A *Streptomyces venezuelae* Cell-Free Toolkit for Synthetic Biology

Simon J Moore[1,2,3*], Hung-En Lai[1,2], Soo-Mei Chee[2,5], Ming Toh[1,2], Seth Coode[3], Patrick Capel[4], Christophe Corre[4], Emmanuel LC de los Santos[4], Paul S Freemont[1,2,5,6*]

[1]Centre for Synthetic Biology and Innovation, Imperial College London, South Kensington Campus, Exhibition Road, London, SW7 2AZ, UK
[2]Department Section of Structural and Synthetic Biology, Department of Infectious Disease; Imperial College London, South Kensington Campus, Exhibition Road, London, SW7 2AZ, UK
[3]School of Biosciences, University of Kent, Canterbury, Kent CT2 7NJ, UK
[4]Warwick Integrative Synthetic Biology Centre, School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK;
[5]The London Biofoundry, Imperial College Translation & Innovation Hub, White City Campus, 80 Wood Lane, London W12 0BZ, UK.
[6]UK Dementia Research Institute Care Research and Technology Centre, Imperial College London, Hammersmith Campus, Du Cane Road, London, W12 0N

Keywords
Cell-free synthetic biology, *Streptomyces*, natural products, *in vitro* transcription-translation, cell-free protein synthesis

*Joint corresponding authors: Dr Simon Moore (S.J.R.Moore@kent.ac.uk) and Professor Paul Freemont (p.freemont@imperial.ac.uk)

Abstract

Prokaryotic cell-free coupled transcription-translation (TX-TL) systems are emerging as a powerful tool to examine natural product biosynthetic pathways in a test-tube. The key advantages of this approach are the reduced experimental timescales and controlled reaction conditions. In order to realise this potential, specialised cell-free systems in organisms enriched for biosynthetic gene clusters, with strong protein production and well-characterised synthetic biology tools, is essential. The *Streptomyces* genus is a major source of natural products. To study enzymes and pathways from *Streptomyces*, we originally developed a homologous *Streptomyces* cell-free system to provide a native protein folding environment, a high G+C (%) tRNA pool and an active background metabolism. However, our initial yields were low (36 µg/mL) and showed a high level of batch-to-batch variation. Here, we present an updated high-yield and robust *Streptomyces* TX-TL protocol, reaching up to yields of 266 µg/mL of expressed recombinant protein. To complement this, we rapidly characterise a range of DNA parts with different reporters, express high G+C (%) biosynthetic genes and demonstrate an initial proof of concept for combined transcription, translation and biosynthesis of *Streptomyces* metabolic pathways in a single ‘one-pot’ reaction.

Introduction

Streptomyces bacteria are environmental specialists (e.g. soil, marine, desert) that synthesize rich repertoires of natural products such as antibiotics. Much of this genetic information is locked up and cryptically regulated within biosynthetic gene clusters; regions of genomic DNA that harbor enzymes and other proteins (e.g. transporters, resistance markers). The key limitation in awakening these clusters for natural product discovery, is silent gene expression and recalcitrant genetics. Traditional strategies to overcome this include genetic modification of the host organism to bypass native regulatory elements, or the 'capture' of the cluster and expression in a heterologous host (1). But this can take several weeks to months to complete with varying levels of success: some cryptic clusters remain dormant due to obscure native regulation. Fundamental tools that aid these efforts, are of major interest to the natural product community.

Prokaryotic cell-free coupled transcription-translation systems are emerging as a new tool for studying natural product biosynthesis (2–8). Cell-free transcription-translation uses a crude cell-extract or purified ribosomes and translation factors – the PURE system - in a 'one-pot' reaction (9, 10). *E. coli* cell-extracts, referred to as either TX-TL (5, 11, 12) or cell-free protein synthesis (CFPS) (4, 13), are low-cost, straightforward to prepare and provide high recombinant protein yields, up to 2300 µg/mL (14). Moreover, metabolism is active, (15) providing ATP regeneration, while amino acid pathways are dynamic; providing additional ATP (through L-glutamate), while some amino acids deplete and are limiting for protein synthesis (16). In summary, TX-TL provides distinct opportunities for natural product biosynthesis: precursors for biosynthesis, direct control (e.g. feed precursors), short experimental timescales (4-24 hours), and stable yields. Moreover, we have shown the potential for automation in TX-TL by screening up to 500 plasmid variants in 24 hours (17).

While *E. coli* TX-TL and the PURE system are promising for natural product biosynthesis (2–4), *E. coli* has limited potential for studying biosynthetic gene clusters from *Streptomyces*, due to a number of genetic and metabolic differences: The codon content between *Streptomyces* (~70% G+C) and *E. coli* (51% G+C) is different; the regulatory sequences that control transcription, post-transcription and overall gene expression is distinct; and secondary metabolism in *E. coli* is not necessarily well-suited, or requires further metabolic engineering. Notwithstanding, *E. coli* synthesis of heterologous proteins can result in poor expression and solubility (5, 18). Therefore, we anticipate that a dedicated *Streptomyces* TX-TL system for homologous protein synthesis, has a number of advantages for studying natural product biosynthesis. As a first step, we originally released a *Streptomyces venezuelae* DSM-40230 (ATCC 10712) TX-TL system, but this had low protein yields (36 µg mL^{-1}) and showed high batch-to-batch variability. We chose *S. venezuelae* ATCC 10712, since it is well-suited to synthetic biology. *S. venezuelae* ATCC 10712 is fast-growing (40 min doubling time) and grows dispersedly in liquid culture; most *Streptomyces spp.* have slower doubling times and clump in mycelial aggregates. Moreover, *S. venezuelae* has a range of synthetic biology tools (19, 20) and is an attractive host for industrial biotechnology (21, 22). In parallel to our studies, the Jewett group (7) also established a *Streptomyces lividans* CFPS system (yields ~50 µg mL^{-1}), which was further optimised (23). A recent update to this system, highlighted the need for adding individual purified translation factors (8) to elevate protein synthesis up to ~400 µg/mL.

Based on this, we rationalized that protein synthesis, in our original proof of concept *S. venezuelae* TX-TL system, could be limited by the use of an energy solution originally optimized for a *E. coli* TX-TL protocol (11). TX-TL requires a cell-extract, a primary and secondary energy source, amino acids, cofactors, molecular crowding agents, additives (e.g. Mg^{2+}, spermidine, folinic acid, tRNA) to direct protein synthesis from a template DNA sequence. The energy source is composed of nucleotide triphosphates to drive initial mRNA and protein synthesis (primary energy source) and commonly 3-phosphoglyceric acid (3-PGA) or phosphoenolpyruvate (PEP) as the secondary energy source. 3-PGA or PEP provide ATP regeneration to leverage extended protein synthesis. Potentially primary metabolism could be activated in TX-TL to provide reducing equivalents (e.g. NADH, FADH), extra energy and building blocks (e.g. amino acids, malonyl-CoA), for natural product biosynthetic pathways, as shown in cell-extract metabolic engineering (24). In this work, we focus on upgrading our *S. venezuelae* system, to elevate protein synthesis and demonstrate its broader potential for cell-free synthetic biology – for characterising DNA parts and activating some model biosynthetic pathways. To achieve this, we made some simple modifications to the system, allowing yield up to 266 µg/mL of expressed recombinant proteins and demonstrate combined transcription-translation and biosynthesis of some example natural product pathways, namely melanin and haem biosynthesis. We also describe an easy-to-follow protocol that simply requires three components: DNA, cell-extract and a master mix that we describe in detail. We believe this generic *Streptomyces* TX-TL toolkit will be of broad interest to the natural product community, complementing experimental wet-lab tools for genome mining studies.

Results and discussion

A high-yield *Streptomyces* TX-TL protocol

To provide an improved *Streptomyces* TX-TL toolkit for synthesis of high G+C (%) genes and pathways from *Streptomyces* spp. and related genomes, a key priority was to optimise protein production and provide a straightforward protocol with minimal batch variation, for ease of repeatability. Since bacterial transcription and translation is coupled, either these steps, physical parameters or components from the energy solution, limit overall TX-TL activity. Therefore, to keep our protocol streamlined, we made the following changes: promoter strength, energy solution, ATP regeneration and RNAse inhibition – in doing so, we obtained a high-yield protocol with minimal variation between different cell-extract batches (Figure 1A).

Promoter strength – Previously, we used the *kasOp** promoter to drive mRNA synthesis in *Streptomyces* TX-TL. This made up to 1.3 μM (36 μg mL^{-1}) of the model superfolder green fluorescence protein (sfGFP) in our previous work (5). Promoter strength is a key limiting factor in heterologous expression systems. *kasOp** is a strong *Streptomyces* constitutive promoter, originally derived from the *kasO*/*cpkO*/*sco6280* promoter, with core -35 and -10 boxes of TTGACN and TAGART, respectively (25). The *kasOp** promoter is active in a range of *Streptomyces* spp. through the endogenous RNA polymerases and HrdB house-keeping Sigma factor (25). Bai *et al* developed a synthetic promoter library based around *kasOp** using fluorescence-activated cell sorting (FACS) to quantity *S. venezuelae* protoplasts (19). This included the isolation of a synthetic promoter 44 (SP44), which is 1.87-fold stronger than *kasOp** (19). We used *Streptomyces* TX-TL to test a panel of promoters developed by Bai *et al*, with SP44 the strongest (2.63 μM sfGFP) and 2.2-fold more active than *kasOp** (Figure 1B). We also repeated this across four independent cell-extract batches, but still observed strong batch variation. However, SP44 provided a stronger reporter plasmid to continue the optimisation process.

Energy solution – Next, we focused on developing a minimal energy solution (MES) to identify any non-essential components. The standard *E. coli* TX-TL energy solution used previously (5), is composed of HEPES buffer, ions (e.g. Mg-glutamate, K-glutamate), nucleotide triphosphates (NTPs - ATP, GTP, CTP and UTP), secondary energy source [typically 3-phosphoglyceric acid (3-PGA) or phosphoenolpyruvate (PEP)], amino acids, molecular crowding agent and a number of additives (13). To establish a MES for *Streptomyces* TX-TL, we first eliminated a number of non-essential components from the energy solution. This included coenzyme A, tRNA (*E. coli*), NAD, cAMP, folinic acid and spermidine (Figure S1A). While we did initially observe a positive response with cAMP, after several repeats in batches, this effect was not repeatable. For the HEPES buffer component, this was non-inhibitory (10-100 mM) and provides optimum activity between pH 8-9 (Figure S1B). For the secondary energy source, we found 3-PGA was essential; its removal decreased sfGFP synthesis by 98% (Figure S1C). We tried to replace 3-PGA with alternative secondary energy sources but observed only minimal activity: maltose (0.13 μM), sucrose (0.15 μM) and pyruvate (0.17 μM). Other potential sources such as glucose (with phosphate), PEP and succinate were inactive (Figure S1C). 3-PGA is the preferred secondary energy source in a range of non-model cell-extract hosts (26), due to its chemical stability and high energy potential, with an optimum concentration of 30 mM (Figure S1D). For the primary energy source (NTPs), there was some basal activity

without additional NTPs but 3 mM ATP/GTP and 1.5 mM CTP/UTP provides peak activity (Figure S1E). Next, surprisingly the removal of amino acids only decreased sfGFP synthesis by 45%, with 0.5-1.5 mM amino acids providing peak activity (data not shown). For metal salt screen, we found that $MgCl_2$, Mg-glutamate or Mg-acetate were all active (Figure S1F), while high levels of K-glutamate (150-200 mM) stimulated increased sfGFP synthesis (Figure S1F). This is possibly due to additional ATP regeneration via entry of α-ketoglutarate into the TCA cycle, as previously shown (27). Lastly, while we observed reasonable activity without PEG, 1% (w/v) PEG 6K was optimum, providing a 44% rise in activity (Figure S1G); it is however desirable to omit PEG for downstream natural product analytical purposes (e.g. LC-MS). Finally, based on these observations, we optimised our basic *Streptomyces* TX-TL MES system by individually fine-tuning the concentration of its core components (3-PGA, NTPs), while leaving DNA (40 nM), Mg-glutamate (4 mM), K-glutamate (150 mM), amino acids (1.25 mM) and PEG 6K (1%) constant. 3-PGA was most optimum at 30 mM, while the NTP level (ratio of 2:1 ATP/GTP:CTP/UTP), showed biphasic activity, peaking at 3 mM ATP/GTP, with full inhibition at 4 mM. Specific data on Mg-glutamate and K-glutamate optimisation with four different cell-extract batches is presented in Figure S2. As a combined result of this optimisation process, sfGFP synthesis was increased to 4 μM representing a 52% increase.

Additional ATP regeneration pathways - In a previous study, Caschera *et al* highlighted that other glycolytic enzymes function in *E. coli* TX-TL, using the disaccharide maltose (or maltodextrin) combined with 3-PGA to prolong protein synthesis up to 10 hours, through inorganic phosphate recycling (14). We investigated whether this part of metabolism is functional in *Streptomyces* TX-TL. Therefore, we tested the *Streptomyces* MES system (with 3-PGA) with maltose, glucose, glucose-6-phosphate (G6P) or fructose-1,6-phosphate (F16P). Interestingly, maltose, glucose, Glc6P, and F16P all prolonged the length of *Streptomyces* TX-TL activity from 2 to 3 hours. This was maximal with 5 mM Glc6P and 30 mM 3-PGA (Figure 1C), at an optimum temperature of 26°C (Figure 1D). All together we observed a 59% increase in sfGFP synthesis to 6.37 μM, but lower levels of NTP are required (Figure 1C) – equivalent to 1 mM ATP/GTP and 0.5 mM CTP/UTP. We speculate this could be related to ATP regulation of the glycolytic enzymes (e.g. hexokinase, fructokinase), leading to rapid depletion of ATP and inhibition of protein synthesis. Although this requires further investigation – there is limited literature on specific glycolytic enzymes from *Streptomyces*.

RNAse inhibition – As a final addition to the system, we tested the inexpensive RNAse inhibitor, polyvinylsulphonic acid (PVSA). Recently, PVSA, an RNA-mimetic, was shown to improve mRNA stability in *E. coli* TX-TL, but not increase protein synthesis (28). In *Streptomyces* TX-TL, 1 mg/mL PVSA increased sfGFP synthesis up to 5.87 μM, in the basic MES system (Figure 1E). But while we observed individual improvements with either the PVSA RNase inhibitor or the blended G6P/3-PGA secondary energy source, in combination, there was no significant additive effect with PVSA and G6P/3-PGA together. This suggested that other rate-limiting factors are at play.

In summary, we have made a specific energy solution for *Streptomyces* TX-TL with an overall 6-fold improvement in the system, partly attributed to combined use of 3PGA and G6P as secondary energy sources (Figure 1F). Furthermore, we find this can

217 combine with remaining reaction components into a single *Streptomyces* master mix
218 (SMM) solution, further streamlining the reaction process. With this simple
219 modification, the TX-TL reaction requires three single components that minimises
220 batch variation: SMM solution, plasmid DNA and the cell-extract. Next, we sought to
221 demonstrate the use of this simplified system for the testing of plasmid tools and
222 regulatory elements for *Streptomyces* synthetic biology.
223
224 **Cell-free characterisation of *Streptomyces* genetic tools for synthetic biology**
225 It is highly desirable to characterise standard DNA parts using rapid and iterative
226 design-build-test-learn cycles - the central paradigm of synthetic biology. For
227 *Streptomyces* and related strains, either conjugation or protoplast transformation is
228 typically used to transfer self-replicating and integrative plasmids for the testing of DNA
229 parts for *Streptomyces* synthetic biology (19, 20, 25). DNA parts are small modular
230 regulatory elements (e.g. promoter, insulator, tags, RBS, ORF, terminator) that
231 facilitate downstream combinatorial DNA assembly workflows (e.g. Golden Gate) for
232 refactoring gene expression pathways. While there are different approaches to
233 quantitate gene expression (20, 29), Bai *et al* (19) recently applied a lysozyme method,
234 to study single-cell gene expression quantitation of *S. venezuelae* ATCC 10712
235 protoplasts using fluorescence-activated cell sorting (FACS).
236
237 We next tested the promoter and RBS elements from Bai *et al* (19) in *Streptomyces*
238 TX-TL, as well as other important regulatory elements: alternative start codons (29)
239 and terminators (30). Firstly, we built these DNA parts to be compatible with our
240 previous DNA assembly method EcoFlex (33). For this we had to modify the promoter
241 consensus (prefix renamed, e.g. SP44a instead of SP44), to remove an internal BsmBI
242 site to permit MoClo assembly. In addition, to provide comparative *in vivo* data, we
243 built a new destination vector (cured of BsmBI and BsaI sites) from pAV-*gapdh* from
244 Phelan *et al* (20) and renamed this StrepFlex (pSF1). pAV-*gapdh* is an integrative
245 shuttle vector, developed as a synthetic biology plasmid tool for *S. venezuelae* (20).
246 Firstly, for the promoter library (*kasOp**, SP15a, SP19a, SP23a, SP25a, SP30a,
247 SP40a, SP44a and *ermEp**), we assembled this with the RiboJ insulator, R15 capsid
248 RBS, mScarlet-I and the Bba_B0015 terminator. For the RBS library, *kasOp** was
249 used for the promoter. For the promoter variants, activity ranged from 5% (*ermEp**) to
250 100% (*kasOp**). In contrast to earlier results (Figure 1B), the BsmBI cured promoters
251 were about 30-50% less active across the library. For the activity of the RBS variants,
252 this ranged from 0.7% (SR9) to 117% (SR39) activity relative to the R15 capsid RBS
253 (Figure 2B). We also tested two-dimensional promoter and RBS space with sfGFP
254 (Figure 2C). Lastly, to provide *in vivo* data, we characterised the mScarlet-I promoter
255 and RBS plasmids (from Figure 2A and 2B) in *S. venezuelae* ATCC 10712 (Figure
256 S3), following the approach by Phelan *et al* (20). Interestingly, there were some
257 significant outliers particularly for the RBS library; SR39 (along with the *E. coli* PET-
258 RBS) was the strongest RBS in contrast to SR40, which was unexpectedly weaker
259 both *in vitro* and *in vivo*. In addition, SR4, an expected weak RBS, was strong in both
260 *in vitro* and *in vivo* measurements (Figure 2A-2B, Figure S3). This may reflect
261 differences in the upstream 5'-untranslated (5'-UTR) region and the use of a different
262 fluorescence reporter (mScarlet-I), in comparison to Bai *et al* (19). However, on the
263 whole the promoter and RBS strengths characterised were broadly consistent with the
264 original publication (19).
265

266 Another important, but perhaps underused regulator of gene expression in the context
267 of synthetic biology, is the start codon. In most bacteria, ATG is the preferred codon
268 for translation initiation through fMet-tRNA. Previously, Myronovskyi *et al* used a β-
269 glucuronidase (GUS) reporter to show that the TTG codon was stronger than ATG for
270 translation initiation, by almost 2-fold in both *Streptomyces albus* J1074 and
271 *Streptomyces sp.* Tu6071 (29). Using sfGFP and mScarlet-I as reporters, our findings
272 suggest that for *S. venezuelae* at least, ATG is equivalent in strength to TTG, followed
273 by CTG and lastly GTG as the weakest (Figure 2C). This also likely changes with
274 coding sequences and 5'-UTR. In comparison, for *E. coli* the order of strength goes as
275 follows: ATG > GTG > TTG > CTG (31). We expect this differs due to high GC codon
276 bias in *Streptomyces*. Despite the use of different experimental conditions, our results
277 confirm that TTG is a strong alternative start codon and that GTG is weak, although
278 the role of this is unclear and intriguing due its high frequency in *Streptomyces*
279 genomes (32).
280
281 Lastly, to the best of our knowledge, no studies have so far reported the use of
282 terminators for controlling pathway expression in *Streptomyces*. Using the same
283 experimental format as we previously used in EcoFlex (33), we tested a selection of
284 Rho-independent terminators from the iGEM catalogue (Bba_B0012, Bba_B0015)
285 and from Chen *et al* (34) in *S. venezuelae* TX-TL (Figure 2D). These activities strongly
286 follow our previous observations in *E. coli* cell-free (33) and these terminators provide
287 a final tool for the design and assembly of synthetic metabolic pathways. These
288 terminators were also designed to prevent repetition in DNA elements and protect
289 against homologous recombination as previously highlighted (34, 35). For now, our
290 TX-TL system demonstrates proof-of-concept data for prototyping DNA parts in
291 *Streptomyces*, which may in future provide a platform for designing pathways – for
292 refactoring precise synthetic biology control of metabolic pathways in engineered
293 *Streptomyces* strains.
294
295 **TX-TL synthesis of high G+C (%) genes**
296 Previously, we found our *Streptomyces* TX-TL system was most active with 40 nM of
297 plasmid DNA, using the *kasOp**-sfGFP reporter, to saturate protein synthesis (28). In
298 comparison to *E. coli* TX-TL, protein synthesis is saturated at around 5-10 nM of
299 reporter DNA, which varies with different promoters and sigma factors (36). We
300 questioned whether DNA degradation led to this discrepancy; most *Streptomyces spp.*
301 degrade methylated plasmid DNA with endonucleases (37). To compare methylated
302 and unmethylated plasmid DNA for methylation-specific endonucleases, we tested
303 unmethylated and methylated SP44-sfGFP reporter plasmid in *Streptomyces* TX-TL.
304 Interestingly, there was no major change in sfGFP synthesis between unmethylated
305 or methylated plasmid DNA, across different DNA concentrations (Figure S4). We also
306 tested relative plasmid DNA stability in *S. venezuelae* cell-extracts, with the standard
307 MES energy solution, and incubated at different time-lengths, followed by re-extraction
308 of the plasmid DNA (using the Qiagen plasmid DNA purification kit), which was then
309 separated and visualised on a 1% (w/v) agarose gel. This indicated that methylated
310 plasmid DNA is stable, during the time (0-4 hr) when TX-TL is active (Figure S4).
311 Further to this, we also tested linear DNA for exonuclease activity. To protect the
312 coding sequences, we PCR amplified about 150-250 bases upstream and
313 downstream of the coding parts, using the standard SP44-sfGFP reporter plasmid. But
314 in the TX-TL reaction, linear DNA was 95% less active than circular DNA, at 40 nM of

DNA (Figure S5). This suggests the *S. venezuelae* cell-extract has exonuclease activity, but endonuclease activity is minimal.

Since circular DNA degradation was not a limiting factor, we tested different fluorescent proteins (Figure 3A) to determine if the optimum plasmid DNA concentration, for protein synthesis, changes. Firstly, we tested mVenus-I and mScarlet-I, combined with the strong SP44 promoter, in comparison to the SP44-sfGFP reporter. The maximum yields achieved for these three proteins were: 6.48 µM sfGFP (174 µg/mL), 9.50 µM mScarlet-I (266 µg/mL) and 7.72 µM mVenus-I (224 µg/mL). Both SP44-mScarlet-I and SP44-mVenus saturated protein synthesis with a lower DNA template (10 nM) (Figure 3B). This was surprising, since the coding sequence of mVenus is 96% identical to sfGFP, with the exception of 30 mutations and an additional GTG (valine) at the second codon for mVenus-I. We speculate that this alters either mRNA stability or translation initiation, although this requires further investigation. Secondly, we also tested the robustness of the system for other proteins from high G+C (%) genes (Figure 3C), using the oxytetracycline enzymes (OxyA, -B, -C, -D, -J, -K, -N and -T) from *Streptomyces rimosus* that were previously only detectable by Western blotting in our original publication (30), as well as three non-ribosomal peptide synthetases (NRPS); this included the TxtA and TxtB NRPS enzymes from thaxtomin A biosynthesis in *Streptomyces scabiei* and an uncharacterised NRPS (NH08_RS0107360) from *S. rimosus*. With the exception of TxtA, most enzymes were discernible by either SDS-PAGE (Figure 3C), or for OxyA (47 kDa) and TxtB (162 kDa), low levels (<0.5 µM) were detected by Western blotting using an anti-Flag tag (data not shown). We also incorporated a C-terminal tetracysteine tag with the oxytetracycline enzymes and mScarlet-I, using the fluorogenic biarsenical dye fluorescein arsenical hairpin binder-ethanedithiol (FlAsH-EDT_2), to measure real-time nascent protein synthesis (Figure 3D). Most oxytetracycline enzymes showed a significant increase in FlAsH-EDT_2 fluorescence ($P < 0.05$), peaking at 120 min, with only OxyN producing the weakest response ($P = 0.056$), although clearly detectable by PAGE or Western. For mScarlet-I, the time-lag between the fluorescence signals for FlAsH-EDT2 (immature protein) and mScarlet-I (mature protein), allowed us to estimate a maturation time of 40 min for mScarlet-I (Figure 3E); this is in close agreement to a literature value of 36 min, calculated *in vivo* (38). In summary, our *Streptomyces* TX-TL system is robust for expression of high G+C (%) genes, has multiple tools (e.g. tags, plasmid systems) and is comparable to other bacterial TX-TL systems.

Transcription, translation and biosynthesis

The next step was to reconstitute a biosynthetic pathway in *Streptomyces* TX-TL system. Initially, to show the synthesis of a single enzyme and its activity, we selected the GUS reporter enzyme. We synthesised the enzyme in the TX-TL reaction from 40 nM SP44-*gus*, left for 4 hrs at 30°C. The GUS enzyme showed a clear band on SDS-PAGE at the expected size of 68 kDa (Figure S6). To test for GUS activity, an equal volume of the TX-TL extract, as well as a negative control reaction, was mixed with X-GlcA substrate in increasing concentrations (10, 25 and 50 mg/mL). Only the extract from the SP44-*gus* reaction developed a deep blue pigment, within minutes, indicating strong GUS activity (Figure 4A).

Next, we selected two metabolic pathways to provide a further test for the TX-TL system. We selected two operons from *S. venezuelae* encoding the melanin and early-

stage haem biosynthetic pathways to provide a discernible output for testing (fluorescence and or colorimetric). Also, both operons were selected from *S. venezuelae*, to improve expression in TX-TL, since the codon usage is adapted to this host.

Melanin is a natural pigment that absorbs ultraviolet (UV) light to protect cells from DNA damage. Recently, Matoba *et al* studied the mechanism of tyrosinase and the role of the "caddie" protein from *Streptomyces castaneoglobisporus* HUT6202 (39). Tyrosinase, TyrC, catalyses the rate-limiting step in melanin biosynthesis: it oxidises the phenol group (in L-tyrosine) into the *orthro*-quinone intermediate, which enters an autocatalytic cascade into the melanin pathway. TyrC is dicopper-dependent, with each Cu(II) atom coordinated by three His residues, which is transferred by MelC1, a small (12.8 kDa) metallochaperone, which transfers copper to the active site. The *S. venezuelae* tyrosinase operon encodes both MelC1 and TyrC. Using our *Streptomyces* TX-TL and SP44-*melC1-tyrC*, from denaturing PAGE, we saw clear synthesis of TyrC at approximately 34 kDa (expected 31.4 kDa) although MelC1 was indistinguishable (Figure S7). In terms of activity, we observed brown pigment formation after ~2 hours, only with the addition of 1 mM $CuCl_2$ (Figure 4B). This indicated melanin biosynthesis was active, despite the apparent absence of MelC1. Without the addition of $CuCl_2$, or the tyrosinase plasmid, the cell-extracts remained clear. Previously, Matoba *et al* showed insertion of Cu(II) into TyrC by MelC1 involves a transient interaction, and that MelC1 is unstable and forms aggregates that are difficult to detect with PAGE (39). Also, apo-TyrC is inactive with Cu(II) alone, which suggests that our TX-TL system supports the synthesis of both TyrC and MelC1.

Lastly, we tested a three-gene biosynthetic operon (*hemC-hemD*/*cysG^A-hemB*) that catalyses the early-stages of haem biosynthesis (40). This pathway was selected since it contains a known fluorescence reporter enzyme CysG^A (41) (also referred to CobA), a methyltransferase naturally fused as HemD/CysG^A. We added a pTU1-A-SP44-*hemC-hemD*/*cysG^A-hemB* (pTU1-A-SP44-*hem*) plasmid into the TX-TL reaction, as well as a negative control plasmid (pTU1-A), both with and without 5-aminolevulinic acid (5-ALA), the substrate for the pathway. In the presence of pTU1-A and 1 mM ALA, there was some minor background fluorescence (Figure 4C), which we expected since haem biosynthesis is essential. In contrast, with pTU1-A-SP44-*hem* and 1 mM ALA, this generated strong red fluorescence, 20-fold higher than background levels in the control reaction (pTU1-A and 1 mM ALA). For protein synthesis, while we could detect HemB (35 kDa) and HemC (38 kDa), the fusion protein HemD/CysG^A was less clear, with other major bands at the expected mass – 58.3 kDa (Figure S6). To verify pathway function, we ran a semi-continuous reaction (42) to facilitate purification by separating the haem intermediates from the cell-extract proteins (inset image in Figure 4C). Interestingly, LC-MS analysis detected the air-oxidised product of the HemD enzyme (uroporphyrinogen III - 837 m/z), observed as a 6-electron oxidised uroporphyrin III (red fluorescent) intermediate at 831 *m/z*; typical for these air-sensitive intermediates (Figure S8). These results show that our *S. venezuelae* TX-TL system can support the synthesis of at least three enzymes from plasmid DNA in a combined 'one-pot' combined translation, translation and enzymatic pathway.

Conclusions

Our study complements a recent surge in interest in the use of cell-free systems for the study of biosynthetic pathways (2, 4, 23, 42). Here we wanted to expand the palette of plasmid tools for the further development of *S. venezuelae* as a synthetic biology chassis by developing an optimised streptomyces TX-TL toolkit (5, 20, 21). Our combined findings show an order of magnitude improvement in protein synthesis over our original streptomyces TX-TL system, without the requirement of any genetic modifications to either limit RNA degradation or increase translation rates; indeed, Xu *et al* recently show this is clear rate-limiting step for other *Streptomyces* cell-free systems (8). In conclusion, our results realise the early-stage potential of *Streptomyces* for the study of synthetic biology for natural products, providing a native prototyping environment for developing synthetic biology tools (e.g. promoters/RBS) and also for the study of biosynthetic pathways from these organisms. We also demonstrate that the semi-continuous system permits reasonable milligram scale-up of biosynthetic metabolites and a clean route to purification and analysis.

428 **Methods**
429
430 **Molecular biology**
431 All plasmids were either prepared using EcoFlex cloning or by routine materials and
432 methods, as previously described (34). For PCR of high GC genes and operons, Q5
433 polymerase (NEB, UK) was used, using standard cycling or touchdown (72-59°C
434 annealing) and the addition of 5% DMSO. For tricky amplicons, the protocol was
435 modified with an annealing time of 30 seconds and elongation temperature reduced
436 to 68°C. The following bacterial strains were used: *S. venezuelae* DSM-40230 and *E.*
437 *coli* DH10β. Unmethylated plasmid DNA was prepared from an *E. coli dam*$^-$ *dcm*$^-$
438 mutant (C2925) obtained from NEB. Plasmids and oligonucleotides are listed in Table
439 S1-S5.
440
441 **Preparation of cell-extracts**
442 *S. venezuelae* ATCC 10712 was grown in GYM (prepared in distilled water). The cell-
443 extracts were prepared as described previously (33), with the exception that β-
444 mercaptoethanol was removed from the wash buffers and replaced with 2 mM
445 dithiothreitol.
446
447 **Energy solution and reaction conditions**
448 The reaction mixture contained: 8 mg/mL cell-extract, 40 nM DNA template, 25 mM
449 HEPES, 1 mM ATP, 1 mM GTP, 0.5 mM UTP, 0.5 mM CTP, 30 mM 3-PGA, 5 mM
450 glucose-6-phosphate, 1.5 mM amino acids (1.25 mM L-leucine), 4 mM Mg-glutamate,
451 150 mM K-glutamate, 1% (w/v) PEG6K and 5 mg/mL PVSA. All reactions were
452 incubated at 30°C, 40 nM pTU1-A-SP44-sfGFP and the MES or SMM buffer system,
453 unless otherwise stated. At least three technical repeats were prepared (for
454 fluorescence measurements) and repeated with at least two independent cell-extract
455 batches (from A4-A7) prepared on separated days.
456
457 **Denaturing PAGE**
458 40 µL cell-free reaction (30 µL SMM + 10 µL plasmid DNA) was incubated in a 2 mL
459 tube at 25-30°C (no shaking) for 6 hours. To precipitate proteins, 1 mL ice-cold 100%
460 (v/v) acetone was added. Samples were placed at -20°C for 30 mins, before
461 centrifugation at 18,000 × *g*, 4°C for 10 min. The supernatant was removed, and the
462 pellet was washed with 1 mL ice-cold 70% (v/v) acetone. Centrifugation and
463 supernatant removal steps were then repeated. The pellet was air-dried, before re-
464 suspending in 30 µL ddH_2O and 10 µL 4X NuPAGE lithium dodecyl sulphate (LDS)
465 sample buffer (ThermoFisher) and boiled at 100°C for 5 min. To ensure the pellet was
466 solubilized, samples were aspirated with a pipette five times, and if necessary (a visible
467 pellet remaining), left for an additional 5 min at 100°C. 10-100 µg total protein was
468 then separated with a 4-12% (v/v) gradient Bis-Tris gel (ThermoFisher) run in MES
469 buffer system. Proteins were stained with InstantBlue (Generon), destained with
470 ddH_2O and images were recorded with the ChemiDoc XRS imaging system (Biorad).
471
472 **TX-TL fluorescence measurements**
473 10 µL cell-free reactions were prepared in a 384-well black ⌈Clear®, F-bottom, low-
474 binding plate (Greiner). Reactions were measured as a triplicate technical repeat and
475 at least repeated with cell-extracts prepared from two separate days. were measured
476 in a 384-well plate. The plate was sealed with aluminium film, SILVERseal™ (Greiner),
477 briefly centrifuged at 2,000 × *g* for 10 seconds. Real-time plate measurements were

recorded in a CLARIOStar© plate reader (BMG Labtech, Germany) at 30°C with 10 seconds of shaking at 500 rpm prior to measurements, using either standard filters (Omega) or monochromator settings (CLARIOStar). Purified sfGFP, mVenus-I and mScarlet-I standards were purified, as described previously (33), to estimate protein concentration during real-time fluorescence measurements.

Mass spectrometry analysis

TX-TL reactions were prepared as two components (A and B) in a semi-continuous reaction as follows: Component A - 100 µL of standard TX-TL reaction, in the absence of PEG, was injected into a Thermo Scientific Pierce 3.5K MWCO 96-well microdialysis device; Component B - 1.5 mL SMM solution with 1 mg/mL carbenicillin, in a 2.5 mL tube. The microdialysis cassette was placed inside the 2.5 mL tube and incubated at 30°C for 24 hrs with shaking (1000 rpm). Samples were acidified with 1% (v/v) HCl, centrifuged at 18,000 x *g* for 25 min at room temperature. The supernatant was loaded onto a pre-equilibrated, with 1% (v/v) HCl, Sep-Pak C-18 (50 mg sorbent) solid-phase extraction cartridge (Waters), washed with 10 mL of 10% (v/v) ethanol and eluted with 2 mL of 50% (v/v) ethanol. All solutions were acidified with 1% (v/v) HCl. Eluted samples were dried under vacuum, at room temperature, using an Eppendorf Concentrator Plus. Samples were dissolved in 150 µL 1% HCl and centrifuged again at 18,000 x *g* for 25 min at room temperature. 1 µL of supernatant was then analysed by LC-MS, performed with an Agilent 1290 Infinity system with an online diode array detector in combination with a Bruker 6500 quadruple time-of-flight (Q-ToF) mass spectrometer. An Agilent Extend-C18 2.1 x 50mm (1.8 µm particle size) column was used at a temperature of 40°C with a buffer flow rate of 0.5 mL^{-1} min^{-1}. LC was performed with a gradient of buffer A (0.1% formic acid in water) and buffer B (0.1% formic acid in acetonitrile). Separation was achieved using 2% buffer B for 0.6 min, followed by a linear gradient to 100% buffer B from 0.6 - 4.6 min, which was held at 100% buffer B from 4.6 - 5.6 min followed by a return to 2% buffer B from 5.6 - 6.6 min, along with 1 min post run. Spectra were recorded between a mass range of 50-1700 *m/z* at a rate of 10 spectra per second in positive polarity.

508 **References**

509 1. P. J. Rutledge, G. L. Challis, Discovery of microbial natural products by
510 activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13**, 509–523
511 (2015).

512 2. Y. Khatri, *et al.*, Multicomponent Microscale Biosynthesis of Unnatural
513 Cyanobacterial Indole Alkaloids. *ACS Synth. Biol.* (2020)
514 https:/doi.org/10.1021/acssynbio.0c00038.

515 3. L. Zhuang, *et al.*, Total in vitro biosynthesis of the nonribosomal macrolactone
516 peptide valinomycin. *Metab. Eng.* **60**, 37–44 (2020).

517 4. A. W. Goering, *et al.*, In vitro reconstruction of nonribosomal peptide
518 biosynthesis directly from DNA using cell-free protein synthesis. *ACS Synth. Biol.* **6**,
519 39–44 (2017).

520 5. S. J. Moore, H.-E. Lai, H. Needham, K. M. Polizzi, P. S. Freemont,
521 \textlessi\textgreaterStreptomyces venezuelae\textless/i\textgreater TX-TL - a next
522 generation cell-free synthetic biology tool. *Biotechnol. J.*, 1600678 (2017).

523 6. T. Ozaki, *et al.*, Dissection of goadsporin biosynthesis by in vitro reconstitution
524 leading to designer analogues expressed in vivo. *Nat. Commun.* **8**, 14207 (2017).

525 7. J. Li, H. Wang, Y. C. Kwon, M. C. Jewett, Establishing a high yielding
526 streptomyces-based cell-free protein synthesis system. *Biotechnol. Bioeng.* **114**,
527 1343–1353 (2017).

528 8. H. Xu, W.-Q. Liu, J. Li, Translation Related Factors Improve the Productivity of a
529 Streptomyces-Based Cell-Free Protein Synthesis System. *ACS Synth. Biol.* **9**, 1221–
530 1224 (2020).

531 9. Y. Kuruma, T. Ueda, The PURE system for the cell-free synthesis of membrane
532 proteins. *Nat. Protoc.* **10**, 1328–1344 (2015).

533 10. B. Lavickova, S. J. Maerkl, A Simple, Robust, and Low-Cost Method To Produce
534 the PURE Cell-Free System. *ACS Synth. Biol.* **8**, 455–462 (2019).

535 11. Z. Z. Sun, *et al.*, Protocols for implementing an Escherichia coli based TX-TL
536 cell-free expression system for synthetic biology. *J. Vis. Exp.* **50762**, e50762 (2013).

537 12. M. K. Takahashi, *et al.*, Rapidly Characterizing the Fast Dynamics of RNA
538 Genetic Circuitry with Cell-Free Transcription-Translation (TX-TL) Systems. *ACS*
539 *Synth. Biol.* **4**, 503–515 (2015).

540 13. N. E. Gregorio, M. Z. Levine, J. P. Oza, A User's Guide to Cell-Free Protein
541 Synthesis. *Methods Protoc* **2** (2019).

542 14. F. Caschera, V. Noireaux, Synthesis of 2.3 mg/ml of protein with an all
543 Escherichia coli cell-free transcription-translation system. *Biochimie* **99**, 162–168
544 (2014).

15. M. C. Jewett, J. R. Swartz, Mimicking the Escherichia coli Cytoplasmic Environment Activates Long-Lived and Efficient Cell-Free Protein Synthesis. *Biotechnol. Bioeng.* **86**, 19–26 (2004).

16. J. F. Zawada, *et al.*, Microscale to manufacturing scale-up of cell-free cytokine production-a new approach for shortening protein production development timelines. *Biotechnol. Bioeng.* **108**, 1570–1578 (2011).

17. S. J. Moore, *et al.*, Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4340–E4349 (2018).

18. Q. Cheng, L. Xiang, M. Izumikawa, D. Meluzzi, B. S. Moore, Enzymatic total synthesis of enterocin polyketides. *Nat. Chem. Biol.* **3**, 557–558 (2007).

19. C. Bai, *et al.*, Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in Streptomyces. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 12181–12186 (2015).

20. R. M. Phelan, *et al.*, Development of next generation synthetic biology tools for use in Streptomyces venezuelae. *ACS Synth. Biol.*, acssynbio.6b00202 (2016).

21. R. M. Phelan, O. N. Sekurova, J. D. Keasling, S. B. Zotchev, Engineering terpene biosynthesis in streptomyces for production of the advanced biofuel precursor bisabolene. *ACS Synth. Biol.* **4**, 393–399 (2015).

22. A. R. Han, *et al.*, Development of a Streptomyces venezuelae-based combinatorial biosynthetic system for the production of glycosylated derivatives of doxorubicin and its biosynthetic intermediates. *Appl. Environ. Microbiol.* **77**, 4912–4923 (2011).

23. J. Li, H. Wang, M. C. Jewett, Expanding the palette of Streptomyces-based cell-free protein synthesis systems with enhanced yields. *Biochem. Eng. J.* (2018).

24. Q. M. Dudley, C. J. Nash, M. C. Jewett, Cell-free biosynthesis of limonene using enzyme-enriched Escherichia coli lysates . *Synth. Biol.* **4** (2019).

25. W. Wang, *et al.*, An engineered strong promoter for streptomycetes. *Appl. Environ. Microbiol.* **79**, 4484–4492 (2013).

26. S. S. Yim, *et al.*, Multiplex transcriptional characterizations across diverse bacterial species using cell-free systems. *Mol. Syst. Biol.* **15**, e8875 (2019).

27. M. C. Jewett, K. a. Calhoun, A. Voloshin, J. J. Wuu, J. R. Swartz, An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol.* **4**, 220 (2008).

28. C. C. Earl, M. T. Smith, R. A. Lease, B. C. Bundy, Polyvinylsulfonic acid: A Low-cost RNase inhibitor for enhanced RNA preservation and cell-free protein translation. *Bioengineered* **9**, 90–97 (2018).

29. M. Myronovskyi, E. Welle, V. Fedorenko, A. Luzhetskyy, β-Glucuronidase as a Sensitive and Versatile Reporter in Actinomycetes. *Appl. Environ. Microbiol.* **77**, 5370–5383 (2011).

30. K. Temme, D. Zhao, C. a. Voigt, Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proceedings of the National Academy of Sciences* **109**, 7085–7090 (2012).

31. A. Hecht, *et al.*, Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids Res.* **45**, 3615–3626 (2017).

32. M. J. Bibb, V. Molle, M. J. Buttner, sigma(BldN), an extracytoplasmic function RNA polymerase sigma factor required for aerial mycelium formation in Streptomyces coelicolor A3(2). *J. Bacteriol.* **182**, 4606–4616 (2000).

33. S. J. Moore, *et al.*, EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology. *ACS Synth. Biol.* **5**, 1059–1069 (2016).

34. Y.-J. Chen, *et al.*, Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **10**, 659–664 (2013).

35. A. A. K. Nielsen, *et al.*, Genetic circuit design automation. *Science* **352**, 6281 (2016).

36. J. Garamella, R. Marshall, M. Rustad, V. Noireaux, The all E. coli TX-TL Toolbox 2.0: a platform for cell-free synthetic biology. *ACS Synth. Biol.* **5**, 344–355 (2016).

37. G. GonzÃ¡lez-CerÃ³n, O. J. Miranda-Olivares, L. ServÃ-n-GonzÃ¡lez, Characterization of the methyl-specific restriction system of \textlessi\textgreaterStreptomyces coelicolor\textless/i\textgreater A3(2) and of the role played by laterally acquired nucleases. *FEMS Microbiol. Lett.* **301**, 35–43 (2009).

38. D. S. Bindels, *et al.*, mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat. Methods* **14**, 53–56 (2017).

39. Y. Matoba, *et al.*, Catalytic mechanism of the tyrosinase reaction toward the Tyr98 residue in the caddie protein. *PLoS Biol.* **16**, e3000077 (2018).

40. I. U. Heinemann, M. Jahn, D. Jahn, The biochemistry of heme biosynthesis. *Arch. Biochem. Biophys.* **474**, 238–251 (2008).

41. S. Wildt, U. Deuschle, cobA, a red fluorescent transcriptional reporter for Escherichia coli, yeast, and mammalian cells. *Nat. Biotechnol.* **17**, 1175–1178 (1999).

42. I. Siebels, *et al.*, Cell-Free Synthesis of Natural Compounds from Genomic DNA of Biosynthetic Gene Clusters. *ACS Synth. Biol.* **9**, 2418–2426 (2020).

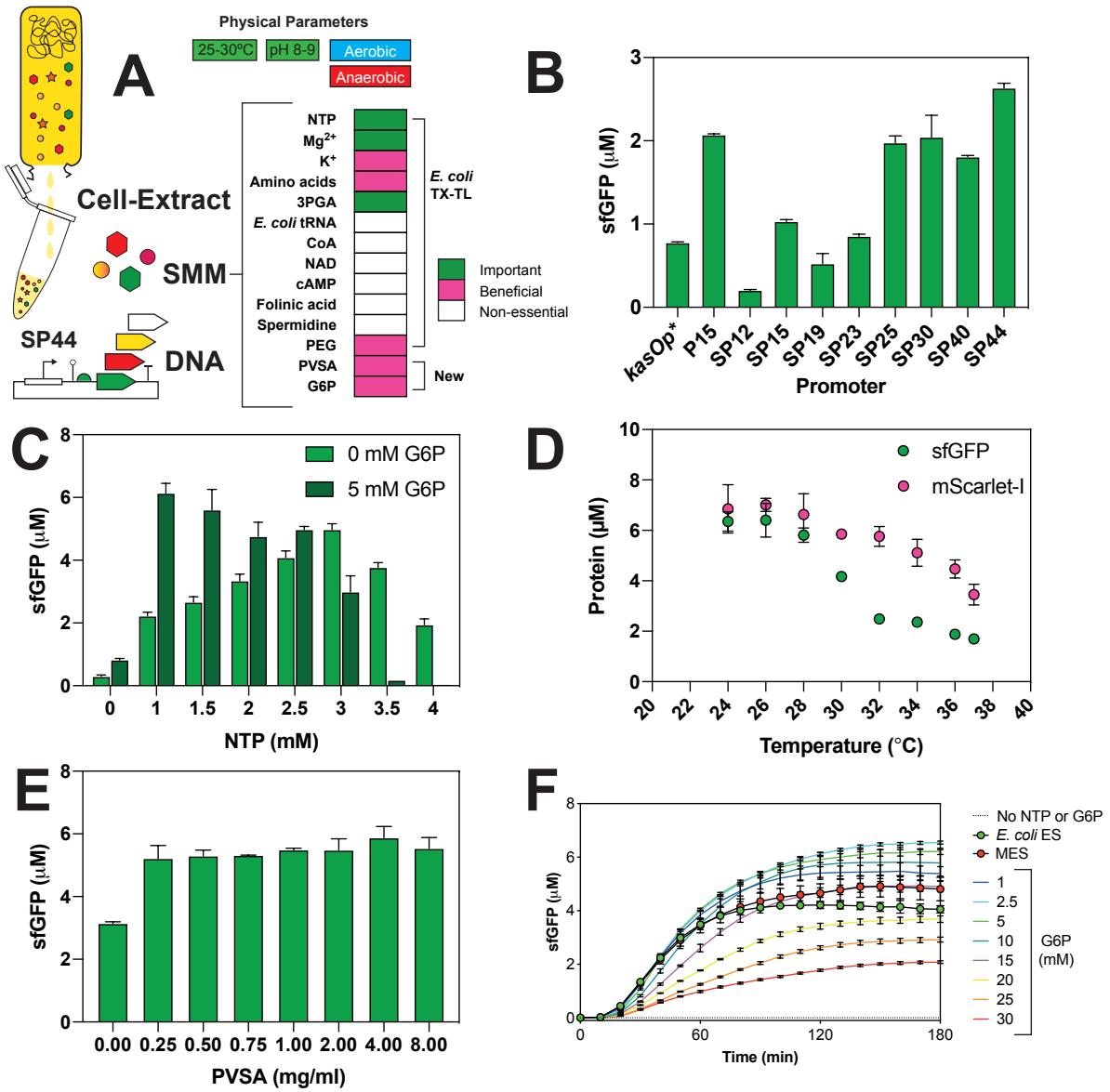

Figure 1. Overview of *Streptomyces* TX-TL optimisation. (A) Outline of physical and biochemical parameters of *Streptomyces* TX-TL and SMM buffer system. Optimisation of: (B) Promoter strength from Bai *et al* (19); (C) primary energy source; (D) Temperature; (E) PVSA; and (F) G6P.

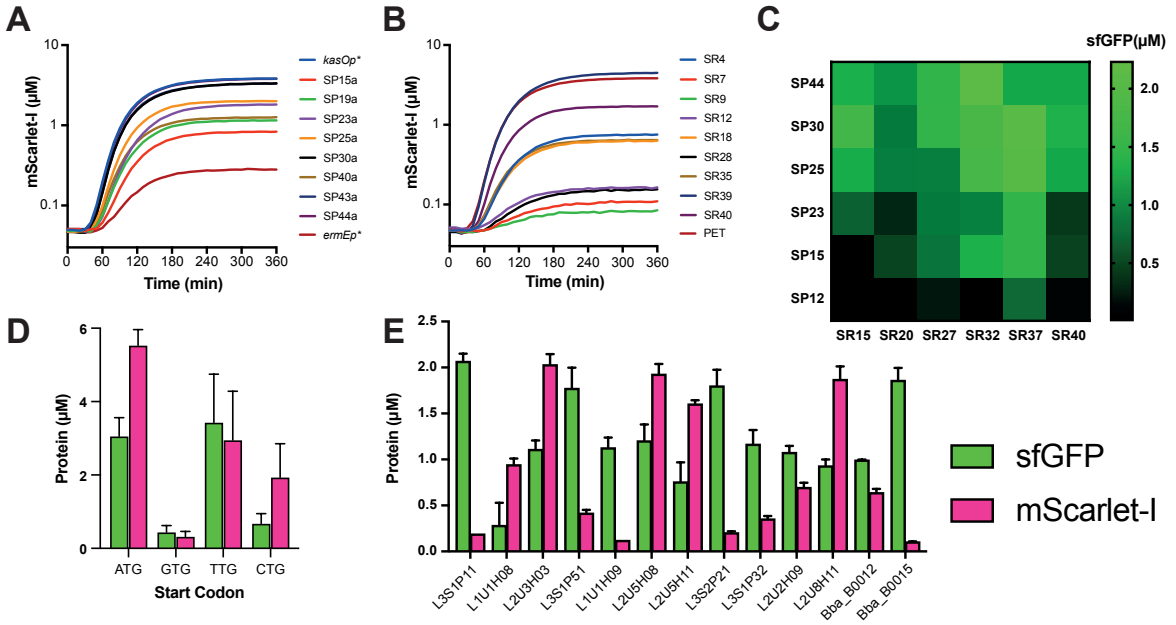

Figure 2. Part characterisation of *Streptomyces* regulatory elements: (A) Promoter-R15 capsid RBS-mScarlet-I; (B) *kasOp**-RBS-mScarlet-I; (C) Promoter-RBS-sfGFP combinations; (D) variable start codons (with sfGFP and mScarlet-I); and (E) variable Rho-independent terminators from Chen *et al* (34). For terminator plasmid design, see Moore *et al* (33). 40 nM plasmid DNA was incubated in the optimised reaction conditions at 28°C as a technical triplicate repeat and repeated on two separate days. Unless otherwise stated, the SP44 promoter, PET RBS and Bba_B0015 were used in constructs, assembled into either pTU1-A (*E. coli*) or pSF-1 (*E. coli* and *Streptomyces* shuttle vector).

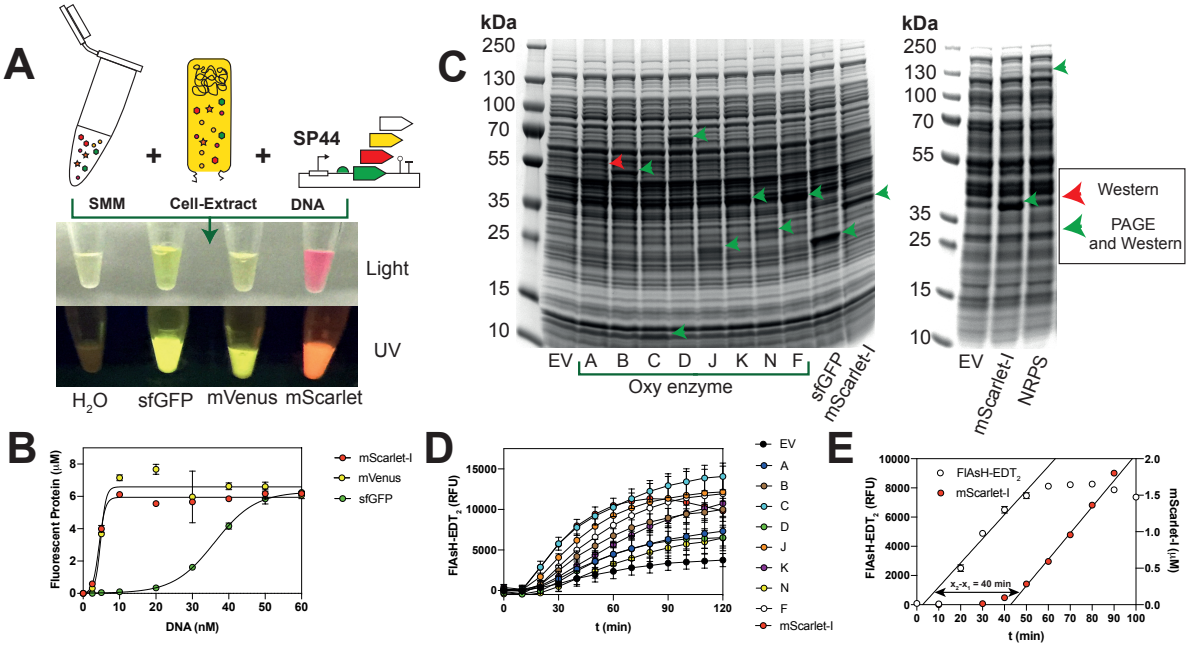

Figure 3. Robust and high-yield synthesis of high G+C (%) genes. (A) Synthesis of codon-optimised fluorescence proteins. (B) Denaturing PAGE of oxytetracycline biosynthetic proteins, fluorescence proteins and a representative NRPS (NH08_RS0107360 from *S. rimosus*). (C) Saturation of protein synthesis for sfGFP, mVenus and mScarlet-I with increasing DNA concentrations. (D) Real-time detection of protein synthesis with C-terminal FlAsH-EDT_2 tag system. (E) Estimation of mScarlet-I maturation time with real-time measurement of immature and mature protein synthesis.

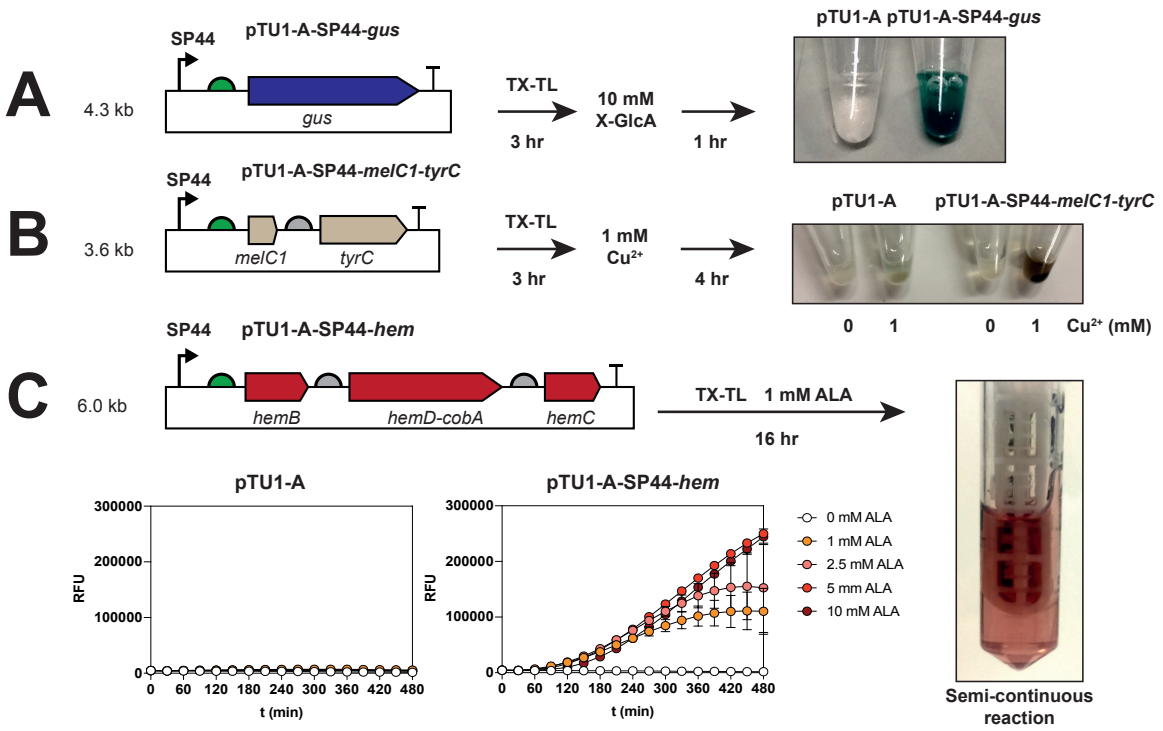

Figure 4. *Streptomyces* cell-free transcription, translation and biosynthesis. (A) Codon-optimised *E. coli* MG1655 GUS enzyme. (B) *S. venezuelae* DSM-40230 tyrosinase (TyrC) and copper metallochaperone (MelC1). (C) *S. venezuelae* DSM-40230 early-stage haem biosynthetic pathway, HemB, HemD-CysG^{A} and HemC. Addition of individual substrates and approximate timescales are indicated within the diagram. For details of batch and semi-continuous reaction conditions, please refer to methods.