

## Prototyping of *Bacillus megaterium* genetic elements through automated cell-free characterization and Bayesian modelling

Simon J. Moore<sup>1,2†</sup>, James T. MacDonald<sup>1,2†</sup>, Sarah Wienecke<sup>3</sup>, Nicolas Kyllis<sup>1,2</sup>, Karen M. Polizzi<sup>1</sup>, Rebekka Biedendieck<sup>3\*</sup> and Paul S. Freemont<sup>1,2\*</sup>

<sup>1</sup>Centre for Synthetic Biology and Innovation, South Kensington Campus, London, United Kingdom;

<sup>2</sup>Department of Medicine, South Kensington Campus, London, United Kingdom;

<sup>3</sup>Institute of Microbiology and Braunschweig Integrated Centre of Systems Biology (BRICS), Technische Universität Braunschweig, Braunschweig, Germany

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

\*To whom correspondence should be addressed:

Professor Paul Freemont  
Tel.: +44 (0) 2075945327  
Fax: +44 (0) 207 594 30  
E-mail address: p.freemont@imperial.ac.uk

Dr. Rebekka Biedendieck  
Tel.: +49 531 3915812  
Fax: +49 531 3915854  
E-mail address: r.biedendieck@tu-braunschweig.de

### Keywords

Cell-free, *Bacillus megaterium*, Bayesian statistics, Synthetic biology, automation, factorial experimental design

## Abstract

Automation and factorial experimental design together with cell-free *in vitro* transcription-translation systems offers a new route to the precise characterization of regulatory components. This now presents a new opportunity to illuminate the genetic circuitry from arcane microbial chassis, which are difficult to assess *in vivo*. One such host, *Bacillus megaterium*, is a giant microbe with industrial potential as a producer of recombinant proteins at gram per litre scale. Herein, we establish a *B. megaterium* cell-free platform and characterize a refactored xylose-repressor circuit using acoustic liquid handling robotics to simultaneously monitor 324 reactions *in vitro*. To accurately describe the system, we have applied a Bayesian statistical approach to infer model parameters by simultaneously using information from multiple experimental conditions. These developments now open up a new approach for the rapid and accurate characterization of genetic circuitry using cell-free reactions from unusual microbial cell chassis for bespoke applications.

## Introduction

Cell-free *in vitro* transcription and translation (ITT) systems utilise cell extracts (1–4) or purified components (5) to synthesise proteins encoded by DNA circuits. Recently, a renaissance in cell-free has captured the imagination of synthetic biologists to explore genetic pathway design (6–9), develop on-site viral biosensors (10), produce antibodies in bioreactors (11) and engineer microfluidic biochip devices (12).

One unexplored area of cell-free protein synthesis is the ability to study the genetic regulatory components of non-model microbial hosts directly in a test tube. Arcane microbial hosts have the potential to provide significant advantages over traditional model organisms and the ability to genetically engineer these microbes is likely to become increasingly important (13). In particular, organisms adapted for growth in extreme environments or able to produce valuable niche biomolecules at high yields from inexpensive substrates are of particular interest (14–17). However, these microbial systems are often not very amenable to genetic manipulation and there is a need to expand the available toolbox of characterized genetic parts. Here, we have developed a cell-extract based ITT system for one such arcane microorganism, *Bacillus megaterium*. We have combined this new ITT system with factorial experimental design, high-throughput automation via acoustic liquid handling robotics and Bayesian model parameter inference methods to create a platform to enable the rigorous and rapid characterization of novel genetic parts. A significant advantage of cell-free systems is that the defined reaction conditions permit the accurate characterization of genetic parts using sophisticated modelling techniques (7, 8, 18, 19). However, such approaches have previously generally been applied to *Escherichia coli* gene expression circuits.

*Bacillus megaterium* is a giant microbe with the potential for biotechnology applications (20, 21). Industrially it has been used to produce penicillin G amidase (22),  $\beta$ -amylases and vitamin B<sub>12</sub> (23–25). In comparison to its closely related cousin *Bacillus subtilis*, *B. megaterium* is relatively uncharacterized but

provides major advantages such as stable plasmid maintenance (26), minimal neutral-alkaline protease activity (27) and the ability to metabolise low cost precursors (20, 22, 28). Moreover, *B. megaterium* also offers a strong native Sec-dependent secretion apparatus (29, 30), which is a desirable feature for downstream processing, whilst for recombinant gene expression the xylose-inducible promoter system (31–33) produces recombinant proteins to the gram per litre scale (28, 30). In addition, protein purification and secretion tools are available (33, 34). From these tools, *B. megaterium* has recently been utilised to access challenging proteins (Fe-S, corrinoid enzymes), such as the reductive dehalogenase that were previously insoluble or inactive from using the routine “workhorse” *E. coli* (25, 35–37).

One barrier to the widespread use of *B. megaterium* is a low-efficiency protoplast transformation procedure (38). Our automated ITT and modelling platform enables the rapid validation of gene expression tools. We tested a designed xylose-repressor circuit by combinatorially varying three experimental conditions (DNA, repressor and inducer concentrations), with the automated generation of transfer instructions for an acoustic liquid handling robot to rapidly set up 324 separate *in vitro* reactions. To our knowledge, this demonstrates the first fully automated platform for  $\mu\text{L}$  scale cell-free reaction screens. To model the obtained experimental data, we used a Bayesian statistical inference scheme to rigorously infer unknown ITT kinetic parameters that describe the cell free reaction. The development of a *B. megaterium* ITT system serves as a model for rapidly prototyping novel genetic parts from a relatively uncharacterized host thus providing a basis for future engineering of this model organism. Our platform also provides an integrated high-throughput experimental and modelling framework for characterizing cell extracts from other unusual chassis.

## Methods

### Strains and Plasmids

*B. megaterium* wild-type strain DSM319 was used for preparing the cell-free extract (66), *E. coli* strain BL21 (DE3) Star<sup>TM</sup>-pLysS (LifeTechnologies, Darmstadt, Germany) for the recombinant production of the His<sub>6</sub>-tagged *B. megaterium* DSM319 repressor protein XylR (see Supplementary text) and *E. coli* strain DH10B (Invitrogen, San Diego, USA) for routine cloning. All the plasmids and oligos used in this study are listed in Supplementary Tables S1-3, respectively. For construction of plasmids, please see Supplementary text.

### Growth and recombinant protein production in *B. megaterium*

We utilised an improved fluorescence variant (GFP<sup>+</sup> - F64L/ S65T/ Q80R/ F99S/ M153T/ V163A) of the wild-type GFP (67), herein simply referred to as GFP, to monitor protein production in *B. megaterium* ITT, along with mCherry. Protoplasted *B. megaterium* cells were individually transformed with the corresponding plasmids and cultivated as described before (68).

### *B. megaterium* ITT extract preparation and reaction conditions

A *B. megaterium* DSM319 ITT system was developed based on the *E. coli* method of Sun *et al* (40). Full details and modifications are outlined in the Supplementary text and Table S4. This method yielded between 27-34 mg mL<sup>-1</sup> of cell-free protein extract. All ITT experiments were repeated a minimum of two times, whilst individual assays were collected a triplicate of technical repeats.

### High-throughout automated robotics and experimental set-up

For automation, an Echo<sup>®</sup>525 liquid handler (Labcyte, Inc.) was combined with a factorial experimental design approach using a script to automate transfer volume instructions and well positions (see Supplementary text). Liquid droplets were transferred as multiples of 25 nL to a final volume of 10 µL as technical triplicate repeat. Plates were sealed with Breathe-Easy<sup>®</sup> sealing membrane (Sigma) and briefly centrifuged at 1,000 × *g* for 10 seconds. A CLARIOStar plate reader (BMG Labtech, Germany) was used for cell-free incubations and

fluorescence measurements. Standard measurements were recorded every 10 min for 40 cycles at 30°C with 10 sec of 250 rpm orbital shaking prior to measurement. Depending on sample size, RNA measurements were recorded 30, 60 or 120 seconds. Fluorescence settings are provided in the SI text.

### Bayesian statistical modelling

The cell-free transcription and translation reactions were modelled using a system of ordinary differential equations (ODEs).

Equation 1

$$\begin{aligned} \frac{d[GFP_{mat}]}{dt} &= k_{mat} \cdot [GFP] \\ \frac{d[GFP]}{dt} &= V_{TL} \cdot \frac{\frac{[mRNA]}{K_{RBS}}}{1 + \frac{[mRNA]}{K_{RBS}}} - k_{mat} \cdot [GFP] \\ \frac{d[mRNA]}{dt} &= V_{TX} \cdot \frac{\left(\frac{h \cdot [DNA]}{K_{prom}}\right) \left(\frac{[mRNA_{cap}]}{K_{mRNA_{cap}}}\right)}{\left(1 + \frac{h \cdot [DNA]}{K_{prom}}\right) \left(1 + \frac{[mRNA_{cap}]}{K_{mRNA_{cap}}}\right)} - d_{mRNA} \cdot [mRNA] \\ \frac{d[mRNA_{cap}]}{dt} &= -V_{TX} \cdot \frac{\left(\frac{h \cdot [DNA]}{K_{prom}}\right) \left(\frac{[mRNA_{cap}]}{K_{mRNA_{cap}}}\right)}{\left(1 + \frac{h \cdot [DNA]}{K_{prom}}\right) \left(1 + \frac{[mRNA_{cap}]}{K_{mRNA_{cap}}}\right)} - d_{mRNA_{cap}} \cdot [mRNA_{cap}] \end{aligned}$$

where,

$$h = \frac{1}{1 + \left( \frac{[XylR] \cdot \frac{1}{1 + \left( \frac{[xylose]}{K_{xylose}} \right)^{n_{xylose}}}}{K_{XylR}} \right)^{n_{XylR}}}$$

The model consisted of 5 chemical reactions, 4 species (Supplementary Table S5), and 15 parameters (Supplementary Table S6). The regeneration of ATP, GTP, CTP and UTP from the secondary energy source was not explicitly modelled and these were all lumped into a single species, mRNA<sup>cap</sup> (mRNA capacity). It was assumed that NTPs were also consumed by non-productive side-reactions and this was modelled as a first-order degradation reaction of mRNA<sup>cap</sup>. The transcription rate was modelled assuming multi-substrate

Michaelis-Menten kinetics with random order binding of DNA template and mRNA<sup>cap</sup>. For the xylose inducible promoter, the fraction of free DNA was approximated using the Hill equation. The mRNA degradation reaction was assumed to be a first order reaction. Translation rate was modelled using Michaelis-Menten kinetics with mRNA as the only substrate. Amino acids were assumed to be in excess so were not explicitly considered. As in previous models of cell-free reactions, consumption of ATP and GTP during translation was not considered (18). The translation reaction produced non-fluorescent immature GFP and this was assumed to mature in a first-order reaction into the detectable fluorescent species, GFP<sub>mat</sub>. The full system of ODEs is shown in the equation 1.

Parameters and initial values (for [mRNA<sub>cap</sub>]) were inferred using Markov chain Monte Carlo (MCMC) implemented in a C++ program using Boost.odeint library to numerically integrate the system of ODEs. A multivariate normal distribution was used as the proposal distribution during MCMC and moves were accepted following the Metropolis criterion. A uniform or a multivariate Gaussian prior distribution was used for all parameters. Experimental errors were assumed to be independent and normally distributed. The log-likelihood function was therefore defined as the sum of logs of the normal probability density function where the means and standard deviations were taken from the experimentally derived values for each data point. Model parameters were inferred simultaneously from multiple experimental datasets by generating multiple trajectories from the same system of ODEs and evaluating the log-likelihood function. Initial values for all species other than [mRNA<sub>cap</sub>] were set 0. Although initial NTP concentrations in the reaction are known, the overall NTP capacity of the system was unknown as the efficiency of NTP regeneration from the secondary fuel source was unknown. For this reason, the initial value of [mRNA<sub>cap</sub>] was inferred. In most experiments, only the mature GFP concentration was measured. Where the mRNA concentration was experimentally determined, this was also included in the log-likelihood function. Source code is available from [https://github.com/jmacdona/B\\_megaterium\\_modelling](https://github.com/jmacdona/B_megaterium_modelling).

## Results

### Isolation of a strong constitutive promoter for *B. megaterium* ITT

To achieve quantitative fluorescence measurements in ITT reactions, a strong promoter system is required. To initially prototype a *B. megaterium* ITT system without regulatory components (e.g. repressor/activator), we first isolated and characterized a strong constitutive  $\sigma^A$  promoter from *B. megaterium*. Previously, a strong  $\sigma^A$  promoter was isolated from the pyruvate dehydrogenase (*pdh*) operon in *Bacillus subtilis* (39). For *B. megaterium*, we isolated the equivalent promoter from the *pdhABCD* (*bmd\_1326-1329*) gene cluster by PCR along with the RBS of *pdhA* (*bmd\_1326*) and used this to control the expression of *gfp* and mCherry. To initially validate the performance of gene expression *in vivo*, the optimised xylose-inducible promoter plasmid GFP variants (30) were used for comparison and tested under plate reader conditions. As expected, the growth rate of the empty vector control and non-induced xylose promoter recombinant strains was faster than the PDH or xylose-promoter strains, with all strains reaching the stationary phase after approximately 8 hours of growth (Supplementary Figure S1A). As a comparison between plasmid strains, normalised activity (RFU OD<sub>600</sub><sup>-1</sup> hr<sup>-1</sup>) for the PDH promoter was 1.4-fold higher than xylose-inducible promoter plasmid strains (Supplementary Figure S1B). For SDS-PAGE analysis, intracellular proteins were analysed from 100 mL batch-flask cultures for a qualitative assessment of GFP and mCherry production from xylose and PDH promoter plasmid strains (Figure 1A).

### *B. megaterium* ITT activity

To test the xylose and PDH promoter systems *in vitro*, a *B. megaterium* cell-free extract was developed by using an *E. coli* S30 protocol for guidance (40). Extracts were prepared through four stages of preparation – growth, cell-disruption, heated incubation (“run-off” period) and dialysis. Following the S30 method with minor modifications (see Supplementary text) the cells were lysed with sonication followed by a 1.5 hour “run-off” period at 37°C and 3 hour dialysis. From this, significant levels of GFP synthesis were detected with 10 mg mL<sup>-1</sup> of protein extract incubated with 10 nM of plasmid DNA, 1.5 mM amino

acids and a 3-PGA energy buffer system. From this initial test, 1.4  $\mu\text{M}$  of GFP was produced within 3 hours at 30°C. To optimise the method further, the sonication, “run-off” and dialysis stages of extract preparation were individually tested. Maximum activity was observed with an energy input of 154  $\text{kJ mL}^{-1}$  for sonication and a “run-off” period of 1 hour at 37°C. These two modifications increased GFP batch synthesis up to 2.8  $\mu\text{M}$  (Supplementary Figure S2). In addition, we also tested alternative incubation temperatures and time periods, but found the 1 hour incubation at 37°C to be optimal. Further details of optimisation are outlined in the Supplementary text. As a third preparation step, dialysis is used to remove small molecules, however, it is not always essential for the preparation of all *E. coli* cell-free extracts (41, 42). For *B. megaterium* DSM319, dialysis did not significantly affect ITT activity and therefore we consider it is not essential for this specific extract preparation. With further optimisation of Mg-glutamate (2 mM) and K-glutamate buffer (50 mM) conditions with 3-PGA as the energy source, maximum levels of 4.96  $\mu\text{M}$  GFP (0.14  $\text{mg mL}^{-1}$ ) are produced from 10 nM pRBBm258 PDH constitutive plasmid, with a band corresponding to GFP observable on SDS-PAGE (Figure 1B). This represents a 3.5-fold improvement from the initial test condition. With the xylose-inducible plasmid lacking the repressor XylR (pKMMBm5), maximum GFP batch synthesis of 2.48  $\mu\text{M}$  is observed (Figure 2C). Maximal levels of protein production were saturated at approximately 5 nM of plasmid DNA for both promoter systems (Figure 1C). For comparison, fully optimised *E. coli* ITT systems can produce up to 80  $\mu\text{M}$  (2.3  $\text{mg mL}^{-1}$ ) of GFP in batch synthesis (1, 43). We also tested mCherry production *in vitro*, but production was limited to 0.17  $\mu\text{M}$  under optimised conditions (Supplementary Figure S3 and Figure S4). This could be due to a high rare codon frequency for the *mCherry* gene (see Supplementary text).

### **Real-time mRNA and protein monitoring of the xylose-repressor system *in vitro***

The xylose-inducible promoter is derived from the native *B. megaterium* *xylABT* operon, which encodes the xylose isomerase (*xylA*), xylulokinase (*xylB*) and permease (*xylT*) genes (31, 44). Expression of the *xyl* operon is repressed from

the divergently transcribed repressor XylR. In the absence of xylose, XylR can bind to two operator sequences ( $O_L/O_R$ ) that overlap by 4 bp providing tight transcriptional gene expression control. Xylose binding to XylR releases the protein to allow transcription to start with gene expression induced up to 350-fold in *B. megaterium* DSM319, thus acting as a sensing mechanism for the cell to switch metabolism and utilise xylose as a carbon source (31). To study this system in cell-free reactions, we utilised a xylose reporter plasmid that lacks the XylR regulator (pKMMBm5). We decided to utilise the standard 3-PGA buffer for energy regeneration. Previously, this plasmid was shown to constitutively produce GFP at similar levels in wild-type *B. megaterium* DSM319 and WH325  $\Delta xylR$  mutant strains (45). This demonstrated that under standard growth conditions, the levels of XylR produced from the genomic copy of the gene were insufficient to repress  $P_{xyl}$  controlled GFP expression from the medium copy plasmid pKMMBm5 (45).

In order to characterize the kinetics of the xylose promoter in the absence of the repressor XylR, we wanted to use our cell-free system to simultaneously monitor transcription and translation *in vitro* (Figure 2). To monitor mRNA expression, we modified the 3' untranslated region (UTR) to include a malachite green aptamer (MGapt) for real-time fluorescence detection of mRNA synthesis as previously described (46–51) (Supplementary Figure S5). mRNA degradation was estimated by measuring fluorescence for 2 hours and the order of mRNA degradation followed a single-phase exponential decay rate (Supplementary Figure S6 and Figure S7). A half-life estimated at 15.6 min is similar (18 min) to previously observed *E. coli* cell-free measurements (7).

We next tested a range of DNA concentrations in cell-free reactions with simultaneous measurement of MGapt and GFP fluorescence. With increasing DNA concentration, the rate of MGapt fluorescence rises rapidly within 6 min up to a peak concentration of approximately 235 nM of GFP-MGapt transcript at 25-27 min after the start of incubation. Thereafter, the signal decays suggesting that either nucleotide or energy regeneration becomes rate limiting for further mRNA synthesis. Spiking the reaction with a purified GFP-MGapt transcript compared to the equivalent plasmid control, shows that mRNA levels are more

stable throughout the assay when expressed from plasmid DNA, suggesting that continuous transcription provides mRNA substrate for translation throughout the reaction time-period (Supplementary Figure S8).

### **Bayesian modelling of the xylose-repressor system**

To characterize the behaviour of the xylose-repressor system in more detail, we chose to model the system using ordinary differential equations (ODEs) with 4 species and 15 parameters (Supplementary Tables S1 and S2). These parameters were simultaneously inferred from multiple experimental data sets using a Bayesian Markov chain Monte Carlo (MCMC) approach (see Methods for details). A posterior distribution of the model parameters was inferred from the GFP and mRNA time course data sets with the DNA template (pKMMBm5-MGapt) titrated at a range of different concentrations. These parameters were simultaneously inferred from multiple experimental datasets and chemical species (mature GFP and mRNA) using a combined log-likelihood function (Supplementary equation S1). Multiple MCMC runs were generated from different parameter starting points and convergence was confirmed for all parameters using the Gelman-Rubin diagnostic(52). Simulated trajectories from the model closely matched the experimental time-course data for both GFP and mRNA concentration at different DNA template concentrations (Figure 3 and Supplementary Figure S9 and Figure S10), whilst the posterior distribution was found to be unimodal (Supplementary Figure S11). The reaction machinery appeared to be saturated at around 5 nM template DNA with no further increase in total protein produced, however, mRNA expression continued increasing with higher DNA concentrations. This implied that the ribosomes were saturated with mRNAs but the host RNA polymerase was not saturated with DNA template.

### **Liquid handling robotics of cell-free reactions for augmented full factorial inducible promoter characterization**

A second, larger-scale experiment was carried out on the XylR repressible promoter to separately characterize its kinetics in cell-free reactions. This experiment was designed to simultaneously monitor GFP translation with 108 unique conditions in triplicate (324 total reactions) made up of an augmented full factorial combination of different concentrations of purified recombinant XylR

repressor protein (0-1000 nM; Supplementary Figure S15), D-xylose (0-1000  $\mu$ M), and pKMMBm5-MGapt DNA template (1, 2 and 5 nM) on a 384 well microtiter plate using an acoustic liquid handling robot (Figure 4A).

In the absence of D-xylose and XylR<sup>His</sup>, GFP is constitutively produced and slightly elevated by the addition of 0.5 mM D-xylose. We also verified that the *B. megaterium* ITT system is not inhibited by high D-xylose concentrations (Supplementary Figure S16). When XylR<sup>His</sup> is titrated in the absence of D-xylose, a repression effect on GFP production is observed, as expected, but this effect is variable with DNA concentration (Figure 4B). Following the titration of D-xylose in the presence of 0.1 to 1.0  $\mu$ M XylR<sup>His</sup>, GFP expression is recovered, which is consistent with the release of the XylR<sup>His</sup> repressor from the promoter region upon binding of D-xylose (Figure 4B).

Kinetic data from the first set of experiments was incorporated into the model by fitting a multivariate Gaussian distribution to the posterior distribution of the first experiment (Supplementary Figure S11). This was then used as the prior distribution for inferring model parameters using data from the second set of experiments with the same MCMC method (Fig 4C, Supplementary Tables S1-2 and Figure S17). Over this diverse range of conditions, both the simulated final GFP concentrations (Figure 4B) and the time-course trajectories were found to fit the experimental data remarkably well (Supplementary Figure S12-14). The maximal translation rate was inferred to be approximately 27-47 nM (GFP) min<sup>-1</sup>, which is similar to previously determined figures for *E. coli* cell-free reactions (53), whilst the maximal transcription rate was inferred at between 310-380 nM (mRNA) min<sup>-1</sup>. Previously, the maturation rate of GFP+ (a synonym for GFPuv3) was measured to be 0.066 min<sup>-1</sup> (54). We independently infer the maturation rate to between 0.047-0.057 min<sup>-1</sup>. The difference may reflect inaccuracies in the model or a lower oxygen concentration could limit maturation in the reaction mixture. Moreover, the XylR repressor was found to cooperatively bind to its operator sequence (Hill coefficient 2.1-2.2), while xylose appears to cooperatively bind XylR (Hill coefficient 1.4-1.5).

### **Succinate as an alternative substrate for energy regeneration**

For future scaled-up cell-free protein production in *B. megaterium*, a minimal, low-cost energy system for extract preparation is required (55). For *B. megaterium* cell-free we initially investigated the importance of each individual component (42) in the standard energy buffer for *B. megaterium* ITT activity. Firstly, in the absence of NTPs, neither mRNA nor GFP is synthesised. Interestingly, if tRNA's, NAD<sup>+</sup>, coenzyme A and folinic acid were removed, this did not significantly affect ITT activity in *B. megaterium* extracts (Supplementary Figure S18). In addition, the removal of 3-PGA only led to a 75% reduction in GFP synthesis, whilst spermidine and cAMP were considered non-essential, but beneficial for activity. Next, we screened potential energy substrates including glucose, maltose, PEP, pyruvate, succinate and glutamate to replace 3-PGA as a sole energy source. The residual activity from the absence of 3-PGA is likely to be due to K-glutamate or other possible amino acid catabolic pathways. Glutamate has previously been shown as a low-cost energy source in *E. coli*, where it enters the Krebs cycle through  $\alpha$ -ketoglutarate (42). For alternative energy sources in *B. megaterium*, we observed strong activity for pyruvate and succinate, however, no activity was observed with glucose, maltose or phosphoenolpyruvate (Supplementary Figure S19). Focusing on succinate, we observed a strong response in GFP production in response to an increasing concentration of the substrate, which peaked in activity at 50 mM Na-succinate (Supplementary Figure S20). This effect was also pH-dependent (Supplementary Figure S21) with maximal activity of GFP synthesis observed at pH 8.2, although RNA synthesis appeared to optimal at lower ranges around pH 7.0.

## Discussion

There is a growing interest in exploring arcane microbial platforms for future cellular factories (13). However, the advantages of using such hosts are often counter balanced by a lack of genetic tools, poor genetic tractability or fastidious growth. Generally to engineer bespoke hosts for a specific industrial process, classical strain evolution by mutagenesis is utilised to optimise production of fine chemicals, amino acids and vitamins (23).

Recently, cell-free ITT platforms have been established for prototyping genetic parts and testing metabolic pathway designs (1, 6, 49). Here we have utilised the advantages of cell-free ITT to establish a *B. megaterium* cell-free platform for prototyping genetic parts and pathway designs for a host that is difficult to engineer but has the potential for industrial scale catalysis of low-cost building blocks (e.g. crude glycerol, sugar beet) into value-added products (20, 56). Moreover, there is a general opportunity in synthetic biology to utilise alternative chassis for bespoke applications, for example accessing challenging recombinant enzymes (25, 36).

To provide a robust assessment of our *B. megaterium* cell-free platform, we also describe a factorial experimental design and modelling approach to delineate the kinetic behaviour cell free reactions through automated experimental testing and Bayesian statistical parameter inference. To our knowledge, our use of automation and Bayesian modelling are the first time this has been applied to cell-free prototyping thereby generating complex datasets for refined and accurate analysis allowing the design of robust synthetic genetic circuits using Bayesian design methods (57). In order to test our cell free extract preparations we isolated and characterized a *B. megaterium* specific  $\sigma^A$  constitutive promoter along with the xylose-inducible promoter. In comparison to a fully optimised *E. coli* cell-free reaction using  $\sigma^{70}$  constitutive promoter or T7 promoter (43), our initial *B. megaterium* extracts produce approximately 10-fold less GFP using the  $\sigma^A$  promoter. Significantly, the GFP production capacity of *B. megaterium in vivo* is comparable (gram per litre scale) to optimised *E. coli* strains (58, 59). This suggests that the current productivity of our cell-free system is below the true

potential for this host. From our experimental and modelling data, mRNA expression is not rate limiting but the translation apparatus appears to be saturated above 5 nM of plasmid DNA. To optimise gene expression in cell-free systems, current strategies have focused optimising energy regeneration (60), mRNA expression (61) or ribosome enrichment (62, 63). Therefore for the moment at least, it would seem that either the aminoacyl-tRNA synthetases or ribosome activity is rate limiting for *B. megaterium* ITT and therefore a key starting point for further investigation. Another unknown factor is the availability or stability of the tRNA derived from the host extract. However for the purpose of characterizing our *B. megaterium* ITT system the expression levels of our test plasmids give strong and discernable signals.

Efficient energy regeneration is an essential component of a cell-free extract to replenish ATP and nucleotide supplies for protein synthesis from DNA (60). To our knowledge we have established for the first time that succinate can act as a substrate for energy regeneration in cell free extracts (Figure 5). Succinate dehydrogenase in bacteria is found as a three-component (ShdABC) membrane bound complex and is unique since it is the only Krebs cycle enzyme step that also participates in the electron transport chain as part of respiratory complex II. The *Bacillus subtilis* SdhABC complex has previously been characterized by electron paramagnetic resonance and shown to catalyse a one-step oxidation of succinate to fumarate, using an electron tunnelling mechanism (64). It has been proposed that inverted vesicles derived from the inner membrane of *E. coli* form during cell-free extract preparation and contain several active members of the electron transport chain and ATP synthase to regenerate energy through oxidative phosphorylation (65). Our observed ITT activity with 3-PGA, glutamate and pyruvate suggest that inverted vesicles are also likely to form in *B. megaterium* cell-free extracts. The use of succinate provides a potential low-cost route for scaling-up *B. megaterium* cell-free systems in the future, which is an essential requirement of cost-efficient production methods (55).

To begin studying genetic circuits in *B. megaterium* cell-free we investigated the xylose-repressor system, which switches on gene expression in response to xylose. Aside from providing high-inducible levels of expression in the presence

of D-xylose, this XylR based regulatory system provides strong *trans* regulatory control in the absence of xylose (31). Interestingly, by reconstituting this system *in vitro* using purified XylR, we also observe tight regulatory control of GFP expression, with our Bayesian model estimating the binding  $K_D$  for XylR and xylose at 19.6 nM and 19.3  $\mu$ M, respectively. This demonstrates a simple reconstitution and modelling characterization of an inducible regulatory element in *B. megaterium* cell-free extracts and thus provides a basis for future prototyping of *B. megaterium* specific regulatory elements. These rigorously characterized elements can then be rationally composed into more complex circuits (7, 47, 57).

Cell-free synthetic biology is a growing area of research for rapid part and pathways design prototyping to large industrial scale protein manufacturing (55). Our demonstration of *B. megaterium* as new host for cell-free synthesis using automated nanolitre scale liquid robotics combined with a rigorous factorial experimental design and Bayesian modelling approach opens up the possibility of systematically exploring less-well studied chassis for cell-free synthetic biology applications.

## **Funding**

This work was supported by the EPSRC [EP/K038648/1 to S.M., EP/K034359/1 to J.M.]; BBSRC Case Studentship [DTG BB/F017324/1 to N.K.] and the German Research Foundation (DFG) priority program [SPP1617 to S.W. and R.B.].

### **Acknowledgements**

SW and RB would like to thank Dieter Jahn for support and discussion.

## References

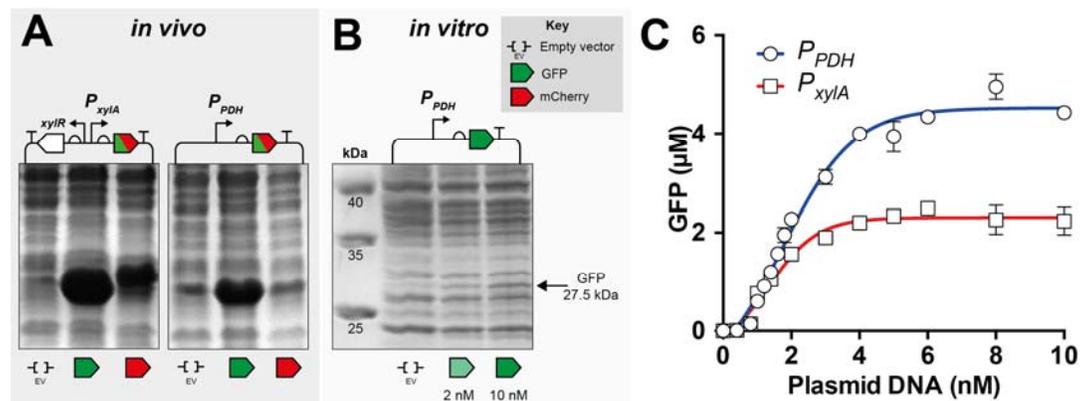
1. Garamella J, Marshall R, Rustad M, Noireaux V (2016) The all E. coli TX-TL Toolbox 2.0: a platform for cell-free synthetic biology. *ACS Synth Biol*. doi:10.1021/acssynbio.5b00296.
2. Siegal-Gaskins D, Tuza ZA, Kim J, Noireaux V, Murray RM (2014) Gene circuit performance characterization and resource usage in a cell-free “breadboard”. *ACS Synth Biol* 3(6):416–25.
3. Jewett MC, Swartz JR (2004) Mimicking the Escherichia coli Cytoplasmic Environment Activates Long-Lived and Efficient Cell-Free Protein Synthesis. *Biotechnol Bioeng* 86(1):19–26.
4. Swartz JR (2009) Universal cell-free protein synthesis. *Nat Biotechnol* 27(8):731–732.
5. Kuruma Y, Ueda T (2015) The PURE system for the cell-free synthesis of membrane proteins. *Nat Protoc* 10(9):1328–1344.
6. Takahashi MK, et al. (2015) Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription-Translation (TX-TL) Systems. *ACS Synth Biol* 4(5):503–515.
7. Siegal-Gaskins D, Tuza ZA, Kim J, Noireaux V, Murray RM (2014) Gene circuit performance characterization and resource usage in a cell-free “breadboard.” *ACS Synth Biol* 3(6):416–425.
8. Geertz M, Shore D, Maerkl SJ (2012) Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc Natl Acad Sci* 109(41):16540–16545.
9. Niederholtmeyer H, et al. (2015) Rapid cell-free forward engineering of novel genetic ring oscillators. *Elife* 4(October):e09771.
10. Pardee K, et al. (2014) Paper-based synthetic gene networks. *Cell* 159(4):940–954.
11. Groff D, et al. (2014) Engineering toward a bacterial “endoplasmic reticulum” for the rapid expression of immunoglobulin proteins. (June):671–678.
12. Niederholtmeyer H, Stepanova V, Maerkl SJ (2013) Implementation of cell-free biological networks at steady state. *Proc Natl Acad Sci U S A* 110(40):15985–90.
13. Eisenstein M (2016) Living factories of the future. *Nature* 531(7594):401–403.
14. Florea M, et al. (2016) Engineering control of bacterial cellulose production using a genetic toolkit and a new cellulose-producing strain. *Proc Natl Acad Sci U S A* 113(24):E3431–40.
15. Phelan RM, Sekurova ON, Keasling JD, Zotchev SB (2015) Engineering terpene biosynthesis in streptomyces for production of the advanced biofuel precursor bisabolene. *ACS Synth Biol* 4(4):393–399.
16. He J, Van Treeck B, Nguyen HB, Melançon CE (2016) Development of an Unnatural Amino Acid Incorporation System in the Actinobacterial Natural Product Producer *Streptomyces venezuelae* ATCC 15439. *ACS Synth Biol* 5(2):125–132.
17. Florea M, Reeve B, Abbott J, Freemont PS, Ellis T (2016) Genome sequence and plasmid transformation of the model high-yield bacterial cellulose producer *Gluconacetobacter hansenii* ATCC 53582. *Sci Rep* 6:23635.
18. Tuza ZA, Singhal V, Kim J, Murray RM (2013) An in silico modeling

- toolbox for rapid prototyping of circuits in a biomolecular “breadboard” system. *Proceedings of the IEEE Conference on Decision and Control* (IEEE), pp 1404–1410.
19. Niederholtmeyer H, Xu L, Maerkl SJ (2013) Real-time mRNA measurement during an in vitro transcription and translation reaction using binary probes. *ACS Synth Biol* 2(8):411–7.
  20. Korneli C, David F, Biedendieck R, Jahn D, Wittmann C (2013) Getting the big beast to work-Systems biotechnology of *Bacillus megaterium* for novel high-value proteins. *J Biotechnol* 163(2):87–96.
  21. Bunk B, et al. (2010) A short story about a big magic bug. *Bioeng Bugs* 1(2):85–91.
  22. Yang Y, et al. (2006) High yield recombinant penicillin G amidase production and export into the growth medium using *Bacillus megaterium*. *Microb Cell Fact* 5:36.
  23. Martens JH, Barg H, Warren MJ, Jahn D (2002) Microbial production of vitamin B12. *Appl Microbiol Biotechnol* 58(3):275–285.
  24. Raux E, Lanois A, Warren MJ, Rambach A, Thermes C (1998) Cobalamin (vitamin B12) biosynthesis: identification and characterization of a *Bacillus megaterium* cobl operon. *Biochem J* 335 ( Pt 1:159–166.
  25. Moore SJ, et al. (2013) Elucidation of the anaerobic pathway for the corrin component of cobalamin (vitamin B12). *Proc Natl Acad Sci U S A* 110(37):14906–14911.
  26. Yang G, et al. (2008) Expression of recombinant *Clostridium difficile* toxin A and B in *Bacillus megaterium*. *BMC Microbiol* 8:192.
  27. Wittchen KD, Meinhardt F (1995) Inactivation of the major extracellular protease from *Bacillus megaterium* DSM319 by gene replacement. *Appl Microbiol Biotechnol* 42(6):871–877.
  28. Korneli C, Biedendieck R, David F, Jahn D, Wittmann C (2013) High yield production of extracellular recombinant levansucrase by *Bacillus megaterium*. *Appl Microbiol Biotechnol* 97(8):3343–53.
  29. Malten M, Hollmann R, Deckwer WD, Jahn D (2005) Production and secretion of recombinant *Leuconostoc mesenteroides* dextransucrase DsrS in *Bacillus megaterium*. *Biotechnol Bioeng* 89(2):206–218.
  30. Stammen S, et al. (2010) High-yield intra- And extracellular protein production using *Bacillus megaterium*. *Appl Environ Microbiol* 76(12):4037–46.
  31. Rygus T, Hillen W (1991) Inducible high-level expression of heterologous genes in *Bacillus megaterium* using the regulatory elements of the xylose-utilization operon. *Appl Microbiol Biotechnol* 35(5):594–9.
  32. Rygus T, Hillen W (1992) Catabolite repression of the xyl operon in *Bacillus megaterium*. *J Bacteriol* 174(9):3049–3055.
  33. Biedendieck R, Yang Y, Deckwer W-DD, Malten M, Jahn D (2007) Plasmid system for the intracellular production and purification of affinity-tagged proteins in *Bacillus megaterium*. *Biotechnol Bioeng* 96(3):525–537.
  34. Malten M, et al. (2006) A *Bacillus megaterium* plasmid system for the production, export, and one-step purification of affinity-tagged heterologous levansucrase from growth medium. *Appl Environ Microbiol* 72(2):1677–9.
  35. Moore SJ, et al. (2013) Characterization of the enzyme CbiH60 involved in anaerobic ring contraction of the cobalamin (vitamin B12) biosynthetic

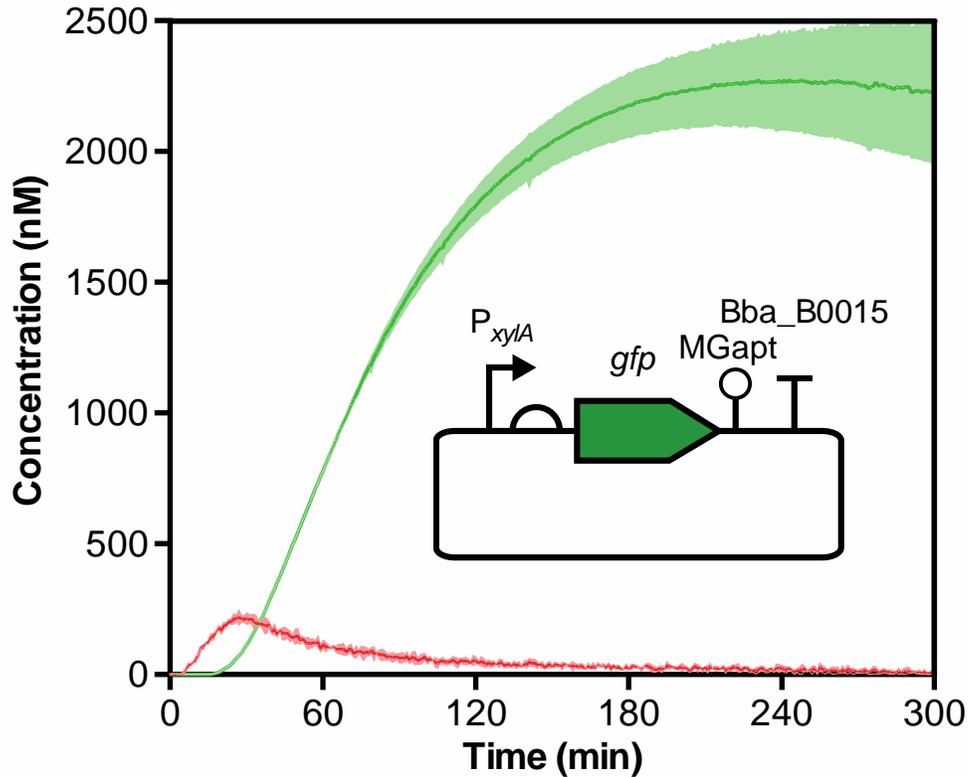
- pathway. *J Biol Chem* 288(1):297–305.
36. Payne KAP, et al. (2015) Reductive dehalogenase structure suggests a mechanism for B12-dependent dehalogenation. *Nature* 517(7535):513–6.
  37. Payne KAP, et al. (2015) Epoxyqueuosine reductase structure suggests a mechanism for cobalamin-dependent tRNA modification. *J Biol Chem* 290(46):27572–27581.
  38. Brey RN, Banner CD, Wolf JB (1986) Cloning of multiple genes involved with cobalamin (Vitamin B12) biosynthesis in *Bacillus megaterium*. *J Bacteriol* 167(2):623–30.
  39. Tojo S, Kumamoto K, Hirooka K, Fujita Y (2010) Heavy involvement of stringent transcription control depending on the adenine or guanine species of the transcription initiation site in glucose and pyruvate metabolism in *Bacillus subtilis*. *J Bacteriol* 192(6):1573–85.
  40. Sun ZZ, et al. (2013) Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J Vis Exp* 50762(79):e50762.
  41. Kwon Y-C, Jewett MC (2015) High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci Rep* 5:8663.
  42. Cai Q, et al. (2015) A simplified and robust protocol for immunoglobulin expression in *Escherichia coli* cell-free protein synthesis systems. *Biotechnol Prog* 31(3):823–831.
  43. Caschera F, Noireaux V (2014) Synthesis of 2.3 mg/ml of protein with an all *Escherichia coli* cell-free transcription-translation system. *Biochimie* 99(1):162–168.
  44. Rygus T, et al. (1991) Molecular cloning, structure, promoters and regulatory elements for transcription of the *Bacillus licheniformis* encoded regulon for xylose utilization. *Arch Microbiol* 155(6):535–42.
  45. Münch KM, et al. (2015) Polar fixation of plasmids during recombinant protein production in *Bacillus megaterium* results in population heterogeneity. *Appl Environ Microbiol* 81(17):5976–5986.
  46. Da Costa JB, Andreiev AI, Dieckmann T (2013) Thermodynamics and kinetics of adaptive binding in the malachite green RNA aptamer. *Biochemistry* 52(38):6575–6583.
  47. Kim J, Khetarpal I, Sen S, Murray RM (2014) Synthetic circuit for exact adaptation and fold-change detection. *Nucleic Acids Res*. doi:10.1093/nar/gku233.
  48. Franco E, et al. (2011) PNAS Plus: Timing molecular motion and production with a synthetic transcriptional clock. *Proc Natl Acad Sci* 108(40):E784–E793.
  49. Takahashi MK, et al. (2015) Characterizing and prototyping genetic networks with cell-free transcription-translation reactions. *Methods* 86:60–72.
  50. Takahashi MK, et al. (2015) Characterizing and prototyping genetic networks with cell-free transcription-translation reactions. *Methods* 86:60–72.
  51. Wittmann A, Suess B (2012) Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett* 586(15):2076–2083.
  52. Gelman A, Rubin DB (1992) Inference from Iterative Simulation Using Multiple Sequences. *Stat Sci* 7(4):457–472.

53. Karzbrun E, Shin J, Bar-Ziv RH, Noireaux V (2011) Coarse-grained dynamics of protein synthesis in a cell-free system. *Phys Rev Lett* 106(4):048104.
54. Iizuka R, Yamagishi-Shirasaki M, Funatsu T (2011) Kinetic study of de novo chromophore maturation of fluorescent proteins. *Anal Biochem* 414(2):173–8.
55. Zawada JF, et al. (2011) Microscale to manufacturing scale-up of cell-free cytokine production—a new approach for shortening protein production development timelines. *Biotechnol Bioeng* 108(7):1570–1578.
56. Vary PS, et al. (2007) *Bacillus megaterium*—from simple soil bacterium to industrial protein production host. *Appl Microbiol Biotechnol* 76(5):957–967.
57. Barnes CP, Silk D, Sheng X, Stumpf MPH (2011) Bayesian design of synthetic biological systems. *Proc Natl Acad Sci* 108(37):15190–15195.
58. Dürrschmid K, et al. (2008) Monitoring of transcriptome and proteome profiles to investigate the cellular response of *E. coli* towards recombinant protein expression under defined chemostat conditions. *J Biotechnol* 135(1):34–44.
59. Kim H-J, Kwon YD, Lee SY, Kim P (2012) An engineered *Escherichia coli* having a high intracellular level of ATP and enhanced recombinant protein production. *Appl Microbiol Biotechnol* 94(4):1079–86.
60. Voloshin AM, Swartz JR (2005) Efficient and scalable method for scaling up cell free protein synthesis in batch mode. *Biotechnol Bioeng* 91(4):516–521.
61. Shin J, Noireaux V (2010) Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J Biol Eng* 4(1):8.
62. Liu Y, Fritz BR, Anderson MJ, Schoborg J a, Jewett MC (2014) Characterizing and alleviating substrate limitations for improved in vitro ribosome construction. *ACS Synth Biol*. doi:10.1021/sb5002467.
63. Fritz BR, Jewett MC (2014) The impact of transcriptional tuning on in vitro integrated rRNA transcription and ribosome construction. *Nucleic Acids Res* 42(10):6774–6785.
64. Hägerhäll C, Aasa R, von Wachenfeldt C, Hederstedt L (1992) Two hemes in *Bacillus subtilis* succinate:menaquinone oxidoreductase (complex II). *Biochemistry* 31(32):7411–21.
65. Jewett MC, Calhoun K a, Voloshin A, Wu JJ, Swartz JR (2008) An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol Syst Biol* 4(220):220.
66. Eppinger M, et al. (2011) Genome sequences of the biotechnologically important *Bacillus megaterium* strains QM B1551 and DSM319. *J Bacteriol* 193(16):4199–213.
67. Scholz O, Thiel A, Hillen W, Niederweis M (2000) Quantitative analysis of gene expression with an improved green fluorescent protein. *Biotechnol Bioeng* 1570:1565–1570.
68. Biedendieck R, et al. (2011) Systems biology of recombinant protein production using *Bacillus megaterium*. *Methods Enzymol* 500:165–195.

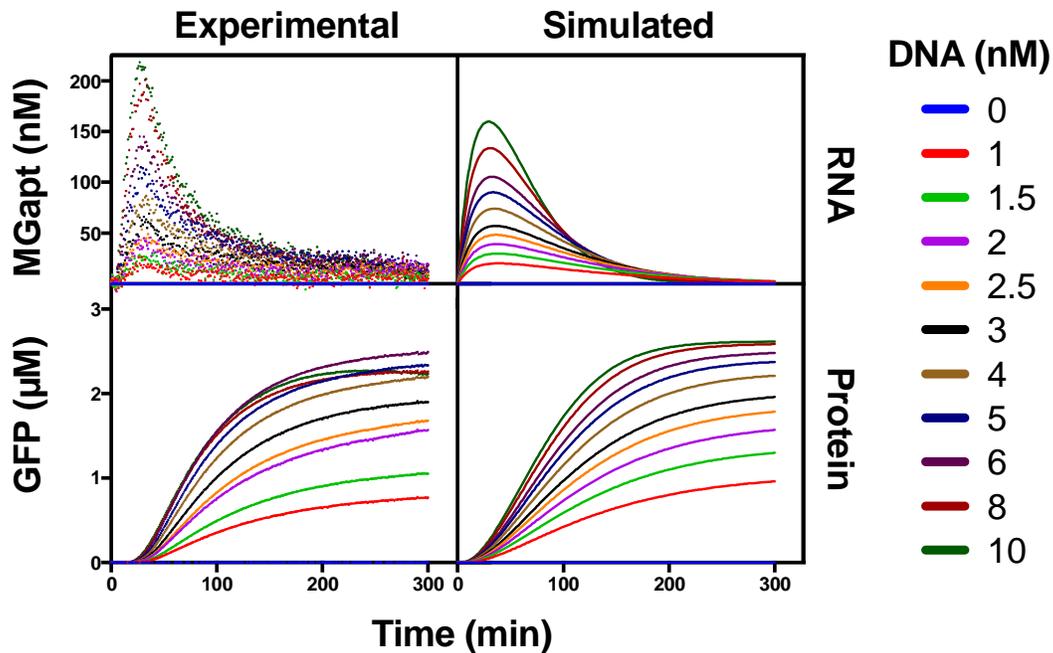
## Figures



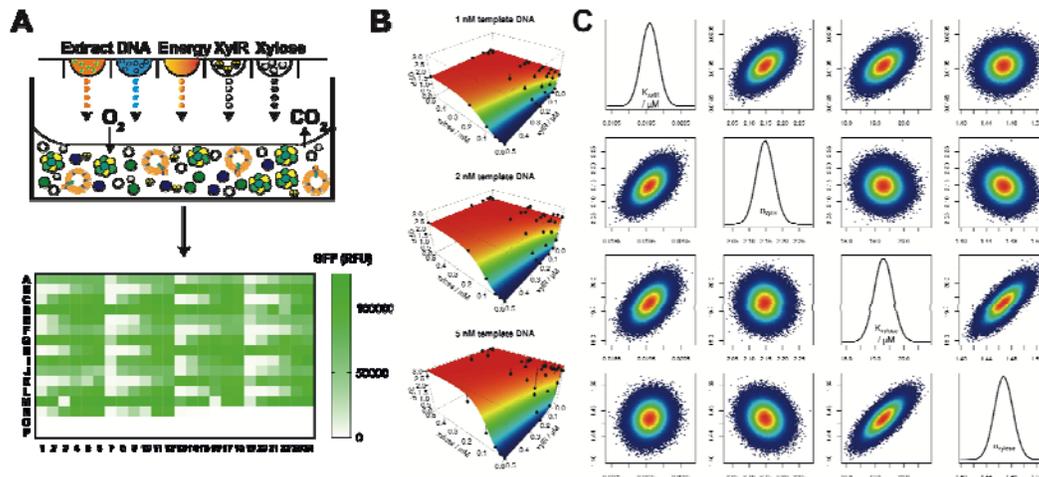
**Figure 1.** Comparison of the promoter systems in *B. megaterium* cells or ITT. **(A)** Recombinant *B. megaterium* DSM319 carrying the plasmids pSSBm85 ( $P_{xylA}$ -*gfp*), pRBBm258 ( $P_{PDH}$ -*gfp*), pSWBm19 ( $P_{xylA}$ -*mCherry*) and pRBBm267 ( $P_{PDH}$ -*mCherry*), respectively, were aerobically cultivated in LB medium. Intracellular proteins were analysed by SDS-PAGE. **(B)** SDS-PAGE analysis of *B. megaterium* cell-free incubations with either empty vector (EV – pMM1520) or pRBBm258 plasmid DNA. **(C)** End-point GFP values ( $\mu\text{M}$ ) with increasing plasmid DNA concentration (nM) for the pRBBm258 and pKMMBm5 (lacking XylR repressor) plasmid.



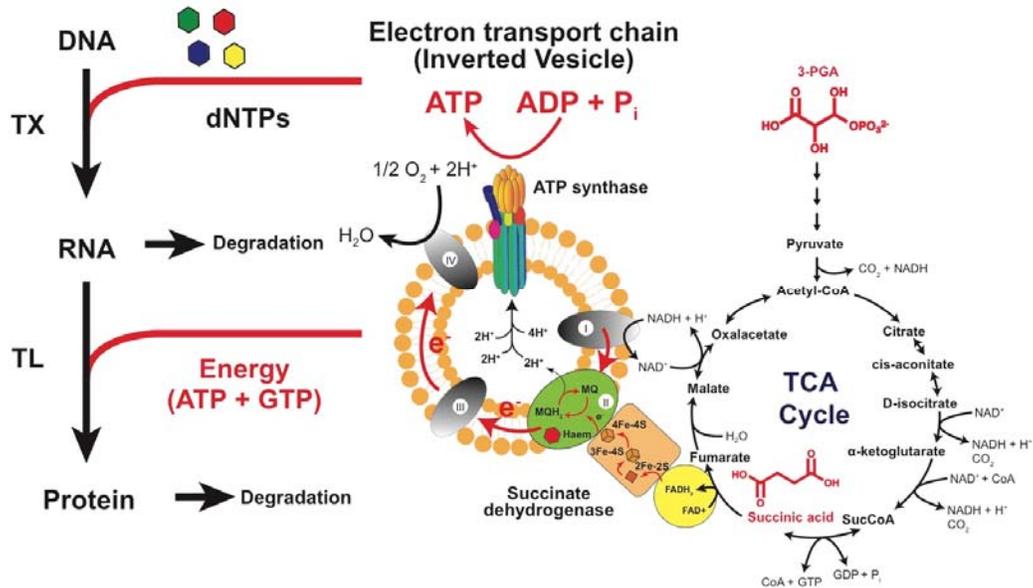
**Figure 2.** *B. megaterium* real-time ITT time course of mRNA and protein production. Design and experimental fluorescence time-course of 10 nM pKMMBm5-MGapt plasmid with mRNA (red line) and GFP (green line) reporters under standard conditions with 3-PGA energy buffer at 30°C. Measurements recorded every minute. Standard deviation is representative of a triplicate technical repeat and is displayed as a shaded area for each curve.



**Figure 3.** Transcription and translation of the xylose-inducible promoter. Cell-free extracts ( $10 \text{ mg mL}^{-1}$ ) were incubated at  $30^\circ\text{C}$  for 6 hours with a range of concentrations (1-10 nM) of the pKMMBm5-MGapt plasmid. Fluorescence data was collected every 60 seconds for MGapt (mRNA) and GFP (protein) signals. Experimental data were modelled using ordinary differential equations with a system of 4 species and 15 parameters detailed in equation 1 and Table S1 and 2, and parameter were inferred using Markov chain Monte Carlo.



**Figure 4.** Cell-free prototype validation and experimental augmentation of the xylose-inducible promoter system. Experimental and Bayesian modelling characterization of the xylose-inducible promoter system in cell-free. **(A)** D-xylose (mM), purified XyIR<sup>His</sup> ( $\mu\text{M}$ ) and pKMMBm5 plasmid DNA (nM) was titrated in a cell-free reaction using an Echo® 525 acoustic liquid handling system (LabCyte) with 25 nL transfer droplets. The end-point value of GFP is provided for this experimental dataset as a surface heat map for 324 conditions. Experimental data was simultaneously modelled with 4 species and 15 parameter variables with MCMC analysis. **(B)** The maximum likelihood simulated end-point GFP concentrations are shown as a surface with the experimental end-point GFP concentrations shown as black points. The distance between the modelled surface and the experimental values is shown as a connecting black line. **(C)** The inferred posterior distribution over the  $K_D$  and Hill-coefficients for binding of XyIR and xylose.



**Figure 5.** Energy regeneration and ITT reaction components in *B. megaterium*. A route for energy regeneration through either 3-PGA or succinic acid as energy substrates for ATP synthesis. Theoretical model of an inverted vesicle harbouring the electron transport chain for providing energy for transcription and translation stages.