nature biotechnology

Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes

Brian F Pfleger[1], Douglas J Pitera[1], Christina D Smolke[1,4], & Jay D Keasling[1–3]

Many applications of synthetic biology require the balanced expression of multiple genes. Although operons facilitate coordinated expression of multiple genes in prokaryotes and eukaryotes, coordinating the many post-transcriptional processes that determine the relative levels of gene expression in operons by *a priori* design remains a challenge. We describe a method for tuning the expression of multiple genes within operons by generating libraries of tunable intergenic regions (TIGRs), recombining various post-transcriptional control elements and screening for the desired relative expression levels. TIGRs can vary the relative expression of two reporter genes over a 100-fold range and balance expression of three genes in an operon that encodes a heterologous mevalonate biosynthetic pathway, resulting in a sevenfold increase in mevalonate production. This technology should be useful for optimizing the expression of multiple genes in synthetic operons, both in prokaryotes and eukaryotes.

The synthesis of natural or unnatural products in microorganisms usually involves the introduction of several genes encoding the enzymes of a metabolic pathway[1,2]. Often, in order to produce these molecules at commercially relevant levels, the genes must be expressed at appropriately balanced levels to avoid the accumulation of toxic intermediates or bottlenecks that result in growth inhibition or suboptimal yields. Similarly, manipulation of multisubunit proteins (for example, F_1F_0-ATPase, proteasomes and ion channels) usually requires coordinated expression of several genes to produce the subunits at the appropriate stoichiometries[3]. Development of tightly regulated, well-characterized molecular components for the construction of optimized biological pathways is a central challenge in synthetic biology[4]. It is nearly impossible to predict the necessary strengths of the promoters and ribosome binding sites (RBSs) required to balance and coordinate the expression of multiple genes.

Grouping multiple, related genes into operons, as is done naturally in prokaryotes[5], is a convenient means for regulating several genes simultaneously without the need for multiple promoters. Internal ribosomal entry sequences (IRESs) from eukaryotic viruses and host stress response pathways perform a similar function and have been harnessed to create operons for heterologous expression of genes in eukaryotes[6–9]. With a single promoter controlling the transcription of several genes, relative expression of each open reading frame in the operon is controlled by altering post-transcriptional processes such as transcription termination[10,11], mRNA stability[12,13] and translation initiation[14–16]. Sequences inserted into the intergenic regions of bacterial operons can direct the processing and segmental stability of a transcript containing multiple coding regions[17,18]. This type of directed mRNA processing results in differential production of the proteins encoded in the operon depending on the nature of the intergenic region between the coding regions.

One of the major obstacles to designing and implementing this type of control is the difficulty in decoupling the many interrelated variables involved in post-transcriptional regulation[19]. We previously demonstrated that it is possible to differentially control the protein levels encoded by two or more genes in an operon using intergenic-region sequences[17,18]. Here, we simultaneously tune the expression of several genes within operons by generating and screening large libraries of TIGRs containing control elements that include mRNA secondary structures, RNase cleavage sites and RBS sequestering sequences. An operon reporter system (**Fig. 1a**) containing the genes encoding the red fluorescent protein DsRed (rfp_{EC})[20,21] and the green fluorescent protein GFP (gfp_{UV}) facilitates high-throughput measurement of relative gene expression resulting from the TIGR libraries.

A large library of TIGR sequences ($>10^4$) was assembled combinatorially from four sets of oligonucleotides (**Supplementary Table 1** online) using PCR. Each oligonucleotide contained two 15-nt sequences that hybridized to a corresponding sequence in the neighboring oligonucleotide, such that a series of chimeric DNA molecules containing oligonucleotides from each of the four sets was created after several rounds of PCR (**Fig. 1b**). Between the hybridization sequences at either end of each oligonucleotide was a variable sequence that provided the diversity of features designed into the library. PCR amplification of this DNA pool with end-specific oligonucleotides enriched the population with full-length TIGRs containing a member from each set of oligonucleotides (**Fig. 1b**). Specific restriction sites incorporated into the amplification primers were used to clone the TIGR library between the two reporter genes.

The TIGR pool that resulted from the assembly of the oligonucleotides was designed to contain three regions, two variable hairpin

[1]Department of Chemical Engineering, University of California Berkeley, California 94720-1462, USA. [2]Department of Bioengineering, University of California Berkeley, California 94720-1762, USA. [3]Synthetic Biology Department, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [4]Present address: Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA. Correspondence should be addressed to J.D.K. (keasling@berkeley.edu).

Received 25 April; accepted 21 May; published online 16 July 2006; doi:10.1038/nbt1226

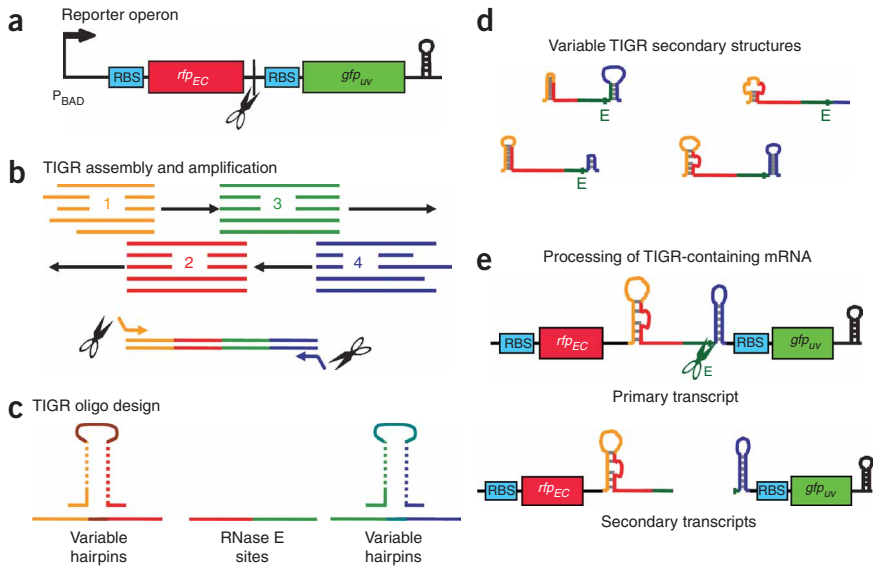

Figure 1 TIGR assembly and reporter operon. (**a**) The reporter plasmid p70RG includes the reporter genes rfp_{EC} and gfp_{UV} downstream of the P_{BAD} promoter. The black scissors indicate the location of the cloning site used to insert the library of TIGRs. (**b**) TIGR assembly reaction. Members of each region (1–4) anneal to members of neighboring regions and are extended by PCR. Eventually full-length TIGRs containing members of each region are assembled and then amplified using end-specific primers containing the restriction sites for cloning. (**c**) The oligonucleotides used for TIGR construction were designed to make three separate regions. 1 and 2 form 5′ hairpins, 2 and 3 form the single-stranded region with RNase E sites and 3 and 4 form a 3′ hairpin. See **Supplementary Methods** for details. (**d**) When transcribed these sequences form a variety of structures incorporating many elements that affect gene expression. (**e**) The TIGR sequence is designed to be processed at a cleavage site between two secondary structures. Cleavage results in two independent secondary transcripts whose stability is determined by the remaining TIGR sequence.

sequences flanking a single-stranded region incorporating various RNase E sites[22,23] (**Fig. 1c**,**d**). When transcribed, those TIGR sequences that contained a strong endonuclease site would be cleaved to generate two secondary transcripts whose stabilities could be individually modulated by the secondary structures flanking the RNase site[17] (**Fig. 1e**). The TIGR sequences also incorporated mRNA secondary structures of various lengths, GC contents, asymmetries and mismatched bulges (**Supplementary Table 2** online). More than 10^4 possible TIGRs were generated from the 9 to 11 oligonucleotides in each of the four sets. Although inclusion of degeneracies in some of the oligonucleotides and the use of error-prone PCR conditions increased the number of possible sequence combinations, the actual size of the library was likely determined by the number of clones obtained after electroporation.

Cells harboring the reporter libraries produced a wide range of fluorescence phenotypes (**Fig. 2a**). Highly fluorescent cells were isolated by fluorescence-activated cell sorting (FACS) to further characterize the library. Cultures of the sorted cells were grown in 96-well plates to determine the fluorescence of each construct after 24 h. The relative fluorescence ratio, red to green, varied by two orders of magnitude (from 45:1 DsRed/GFP to 3:1 GFP/DsRed; **Fig. 2b**), depending on the intergenic region. As expected, the distributions of relative mRNA levels and fluorescence ratios differed, indicating that TIGRs may also affect translation (**Fig. 2c**,**d**). TIGR sequences generally had a stronger influence on the expression of the gene 3′ to the TIGR than on the gene 5′ to the TIGR (**Supplementary Figs. 1** and **2** online).

The operons from fifteen plasmids were sequenced to characterize mechanisms responsible for the resulting expression ratios (**Supplementary Table 3** online). Many of the sequences were predicted to form secondary structures (as predicted by the Mfold web server[24]) with variably sized hairpins and single-stranded regions. The secondary structures and the size distribution of transcripts (assessed from northern blot analysis; **Fig. 3**) suggest that at least three mechanisms are responsible for the observed differences in gene expression in the library. These mechanisms—differential mRNA processing (**Fig. 3a–c**), transcription termination (**Fig. 3d–g**) and RBS sequestration (**Fig. 3h**,**i**)—are discussed below.

Regarding intergenic cleavage and differential processing of secondary mRNAs, the incorporation of RNase sites between the two coding regions of a transcript (**Fig. 3a**) has been shown to decouple the stability of the two coding regions and thus the production of the corresponding proteins[17]. Of the TIGR library samples studied, most had two distinct transcripts with lengths of approximately 1,800 and 800 bases, corresponding to the predicted primary (gfp_{UV} and rfp_{EC} coding regions on one transcript) and secondary (gfp_{UV} or rfp_{EC} coding region alone) transcripts (**Fig. 3a–c**). This indicates processing of the primary transcripts by endonucleases at cleavage sites placed in the TIGR between the coding regions. Cleavage in the TIGR would result in two independent secondary transcripts whose stability, and ultimately the amount of protein produced from them, would be dictated by the remaining TIGR sequences at the 3′- and 5′-ends of the separated transcripts. Differences in the intensities of the primary and secondary transcripts support differential transcript stability (**Fig. 3b**,**c**).

Regarding transcription termination, large ratios in the expression of the first gene relative to that of the second gene could be generated by increasing the frequency of termination before transcription of the 3′ gene. In the TIGR library, the overall distribution of fluorescence ratios was skewed with nearly 70% of the clones displaying more red fluorescence (first gene) than green fluorescence (second gene) (**Fig. 2b**). Premature transcription termination due to large intergenic secondary structures between the coding regions, best demonstrated by sample 1 (**Fig. 3d**), is the most likely explanation for the skewed expression in favor of the first gene. The sample 1 TIGR is predicted to comprise two very large hairpins with over 20 bp in each stem (**Fig. 3e**). Northern blot analyses revealed a large quantity of stable rfp_{EC} transcript corresponding to the size of a single gene (**Fig. 3f**) and very little gfp_{UV} *transcript, either full-length or containing* gfp_{UV} alone (**Fig. 3g** and **Supplementary Fig. 1** online). The strong hairpins in this TIGR may promote premature transcription termination and/or protect against exoribonucleases[25]. Either mechanism would increase the ratio of abundance of the first gene product relative to that of the second gene product (red/green).

Finally, RBS sequestration resulted in a number of samples showing a substantial difference between the relative mRNA and protein levels (**Fig. 2c**,**d** and **Supplementary Fig. 1**). In many of these samples, the gfp_{UV} RBS was part of a secondary structure, such that *cis* base-pairing

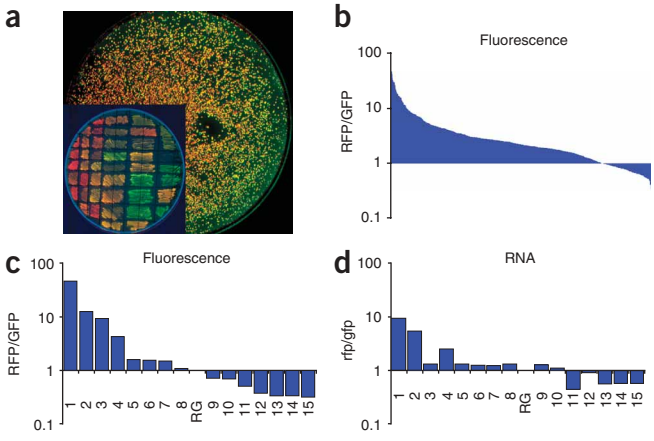

Figure 2 Expression from TIGR library in the original operon vector p70RG. (**a**) Colonies of fluorescent strains imaged using a laser scanner that detects fluorescence at 526 nm and 580 nm. The image is an overlay of the two signals. The inset is a selection of the colonies exposed to UV light at 295 nm to excite GFP fluorescence. (**b**) DsRed:GFP fluorescence ratios of the library after 24 h of growth. (**c**) Fifteen clones were assayed for fluorescence during exponential growth. Shown are DsRed/GFP fluorescence ratios normalized to the fluorescence of p70RG. (**d**) The ratios of rfp_{EC} and gfp_{UV} mRNA from exponentially growing cells were determined by real-time PCR and dot blot hybridization. Shown are the rfp_{EC}:gfp_{UV} mRNA ratios determined by real-time PCR. Note the similarity in the distributions of the two samples, indicating that the TIGR alters mRNA levels that in turn change fluorescence ratios.

of the RBS may have prevented the ribosome from loading onto the transcript. Secondary structures incorporating the RBS reduce the rate of translation initiation[14–16,26]. Sample 3, which best illustrates this mechanism (**Fig. 3h**), consists of four hairpins, the last of which incorporates the RBS at the base of its stem (**Fig. 3i**). Despite the significant levels of gfp_{UV} mRNA present (both full-length and secondary transcripts containing gfp_{UV} only; **Fig. 3c**) there was very little green fluorescence (**Fig. 2c** and **Supplementary Fig. 1** online), suggesting that the ribosome was not able to load onto the RBS 5′ of the second coding region.

After successful demonstration of the ability to control the expression of reporter genes in an operon, we applied the TIGR approach to optimize flux through a heterologous mevalonate pathway, introduced into *Escherichia coli* to produce amorpha-4,11-diene, a precursor to the antimalarial drug artemisinin[2]. The heterologous mevalonate pathway, which comprises three *Saccharomyces cerevisiae* genes, transforms acetyl-CoA into the isoprenoid precursors, isopentenyl pyrophosphate and dimethylallyl pyrophosphate (**Fig. 4a**). Although an operon consisting of the first three genes in the mevalonate pathway (*atoB*, *HMGS*, t*HMGR*; referred to collectively as MevT; **Fig. 4b**) appears to limit flux, overexpression of this operon reduced both growth and product formation, potentially owing to the toxicity of imbalanced gene expression (D.J.P., unpublished data). Directed approaches to remedy this imbalance would involve the difficult and time-consuming construction and analysis of many clones, using various approaches to alter gene copy number and/or rates of transcriptional or translational initiation.

Instead, application of the combinatorial approach described above reduced pathway optimization time and increased the number and variety of mechanisms explored to balance the MevT pathway. Specifically, the TIGR libraries were used to generate a series of operons that were subsequently screened for increased mevalonate production. A megaprimer PCR approach (**Supplementary Fig. 3** online)[27] was used to simultaneously place TIGR libraries between the first and second genes and between the second and third genes of the MevT operon (**Fig. 4c**). Functional operons from the libraries were selected by transforming the plasmid library into *E. coli* DP5, a strain engineered to be auxotrophic for mevalonate (B.F.P., D.J.P., J.D. Newman, V.J.J. Martin & J.D.K., data not shown). The plasmids isolated from this enriched pool were recovered, transformed into a production strain (DH10B) and screened to identify the highest-producing version by using a mevalonate auxotroph transformed with a plasmid harboring a constitutively expressed *gfp* as a biosensor for mevalonate (B.F.P., D.J.P., J.D. Newman, V.J.J. Martin & J.D.K., data

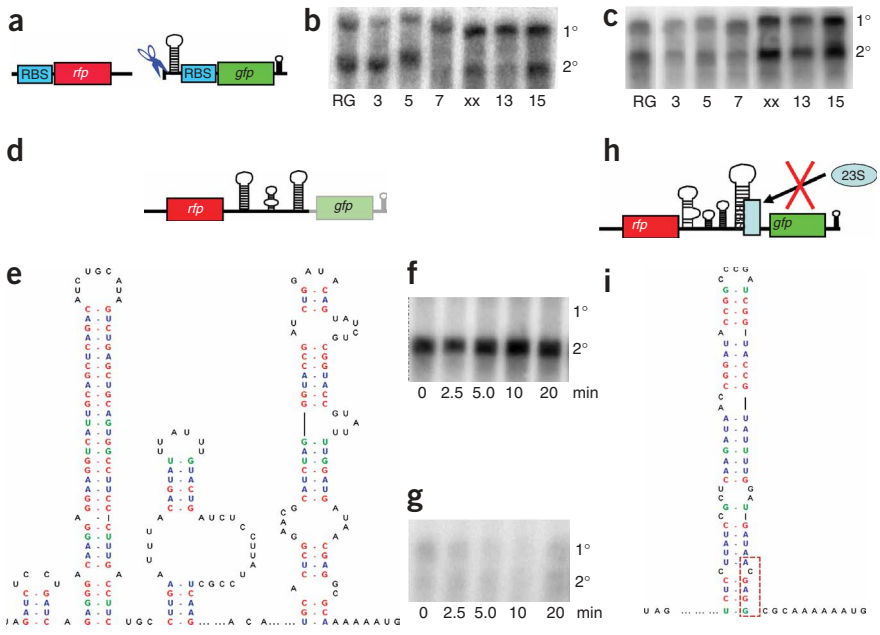

Figure 3 TIGR effects on expression. (**a**) Differential RNase protection mechanism of TIGR samples. Cleavage of the TIGR generates two secondary transcripts. (**b**) Northern blot of total RNA from exponentially growing cultures of numbered clones probed for rfp_{EC}. (**c**) Northern blots of the same RNA in **b** probed for gfp_{UV}. (**d**) Premature transcription termination mechanism of sample 1. The large intergenic-region structure prevents transcription of the gfp_{UV} gene. (**e**) The predicted secondary structure of sample 1's TIGR contains two large hairpins. (**f**) rfp_{EC}-probed northern blot of total RNA harvested at various time points (bottom) after transcription was stopped by adding rifampicin. (**g**) Northern blots of the same RNA in **f** probed for gfp_{UV}. The primary (1°) and secondary (2°) transcripts are labeled accordingly. (**h**) RBS sequestration mechanism of sample 3. (**i**) Predicted secondary structure of the last hairpin of sample 3's TIGR. Note the location of the RBS at the base of the hairpin (dashed box). Despite the gfp_{UV} mRNA present in (**c**, sample 3), there was less green fluorescence relative to p70RG than expected (**Fig. 2c**,**d** and **Supplementary Fig. 1** online).

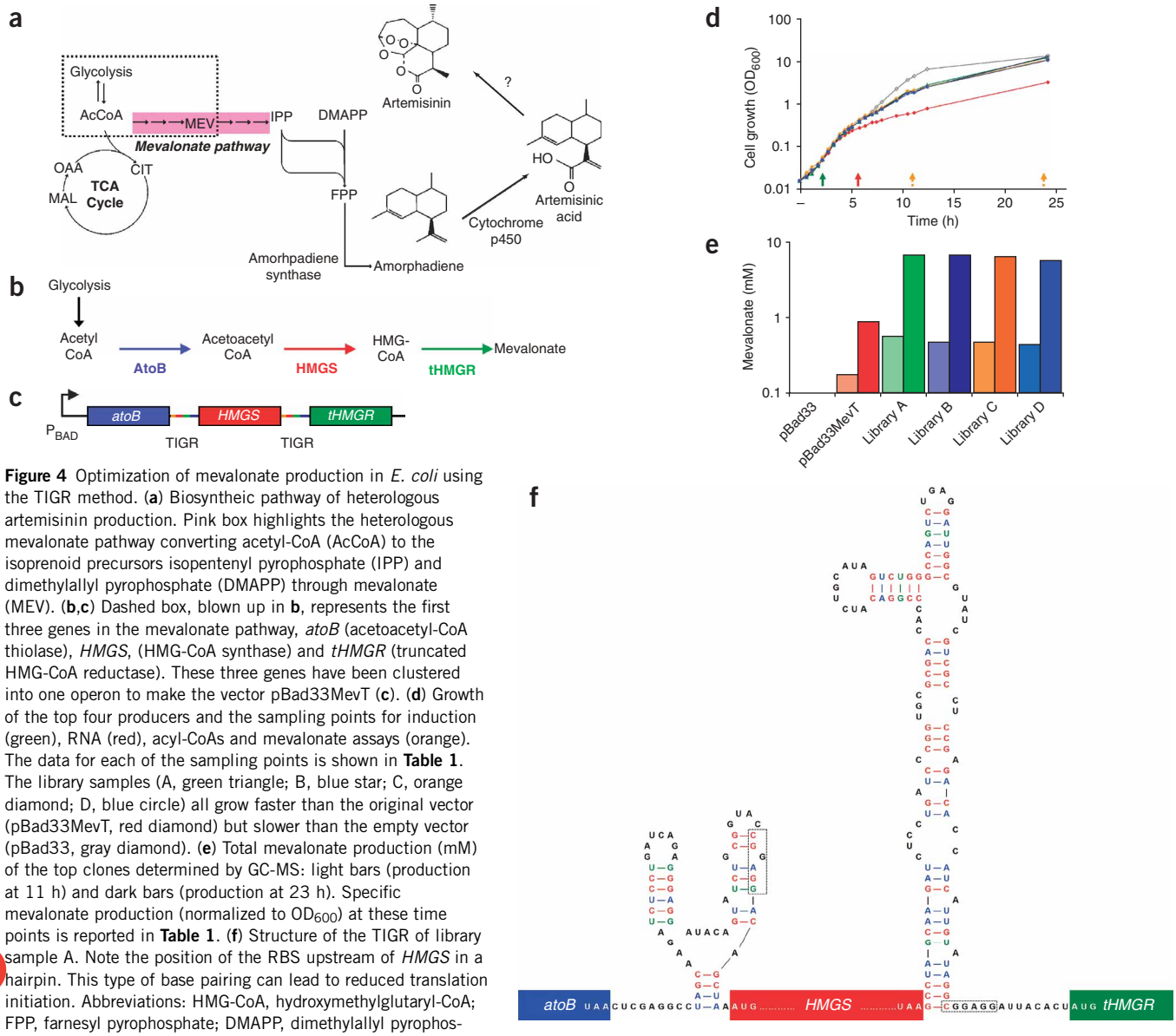

Figure 4 Optimization of mevalonate production in *E. coli* using the TIGR method. (**a**) Biosyntheic pathway of heterologous artemisinin production. Pink box highlights the heterologous mevalonate pathway converting acetyl-CoA (AcCoA) to the isoprenoid precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) through mevalonate (MEV). (**b**,**c**) Dashed box, blown up in **b**, represents the first three genes in the mevalonate pathway, *atoB* (acetoacetyl-CoA thiolase), *HMGS*, (HMG-CoA synthase) and *tHMGR* (truncated HMG-CoA reductase). These three genes have been clustered into one operon to make the vector pBad33MevT (**c**). (**d**) Growth of the top four producers and the sampling points for induction (green), RNA (red), acyl-CoAs and mevalonate assays (orange). The data for each of the sampling points is shown in **Table 1**. The library samples (A, green triangle; B, blue star; C, orange diamond; D, blue circle) all grow faster than the original vector (pBad33MevT, red diamond) but slower than the empty vector (pBad33, gray diamond). (**e**) Total mevalonate production (mM) of the top clones determined by GC-MS: light bars (production at 11 h) and dark bars (production at 23 h). Specific mevalonate production (normalized to OD_{600}) at these time points is reported in **Table 1**. (**f**) Structure of the TIGR of library sample A. Note the position of the RBS upstream of *HMGS* in a hairpin. This type of base pairing can lead to reduced translation initiation. Abbreviations: HMG-CoA, hydroxymethylglutaryl-CoA; FPP, farnesyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; CIT, citrate: MAL, Malate; OAA, oxaloacetate.

not shown). This fluorescence-based screen incorporates a GFP-producing sensor strain whose growth is dependent on the level of mevalonate secreted into the surrounding medium by the producer strain. Of the more than 600 colonies screened in this manner, 18% were significantly (more than three standard deviations greater, calculated from two sets of triplicate controls on each plate) more fluorescent than the original operon (control without TIGRs).

The four best producers were examined to determine the mechanism responsible for the increased mevalonate production. The most obvious difference between the library samples and the pBad33MevT control was the improved growth of the library strains. Cultures of uninduced DH10B harboring pBad33 (empty vector) or pBad33MevT (original pathway, no TIGRs) grew similarly. However, once induced, cultures harboring pBad33MevT grew much slower than those harboring pBad33. Although the final culture densities were similar, the empty vector reached stationary phase nearly 12 h earlier (**Fig. 4d**). Uninduced cultures of DH10B harboring one of each of the library plasmids grew similarly to the other uninduced strains. However, once induced, they grew faster than pBad33MevT and slower than pBad33. Mevalonate production from the library strains was sevenfold greater than that from pBad33MevT, averaging 6 mM after 24 h (**Fig. 4e**). The specific production (**Table 1**) of the library strains was twice that of the pBad33MevT control, indicating that both the improved growth and increased flux through the pathway contributed to the increased production.

The concentrations of intracellular acyl-CoA species were measured to determine if the TIGR changed the relative carbon flux through any of the three steps in the pathway. The biggest difference between the strains was the concentration of acetyl-CoA, a central metabolic intermediate (**Table 1**). The strain containing the original mevalonate pathway (pBAD33MevT) had much lower acetyl-CoA levels than the control strain and the strains containing the evolved pathway. The fact that the strains producing more mevalonate had more acetyl-CoA was unexpected but correlates with the improved growth seen in the

library samples. The other acyl-CoAs in the pathway (acetoacetyl-CoA, HMG-CoA) were not substantially changed among the strains expressing mevalonate pathway genes. The data suggest that the improved production was due primarily to increased growth correlated with the availability of acetyl-CoA.

Transcripts from each strain were examined using dot blots to determine whether changes in TIGR sequence altered the relative expression levels of the three genes. The *HMGS* and *tHMGR* transcripts were noticeably reduced in the library strains in comparison to the transcript expressed from pBad33MevT (**Table 1**). Corroborating the mRNA data, the tHMGR enzyme activity was substantially reduced in library sample A compared to pBad33MevT (**Table 1**). The reduction in expression was not expected, but suggests that reduced expression of *HMGS* and *tHMGR* was responsible for the improved growth and increased mevalonate production.

The sequences of the TIGRs (**Supplementary Table 3**) from each of the top four producers were predicted to form a small set of similar structures. The TIGR between *HMGS* and *tHMGR* was predicted to form either a large hairpin or a short single-stranded structure, whereas the TIGR between *atoB* and *HMGS* contained two small hairpins separated by a few single-stranded bases (**Fig. 4e**). The second of these hairpins appears to sequester the RBS, thereby downregulating production of HMGS. Reduced translation rates can in turn affect mRNA stability[19], and therefore message levels, by decreasing transcript coverage by ribosomes and leaving the transcript open to RNase degradation. Therefore, reduced HMGS translation rates could lead to the reduction in *HMGS* and *tHMGR* message, which would correlate with the improved growth and increased mevalonate production.

The ability to access multiple regulatory mechanisms simultaneously without specific design is a major strength of the TIGR approach. As the mechanisms that underlie the sevenfold increase in mevalonate production associated with the improved version of pBad33MevT are counterintuitive, reduction of HMGS and tHMGR activities would not have been a priority in construct design. Most importantly, incorporation of our optimized operon into the prokaryotic strain engineered to produce artemisinin or its precursor (amorphadiene) may reduce the cost of drug production using this pathway. Similarly, use of TIGRs containing IRES elements and RNase sites may also be valuable for designing operons used in pathway engineering, multisubunit protein production and gene therapy in eukaryotes.

Table 1 Expression analysis of select mevalonate producers

		pBad33	pBad33MevT	Sample A	Sample B	Sample C	Sample D
mRNA levels (ng)							
atoB		1.5	6.0	5.7	6.8	6.1	8.6
hmgs		0.2	4.3	2.3	2.1	2.2	3.5
thmgr		0.3	6.3	3.1	3.9	3.0	4.0
Enzyme activity (μmol/min/mg protein)							
tHMGR		<0	3,000	1,100	ND	ND	ND
Time (h)	Metabolite concentrations (nM normalized to OD)						
5	CoA	292	326	310	310	284	315
	Acetyl CoA	826	<0	126	123	184	129
	AA-CoA	24.4	14.6	19.2	19.8	18.2	17.6
	HMG-CoA	7.6	19.9	37.2	23.4	37.8	25.7
	Mevalonate	<0	214,000	172,000	ND	151,000	177,000
11	CoA	176	104	283	354	208	288
	Acetyl CoA	627	47	774	678	587	670
	AA-CoA	0.8	7.4	2.0	2.6	2.1	2.1
	HMG-CoA	0.4	20.7	21.9	19.3	12.8	9.7
	Mevalonate	<0	283,000	279,000	243,000	222,000	240,000
24	Mevalonate	<0	276,000	565,000	524,000	596,000	534,000

METHODS

General. Media and chemicals were purchased from Sigma and Fisher Scientific. Enzymes were purchased from Roche, Promega and New England Biolabs. Oligonucleotides (**Supplementary Table 1**) were purchased from Operon Biotechnologies. C-medium[17] was supplemented with 3.4% glycerol, 1% Casamino acids (Difco) and micronutrients. Carbenecillin and chloramphenicol were used at concentrations of 50 μg/ml and 34 μg/ml, respectively, and cells were grown at 37 °C, unless otherwise indicated. The strains and plasmids used in this study are listed in **Supplementary Table 4**.

Assembly of TIGR libraries. Cloning of the initial reporter vectors and mevalonate pathway constructs are described in the **Supplementary Methods**. TIGRs were synthesized using PCR to assemble oligonucleotides into chimeric DNA sequences. Four-hundred (400) picomoles of an equimolar oligonucleotide mixture were added to a mixture containing 2.5 units of AmpliTaq Gold polymerase (Applied Biosystems). The assembly was conducted over 35 rounds of 15 s at 95 °C, 30 s at 72 °C and 20 + 5 s/cycle at 72 °C. The resulting assembly products were purified with a nucleotide removal column (Qiagen) and amplified using end-specific primers containing *Bgl*II and *Asp*718 restriction sites. The amplified libraries were subcloned into p70RG, the ligation products were electroporated into *E. coli* DH10B, and the resulting transformants were plated on LB agar with carbenecillin.

Reporter library screening. The p70RG library transformants were collected from agar plates in 5 ml of PBS. The cell suspensions were diluted to 10^7 cells/ml before screening through FACS on a Beckman-Coulter EPICS Elite Sorter. The extreme 30% representing highly red, highly green, and highly green and red cells were separated from the remaining cell population. Ten million events were collected and resorted to remove any undesired cells that were carried over in the initial sort. The sorted populations were plated on LB agar with carbenecillin and individual colonies were subsequently grown overnight in C-medium. Cultures were back diluted 1:100 into fresh C-medium with carbenecillin and 0.2% arabinose. After growing for 24 h, the OD_{600} and GFP and DsRed fluorescence were measured using a Tecan Safire plate reader. GFP and DsRed were measured at excitation/emission wavelengths of 400 nm/510 nm and 558 nm/583 nm, respectively. Each fluorescence value was normalized to the number of cells by dividing by the OD_{600}.

Fifteen selected members of the intergenic-region library were grown overnight in LB medium with carbenicillin and inoculated into C medium with carbenicillin to an OD_{600} of 0.016. At an OD_{600} of 0.05, the cultures were induced with 0.2% arabinose. At an OD_{600} of 0.4, fluorescence and mRNA levels were determined. These values were normalized to the values generated by the p70RG control operon and presented above (**Fig. 2c**,**d**).

RNA methods. Messenger RNA analysis was performed by dot-blot hybridization, northern blot hybridization and real-time PCR. Details for these methods are found in the **Supplementary Methods**. Briefly, total RNA was isolated using a RiboPure-Bacteria kit (Ambion) and quantified on a Bioanalyzer Total RNA Nanochip (Agilent Technologies). Dot and northern blots were generated according to standard protocols[28]. The construction of probe templates is described in the **Supplementary Methods**. Probes were synthesized by *in vitro*

transcription from these gel extracted templates with SP6 RNA polymerase (Promega) in the presence of [^{32}P]-labeled α-CTP (PerkinElmer) and unlabeled nucleotide triphosphates (Promega) according to the manufacturer's instructions. All probes are specific for their own genes and did not generate any cross reactivity to the other transcripts.

Megaprimer construction of TIGR-HMGS-TIGR libraries. Construction of intergenic-region libraries between three genes was performed using a megaprimer reaction[27]. Previously amplified libraries were reamplified with one of two primer sets: MevT-A, MevT-B or MevT-C, MevT-D in a standard PCR. These primers contain either a restriction site for subsequent cloning or a 5′ tail complementary to *HMGS*. The products of these reactions were purified and used as megaprimers to amplify *HMGS* in a subsequent PCR (**Supplementary Fig. 3**). The resulting fragments had the form *Xho*I–IGR–HMGS–IGR–*Not*I. These fragments were cloned into pBad33MevT-L, a vector containing the remaining genes of the MevT pathway, *AtoB* and t*HMGR*. The ligations were transformed into competent DH10B (Invitrogen) and plated onto LB agar plates containing chloramphenicol and 0.1% glucose. Transformants were pooled as above and their plasmids isolated. Functional operons were selected by transforming the resulting plasmid pool into *E. coli* DP5, a mevalonate auxotroph (**Supplementary Table 4**, and B.F.P., D.J.P., J.D. Newman, V.J.J. Martin & J.D.K., unpublished data) and plating on the appropriate media. The plasmids were then isolated from pooled cultures of the surviving colonies and transformed into DH10B for further screening.

Biosensor screening of mevalonate producing libraries. Colonies containing functional operons were transferred into 96-well plates and grown overnight in C-medium with chloramphenicol and 0.1% glucose. Cultures were back diluted 1:100 into fresh C-medium with chloramphenicol and 0.2% arabinose. After 24 h, the cells were pelleted and the spent medium collected. A culture of the biosensor cells was grown overnight in C-medium with 50 μg/l kanamycin and 1 mM mevalonate and back diluted to an OD_{600} of 0.02. One-hundred ninety (190) μl of this culture was combined with either 10 μl of the spent media or 10 μl of a 1:10 dilution of the spent media in separate screening plates. Two wells per plate were run in triplicate as internal controls. Mevalonate controls between 100 μM to 2 mM and the highest mevalonate producer were run on each plate. The biosensor plates were grown for 48 h, during which the GFP fluorescence was periodically measured using a Typhoon laser scanner. Samples were clustered into groups based on their relative fluorescence compared to the best mevalonate producer and average standard deviation from the triplicate wells. The mevalonate producing cells corresponding to the highest fluorescent biosensor wells were subjected to further analysis.

MevT expression analysis. High mevalonate–producing library members were assayed for cell growth, mRNA levels, enzyme activity, intracellular acyl-coenzyme A levels and mevalonate concentrations. Cultures were inoculated to an OD_{600} of 0.016 from glucose-repressed overnight cultures, grown to an OD_{600} of 0.05 and induced with 0.2% arabinose. Dot blots and northern blots were prepared in duplicate from total RNA isolated at an OD_{600} of 0.4. Blots were probed with appropriate radiolabeled probes generated as described in the **Supplementary Methods** and **Supplementary Table 1**. tHMGR protein levels were assayed enzymatically by monitoring the disappearance of NADPH (tHMGR cofactor) by measuring the absorbance at 340 nm[29]. Mevalonate levels were determined by gas chromatography–mass spectrometry (GC-MS) as described in the **Supplementary Methods**. Acyl-CoA levels were determined by liquid chromatography–mass spectrometry (LC-MS) analysis of cell extracts described in the **Supplementary Methods**.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

The authors would like to acknowledge Vincent J.J. Martin, Jack D. Newman, Katherine D. McMahon, Sydnor T. Withers and Wesley D. Marner II for their constructive comments, Hector Nolla for performing FACS experiments, and Chris J. Paddon for performing the HMGR assays. This research was conducted under the sponsorship of the Institute for OneWorld Health, through the generous support of The Bill and Melinda Gates Foundation and by National Science Foundation Grant no. BES 9906405. D.J.P. is the recipient of a National Science Foundation graduate fellowship.

AUTHOR CONTRIBUTIONS

B.F.P. designed and conducted experiments; D.J.P. designed and conducted experiments; C.D.S. designed experiments; J.D.K. designed experiments.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Khosla, C. & Keasling, J.D. Metabolic engineering for drug discovery and development. *Nat. Rev. Drug Discov.* **2**, 1019–1025 (2003).
2. Martin, V.J., Pitera, D.J., Withers, S.T., Newman, J.D. & Keasling, J.D. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **21**, 796–802 (2003).
3. White, M.M. Pretty subunits all in a row: using concatenated subunit constructs to force the expression of receptors with defined stoichiometry and spatial arrangement. *Mol. Pharmacol.* **69**, 407–410 (2006).
4. Endy, D. Foundations for engineering biology. *Nature* **438**, 449–453 (2005).
5. Baga, M., Goransson, M., Normark, S. & Uhlin, B.E. Processed mRNA with differential stability in the regulation of *E. coli* pilin gene expression. *Cell* **52**, 197–206 (1988).
6. Komar, A.A. & Hatzoglou, M. Internal ribosome entry sites in cellular mRNAs: mystery of their existence. *J. Biol. Chem.* **280**, 23425–23428 (2005).
7. Martin, P., Albagli, O., Poggi, M.C., Boulukos, K.E. & Pognonec, P. Development of a new bicistronic retroviral vector with strong IRES activity. *BMC Biotechnol.* **6**, 4 (2006).
8. Chappell, S.A. & Mauro, V.P. The internal ribosome entry site (IRES) contained within the RNA-binding motif protein 3 (*Rbm3*) mRNA is composed of functionally distinct elements. *J. Biol. Chem.* **278**, 33793–33800 (2003).
9. Fernandez-Miragall, O., Ramos, R., Ramajo, J. & Martinez-Salas, E. Evidence of reciprocal tertiary interactions between conserved motifs involved in organizing RNA structure essential for internal initiation of translation. *RNA* **12**, 223–234 (2006).
10. Nudler, E. & Gottesman, M.E. Transcription termination and anti-termination in *E. coli*. *Genes Cells* **7**, 755–768 (2002).
11. Mandal, M. & Breaker, R.R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**, 451–463 (2004).
12. Arraiano, C.M. & Maquat, L.E. Post-transcriptional control of gene expression: effectors of mRNA decay. *Mol. Microbiol.* **49**, 267–276 (2003).
13. Kushner, S.R. mRNA decay in prokaryotes and eukaryotes: different approaches to a similar problem. *IUBMB Life* **56**, 585–594 (2004).
14. de Smit, M.H. & van Duin, J. Control of translation by mRNA secondary structure in *Escherichia coli*. A quantitative analysis of literature data. *J. Mol. Biol.* **244**, 144–150 (1994).
15. Isaacs, F.J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* **22**, 841–847 (2004).
16. Majdalani, N., Vanderpool, C.K. & Gottesman, S. Bacterial small RNA regulators. *Crit. Rev. Biochem. Mol. Biol.* **40**, 93–113 (2005).
17. Smolke, C.D., Carrier, T.A. & Keasling, J.D. Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. *Appl. Environ. Microbiol.* **66**, 5399–5405 (2000).
18. Smolke, C.D. & Keasling, J.D. Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon. *Biotechnol. Bioeng.* **80**, 762–776 (2002).
19. Deana, A. & Belasco, J.G. Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev.* **19**, 2526–2533 (2005).
20. Matz, M.V. *et al.* Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.* **17**, 969–973 (1999).
21. Pfleger, B.F., Fawzi, N.J. & Keasling, J.D. Optimization of DsRed production in *Escherichia coli*: effect of ribosome binding site sequestration on translation efficiency. *Biotechnol. Bioeng.* **92**, 553–558 (2005).
22. Naureckiene, S. & Uhlin, B.E. In vitro analysis of mRNA processing by RNase E in the pap operon of *Escherichia coli*. *Mol. Microbiol.* **21**, 55–68 (1996).
23. Kaberdin, V.R. Probing the substrate specificity of *Escherichia coli* RNase E using a novel oligonucleotide-based assay. *Nucleic Acids Res.* **31**, 4710–4716 (2003).
24. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
25. Spickler, C. & Mackie, G.A. Action of RNase II and polynucleotide phosphorylase against RNAs containing stem-loops of defined structure. *J. Bacteriol.* **182**, 2422–2427 (2000).
26. Gottesman, S. The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu. Rev. Microbiol.* **58**, 303–328 (2004).
27. Burke, E. & Barik, S. Megaprimer PCR: application in mutagenesis and gene fusion. *Methods Mol. Biol.* **226**, 525–532 (2003).
28. Sambrook, J., Fritsch, E.F. & Maniatis, T.. *Molecular Cloning A Laboratory Manual.* (Cold Spring Harbor Press, Cold Spring Harbor, NY, 1989).
29. Rodwell, V.W. *et al.* 3-Hydroxy-3-methylglutaryl-CoA reductase. *Methods Enzymol.* **324**, 259–280 (2000).