ACS Synthetic Biology

pubs.acs.org/synthbio

Research Article

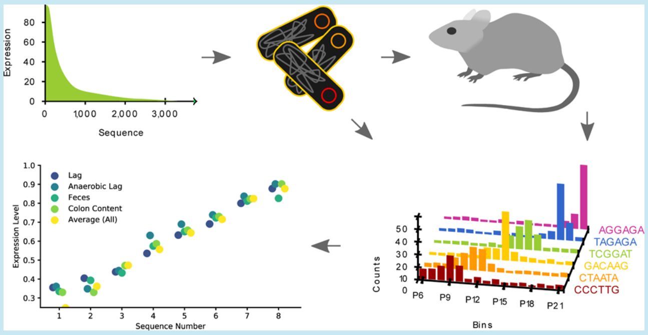
Escherichia coli Promoters with Consistent Expression throughout the Murine Gut

Jeremy Armetta,[+] Michael Schantz-Klausen,[+] Denis Shepelin, Ruben Vazquez-Uribe, Martin Iain Bahl, Martin Frederik Laursen, Tine Rask Licht, and Morten O.A. Sommer*

Cite This: *ACS Synth. Biol.* 2021, 10, 3359−3368

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Advanced microbial therapeutics have great potential as a novel modality to diagnose and treat a wide range of diseases. Yet, to realize this potential, robust parts for regulating gene expression and consequent therapeutic activity in situ are needed. In this study, we characterized the expression level of more than 8000 variants of the *Escherichia coli* sigma factor 70 ($\sigma70$) promoter in a range of different environmental conditions and growth states using fluorescence-activated cell sorting and deep sequencing. Sampled conditions include aerobic and anaerobic culture in the laboratory as well as growth in several locations of the murine gastrointestinal tract. We found that $\sigma70$ promoters in *E. coli* generally maintain consistent expression levels across the murine gut (R^2: 0.55−0.85, p value < 1×10^{-5}), suggesting a limited environmental influence but a higher variability between in vitro and in vivo expression levels, highlighting the challenges of translating in vitro promoter activity to in vivo applications. Based on these data, we design the Schantzetta library, composed of eight promoters spanning a wide expression range and displaying a high degree of robustness in both laboratory and in vivo conditions (R^2 = 0.98, p = 0.000827). This study provides a systematic assessment of the $\sigma70$ promoter activity in *E. coli* as it transits the murine gut leading to the definition of robust expression cassettes that could be a valuable tool for reliable engineering and development of advanced microbial therapeutics.

KEYWORDS: *probiotics, promoter, gut, microbiota, engineering, flow-seq*

INTRODUCTION

Synthetic biology has delivered a wide range of powerful tools and methods enabling researchers and engineers to design microorganisms with a bottom-up approach.[1] These tools have to a large extent been confined to controlled conditions of the laboratory. Yet, synthetic biology is now delivering heavily engineered microorganisms for biobased chemical[2] or therapeutic molecule production.[3,4] Another promising application of synthetic biology lies in medicine where engineered cells are being developed as therapeutics and diagnostics.[5] Building on the growing data supporting the importance of the gut microbiota in human health,[6,7] advanced microbial therapeutics (AMTs) stand out as a powerful approach to treat numerous diseases.[8] Engineered strains of probiotic *Escherichia coli* Nissle 1917 have been used to generate promising results as a treatment of phenylketonuria and hyperammonemia.[9,10] However, despite recent progress, it remains challenging to engineer microbes with predictable phenotypes in the dynamic and complex environments of mammalian hosts. Indeed, the degree of correspondence between the output of defined genetic parts in the laboratory and in the host is poorly understood. Even for commonly used probiotic organisms, such as *E. coli* Nissle 1917, knowledge about regulatory elements rely mostly on in vitro characterization.[11−18] The lack of a robust and predictable expression cassettes for use in vivo strongly hampers the development AMTs.

The use of fluorescence activated cell sorting (FACS) to sort diverse regulatory libraries based on their protein expression output followed by deep sequencing, frequently referred to as flow-seq, is well suited for investigating the output of regulatory elements.[11,13,14,18] Indeed, flow-seq is very useful for characterization of parts for use in synthetic biology in bacterial[6−11] and eukaryotic hosts.[16,17,19,20] Metagenomic mining and large-scale DNA synthesis has been used in combination to assess the activity of novel regulatory elements in the laboratory.[21] Yet, flow-seq has not been widely applied to study the activity of parts-libraries in vivo.[22]

Received: July 14, 2021
Published: November 29, 2021

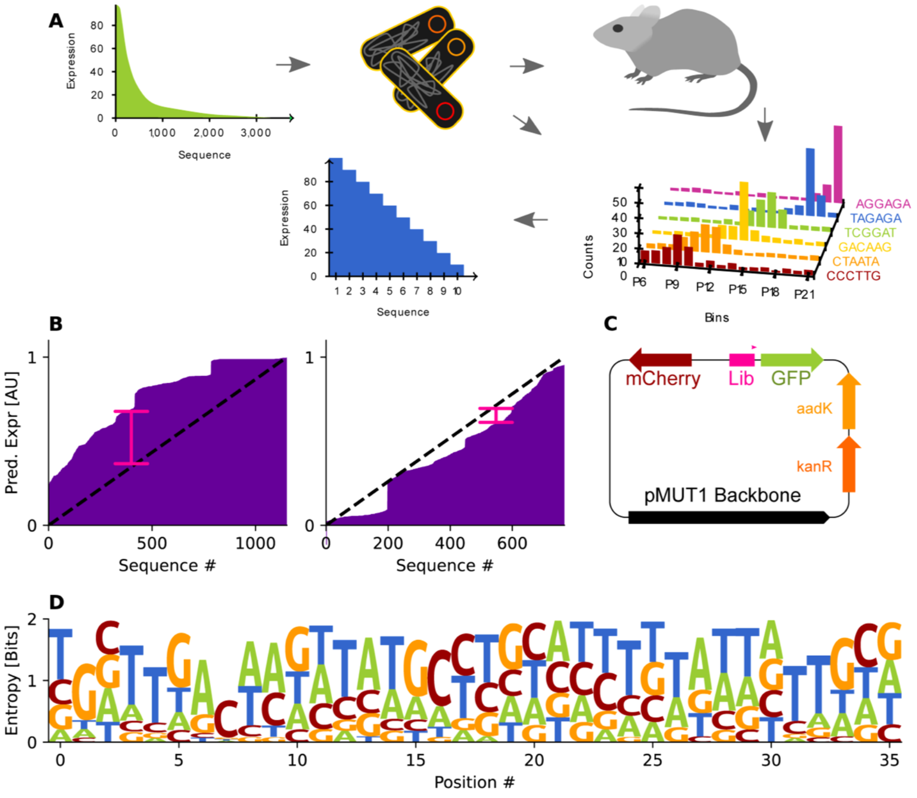


Figure 1. Design and generation of the σ^{70} promoter libraries. (A) Schematic workflow of the study. A semi-random library containing a large number sequences is created and transformed into *E. coli* Nissle and then inoculated into mice. In vitro and in vivo expression levels of individual sequences are determined using flow-seq. (B) Example of two promoter libraries evaluated by sorting every sequence according to their predicted expression level (purple) and comparing the vertical distance (pink) to the ideal library (black dashed line). Predicted expression is measured in arbitrary units (AU) ranging from 0 (no expression) to 1, which is the maximally observed expression level. (C) Plasmid construct used to measure the expression level of the promoters in the library. The plasmid is based on the cryptic pMUT1 plasmid backbone, since this particular backbone has displayed remarkable stability in *E. coli* Nissle. mCherry is expressed from a constant promoter and GFP under the control of the library and used to quantify the expression level of each promoter in the library. A kanamycin resistance gene (*kanR*) is used as a selection marker for cloning, and *aadK* is used to confer streptomycin resistance. (D) Logo plot of the final combined σ^{70} library showing the high diversity of the design. Letter height is scaled by total entropy, not information content.

Recent efforts on microbiome expression machinery mainly focused on *Bacteroides* species due to its abundance in the gut, long term colonization potential, and poorly characterized regulatory elements.[23] A set of inducible promoters and regulatory circuits was developed for *Bacteroides thetaiotaomicron* and demonstrated to function in vivo based on stool-based measurements.[24] Additionally, a novel phage promoter was used to regulate fluorescent markers in several different *Bacteroides* species in vivo.[25]

However, it remains poorly understood to which degree regulatory elements function differently in distinct sites of the mammalian gastrointestinal (GI) tract and to what extent the output in vivo is correlated to the output under controlled laboratory conditions in vitro. Indeed, the growth rate, oxygen availability, and nutrient availability varies substantially in the gastrointestinal tract, which could give rise to divergent outputs of regulatory elements.[26−29] One study started to address this gap by profiling the activity of 30 constitutive promoters in *E. coli* Nissle, suggesting significant differences in promoter expression depending on the gut site.[22] Consequently, building a defined set of *E. coli* gene expression cassettes that function reliably and predictably in the gastrointestinal tract would be desirable.

We hypothesized that flow-seq could be used as a method to efficiently measure expression levels of reporter protein directly in samples taken from mice inoculated with the probiotic *E. coli* Nissle 1917 strain[30] and sought to establish flow-seq as a high throughout method to investigate the behavior of genetic elements in *E. coli* and other commensal species in vivo.[25]

As a basis for this investigation, to establish a defined set of expression cassettes, we chose the *E. coli* housekeeping promoter σ^{70} for our study. The σ^{70} is a ubiquitous housekeeping promoter across several bacterial genus displaying an always-on transcriptional activity. Accordingly, it is commonly used in bacterial synthetic biology, e.g., the Anderson promoter library (iGEM registry parts J23100 through J23119). The core promoter sequence of σ^{70} has been extensively characterized in vitro, enabling rational design of diverse promoter libraries.[12,18,31−33]

To bridge the gap between laboratory and host environments, we generated a σ^{70} promoter library in *E. coli* Nissle and investigated the expression of over 8000 promoter constructs under aerobic and anaerobic conditions in vitro as well as in samples collected throughout the murine GI tract.

■ RESULTS

Designing a σ^{70} Promoter Library that Evenly Spans the Widest Range of Expression Levels in *E. coli*. In order to design a library exploring a wider and more diverse sequence space than previous efforts,[34,35] a two-step approach

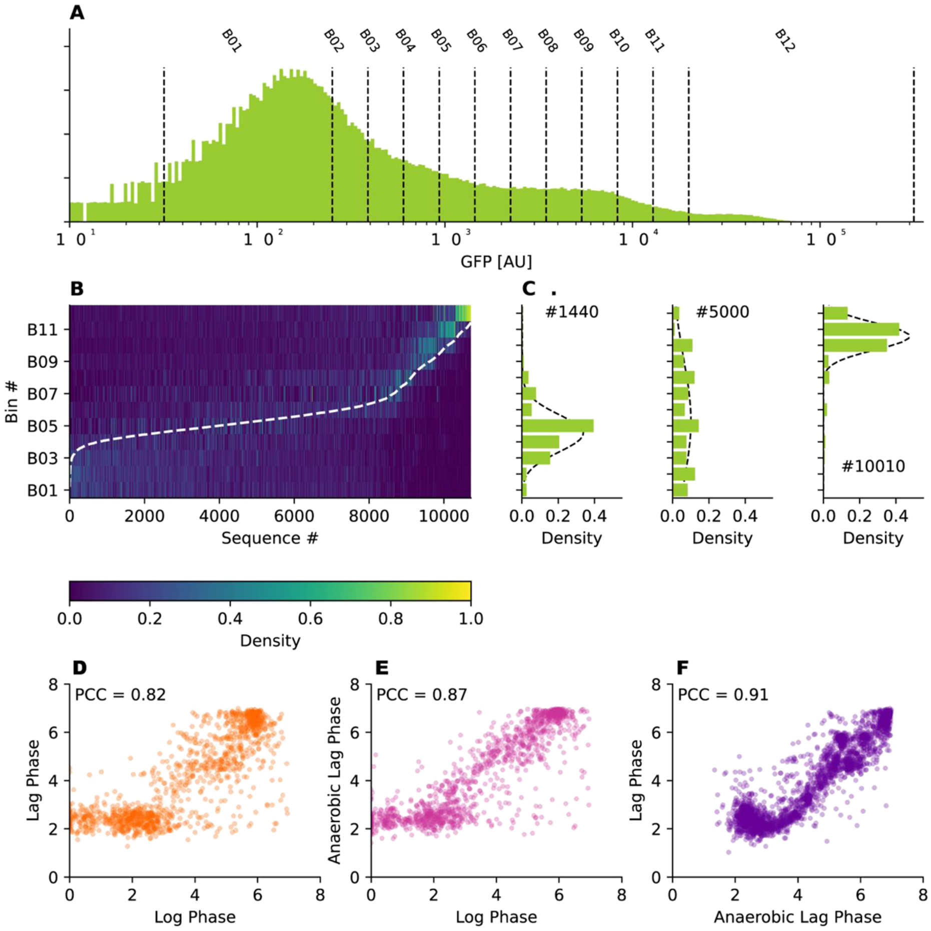


Figure 2. Sorting and sequencing the promoter libraries in vitro. (A) The GFP fluorescence is divided into 12 bins to capture different expression levels. (B) After sequencing, each occurrence of a sequence is counted across all bins. The heatmap shows the density. The dashed line is the calculated expression level. (C) A few examples of sequences are shown displaying different distributions from a well-behaved low expression (left), a high deviation (middle), and a well-behaved high expression level. The dashed line shows a Gaussian curve fitted over the bin distribution. (D−F) Pairwise comparison of the different in vitro expression datasets.

was taken. First, raw data from a previous in vitro study was re-analyzed to couple each sequence to an expression level. This data was used as training data for a random forest model,[14] as these machine learning algorithms demonstrate a very high time-to-performance ratio and robustness against overfitting, making them particularly suitable for such genomic applications. A scoring method was set up to evaluate a potential sub-library composed of a degenerate sequence. The expression level for each individual sequence was predicted using the random forest model; expression levels were sorted and ordered, to then be compared to an ideal library distribution using the Kolmogorov−Smirnov distance,[36] which represents a quantitative measure of resemblance between the two distributions and is a reliable indicator of library quality (Figure 1B).[37] Libraries with a low Kolmogorov−Smirnov distance are more similar to the ideal library distribution, and therefore display a more even expression coverage.

Second, a simulated evolution algorithm was designed to create sub-libraries displaying an even expression coverage by generating successions of virtual populations of potential sub-libraries ranked by the scoring method. The simulated evolution randomly initiates a population of degenerate sequences with the target diversity, ranking them based on the objective function and passing high-performing sequences along to the next iteration. To ensure a high diversity in the final library, an extra penalty was added to the objective function de-prioritizing bases already present in the same position on a previously designed sub-library. The final library was composed of a combination of nine degenerate sub-libraries (Figure 1C), which showed a broad expression distribution in our model indicating the presence of several different expression levels. The individual sub-libraries contained between 512 and 1152 unique promoter sequences, summing to a total library diversity of 8704 individual sequences (Figure 1D).

Construction of Diverse Promoter Libraries for Assessment of In Vivo Gene Expression. In vivo flow-seq presents some unique technical challenges including isolation of the target organism from the other microbial constituents of the microbiome and particles in mouse fecal matter, maturation of the fluorescent protein under the environmental conditions of the gut, and colonization of the target organism in vivo.

E. coli Nissle 1917 was chosen as the wild-type *E. coli* chassis due to its proven safety and common use as a probiotic and engineered microbial cell therapy.[30] *E. coli* Nissle harbors a native cryptic plasmid pMUT1,[38] which has been shown to be highly stable.[39] Therefore, *E. coli* Nissle was cured from its native pMUT1, to leverage pMUT1 for the strain engineering. The *kanR* kanamycin resistance gene was added to the plasmid

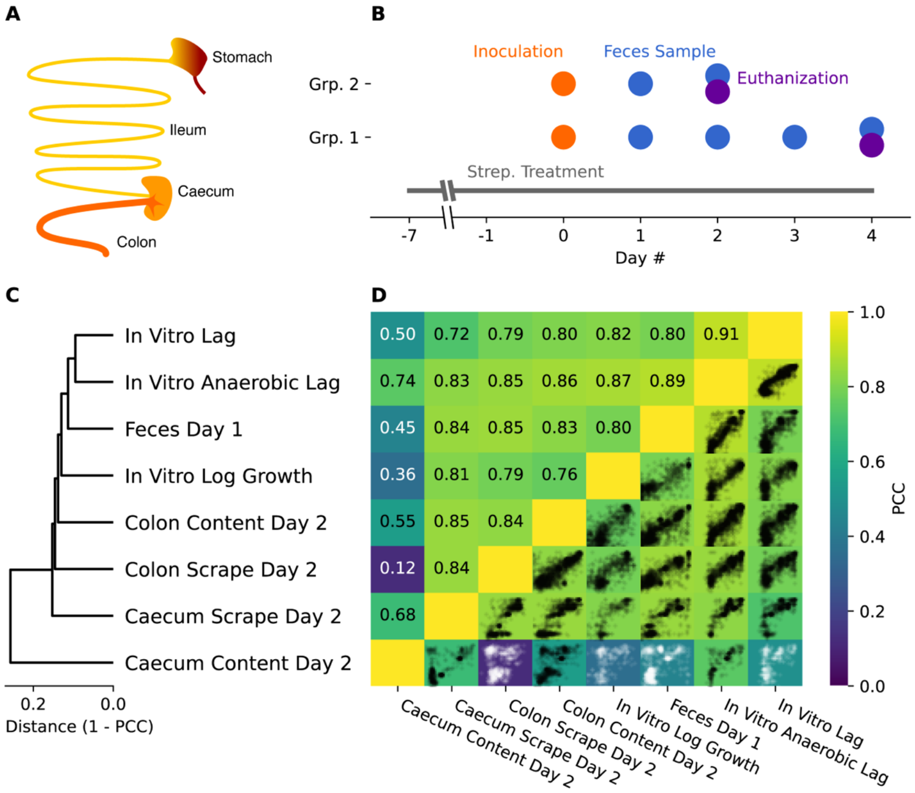


Figure 3. Setup and results from the in vivo experiments. (A) Sampling locations within the murine GI tract. (B) Experiment design. 12 mice in 6 cages were treated with streptomycin for the duration of the experiment. At day 0, mice were inoculated with the *E. coli* Nissle libraries and feces samples were collected from cages throughout the experiments. The mice were divided into two groups, group 1 being euthanized on day 2 and group 2 on day 4 to perform a section and extract samples from the intestines. (C) Dendrogram of the locations created using hierarchical clustering location-to-location Pearson's *R* correlation as a distance measurement. (D) Correlation matrix of the locations compared to each other, sorted using the dendrogram from the clustering. The lower triangle displays the scatter plots of expression values recorded at each location. Numbers in the upper triangle and the color of each field are the Pearson's *R* values calculated based on the scatter plots.

for cloning purposes along the *aadK* streptomycin resistance gene that was included to enable the strain to resist streptomycin pre-treatment used in the in vivo experiments.[40]

Since GFP needs oxygen to mature,[41] alternative reporters were investigated, including the RNA-based spinach aptamer and flavin-based fluorescent proteins.[42] However, these reporters were too dim to measure over the high autofluorescence around green wavelengths in mouse feces using the present experimental setup. Experiments were then carried out to explore a post-maturation step for sfGFP.[43] It was found that even the few minutes it took to prepare samples for flow-cytometry were sufficient to mature the fluorescence signal from sfGFP allowing for cell sorting by fluorescence intensity. This was also the case for cells grown completely anaerobic in deoxygenated medium (Figure S1). To deal with the high autofluorescence from mouse feces in the GFP channel, an mCherry[44] red fluorescent protein was added with constitutive expression as a way to identify the engineered *E. coli* Nissle particles in the flow cytometer (Figure 1C).[45] mCherry fluorescence displays bimodal distribution on some of the sublibraries (Figure S1), possibly due to some promoters having activity in the reverse direction or plasmid supercoiling. GFP fluorescence levels were not normalized to mCherry fluorescence levels to preserve potential promoter's broad expression profiles.

Oxygen Availability and Growth Phase Has Limited Impact on Promoter Activity In Vitro. A major difference between typical laboratory test conditions and the mammalian gastrointestinal tract is the oxygen level. To assess the impact of oxygen on the protein expression from the σ^{70} promoter libraries, we cultured *E. coli* Nissle, harboring this library under aerobic and anaerobic conditions. Following an overnight culture, the libraries where sorted based on the GFP expression levels into 12 bins using FACS (Figure 2A). Cells in each bin were recovered overnight, after which pMUT1 plasmids were extracted. The region of the pMUT1 plasmid containing the σ^{70} promoter library was PCR amplified and sequenced using Illumina MiSeq. Expression levels for each sequence were calculated by fitting the mean of the distribution across flow cytometry bins[11] (dashed line on Figure 2B,C).

Protein expressions from promoter sequences in vitro under aerobic and anaerobic conditions are strongly correlated. A total of 2865 recovered sequences were found to overlap between all three libraries. Considering only library sequences with enough reads (≥25 reads) to calculate the expression levels, 6899 individual sequences (79.3% of the total library diversity) were recovered for overnight aerobic incubation, 2591 individual sequences (29.8% of the total library diversity) for mid-log aerobic incubation, and 7039 individual sequences (80.9% of the total library diversity) anaerobic overnight incubation. Despite an overall difference in reporter fluorescence distribution (Figure S1), the analysis of expression levels showed that promoters generally behaved similarly in all three conditions, with Pearson's *R* correlation from 0.82 to 0.91 (p value < 10^{-5}, Figure 2D−F). Accordingly, it can be concluded that the activity of the σ^{70} promoter

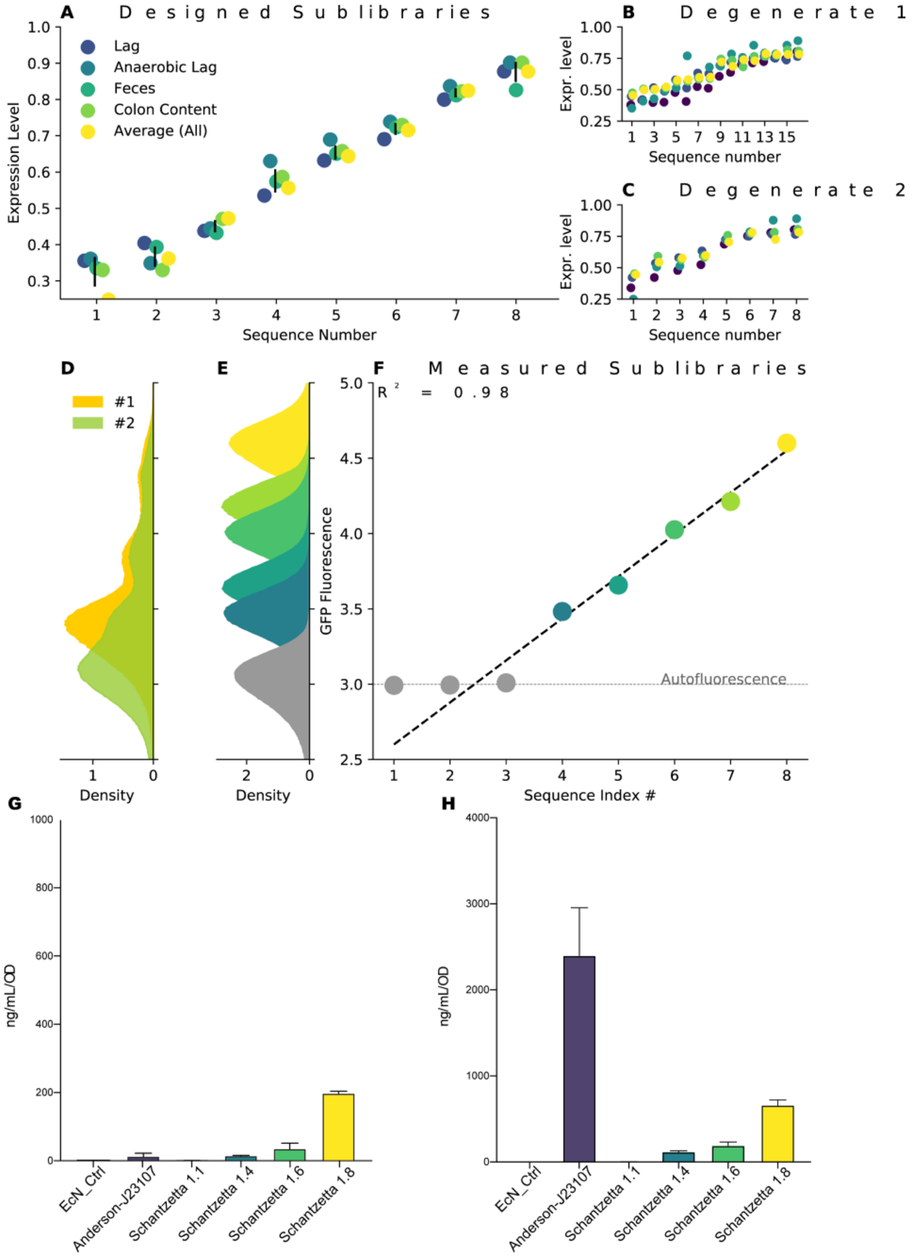


Figure 4. Determination of promoter libraries most suitable for predictable engineering across different environments. (A) Schantzetta library composed of eight promoter sequences with the highest degree of consistency between in vitro and in vivo environments. Chosen from lowest to highest with even spacing between them. Expression level is defined as in Figure 2B,C and compressed to the 0−1 scale. (B, C) Degenerate 1 and 2 sequences are determined in a similar fashion as the eight promoter sequences but must fit on a single degenerate primer for easy cloning. (D) Degenerate libraries re-cloned and measured in a flow cytometer in *E. coli* Nissle in the stationary phase, showing fluorescence at both low and high levels of expression. (E) Eight peaks of the individual promoter libraries measured in a flow cytometer in *E. coli* Nissle in the stationary phase. (F) Average expression of the eight individual promoter libraries compared with their sequence index. The three lowest expression levels show the same mean expression as the negative control, meaning that they do not show any expression above autofluorescence. The dashed line shows a linear regression of those promoter level measurements above negative control illustrating the high coefficient of determination (R^2 = 0.98, p value < 10^{-5}). (G, H) In vitro expression in *E. Coli* Nissle of human hormones GLP1 (G) and IGF1 (H) under four promoters from the Schantzetta library (1.1, 1.4, 1.6, and 1.8). Expressions can be compared between a popular medium strength promoter from the Anderson library (J23107).

libraries is similar across these three different in vitro conditions.

σ^{70} Promoter Expression Levels Show Little Variation throughout the Murine GI Tract. To assess the activity of the promoter libraries in the murine gut (Figure 3A), *E. coli* Nissle cultures harboring the σ^{70} promoter libraries were administered by oral gavage to streptomycin-treated mice and subsequently analyzed with flow-seq to determine the promoter expression levels. The experiment was set up using 12 mice, with two mice per cage. After a 7 day pre-treatment period with streptomycin, the mice were inoculated (on day 0) with the library. On all days, fecal samples were taken and processed, and on days 2 and 4, each half of the mice were sacrificed and content and scrape samples were taken from ileum, caecum, and colon (Figure 3B).

Each sample was sorted by GFP fluorescence into 8 or 12 bins, depending on the degree of biomass and fluorescent cells recovered. The degree of recovery varied greatly due to the nature of the samples. Unfortunately, ileum content and scrape bacterial load was insufficient to perform cell sorting. Between 366 and 1600 sequences per conditions passed the filters for expression level analysis leading to 78 promoters assessed in all conditions, with a sufficient number of reads in each condition (≥25 reads). No samples were recovered from days 3 and 4 due to washout of bacteria from the mouse gut (Figure S2).

To compare samples with each other, the average expression level for each location was first used to calculate Pearson's R between all locations including the in vitro samples (Figure 3C,D). The promoter expression level showed a relatively high correlation (R^2: 0.55−0.85, p value $< 1 \times 10^{-5}$, except for caecum content vs colon scrape comparison yielding Pearson R equal to 0.12) between different locations in the gastrointestinal tract using this test. The expression levels across locations show very little variation, with a mean standard deviation of 0.739 expression levels out of eight expression levels (arbitrary units). This suggests a limited environmental influence over promoter expression with cecal samples being most distinct from other locations with regards to promoter expression levels.

The correlation between in vitro locations and in vivo expression levels was lower (R^2: 0.44−0.85, p value $< 1 \times 10^{-5}$) for the σ70 promoter. These data highlight the challenges of translating in vitro promoter activity to in vivo applications.

A Defined Set of *E. Coli Promoters* with Predictable Expression throughout the Murine GI Tract. To engineer AMTs using *E. coli* Nissle as a chassis, a set of gene expression cassettes with predictable protein expression throughout the gastrointestinal tract would be beneficial. To support such future engineering efforts, a number of sequence libraries were extracted from the expression level data. The first library (Schantzetta1) comprised eight sequences selected to span expression levels uniformly distributed from lowest to highest in the dataset. These eight sequences were further chosen to have the lowest degree of variation in expression levels across all locations in the murine gastrointestinal tract and in vitro conditions (Figure 4A). Two additional libraries were selected under the same constraints but with the additional condition that they could be represented on a degenerate primer. The two degenerate libraries contain 8 (Schantzetta2a) and 16 sequences (Schantzetta2b) (Figure 4B,C and Table S1).

To verify the expression level of Schantzetta1, each sequence was re-synthesized and cloned into a fresh stock of the original plasmid. Isolated clones of each of the individual sequences were measured separately (Figure 4D,E), and the degenerate libraries were measured as a whole library on a flow-cytometer (Figure 4F). The libraries showed the expected wide distribution of expression levels. The three sequences with the lowest expression level (Schantzetta1.1−3) could not be sufficiently resolved on our plate reader due to background fluorescence, yet the five sequences with the highest expression levels (Schantzetta1.4−8) showed a very high correlation with the previous flow seq data ($R^2 = 0.98$, $p = 0.000827$).

To test the suitability of the Schantzetta library to produce therapeutic compounds, we tested in vitro some promoters of the Schantzetta1 library for the expression of two human hormones: GLP1 and IGF1. While the expression level did not follow the same quantitative expression levels as observed for GFP, we observed a progressive concentration increase for the two peptides as the promoter index increased (Figure 4G,H). Similar to the results with GFP (Figure 4F), the hormone concentration was below the detection level for the Schantzetta1.1-based construct (Figure 4G,H). Notably, the absolute production levels of these two specific proteins appear to be generally similar for the Schantzetta promoters, compared to the Anderson-based constructs exhibiting an important variation between GLP1 and IGF1, around 50-fold.

■ DISCUSSION

With the emerging application of synthetic biology to create programmable bacterial therapeutics, also referred to as AMTs, the need is growing for parts with robust gene expression in vivo. Previous studies[22] have demonstrated that flow-seq is suitable for the high-throughput study of promoters, highlighting a high correlation between GFP protein levels and fluorescence independent of barcoding, as well as a good correlation between the RNA/DNA ratio and GFP fluorescence. Consequently, we leveraged flow-seq to characterize the expression of a 8704 σ^{70} promoter library in *E. coli* Nissle under a range of in vitro and in vivo conditions. We showed that most of the promoters demonstrate a similar expression along the murine GI tract but observed important differences when comparing the in vitro and in vivo expression, in agreement with a previous study of 30 promoters.[22] Based on the data from this study, we designed the Schantzetta library, comprising promoter sequences with predictable, robust, and consistent gene expression across the murine gastrointestinal tract and in the laboratory. Therefore, the Schantzetta library offers valuable tools to allow precision engineering of the next generation probiotics. In particular, the Schantzetta library could be leveraged to precisely and reliably adjust the constant expression of proteins requiring a specific concentration in the host, such as hormonal and neurotransmitter-based therapeutics, and could be an attractive solution for the engineering of strains with antipathogenic activities.

The Schantzetta constitutive promoters provide a complementary tool to the broadly used inducible promoters such as FNR, pTET, and pBad. The Schantzetta library, while not offering the ease of direct expression tunability, allows a constant predictable gene expression. Indeed, availability and regular delivery of the inducer in the gut can be challenging in vivo to control the variability in gene expression compared to constitutive promoter. Moreover, finding the appropriate inducer remains troublesome, as the most characterized inducible systems can be toxic for the host and its microbiome (tetracycline) or processed natively by other microbiome members (polysaccharides).

The context effects of the Schantzetta promoters were not investigated, as the intention of the study was to focus on the functional relationships in the core promoter sequence as well as producing promoter sequences intended for use in synthetic biology in which the context is well defined. Because this study aims at providing new tools rather than demonstrating therapeutic efficacy, the expression of therapeutic molecules in antibiotic-free or disease mouse models were not investigated. Additionally, precision engineering on plasmids accelerates the development and testing of candidate therapeutics before a potential genomic integration of the final design. Subsequent studies are needed to explore the impact of promoter context, as well as the influence of animal model breed or diet, on the in vivo expression level and persistence of the strain in the host. Still, we believe these data constitute a useful resource that will contribute to increasing the in vivo reliability of AMTs.

■ MATERIALS AND METHODS

σ^{70} Promoter Library Design. Data was fetched from the NCBI Sequence Read Archive accession number SRA012345. Sequences from the rnap-wt and full-wt experiments were analyzed by counting occurrences in each bin 1−9 in the experiments and calculating the final expression value as $x_s = \sum_b e_b \cdot c_{b,s} / \sum_b c_{b,s}$ with $c_{b,s} = n_{b,s} / \sum_s n_{b,s}$, where $n_{b,s}$ is the count of sequence s in bin b and e_b is the mean fluorescence value in bin b.

A total of 36 base pair long sequences from rnap-wt and full-wt were then extracted and encoded using a one-hot encoding scheme, i.e., representing A, T, G, and C with the codes [1,0,0,0], [0,1,0,0], [0,0,1,0], and [0,0,0,1], respectively. The final input vector was then a concatenation of 36 codes corresponding to the base at each sequence position, resulting in 144 input features in total. The random forest regression implementation from scikit-learn was used to train a model containing 250 trees using otherwise default parameters on the rnap-wt dataset and using the full-wt dataset as a hold-out test set to ascertain model performance.

To generate the degenerate libraries, a simulated evolution algorithm was implemented performing the following steps: First, 1000 libraries are randomly generated and scored. Then, for each generation: (1) Retain the top 20% libraries, add this to the parents list. (2) Randomly select 5% of the remaining libraries and add them to the parents list. (3) Select 1% of sequences from the parents list and mutate them. (4) Add all the libraries from the parents to a "new population" list. (5) Fill up the new population list until it contains 1000 libraries by picking two random libraries from parents list and ″mate″ them.

The simulation usually converges in 50−100 generations, where each new generation sees no improvement to the objective function. The sequence libraries are represented as a 36 × 4 matrix. Each column is a sequence position with four Booleans for A, T, G, and C. To mutate a sequence, 5% of positions in the matrix are randomly set to false and 5% are randomly set to true. If a column contains 4× false, it is reverted to its original state. To mate two sequences, random sections of each library are put together to create the child. After each mutation or mating, the library is corrected to the target diversity by adding or subtracting bases until it reaches the desired diversity.

Each library is scored using an objective function, which calculates the maximum distance between the actual distribution and the optimal distribution, where the optimal distribution is a straight line from 0 to maximum expression.

Cloning Procedures. The final plasmid (pMSKL023 in Table S2) was assembled with the NEBuilder HiFi kit (New England BioLabs, Inc.) using the manufacturer's instructions and cloned into One Shot TOP10 Electrocompetent *E. coli* (Invitrogen) the using manufacturer's instructions. The pMUT1 backbone originally isolated from *E. coli* Nissle 1917 was donated by Dr. Mareike Bongers. *aadK* was donated by Dr. Andreas Porse.

Libraries were cloned by PCR amplifying from pMSKL023 with a primer containing the degenerate library on a primer with overhang (oMSKL203,206,209,211−216) and a reverse primer (oMSKL133). The single fragment was purified and assembled with the NEBuilder HiFi kit (New England BioLabs, Inc.) using the manufacturer's instructions and cloned into One Shot TOP10 Electrocompetent *E. coli* (Invitrogen) using a modification of the manufacturer's instructions: After an hour of recovery, the whole cell suspension is added to 12 mL 2 × YT media (Sigma-Aldrich) with 100 μg/mL kanamycin and grown overnight at 37 C in a shaking incubator.

Plasmid libraries were then purified using the Macherey−Nagel NucleoSpin Plasmid kit according to the manufacturer's instructions and electroporated in an *E. coli* Nissle 1917 strain previously cured of the native pMUT1 plasmid donated by Mareike Bongers.

***In Vitro* Flow-Seq.** A 250 μL overnight cell suspension of *E. coli* Nissle containing the library was grown overnight in 15 mL 2 × YT media (Sigma-Aldrich) in aerobic conditions and anaerobic conditions. A 250 μL cell suspension grown aerobic were further diluted in 15 mL 2 × YT media and grown to an OD_{600} of 0.4. Cell suspensions were diluted 1000× into PBS (from 10× stock solution, pH 7.4, Invitrogen) with 0.5% v/v tween-20 (*E. coli* Nissle will aggressively clump and adhere to the FACS instrument without tween causing blockages). Diluted cell suspensions were added to a BD FACSAria II for sorting.

Using the aerobic overnight as a reference, two gates were created on the FITC channel log-histogram at the lowest and highest fluorescence levels. Ten gates were created with equal spacing between the upper and lower gate, making a total of 12 bins sorting on the FITC channel. For each bin, up to 1,000,000 million cells were sorted in a new tube with 500 μL 2 × YT media, with sorting stopped after 10 min unless a bin had less than 25,000 sorting events.

A variation of the "16S Metagenomic Sequencing Library Preparation" protocol was used to sequence the sorted libraries.[46] Sorted cell suspensions were grown overnight in 5 mL 2 × YT media, after which a 221 bp DNA fragment was PCR amplified directly from 1 μL cell culture using primers oMSKL219 and oMSKL220 for eight cycles using Phusion PCR master mix (Invitrogen) according to the manufacturer's instructions (three-step protocol with a melting temperature of 60 °C, 10 s extension time) in 25 μL of total reaction volume. The remaining primers were digested by adding 0.5 μL exonuclease I (20 U/μL, Thermo Scientific) directly to the PCR reaction mixture and incubating for 15 min at 37 C and inactivating for 15 min at 85 °C in a PCR cycler. Nextera XT (Illumina) sequencing adapters were added to PCR fragments by adding 1.25 μL of each forward and reverse adapter and re-running the PCR program for 12 cycles with a primer melting temperature of 65 °C.

PCR fragments were then purified and normalized using Just-a-Plate 96 PCR Purification and Normalization Kit (Charm Biotech) according to the manufacturer's instructions. Sequencing libraries were then pooled and sequenced on an Illumina MiSeq.

Mouse Experiments. Mouse experiments were approved by the Danish Animal Experiments Inspectorate (license number 2015-15-0201-00553) and carried out in accordance with existing Danish guidelines for experimental animal welfare. Twelve male NMRI outbred mice (Taconic Europe), 6 weeks of age, were divided into six cages and two groups. The mice were housed in type III Makrolon cages (Techniplast, Varese, Italy) containing a bedding, nesting material, hiding place, and wooden block and were fed standard Altromin 1314 chow (Brogaarden, Gentofte, Denmark). The experiment took place over 12 days, where all mice received 5 mg/mL streptomycin in sterile drinking water over the entire period. After a pretreatment period of 7 days, each mouse was inoculated with 100 μL of overnight *E. coli* Nissle library culture washed and concentrated to OD_{600} = 10 with gavage. On each day, at approximately 24 h intervals, feces were collected from the cages. On the third day after inoculation, the mice from group 1 were euthanized and dissected, and on the fifth day group 2 was euthanized and dissected.

After euthanasia, samples were taken from the ileum, caecum, and colon. Content of each gut sample was extracted and stored in 1 mL PBS (from 10× stock solution, pH 7.4, Invitrogen) on ice before being run through the flow cytometer within an hour. Tissue samples were rinsed in a saline solution (0.90% NaCl), and mucus was scraped off and stored in 0.5 mL PBS on ice.

In Vivo Flow-Seq. Feces and tissue samples were dissolved in a 1.5 mL PBS-Tween (from 10× stock solution, pH 7.4, Invitrogen with 0.5% v/v tween-20) on ice and manually homogenized. Solids were spun down at 500 × *g*, and the supernatant was filtered through a 40 μm cell strainer integrated into the lid of a test tube (Corning Falcon Test Tube with Cell Strainer Snap Cap, Fisher Scientific). Filtered samples were then sorted as soon as possible with the same settings as the in vitro samples on the BD FACSAria II, with the exception that a gate was added on the FSC/SSC scatter plot and on the mCherry channel (561 nm laser, 610/20 nm bandpass filter) to isolate *E. coli* Nissle. On days 1 and 3, 12 gates were used to sort green fluorescence, and on days 2 and 4, and an eight gate setup was used instead to save time. Samples were sequenced using the same protocol as the in vitro libraries.

Quantification of Expression Levels. To quantify protein expression levels of samples, a script was written to count promoter sequences. Expression levels calculate the final expression value as $x_s = \sum_b b \cdot c_{b,s} / \sum_b c_{b,s}$ with $c_{b,s} = n_{b,s} / \sum_s n_{b,s}$, where $n_{b,s}$ is the count of sequence s in bin b.

Sample Comparison. All samples were compared against one another. For each pair of samples with quantified expression levels, all sequences were extracted that were present in both samples. A scatterplot was created of expressions levels as measured in each sample. The correlation was quantified using Pearson's *R* value.

Extracting High Quality Sequences. Individual sequences were extracted using a brute force search. Eight expression levels were predefined, and for each promoter in the Aerobic Stationary, Anaerobic Stationary, Feces 1 + 2, and Colon Content Mouse 6 libraries, the distance of the measured promoter to the desired expression level was calculated, and the promoter with the lowest average distance was chosen to represent the expression level.

Degenerate libraries were created in a similar fashion to the initial promoter library design, i.e., using an evolutionary algorithm minimizing the objective function defined by the longest distance at a target expression level to the average of the expression levels recorded in the samples mentioned above. Promoter sequences generated, which did not have a recorded expression level was set to have an expression of −1, to discourage promoter sequences without recorded expression levels.

Newly extracted sequences and degenerate promoters were ordered on primers and cloned into a fresh pMSKL023 stock using the NEBuilder HiFi kit (New England BioLabs, Inc.) using the manufacturer's instructions in One Shot TOP10 Electrocompetent *E. coli* (Invitrogen), purified and electroporated into *E. coli* Nissle 1917. A 250 μL overnight cell suspension grown from sequence-verified colonies was grown overnight in 15 mL 2 × YT media (Sigma-Aldrich), in aerobic conditions and run in a Sony SH800S Cell Sorter (Sony Biotechnology Inc.) in quantification-only mode.

Construction of IGF1 and GLP-1 Plasmids and In Vitro Expression. The plasmids pHH06 (IGF1) and pHH10 (GLP-1) were constructed using a codon optimized version of the human peptides (Uniprot identifier PRO_0000015664 and PRO_0000011258, respectively). The cDNA of the genes including the constitutive promoter J23107 (http://parts.igem.org/Promoters/Catalog/Anderson), 5'UTR, and OmpA secretion tag were synthesized from a commercial source (IDT). The primers for cloning were designed with the AMUSER web tool (http://www.cbs.dtu.dk/services/AMUSER/). Phusion U Polymerase from Thermofisher Scientific was used for amplification of backbone and DNA fragments. The plasmids pHH06-J23107 and pHH10-J23107 were constructed by inserting the cDNA into the backbone of pMUT1 plasmid used in this study. Construction of the different plasmids was done by using USER cloning.[47] The Schantzetta promoters 1.1, 1.4, 1.6, and 1.8 were introduced using long primers synthetized by IDT and pHH06-J23107 or pHH10-J23107 as the DNA template (Table S2). All constructed plasmids were confirmed by Sanger sequencing and transformed into *E. coli* Nissle 1917.

Individual colonies for each construct were inoculated in 3 mL of LB media containing 50 μg/mL kanamycin and incubated at 37 °C overnight. Next day, the cultures were diluted 100-fold in new LB media containing 50 μg /mL kanamycin at desired volumes. Strains were cultured aerobically in a shaker at 200 rpm at 37 °C. Aliquots of the culture were collected at regular intervals and centrifuged at 10,000 × *g* for 5 min. The supernatant was collected and stored at −20 °C for later analysis. The samples were subjected for the GLP-1 or IGF1 assay by ELISA as described by the manufacturer's protocol (Abcam product: ab184857 and ab211651, respectively).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.1c00325.

> (Figure S1) Histograms of mCherry and GFP fluorescence levels in the in vitro experiments; (Figure S2) fraction of particles in the FACS passing the forward- and side-scatter gate showing mCherry fluorescence; (Table S1) promoter sequences of the Schantzetta library (1.1 to 1.8), by relative strength order, as well as the degenerate libraries 1 and 2 sequences; (Table S2) strains and plasmids used in this study (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Morten O.A. Sommer − *Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-*

2800 Lyngby, Denmark; orcid.org/0000-0003-4005-5674; Email: msom@bio.dtu.dk

Authors

Jeremy Armetta – *Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Lyngby, Denmark*

Michael Schantz-Klausen – *Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Lyngby, Denmark*

Denis Shepelin – *Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Lyngby, Denmark*

Ruben Vazquez-Uribe – *Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Lyngby, Denmark*

Martin Iain Bahl – *National Food Institute, Technical University of Denmark, DK-2800 Lyngby, Denmark*

Martin Frederik Laursen – *National Food Institute, Technical University of Denmark, DK-2800 Lyngby, Denmark*

Tine Rask Licht – *National Food Institute, Technical University of Denmark, DK-2800 Lyngby, Denmark*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acssynbio.1c00325

Author Contributions

+J.A. and M.S.-K. contributed equally to the work

Author Contributions

M.S.K. conceived and planned the experiments. J.A., M.S.K., and R.V.U. performed the in vitro experiments. J.A., M.S.K., M.I.B., M.F.L., and T.R.L. contributed to the in vivo experiments. J.A., M.S.K., and D.S. analyzed the data. M.O.A.S. supervised the study. J.A., M.S.K., D.S., and M.O.A.S. wrote the final manuscript in consultation with R.V.U., M.I.B., M.F.L., and T.R.L.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was funded by The Novo Nordisk Foundation under NFF grant number: NNF10CC1016517 and The Novo Nordisk Foundation, Challenge programme, CaMiT under grant agreement: NNF17CO0028232.

■ REFERENCES

(1) Wang, Y.-H.; Wei, K. Y.; et al. Synthetic Biology: Advancing the Design of Diverse Genetic Systems. *Annu. Rev. Chem. Biomol. Eng.* **2013**, *4*, 69–102.

(2) Ford, T. J.; Silver, P. A. Synthetic biology expands chemical control of microorganisms. *Curr. Opin. Chem. Biol.* **2015**, *28*, 20–28.

(3) Luo, X.; Reiter, M. A.; et al. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **2019**, *567*, 123.

(4) Dou, J.; Bennett, M. R. Synthetic Biology and the Gut Microbiome. *Biotechnol. J.* **2018**, *13*, 1700159.

(5) Riglar, D. T.; Silver, P. A. Engineering bacteria for diagnostic and therapeutic applications. *Nat. Rev. Microbiol.* **2018**, *16*, 214–225.

(6) Rooks, M. G.; Garrett, W. S. Gut microbiota, metabolites and host immunity. *Nat. Rev. Immunol.* **2016**, *16*, 341–352.

(7) Valles-Colomer, M.; Falony, G.; et al. The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat. Microbiol.* **2019**, *4*, 623–632.

(8) Mimee, M.; Citorik, R. J.; et al. Microbiome therapeutics — Advances and challenges. *Adv. Drug Delivery Rev.* **2016**, *105*, 44–54.

(9) Kurtz, C. B.; Millet, Y. A.; Puurunen, M. K.; Perreault, M.; Charbonneau, M. R.; Isabella, V. M.; Kotula, J. W.; Antipov, E.; Dagon, Y.; Denney, W. S.; Wagner, D. A.; West, K. A.; Degar, A. J.; Brennan, A. M.; Miller, P. F. An engineered E. coli Nissle improves hyperammonemia and survival in mice and shows dose-dependent exposure in healthy humans. *Sci. Transl. Med.* **2019**, *11*, No. eaau7975.

(10) Isabella, V. M.; Ha, B. N.; et al. Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. *Nat. Biotechnol.* **2018**, *36*, 857–864.

(11) Bonde, M. T.; Pedersen, M.; et al. Predictable tuning of protein expression in bacteria. *Nat. Methods* **2016**, *13*, 233–236.

(12) Hawley, D. K.; McClure, W. R. Compilation and analysis of Escherichia coli promoter DNA sequences. *Nucleic Acids Res.* **1983**, *11*, 2237–2255.

(13) Kinney, J. B.; Murugan, A.; et al. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc. Natl. Acad. Sci.* **2010**, *107*, 9158–9163.

(14) Kosuri, S.; Goodman, D. B.; Cambray, G.; Mutalik, V. K.; Gao, Y.; Arkin, A. P.; Endy, D.; Church, G. M. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14024–14029.

(15) Peterman, N.; Lavi-Itzkovitz, A.; et al. Large-scale mapping of sequence-function relations in small regulatory RNAs reveals plasticity and modularity. *Nucleic Acids Res.* **2014**, *42*, 12177–12188.

(16) Sharon, E.; Kalma, Y.; et al. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **2012**, *30*, 521–530.

(17) Sharon, E.; van Dijk, D.; et al. Probing the effect of promoters on noise in gene expression using thousands of designed sequences. *Genome Res.* **2014**, *24*, 1698–1706.

(18) Urtecho, G.; Tripp, A. D.; Insigne, K. D.; Kim, H.; Kosuri, S. Systematic Dissection of Sequence Elements Controlling σ70 Promoters Using a Genomically Encoded Multiplexed Reporter Assay in Escherichia coli. *Biochemistry* **2019**, *58*, 1539–1551.

(19) Petersen, S. D.; Zhang, J.; et al. Modular 5′-UTR hexamers for context-independent tuning of protein expression in eukaryotes. *Nucleic Acids Res.* **2018**, e127.

(20) Noderer, W. L.; Flockhart, R. J.; Bhaduri, A.; Diaz de Arce, A. J.; Zhang, J.; Khavari, P. A.; Wang, C. L. Quantitative analysis of mammalian translation initiation sites by FACS-seq. *Mol. Syst. Biol.* **2014**, *10*, 748.

(21) Johns, N. I.; Gomes, A. L. C. C.; Yim, S. S.; Yang, A.; Blazejewski, T.; Smillie, C. S.; Smith, M. B.; Alm, E. J.; Kosuri, S.; Wang, H. H. Metagenomic mining of regulatory elements enables programmable species-selective gene expression. *Nat. Methods* **2018**, *15*, 323–329.

(22) Crook, N.; Ferreiro, A.; et al. Transcript barcoding illuminates the expression level of synthetic constructs in E. coli Nissle residing in the mammalian gut. *ACS Synth. Biol.* **2020**, *9*, 1010.

(23) Lim, B.; Zimmermann, M.; Barry, N. A.; Goodman, A. L. Engineered Regulatory Systems Modulate Gene Expression of Human Commensals in the Gut. *Cell* **2017**, *169*, 547–558.e15.

(24) Mimee, M.; Tucker, A. C.; Voigt, C. A.; Lu, T. K. Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota. *Cell Syst.* **2015**, *1*, 62–71.

(25) Whitaker, W. R.; Shepherd, E. S.; Sonnenburg, J. L. Tunable Expression Tools Enable Single-Cell Strain Distinction in the Gut Microbiome. *Cell* **2017**, *169*, 538–546.e12.

(26) Poulsen, L. K.; Licht, T. R.; et al. Physiological state of Escherichia coli BJ4 growing in the large intestines of streptomycin-treated mice. *J. Bacteriol.* **1995**, *177*, 5840–5845.

(27) Chang, D.-E.; Smalley, D. J.; et al. Carbon nutrition of {Escherichia} coli in the mouse intestine. *Proc. Natl. Acad. Sci.* **2004**, *101*, 7427–7432.

(28) Hooper, L. V.; Midtvedt, T.; Gordon, J. I. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu. Rev. Nutr.* **2002**, *22*, 283−307.

(29) Alpert, C.; Scheel, J.; Engst, W.; Loh, G.; Blaut, M. Adaptation of protein expression by Escherichia coli in the gastrointestinal tract of gnotobiotic mice. *Environ. Microbiol.* **2009**, *11*, 751−761.

(30) Grozdanov, L.; Raasch, C.; Schulze, J..; Sonnenborn, U.; Gottschalk, G.; Hacker, J.; Dobrindt, U. Analysis of the Genome Structure of the Nonpathogenic Probiotic Escherichia coli Strain Nissle 1917. *J. Bacteriol.* **2004**, *186*, 5432−5441.

(31) Vvedenskaya, I. O.; Zhang, Y.; et al. Massively Systematic Transcript End Readout, "MASTER:" Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields. *Mol. Cell* **2015**, *60*, 953−965.

(32) Yus, E.; Yang, J.-S.; et al. A reporter system coupled with high-throughput sequencing unveils key bacterial transcription and translation determinants. *Nat. Commun.* **2017**, *8*, 368.

(33) Winkelman, J. T.; Vvedenskaya, I. O.; Zhang, Y.; Zhang, Y.; Bird, J. G.; Taylor, D. M.; Gourse, R. L.; Ebright, R. H.; Nickels, B. E. Multiplexed protein-DNA cross-linking: Scrunching in transcription start site selection. *Science* **2016**, *351*, 1090−1093.

(34) Einav, T.; Phillips, R. How the avidity of polymerase binding to the −35/−10 promoter sites affects gene expression. *Proc. Natl. Acad. Sci.* **2019**, *116*, 13340−13345.

(35) Brewster, R. C.; Jones, D. L.; et al. Tuning Promoter Strength through RNA Polymerase Binding Site Design in Escherichia coli. *PLoS Comput. Biol.* **2012**, *8*, No. e1002811.

(36) Massey, F. J., Jr. The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68−78.

(37) Jeschek, M.; Gerngross, D.; et al. Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nat. Commun.* **2016**, *7*, 11163.

(38) Blum-Oehler, G.; Oswald, S.; et al. Development of strain-specific PCR reactions for the detection of the probiotic Escherichia coli strain Nissle 1917 in fecal samples. *Res. Microbiol.* **2003**, *154*, 59−66.

(39) Zainuddin, H. S.; Bai, Y.; Mansell, T. J. CRISPR-based curing and analysis of metabolic burden of cryptic plasmids in Escherichia coli Nissle 1917. *Eng. Life Sci.* **2019**, *19*, 478−485.

(40) Porse, A.; Gumpert, H.; Kubicek-Sutherland, J. Z.; Karami, N.; Adlerberth, I.; Wold, A. E.; Andersson, D. I.; Sommer, M. O. A. Genome Dynamics of Escherichia coli during Antibiotic Treatment: Transfer, Loss, and Persistence of Genetic Elements In situ of the Infant Gut. *Front. Cell. Infect. Microbiol.* **2017**, *7*, 126.

(41) Craggs, T. D. Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev.* **2009**, *38*, 2865.

(42) Buckley, A. M.; Petersen, J.; et al. LOV-based reporters for fluorescence imaging. *Curr. Opin. Chem. Biol.* **2015**, *27*, 39−45.

(43) Pédelacq, J.-D.; Cabantous, S.; et al. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **2006**, *24*, 79−88.

(44) Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N. G.; Palmer, A. E.; Tsien, R. Y. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* **2004**, *22*, 1567−1572.

(45) Yeung, E.; Dy, A. J.; Martin, K. B.; Ng, A. H.; Del Vecchio, D.; Beck, J. L.; Collins, J. J.; Murray, R. M. Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks. *Cell Syst.* **2017**, *5*, 11−24.e12.

(46) Fadrosh, D. W.; Ma, B.; et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2014**, *2*, 6.

(47) Genee, H. J.; Bonde, M. T.; et al. Software-Supported USER Cloning Strategies for Site-Directed Mutagenesis and DNA Assembly. *ACS Synth. Biol.* **2015**, *4*, 342−349.