioLector micro-fermentation system in 48-well Flowerplates® (m2p-labs, Baesweiler, Germany) as described by Torok et al. (23). The synthetic Feed in Time (FIT) fed-batch medium with glucose and dextran as carbon sources (m2p-labs GmbH, Baesweiler, Germany) was used. Immediately prior to inoculation 0.6 % (v/v) of the glucose releasing enzyme mix (EnzMix) was added. The GFPmut3.1 expression level was monitored at an excitation of 488 nm and an emission of 520 nm. The signal is given in relative fluorescence units [rfu]. The cycle time for all parameters was 20 min. The initial cell density was equivalent to an optical density of OD_{600} = 0.3. For inoculation, a deep frozen (-80 °C) working cell bank (WCB) ($O\ddot{u}_{\beta oo}$ = 2) was thawed and biomass was harvested by centrifugation (7500 rpm, 5 min). Cells were washed with 500 pL of the corresponding medium to remove residual glycerol and centrifuged; then, pellets were re-suspended in the total cultivation medium. All cultivations were prepared in three replicates at 30 °C for 22 h. Recombinant gene

expression was induced with 0.005 mM, 0.01 mM or 0.5 mM IPTG, respectively, 10 h after start of cultivation.

For fed-batch fermentations, cells were grown in a 1.5 L (1.2 L working volume, 0.4 L minimal volume) DASGIP® Parallel Bioreactor System (Eppendorf AG, DE) equipped with standard control units. The pH was maintained at 7.0 ± 0.05 by addition of 12.5 % ammonia solution (Thermo Fisher Scientific, MA/USA); the temperature was maintained at 37 ± 0.5 °C during batch phase and was decreased to 30 ± 0.5 °C during feed phase. The dissolved oxygen (O2) level was stabilized above 30 % saturation by controlling stirrer speed and aeration rate. Foaming was suppressed by addition of antifoam suspension (Glanapon, 2000, Bussetti, AT). For inoculation, a deep-frozen (-80 °C) working cell-bank vial was thawed and 1 ml (optical density at 600 nm = 1) was transferred aseptically to the bioreactor.

Feeding was imitated when the culture, grown to 6 g cell dry mass (CDM) in 0.6 L batch medium, entered the stationary phase. A fed-batch regime with an exponential carbon-limited substrate feed was used to provide a constant growth rate of 0.1 /h over 2.5 doubling times. The substrate feed was controlled by increasing the pump speed according to the exponential growth algorithm, $x = \chi_0 e^{\mu t}$, with superimposed feedback control of weight loss in the substrate bottle. The CDW yield coefficient on glucose was 0.3 g/g and the feed medium provided glucose and components sufficient to yield an additional 32 g of CDW. Induction of the expression system was performed by adding Isopropyl-b-D-thiogalactopyranoside (IPTG) to the reactor to yield a concentration of 10 µmol / g CDW. Preparation and composition of the minimal medium used in this experiment was previously described (17).

Strains

BL21^{Q} - in short: BQ

For the integration of the lacI^{Q} promoter in *E. coli* BL21 (New England BioLabs®Inc., MA/USA), the plasmid pETAmp-laclq was constructed. This plasmid contains the ampicillin resistance gene, flanked by FRT sites and the lad gene controlled by the lacI^{Q} promoter. The ampicillin resistance gene was amplified from pET1 1a using the overhang PCR technique in order to add FRT sites and the restriction sites BamHI (5') and Kpnl (3'). Following primers were used: BamHI-FRT-Amp-for and Kpnl-FRT-Amp-rev.

The pBR322 ori and the lad gene were amplified from pET30a using the overhang PCR technique in order to add a C -> T mutation within the lad promoter and

the restriction sites KpnI (5') and BamHI (3'). Following primers were used: KpnI-pBR322-for and BamHI-laciq-rev.

Linear DNA cartridges for genome integration were amplified using the Q5® High-Fidelity DNA Polymerase (New England BioLabs®Inc., MA/USA), according to the manufacturer's manual. Following primers were used: GI-laclq-for and GI-laclq-rev.

Integration into the bacterial chromosome occurred at the *lac*-operon site of *E. coli* BL21 (New England BioLabs®Inc., MA/USA), which carries the pSIM5 plasmid, as described by Sharan et al. (26).

Screening of positive clones and amplification of the integrated DNA cartridge was performed by basic colony PCR technique, using OneTaq® DNA Polymerase (New England BioLabs®Inc., MA/USA), according to the manufacturer's manual. Following primers were used: lacI/1_ext and laci/2_ext.

Primer AmpStop was used for sequencing the amplified DNA integration cartridge.

BL21 Q ::TN7<1 lacOAI -GFPmut3.1 -tZ> - in short: BQ<1lacO-A1>

The sequence of the **T7AI** promoter was adopted from (18) (designated as **P A-1/04)** and contains a 2 bp truncated lac01 sequence between the -10 and -35 promoter region. This promoter was ordered as gBlocks® Gene Fragment (Integrated DNA Technologies, IA/USA), containing a 5' spacer sequence from pET30a and the restriction sites SphI (5') and XbaI (3') and subsequently cloned into the pET30a-cer-tZENIT-GFPmut3.1 backbone. The new plasmid was designated as pETk1 lacOA1tZ.c-GFPmut3.1 .

Linear DNA cartridges for genome integration were amplified using the Q5® High-Fidelity DNA Polymerase (New England BioLabs®Inc., MA/USA), according to the manufacturer's manual. Following primers were used: TN7_1_pET30aw/oKanR_for and TN7_2_pET30a_for.

Integration into the bacterial chromosome occurred at the attTN7 site of *E. coli* BL21 Q, which carries the pSIM5 plasmid, as described by Sharan et al. (26).

Following primers were used for screening of positive clones: TN7/1_ext and TN7/2_ext.

Primer seq_MCS-for and seq_MCS-rev were used for sequencing the amplified DNA integration cartridge.

BL21Q::TN7<1lacOT5-GFPmut3.1-tZ> - in short: BQ<1lacO-T5>

The sequence **T5N25** promoter was adopted from (18) and contains a 2 bp truncated lac01 sequence between the -10 and -35 promoter region. The initial

transcribed sequence (ITS) between +1 and +20 of **T5N25** was exchanged by the ITS of **T7AI** (21). This promoter was ordered as gBlocks® Gene Fragment (Integrated DNA Technologies, IA/USA), containing a 5' spacer sequence from pET30a and the restriction sites SphI (5') and XbaI (3') and subsequently cloned into the pET30a-cer-tZENIT-GFPmut3.1 backbone. The new plasmid was designated as pETk1 lacOT5tZ.c-GFPmut3.1 .

BL21 ::TN7<2lacOA1-GFPmut3.1-tZ> and BL21 ::TN7<2lacOT5-GFPmut3.1-tZ> - in short: B<2lacO-A1> and B<2lacO-T5>

Besides an increased level of lacI by the lacI^Q promoter, a second lacO can reduce the basal expression, by enabling DNA loop formation. For the addition of a second lac01 sequence, 62 bp upstream of the first lac01 , an overhang PCR was performed with the templates pETk1 lacOA1tZ.c-GFPmut3.1 or pETk1 lacOT5tZ.c-GFPmut3.1 , respectively. The forward primer (2lacO-for) contains the *lac*-operator and the restriction site SphI (5'), the reverse primer (2lacO-rev) contains the restriction site NdeI (3'). The new plasmids were designated as pETk2lacOA1tZ.c-GFPmut3.1 and pETk2lacOT5tZ.c-GFPmut3.1 .

Integration into the bacterial chromosome occurred at the attTN7 site of *E. coli* BL21 (New England BioLabs®Inc., MA/USA).

Amplification of linear DNA cartridge and screening was carried out as previously described.

Construction and characterization of promoter/operator combinations.

Basic cloning methods like restriction endonuclease (REN) digest, agarose gel electrophoresis (AGE), ligation and transformation of *E. coli* plasmids were carried out according to Sambrook et al. (24). For the integration of the lacI^Q promoter in *E. coli* BL21 (New England BioLabs®Inc., MA/USA), the plasmid pETAmp-laclq was constructed. This plasmid contains the ampicillin resistance gene, flanked by FRT sites and the lacI gene controlled by the lacI^Q promoter (25). The pBR322 ori and the lacI gene were amplified from pET30a using the overhang PCR technique in order to add a C -> T mutation within the lacI promoter. The linear lacI^Q DNA cartridge for genome integration was amplified using the Q5® High-Fidelity DNA Polymerase (New England BioLabs®Inc., MA/USA), according to the manufacturer's manual. Integration into the bacterial chromosome occurred at the lac-operon site of *E. coli* BL21 , which carries the pSIM5 plasmid, as described by Sharan et al. (26). This strain got the designation BL21^Q. The sequences of the **T7AI** and the **T5$_{N25}$** promoter were adopted from Lanzer

and Bujard (18) (designated as $_{PAI/O4}$ and $_{PN25/04)}$ and contain a 2 bp truncated lac01 sequence between the -10 and -35 promoter region. These promoters were ordered as gBlocks® Gene Fragments (Integrated DNA Technologies, IA/USA), containing a 5' spacer sequence from pET30a and the restriction sites SphI (5') and XbaI (3') and subsequently cloned into the pET30a-cer-tZENIT-GFPmut3.1 backbone. The tZENIT terminator is described elsewhere (27). A second lac01 sequence, 62 bp upstream of the first lac01 , was added via overhang PCR. The 3lacO-T5 promoter/operator combination was adopted from pJexpress 401-406 (T5) vector from ATUM (CA/USA). Linear DNA cartridges were integrated into the bacterial chromosome at the attTN7 site of *E. coli* BL21 or BL21^{Q}.

GFPmut3.1 off-line expression analysis and quantification

Recombinant GFPmut3.1 was quantified by ELISA according to Reischer et al. (28). SDS-PAGE analysis was performed as previously described (29).

Flow cytometry

A Gallios flow cytometer (Beckman Coulter, CA/USA) was used to determine the fraction of GFPmut3.1 -producing cells. Cells were harvested 12h after induction and then diluted 1/2025 in PBS. Excitation of GFPmut3.1 fluorescence was performed using an OPSL Sapphire Laser at 488 nm, with subsequent emission being measured through use of the FL1 Channel (505-545). Data were recorded for 15000 cells per sample at ~ 300 events/sec and analyzed with Kaluza analysis software (Beckman Coulter).

LacI western blot and quantification

Cell extracts obtained from ~1.2 x 10^{7} BL21-wt and B<2lacO-A1> cells were separated by SDS-PAGE as previously described (29). After separation, the proteins were bloted on the provided membrane using the iBIot® Dry Blotting System according to the manufacture's manual (Invitrogen™/ Thermo Fisher Scientific, CA/USA). Subsequently, proteins were blocked 4 hours at room temperature with 3 % nonfat dry milk in PBST (1x PBS Dulbecco and 0.05 % Tween 20). The blot was then incubated with primary antibody (1:1000 anti-LacI Antibody, clone 9A5 (Sigma-Adrich/ Merck, MO/USA) 1 hour at room temperature. It was then incubated with alkaline phosphatase conjugated secondary antibody (1:2000 Anti-Mouse IgG (whole molecule) - Sigma A51 53 (Sigma-Adrich/ Merck, MO/USA) for 1 hour at room temperature and developed with SigmaFAST™ BCIP®/NPT tablets (Sigma-Adrich/ Merck, MO/USA) according to the manufacturer's manual. Band intensities were quantified with ImageQuant TL software (GE Healthcare, IL/USA).

Table 1. Primers used in the Examples. Underlined: binding part of overhang primers, italic: overhang, bold uppercase letters: restriction sites, lowercase letters: lacOI , bold lower-case letter: FRT-sites, underlined bold uppercase: C->T mutation in $lacI^Q$ promoter.

Name	Sequence 5' – 3'
BamHI-FRT-Amp-for	*CAAGTCG**GATC**CGAT**gaagttcctattctctagaaagtataggaacttcc*** AGAAAAAAAGGATCTCAAGAAG (SEQ ID NO:7)
KpnI-FRT-Amp-rev	*ACGGGGTCG**GTACC**CCT**gaagttcctatactttctagagaataggaacttc*** GTTAGCAATTTAACTGTGATAAAC (SEQ ID NO:8)
KpnI-pBR322-for	*AGGG**GTAC**C* GACCCCGTAGAAAAGATCAAAGGATC (SEQ ID NO:9)
BamHI-laciq-rev	*ATCG**GATC**CGACATCC*CGGACACCATCGAATGGTGCAAAAC (SEQ ID NO:10)
GI-lacIq-for	CGTTACTGGTTTCACATTCACCAC (SEQ ID NO:11)
GI-lacIq-rev	*CGCAGGCTATTCTGGTGGCCGGAAGGCGAAGCGGCATGCATTTACGTTGA* CCTTTGATCTTTTCTACGGGGTCGG (SEQ ID NO:12)
lacI/1_ext	CGTAAAAATGCGCTCAGGTCAAATTCAG (SEQ ID NO:13)
laci/2_ext	CAGATCGAAGAAGGGGTTGAATCGC (SEQ ID NO:14)
AmpStop	TCAGGCAACTATGGATGAAC (SEQ ID NO:15)
TN7_1_pET30aw/oKanR_for	*AGATGACGGTTTGTCACATGGAGTTGGCAGGATGTTTGATTAAAAACATA* GTAGTAGGTTGAGGCCGTTG (SEQ ID NO:16)
TN7_2_pET30a_for	*CAGCCGCGTAACCTGGCAAAATCGGTTACGGTTGAGTAATAAATGGATGC* GAAGATCCTTTGATCTTTTCTACG (SEQ ID NO:17)
TN7/1_ext	ACCGGCGCAGGGAAGG (SEQ ID NO:18)
TN7/2_ext	TGGCGCTAATTGATGCCG (SEQ ID NO:19)
2lacO-for	*GT**GCATGC**tTACACGTACTTAGTCGCTGAAaattgtgagcggataacaatt* CCATACCCACGCCGAAA (SEQ ID NO:20)
2lacO-rev	CTTTGCT**CATATG**TATATCTCCTTC (SEQ ID NO:21)
seq_MCS-for	GTAGTAGGTTGAGGCCGTTG (SEQ ID NO:22)
seq_MCS-rev	CGGATATAGTTCCTCCTTTCAG (SEQ ID NO:23)

Table 2. gBlocks® Gene Fragments used in the Examples. bold uppercase letters: restriction sites, bold and italic: -35 and -10 region, underlined: lacOI*, bold lowercase letters: native ITS of $T7_{A1}$ promoter.

Name	Sequence 5' – 3'
T5A1	GAATGGT**GCATGC**AAGGAGATGGCGCCCAACAGTCCCCCGGCCACGG GGCCTGCCACCATACCCACGCCGAAACAAGATCATAAAAAATTTAT***TTGC TT***TGTGAGCGGATAACAAT***TATAAT***AGATTC**atcgagagggacacggcgaa**ctct **aga**ACGGATATAGTCCTTCAG (SEQ ID NO:24)
A1A1	GAATGGT**GCATGC**AAGGAGATGGCGCCCAACAGTCCCCCGGCCACGG GGCCTGCCACCATACCCACGCCGAAACAAGTTTATCAAAAAGAGTG***TTG ACT***TGTGAGCGGATAACAAT***GATACT***TAGATTC**atcgagagggacacggcgaa** c**tctaga**ACGGATATAGTCCTTCAG (SEQ ID NO:25)

Table 3. Promoter sequences used in the Examples. Promoter sequences were cloned into pET30a-cer plasmid via SphI and NdeI restriction sites. Italic upper-case letters: restriction sites, lower case letters: lac operators, underlined: core promoter sequence, italic bold upper-case letters: -35 and -10 promoter elements, italic bold lower case letters: ribosomal binding site, bold upper case letters: +1 T7A1 +20 initial transcribed sequence.

Name	Sequence 5' – 3'
3lacO-T5	*GCATGC* TTACACGTACTTAGTCGCTGAA aattgtgagcggataacaatt ACGAGCTTCATGCACAGTTAA ATCATAAAAAATTTAT ***TTGCTT*** tgtgagcggataacaat ***TATAAT*** A tgtggaattgtgagcgctcacaattccaca ACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAG ***aaggag*** ATATA *CATATG* (SEQ ID NO:28)
2lacO-T5	*GCATGC* TTACACGTACTTAGTCGCTGAA aattgtgagcggataacaatt CCATACCCACGCCGAAACAAG ATCATAAAAAATTTAT ***TTGCTT*** tgtgagcggataacaat ***TATAAT*** AGATTC **ATCGAGAGGGACACGGCGAA** CTCTAGAAATAATTTTGTTTAACTTTAAG ***aaggag*** ATATA *CATATG* (SEQ ID NO:29)
1lacO-T5	*GCATGC*AAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGC CACCATACCCACGCCGAAACAAG ATCATAAAAAATTTAT ***TTGCTT*** tgtgagcggataacaat ***TATAAT*** AGATTC **ATCGAGAGGGACACGGCGAA** CTCTAGAAATAATTTTGTTTAACTTTAAG ***aaggag*** ATATA *CATATG* (SEQ ID NO:30)
2lacO-A1	*GCATGC* TTACACGTACTTAGTCGCTGAA aattgtgagcggataacaatt CCATACCCACGCCGAAACAAG ATCATAAAAAAGAGTG ***TTGACT*** tgtgagcggataacaat ***GATACT*** TGATTC **ATCGAGAGGGACACGGCGAA** CTCTAGAAATAATTTTGTTTAACTTTAAG ***aaggag*** ATATA *CATATG* (SEQ ID NO:31)
1lacO-A1	*GCATGC*AAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGC CACCATACCCACGCCGAAACAAG TTTATCAAAAAGAGTG ***TTGACT*** tgtgagcggataacaat ***GATACT*** TAGATTC **ATCGAGAGGGACACGGCGAA** CTCTAGAAATAATTTTGTTTAACTTTAAG ***aaggag*** ATATA *CATATG* (SEQ ID NO:32)

T7	*GCATGC*AAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGC CACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGC CCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGC ACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGA TCGAGATCGATCTCGATCCCGCGAAAT TAATACGACTCACTATAGG ggaattgtgagcggataacaattcc CCTCTAGAAATAATTTTGTTTAACTTTAAG ***aaggag*** ATATA *CATATG* (SEQ ID NO:33)

Example 2: Productivity of Host RNAP Dependent Promoters/Operator combinations

The T7 expression system is known to provide high expression rates, even from a single target gene copy, integrated into the *E. coli* genome. First it was tested whether the same productivity can be reached by σ^{70} *E. coli* RNAP dependent promoters in the same experimental set-up. Therefore, plasmid-free and plasmid-based T5N25 and T7A1 promoter/operator combinations were compared with the T7 expression system. The cells were grown in fed-batch like conditions in micro-titer fermentations over a period of 22 hours. Expression of GFP was induced by a single pulse of IPTG of 0.5 mmol/L after 10 hours.

In all promoter/operator combinations, the cells were able to maintain growth during the production period of 12 hours in the micro-titer fermentations. An average growth rate of $\mu = 0.05\ h^{-1}$ allowed for direct comparison of the T7 and the host RNAP dependent promoters.

In plasmid-based expression systems, results from on-line fluorescence measurements of GFPmut3.1 were in a similar range as the T7 expression system for all promoter/operator combinations, except for B(3lacO-T5). (Figure 2B). These results were confirmed by SDS-PAGE analyses. However, in genome-integrated expression systems, quite distinctive differences of the respective promoter/operator combinations could be observed (Figure 2A). As compared to the T5 expression systems, GFPmut3.1 yields were 1.5-fold higher in the A1 expression systems. In the genome-integrated T7 expression system, induction of GFP gene expression led to 145 rfu and a specific product concentration (Yp/x) of ~ 135 mg/g soluble GFPmut3.1 and negligible amounts inclusion bodies (IBs). The same experiment with the A1 expression systems yielded almost 50 rfu and 37 mg/g soluble GFPmut3.1 without IBs.

The observed reduced productivity of B(3lacO-T5) and B<3lacO-T5> may result from the perfectly symmetric /ac-operator (sym-lacO) (7) at the initial transcribed sequence (ITS) which has an influence on promoter escape and therefore, productivity

(21). This effect was less visible in the plasmid-based 3lacO-T5 expression system, where the high plasmid copy number compensates for the reduced promoter activity. However, since in the plasmid-free expression system, the promoter activity was quite low, the three lacO version was dismissed for the A1 promoter. For one and two lacO promoter/operator combinations, the sym-lacO was replaced by the native ITS of the A1 promoter (+1 - +20). This resulted in a 2.4-fold increase in productivity in case of the T5 promoter. However, a reduction in lacO binding sites leads inevitably to increased basal expression.

Example 3: Basal Expression in Host RNAP Dependent Expression Systems

For challenging proteins even low basal expression can have adverse effects on host metabolism. Sometimes transformation of plasmids or integration cartridges lead to toxicity and it is difficult to obtain transformants. Therefore, tightness of gene regulation is an important quality criterion of expression systems.

In plasmid-based systems, promoters that were controlled by one /ac-operator (1lacO) showed the highest basal expression at a level of ~ 10 rfu, especially under C-limited conditions. The addition of a second lacO (2lacO) or the increase of the inhibitor Lad by introducing the lacI^{Q} promoter reduced the basal expression of the A1 promoter to 50%. In case of the T5 promoter, only the combination of three lac-operators (3lacO) reduced basal expression to almost 0 rfu. In contrast to the plasmid-based expression systems, in all genome integrated systems a significant impact of the promoter/operator combination on systems leakiness could be observed. Both, the increase of Lad molecules or the addition of a second lacO reduced the basal expression of A1 expression systems from 14 rfu to nearly no significant background expression and without reduction in productivity (Figure 2A). Although, both promoters contain the *lac-operators* in the identical position, only an increased level of Lad molecules or three *lac-operators* reduced basal expression of T5 expression systems sufficiently. The $_{T7A1}$ promoter is recognized by RNAP only half as efficiently as $_{T5N25}$ (20) and as one /ac-operator is located within the promoter sequence between the -10 and -35 promoter elements, host RNAP and /ac-repressor compete each other for their respective binding site which determines how efficiently promoter activity is controlled by repressors.

Example 4: Control of Recombinant Gene Expression Rate

Transcription rate control, also referred to as fine-tuning of protein production or “tunability” is highly relevant in bioprocessing. Optimal bioprocesses are designed to

maximally exploit the cells' synthesizing capacity during a maximal long period, yielding proper folded and processed protein. Depending on the physical properties and metabolic requirements of the desired product, the transcription rates must be adapted, to be in accordance with RNA stability, translation efficiency, folding, transport an all other interactions within the system.

To evaluate the tunability of the promoter/ operator combination described herein, a series of fed-batch like microtiter cultures at varying IPTG levels were tested and compared to the plasmid-free *11* expression system. Induction was performed using a single pulse of 0.005, 0.01 and 0.5 mM IPTG. On-line fluorescence measurement and end-point flow cytometry analysis were used to characterize the different promoter/operator combinations.

Expression systems, controlled by one lacO for gene regulation, exhibited not only the highest basal expression but also the least pronounced graduation of GFPmut3.1 expression at the given inducer concentrations (Figure 3C, F). Although, promoters with two lacOs showed sufficiently low basal expression, they produced significantly less at lower inducer concentrations (Figure 3B, E). The promoter/operator combinations 3lacO-T5 and 2lacO-A1 lead to a complete production stop of recombinant GFP after a certain time, independently of inducer concentration (Figure 3 A, E). This behavior was not observed in promoter/operator combinations with only one lacO. Promoters controlled by one lacO, the $lacI^Q$ (Figure 3D, G) and the *11* expression system (Figure 3H) combine the desired properties of low systems leakiness and tunability.

However, the *11* expression system is known to exhibit an "all-or-none" behavior, where the reduced expression level in partially induced cultures is the result of the formation of subpopulations of fully induced and non-induced cells, as reviewed in (22). To answer the question, if single-cell tunability in host RNAP dependent expression systems is possible, flow cytometry analysis of all promoter/operator combinations was performed. As shown in Figure 4, the genome-integrated *11* expression system exhibits no homogeneous population in partially induced cultures. In fact, a mixture of fully, partially and not induced cells was found particularly at very low inducer concentrations. In the E3<2lacO-A1 > expression system, the flow cytometry analysis revealed two distinct sub-populations of producing and non-producing cells (Fig. 4), as these expression systems stopped their productivity, but still continued to grow. This behavior was also observed in E3<3lacO-T5>. This was different for E3Q<1 lacO-A1 >, where the induction of GFP resulted in homogenous populations at any given IPTG concentration (Fig 4).

Based on these findings, it appears that the complete stop in productivity of all other expression systems when partially induced is associated with the autoregulation of the *lac* inhibitor. The *lac*-operon is regulated by 3 lacO binding sites (Figure 5A). The Lad molecule binds to either Iac01 and Iac03 or Iac01 and Iac02. Lac03 overlaps with the 3' end of the lad gene. The binding of Lad to Iac01 and Iac03 causes a loop formation of the DNA and results in truncated lad mRNA molecules, which are digested by the cell. This results in a constant level of ~40 molecules in fully induced cells and ~15 molecules in non-induced cells.

If the binding constant (K_a) of Lad to lacO at the gene of interest (GOI) is higher than the binding constant to the lacO at the lac-operon, the first Lad molecules, which are not inactivated by IPTG will preferentially bind to the lacO binding sites of the GOI instead of the Iac03/lac01 on the lac-operon. Hence, autoregulation of Lad does not intervene and more Lad molecules are being produced (Figure 5B). The whole system becomes over regulated and results in a complete stop in production.

To support this hypothesis, the effect of autoregulation on Lad levels of B<2lacO-A1> and BL21 wild-type (BL21-wt) cells was compared. The Lad content of non-induced, partially and fully induced cells was estimated using western blot analysis. The band intensities were quantified and normalized with the cell number (Figure 7).

In fully induced BL21 wild-type cells, the amount of Lad molecules was 3.5-fold, compared to non-induced BL21 wild-type cells. Partially induction with 0.01 mM IPTG only led to a 0.3-fold increase. The fold change of 3.5 in fully induced BL21-wt cells is in accordance with the results of Semsey et al., who measured on average 15 Lad molecules per cell in the absence of inducer and ~40 molecules in fully induced cells (11). In B<2lacO-A1>, Lad amounts of non-induced and partially induced cells were clearly higher corn-pared to BL21 wild-type. Lad yields were 2.3-fold in the absence of inducer and 2.7-fold in partially induced cells relative to BL21-wt. In fully induced cells, Lad yields were 4.0-fold, which corresponds with the fully induced wild-type BL21.

Although the addition of 0.01 mM IPTG results in almost half-maximal GFPmut3.1 expression (Figure 3), it has almost no influence on Lad levels. Obviously, Lad is still able to bind to Iac01/lac03 in the lac operon, hence maintaining its autoregulation under these conditions. In addition to that, the *lad* gene is transcribed from a weak promoter resulting in about one new mRNA per cell generation (38), unlike the strong **T7AI** promoter. Yet, the high Lad levels in non-induced and partially induced B<2lacO-A1>

cells clearly support our hypothesis of the impact of Lad autoregulation on expression rate control in genome-integrated *E. coli* production strains as depicted in (Figure).

The effect of Lad autoregulation was only observed in genome-integrated host RNAP dependent expression systems, which are controlled by two or three lac operators. However, this effect was not observed in plasmid-based host RNAP dependent expression systems or in the conventional T7 expression system. The reason for this can be seen in the balance of lac operators to Lad concentration. The T7 expression system harbors a further lad gene sequence within its DE3 lysogen, thus theoretically a doubling of the Lad concentration per cell. The plasmid-based expression systems used in this work are based on the pET plasmid system that encode a further lad gene sequence. That in turn results in further 15-20 lad gene sequences, depending on the plasmid copy number. However, the effect of Lad autoregulation on partially induced cells can also be observed in plasmid-based expression systems as seen in the case of *E. coli* pAVEway™ expression system from Fujifilm Diosynth Biotechnologies (NC/USA). In this plasmid-based expression system, transcription control is enabled by two perfectly symmetric lac operators, one positioned upstream of the T7A3 promoter and one downstream. The high affinity of Lad to the symmetric lac operators combined with the ability of DNA loop formation results in very low basal expression but exhibits also a complete stop in productivity in partially induced cultures.

Considering the autoregulation of the lac-inhibitor, a promoter/operator combination, which fulfils the desired properties such as high expression rate, negligible basal expression and true control of expression rate even at low inductor concentrations without a complete stop of productivity could successfully be identified.

Conclusion

The regulation of transcription in *E. coli* is receiving considerable attention because it is the first step in the process of recombinant protein production. Transcription control allows a cell to assign its resources towards the production of the recombinant protein and a tight and tunable control is essential for successful bioprocesses. It is evidenced herein that in plasmid-free expression systems, the regulatory elements of the *lac*-operon must be well balanced to control host RNAP dependent promoters. Three lac-operators reduce basal-expression to negligible amounts, but also the recombination production rate. The perfectly symmetric lacO in the initial transcribed sequence (ITS) hampers promoter escape of the RNAP. As shown by Hsu et al., the wild-type ITS of

T7A1 exhibits an enrichment of purines and one of the best promoter escape properties (21).

Promoters containing only one lacO exhibit considerable higher promoter strength, but also higher systems leakiness. In promoter/operator combinations containing two lacOs, the two lac01 in a distance of 62 bp at the site of the GOI exhibit a very strong binding affinity to the repressor molecule and thus prevent lad autoregulation which results in a complete stop in productivity in partially induced cells. However, the binding affinity can be reduced by the use of less symmetric lacOs like lac03 or lac02 or by varying the distance between them (see Example 5).

As demonstrated herein, the combination of one lacO with an increased level of intracellular Lad caused by the lacI^{Q} promoter results in high expression rates, low basal expression and true tunability on a cellular level. Thus, this novel expression system is specifically suitable for the production of challenging proteins, as there is no plasmid-mediated metabolic load and by using the host RNAP the genetic stability increases.

Importantly, the inducible system described herein demonstrates significantly improved expression rates, reduced basal expression and true tunability compared to the T7 expression system (see e.g. Figures 3 and 4). The inducible expression system described herein fulfills all desired properties that are required for an efficient expression system, such as high expression rate, negligible basal expression and true control of expression rate that is steplessly adjustable, even at low inducer concentrations.

Example 5: Control of Recombinant Gene Expression Rate in an Inducible Expression System Comprising Two lacOs.

Strains: BL21 ::TN7<2lac0.xxA1-GFPmut3.1-tZ> and E3L21 ::TN7<2lac0.xxT5-GFPmut3.1-tZ> - in short: E3<2lac0.xx-A1 > and E3<2lac0.xx-T5>

For the addition of a second lac01 sequence at a bigger distance to the first lac01 than 62bp, an overhang PCR is performed using the templates pETk1 lacOA1tZ.c-GFPmut3.1 or pETk1 lacOT5tZ.c-GFPmut3.1 , respectively. The two lac01 operators are 92, 103, 114 or 125 bp apart. The forward primers 2lac0.92-for, 2lacO.103-for, 2lac0.1 14-for and 2lac0.125-for contain the lac-operator and the restriction site SphI (5'), the reverse primer (2lacO-rev) contains the restriction site NdeI (3'). The new plasmids are designated as pETk2lac0.92A1tZ.c-GFPmut3.1 , pETk2lacO.103A1tZ.c-GFPmut3.1 , pETk2lac0.1 14A1tZ.c-GFPmut3.1 , pETk2lac0.125A1tZ.c-GFPmut3.1 and pETk2lac0.92T5tZ.c-GFPmut3. 1, pETk2lacO. 103T5tZ.c-GFPmut3. 1, pETk2lacO. 114T5tZ.c-GFPmut3. 1, pETk2lacO. 125T5tZ.c-GFPmut3. 1.

Integration into the bacterial chromosome occurs at the attTN7 site of *E. coli* BL21 (New England BioLabs®Inc., MA/USA).

Amplification of linear DNA cartridge and screening is carried out as described above.

Example 6: Fab production using BQ<1lacO-A1 > in Fed-Batch Culture

The T7 based expression system shows a unique strength sufficient for high expression rates even from a single copy. For systems with a single copy of the GOI under control of a host RNAP specific promotor significantly decreased expression rates are expected. Consequently, such systems will not be competitive in case when recombinant proteins must be produced at high levels. The situation is different for antibody fragments and other challenging proteins where the final product yield is definitely not determined by the strength of the promoter system but by currently un-identified reasons. To investigate these aspects, the BQ<1 lacO-A1 > expression system was selected for the production of the leader/Fab combination dsbA/ FTN2 (dFTN2) and was compared with B3<T7> producing the same leader/Fab combination. The cells were grown in fed-batch mode at a constant growth rate of 0.1/h feed of defined medium. In the experiment the amount of cell dry weight to be produced is pre-defined to 40 g CDW. Recombinant gene expression was induced by single pulse of IPTG of 10 pmol/gCDW at 0.5 doublings past feed start.

The results in Figure 8 and Figure 9 are given in total specific content of recombinant Fab per cell dry weight (mg/g), which is the sum of extra-cellular Fab measured in the fermentation supernatant and cellular Fab. In the T7-based system (Figure 8), induction of dFTN2 expression led to a maximum cellular Fab concentration of 1.8 mg/g 11 hours after induction and dropped to 0.7 mg/g at end of fermentation (Figure 8, open diamonds). At this time period, extra-cellular Fab increased from almost 0 mg/g to 2.2 mg/g (Figure 8, open triangles). This results in a maximum total Fab concentration of 3.5 mg/g 15 hours after induction which dropped to 2.1 mg/g at the end of fermentation (Figure 8, black dot). The increase of extra-cellular Fab in the fermentation supernatant can be attributed to cell lysis, which could be verified by measuring the DNA content in the fermentation supernatant.

The same experiment with the BQ<1 lacO-A1 > expression system yielded significantly improved results (Figure 9). The content of cellular Fab could be maintained at 2.5 mg/g during the whole fermentation (Figure 9, open diamonds). Extra-cellular Fab content increased to 2.4 mg/g at the end of fermentation (Figure 9, open triangle). This

results in a maximum total Fab concentration of 4.7 mg/g at the end of fermentation (Figure 9, black dot). Although the relative promoter strength of 1lacO-A1 is about 30 % compared to T7, this expression system yielded the same amount of total Fab as the strong T7 expression system until 15 hours after feed start and exceeded the T7 system at the end of fermentation by factor 2. These results clearly show, that a reduced promoter strength can be beneficial for the production of challenging proteins, as it decreases the metabolic burden of the cell and stress-induced proteolysis.

REFERENCES

1. Angius, F., Ilioaia, O., Amrani, A., Suisse, A., Rosset, L , Legrand, A., Abou-Hamdan, A., Uzan, M., Zito, F., and Miroux, B. (2018) A novel regulation mechanism of the T7 RNA polymerase based expression system improves overproduction and folding of membrane proteins, *Scientific reports 8,* 8572.
2. Chia-Chang Hsu, O. R. T. T. a. T. W. O. (201 5) Periplasmic expression in and release of Fab fragments from Escherichia coli using stress minimization, *Journal of Chemical Technology & Biotechnology 91.*
3. Saida, F., Uzan, M., Odaert, B., and Bontems, F. (2006) Expression of highly toxic genes in E. coli: special strategies and genetic tools, *Current protein & peptide science* 7, 47-56.
4. Riggs, A. D., and Bourgeois, S. (1968) On the assay, isolation and characterization of the lac repressor, *J Mol Biol 34,* 361-364.
5. Barkley, M. D., Riggs, A. D., Jobe, A., and Burgeois, S. (1975) Interaction of effecting ligands with lac repressor and repressor-operator complex, *Biochemistry 14,* 1700-1712.
6. Oehler, S., Eismann, E. R., Kramer, H., and Muller-Hill, B. (1990) The three operators of the lac operon cooperate in repression, *EMBO J 9,* 973-979.
7. Sadler, J. R., Sasmor, H., and Betz, J. L. (1983) A perfectly symmetric lac operator binds the lac repressor very tightly, *Proc Natl Acad Sci U S A 80,* 6785-6789.
8. Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B., and Muller-Hill, B. (1994) Quality and position of the three lac operators of E. coli define efficiency of repression, *EMBO J 13,* 3348-3355.
9. Mossing, M. C., and Record, M. T., Jr. (1986) Upstream operators enhance repression of the lac promoter, *Science 233,* 889-892.
10. Reznikoff, W. S., Winter, R. B., and Hurley, C. K. (1974) The location of the repressor binding sites in the lac operon, *Proc Natl Acad Sci U S A 71,* 2314-2318.
11. Semsey, S., Jauffred, L., Csiszovszki, Z., Erdossy, J., Steger, V., Hansen, S., and Krishna, S. (201 3) The effect of Lad autoregulation on the performance of the lactose utilization system in Escherichia coli, *Nucleic Acids Res 41,* 6381-6390.
12. Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in Escherichia coli: advances and challenges, *Frontiers in microbiology 5,* 172.
13. Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *J Mol Biol 189,* 113-1 30.
14. Mairhofer, J., Scharl, T., Marisch, K., Cserjan-Puschmann, M., and Striedner, G. (201 3) Comparative transcription profiling and in-depth characterization of plasmid-based and plasmid-free Escherichia coli expression systems under production conditions, *Applied and environmental microbiology 79,* 3802-3812.

15. Glick, B. R. (1995) Metabolic load and heterologous gene expression, *Biotechnology advances 13,* 247-261 .
16. Marchand, I., Nicholson, A. W., and Dreyfus, M. (2001) High-level autoenhanced expression of a single-copy gene in Escherichia coli: overproduction of bacteriophage T7 protein kinase directed by T7 late genetic elements, *Gene 262,* 231-238.
17. Striedner, G., Pfaffenzeller, I., Markus, L , Nemecek, S., Grabherr, R., and Bayer, K. (2010) Plasmid-free T7-based Escherichia coli expression systems, *Biotechnology and bioengineering 105,* 786-794.
18. Lanzer, M., and Bujard, H. (1988) Promoters largely determine the efficiency of repressor action, *Proc Natl Acad Sci U S A 85,* 8973-8977.
19. Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP), *Gene 173,* 33-38.
20. Deuschle, U., Kammerer, W., Gentz, R., and Bujard, H. (1986) Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures, *EMBO J 5,* 2987-2994.
21. Hsu, L. M., Cobb, I. M., Ozmore, J. R., Khoo, M., Nahm, G., Xia, L., Bao, Y., and Ahn, C. (2006) Initial transcribed sequence mutations specifically affect promoter escape properties, *Biochemistry 45,* 8841-8854.
22. Marschall, L., Sagmeister, P., and Herwig, C. (2017) Tunable recombinant protein expression in E. coli: promoter systems and genetic constraints, *Applied microbiology and biotechnology 101,* 501-512.
23. Toeroek, C., Cserjan-Puschmann, M., Bayer, K., and Striedner, G. (2015) Fed-batch like cultivation in a micro-bioreactor: screening conditions relevant for Escherichia coli based production processes, *SpringerPlus 4,* 490.
24. Green, J. F. S. a. M. R. (2012) *Molecular Cloning: A Laboratory Manual,* Vol. 4, Cold Spring Harbor Laboratory Press, U.S.
25. Mullerhill, B., Crapo, L., and Gilbert, W. (1968) Mutants That Make More Lac Repressor, *P Natl Acad Sci USA 59,* 1259-+.
26. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., and Court, D. L. (2009) Recombineering: a homologous recombination-based method of genetic engineering, *Nat Protoc 4,* 206-223.
27. Mairhofer, J., Wittwer, A., Cserjan-Puschmann, M., and Striedner, G. (2015) Preventing T7 RNA polymerase read-through transcription-A synthetic termination signal capable of improving bioprocess stability, *ACS Synth Biol 4,* 265-273.
28. Reischer, H., Schotola, L, Striedner, G., Potschacher, F., and Bayer, K. (2004) Evaluation of the GFP signal and its aptitude for novel on-line monitoring strategies of recombinant fermentation processes, *Journal of biotechnology 108,* 115-125.
29. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 221,* 680-685.
30. Waddell et al., Tn& transposition: Recognition of the attTn7 target sequence, *Proc. Natl. Acad. Sci. USA* (1989), vol. 86, pp. 3958-3962.

31. Muyrers, J. P. P. Zhang, Y. and Stewart A. F. Introducing Red®/ET® Recombination: DNA Engineering for the 21 Century. Gene Cloning & Expression Technologies; 2002, edited by Michael P. Weiner & Quinn Lu, *Biotechniques PRESS '02*
32. Sternberg N and Hoess R. The molecular genetics of bacteriophage P1. *Annu Rev. Genet.* 1983; 17 123-54.
33. Gentz and Bujard, Promoters Recognized by Escherichia coli RNA Polymerase Selected by Functions: Highly Efficient Promoters from Bacteriophage T5, *Journal of Bacteriology,* 1985, 164(1):70-77.
34. Penumetcha et al., Improving the Lac system for synthetic biology, *BIOS: A Quarterly Journal of Biology,* 2010, 81(1):7-15.
35. Oehler et al., The three operators of the lac operon cooperate in repression, *EMBO Journal,* 1990, 9(4):973-979.
36. Muller et al., Repression of lac Promoter as a Function of Distance, Phase and Quality of an Auxiliary lac Operator, *J. Mol. Biol.,* 1996, 257:21-29.
37. Dubendorff and Studier, Controlling Basal Expression in an Inducible T7 Expression System by Blocking the Target T7 Promoter with lac Repressor, *J. Mol. Biol.,* 1991 , 2019:45-59.
38. Hawley and McClure, Compilation and analysis of Escherichia coli promoter DNA sequences, *Nucleic Acids Research,* 1983, 11(8):2237-2255.

CLAIMS

1. A genome-based expression system for production of a protein of interest (POI) in a prokaryotic host, comprising at least

a) an RNA polymerase (RNAP) gene,

b) a gene encoding a POI, comprising

- a coding sequence,

- a promoter operably linked to said coding sequence, wherein said promoter is recognized by the RNAP expressed from a), and

- at least one lac operator (lacO) within the sequence of said promoter; and

c) a *lacI* gene encoding a lac repressor protein (Lad) comprising

- a coding sequence,

- a lad promoter operably linked to the lad coding sequence, wherein the lad promoter is a wild-type lad promoter or a lad promoter which increases Lad expression;

wherein the expression rate of the POI is regulated by an inducer binding LacI.

2. The genome-based expression system of claim 1, wherein the gene encoding a POI contains (i) one lacO within the sequence of the promoter or (ii) one lacO within the sequence of the promoter and one lacO upstream of the first lacO.

3. The genome-based expression system of claim 1 or 2, wherein the gene encoding a POI contains one lacO within the sequence of the promoter, and the lacI promoter is a promoter which increases LacI expression.

4. The genome-based expression system of any one of claims 1 to 3, wherein the gene encoding a POI contains one lacO within the sequence of the promoter and one lacO upstream of the first lacO, and the lacI promoter is a promoter which increases LacI expression.

5. The genome-based expression system of any one of claims 1 to 4, wherein the prokaryotic host is *Escherichia coli* (*E.coli*), preferably the host is *E.coli* of the strain BL21 or K-12.

6. The genome-based expression system of any one of claims 1 to 5, wherein the RNAP is a heterologous or homologous RNAP, preferably the RNAP is an RNAP homologous to the host, specifically it is an *E.coli* RNA polymerase, preferably the σ^{70} E.coli RNA polymerase.

7. The genome-based expression system of any one of claims 1 to 6, wherein the promoter in b) of claim 1 is selected from the group consisting of T5, T5N25, T7A1, T7A2, T7A3, lac, lacUV5, tac and trc.

8. The genome-based expression system of any one of claims 1 to 7, wherein the lad promoter which increases Lad expression is the lacI^Q promoter comprising SEQ ID NO:1.

9. The genome-based expression system of any one of claims 1 to 8, wherein the lac operator is a lac01 comprising SEQ ID NO:3, lac02 comprising SEQ ID NO:4 or lac03 comprising SEQ ID NO:5 or a functional variant thereof with at least 65% sequence identity or a perfectly symmetric lacO.

10. The genome-based expression system of any one of claims 1 to 9, wherein said promoter operably linked to the coding sequence encoding the protein of interest comprises an initial transcribed sequence (ITS), preferably a native T7A1 initial transcribed sequence comprising SEQ ID NO:2.

11. The genome-based expression system of any one of claims 1 to 10, wherein the inducer is selected from the group consisting of isopropylthiogalactoside (IPTG), lactose, methyl^-D-thiogalactoside, phenyl^-D-galactose and ortho-nitrophenyl^-galactoside (ONPG).

12. The genome-based expression system of any one of claims 1 to 11, wherein the gene encoding the POI contains one lacOI operator within the sequence of the promoter operably linked to the coding sequence and the native T7A1 initial transcribed sequence comprising SEQ ID NO:2, and wherein the lad promoter is a lacI^Q promoter.

13. The genome-based expression system of any one of claims 1 to 11, wherein the gene encoding the POI contains two lac operators which are at least 92 or 94 base pairs (bps) apart, preferably 103, 105, 114, 116, 125, 127, 134, 136, 138 or 149 bps apart, wherein one lac operator is located within the sequence of the promoter operably linked to the coding sequence and the second lac operator is upstream of the promoter.

14. A method of plasmid-free manufacturing of a protein of interest in a prokaryotic host, using the genome-based expression system of any one of claims 1 to 13, comprising the steps of

a) cultivating the host cells and inducing expression of the gene encoding the POI by addition of an inducer,

b) harvesting the POI,

c) isolating and purifying the POI, and optionally

d) modifying, and

e) formulating the POI.

15. An expression cassette comprising at least one heterologous gene configured to produce at least one heterologous POI, including

a) one or more coding sequences encoding the one or more POI,

b) a promoter operably linked to the one or more coding sequences, and

c) at least one lac operator (lacO) within the sequence of said promoter;

wherein the affinity of lad to lacO of c) is lower than the affinity of lad to the lac operators Iac01 and Iac03 of the endogenous lac operon of a host cell.

16. The expression cassette of claim 15, wherein the heterologous gene configured to produce at least one heterologous POI includes two lac operators, which are at least 92 or 94 bp apart, wherein one lac operator is located within the sequence of the promoter and the second lac operator is upstream of the promoter.

17. The expression cassette of claim 15 or 16, further comprising a heterologous lad promoter, which is the $lacI^Q$ promoter comprising SEQ ID NO:1 and wherein the heterologous gene configured to produce at least one heterologous POI comprises a Iac01 operator within the sequence of the promoter operably linked to the coding sequence and a native T7A1 initial transcribed sequence comprising SEQ ID NO:2.

18. A method of manufacturing of a POI in a prokaryotic host on a manufacturing scale, using the expression cassette of any one of claims 15 to 17, comprising the steps of

a) integrating the expression cassette into the chromosome of the prokaryotic host,

b) cultivating the host cells and inducing expression of the gene encoding the POI by addition of an inducer,

c) harvesting the POI, and

d) isolating and purifying the POI, and optionally

e) modifying and

f) formulating the POI.

FIGURES

Fig. 1

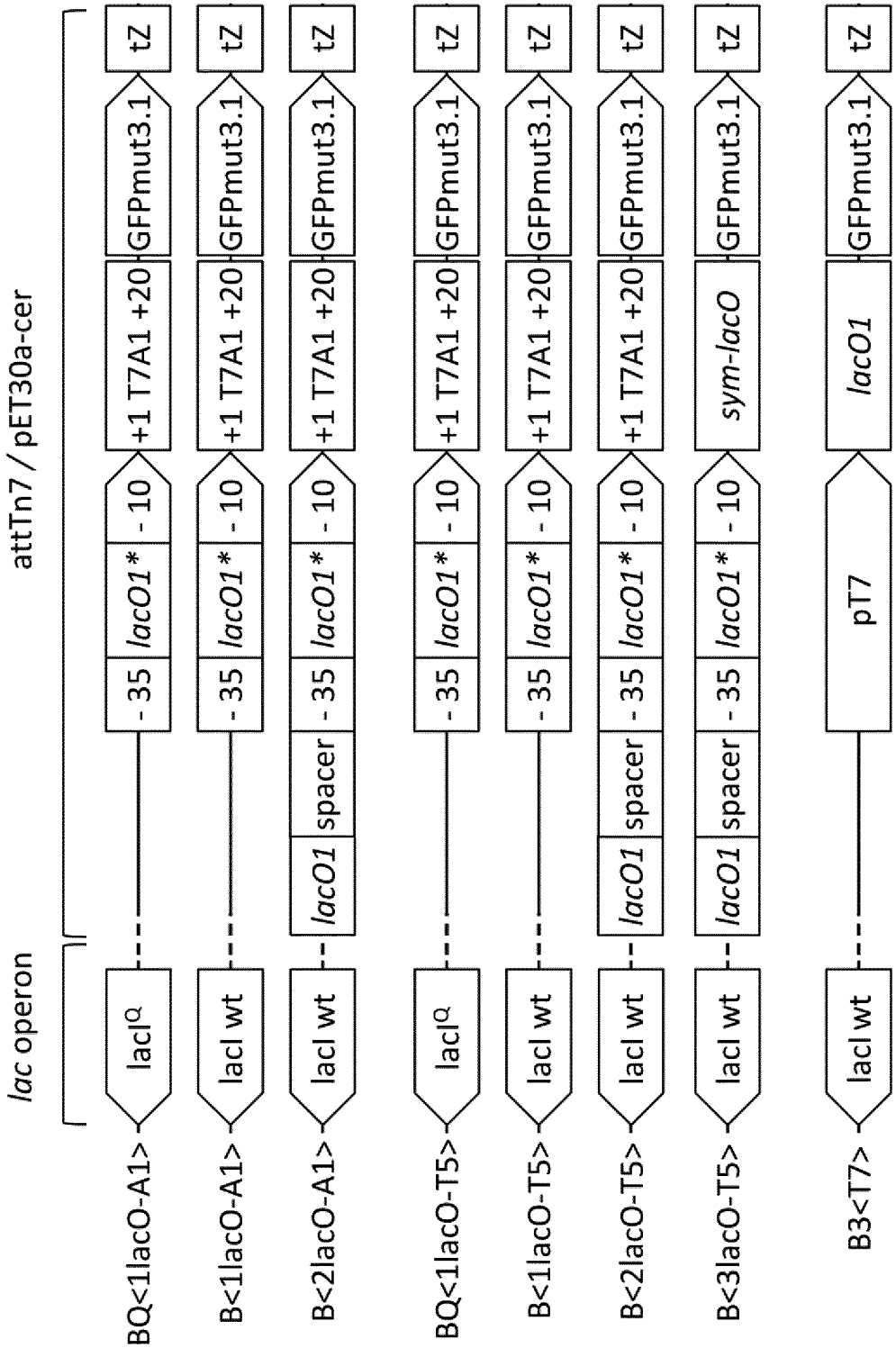


Fig. 2

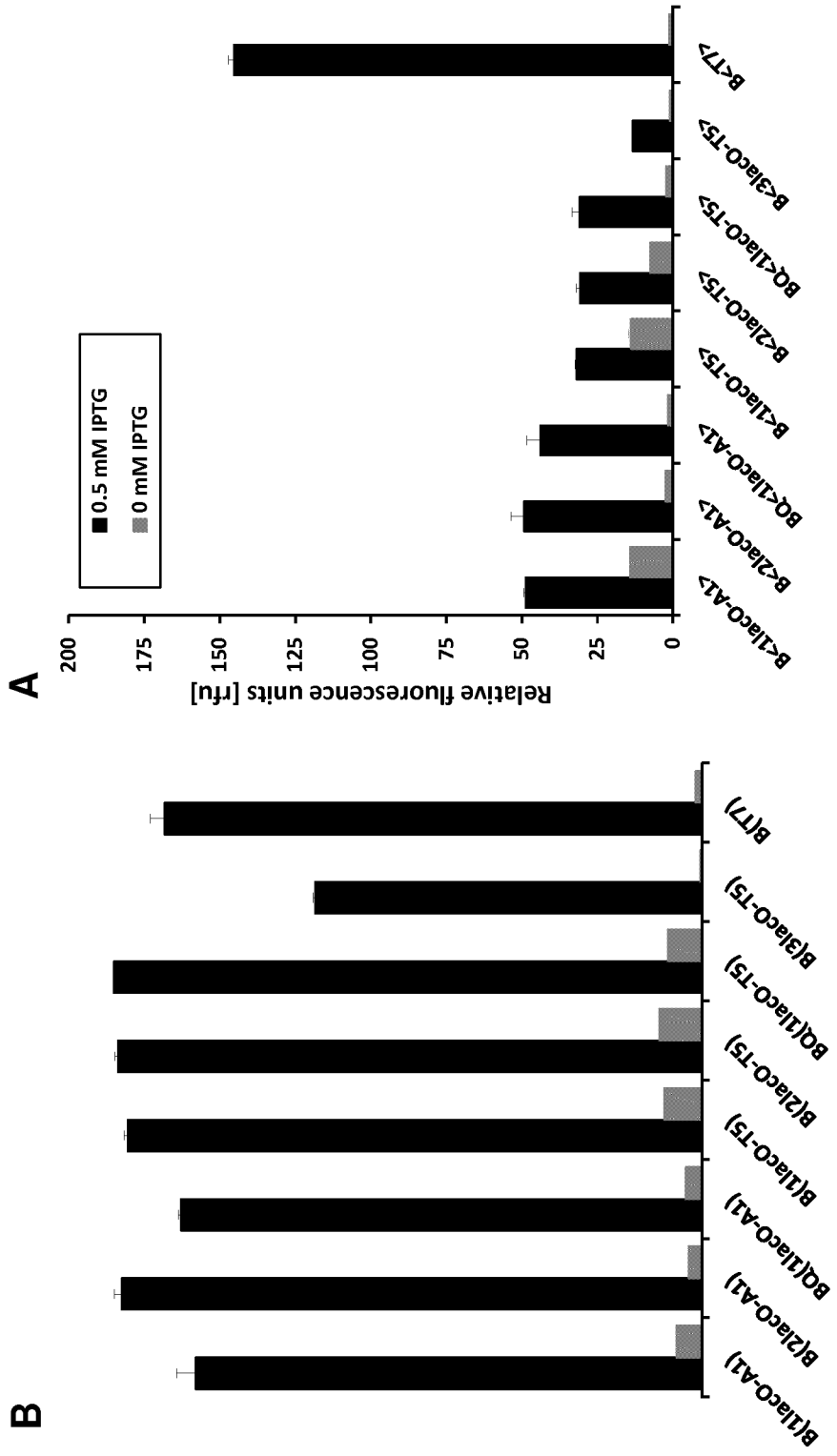

A
Relative fluorescence units [rfu]
0
25
50
75
100
125
150
175
200
0.5 mM IPTG
0 mM IPTG
B<1lacO-A1>
B<2lacO-A1>
BO<1lacO-A1>
B<1lacO-T5>
B<2lacO-T5>
BO<1lacO-T5>
B<3lacO-T5>
B<T7>
B
B(1lacO-A1)
B(2lacO-A1)
BO(1lacO-A1)
B(1lacO-T5)
B(2lacO-T5)
BO(1lacO-T5)
B(3lacO-T5)
B(T7)

Fig. 3

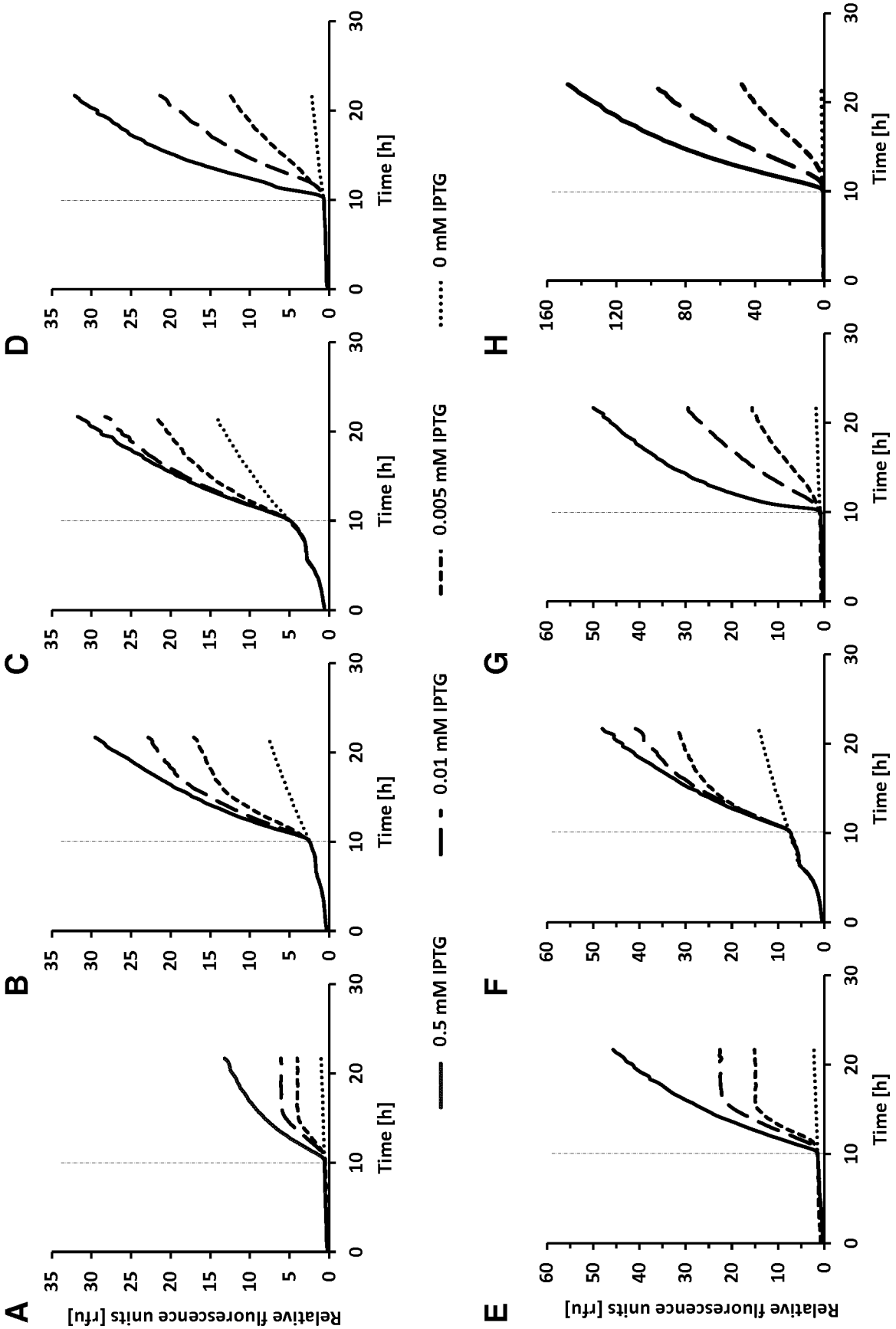

A
B
C
D
E
F
G
H
Relative fluorescence units [rfu]
Time [h]
0.5 mM IPTG
0.01 mM IPTG
0.005 mM IPTG
0 mM IPTG

Fig. 4

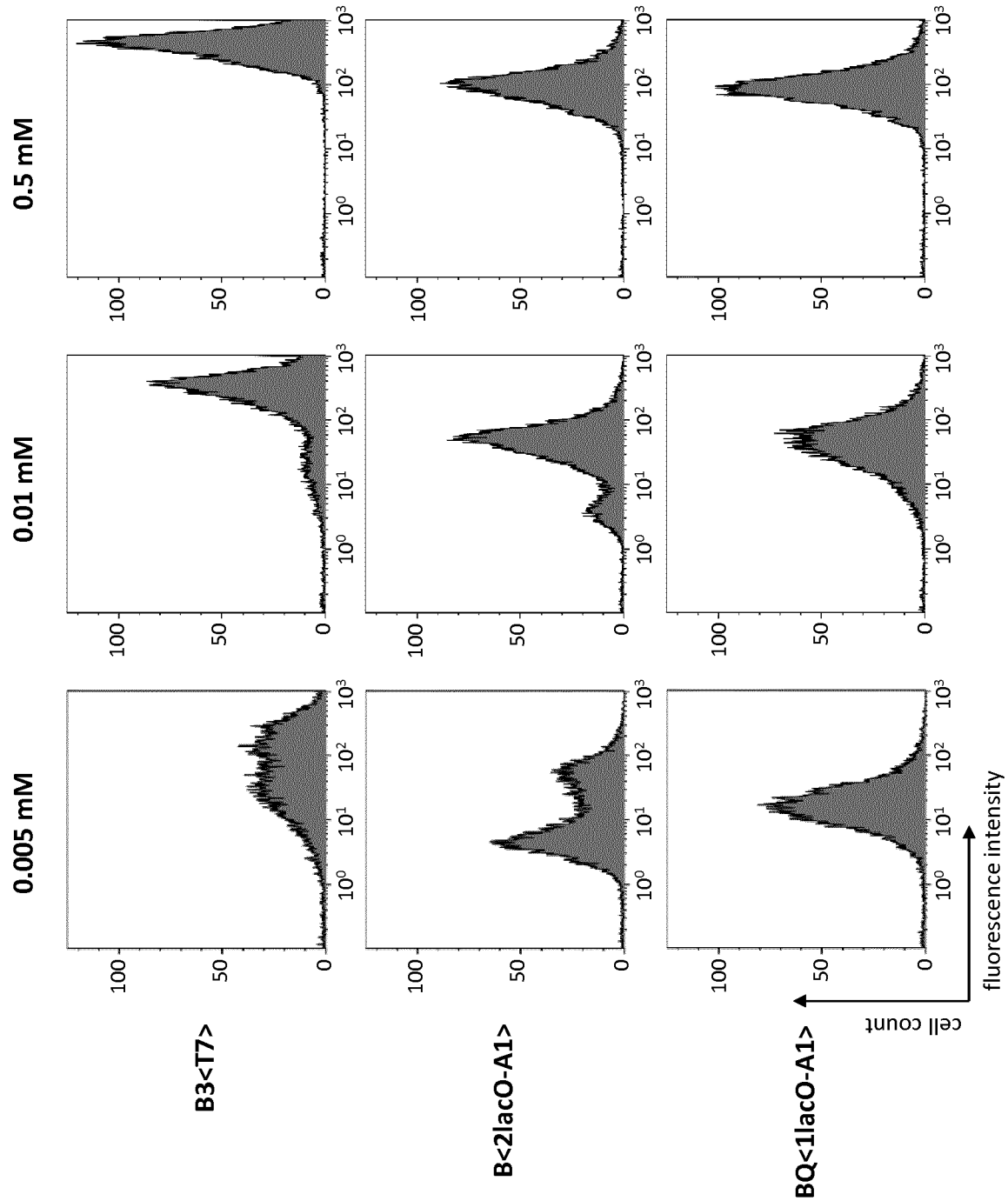

0.5 mM
0.01 mM
0.005 mM
B3<T7>
B<2lacO-A1>
BQ<1lacO-A1>
cell count
fluorescence intensity

Fig. 5

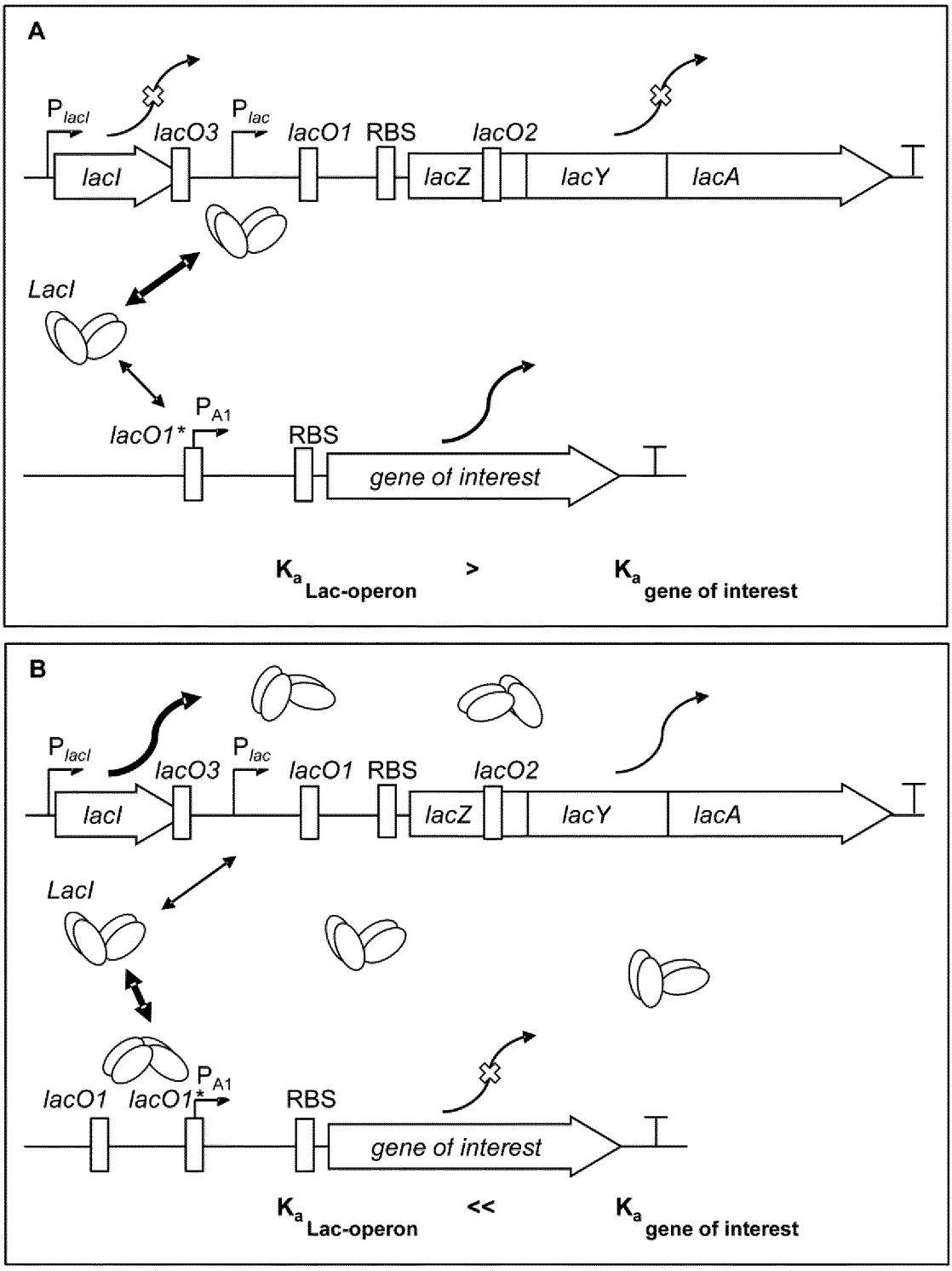

A
P lacI
lacO3
P lac
lacO1
RBS
lacO2
lacI
lacZ
lacY
lacA
LacI
P A1
lacO1*
RBS
gene of interest
K a Lac-operon
>
K a gene of interest
B
P lacI
lacO3
P lac
lacO1
RBS
lacO2
lacI
lacZ
lacY
lacA
LacI
lacO1
lacO1*
P A1
RBS
gene of interest
K a Lac-operon
<<
K a gene of interest

Fig. 6

SEQ ID NO:1 *(lacI^{Q} promoter)*
CGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCaCCATTCGATGGTGTC
CG

SEQ ID NO:2 *(ITS of T7A1)*
ATCGAGAGGGACACGGCGAA

SEQ ID NO:3 (lacO1)
AATTGTGAGCGGATAACAATT

SEQ ID NO:4 (lacO2)
AAaTGTGAGCGagTAACAAcc

SEQ ID NO:5 (lacO3)
ggcaGTGAGCGcAacgCAATT

SEQ ID NO:6 (truncated lacO1)
TTGTGAGCGGATAACAATT

Fig. 6 continued

SEQ ID NO:26 (*lacI* gene)

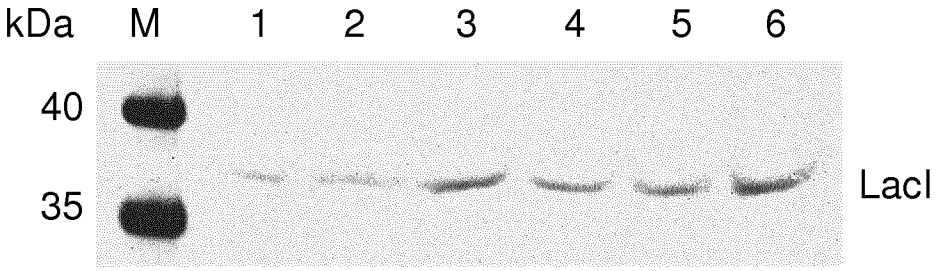
ATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGG
CAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCC
GTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGT
GGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGC
ACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGA
CCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTT
GATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTAC
GCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTT
AGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAA
ATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGT
GCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCA
CTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTA
CCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGATA
CCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCG
CCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGC
GGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCT
GGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCA
GCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGA

SEQ ID NO:27 (LacI amino acid sequence)

MAELNYIPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVV
VSMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVPALFLDVSDQ
TPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPI
AEREGDWSAMSGFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADIS
VVGYDDTEDSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRK
TTLAPNTQTASPRALADSLMQLARQVSRLESGQ

Fig. 7

A


kDa
M
1
2
3
4
5
6
40
35
LacI

B

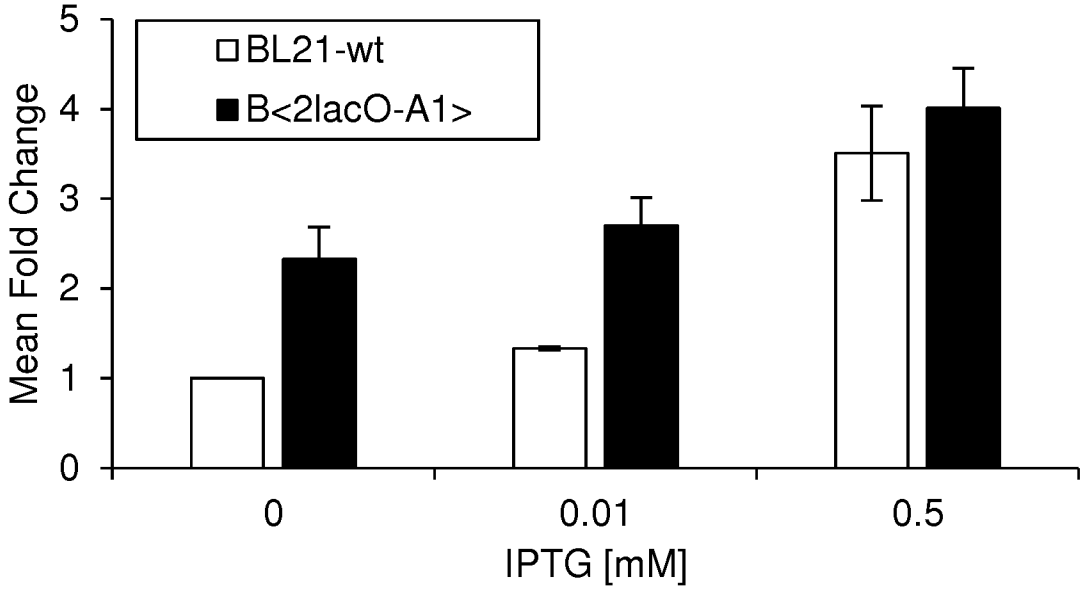

BL21-wt
B<2lacO-A1>
Mean Fold Change
0
1
2
3
4
5
0
0.01
0.5
IPTG [mM]

Fig. 8

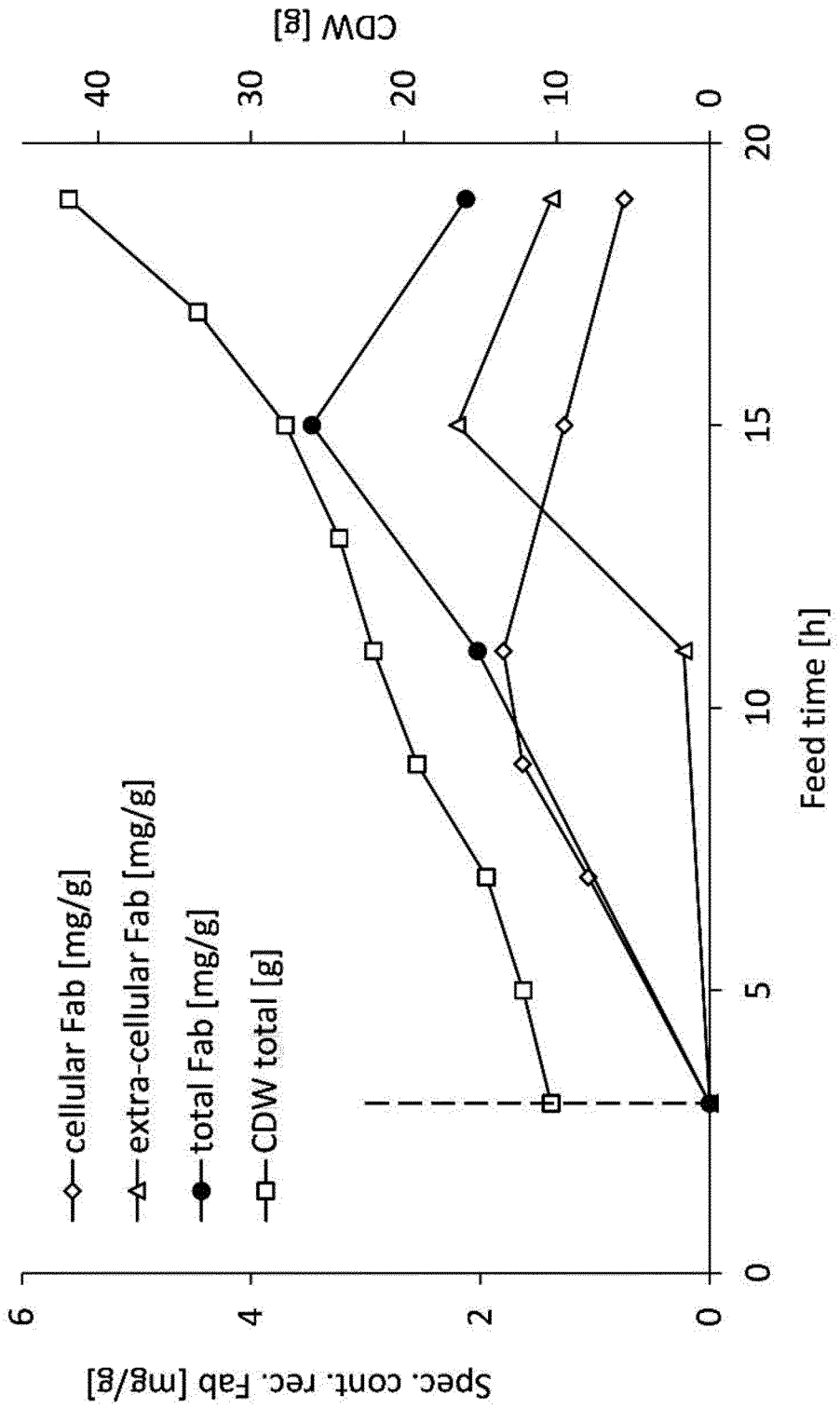
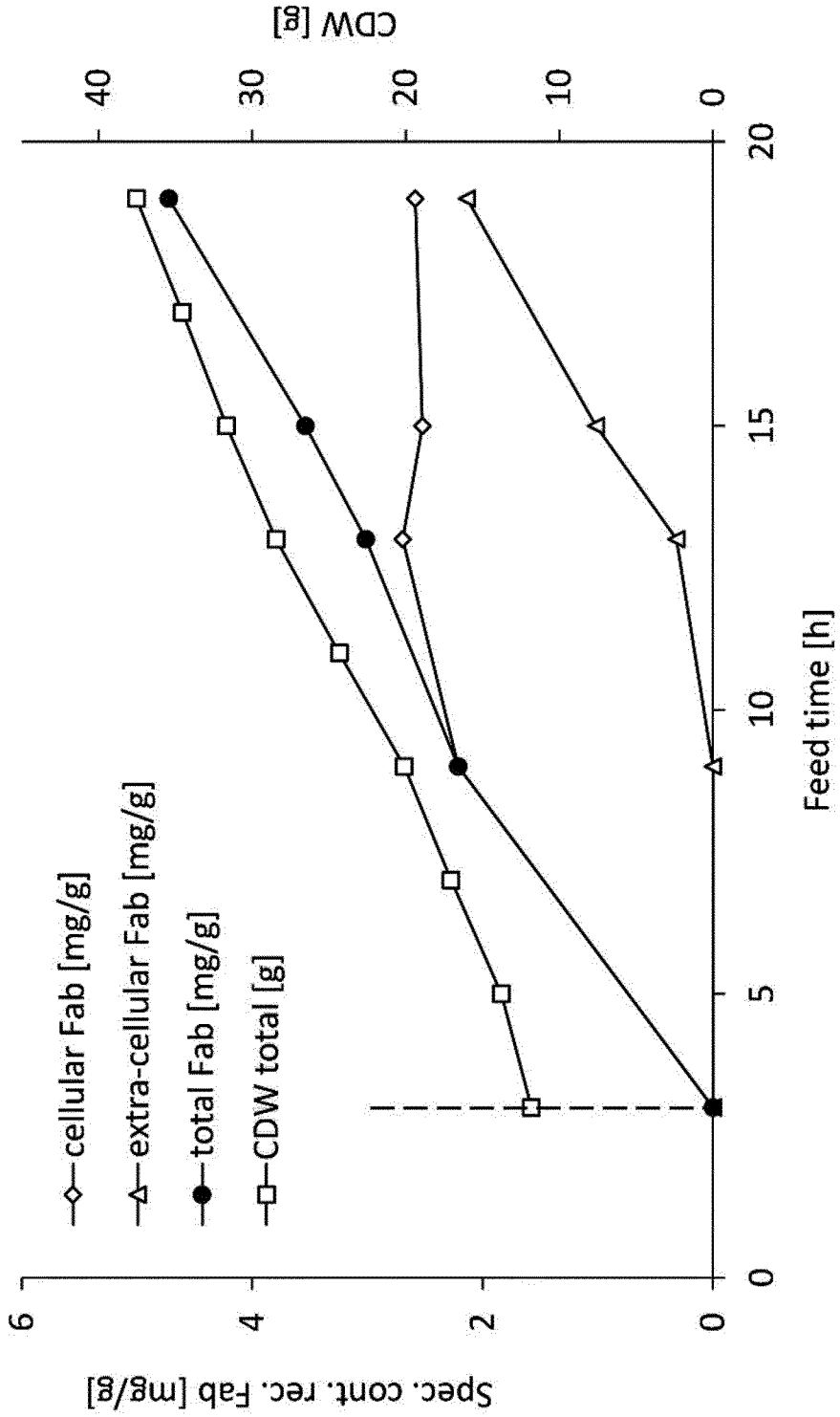

Spec. cont. rec. Fab [mg/g]
0
2
4
6
CDW [g]
0
10
20
30
40
0
5
10
15
20
Feed time [h]
cellular Fab [mg/g]
extra-cellular Fab [mg/g]
total Fab [mg/g]
CDW total [g]

Fig. 9


Spec. cont. rec. Fab [mg/g]
0
2
4
6
Feed time [h]
0
5
10
15
20
CDW [g]
0
10
20
30
40
cellular Fab [mg/g]
extra-cellular Fab [mg/g]
total Fab [mg/g]
CDW total [g]

INTERNATIONAL SEARCH REPORT

International application No
PCT/ EP20 19/074239

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N 15/63 C12N 15/70
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2008/142028 A1 (BOEHRINGER INGELHEIM RCV GMBH [AT]; SANDOZ AG [CH]; STRIEDNER GERALD [] 27 November 2008 (2008-11-27) the whole document -----	1-18
A	GERALD STRIEDNER ET AL: "Plasmid-free T7-based Escherichia coli expression systems", BIOTECHNOLOGY AND BIOENGINEERING, vol. 105, no. 4, 4 November 2009 (2009-11-04), pages 786-794, XP055180766, ISSN: 0006-3592, DOI: 10.1002/bit.22598 page 791, right-hand column -----	1-18
	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) orwhich is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 September 2019	04/10/2019

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Weinberg, Suzanna

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PALLAVI PENUMETCHA ET AL: "Improving the Lac System for Synthetic Biology", BIOS: A QUARTERLY JOURNAL OF BIOLOGY, vol. 81, no. 1, 1 March 2010 (2010-03-01), pages 7-15, XP055548761, United States ISSN: 0005-3155, DOI: 10.1893/011.081.0104 page 8, right-hand column	1-18
A	LILIAN M. HSU ET AL: "Initial Transcribed Sequence Mutations Specifically Affect Promoter Escape Properties +", BIOCHEMISTRY, vol. 45, no. 29, 1 July 2006 (2006-07-01), pages 8841-8854, XP055549934, US ISSN: 0006-2960, DOI: 10.1021/bi060247u cited in the application abstract	1-18
A	S. OEHLER ET AL: "The three operators of the lac operon cooperate in repression.", EMBO (EUROPEAN MOLECULAR BIOLOGY ORGANIZATION) JOURNAL, vol. 9, no. 4, 1 April 1990 (1990-04-01), pages 973-979, XP055551859, DE ISSN: 0261-4189, DOI: 10.1002/j.1460-2075.1990.tb08199.x the whole document	1-18
A	JOHANNES MÜLLER ET AL: "Repression of lac Promoter as a Function of Distance, Phase and Quality of an Auxiliary lac Operator", J. MOL. BIOL, vol. 257, 1 January 1996 (1996-01-01), pages 21-29, XP055549889, DOI: https://doi.org/10.1006/jmbi.1996.0143 the whole document	1-18
A	DUBENDORF J W ET AL: "Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor", JOURNAL OF MOLECULAR BIOLOGY, ACADEMIC PRESS, UNITED KINGDOM, vol. 219, no. 1, 5 May 1991 (1991-05-05), pages 45-59, XP024021410, ISSN: 0022-2836, DOI: 10.1016/0022-2836(91)90856-2 [retrieved on 1991-05-05] the whole document	1-18

-/--

2

International application No
PCT/ EP20 19/074239

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 03/050240 A2 (LILLY CO ELI [US]; FRYE CHRISTOPHER CARL [US]; HERSHBERGER CHARLES LEE) 19 June 2003 (2003-06-19) page 2 - page 3	1-18
A	KEITH E J TYO ET AL: "Stabilized gene duplication enables long-term selection-free heterologous pathway expression", NATURE BIOTECHNOLOGY, vol. 27, no. 8, 26 July 2009 (2009-07-26), pages 760-765, XP055551987, ISSN: 1087-0156, DOI: 10.1038/nbt.1555 the whole document	1-18

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2019/074239

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008142028 A1	27-11-2008	AU 2008252990 A1	27-11-2008
		CA 2687170 A1	27-11-2008
		EP 2160469 A1	10-03-2010
		EP 3399042 A1	07-11-2018
		JP 5837747 B2	24-12-2015
		JP 5996052 B2	21-09-2016
		JP 2010527239 A	12-08-2010
		JP 2015221045 A	10-12-2015
		KR 20100017315 A	16-02-2010
		US 2011045531 A1	24-02-2011
		US 2017253900 A1	07-09-2017
		WO 2008142028 A1	27-11-2008
WO 03050240 A2	19-06-2003	AU 2002348205 A1	23-06-2003
		BR 0214542 A	16-08-2005
		CA 2467505 A1	19-06-2003
		CN 1604960 A	06-04-2005
		EP 1456349 A2	15-09-2004
		JP 2006502691 A	26-01-2006
		MX PA04005717 A	03-06-2005
		US 2005009029 A1	13-01-2005
		WO 03050240 A2	19-06-2003