The Spinach RNA aptamer as a characterisation tool for synthetic biology

Georgios Pothoulakis[1,2], Benjamin Reeve[1,2] and Tom Ellis*[1,2]

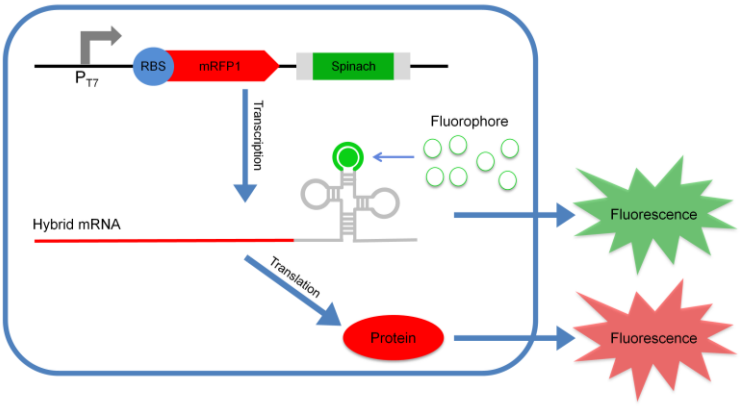
[1] Centre for Synthetic Biology and Innovation, Imperial College London, SW7 2AZ, UK

[2] Department of Bioengineering, Imperial College London, SW7 2AZ, UK

* Corresponding Author – t.ellis@imperial.ac.uk

Abstract

Characterisation of genetic control elements is essential for the predictable engineering of synthetic biology systems. The current standard for *in vivo* characterisation of control elements is through the use of fluorescent reporter proteins such as green fluorescent protein (GFP). Gene expression, however, involves not only protein production but also the production of messenger RNA (mRNA). Here we present the use of the Spinach aptamer sequence, an RNA mimic of GFP, as a tool to characterise mRNA expression in *Escherichia coli*. We show how the aptamer can be incorporated into gene expression cassettes and how co-expressing it with a red fluorescent protein (mRFP1) allows, for the first time, simultaneous measurement of mRNA and protein levels from engineered constructs. Using flow cytometry, we apply this tool here to evaluate ribosome binding site sequences and promoters and use it to highlight the differences in the temporal behaviour of transcription and translation.



Keywords

Measurement, Spinach RNA, Aptamer, Promoter, Ribosome Binding Site, Fluorescence

Accurate characterisation of gene expression control elements is an important part of synthetic biology (*1*). Quantitative and qualitative data informs part choice and is used to define key parameters for predictive models of gene networks and metabolic pathways (*2-4*). Currently, fluorescent proteins such as GFP are the standard tool used to measure gene expression *in vivo*. Linking the expression of these proteins directly to engineered devices allows quantification of their output using high-throughput and high-sensitivity equipment such as fluorescence-enabled multiwell plate readers and flow cytometers (*5*). A major limitation of fluorescent proteins, however, is that they report only on cellular protein production. Gene expression is a multi-step process where DNA is converted to messenger RNA (mRNA) and then to protein, and different control elements act at different stages within this: promoters controlling transcription, and the ribosome binding site (RBS) controlling translation initiation rates. Quantifying gene expression using fluorescent proteins measures only the final product of these stages and does not interrogate them separately. Therefore, there is a paucity of data for gene expression models which ideally would capture transcription and translation independently.

In 2011, researchers at Cornell University described for the first time an RNA sequence that itself could mimic GFP in living cells (*6*). This RNA sequence, known as the Spinach aptamer, fluoresces when expressed in live *Escherichia coli* and mammalian cells in the presence of the provided fluorophore molecule, 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI). Given that the Spinach aptamer requires only a provided fluorophore and a short RNA sequence, it offers an elegant method for characterising the transcriptional step of gene expression separately from translation. We describe here how the Spinach aptamer, when included in a construct expressing a fluorescent protein, can be used to simultaneously measure transcription and translation of mRNAs, thus providing a new tool for gene expression characterisation.

Results and Discussion

To demonstrate the utility of Spinach for measuring gene expression, we first sought to integrate the Spinach aptamer sequence into an mRNA actively translated in *E. coli*. Previously, Spinach has been expressed in *E. coli* from the pET28c-Spinach plasmid which uses the T7 promoter to express a 171nt tRNA scaffold containing the Spinach aptamer (*6*). To create a plasmid that produces a translated mRNA, we inserted the coding sequence (with RBS) for the monomeric red fluorescent protein, mRFP1 (*7*), into pET28c-Spinach downstream of the T7 promoter and upstream of the Spinach aptamer. Both this new construct (pGPR001) and pET28c-Spinach were transformed into BL21(DE3) *E. coli* cells and grown in liquid media at 37°C before being induced with IPTG to initiate T7 RNA polymerase-driven gene expression. 2 hours post-induction, cells were assayed for Spinach and mRFP1 expression by fluorescence microscopy (Figure 1A) and cells from the same cultures were analysed simultaneously using two-colour flow cytometry (Figures 1B and 1D). In the presence of the fluorophore DFHBI,

green fluorescence from the Spinach aptamer can be seen from cells with both the pET28c-Spinach and pGPR001 plasmids. Red fluorescence from the protein mRFP1 is also seen from pGPR001. This verifies that the Spinach aptamer can be incorporated into a translated mRNA and measured *in vivo*. Spinach has previously only been measured in *E. coli* using fluorescent microscopy (*6, 8*). The ability to measure Spinach with analytical flow cytometry allows it to be used for rapid, high-quality gene expression quantification. Flow cytometry also helps identify multiple populations within samples. As can be seen from Figures 1B and 1C, not all cells in an induced culture fluoresce. This can also be seen in the microscopy images (Figure 1A, arrows). This OFF-population consists of cells not expressing either green or red fluorescence, and so we attribute this to cells losing plasmid following cell division or acquiring mutations at the chromosomal gene encoding T7 RNA polymerase (*9*). The burden of high expression of recombinant DNA may be promoting both during our experiments. With flow cytometry, this population can however be readily resolved, allowing Spinach to be measured despite bimodal distributions.

The combination of mRFP1, Spinach and two-colour flow cytometry enables temporal characteristics to be captured during gene expression. To demonstrate this we took a time-series of green and red fluorescence measurements during a 3 hour induction of Spinach and mRFP1-Spinach expression from the T7 promoter (Figure 2A). IPTG added at time zero activates expression of chromosomally-encoded T7 RNA polymerase, which in turn transcribes the constructs (Figure 1C). Transcript expression can be detected 15 mins following induction for Spinach alone, and 30 to 45 mins for mRFP1-Spinach. Transcript copies per cell rise rapidly within the first hour and then stabilise, in contrast to protein copies per cell which rise continually from 30 mins onwards (Figure S1). These data underline the fundamental differences between mRNA and proteins in terms of their half-lives in the cell, with mRNA turning-over in minute timescales and proteins turning-over in hours. It also identifies that transcription of the aptamer alone (pET28c-Spinach) not only yields ~5-fold more green fluorescence than when transcribed in the 3' untranslated region (3'UTR) of the mRFP1 mRNA (pGPR001), but that measurable transcription emerges faster. Slower emergence of green fluorescence from pGPR001 is likely to be a direct consequence of the aptamer not being immediately transcribed by the RNA polymerase. In order to investigate this effect and determine the optimum placement of Spinach within an mRNA we next assembled a series of modified constructs placing the Spinach aptamer at different positions within mRNAs and quantified their fluorescence using flow cytometry as before (Figure 2B). Three alternative designs failed to yield greater green fluorescence per cell than pGPR001 which places Spinach in the 3'UTR. Relocating Spinach to the 5'UTR (pGPR02) gave decreased expression, despite this region being immediately transcribed during transcription. Incorporating two copies of Spinach in tandem in the 3'UTR (pGPR03) also led to decreased fluorescence, presumably due to local homology between neighbouring aptamer sequences. Placing Spinach within a bicistronic operon to shield it from cellular ribonucleases (pGPR04) also failed to increase green fluorescence, and did not yield greater red fluorescence either. Despite care being taken in all designs to insulate parts

with spacer sequences, it is likely that these observations are due to aptamer mis-folding, and this may be being promoted in all designs by the presence of translation machinery enzymes with helicase activities, notably the ribosome (*10*). Mis-folding of mRNA may also lead to changes in mRNA degradation rates and to stalled transcripts, which may block further gene expression.

Crucially, Spinach also allows control elements such as the promoter and RBS to be assessed separately using the same construct. To demonstrate this we designed three RBS sequences (R1K, R9K and R148K) of different predicted strengths using the Salis Lab RBS Calculator (*11, 12*) and incorporated these within T7-expressed pGPR01. Measured green fluorescence from these constructs (i.e. mRNA levels) remained equivalent to that of the parent construct (pGPR01) but red fluorescence (i.e. mRFP1 expression) varied in accordance to the predicted RBS strength (Figure 3A), except when the strongest RBS (R148K) was used, where green fluorescence slightly decreased (pGPR07). In this case, plasmid or transcript copies per cell may be being reduced by the stringent response, a cellular stress response to amino acid starvation that can be induced by the over-expression of recombinant proteins (*13-15*). The use of the Spinach aptamer provides a unique tool with which to identify such effects.

Replacing the T7 promoter with the T5 promoter, which utilises the host RNA polymerase rather than T7 RNA polymerase (*16*), allowed us to also investigate how different promoters affect expression. An attempt to detect Spinach fluorescence from the well-characterised *E. coli* AraBAD promoter (*17*) did not yield green fluorescence above background (data not shown), but with the T5 promoter, green fluorescence was detectable (Figure 3B). A further comparison of Spinach fluorescence from the T7 and T5 promoters without mRFP1 (pET28c-Spinach and pGPS01) confirmed T5 to be approximately one third the strength of T7 (Figure 3C). Despite the low values of green fluorescence from the T5 promoter, altering the T5-mRFP1 construct with two designed RBS sequences (R38K and R148K) allowed us to identify a case where two constructs composed of different parts (pGPR01: T7-mRFP1-Spinach, and pGPR10: T5-R148K-mRFP1-Spinach) produce equivalent protein output but do so from different transcript levels.

Our work here shows that a construct consisting of mRFP1-Spinach allows for the first time, direct and simultaneous *in vivo* measurement of transcript and protein production from the same mRNA. We have demonstrated this with transcripts expressed by both T7 RNA polymerase and by the native *E. coli* transcription machinery. To aid in more accurate models of gene expression, it has been previously proposed that transcription and translation be measured separately, with transcription rates characterised as Polymerases per second (PoPS) and translation rates characterised as Ribosomes per second (RiPS) (*18*). Measurement of mRNAs containing Spinach brings this proposition closer to reality, providing direct characterisation data for both steps in gene expression.

While it is clear that there are current limitations to the use of the Spinach aptamer as a characterisation tool, future improved designs will likely resolve issues of sensitivity and mis-folding. Interrogation of the RNA folding within Spinach-containing mRNAs using new methods such as SHAPE-Seq (*19*) may shed light on design strategies that ensure that the aptamer correctly folds and is efficiently insulated from local sequences, ribosomes and ribonucleases. The use of Spinach and future variants as measurement tools in synthetic biology offers great potential benefits. The order-of-magnitude faster turnover of RNA compared to protein in cells makes sensitivity more challenging but offers more accurate temporal measurements and a new tool to investigate mRNA degradation. As we demonstrate here, mRFP1-Spinach constructs can be used to evaluate designed RBS sequences and Spinach alone can also be used to measure promoter strength without requiring translation, which itself imposes a cost to the cell (*15*). Perhaps the most interesting insights to be gained using Spinach will come from simultaneously inspecting the individual effects on transcription and translation caused by changes in cell state, such as growth phase, or by design features other than the promoter and RBS that can lead to changes in gene expression, such as codon usage.

Methods

pET28c-Spinach plasmid was a kind gift of Samie Jaffrey (Cornell University) and contains the Spinach aptamer inside a tRNA scaffold (*20*). All other plasmid constructs in this study were derived from this. The mRFP1 gene for the pGPR plasmids was amplified by PCR from part BBa_E1010 from the Registry of Standard Biological Parts (http://partsregistry.org), with the forward primer encoding an RBS sequence (RBS1). In all constructs random, synthetic 30 nucleotide spacer sequences were added between the promoter, the RBS-mRFP1 region and the scaffold-held Spinach aptamer in order to insulate parts from local folding and ribosome binding. The R1K, R9K, R38K and R148K RBS sequences were designed using the Salis Lab RBS Calculator (*11, 12*) and have predicted strengths with mRFP1 of 938, 9166, 38394 and 148124 A.U., respectively. All constructs were assembled scarlessly *in vitro* from PCR-amplified parts using Gibson Assembly (*21*) before transformation into BL21(DE3) *E. coli* cells. Details of all constructs and sequences of all parts are provided in the supporting information.

For the Spinach and mRFP1 measurements, cells were grown in shaking liquid culture (LB+kanamycin media supplemented with 1% glucose) at 30°C overnight. Following this, 150 µl of culture was diluted in 3 ml of LB+kanamycin media and grown at 37°C until at OD_{600}= 0.4. IPTG was then added to 1 mM final concentration. Incubation with shaking was continued at 37°C for 2 hours (except for the time course). To measure fluorescence, 10 µl of culture was diluted into 100 µl LB and DFHBI (Lucerna Technologies) was added to 200 µM final concentration. Samples were incubated at 37°C with shaking for 10 min, placed on ice for 10 min and then analysed.

Flow cytometry analysis used a modified two-colour Becton Dickinson FACScan flow cytometer measuring 2 µl of each sample in 1 ml of water for 45 seconds (approx. 80000 cells). A 488 nm laser was used for excitation of green fluorescence detecting through a 530 nm band pass filter (FL1) with gain 890. Red fluorescence was excited with a 561 nm laser and 610 nm filter (FL5) with gain 850. Data were analysed using Flow-Jo (Tree Star), gating samples for forward scatter and side scatter and excluding FL1 and FL5 values below 2×10^0 A.U.

Fluorescent images were taken through a 60x CPI60 objective mounted on a Nikon Eclipse Ti inverted microscope, with live cells imaged on a 1% M9-agar pad slide. Excitation, emission filters and exposures were respectively 480 nm, 535 nm, 4000 ms for the GFP channel (Spinach) and 532 nm, 590 nm, 1000 ms for Cy3 channel (mRFP1). NIS-Elements Microscope Imaging Software (Nikon) was used for capture and ImageJ (NIH) was used for image presentation.

Supporting Information Available: Table S1 (Constructs), Table S2 (RBS Sequences), Table S3 (Spacer Sequences), Figure S1 (Time Series Peak Values). This material is available free of charge via the Internet at http://pubs.acs.org.

Abbreviations. DFHBI: 5-difluoro-4-hydroxybenzylidene imidazolinone, IPTG: Isopropyl β-D-1-thiogalactopyranoside, RBS: Ribosome Binding Site, PCR: Polymerase Chain Reaction, LB: Luria-Bertani medium. A.U.: Arbitrary Units

Author Information. Department of Bioengineering, Imperial College London, SW7 2AZ, UK

Acknowledgments. The authors wish to thank Samie Jaffrey (Cornell University) and Fabio Chizzolini and Sheref Mansy (University of Trento) for DNA parts and correspondence. We also thank Rhys Algar, Felix Jonas, Karen Polizzi and all members of CSYNBI at Imperial College London. Work at CSYNBI is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) and BR is additionally co-funded by TMO Renewables Ltd.

References

1. Purnick, P. E., and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems, *Nat Rev Mol Cell Biol 10*, 410-422.
2. Ellis, T., Wang, X., and Collins, J. J. (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions, *Nat Biotechnol 27*, 465-471.
3. Oyarzun, D. A., and Stan, G. B. (2013) Synthetic gene circuits for metabolic control: design trade-offs and constraints, *J R Soc Interface 10 (78),* 20120671-20120675

4. Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases, *Nucleic Acids Res 40*, 8773-8781.
5. Canton, B., Labno, A., and Endy, D. (2008) Refinement and standardization of synthetic biological parts and devices, *Nat Biotechnol 26*, 787-793.
6. Paige, J. S., Wu, K. Y., and Jaffrey, S. R. (2011) RNA mimics of green fluorescent protein, *Science 333*, 642-646.
7. Campbell, R. E., Tour, O., Palmer, A. E., Steinbach, P. A., Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (2002) A monomeric red fluorescent protein, *Proc Natl Acad Sci U S A 99*, 7877-7882.
8. Paige, J. S., Nguyen-Duc, T., Song, W., and Jaffrey, S. R. (2012) Fluorescence imaging of cellular metabolites with RNA, *Science 335*, 1194.
9. Vethanayagam, J. G., and Flower, A. M. (2005) Decreased gene expression from T7 promoters may be due to impaired production of active T7 RNA polymerase, *Microb Cell Fact 4*, 3.
10. Takyar, S., Hickerson, R. P., and Noller, H. F. (2005) mRNA helicase activity of the ribosome, *Cell 120*, 49-58.
11. Salis, H. M. (2011) The ribosome binding site calculator, *Methods Enzymol 498*, 19-42.
12. Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression, *Nat Biotechnol 27*, 946-950.
13. Wegrzyn, G. (1999) Replication of plasmids during bacterial response to amino acid starvation, *Plasmid 41*, 1-16.
14. Yamagishi, M., Cole, J. R., Nomura, M., Studier, F. W., and Dunn, J. J. (1987) Stringent control in Escherichia coli applies also to transcription by T7 RNA polymerase, *J Biol Chem 262*, 3940-3943.
15. Shachrai, I., Zaslaver, A., Alon, U., and Dekel, E. (2010) Cost of unneeded proteins in E. coli is reduced after several generations in exponential growth, *Mol Cell 38*, 758-767.
16. Gentz, R., and Bujard, H. (1985) Promoters recognized by Escherichia coli RNA polymerase selected by function: highly efficient promoters from bacteriophage T5, *J Bacteriol 164*, 70-77.
17. Schleif, R. (2010) AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action, *FEMS Microbiol Rev 34*, 779-796.
18. Marchisio, M. A., and Stelling, J. (2008) Computational design of synthetic gene circuits with composable parts, *Bioinformatics 24*, 1903-1910.
19. Lucks, J. B., Mortimer, S. A., Trapnell, C., Luo, S., Aviran, S., Schroth, G. P., Pachter, L., Doudna, J. A., and Arkin, A. P. (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq), *Proc Natl Acad Sci U S A 108*, 11063-11068.
20. Ponchon, L., and Dardel, F. (2007) Recombinant RNA technology: the tRNA scaffold, *Nat Methods 4*, 571-576.
21. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nat Methods 6*, 343-345.

FIGURE LEGENDS

Figure 1: Parallel measurement of Spinach and mRFP1-Spinach expression by fluorescence microscopy and flow cytometry. A) Live-cell fluorescence (GFP/Green and Cy3/Red) and Phase images of BL21(DE3) *E. coli* cells carrying either the pET28c-Spinach (T7-Spinach) or pGPR01 (T7-mRFP1-Spinach) plasmid in the presence or absence of 200 μM DFHBI. Images are taken after a 2 hour induction with IPTG. White arrows highlight OFF-population cells not showing any fluorescence. **B)** Flow cytometry analysis showing FL1/FL5 area plots for cells from the same samples as those imaged by microscopy. OFF-populations of cells showing no fluorescence are seen in the lower left corner. Vertical and horizontal lines show exclusion of FL1 and FL5 values below $2x10^0$ A.U. **C)** Diagram of gene expression control for pGPR01. T7 RNA pol (polymerase) is chromosomally encoded in BL21(DE3) cells and is inhibited in the presence of glucose and absence of IPTG. **D)** Individual FL1 (green fluorescence) and FL5 (red fluorescence) histograms for the gated data shown in B. Background fluorescence for each is indicated by a line plotted through the peaks of negative control samples (i.e. pET28c-Spinach with no DFHBI). The fluorescence peak (modal value) is given for each histogram.

Figure 2: Temporal characteristics and optimal placement of the Spinach aptamer within a synthetic mRNA. A) Green fluorescence (FL1) and red fluorescence (FL5) histograms of *E. coli* from flow cytometry are shown with peak values as in Figure 1D. Histograms show sampling over a 3 hour induction period from cells expressing from the pET28c-Spinach and pGPR01 plasmids. In cases of bimodality (multiple peaks), the peak value given is that for the ON population. **B)** Histograms of *E. coli* cells hosting either pET28c-Spinach, pGPR01 or variations of the pGPR01 plasmid where the RBS1-mRFP1 (red ball and arrow) and Spinach (cross) parts have been rearranged in constructs expressed from the T7 promoter (blue right-angled arrow). Presence and absence of DFHBI are respectively noted by (+) and (-) and by green and grey crosses in the construct diagrams. Shaded histograms indicate those present as positive and negative controls.

Figure 3: Characterisation of designed ribosome binding site sequences and promoters. A) Green fluorescence (FL1) and red fluorescence (FL5) histograms of *E. coli* expressing mRNA from versions of pGPR01 with designed RBS sequences (R1K, R9K, R148K) replacing RBS1. **B)** Histograms for a version of pGPR01 where the T5 promoter replaces the T7 promoter (pGPR08) and for when RBS1 is replaced with designed RBS sequences (R38K, R148K) for this case. **C)** Histograms comparing T7 promoter-driven Spinach expression (pET28c-Spinach) with T5 promoter-driven expression (pGPS01). Construct diagrams depict the T7 and T5 promoters as blue and orange right-angled arrows, respectively, mRFP1 as a red arrow and Spinach as a cross. RBS1 is a red ball and designed RBS sequences are shaded blue balls. Presence and absence of DFHBI are respectively noted by (+) and (-) and by green and grey crosses in the construct diagrams. Shaded histograms indicate those present as positive and negative controls.