Adaptive evolution in producing microtiter cultivations generates genetically stable *Escherichia coli* production hosts for continuous bioprocessing

Artur Schuller[1], Monika Cserjan-Puschmann[1], Christoph Köppl[1], Reingard Grabherr[1], Martin Wagenknecht[2], Matteo Schiavinato[3], Juliane Dohm[3], Heinz Himmelbauer[3], and Gerald Striedner[1]

[1]Christian Doppler Laboratory for production of next-level biopharmaceuticals in E. coli, University of Natural Resources and Life Sciences
[2]Boehringer Ingelheim RCV GmbH und Co KG
[3]Institute of Computational Biology, University of Natural Resources and Life Sciences

August 17, 2020

Abstract

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. We investigated 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. We conducted iterated fed-batch-like microbioreactor cultivations under production conditions. We used the easy-to-produce green fluorescent protein (GFP) and a challenging antigen-binding fragment (Fab) as model proteins, and BL21(DE3) and BL21[Q] strains as expression hosts. In comparative whole genome sequencing analyses, we identified mutations that allowed cells to grow unhindered despite recombinant protein production. 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 remained stable in the production of GFP in long-term cultivations. For the production of Fab, mutations in *lacI* of the BL21[Q] derivatives had positive effects on long-term stability. Our 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 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 such chemostat cultivations, cells are maintained in a steady-state growth environment by adding fresh medium to the reactor at 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 always the same, which can be advantageous in terms of product quality.

1

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 analogue isopropyl-D-1-thiogalactopyranoside (IPTG).[9]

Genetic heterogeneity caused by metabolic 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 metabolism, 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 (IS), and by deleting *recA* or error-prone DNA polymerase genes.[16] Another promising strategy uses the directed 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 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 full functionality of the host RNAP is required for cell growth.[18]

In combination with a directed evolution approach, this would allow characterization of mutated production hosts, which have adapted themselves to the metabolic load triggered by recombinant protein production and could potentially enable continuous protein production. Moreover, a directed 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 directed evolution to produce cell membrane proteins that are toxic to host cells. These strains are currently widely used for 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 metabolic load 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 GFP-mut3.1 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 lab-scale bioreactors with the above-mentioned clones and additional robust production strains obtained by the directed evolution approach.

2 Materials and methods

2.1 Construction of expression systems

2.1.1 Strains

E. coli K-12 NEB5-α [φηυΑ2Δ(αργΦ-λαςΖ)Υ169 πηοΑ γλν ˘44 Φ80 Δ(λαςΖ)Μ15 γψρΑ96 ρεςΑ1 ρελΑ1 ενδΑ1 τηι-1 ησδΡ17] 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) [φηυΑ2 [λον] ομπΤ γαλ (λ ΔΕ3) [δςμ] [;]ησδΣ λ ΔΕ3 = λ σΒαμΗΙο [;]ΕςοΡΙ-Β ιντ::(λαςΙ::ΠλαςΥ˘5::Τ7 γενε1) ι21 [;]νιν5] (NEB). Ηερεαφτερ, τηις στραιν ις ρεφερρεδ το ας B3<T7-γενε >. Λινεαρ ΔΝΑ ςαρτριδγες ςοντρολλεδ βψ τηε Α1 προμοτερ ωερε ιντεγρατεδ ιντο τηε βαςτεριαλ ςηρομοσομε ατ τηε αττΤν7 σιτε οφ *E. ςολι* ΒΛ21 [φηυΑ2 [λον] ομπΤ γαλ [δςμ] ΔhsdS] (NEB) containing the *lacI^Q* promoter as described by Glascock and Weickert.[25] Briefly, the pETAmp-lacIq plasmid was constructed for 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 glucose and 16.5 g/L dextran as carbon sources (m2p-labs GmbH, Baesweiler, Germany) was used with the following additions (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 $MgSO_4·7H_2O$, and 0.0065 Thiamin·HCl.

The trace element solution contained (mg/L): 0.36 $ZnSO_4·7H_2O$, 0.33 $CuSO_4·5H_2O$, 0.20 $MnSO_4·H_2O$, 27.30 $FeCl_3·6H_2O$, 21.84 Titriplex III, 0.36 $CoCl_2·6H_2O$, and 1.31 $CaCl_2·2H_2O$. 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 emission of 520 nm. The signal is given in relative fluorescence units (rfu). The cell dry matter (CDM, given in g/L) 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 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 re-suspended 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 mM.

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 $\pm$ 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 pre-culture shake flask containing 25 mL M9ZB [30] for cultivation for at least 8 hours at 37°C. Subsequently, a volume equivalent to 25 OD$_{600}$ units (25/OD$_{600}$ = volume in mL) was transferred aseptically to the bioreactors.

Feeding was initiated when the culture, grown to 10 g/L 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 μ = 0.1/h over 19 hours or 2.74 doublings. The substrate feed was controlled by increasing the pump speed according to the exponential growth algorithm, $x = x_0 e^{\mu\tau}$, ωιτη συπεριμποσεδ φεεδβαςχ ςοντρολ οφ ωειγητ λοσς ιν τηε συβστρατε βοττλε. Τηε ˚ΔΜ ψιελδ ςοεφφιςιεντ ον γλυςοσε ωας 0.33 γ/γ. Τηερεφορε, τηε φεεδ μεδιυμ προιδεδ 66 γ/Λ γλυςοσε ανδ συφφιςιεντ ςομπονεντς το ψιελδ α φιναλ ˚ΔΜ ςονςεντρατιον οφ 30 γ/Λ ιν 1.2 Λ. Τηε εξπρεσσιον σψστεμ ωας ινδυςεδ βψ αδδινγ ΙΠΤΓ το τηε ρεαςτορ το τηε ψιελδ α ςονςεντρατιον οφ 10 μmol/g 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 CDM and a dilution rate of D = 0.1/h 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 medium designed to produce another 15.4 g of CDM, which corresponds to a final CDM concentration of 30 g/L. In the next step, the bioreactor was shifted to chemostat mode with a dilution rate of 67 mL/h. The medium was used as continuous feed to the reactor to provide nutrients to maintain a CDM concentration of 30 g/L. The temperature was decreased to 30°C and recombinant protein production induced with 10 μmol IPTG/g 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 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 Off-line measurement of GFP fluorescence

Off-line fluorescence measurements were made using a Tecan analyzer infinite 200Pro. For GFPmut3.1, the excitation wavelength was 485 nm and the emission wavelength 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 hours and samples from chemostat cultivations every 24 hours. The sample volume for intracellular protein quantification was calculated as 3.5/OD$_{600}$ = volume in mL, 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 ELISA 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 were 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 hour 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, 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 (ThermoFisher, Waltham, MA, USA). Quantification was performed using a Qubit 3.0 Fluorometer and a dsDNA BR assay kit (ThermoFisher, Waltham, MA, USA). 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 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 8 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). We generated four reference sequences that represented our 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, we used breseq [35] (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 a coverage of 150× using 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 Directed evolution and isolation of protein production strains

We applied repeated fed-batch-like cultivations in a 48-well microbioreactor system in order to characterize mutation patterns triggered by recombinant protein production under long-term 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 CDM at the beginning of each passage, which corresponds to full induction of recombinant protein production. Fab-producing cells were induced after 6 hours of the first passage and cultivated in medium that already contained 10 μmol IPTG/g 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 over a period of six passages and a total of 42 generations. Fab-producing cells were cultivated over a period of 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). The cells were able to reach the calculated end biomass of 6 g/L CDM in each of the six passages. Similarly, the productivity remained unchanged and was always approximately 25 rfu/g CDM at the end of each passage (Figure S2). Flow cytometry revealed continuous homogeneous populations of all 24 measured cultivations, as depicted for clones #E1 and #E2 in Figure 1B.

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 a final biomass of only 1.5 g/L CDM. However, all 24 T7 strains were able to reach the calculated end biomass of 6 g/L 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 over the period of 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-cell colonies from this mixture, we streaked the cells on agar plates after passage 6 and analyzed the isolated colonies by flow cytometry. We isolated 10 homogeneous subpopulations from weak producing (subclone 1, S1) to strong producing (subclone 10, S10) cells (Figure 1C).
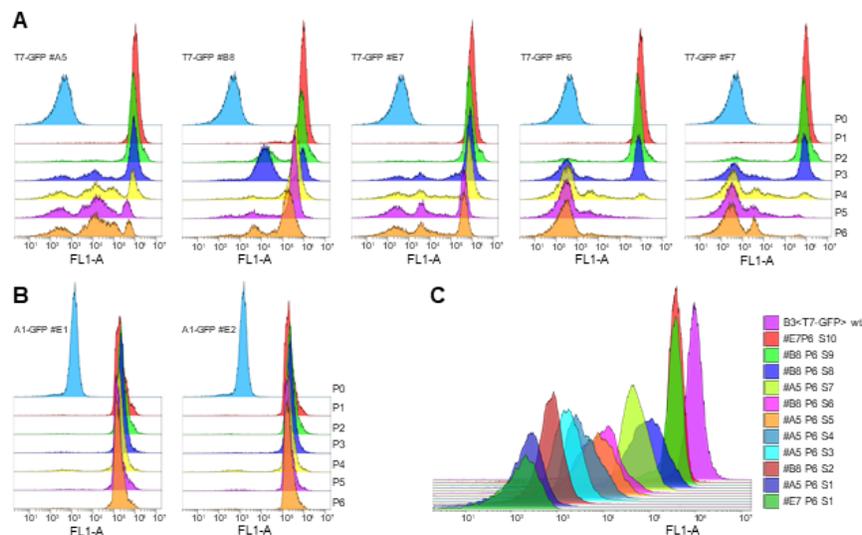
Figure 1. Analysis of single cell expression using GFP as a model protein. (A-B) Flow cytometry over the course of six passages. (A) Derivatives of B3<T7-GFP>. (B) Derivatives of BQ<A1-GFP>. (C) Derivatives of B3<T7-GFP> separated and categorized according to the strength of their expression. S1: weak producing to S10: strong producing subclones.

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 exhibited reduced growth rates in passage 2 but were able to recover and exhibited unrestricted growth in passage 3 (Figure 2A). 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 and 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; 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.

7
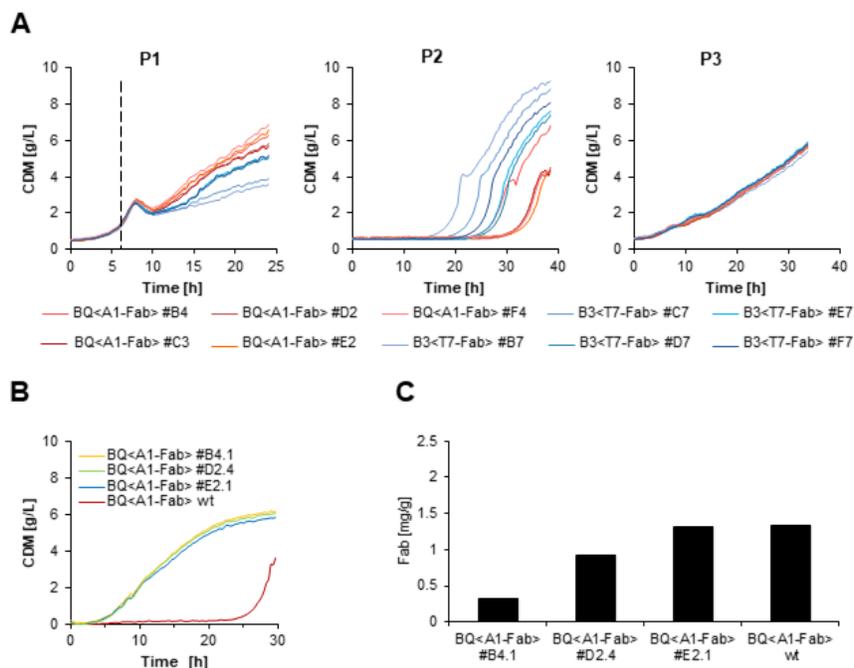
Figure 2. Growth characteristics and product formation of Fab-producing strains. (A) Growth behavior of BQ<A1-Fab> (red lines) and B3<T7-Fab> (blue lines) over the course of three passages. The dashed vertical line indicates the time point of induction. (B) Cells were fractionated on agar plates and 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 CDM].

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 approximately 1×10^{-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. We applied read mapping and a *de novo* assembly approach to examine SNPs and short indels, as well as 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 directed 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 exactly 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 has no influence on the basal expression or tunability of the A1 expression system (Figure S3).

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

8

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.

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.[42, 43] 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 analogues. Cells overexpressing SetA exhibit decreased accumulation of lactose and IPTG.[44] We speculate that this mutation increases the SetA expression rate, preventing intracellular accumulation of IPTG and reducing the induction of the GOI.

Derivative #A5P6 S5 has a T-C mutation within *ptsI* , leading to a 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 [45]. 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 induction of the GOI.

In addition to a mutation within the T7 RNAP gene, derivative #A5P6 S7 also exhibited a mutation within *malP* . The $(T)_{5?4}$ mutation causes a frameshift and probably complete loss of function. MalP is involved in maltose metabolism and has high affinity for short, linear α-1,4 linked oligoglucosides.[46] 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.[47]

A mutation in the type III CoA transferase gene, *caiB*[48], 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 induction of the GOI.[49] The metabolic 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 clearly attributed in the case of *ptsI* , as this mutation does not occur in combination with a mutation in the T7 promoter or in the T7 RNAP gene. Unfortunately, this is not the case with mutations in *setA* and *malP* . However, the mutations in *srlR* and *gntT* clearly have a positive influence on the growth rate under conditions of recombinant protein production.

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	Δ 13 bp	Intergenic	-73 from *gfpmut3.1*	T7 promoter
	A5P6 S1	743064	Δ 10,872 bp	Coding	-	Part of λ DE3
	B8P6 S2	43017	GGT - GTT (G98V)	Coding	*caiB*	Type III coA transfera

9

Strain	Clone	Position	Mutation	Region	Gene	Function
		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
	A5P6 S4	753056	TTC - CTC (F882L)	Coding	T7 RNAP gene	Enterobacteria phage
	A5P6 S5	2407211	CTG - CCG (L256P)	Coding	*ptsI*	Bacterial phosphotran
	B8P6 S6	43017	GGT - GTT (G98V)	Coding	*caiB*	Type III coA transfera
		3800859	+C	Intergenic	-81 from *gfpmut3.1*	T7 promoter
	A5P6 S7	751679	CGC - TGC (R423C)	Coding	T7 RNAP gene	Enterobacteria phage
		3412013	$(T)_{5-4}$	Coding	*malP*	Maltodextrin phospho
	B8P6 S8	43017	GGT - GTT (G98V)	Coding	*caiB*	Type III coA transfera
		2669647	Δ 2 bp	Coding	*srlR*	GutR transcriptional
	B8P6 S9	43017	GGT - GTT (G98V)	Coding	*caiB*	Type III coA transfera
		335936	C - T	Intergenic	-104 from *lacZ*	Upstream region of pI
		3406745	C - T	Intergenic	-168 from *gntT*	*gntT* operator site
	E7P7 S10	335936	C - T	Intergenic	-104 from *lacZ*	Upstream region of *pI*
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

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] In addition, 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. In addition, 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.[50, 51] 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). As a result, cells have tuned themselves to the maximum tolerated productivity. This seems to be a 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.

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 R
	D7	752266	Δ 11 bp	Coding	T7 RNAP gene	Enterobacteria phage T7 R
	F7	752331	GGG-GAG (G640E)	Coding	T7 RNAP gene	Enterobacteria phage T7 R
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*

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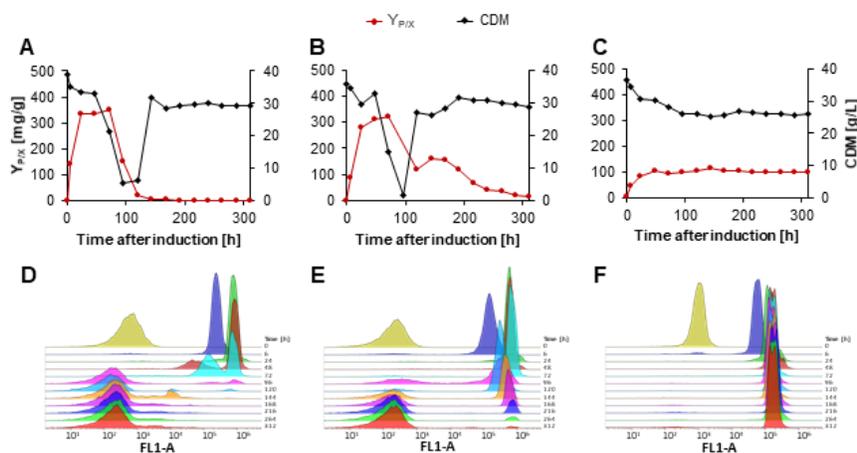
Figure 3. Process characteristics of GFP-producing strains in chemostat culture. (A-C) Courses of the specific content of GFP ($Y_{P/X}$) and biomass. (D-F) Single-cell expression analysis. (A, D) B3<T7-GFP> wt, (B, E) B3<T7-GFP> #E7P6 S10, (C, F) BQ<A1-GFP> wt.

For B3<T7-GFP> wt, induction with IPTG resulted in very high specific content of 348 mg GFP/g CDM (Figure 3A), which was clearly higher than the product titer from a fed-batch culture with the same final biomass of 30 g/L (Figure 4A, Table 3). This productivity could be maintained for 72 hours 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 approximately 120 hours, the biomass recovered and increased to the intended 30 g/L CDM, but without product formation. The specific GFP content remained at 0 mg/g 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 after 72 hours in chemostat mode. After 120 hours, the cells recovered to the intended biomass of 30 g/L CDM. Unlike its wild-type ancestor, the specific GFP

11

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 over the course of 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 an extraordinarily high stability (Figure 3C). Over the entire cultivation period of 312 hours and 45 generations, the specific product content remained constant at approximately 100 mg GFP/g 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 apparently represents such a low metabolic load 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.
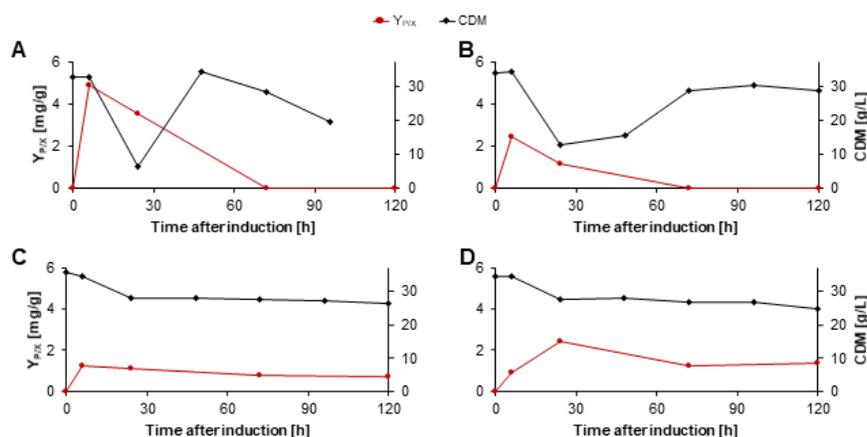
Figure 4. Process characteristics of Fab-producing strains in chemostat culture. Courses of the specific content of Fab ($Y_{P/X}$) and biomass. (A) B3<T7-Fab> wt, (B) BQ<A1-Fab> wt, (C) BQ<A1-Fab> #D2.4, (D) BQ<A1-Fab> #E2.1.

Long-term chemostat cultivation with Fab were performed with two BQ<A1-Fab> variants, #D2.4 and #E2.1 (Figure 4C and D). The wild-type non-mutated strains B3<T7-Fab> wt and BQ<A1-Fab> wt served as reference (Figure 4A and 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 hours after induction and concomitant reduction of the product. The maximum amount of product was 4.9 mg/g and 2.5 mg/g 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 and D, 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 degree of metabolic load 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 approximately 1.0 mg Fab/g CDM over a total of 120 hours, which corresponds to 17.3 generations. Derivative #E2.1 produced even more at 1.6 mg/g. Nevertheless, long-term stability was evidently accompanied by a reduction in productivity. Compared to the conventional fed-batch processing, in which BQ<A1-Fab> wt produced 2.7 mg/g (Figure 4D), the product titers of the corresponding derivatives were clearly lower, approximately half in the case of the best producing mutant #E2.1.

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

Strain	Cultivation mode	Max. volumetric yield [g/L][a]	Process time [b] [h]	STY [mg/
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), 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.

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

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

For calculation of the process time for both process types, we added a down time of 12 hours 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 hours.

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 hours. Chemostat cultivation can be extended to 48 hours 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, a short-term chemostat cultivation of 48 hours could be interesting for easy-to-produce proteins. With 874 mg/L/h, the STY of B3<T7-GFP>wt in the chemostat

13

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/h in fed-batch cultivation was relatively low compared to the STY of 286 mg/L/h of the conventional T7 expression system. In addition to the short production time of 19 hours, 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 hours, 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 mode, 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 g/L to 0.035 g/L, 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/h in fed-batch mode to 4.33 g/L/h 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 metabolic load 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. As already mentioned, the time in which the actual fermentation occurs and recombinant protein is produced in fed-batch cultivation is relatively low. In addition, the cost of CIP and SIP must be included. Although these process steps must also be carried out in a 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 collapse 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, 52, 53]

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 more severe metabolic load on cell metabolism, 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 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

14

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 obviously physiologically acceptable, level. For continuous recombinant protein production in *E. coli* , cells have to adapt to their maximum tolerable level of recombinant protein production. Therefore, the mutations found here do not generally apply to all challenging or easy-to-produce recombinant proteins, but can vary from protein to protein.

Regarding industrial regulations, the US 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 a directed 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.

Acknowledgement

Author contributions:

AS, MCP, RG, JCD, HH, and GS designed the research; AS, CK, and MS performed the research; AS, MCP, RG, MW, JCD, MS, HH, and GS analyzed the data; AS and GS wrote the paper, with contributions from MS, JCD, MCP, and MW. All authors approved the final version of the manuscript.

Conflict of interest

The authors declare no commercial or financial conflict of interest.

5 References

[1] J. Walther, R. Godawat, C. Hwang, Y. Abe, A. Sinclair, K. Konstantinov, *J. Biotechnol.* **2015** , *213* , 3-12.

[2] N. Ziv, N. J. Brandt, D. Gresham, *JoVE* **2013** , e50168.

[3] K. Peebo, P. Neubauer, *Microorganisms* **2018** ,*6* .

[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. Biotech.* **2019** , *7* .

[6] A. Schmideder, D. Weuster-Botz, *J. Ind. Microbiol. Biotechnol.* **2017** , *44* , 1021-1029.

[7] F. W. Studier, B. A. Moffatt, *J. Mol. Biol.* **1986** , *189* , 113-130.

[8] S. B. Noronha, H. J. C. Yeh, T. F. Spande, J. Shiloach, *Biotechnol. Bioeng.* **2000** , *68* , 316-327.

[9] G. L. Rosano, E. A. Ceccarelli, *Front. Microbiol.* **2014** , *5* , 172.

[10] O. Zelder, B. Hauer, *Curr. Opin. Microbiol.* **2000** , *3* , 248-251.

[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* , 869-876.

[13] B. R. Glick, *Biotechnol. Adv.* **1995** , *13* , 247-261.

[14] L. De Gelder, J. M. Ponciano, P. Joyce, E. M. Top, *Microbiology* **2007** , *153* , 452-463.

[15] G. Striedner, I. Pfaffenzeller, L. Markus, S. Nemecek, R. Grabherr, K. Bayer, *Biotechnol. Bioeng.* **2010** , *105* , 786-794.

[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* , 9236-9250.

[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* , 289-298.

[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* , 1243-1247.

[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 Expr. Purif.* **2004** , *37* , 203-206.

[23] R. P. Maharjan, T. Ferenci, *PLoS Biol.* **2017** , *15* , e2001477.

[24] R. P. Maharjan, T. Ferenci, *Environ. Microbiol. Rep.* **2018** , *10* , 626-633.

[25] C. B. Glascock, M. J. Weickert, *Gene* **1998** , *223* , 221-231.

[26] S. K. Sharan, L. C. Thomason, S. G. Kuznetsov, D. L. Court, *Nat. Protoc.* **2009** , *4* , 206-223.

[27] B. P. Cormack, R. H. Valdivia, S. Falkow, *Gene* **1996** , *173* , 33-38.

[28] M. Fink, S. Vazulka, E. Egger, J. Jarmer, R. Grabherr, M. Cserjan-Puschmann, G. Striedner, *Biotechnol. J.* **2019** , *14* , e1800637.

[29] C. Toeroek, M. Cserjan-Puschmann, K. Bayer, G. Striedner, *SpringerPlus* **2015** , *4* .

[30] F. W. Studier, *Protein Expr. Purif.* **2005** , *41* , 207-234.

[31] H. Reischer, I. Schotola, G. Striedner, F. Potschacher, K. Bayer, *J. Biotechnol.* **2004** , *108* , 115-125.

[32] F. He, *Bio-protocol* **2011** , *1* , e97.

[33] A. M. Bolger, M. Lohse, B. Usadel, *Bioinformatics* **2014** , *30* , 2114-2120.

[34] H. Jeong, H. J. Kim, S. J. Lee, *Genome Announc.* **2015** , *3* , e00134-00115.

[35] D. E. Deatherage, J. E. Barrick, *Methods Mol. Biol.* **2014**, *1151*, 165-188.

[36] D. Kim, B. Langmead, S. L. Salzberg, *Nat. Methods* **2015**, *12*, 357-360.

[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*, 1422-1423.

[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. U S A* **2012**, *109*, E2774-2783.

[41] R. Maharjan, T. Ferenci, *Mol. Biol. Evol.* **2015**, *32*, 380-391.

[42] A. Porco, N. Peekhaus, C. Bausch, S. Tong, T. Isturiz, T. Conway, *J. Bacteriol.* **1997**, *179*, 1584-1590.

[43] N. Peekhaus, T. Conway, *J. Bacteriol.* **1998**, *180*, 1777-1785.

[44] J. Y. Liu, P. F. Miller, M. Gosink, E. R. Olson, *Mol. Microbiol.* **1999**, *31*, 1845-1851.

[45] R. D. Simoni, S. Roseman, M. H. Saier, Jr., *J. Biol. Chem.* **1976**, *251*, 6584-6597.

[46] R. Schinzel, B. Nidetzky, *FEMS Microbiol Lett* **1999**, *171*, 73-79.

[47] M. Yamada, M. H. Saier, *J. Mol. Biol.* **1988**, *203*, 569-583.

[48] T. Elssner, C. Engemann, K. Baumgart, H. P. Kleber, *Biochemistry* **2001**, *40*, 11140-11148.

[49] A. Fernandez-Castane, C. E. Vine, G. Caminal, J. Lopez-Santin, *J. Biotechnol.* **2012**, *157*, 391-398.

[50] D. A. Siegele, J. C. Hu, W. A. Walter, C. A. Gross, *J. Mol. Biol.* **1989**, *206*, 591-603.

[51] T. Gardella, H. Moyle, M. M. Susskind, *J. Mol. Biol.* **1989**, *206*, 579-590.

[52] J. M. Couto, A. McGarrity, J. Russell, W. T. Sloan, *Microb. Cell Fact.* **2018**, *17*, 8.

[53] N. Rai, L. Huynh, M. Kim, I. Tagkopoulos, *Biotechnol. Bioeng.* **2019**, *116*, 693-703.