ACS Synthetic Biology
Cite This: ACS Synth. Biol. 2019, 8, 1263−1275
Research Article
pubs.acs.org/synthbio

Design and Evaluation of Synthetic Terminators for Regulating Mammalian Cell Transgene Expression

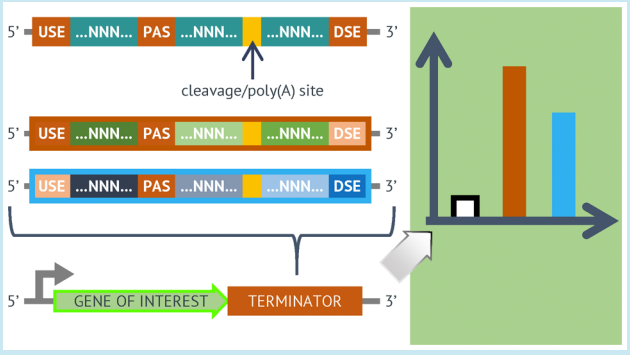
Joseph K. Cheng,[†,§] Nicholas J. Morse,[†] James M. Wagner,[†] Scott K. Tucker,[‡] and Hal S. Alper*[,†,‡]

[†]McKetta Department of Chemical Engineering, The University of Texas at Austin, 200 E Dean Keeton Street, Stop C0400, Austin, Texas 78712, United States

[‡]Institute for Cellular and Molecular Biology, The University of Texas at Austin, 2500 Speedway Avenue, Austin, Texas 78712, United States

Supporting Information

ABSTRACT: Tuning heterologous gene expression in mammalian production hosts has predominantly relied upon engineering the promoter elements driving the transcription of the transgene. Moreover, most regulatory elements have borrowed genetic sequences from viral elements. Here, we generate a set of 10 rational and 30 synthetic terminators derived from nonviral elements and evaluate them in the HT1080 and HEK293 cell lines to demonstrate that they are comparable in terms of tuning gene expression/protein output to the viral SV40 element and often require less sequence footprint. The mode of action of these terminators is determined to be an increase in mRNA half-life. Furthermore, we demonstrate that constructs comprising completely nonviral regulatory elements (*i.e.*, promoters and terminators) can outperform commonly used, strong viral based elements by nearly 2-fold. Ultimately, this novel set of terminators expanded our genetic toolkit for engineering mammalian host cells.



KEYWORDS: *mRNA half-life, synthetic terminators, mammalian production host, HT1080, HEK293*

The 3′ UTR, or terminator region, is often overlooked in the design of synthetic and transgene elements. For the case of typical mammalian cell hosts, the terminator is usually and simply chosen from a small list of commonly (and previously) reported viral-derived elements.[1] However, such a selection is likely one of convenience rather than function or fundamental understanding of terminator function. Functionally, the terminator can modulate gene expression and its corresponding protein output by controlling the stability of the transgene mRNA and exerting influence on post-transcriptional processing such as intron splicing.[2−4] Furthermore, significant molecular biology efforts have characterized various viral and endogenous terminators[5−11] resulting in a proposed, generic structure for this element.[12] Specifically, this generalized terminator structure is comprised of an upstream sequence element (USE), the highly conserved hexameric polyadenylation signal (PAS), a cleavage/polyadenylation site, and the downstream sequence element (DSE).

This basic terminator structure is similar to those of yeast terminators that have been previously designed by our group for the host *S. cerevisiae* and shown to be transferrable to another yeast, *Y. lipolytica*.[13] Specifically, this prior work has demonstrated that terminators function relatively independent of other elements (short of some sequence-based context effects) in altering the mRNA half-life, facilitating transcriptional stops, and influencing total net protein output from a transcript. Recently, a systematic dissection of the 3′ UTRs from human chemokines verified the regulatory effects of this element on chemokine mRNA abundance and protein production, along with their correlation with mRNA stability in various mammalian cell types.[14] Therefore, the commonality of the conserved regions of the terminator element indicates that unlike promoters, terminator elements can be broadly (and in some cases, modularly) designed.

The modularity and portability of terminators in mammalian cells has been evaluated in a study that investigated the endogenous human gastrin gene element along with the viral SV40 terminator and affirmed improved reporter gene expression in three tested mammalian cell lines.[15] As a result, this work suggested that mammalian host terminators can be rationally designed in a similar fashion to yeast[13,16] for improved translational output. More importantly, these prior results illustrate that the terminator can be designed in an orthogonal manner to the promoter selected for both yeast[13,17] and mammalian cells.[3]

Building upon the molecular biology studies of terminator elements, several key regions and consensus sequences have been explored in their relationship to post-transcriptional regulation.[18] The most highly conserved element is the

Received: July 5, 2018
Published: May 15, 2019

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263−1275

hexameric PAS, with minimal deviation from the AAUAAA consensus.[10,19] The USE contains sequences that interact with splicing factors in a manner that helps promote 3′ end formation.[20] Likewise, the consensus sequence for the DSE is essential for 3′ end formation.[11,21−23] Despite the generic and conserved structure of mammalian terminators, it has been proposed that these elements are not minimal and thus certain elements are dispensable.[24] For example, in the absence of the hexameric PAS, other USEs can compensate by directing the recognition of the appropriate polyadenylation site and poly(A) addition.[25,26] To this end, a synthetic sequence comprised of only the PAS and DSE spaced 22- to 23-bp apart was sufficient to function and terminate transcription.[27] As a result, these elements as well as their spacing and interaction are important factors that guide performance.

Here, we leverage our previous efforts of rational design of synthetic yeast terminators[13] together with these key elements to evaluate a set of novel endogenous and synthetic mammalian terminators. Specifically, we develop a set of 10 endogenous and 30 synthetic terminator variants that have varying capacity to modulate gene expression and translational output through an mRNA half-life based mechanism. We further demonstrate the transferability of these terminators across two mammalian cell lines, HT1080 and HEK293. Finally, by coupling a fully human-derived endogenous with a synthetic terminator, we were able to demonstrate higher output than from a fully viral-derived construct. These results expand the availability of mammalian terminators and facilitates the replacement of viral-derived sequences.

RESULTS AND DISCUSSION

Rational Design and Evaluation of Selected Endogenous Terminators. Many mammalian housekeeping genes are reported to have mRNA exhibiting relatively long half-lives.[28] As a result, these observations suggest that the respective terminator regions of these genes could harbor favorable sequences in their 3′ UTR that confer this attribute. Interestingly, some of these genes (*e.g.*, *EEF1A1*, *GAPDH*, and *ACTB*) are also among the most highly expressed from the genome based on previous microarray expression data analysis.[29] Therefore, we first explored 3′ UTR variants from three of these highly expressed genes with long half-lives (Table 1 and Supporting Information Table S11) for their utility as protein output-enhancing terminators. This impact was measured initially *via* a fluorescent reporter protein, hrGFP. However, the exact sequences for these terminators are not specifically annotated in the genome; thus, we created several variants to evaluate for their function. To accomplish this, terminator sequences from *EEF1A1*, *GAPDH*, and *ACTB* were identified through determining approximate locations of the key terminator elements previously described,[12] most notably the USEs and DSEs (Table 1 and Supporting Information Table S11). Finally, as a control, we also created a similar variation/truncation set (Table 1 and Supporting Information Table S11) of the commonly used, viral SV40 terminator to likewise estimate the impacts of USEs and DSEs on function.

In our prior work,[17] the impact and function of the terminator was most pronounced (*i.e.*, with the highest dynamic range) when paired with moderate to weak strength promoters. Therefore, we chose to pair each of these endogenous terminator sets with a moderate strength promoter derived from *EIF4A1* to drive transient hrGFP reporter expression in the HT1080 model cell line. Protein output was assessed through flow cytometry at a point 48 h post transfection and reported as the geometric mean fluorescence intensity (Figure 2A). The truncation variants of the SV40 viral terminator suggested that the DSE had a greater impact on protein output (Figure 2A). In this regard, of the four SV40 terminator variants tested, only m.SV40pA.1 affected protein output negatively relative to the full-length terminator (f.SV40pA) (Figure 2A). In contrast, the variant analysis from the endogenous genes did not yield any conclusive evidence for the USE or DSE as the key driver. Nevertheless, this set led to the identification of three endogenous terminator sequences (T.GAPDH.2, T.ACTB.3, and T.ACTB.4) that function comparably to the full length 222-bp SV40 terminator (f.SV40pA) (Figure 2A).

Synthetic Terminator Design and Evaluation. Next, we sought to evaluate the use of a synthetic terminator scaffold as an efficacious way to increase net protein output through mRNA enhancement. This section is motivated by prior work demonstrating that rationally designed, synthetic terminators can function in eukaryotes.[13,27] Therefore, we employed the generic mammalian terminator structure[12] as a scaffold (Figure 1A) and created 30 synthetic permutations by modulating the spacer, polyA, USE and DSE elements. To do so, spacer sequences derived from the endogenous sequences for *EEF1A1*, *GAPDH*, and *ACTB* used above as well as from SV40 (Table 2 and Supporting Information Table S12). Variants were created using the polyadenylation site consensus dinucleotide "CA" instead of the native site from *EEF1A1*, *GAPDH*, and *ACTB* to evaluate the impact of this key terminator element. Finally, we explored putative USEs and DSEs from these endogenous genes in addition to their consensus sequences,[12] as well as two DSEs identified in other reports.[24,30] Collectively, the derivatives (levels) of these key elements (factors) are summarized in Figure 1B.

These synthetic terminators were evaluated similarly to the endogenous elements described above through a hrGFP reporter. A total of 14 of the 30 synthetic terminators developed comprising solely endogenous and consensus sequences perform similarly to the common, viral derived f.SV40pA terminator, with one (Tm.synth.13) performing in a manner that is statistically indistinguishable (Supporting Information Table S1). A select subset of these terminators were coexpressed with a constitutive mStrawberry (RFP) reporter to verify robust transfection conditions (Supporting Information Table S2), enabling us to attribute measured differential hrGFP expression to the terminator variant. Collectively, this set of synthetic terminators affords a 5.7-fold dynamic range of hrGFP reporter output (Figure 2A).

Terminators Exhibit Functionality Across Different Genomic Contexts. As these synthetic terminators were evaluated based on a generic scaffold, we sought to evaluate whether these elements would function in a transgene and promoter-independent manner. To do so, we examined a subset of our endogenous and synthetic terminator designs for their ability to modulate the protein output of secreted alkaline phosphatase (SEAP), a secreted reporter protein. Specifically, we selected this particular subset of terminator designs in order to facilitate the dissection of the key elements that comprise this generic terminator scaffold (not necessarily due to the performance of the terminator). Likewise, we evaluated their function when paired with a stronger promoter (1356-bp

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263−1275

Table 1. Native/Endogenous Sequences with Putative Annotations of Terminator Components

terminator	total length	description	USE (T-rich)	spacer 1 (0–20 bp)	spacer 2 (15–30 bp)	poly(A) site	spacer 3 (0–20 bp)	DSE (T- or G/T-rich)
f.SV40pA	222	SV40 late terminator	CAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAAT GCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAA	ccattata agctgc	caagttaacaac	AA	caattgcatt-cattt	TATGTTTCA GGTTCAGGG GGAGGTGTG GGAGGTTTT TTAAAGCAAG-TAAAACCTCTA-CAAATGTGGTA
m.SV40pA.1	62	"minimal" SV40 late terminator, minimal USE, minimal DSE	TGTAA	ccattata agctgc	caagttaacaac	AA	caattgcatt-cattt	TATGTTTC
m.SV40pA.2	92	"minimal" SV40 late terminator, minimal USE, short DSE	TGTAA	ccattat aagctgc	caagttaacaac	AA	caattgcatt-cattt	TATGTTTCAG GTTCAGGGG GAGGTGTGG GAGGTTTTTT
m.SV40pA.3	192	"minimal" SV40 late terminator, full USE, short DSE	CAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAA ATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAA	ccattat aagctgc	caagttaacaac	AA	caattgcatt-cattt	TATGTTTCA GGTTCAGGG GGAGGTGTG GGAGGTTTTTT
m.SV40pA.4	142	"minimal" SV40 late terminator, short USE, full DSE	TGTGATGCTATTGCTTTATTTGTAA	ccattat aagctgc	caagttaacaac	AA	caattgcatt-cattt	TATGTTTCA GGTTCAGGG GGAGGTGTGG GAGGTTTTT TAAAGCAAG TAAAACCTCT ACAAATG TGGTA
T.EEF1A1.1	71	EEF1A1 terminator, short USE, short DSE	TGGTAT	tcattac aaacttgc tcactac	tgaattttaagcttt	AA	gatgaagtggca	TTTCTTTT
T.EEF1A1.2	173	EEF1A1 terminator, short USE, long DSE	TGGTAT	tcattac aaacttgc tcactac	tgaattttaagcttt	AA	gatgaagtggca	TTTCTTTTAA CAGTTACTAT GTTGGAATTG GTTACAAA TTTTGGAGT GGATTTCAAA AGTGAGAGC TAACTTCA GTTGATTT CAAGGTAG TGCTTGGCTT TTTTTGTTTA
T.EEF1A1.3	117	EEF1A1 terminator, long USE, short DSE	TGTGAAACCCAGTGTCTTAGACAACTGTGGCTTGAGCACCACCTGCTGGTAT	tcattac aaacttgct cactac	tgaattttaagcttt	AA	gatgaagtggca	TTTCTTTT
T.EEF1A1.4	219	EEF1A1 terminator,	TGTGAAACCCAGTGTCTTAGACAACTGTGGCTTGAGCACCACCTGCTGGTAT	tcattac aaacttgc tcactac	tgaattttaagcttt	AA	gatgaagtggca	TTTCTTTTA ACAGTTACT ATGTTGGAA

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263–1275

Table 1. continued

terminator	total length	description	USE (T-rich)	spacer 1 (0–20 bp)	spacer 2 (15–30 bp)	poly(A) site	spacer 3 (0–20 bp)	DSE (T- or G/T-rich)
		long USE, long DSE						TTGGTTACA AATTTTGGA GTGGATTT CAAAAGTGA GAGCTAACTT CAGTTGATTT CAAGGTAGT GCTTGGCTTT TTTTGTTTA
T.GAPDH.1	99	GAPDH terminator, short USE	TGTCATGTAC	catc	gtaccctg tgctcaaccagt	TA	cttgtcct gtcttattctag	GGTCTGGGGC AGAGGGGAG GGAAGCTGG GCTTGTGTC
T.GAPDH.2	142	GAPDH terminator, long USE	TGTAGACCCCTTGAAGAGGGGAGGGGCCTAGGGAGCCGCACCTTGTCATGTAC	catc	gtaccct gtgctcaaccagt	TA	cttgtcct gtcttattctag	GGTCTGGGGCA-GAGGGGA GGGAAGCTG GGCTTGTGTC
T.ACTB.1	700	ACTB reference sequence (same for both mRNA variants), full USE sequence prior to PAS, 100-bp DSE after poly (A) site	GCGGACTATGACTTAGTTGCGTTACACCCTTTCTTGACAAAACCTAACTTGCGCAGAAAACAA GATGAGATTGGCATGGCTTTATTTGTTTTTTTTGTTTTGTTTTGGTTTTTTTTTTTTTTT TGGCTTGACTCAGGATTTAAAAACTGGAACGGTGAAGGTGACAGCAGTCGGTTGGAGCGA GCATCCCCCAAAGTTCACAATGTGGCCGAGGACTTTGATTGCACATTGTTGTTTTTTTAATA GTCATTCCAAATATGAGATGCGTTGTTACAGGAAGTCCCTTGCCATCCTAAAAGCCACCCCA CTTCTCTCTAAGGAGAATGGCCCAGTCCTCTCCCAAGTCCACACAGGGGAGGTGATAGCATT GCTTTCGTGTAAATTATGTAATGCAAAATTTTTTTAATCTTCGCCTTAATACTTTTTTATTTTG TTTTATTTTGAATGATGAGCCTTCGTGCCCCCCCCTTCCCCCTTTTTTGTCCCCCAACTTGAGA TGTATGAAGGCTTTTGGTCTCCCTGGGAGTGGGTGGAGGCAGCCAGGGCTTACCTGTAC	actgacttga-gaccagttg	agtgcacac-cttaaaaat	GA	ggccaa gtgtgac	TTTGTGGTG TGGCTGGGT TGGGGGCAG CAGAGGGTG AACCCTGCAG-GAGGGTG AACCCTGCA AAAGGGTGG GGCAGTGG GGGCCAAC
T.ACTB.2	637	ACTB reference sequence (same for both mRNA variants), full USE sequence prior to PAS, 37-bp DSE after poly (A) site	GCGGACTATGACTTAGTTGCGTTACACCCTTTCTTGACAAAACCTAACTTGCGCAGAAAACAA GATGAGATTGGCATGGCTTTATTTGTTTTTTTTGTTTTGTTTTGGTTTTTTTTTTTTTTT TGGCTTGACTCAGGATTTAAAAACTGGAACGGTGAAGGTGACAGCAGTCGGTTGGAGCGA GCATCCCCCAAAGTTCACAATGTGGCCGAGGACTTTGATTGCACATTGTTGTTTTTT TAATAGTCATTCCAAATATGAGATGCGTTGTTACAGGAAGTCCCTTGCCATCCTAAAAGCCA CCCCACTTCTCTCTAAGGAGAATGGCCCAGTCCTCTCCCAAGTCCACACAGGGGAGGTGA TAGCATTGCTTTCGTGTAAATTATGTAATGCAAAATTTTTTTAATCTTCGCCTTAATACTTTT TTATTTTGTTTTATTTTGAATGATGAGCCTTCGTGCCCCCCCTTCCCCCTTTTTTGTCCCCC AACTTGAGATGTATGAAGGCTTTTGGTCTCCCTGGGAGTGGGTGGAGGCAGCCAGG GCTTACCTGTAC	actgacttga-gaccagttg	agtgcacac-cttaaaaat	GA	ggccaa gtgtgac	TTTGTGGTG TGGCTGGGTT GGGGG
T.ACTB.3	333	ACTB reference sequence, short USE sequence prior to PAS for both mRNA variants, 100-bp after poly (A) site	TTGCTTTCGTGTAAATTATGTAATGCAAAATTTTTTTAATCTTCGCCTTAATACTTTTTTATT TTGTTTTATTTTGAATGATGAGCCTTCGTGCCCCCCCTTCCCCCTTTTTTGTCCCCCAACTT GAGATGTATGAAGGCTTTTGGTCTCCCTGGGAGTGGGTGGAGGCAGCCAGGGC TTACCTGTAC	actgacttga-gaccagttg	agtgcacacct-taaaaat	GA	ggccaa gtgtgac	TTTGTGGTGT GGCTGGGTTG GGGGCAGCA GAGGGTGAAC-CCTGCAGG AGGGTGAAC CCTGCAAAA GGGTGGGGC AGTGGGG GCCAAC
T.ACTB.4	270	ACTB reference sequence, short USE	TTGCTTTCGTGTAAATTATGTAATGCAAAATTTTTTTAATCTTCGCCTTAATACTTTTTTATTT TGTTTTATTTTGAATGATGAGCCTTCGTGCCCCCCCTTCCCCCTTTTTTGTCCCCCAACT TGAGATGTATGAAGGCTTTTGGTCTCCCTGGGAGTGGGTGGAGGCAGCCAGGGCT TACCTGTAC	actgacttga-gaccagttg	agtgcacac-cttaaaaat	GA	ggccaa gtgtgac	TTTGTGGTG TGGCTGGGT sTGGGGG

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263−1275

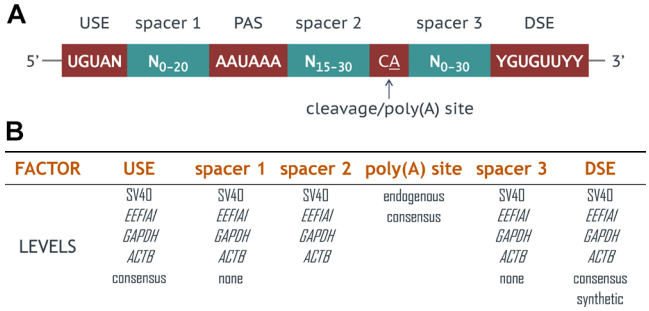


Figure 1. Generic structure of the mammalian terminator and variants evaluated in this study. (A) A depiction of the generic structure of mammalian terminators as previously described.[12] USE represents the upstream sequence element based on a consensus sequence,[20] PAS represents the highly conserved hexameric polyadenylation signal,[10] and DSE represents the downstream sequence element.[22] (B) Depiction of key terminator elements and their variants investigated in this study.

Table 1. continued

terminator	total length	description	USE (T-rich)	spacer 1 (0–20 bp)	spacer 2 (15–30 bp)	poly(A) site	spacer 3 (0–20 bp)	DSE (T- or G/T-rich)
		sequence prior to PAS for both mRNA variants, 37-bp after poly (A) site						

promoter pEEF1A1.1356 derived from *EEF1A1*, commonly referred to as EF1α) in the context of hrGFP output.

First, we evaluated the impact on changing the transgene of interest preceding the terminator. Overall, the evaluated terminators exhibited a similar trend in SEAP output as compared with the hrGFP data (Figure 2B) with some deviations. Specifically, this subset of terminators afforded an 11-fold dynamic range of SEAP output (compared with the 5.7-fold range with hrGFP shown in Figure 2A). Interestingly, the *ACTB*-derived terminator (T.ACTB.4) performance was only minimally influenced by this change transgene sequence (and behaved similarly to the SV40 terminator), whereas the *GAPDH*-derived terminator (T.GAPDH.2) was only half as functional in this context (compare Figure 2A,B). Of the synthetic terminators tested, only Tm.synth.19 and Tm.synth.20 were comparable to SV40 in terms of reporter protein output (Supporting Information Table S3). While Tm.synth.20 was likewise a relatively high performing terminator with hrGFP, the change in function for Tm.synth.19 (compare Figure 2A,B) illustrates a known interaction that can occur between terminators and preceding coding sequence.[13]

Second, we evaluated the impact on terminator performance when the promoter was replaced from the originally used moderate strength promoter (pEIF4A1.636) to the strong endogenous promoter (pEEF1A1.1356). The data demonstrates a reduction in the overall dynamic range of hrGFP level to approximately 3.5-fold (Figure 2c). However, this reduction is expected with stronger promoters as the abundant transcript level (even without half-life enhancement) is beginning to saturate the translation machinery. Additionally, the endogenous terminators T.GAPDH.2 and T.ACTB.4 demonstrated a context dependency, similar to that seen with SEAP above (Figure 2C and Supporting Information Table S4). Finally, under the context of the strong promoter (as highlighted by the marked increase in gMFI when comparing Figure 2A,C), 8 of the 12 selected synthetic terminators behaved comparably to the SV40 terminator (Figure 2C). These small deviations notwithstanding, these results demonstrate that the synthetic terminators in particular are able to function across genetic context.

Terminators Exhibit Functionality in a Different Host Cell Line. As a final test of generalizable function for the

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263–1275

Table 2. Synthetic Terminator Sequences with Putative Annotations of Terminator Components[a]

terminator	total length	description	USE (T-rich)	spacer 1 (0–20 bp)	spacer 2 (15–30 bp)	poly(A) site	spacer 3 (0–20 bp)	DSE (T- or G/T-rich)
Tm.synth.1	113	GAPDH long USE, GAPDH spacer 1, GAPDH spacer 2, GAPDH poly(A) site, GAPDH spacer 3, DSE consensus	TGTAGACCCCTTGAAGAG GGGAGGGGCCTAGG-GAGCCGCACCTT GTCATGTAC	catc	gtaccctg tgctcaaccagt	TA	cttgtcct gtcttattctag	TGTGTTTT
Tm.synth.2	128	GAPDH long USE, GAPDH spacer 1, GAPDH spacer 2, GAPDH poly(A) site, GAPDH spacer 3, Levitt consensus	TGTAGACCCCTTGAAGAG GGGAGGGGCCTAGG-GAGCCGCACCTT GTCATGTAC	catc	gtaccct gtgctcaaccagt	TA	cttgtcct gtcttattctag	TCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.3	136	GAPDH long USE, GAPDH spacer 1, GAPDH spacer 2, GAPDH poly(A) site, GAPDH spacer 3, DSE + Levitt consensus	TGTAGACCCCTTGAAGAGG GGAGGGGCCTAGGGAGC CGCACCTTGTCATGTAC	catc	gtaccctg tgctcaaccagt	TA	cttgtcct gtcttattctag	TGTGTTTTTCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.4	109	4× USE consensus, GAPDH spacer 1, GAPDH spacer 2, GAPDH poly(A) site, GAPDH spacer 3, GAPDH DSE	TGTAATGTAATGTAATG-TAA	catc	gtaccctg tgctcaaccagt	TA	cttgtcctg tcttattctag	GGTCTGGGGCAGAGGGGAGGGAAGCTGGGCTTGTGTC
Tm.synth.5	80	4× USE consensus, GAPDH spacer 1, GAPDH spacer 2, GAPDH poly(A) site, GAPDH spacer 3, DSE consensus	TGTAATGTAATGTAATG-TAA	catc	gtaccctg tgctcaaccagt	TA	cttgtcct gtcttattctag	TGTGTTTT
Tm.synth.6	95	4× USE consensus, GAPDH spacer 1, GAPDH spacer 2, GAPDH poly(A) site, GAPDH spacer 3, Levitt consensus	TGTAATGTAATGTAATG-TAA	catc	gtaccct gtgctcaaccagt	TA	cttgtcct gtcttattctag	TCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.7	103	4× USE consensus, GAPDH spacer 1, GAPDH spacer 2, GAPDH poly(A) site, GAPDH spacer 3, DSE + Levitt consensus	TGTAATGTAATGTAATG-TAA	catc	gtaccctg tgctcaaccagt	TA	cttgtcct gtcttattctag	TGTGTTTTTCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.8	56	4× USE consensus, GAPDH spacer 2, GAPDH poly(A) site, DSE consensus	TGTAATGTAATGTAATG-TAA		gtaccct gtgctcaaccagt	TA		TGTGTTTT
Tm.synth.9	71	4× USE consensus, GAPDH spacer 2, GAPDH poly(A) site, Levitt consensus	TGTAATGTAATGTAATG-TAA		gtaccct gtgctcaaccagt	TA		TCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.10	79	4× USE consensus, GAPDH spacer 2, GAPDH poly(A) site, DSE + Levitt consensus	TGTAATGTAATGTAATG-TAA		gtaccct gtgctcaaccagt	TA		TGTGTTTTTCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.11	71	EEF1A1 short USE, EEF1A1 spacer 1, EEF1A1 spacer 2, EEF1A1 poly(A) site, EEF1A1 spacer 3, DSE consensus	TGGTAT	tcattacaa acttgctcactac	tgaattttaagcttt	AA	gatgaagtggca	TGTGTTTT
Tm.synth.12	86	EEF1A1 short USE, EEF1A1 spacer 1, EEF1A1 spacer 2, EEF1A1 poly(A) site, EEF1A1 spacer 3, Levitt consensus	TGGTAT	tcattaca aacttgctcactac	tgaattttaagcttt	AA	gatgaagtggca	TCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.13	94	EEF1A1 short USE, EEF1A1 spacer 1, EEF1A1 spacer 2, EEF1A1 poly(A) site, EEF1A1 spacer 3, DSE + Levitt consensus	TGGTAT	tcattacaa acttgctcactac	tgaattttaagcttt	AA	gatgaagtggca	TGTGTTTTTCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.14	187	4× USE consensus, EEF1A1 spacer 1, EEF1A1 spacer 2, EEF1A1 poly(A) site, EEF1A1 spacer 3, EEF1A1 long DSE	TGTAATGTAATGTAATG-TAA	tcattac aaacttgctc actac	tgaattttaagcttt	AA	gatgaagtggca	TTTCTTTTAACAGTTACTATGTTGGAATTGGTTACAA ATTTTGGAGTGGATTTCAAAAGTGAGAGCTAAC TTCAGTTGATTTCAAGGTAGTGCTTGGCTTTTTTT GTTTA
Tm.synth.15	85	4× USE consensus, EEF1A1 spacer 1, EEF1A1 spacer 2, EEF1A1 poly(A) site, EEF1A1 spacer 3, DSE consensus	TGTAATGTAATGTAATG-TAA	tcattac aaacttgctc actac	tgaattttaagcttt	AA	gatgaagtggca	TGTGTTTT
Tm.synth.16	100	4× USE consensus, EEF1A1 spacer 1, EEF1A1 spacer 2, EEF1A1 poly(A) site, EEF1A1 spacer 3, Levitt consensus	TGTAATGTAATGTAATG-TAA	tcattacaaacttgc tcactac	tgaattttaagcttt	AA	gatgaagtggca	TCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.17	108	4× USE consensus, EEF1A1 spacer 1, EEF1A1 spacer 2, EEF1A1 poly(A) site, EEF1A1 spacer 3, DSE + Levitt consensus	TGTAATGTAATGTAATG-TAA	tcattacaa acttgctcactac	tgaattttaagcttt	AA	gatgaagtggca	TGTGTTTTTCTGTGTGTTGGTTTTTTGTGTG

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263–1275

Table 2. continued

terminator	total length	description	USE (T-rich)	spacer 1 (0–20 bp)	spacer 2 (15–30 bp)	poly(A) site	spacer 3 (0–20 bp)	DSE (T- or G/T-rich)
Tm.synth.18	51	4× USE consensus, EEF1A1 spacer 2, EEF1A1 poly(A) site, DSE consensus	TGTAATGTAATGTAATG-TAA		tgaattttaagcttt	AA		TGTGTTTT
Tm.synth.19	66	4× USE consensus, EEF1A1 spacer 2, EEF1A1 poly(A) site, Levitt consensus	TGTAATGTAATGTAATG-TAA		tgaattttaagcttt	AA		TCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.20	74	4× USE consensus, EEF1A1 spacer 2, EEF1A1 poly(A) site, DSE + Levitt consensus	TGTAATGTAATGTAATG-TAA		tgaattttaagcttt	AA		TGTGTTTTTCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.21	77	4× USE consensus, ACTB spacer 2, ACTB poly(A) site, DSE + Levitt consensus	TGTAATGTAATGTAATG-TAA		agtgcaca ccttaaaaat	GA		TGTGTTTTTCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.22	77	4× USE consensus, ACTB spacer 2, poly(A) site consensus, DSE + Levitt consensus	TGTAATGTAATGTAATG-TAA		agtgcaca ccttaaaaat	CA		TGTGTTTTTCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.23	56	4× USE consensus, GAPDH spacer 2, poly (A) site consensus, DSE consensus	TGTAATGTAATGTAATG-TAA		gtaccctgt gctcaaccagt	CA		TGTGTTTT
Tm.synth.24	71	4× USE consensus, GAPDH spacer 2, poly (A) site consensus, Levitt consensus	TGTAATGTAATGTAATG-TAA		gtaccct gtgctcaaccagt	CA		TCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.25	79	4× USE consensus, GAPDH spacer 2, poly (A) site consensus, DSE + Levitt consensus	TGTAATGTAATGTAATG-TAA		gtaccctgt gctcaaccagt	CA		TGTGTTTTTCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.26	51	4× USE consensus, EEF1A1 spacer 2, poly (A) site consensus, DSE consensus	TGTAATGTAATGTAATG-TAA		tgaattttaagcttt	CA		TGTGTTTT
Tm.synth.27	66	4× USE consensus, EEF1A1 spacer 2, poly (A) site consensus, Levitt consensus	TGTAATGTAATGTAATG-TAA		tgaattttaagcttt	CA		TCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.28	74	4× USE consensus, EEF1A1 spacer 2, poly (A) site consensus, DSE + Levitt consensus	TGTAATGTAATGTAATG-TAA		tgaattttaagcttt	CA		TGTGTTTTTCTGTGTGTTGGTTTTTTGTGTG
Tm.synth.29	67	4× USE consensus, GAPDH spacer 2, poly (A) site consensus, MC4R DSE[24]	TGTAATGTAATGTAATG-TAA		gtaccctgtg ctcaaccagt	CA		CGTGTTATTCATAAGCATT
Tm.synth.30	61	4× USE consensus, GAPDH spacer 2, poly (A) site consensus, #7 DSE from Pérez Cañadillas *et al.* (CstF-64 RRM)[30]	TGTAATGTAATGTAATG-TAA		gtaccctg tgctcaaccagt	CA		GTTGTGTGTGTTG

[a]Note that blank entries for the spacer 1 or spacer 3 denote a lack of spacer element used.

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263–1275

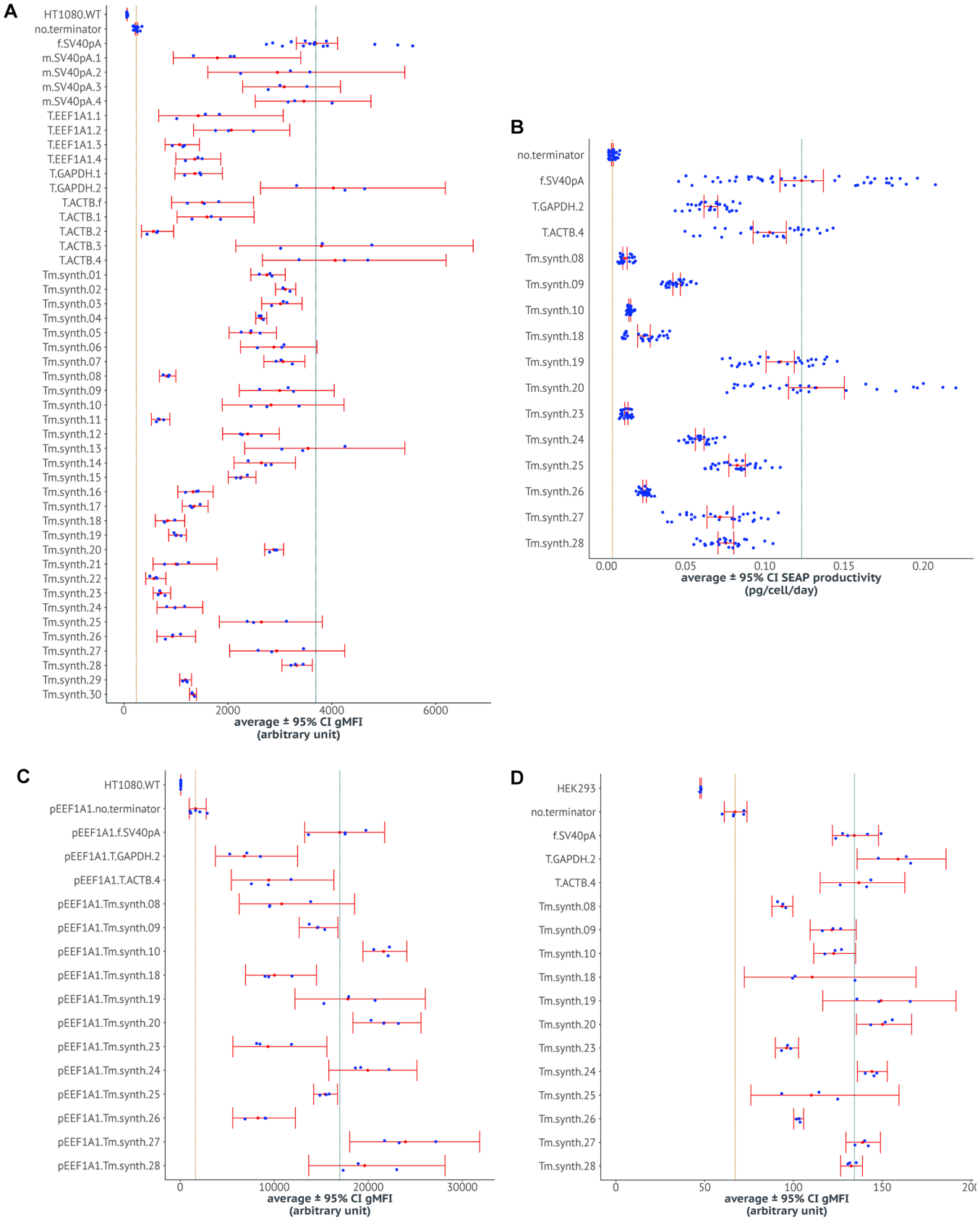


Figure 2. Transient reporter protein output 48 h post-transfection with various terminators cell lines. (A) hrGFP protein output reported as average geometric mean fluorescence intensity (gMFI) in arbitrary units in HT1080. Error bars represent the 95% CI (red bars) of the average gMFI (red point) from at least 3 independent transfections/replicates (blue points). (B) SEAP output reported as average specific productivity in pg/cell/day in HT1080 with subset of terminators. Error bars represent the 95% CI (red bars) of the average specific productivity (red point) from at least 3 independent transfections/replicates (blue points). (C) hrGFP output when pairing various terminators with a strong promoter

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263−1275

Figure 2. continued

(pEEF1A1.1356) as average geometric mean fluorescence intensity (gMFI) in arbitrary units in HT1080. Error bars represent the 95% CI (red bars) of the average gMFI (red point) from at least 3 independent transfections/replicates (blue points). (D) hrGFP output when pairing various terminators with the *EIF4A1*-derived promoter (pEIF4A1.636) as average geometric mean fluorescence intensity (gMFI) in arbitrary units in HEK293. Error bars represent the 95% CI (red bars) of the average gMFI (red point) from at least 3 independent transfections/replicates (blue points).

endogenous and synthetic terminator variants, we sought to verify whether these elements could function in an alternative mammalian host, in this case, HEK293, using the moderate strength promoter (pEIF4A1.636) linked with hrGFP. The resulting data suggests that each terminator variant tested was able to function as a terminator (Figure 2D), and many of these variants behaved comparably to the reference terminator from SV40 (Supporting Information Table S5). A decreased dynamic range of hrGFP output was observed in HEK293 which can be attributed primarily to the lower pEIF4A1.636 promoter functionality in these cells (compare hrGFP output with f.SV40pA in Figure 2A,D). Collectively, the general terminator scaffold used in this work confirms the necessary and sufficient composition for functionality, and tuning these key elements within the scaffold can be an effective strategy to design novel terminators.

Extracting Key Composition Features Leading to Mammalian Terminator Function. In an effort to expand the design-build-test cycle observed in this paper so far, we opted to add a "learn" component in which we investigated the interactions between the key scaffold elements. In particular, we compared hrGFP and SEAP outputs within a subset of the synthetic terminators, specifically interrogating the effects of spacer 2 (between the PAS and the polyadenylation site), the polyadenylation site, and the DSE sequence independently. To perform this analysis, we specifically evaluated Tm.synth.8−10, 18−20, and 23−28 (Table 2 and Supporting Information Table S12), as they contain permutations of these three key terminator elements.

Transient hrGFP output driven by the moderate strength promoter (pEIF4A1.636) was evaluated using ANOVA followed by Tukey's HSD *post hoc* testing to uncover sequence determinants of function. Specifically, this analysis demonstrates that the DSE can significantly impact terminator functionality and protein output ($p < 0.001$). Synthetic DSE sequences were functional from the reporter output data (Figure 1A) and were stronger than the DSE consensus (Supporting Information Table S6). A broader ANOVA across the entire set of terminator designs still revealed significant interactions between the choice of the DSE element, spacer 2, and the polyadenylation site (Supporting Information Table S7). Likewise, an analysis conducted using the SEAP output driven by the pEIF4A1.636 promoter and hrGFP output driven by the pEEF1A1.1356 promoter revealed the same central role that the DSE has on output enhancement (Supporting Information Table S8 and S9).

Furthermore, our analysis suggests that the DSE itself can critically impact the resulting protein output in a sequence-dependent manner and the DSE consensus sequence alone is sufficient for activity (Figure 2A), affirming previous work that a potent DSE can be sufficient for transcript termination.[24] Although the DSE consensus was sufficient for functionality, we observed significantly enhanced protein output with other synthetic DSEs, suggesting that a consensus sequence may not be optimal (Supporting Information Table S6). These observations were particularly evident between the *EEF1A1*-derived terminator variants and its synthetic variants with different DSEs (compare T.EEF1A1.1, T.EEF1A1.2, Tm.synth.11, Tm.synth.12, and Tm.synth.13, Figure 2A). Notably, these resulting protein outputs did not correlate with concomitant changes to DSE sequence length, suggesting that the sequence footprint of terminator would be a poor predictor of its impact on protein output.

Moreover, the data corroborates previous bioinformatics analyses suggesting that the DSE has a greater impact on polyadenylation processing than the USE.[31] As evidence to support this, changing the endogenous USE sequence to the "consensus 4× repeat" did not consistently alter protein output from using the *EEF1A1*-derived (compare T.EEF1A1.2, T.EEF1A1.4, and Tm.synth.14) and *GAPDH*-derived (compare T.GAPDH.1, T.GAPDH.2, and Tm.synth.04) terminators (Figure 2A). Likewise, altering the endogenous polyadenylation site to the dinucleotide consensus sequence (CA) had no measurable impact on protein output (Supporting Information Tables S7−S9). Finally, although certain combinations of the spacer 2, the polyadenylation site, and the DSE had significant impact on protein output, this impact was masked when expression was driven by a strong promoter (Supporting Information Table S7−S9). Ultimately, these results highlight that terminator function and the resulting protein output can be modulated by altering the design of these key elements.

Terminators Function through Increased mRNA Stability. On the basis of prior studies in yeast by our group,[13,17] we hypothesized that the terminators were mechanistically increasing protein output though stabilizing mRNA. In such a model, the mRNA decay constant of a particular gene should have an inverse linear relationship with its corresponding protein output level holding all other factors equal (see Supporting Information for equations). To test this, we performed RT-PCR with RNA isolated at different time points after the addition of actinomycin D, a known transcriptional inhibitor. Six different terminators that span the observed dynamic range (T.GAPDH.1, T.GAPDH.2, Tm.Synth.2, Tm.Synth.15, Tm.Synth.17, and Tm.Synth.22) and the no terminator control were analyzed *via* this approach and normalized to the viral-derived SV40 terminator as the reference point. For each terminator evaluated, decay constants were plotted against the inverse of the fluorescence and led to a statistically significant linear relationship based on Deming regression analysis, estimating the 95% CI of the slope as (17.2, 150) (Figure 3). These observations and this nonzero slope suggest that these terminators are functioning to increase mRNA stability, effectively leading to overall higher protein levels as a result.

Designed Promoter-Terminator Elements Can Outperform Conventional, Viral-Derived Sequences. On the basis of the performance of select endogenous and synthetic terminators compared with the SV40 terminator (Figure 2A), we expected that these novel terminators are suitable replacements viral-derived SV40 terminator. Therefore, this

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263−1275

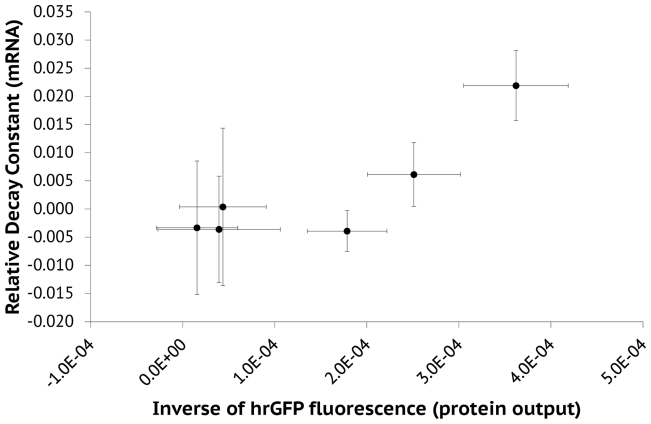


Figure 3. Correlation between the measured mRNA decay constant and the inverse of its measured fluorescence measurement based on the hrGFP reporter. The mRNA decay constants of hrGFP under various terminator contexts were determined with a RT-PCR assay after stopping transcription with actinomycin D. A linear correlation implies mRNA stability is impacting the resulting fluorescence measurement due to the terminator designs. Error bars represent independent transfections (triplicates) in both dimensions. Deming regression was used to determine the statistical significance of the relationship.

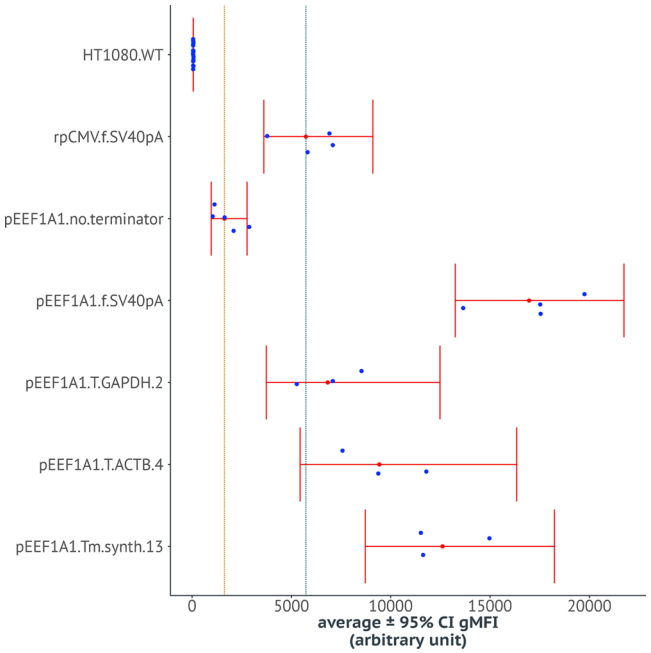


Figure 4. Comparison of strong viral-derived elements with fully endogenous/synthetic elements for regulating gene expression. Transient hrGFP expression 48 h post-transfection in HT1080 cells reported as average geometric mean fluorescence intensity (gMFI) in arbitrary units. Error bars represent the 95% CI (red bars) of the average gMFI (red point) from at least 3 independent transfections/replicates (blue points). rpCMV corresponds to reference CMV promoter, pEEF1A1 corresponds to full length 1356-bp promoter derived from *EEF1A1*.

finding raises the prospect of completely replacing strong viral-derived regulatory elements. To test this premise, we compared a viral-derived transgene cassette consisting of the common CMV promoter to drive hrGFP expression paired with the f.SV40pA terminator. In contrast, we paired fully human-derived endogenous and synthetic terminators (T.GAPDH.2, T.ACTB.4, and Tm.synth.13) with the strong endogenous promoter derived from *EEF1A1* (pEEF1A1.1356). Upon construction and evaluation, the fully endogenous/synthetic constructs were comparable to, and 1.9× stronger (Tm.synth.13) than, the viral-derived construct. These results highlight that viral-derived components are not necessary to obtain high transient transgene expression in HT1080 (Figure 4 and Supporting Information Table S10). Interestingly, these novel variants were both 1.5× larger (333-bp) and nearly 2.4× smaller (94-bp) in sequence length (Tables 1 and 2) relative to the viral terminator (222-bp), again implying that sequence length is not a necessary determinant of strong terminator activity.

CONCLUSION

In this work, we generated and characterized the impact of 10 endogenous and 30 synthetic terminator variants on their ability to modulate gene expression. Critically, all of these variants were functional, and many variants had a significantly reduced DNA sequence footprint compared with the commonly used strong, viral terminator derived from SV40. For the initial evaluation in this study, we opted to use a moderate strength promoter as it enables a broader dynamic range and evaluation of terminator performance. As a regulatory element, the terminator would ideally function in a predictable manner regardless of the transgene and genetic context. However, our subsequent characterizations demonstrated that some context dependency does exist with respect to the transgene and promoter chosen. While the relative measured gene expression levels between terminators were inconsistent when the protein reporter was changed (illustrating a high impact of upstream sequence on mRNA folding), a large dynamic range and highly functional terminators could be identified from this small collection. Thus, these elements can be used as a regulatory element with only a small screening exercise for context. In this regard, these results align with our previous work observing similar phenomena in *S. cerevisiae*.[17]

The central premise for this engineering effort was to replace viral-derived regulatory elements with endogenous and synthetic counterparts in a manner that does not impact function. Specifically, by combining techniques to engineer synthetic promoters[29] along with these endogenous and synthetic terminators, a full conversion of viral-derived regulatory sequences to purely endogenous and synthetic sequences can be realized. Moreover, this replacement was not detrimental to gene expression, but instead we achieved comparable or stronger target gene expression when using endogenous/synthetic elements. The comparable activity offered by these terminator variants with smaller sequence footprint is an attractive feature in modern genetic engineering applications that incorporate multicistron designs, and alleviates the need for redundant terminator sequences in the same vector. While this smaller footprint can also be appreciated when used with DNA virus vectors that have established limitations on cargo capacity,[32] their use with emerging methods of genome engineering can result in a fully nonviral approach to accomplish this task. Technologies such as the *piggyBac* transposon system derived from the cabbage looper moth *Trichoplusia ni* can mediate effective gene transfer even in human primary T cells,[33] enabling the design of novel cellular therapy applications without viral-derived genetic elements.

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263−1275

Collectively, these results highlight the prospects of terminator mutants, but illustrates that further understanding is necessary to remove context dependency. This work uncovered interactions between key terminator elements, but design rules for predictive terminator design were not evident from this work. Similar to previous efforts in yeast,[13,17] this work expands the set of characterized terminators available for tuning gene expression. Ultimately, the ability to use terminators as synthetic regulatory elements in mammalian cells opens the door to precise gene control and metabolic pathway engineering.

METHODS

Plasmid Construction and Preparation for Cell Culture. Plasmids were constructed by a combination of standard molecular biology techniques (PCR, Gibson assembly, In-Fusion HD cloning kit). The essential elements required for plasmid propagation in *E. coli* (origin, beta-lactamase) from pUC19 were subcloned with the conventional SV40 late terminator and multiple cloning sites (*Eco*RI, *NotI*, *NheI*, *PmeI*, *XbaI*, and *NsiI*) using Gibson assembly to generate the base vector from previous work.[29] Subsequently, the SEAP and hrGFP sequences were subcloned from other plasmids into these cloning sites from previous work.[29] Endogenous sequences were cloned from genomic DNA extracted from HT1080 WT culture with Wizard Genomic DNA Purification Kit (Promega Corporation). Synthetic sequences were produced as oligonucleotides (IDT) for assembly. Vector variants containing the SEAP reporter were constructed using Gibson Assembly by replacing the hrGFP reporter. Promoter vector variants were constructed using the In-Fusion HD cloning kit by removing the pEIF4A1.636 promoter *via* *NotI-HF*/*NheI-HF* digestion and subsequent gel extraction of the plasmid backbone. Clones containing plasmid DNA were screened from transformation of Stellar chemically competent cells (Takara Bio). Finally, plasmid DNA was extracted from 150 mL of bacterial culture using the Zymo Research ZymoPURE Maxi Prep kit, and further purified by ethanol precipitation.

Cell Culture and Transient Transfections. Briefly, HT1080 WT and HEK293F cells were maintained at in a humidified 37 °C incubator under 5% CO_2 with 125 mL shake flasks. To establish transient expression cultures in HT1080, three batches of 12×10^6 viable cells were resuspended in RPMI-1640 media for HT1080 (0.75 mL per cuvette) and transfected each with 50 μg of plasmid DNA (harboring our reporter expression cassette with various terminators) using a 4 mm electroporation cuvette and Gene Pulser XCell (BioRad) at 350 V and 950 μF of capacitance. Similarly, in HEK293F, three batches of 12×10^6 viable cells were resuspended in HyClone SFM4Transfx-293 media (0.75 mL per cuvette) and transfected each with 35 μg of plasmid DNA (harboring our reporter expression cassette with various terminators) using a 4 mm electroporation cuvette and Gene Pulser XCell (BioRad) at 374 V and 950 μF of capacitance.

Expression Quantification Using Flow Cytometry and Chemiluminescent Reporter Assay. Briefly, as previously described:[29] $1.0-2.0 \times 10^6$ cells were pelleted and suspended in sterile 1× DPBS 48-h post-transfection for both HT1080 and HEK293 cell types. These cell suspensions were analyzed using the BD LSRII Fortessa (UT Institute of Cellular and Molecular Biology Core Facility) with the following parameters: 340 V FSC, 176 V SSC, 200 V GFP (HT1080) or 180 V GFP (HEK293F). 25 000−30 000 events were collected from a subpopulation that represented live cells (as determined by gating WT cells, see Supporting Information Figure S1 for representative cytometry plots) and these live cells were divided between GFP^- or GFP^+ by setting a threshold such that 1% of the WT cells are considered GFP^+ as an approximation of autofluorescence. For the subset of terminators with mStrawberry coexpression, RFP^+ cells (275 V) were gated from the live cell subpopulation prior to GFP^+/GFP^- discrimination (see Supporting Information Figure S2 for representative cytometry plots). All other samples were analyzed using the same gate for live cell subpopulation and threshold for GFP^+ cells. The supernatant obtained from centrifugation to pellet cells was sampled for analysis using the NovaBright Secreted Placental Alkaline Phosphatase (SEAP) Enzyme Reporter Gene Chemiluminescent Detection System 2.0 (Invitrogen).

Experimental Conditions for Half-Life Determination. Wild-type HT1080 cells were started from a working cell bank and grown for 1 week in CD50/50 media, passaging every 2−3 days. Sensitivity to actinomycin D was determined by growing cells at 4×10^5 cells/mL in fresh CD50/50 media containing 0, 0.5, 1, or 5 μg/mL of actinomycin D in culture duplicate and viability analyzed with a ViCELL XR Cell Viability Analyzer (Beckman Coulter) every 24 h. All three concentrations of actinomycin D failed to observe any growth after 48 h, so 1 μg/mL was chosen for subsequent experiments.

To perform the transfections, wild-type HT1080 were grown from glycerol stock in CD50/50 media, passaging every 2−3 days for a week. Cultures were then expanded in grown in CD50/50 media so as to have 1×10^7 cells per transfection. Cells were centrifuged and resuspended in RPMI prior to transfection. Three transfections were performed separately for each construct with 40 μg of DNA for each transfection. Transfections were done by electroporation at 350 V and 950 μF and incubated at room temperature for 15 min prior to pipetting cells to flasks containing prewarmed CD50/50 media. Cultures were then outgrown and allowed to recover for 60 h.

Small aliquots from each culture were analyzed *via* flow cytometry using a FACS Fortessa (BD) in biological triplicate. Then, 1 μg/mL of actinomycin D was added to each flask and gently mixed. After addition of actinomycin D, aliquots were taken from each flask containing 0.5 to 1.0×10^6 cells at 0, 4, 8, and 12 h. Samples were centrifuged and the media pipetted off before resuspending in 100 μL of sterile PBS and 575 μL of RNALater (Invitrogen) and stored at −20 °C.

Before RNA extraction, samples were thawed at room temperature. RNA was extracted using the PureLink RNA Mini Kit with Trizol (Thermo) following manufacturer's instructions. Concentrations were determined using a Nanodrop 2000 (Thermo Fisher Scientific). Then, 1 μg of RNA was digested with DNase I and purified with the GeneJetRNA Cleanup and Concentration Kit (Thermo), following the manufacturer's instruction for DNase reaction and removal. An RIN was not determined. Reverse transcription was done using the Protoscript II First Strand cDNA Synthesis Kit (NEB) following manufacturer's standard protocol with 20pmol of a gene specific primer toward GFP (Q-hrGFP-R4: CTTGC-TCTTCATCAGGGTGC) and 200 ng of RNA per 20 μL reaction. To test for DNA contamination, one replicate from each of the constructs was run with and without reverse transcriptase. The no RT control was 16−20 cycles separated

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263−1275

from the cDNA samples, suggesting very little DNA contamination. cDNA was then stored at −20 °C.

qPCR was run using Power SYBR Green Master Mix (Thermo Fisher Scientific) according to manufacturer's instructions on a ViiA 7 (Thermo Fisher Scientific) using primers specific toward GFP (Q-hrGFP-F4: GTGGTGTACATGAACGACGG Q-hrGFP-R4: CTTGCTCTTCATCAGGGTGC) and analyzed with a standard curve generated with dilutions of plasmid DNA. Primer efficiency was measured to be 98.3%, R^2 was 0.998, and the linear dynamic range ranged from 2.65×10^8 to 2.65×10^3 copies per μL. Weighted linear regression was used for fitting the standard curve, accounting for errors in technical triplicates at each concentration. Cycle times of samples were within the linear dynamic range. Melt curves showed a single peak for the amplicons. QuantStudio Real-Time PCR Software was used to analyze the amplification curves and extract cycle times.

Software for Data Analyses. ANOVA and Tukey's HSD *post hoc* tests were performed with R: The R Project for Statistical Computing (verified in v3.5.0), and Deming regression analysis was conducted in MS Excel 2013.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.8b00285.

Accession codes of nucleotide sequences; Equations for transcript half-life analysis; Supporting Tables and Figures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (512) 471-4417. E-mail: halper@che.utexas.edu.

ORCID

Hal S. Alper: 0000-0002-8246-8605

Present Address

§Ben Towne Center for Childhood Cancer Research, Seattle Children's Research Institute, 1100 Olive Way, Suite 100, Seattle, Washington 98101, United States.

Author Contributions

H.S.A. conceived the overall experiment/workflow design, J.K.C. conducted the initial protein output experiments and analyses, N.J.M. conducted the experiments and analyzed the data to determine the correlation between protein expression and transcript levels, J.M.W. performed and analyzed the experiments incorporating constitutive reporter expression to verify robust transfection parameters used in the study, and S.K.T. performed the experiments to collect SEAP and hrGFP output with terminator subset. All authors reviewed the results and critically revised and approved of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was partially supported by Shire Human Genetic Therapies and MedImmune, LLC. We thank Annabel Wang (UT-Austin chemical engineering undergraduate) for her assistance with molecular biology work.

■ ABBREVIATIONS

SV40, Simian Virus 40; CMV, cytomegalovirus; hrGFP, humanized *Renilla* green fluorescent protein; SEAP, secreted alkaline phosphatase; RFP, red fluorescent protein, mStrawberry.

■ REFERENCES

(1) Kaufman, R. J. (2000) Overview of vector design for mammalian gene expression. *Mol. Biotechnol. 16*, 151−160.

(2) Proudfoot, N. J., Furger, A., and Dye, M. J. (2002) Integrating mRNA processing with transcription. *Cell 108*, 501−512.

(3) West, S., and Proudfoot, N. J. (2009) Transcriptional termination enhances protein expression in human cells. *Mol. Cell 33*, 354−364.

(4) Antoniou, M., Geraghty, F., Hurst, J., and Grosveld, F. (1998) Efficient 3′-end formation of human β-globin mRNA in vivo requires sequences within the last intron but occurs independently of the splicing reaction. *Nucleic Acids Res. 26*, 721−729.

(5) Baek, K. H., Sato, K., Ito, R., and Agarwal, K. (1986) Rna Polymerase-Ii Transcription Terminates at a Specific DNA-Sequence in a Hela Cell-Free Reaction. *Proc. Natl. Acad. Sci. U. S. A. 83*, 7623−7627.

(6) Sato, K., Ito, R., Baek, K. H., and Agarwal, K. (1986) A specific DNA sequence controls termination of transcription in the gastrin gene. *Mol. Cell. Biol. 6*, 1032−1043.

(7) Connelly, S., and Manley, J. L. (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev. 2*, 440−452.

(8) Carswell, S., and Alwine, J. C. (1989) Efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences. *Mol. Cell. Biol. 9*, 4248−4258.

(9) Schek, N., Cooke, C., and Alwine, J. C. (1992) Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses. *Mol. Cell. Biol. 12*, 5386−5393.

(10) Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J. M., and Gautheret, D. (2000) Patterns of variant polyadenylation signal usage in human genes. *Genome Res. 10*, 1001−1010.

(11) Zarudnaya, M. I., Kolomiets, I. M., Potyahaylo, A. L., and Hovorun, D. M. (2003) Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Res. 31*, 1375−1386.

(12) Proudfoot, N. J. (2011) Ending the message: poly(A) signals then and now. *Genes Dev. 25*, 1770−1782.

(13) Curran, K. A., Morse, N. J., Markham, K. A., Wagman, A. M., Gupta, A., and Alper, H. S. (2015) Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *ACS Synth. Biol. 4*, 824−832.

(14) Zhao, W., and Erle, D. J. (2018) Widespread Effects of Chemokine 3′ Untranslated Regions on mRNA Degradation and Protein Production in Human Cells. *J. Immunol. 201*, 1053−1061.

(15) Kim, D., Kim, J. D., Baek, K., Yoon, Y., and Yoon, J. (2003) Improved mammalian expression systems by manipulating transcriptional termination regions. *Biotechnol. Prog. 19*, 1620−1622.

(16) Morse, N. J., Gopal, M. R., Wagner, J. M., and Alper, H. S. (2017) Yeast Terminator Function Can Be Modulated and Designed on the Basis of Predictions of Nucleosome Occupancy. *ACS Synth. Biol. 6*, 2086−2095.

(17) Curran, K. A., Karim, A. S., Gupta, A., and Alper, H. S. (2013) Use of expression-enhancing terminators in Saccharomyces cerevisiae to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metab. Eng. 19*, 88−97.

(18) West, S., Proudfoot, N. J., and Dye, M. J. (2008) Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Mol. Cell 29*, 600−610.

(19) Proudfoot, N. (1991) Poly(A) signals. *Cell 64*, 671−674.

(20) Danckwardt, S., Kaufmann, I., Gentzel, M., Foerstner, K. U., Gantzert, A. S., Gehring, N. H., Neu-Yilik, G., Bork, P., Keller, W., Wilm, M., Hentze, M. W., and Kulozik, A. E. (2007) Splicing factors

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263−1275

stimulate polyadenylation *via* USEs at non-canonical 3 ′ end formation signals. *EMBO J. 26*, 2658−2669.
(21) Gil, A., and Proudfoot, N. J. (1984) A sequence downstream of AAUAAA is required for rabbit *β*-globin mRNA 3′-end formation. *Nature 312*, 473−474.
(22) McLauchlan, J., Gaffney, D., Whitton, J. L., and Clements, J. B. (1985) The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. *Nucleic Acids Res. 13*, 1347−1368.
(23) Salisbury, J., Hutchison, K. W., and Graber, J. H. (2006) A multispecies comparison of the metazoan 3′-processing downstream elements and the CstF-64 RNA recognition motif. *BMC Genomics 7*, 55.
(24) Nunes, N. M., Li, W., Tian, B., and Furger, A. (2010) A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. *EMBO J. 29*, 1523−1536.
(25) Brown, K. M., and Gilmartin, G. M. (2003) A mechanism for the regulation of pre-mRNA 3′ processing by human cleavage factor Im. *Mol. Cell 12*, 1467−1476.
(26) Venkataraman, K., Brown, K. M., and Gilmartin, G. M. (2005) Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev. 19*, 1315−1327.
(27) Levitt, N., Briggs, D., Gil, A., and Proudfoot, N. J. (1989) Definition of an efficient synthetic poly(A) site. *Genes Dev. 3*, 1019−1025.
(28) Sharova, L. V., Sharov, A. A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M. S. (2009) Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res. 16*, 45−58.
(29) Cheng, J. K., and Alper, H. S. (2016) Transcriptomics-Guided Design of Synthetic Promoters for a Mammalian System. *ACS Synth. Biol. 5*, 1455−1465.
(30) Perez Canadillas, J. M., and Varani, G. (2003) Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J. 22*, 2821−2830.
(31) Legendre, M., and Gautheret, D. (2003) Sequence determinants in human polyadenylation site selection. *BMC Genomics 4*, 7.
(32) Wu, Z., Yang, H., and Colosi, P. (2010) Effect of genome size on AAV vector packaging. *Mol. Ther. 18*, 80−86.
(33) Huang, X., Guo, H., Tammana, S., Jung, Y. C., Mellgren, E., Bassi, P., Cao, Q., Tu, Z. J., Kim, Y. C., Ekker, S. C., Wu, X., Wang, S. M., and Zhou, X. (2010) Gene transfer efficiency and genome-wide integration profiling of Sleeping Beauty, Tol2, and piggyBac transposons in human primary T cells. *Mol. Ther. 18*, 1803−1813.

DOI: 10.1021/acssynbio.8b00285
ACS Synth. Biol. 2019, 8, 1263−1275