ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

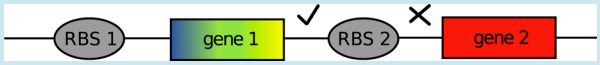
Fluorescent Proteins and *in Vitro* Genetic Organization for Cell-Free Synthetic Biology

Roberta Lentini, Michele Forlin, Laura Martini, Cristina Del Bianco, Amy C. Spencer, Domenica Torino, and Sheref S. Mansy*

CIBIO, University of Trento, via delle Regole 101, 38123 Mattarello (TN), Italy

Supporting Information

ABSTRACT: To facilitate the construction of cell-free genetic devices, we evaluated the ability of 17 different fluorescent proteins to give easily detectable fluorescence signals in real-time from *in vitro* transcription-translation reactions with a minimal system consisting of T7 RNA polymerase and *E. coli* translation machinery, i.e., the PUREsystem. The data were used to construct a ratiometric fluorescence assay to quantify the effect of genetic organization on *in vitro* expression levels. Synthetic operons with varied spacing and sequence composition between two genes that coded for fluorescent proteins were then assembled. The resulting data indicated which restriction sites and where the restriction sites should be placed in order to build genetic devices in a manner that does not interfere with protein expression. Other simple design rules were identified, such as the spacing and sequence composition influences of regions upstream and downstream of ribosome binding sites and the ability of non-AUG start codons to function *in vitro*.



KEYWORDS: *cell-free, fluorescent protein, transcription-translation, ribosome binding site, synthetic biology*

The majority of synthetic biology research makes use of a living chassis that provides for the necessary but poorly characterized biological components required for life. Conversely, a smaller community of synthetic biologists has begun to build cell-like systems with a nonliving, cell-free chassis.[1−7] Although the cell-free branch of synthetic biology has progressed more slowly, success could provide for new technologies with several beneficial features. For example, the resulting cellular mimics would consist of fully defined components. Therefore, it should be possible to build a complete mathematical model describing the cellular mimic that could aid in designing new features. Additionally, potentially technologically problematic features of life, such as evolution, could be intentionally removed by building systems that do not replicate.

A significant step forward in allowing for the construction of such well-defined, bottom-up systems came from Ueda and colleagues, who showed that coupled transcription and translation reactions can be mediated by fully defined components *in vitro*.[8] Their system, hereafter referred to as the PUREsystem, consisted of T7 RNA polymerase and *Escherichia coli* translation machinery. Subsequent work demonstrated the compatibility of the PUREsystem with liposomes[9,10] and with the expression of gene networks.[11] Nevertheless, there has been little attempt to better define the influences of genetic organization on protein output with purified transcription-translation machinery. Recently, a S30 *E. coli* cell extract translation system and the PUREsystem were used to determine the influences of different ribosome binding sites and transcriptional repressors on the synthesis of eGFP.[12,13]

Although much is known about natural, *in vivo* genetics, much still remains unresolved. For example, the refactoring of the T7 genome was successful in the sense that viable bacteriophage were produced; however, the refactored bacteriophage was significantly less infective.[14] Similar challenges are routinely encountered when genetic elements are inserted into organisms to engineer new circuitry. Typically, many permutations are required before desired function is achieved.[15] The situation is perhaps even more challenging for systems that exploit a cell-free chassis since biological parts are evolved to function optimally under the chemical conditions found *in vivo*. *In vitro* conditions are undoubtedly different. Further, unidentified molecular components necessary for activity *in vivo* may be missing from *in vitro* constructions. The design and implementation of predictable, genetically encoded cell-free systems is difficult because of the lack of cell-free chassis data coupled with an incomplete understanding of natural, *in vivo* genetics.

Here we sought to identify some practical rules for the construction of genetically encoded, cell-free systems. First, 17 different fluorescent proteins were screened for their ability to generate easily detectable fluorescence signals after *in vitro* transcription and translation with the PUREsystem. Fluorescent proteins then were expressed from a bicistronic construct to identify fluorescent protein pairs that could be used to quantify the influences of genetic organization on protein production. A series of synthetic operons that differed in the spacing and sequence between the two encoded genes, the spacing and sequence between the ribosome binding site and the start codon, and the influence of the first nucleotide position of the start codon on *in vitro* expression levels was assessed with the developed ratiometric fluorescence assay. We

Received: January 17, 2013
Published: March 8, 2013

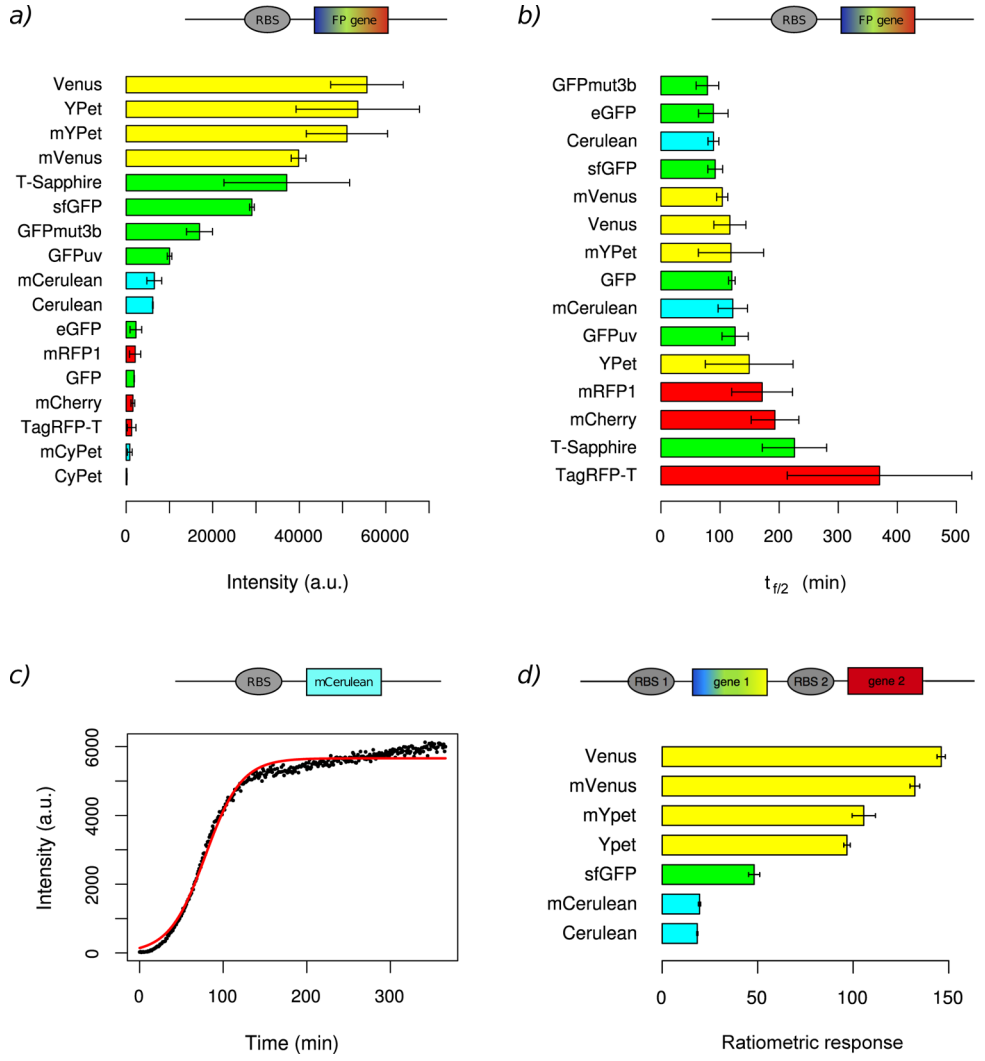


Figure 1. Fluorescence profiles of *in vitro* expressed genetic constructs at 37 °C with the PUREsystem. (a) Fluorescence intensities after 6 h of *in vitro* expression for 17 different fluorescent proteins. (b) The $t_{f/2}$ of each fluorescent protein was calculated by fitting the kinetic data to a logistic model as described in the Methods section. The $t_{f/2}$ represents the time at maximum growth. (c) The fitting of mCerulean kinetic data is shown as a representative example. The logistic model estimation is shown in red, while the black points represent measured values. A control reaction without plasmid showed no fluorescence. (d) The ratiometric response of bicistronic constructs after 6 h of *in vitro* expression. The ratiometric response was calculated by dividing the fluorescence arising from the protein encoded by gene 1 by the fluorescence resulting from the gene product of gene 2. Here gene 2 always encoded mCherry. A cartoon above each panel gives a schematic representation of the used constructs. The data shown in panels a and b are from constructs RL001A-RL013A and CD100A-CD103A. Panel c used RL005A, and panel d used RL015A-RL021A. More information on each construct is provided in Supplementary Table S1.

found that a high guanosine content inhibited translation, that sequences 5′ to the ribosome binding site were more amenable to the incorporation of restriction sites for cloning, and that ribosome binding sites were most efficient when separated from the start codon by 4−9 nucleotide positions. GUG, UUG, and CUG were functional as start codons in minimal, reconstituted translation systems, although their associated expression levels were significantly reduced.

RESULTS AND DISCUSSION

***In Vitro* Expression of Fluorescent Proteins.** A total of 17 different fluorescent proteins were tested individually for their ability to give easily detectable fluorescence signals from *in vitro* transcription-translation reactions with the PUREsystem at 37 °C. Of these 17 proteins, four (mCerulean, mCyPet, mVenus, and mYPet) contained a A206K substitution to inhibit dimerization. As seen in Figure 1a, all of the tested constructs produced easily detectable signals above background arising from the fluorescent protein except for CyPet and mCyPet. These two cyan fluorescent proteins gave slightly increased fluorescence when expressed at 30 °C (Supplementary Figure S1). Consistent with the reported brightness of each fluorescent protein,[16] the yellow fluorescent proteins were associated with the most intense fluorescence, followed by the green, cyan, and red fluorescent proteins (Figure 1a). Monomeric versions of Cerulean and YPet gave fluorescence intensities within 5% of their dimeric parent construct. *In vitro* transcribed and translated Venus was 40% more intense than mVenus; however, the error associated with the single fluorescent protein measurements was too large to make meaningful conclusions. This issue was resolved by using a ratiometric method described below. After 6 h of *in vitro* transcription-translation, the mVenus concentration reached 8 μM.

dx.doi.org/10.1021/sb400003y | *ACS Synth. Biol.* 2013, 2, 482−489

Most of the constructs gave sigmoidal shaped kinetic profiles and were complete within 6 h. The exceptions were T-Sapphire and TagRFP-T (Supplementary Figure S2), both of which did not reach their maximal fluorescence within 6 h. The fitting of the kinetic data to a logistic model was used to determine the time point at which the rate of fluorescence increase was maximal, which corresponded to the time required to reach half maximal fluorescence ($t_{f/2}$) (Figure 1b,c). Note that the $t_{f/2}$ includes all of the steps involved in converting the information encoded in DNA to a fluorescence signal and does not solely describe the last oxidation step of chromophore formation.[17] The shortest $t_{f/2}$ value was 79 min for GFPmut3b, and the longest $t_{f/2}$ was over 300 min for TagRFP-T (Supplementary Table S4). The average $t_{f/2}$ values for the expression of cyan, green, yellow, and red fluorescent proteins were 105, 122, 122, and 245 min, respectively. The $t_{f/2}$ was 40% larger for mCerulean than Cerulean, whereas mVenus and mYPet had $t_{f/2}$ values 12% and 26% smaller than Venus and YPet, respectively. On the basis of fluorescence intensity and kinetic data, Cerulean, mCerulean, super folder GFP (sfGFP), Venus, mVenus, YPet, mYPet, mRFP1, and mCherry were selected for further analysis.

To reduce experimental error, we pursued the construction of a ratiometric fluorescence system based on synthetic operons that encoded two fluorescent proteins. In this way the influences of pipetting, lamp performance, and DNA template quality and concentration, among other difficult to control variables, would be removed. To build such a ratiometric system, a red fluorescent protein was desirable because the excitation and emission spectra of red fluorescent proteins are better separated from the fluorescence spectra of other fluorescent proteins. mRFP1 and mCherry were, therefore, tested in bicistronic constructs that additionally encoded sfGFP to evaluate their utility in characterizing expression levels. More specifically, small synthetic operons containing a standard T7 transcriptional promoter, a ribosome binding site (RBS), a gene encoding sfGFP followed by a sequence that encoded the red fluorescent protein and a T7 transcriptional terminator were assembled. All of the fluorescent proteins in these constructs gave reproducible and easily detectable fluorescence signals. After 6 h of expression with purified transcription-translation machinery, the ratio of sfGFP fluorescence to mRFP1 and to mCherry fluorescence was 115.1 ± 6.9 and 49.9 ± 2.4, respectively (Supplementary Figure S3). We chose to use mCherry for the remaining experiments, because mCherry showed more intense fluorescence from the bicistronic construct and because mCherry was shown to be more photostable than mRFP1.[16]

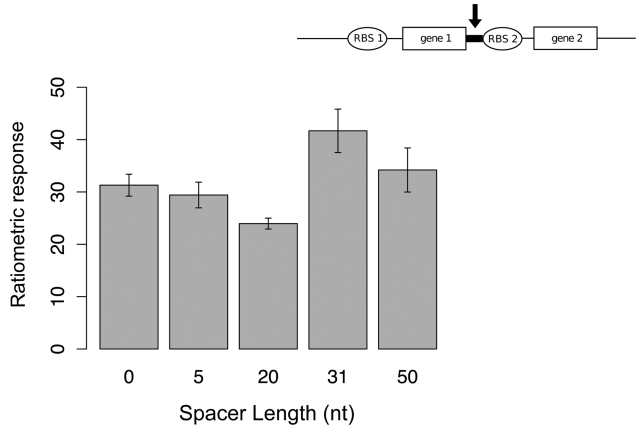
We next assembled six additional synthetic operons that encoded different fluorescent proteins followed by a sequence coding for mCherry. After *in vitro* transcription and translation, the fluorescence profiles were similar to those obtained with the single fluorescent protein constructs in that the yellow fluorescent proteins were the most intense, followed by green, and cyan fluorescent proteins (Figure 1d). However, the error of each ratiometric measurement was significantly reduced (relative standard error <8%) in comparison to the data obtained from the monocistronic, single fluorescent protein constructs (relative standard error <60%, excluding TagRFP-T). The A206K substitution that inhibits protein dimerization had a small effect on fluorescence intensity. More specifically, the ratiometric response, i.e., the fluorescence intensity of the fluorescent protein tested divided by the fluorescence intensity of mCherry, for mVenus, mYPet, and mCerulean were within 10% of the values measured for Venus, YPet, and Cerulean, respectively. The ratiometric response over time showed that stable readings could be taken after 3 h for all constructs tested (Supplementary Figure S4).

It was not clear from the outset which fluorescent proteins would perform well *in vitro* with minimal transcription-translation machinery. Although the physical characteristics of individually purified proteins, such as brightness and photostability, are useful in deciding if a protein could be suitable for a specific application, these parameters are not enough to understand if *in vitro* expression will give a robust, reproducible signal. For example, if *in vitro* produced protein is insoluble, folds slowly, or requires a long period of time for chromophore formation, then that protein would be less useful as an *in vitro* genetic reporter. Even within cells, differences in fluorescent protein behavior have been noted, particularly for multidomain proteins.[18] Despite these difficulties, we found that most of the fluorescent proteins tested function satisfactorily in *in vitro* transcription-translation reactions with the PUREsystem at 37 °C. One exception is CyPet, which fails to give a significant fluorescent output. The fact that CyPet expression at 30 °C gives a better fluorescence signal is consistent with previous reports on the poor folding properties of CyPet.[16] If a fluorescent protein with cyan spectral properties were desired, cerulean would be a better choice. The green fluorescent proteins are generally bright and rapidly give rise to fluorescence signals, e.g., the $t_{f/2}$ of sfGFP is 92 min. sfGFP is particularly amenable to *in vitro* transcription-translation; however, GFPmut3b performs similarly well. GFPmut3b is one of the more common fluorescent proteins used in synthetic biology. Two of the tested green fluorescent proteins fluoresce upon excitation with near-UV light. Of these two, T-Sapphire has a $t_{f/2}$ approximately 100 min longer than that of GFPuv. Therefore, GFPuv would be better for real-time detection assays than T-Sapphire. The yellow fluorescent proteins Venus and YPet are the brightest fluorescent proteins that we tested and have $t_{f/2}$ values below 150 min. Venus and YPet are excellent choices to monitor *in vitro* reactions particularly when low protein output is expected, e.g., when expressing inside of vesicles.[19] YPet is more photostable,[16] which could be important depending upon the nature of the planned experiments. The red fluorescent proteins mCherry and mRFP1 perform similarly well in *in vitro* transcription-translation reactions, but mCherry is more photostable. Although TagRFP-T is a highly photostable red fluorescent protein alternative, the long $t_{f/2}$ of TagRFP-T limits its usefulness.

All of the seven tested double fluorescent protein constructs performed well, and so the choice of fluorescent protein pairs depends on the specifics of the experimental setup. We found that the mVenus-mCherry pair gives easy to detect fluorescence signals and reproducible data without interference between the emission of mVenus and the emission of mCherry. Therefore, the subsequent experiments that probed the effects of genetic organization on protein production were performed with synthetic operons encoding mVenus and mCherry. However, for the remaining experiments the order of the genes was reversed so that mCherry was encoded first followed by mVenus in the bicistronic message. In this way, the lower intensity fluorescent protein, i.e., mCherry, could be used to provide the reference fluorescence signal and the influences of the region between the two genes on the expression of the

brighter fluorescent protein, i.e., mVenus, could be more easily assessed. Nevertheless, care should be taken in interpreting the resulting data. The assay can be used to characterize how changes in DNA sequence influence the ratio of the two encoded proteins. However, the assay does not differentiate between the decrease of expression of gene 1 or the increase of expression of gene 2. In other words, multiple mechanisms can give indistinguishable results.

Influence of Sequences Upstream of the Ribosome Binding Site. The first question we sought to answer was whether the number of nucleotides separating the stop codon of gene 1 from the ribosome biding site of gene 2 influenced gene expression. Therefore, constructs containing 0, 5, 20, 31, and 50 bp spacer sequences between the UAA stop codon of gene 1 and the AAGGAG RBS of gene 2 were tested (Figure 2). Although differences in expression levels were observed, the

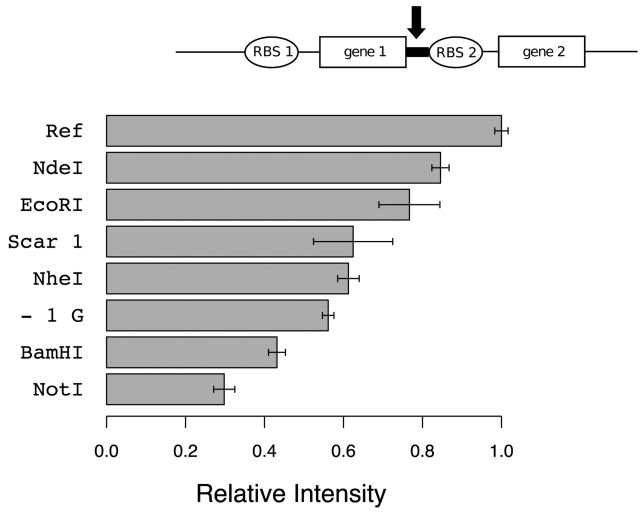


feature	name	RNA sequence
0	RL025A	
5	RL026A	GCGGA
20	RL024A	GCGGAUCCGAAUUCAAUUAG
31	RL027A	GCGGAUCCGAAUUCAAUUAGUUUGAACUUAU
50	RL030A	GCGGAUCCGAAUUCAAUUAGUUUGAACUUAUCAUAGUCAUAUCUUUCAAU

Figure 2. Influence of spacer length between an upstream gene and a downstream ribosome binding site on expression levels. The ratiometric response represents the fluorescence arising from mVenus (encoded by gene 2) divided by the fluorescence of mCherry (encoded by gene 1). Spacer lengths of 0, 5, 20, 31, and 50 nucleotides were tested. The corresponding RNA sequence for the region of interest of each construct is shown below the graph. Each bicistronic construct was expressed *in vitro* with the PUREsystem at 37 °C for 6 h.

differences did not correlate with the length of the spacer. For example, the 5 bp and the 31 bp spacer containing constructs both resulted in higher relative expression of gene 2 when compared with the 20 bp spacer. This suggested that the variance in fluorescence ratios resulted from something other than spacer length, such as sequence composition. For the remainder of the experiments, the 31 bp spacer construct (RL027A) was used as the reference.

Since the length of the spacer between gene 1 and RBS 2 did not appear to be correlated with the expression of gene 2, we wondered if the sequence composition rather than the length was responsible for the observed differences in expression. We decided to investigate the influences of sequence composition by incorporating different restriction sites immediately upstream to RBS 2. In this way we hoped to additionally identify restriction sites useful for the assembly of genetically encoded devices. Therefore, in each of the tested constructs, the 31 bp spacer length was maintained, and sequences containing a NdeI, *Bam*HI, NheI, *Eco*RI, NotI, or a scar site were incorporated. The scar site represented the sequence that results from standard BioBrick assembly in which complementary XbaI and SpeI digested products are ligated.[20] Additionally, the U before the AAGGAG RBS was mutated to a G, since a U residue is capable of base-pairing with 16S rRNA. A significant effect of sequence composition on the amount of protein produced was observed (Figure 3). The



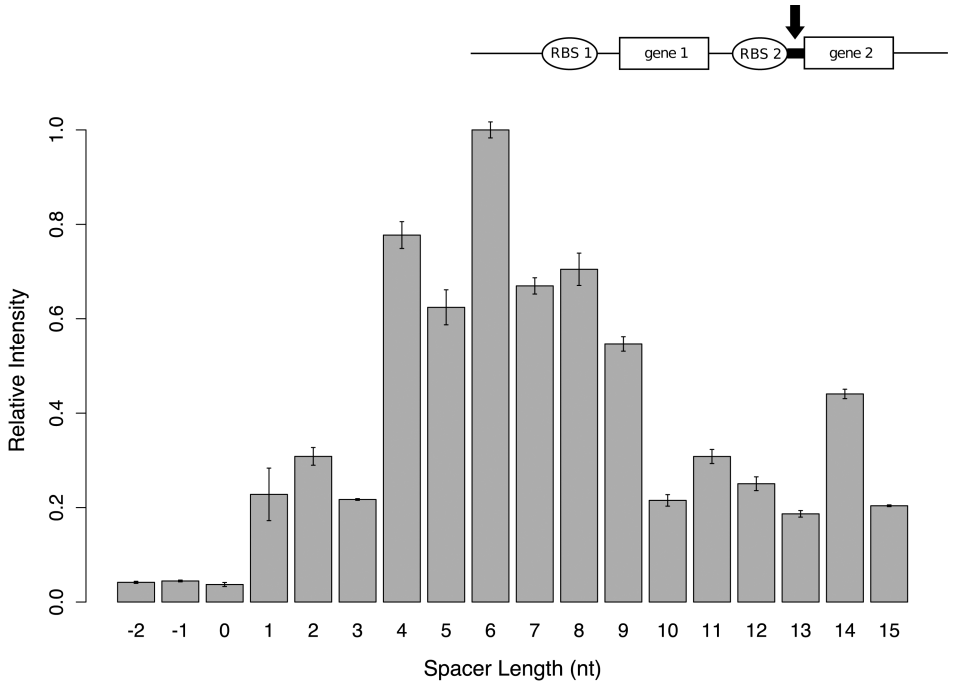
feature	name	RNA sequence
Ref	RL027A	GCGGAUCCGAAUUCAAUUAGUUUGAACUUAU
NdeI	RL036A	GCGGAUCCGAAUUCAAUUAGUUUGACAUAUG
EcoRI	RL039A	GCGGAUCCGAAUUCAAUUAGUUUGAGAAUUC
Scar 1	RL041A	GCGGAUCCGAAUUCAAUUAGUUUUACUAGAG
NheI	RL038A	GCGGAUCCGAAUUCAAUUAGUUUGAGCUAGC
- 1 G	LM021A	GCGGAUCCGAAUUCAAUUAGUUUGAACUUAG
BamHI	RL037A	GCGGAUCCGAAUUCAAUUAGUUUGAGGAUCC
NotI	RL040A	GCGGAUCCGAAUUCAAUUAGUUUGCGGCCGC

Figure 3. Influence of sequence composition upstream of the ribosome binding site on *in vitro* expression levels. The corresponding RNA sequence for the region of interest of each construct is shown below the graph. Underlined positions refer to the introduced feature. Ref refers to the reference construct RL027A, Scar 1 indicates the standard BioBrick scar sequence, and −1 G refers to the introduction of a G immediately prior to RBS 2. Each bicistronic construct was expressed *in vitro* with the PUREsystem at 37 °C for 6 h. Gene 1 encoded mCherry, and gene 2 encoded mVenus. Data are plotted relative to RL027A.

introduction of a NotI site was the most inhibitory, bringing relative expression down by 70% in comparison to the reference RL027A construct. Of the restriction sites tested, NdeI and *Eco*RI restriction site sequences were the most conducive to high expression (84% and 77% relative expression, respectively). Removing the additional base-pair of the RBS, i.e., the U to G mutation, decreased protein production by 44%, consistent with the observed decrease in expression from the 20 bp spacer construct described above that contained the same nucleotide at this position.

Influence of Sequences Downstream of the Ribosome Binding Site. Having probed the influences of the region 5′ to RBS 2, we next investigated the impact of the region 3′ to RBS 2. First, we altered the spacing between RBS 2 and the start codon of gene 2 one nucleotide at a time from −2 to 15 bp. Here the spacing nomenclature followed the aligned spacing described by Chen et al.[21] in which the RBS was aligned with

dx.doi.org/10.1021/sb400003y | *ACS Synth. Biol.* 2013, 2, 482−489



feature	name	RNA sequence	feature	name	RNA sequence
-2	RL032A		7	RL033A	AAUAAUCUA
-1	AS006A	A	8	LM003A	AAUAAUCUAC
0	DT007A	AA	9	RL034A	AAUAAUCUACU
1	DT008A	AAU	10	AS007A	AAUAAUCUACUU
2	DT006A	AAUA	11	LM022A	AAUAAUCUACUUC
3	RL031A	AAUAA	12	LM023A	AAUAAUCUACUUCA
4	AS002A	AAUAAU	13	LM024A	AAUAAUCUACUUCAA
5	AS001A	AAUAAUC	14	RL043A	AAUAAUCUACUUCAAU
6	RL027A	AAUAAUCU	15	RL044A	AAUAAUCUACUUCAAUU

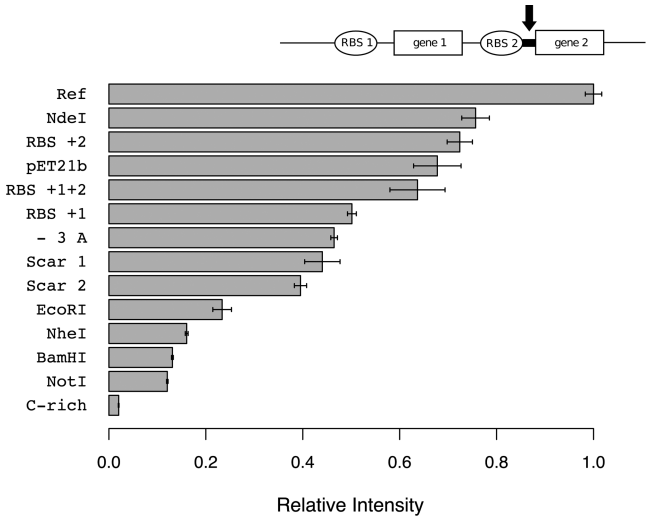
Figure 4. Ribosome binding site spacing. The influence of the aligned spacing between the ribosome binding site and the start codon is shown. The corresponding RNA sequence for the region of interest of each construct is reported below the graph. Each bicistronic construct was expressed *in vitro* with the PURExpress system at 37 °C for 6 h. Gene 1 encoded mCherry, and gene 2 encoded mVenus. Data are plotted in reference to RL027A.

the anti-RBS sequence of the 16S rRNA and the position across from the last position of the anti-RBS was taken as 0 (Supplementary Figure S5). The results were consistent with previous *in vivo* studies,[22] which showed a Gaussian distribution of activity with optimal aligned spacing between 4 and 9 bp (Figure 4). Spacer lengths shorter or longer than this range generally resulted in dramatically decreased protein production. For example, the 3 bp spacer produced 72% less protein than the 4 bp spacer. Similarly, the 10 bp spacer reduced protein synthesis by 60% when compared to the 9 bp spacer construct. For the specific constructs tested in this study, the 6 bp spacer produced the most protein. Since protein expression was detected with the shortest spacer tested on both sides of RBS 2, we also made a minimal construct with a 0 bp spacer between the UAA stop codon of gene 1 and RBS 2 and −2 aligned spacing between RBS 2 and the start codon of gene 2. The synthesis of mVenus from this minimally spaced construct was low but still detectable (3% relative to RL027A).

Next, we evaluated the effect of sequence composition of the region between RBS 2 and the AUG start codon of gene 2 on expression levels. This region of the reference sequence RL027A was designed to be high in A-U content and low in G content because a sequence that is known to facilitate gene expression, i.e., the T7 phage gene 10 leader sequence,[23] has similar characteristics. Sequences that contained the same restriction sites tested above for the region upstream of RBS 2 were placed immediately upstream of the start codon of gene 2. An additional BioBrick scar site also was screened that was shorter and thus thought to interfere less with translation. The presence of an A three nucleotides upstream of the start codon was evaluated since an A at this position is frequently found in prokaryotic and eukaryotic sequences.[24,25] A C-rich sequence was evaluated since a previous *in vitro* study[26] found increased expression associated with high C-content. Finally, mutations that introduced additional base-pairing with the 16S rRNA were added. The data showed a strong influence of sequence composition on protein yields with the NotI restriction site being the most inhibitory, decreasing expression by 87% (Figure 5). The NdeI restriction site was the most conducive to protein synthesis (76% relative expression). Both scar sequences resulting from BioBrick assembly performed similarly, decreasing translation by over 50%. The C-rich sequence greatly decreased protein expression by 98% relative to RL027A. Neither an A residue three nucleotides preceding the start codon nor the expansion of the RBS-anti-RBS base-pairing region increased protein production in the tested constructs.

Finally, we investigated whether other codons could substitute for the AUG start codon. In *E. coli*, GUG and UUG function as start codons at a frequency of 14% and 3%,

dx.doi.org/10.1021/sb400003y | *ACS Synth. Biol.* 2013, 2, 482−489



feature	name	RNA sequence	feature	name	RNA sequence
Ref	RL027A	AAUAAUCU	Scar 1	LM010A	UACUAGAG
NdeI	LM005A	AAUAACAU	Scar 2	LM011A	AAUACUAG
RBS +2	LM014A	AUUAAUCU	EcoRI	LM008A	AAGAAUUC
pET21b	LM016A	AUAUACAU	NheI	LM007A	AAGCUAGC
RBS +1+2	LM015A	GUUAAUCU	BamHI	LM006A	AAGGAUCC
RBS +1	LM013A	GAUAAUCU	NotI	LM009A	GCGGCCGC
- 3 A	LM012A	AAUAAACU	C-rich	RL050A	CCCCCUCC

Figure 5. Influence of sequence composition between the ribosome binding site and the start codon on expression levels. The corresponding RNA sequence for the region of interest of each construct is shown below the bar graph. Ref indicates the reference construct RL027A. Scar 1 is the standard BioBrick scar sequence. Scar 2 is the shorter, alternate scar sequence. −3 A indicates the introduction of an A three positions upstream of the start codon. pET21b is the same spacer sequence found in the expression vector pET21b (Novagen). RBS +1, RBS +2, and RBS +1+2 indicate RBS expansions. Each introduced feature is underlined in the corresponding sequence. Note that only half of the NdeI restriction site is shown since the remaining half overlaps with the start codon. Each bicistronic construct was expressed *in vitro* with the PURExpress system at 37 °C for 6 h. Gene 1 encoded mCherry, and gene 2 encoded mVenus. Data are plotted in reference to RL027A.

respectively.[25] If non-AUG codons can function as start codons in minimally reconstituted systems, then these alternate start codons could be used to control protein levels. Also, knowledge regarding the functionality of non-AUG start codons could help to identify internal RBS-start codon pairs that could potentially interfere with the intended activity of genetic devices. We therefore substituted a GUG, UUG, and CUG in place of the AUG start codon and measured the production of mVenus. All of the alternate start codons produced protein, albeit at a significantly reduced level between 12% and 27% relative to the AUG start codon containing reference construct (Figure 6).

Considerations for the Assembly of *in Vitro* Genetic Systems. To determine if simple rules could be formulated that would facilitate the construction of genetically encoded, cell-free devices, the collected data were statistically analyzed. First, we sought to determine which regions were more amenable to the incorporation of restriction sites. A paired *t* test showed that sequences upstream of RBS 2 had less influence on the protein fluorescence ratios than the sequence between RBS 2 and the start codon (p-value = 0.0145). Next, sequences immediately 5′ and 3′ to RBS 2 (8 bp each) were considered. The resulting data from 22 synthetic operons were fit to multiple regression models that searched for first and second order interactions between base composition that

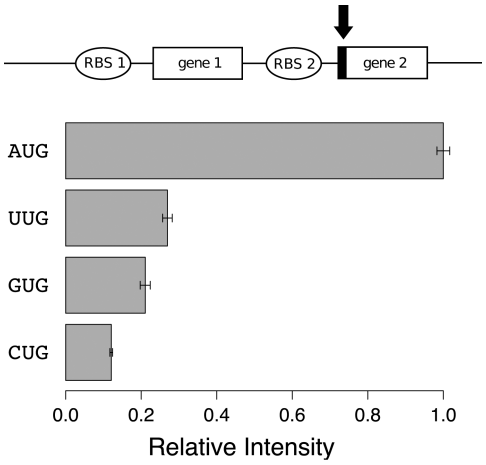


Figure 6. Alternate start codons. The ability of UUG, GUG, and CUG to function as start codons *in vitro* was evaluated. Relative intensities are averages of three replicates and plotted in reference to the AUG start codon containing construct. Each bicistronic construct was expressed *in vitro* with the PURExpress system at 37 °C for 6 h. Gene 1 encoded mCherry, and gene 2 encoded mVenus. The AUG, UUG, GUG, and CUG start codon constructs were RL027A, LM019A, LM018A, and LM020A, respectively.

correlated with the measured fluorescence intensity ratios. The resulting model was statistically significant (F-test p-value = 8.79×10^{-7}) and described almost 75% of the data variability (adjusted r^2 = 0.7453). The estimated parameters (Supplementary Table S1) revealed a strong effect of the G content in sequence composition of the region 5′ to RBS 2 (p-value <0.001). More specifically, a high G content negatively correlated with the fluorescence ratio, whereas combined A-U-rich sequences in the region 3′ to RBS 2 positively correlated with the fluorescence intensity ratio (p-value <0.001).

Taken together, the data indicate that the nucleotide sequence between genes 1 and 2 influence protein production, but not uniformly. The spacing upstream of the RBS is not as strong of a determinant of expression levels as the spacing downstream of the RBS. The one construct that deviates from this trend (RL024A) contains a mutation that decreases the number of potential base-pairs between the mRNA and the 16S rRNA. Most of the constructs tested here contain six to seven potential base-pairing interactions between the Shine-Dalgarno (RBS) and the anti-Shine-Dalgarno site of the ribosome. The introduction of additional base-pairing does not facilitate expression, consistent with previous studies that show that on average *E. coli* mRNA RBS sequences interact with the ribosome via six base-pairs and that the strengthening of the interaction often decreases rather than increases protein synthesis.[27] The optimal aligned spacing between the RBS and the start codon and the functionality of alternate start codons is the same for *in vitro* protein production with the PURExpress system and for natural *E. coli* expression.

Taken together, a few simple rules for the construction of *in vitro* genetic systems can be formulated from the acquired data. Restriction sites should either be placed before the RBS, since this region is more amenable to sequence modification, or a NdeI site that overlaps with the start codon should be exploited. If high protein levels are desired, then the aligned RBS spacing should be between four and nine nucleotides and the spacer sequence should be high in A and T content and low in G content. The use of alternate start codons can be used to

dx.doi.org/10.1021/sb400003y | *ACS Synth. Biol.* 2013, 2, 482−489

significantly reduce protein synthesis, when needed, and the spacing between the end of one gene and the RBS of the next gene is not crucial. Nevertheless, the complexity of transcription and translation ensures that there are many more factors that influence gene expression than was probed here. mRNA can interact with regions of the ribosome other than the 3′-terminus of the 16S rRNA[26−29] and the folding of mRNA significantly affects protein synthesis.[19,30−33] Further studies with purified, *in vitro* systems likely will aid in better understanding these processes and in facilitating the synthesis of more complex cellular mimics.

METHODS

Genetic Constructs. Genes encoding the fluorescent proteins were synthesized by Genscript or Mr. Gene, except for super folder GFP (BBa_I746916), GFPmut3b (BBa_E0040), and mRFP1 (BBa_E1010), which were from the registry of standard biological parts (http://partsregistry.org), and eGFP, which was from Roche. Mutagenesis was either performed by Genscript or through the use of phusion site-directed mutagenesis (Finnzymes). All genes were subcloned into pET21b by restriction digestion and ligation with NdeI and *Bam*HI, except for super folder GFP and GFPmut3b, which used NheI and *Bam*HI sites. All constructs were confirmed by sequencing at Genechron or Eurofins MWG Operon. The DNA sequences of all the constructs used are provided in the Supporting Information (Table S2).

Transcription-Translation Reactions. Plasmids were amplified in *E. coli* DH5α or NovaBlue and purified with Wizard Plus SV Minipreps DNA Purification System (Promega) or QIAprep Spin Miniprep Kit (Qiagen). Subsequently, the DNA was phenol-chloroform extracted, ethanol precipitated, and resuspended in deionized and diethylpyrocarbonate (DEPC) treated water. A 250 ng portion (2 nM final concentration) of DNA was used for each transcription-translation reaction with the PURExpress *in vitro* protein synthesis kit (New England BioLabs) supplemented with 20 units of human placenta RNase inhibitor (New England BioLabs). The final volume of each reaction was 25.5 μL. Reactions were monitored by fluorescence spectroscopy with a Photon Technology International (PTI) QuantaMaster 40 UV−vis spectrofluorometer equipped with two detectors (T-format). Excitation and emission wavelengths were specific for each fluorescent protein (Supplementary Table S3). The reaction components, except for the DNA template, were assembled on ice and then incubated at 37 °C in the spectrofluorometer. Subsequently, the reaction was initiated by the addition of DNA template. Mineral oil was layered on top of each sample to inhibit evaporation during the course of the experiment. Control experiments with GFPmut3b showed that mineral oil did not influence the appearance of fluorescence. Each reaction was repeated at least three times. An Agilent 8453 UV−vis spectrophotometer was used to quantify mVenus protein concentration by using an extinction coefficient at 515 nm of 92,200 M^{-1} s^{-1}.[16]

Data analysis. All statistical analyses used R statistical computing software.[34] The single protein construct fluorescent data were fit to

$$I(t) = \frac{K}{1 + e^{-B(t-t_{f/2})}} \quad (1)$$

where K, B, and $t_{f/2}$ were the upper asymptote, growth rate, and time of maximum growth, respectively (Supplementary Table S4). The parameters were estimated by using a nonlinear least-squares analysis with the Gauss−Newton algorithm. The mean values and standard errors were then calculated from data from three replications. The influence of spacer nucleotide composition on the fluorescence intensity was determined with multiple regression models. The models were estimated and reduced by using stepwise regression with a penalty term that was selected by minimum predictive mean squared error based on repeated cross-validation (10% leave-out). The best predictive models were obtained by using a stringent criterion (twice the Bayesian Information Criterion, BIC). We then estimated the model with such a penalty term on the whole set of operon spacer data. Paired t tests were used to test whether the restriction sites 5′ or 3′ to RBS 2 affected differently fluorescence intensity ratios.

ASSOCIATED CONTENT

Supporting Information

Supplementary tables and figures. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel: +39 0461 28 3438. Fax: +39 0461-283091. E-mail: mansy@science.unitn.it.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

RL015A, RL016A, RL018A, RL020A, and RL027A are available through Addgene. Versions of RL027A modified by the 2012 Trento iGEM team to be BioBrick compatible were deposited as BBa_K731700 and BBa_K731710 in the Registry of Standard Biological Parts (http://partsregistry.org/). We thank the Armenise-Harvard Foundation, the autonomous province of Trento (Ecomm), the Marie-Curie Trentino COFUND (ACS), and CIBIO for funding.

REFERENCES

(1) Forlin, M., Lentini, R., and Mansy, S. S. (2012) Cellular imitations. *Curr. Opin. Chem. Biol. 16*, 586−592.
(2) Forster, A. C., and Church, G. M. (2006) Towards synthesis of a minimal cell. *Mol. Syst. Biol. 2*, 45.
(3) Harris, D. C., and Jewett, M. C. (2012) Cell-free biology: exploiting the interface between synthetic biology and synthetic chemistry. *Curr. Opin. Biotechnol. 23*, 672−678.
(4) Ichihashi, N., Matsuura, T., Kita, H., Sunami, T., Suzuki, H., and Yomo, T. (2010) Constructing partial models of cells. *Cold Spring Harbor Perspect. Biol. 2*, 295−303.
(5) Martos, A., Jimenez, M., Rivas, G., and Schwille, P. (2012) Towards a bottom-up reconstitution of bacterial cell division. *Trends Cell Biol. 22*, 634−643.
(6) Noireaux, V., Maeda, Y. T., and Libchaber, A. (2011) Development of an artificial cell, from self-organization to computation and self-reproduction. *Proc. Natl. Acad. Sci. U.S.A. 108*, 3473−3480.
(7) Stano, P., and Luisi, P. L. (2010) Achievements and open questions in the self-reproduction of vesicles and synthetic minimal cells. *Chem. Commun. (Cambridge) 46*, 3639−3653.
(8) Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-free translation reconstituted with purified components. *Nat. Biotechnol. 19*, 751−755.

(9) Stano, P., Kuruma, Y., Souza, T. P., and Luisi, P. L. (2010) Biosynthesis of proteins inside liposomes. *Methods Mol. Biol. 606*, 127−145.
(10) Sunami, T., Matsuura, T., Suzuki, H., and Yomo, T. (2010) Synthesis of functional proteins within liposomes. *Methods Mol. Biol. 607*, 243−256.
(11) Kita, H., Matsuura, T., Sunami, T., Hosoda, K., Ichihashi, N., Tsukada, K., Urabe, I., and Yomo, T. (2008) Replication of genetic information with self-encoded replicase in liposomes. *ChemBioChem 9*, 2403−2410.
(12) Karig, D. K., Iyer, S., Simpson, M. L., and Doktycz, M. J. (2012) Expression optimization and synthetic gene networks in cell-free systems. *Nucleic Acids Res. 40*, 3763−3774.
(13) Shin, J., and Noireaux, V. (2012) An E. coli cell-free expression toolbox: application to synthetic gene circuits and synthetic cell. *ACS Synth. Biol. 1*, 29−41.
(14) Chan, L. Y., Kosuri, S., and Endy, D. (2005) (2005) Refactoring bacteriophage T7. *Mol. Syst. Biol. 1*, 0018.
(15) Kwok, R. (2010) Five hard truths for synthetic biology. *Nature 463*, 288−290.
(16) Shaner, N. C., Steinbach, P. A., and Tsien, R. Y. (2005) A guide to choosing fluorescent proteins. *Nat. Methods 2*, 905−909.
(17) Iizuka, R., Yamagishi-Shirasaki, M., and Funatsu, T. (2011) Kinetic study of de novo chromophore maturation of fluorescent proteins. *Anal. Biochem. 414*, 173−178.
(18) Chang, H. C., Kaiser, C. M., Hartl, F. U., and Barral, J. M. (2005) De novo folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria. *J. Mol. Biol. 353*, 397−409.
(19) Martini, L., and Mansy, S. S. (2011) Cell-like systems with riboswitch controlled gene expression. *Chem. Commun. (Cambridge) 47*, 10734−10736.
(20) Shetty, R. P., Endy, D., and Knight, T. F., Jr. (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng. 2*, 5.
(21) Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. *Nucleic Acids Res. 22*, 4953−4957.
(22) Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D., and Gold, L. (1992) Translation initiation in Escherichia coli: sequences within the ribosome-binding site. *Mol. Microbiol. 6*, 1219−1229.
(23) Olins, P. O., Devine, C. S., Rangwala, S. H., and Kavka, K. S. (1998) The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in Escherichia coli. *Gene 73*, 227−235.
(24) Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell 44*, 283−292.
(25) Shultzaberger, R. K., Bucheimer, R. E., Rudd, K. E., and Schneider, T. D. (2001) Anatomy of Escherichia coli ribosome binding sites. *J. Mol. Biol. 313*, 215−228.
(26) Barendt, P. A., Shah, N. A., Barendt, G. A., and Sarkar, C. A. (2012) Broad-specificity mRNA-rRNA complementarity in efficient protein translation. *PLoS Genet. 8*, e1002598.
(27) Vimberg, V., Tats, A., Remm, M., and Tenson, T. (2007) Translation initiation region sequence preferences in Escherichia coli. *BMC Mol. Biol. 8*, 100.
(28) Etchegaray, J. P., Xia, B., Jiang, W., and Inouye, M. (1998) Downstream box: a hidden translational enhancer. *Mol. Microbiol. 27*, 873−874.
(29) Boni, I. V., Isaeva, D. M., Musychenko, M. L., and Tzareva, N. V. (1991) Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. *Nucleic Acids Res. 19*, 155−162.
(30) Caschera, F., Bedau, M. A., Buchanan, A., Cawse, J., de Lucrezia, D., Gazzola, G., Hanczyc, M. M., and Packard, N. H. (2011) Coping with complexity: machine learning optimization of cell-free protein synthesis. *Biotechnol. Bioeng. 108*, 2218−2228.
(31) de Smit, M. H., and van Duin, J. (1990) Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci. U.S.A. 87*, 7668−7672.
(32) Kobori, S., Ichihashi, N., Kazuta, Y., Matsuura, T., and Yomo, T. (2012) Kinetic analysis of aptazyme-regulated gene expression in a cell-free translation system: modeling of ligand-dependent and -independent expression. *RNA 18*, 1458−1465.
(33) Winkler, W. C., and Breaker, R. R. (2005) Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol. 59*, 487−517.
(34) R Development Core Team (2011) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna.