Mobius Assembly for Plant Systems uncovers combinatorial interactions among promoters, coding sequences, and terminators in gene regulation

Elif Gediz Kocaoglan[1,2,*], Andreas Andreou[3,4,*], Jessica Nirkko[3,4], Marisol Ochoa-Villarreal[3,4], Gary Loake[3], and Naomi Nakayama[1-4]

Affiliations:

1. Department of Bioengineering, Imperial College London, South Kensington, London, UK
2. Imperial College Centre for Synthetic Biology, Imperial College London, London, UK
3. School of Biological Sciences, University of Edinburgh, King's Buildings, Edinburgh, UK
4. Synthsys Centre for Synthetic and Systems Biology, University of Edinburgh, King's Buildings, Edinburgh, UK

*: These authors contributed equally.

Corresponding author: Naomi Nakayama – n.nakayama@imperial.ac.uk.

ABSTRACT

Plants are the primary biological platforms for producing food, energy, and materials in agriculture; however, they are largely untouched by synthetic biology-driven transformation in bioproduction technologies. Molecular tools for complex, multigene engineering are as yet limited, with development underway to enhance stability and predictivity. Here, we present a new standardized and streamlined toolkit for plant synthetic biology, Mobius Assembly for Plant Systems (MAPS). It is based on small plant binary vectors (pMAPs) that contain a fusion origin of replication that enhances plasmid yield in both *E. coli* and *Agrobacterium*. MAPS includes a new library of promoters and terminators with different activity levels; the parts were made small in size to improve construct stability and transformation efficiency. These promoters and terminators were characterized using a high-throughput protoplast expression assay. Our findings show a significant influence of terminators on gene expression, as the strength of a promoter can change more than 7 folds with the different terminators. Additionally, we have observed that changing the coding sequence changes the relative strength of promoter and terminator pairs, uncovering combinatorial interactions among all parts of a transcriptional unit. We further gained insights into the mechanisms of such interactions by analyzing RNA folding. These results contribute to improving stability, predictability, and orthogonality in synthetic biology of plant systems and beyond.

KEYWORDS

Plant synthetic biology, context dependency, gene regulation, promoters, terminators, protoplast, RNA folding

INTRODUCTION

Since the dawn of civilization, humans have produced food, energy, materials, and drugs by growing plants. Yet, in synthetic and engineering biology applications, plants have remained a minor chassis for bioproduction with far untapped potential. Consequently, the molecular tools and technologies for rational engineering of plant systems are still limited. With the increasing world population and imminent climate crisis, deeply considered, design-led genetic engineering holds the key to a sustainable future, creating transformative solutions for a wide range of applications from agriculture to environmental regeneration[1–4]. One example is the transfer of thirteen genes into rice in a single construct for enhancing anthocyanin production[5]. Anthocyanins are known for their antioxidant properties, and improving the nutritional value of rice by enhancing anthocyanin levels demonstrates the concept of biofortification. From enhancing crop yields, quality, and resilience to improving carbon fixation and biofuel efficiency, synthetic biology-based multigene engineering is a powerful catalyst for futureproofing our society and planet[6–8].

Synthetic biology is thought to confer the precision and predictability necessary for re-designing metabolic biosynthesis, biological developmental pathways, or plant responses to environmental signals. It achieves this through the application of engineering principles: standardization, modularity, and design-build-test cycles[9]. However, as the field matures past its first two decades, having introduced numerous molecular tools and technologies, biology continues to unravel new surprises. The fundamental assumptions, such as orthogonality or context-independent universal functions of genetic parts, are routinely challenged, uncovering new insights and understanding of how biological systems work. This is even the case for the regulation of a single gene, let alone many genes necessary for network or pathway engineering.

Developing synthetic biology tools and technologies for multigene engineering starts with creating a collection (library) of standardized parts. Standardized/standard parts (Phytobricks)[10] can be promoters, coding sequences, terminators, and functional protein tags, for example. These parts together make up a gene or transcriptional unit (TU). Promoters are DNA sequences near the transcription start site (TSS) where RNA polymerase binds, and they are well known for their role in spatiotemporal gene transcription regulation. The coding sequence is the nucleotide sequence of a gene that is transcribed and translated into a protein. Terminators are genetic elements, usually located at the end of a gene or operon, and they terminate transcription. Although often underestimated, the terminator has been shown to exert regulatory influence by controlling transcription arrest, mRNA stability, protecting genes against silencing and tuning other transcription functions[11].

The most commonly used promoters and terminators in plant systems come from the cauliflower mosaic virus (*35S*), Agrobacterium opine genes (*NOS, MAS,* and *OCS*), and more recently plant genomes (e.g., *UBQ10* promoter and *HSP* terminator from Arabidopsis). The shortage of well-characterized regulatory parts results in repeated use of the same sequences when making a multigene construct. This can reduce plasmid stability[12] or even raise the frequencies of homologous-dependent gene silencing[13] due to the highly similar sequences present nearby. Therefore, more promoter and terminator parts will enrich the plant genetic engineering toolkit available in the community and increase our capacity to create genetic modifications. Besides, there is a need for shorter standard parts, especially for single vector multigene delivery, since it will substantially reduce the already large size of multigene constructs. Large constructs tend to make plasmids structurally unstable[14] and decrease the efficiency of agrobacterium-mediated transformation[15].

Once standard genetic parts are built, the next step is the characterization of their activity level.

This involves combining these parts and quantifying the output reporter gene expression. While promoters are generally more well characterized, recently, attention has been also directed towards terminators; diverse arrays of terminators have been characterized in bacteria[16], yeast[17], and moss[18]. However, for plant systems the number of available terminator parts was limited for a long time, and their characterization was performed with the *Nicotiana benthamiana* leaf infiltration system[19,20]. It is worth noting that different species, or even different expression systems within the same species, may exhibit varying levels of expression for the same TU[19,20]. Therefore, in order to achieve higher stability and predictability in multigene engineering, new standard parts should be characterized in different combinations and chassis environments to account for potential context dependency.

As the number of constructs to build and characterize increases, fast and easy DNA assembly methods become even more important molecular construction, which is the foundation of synthetic biology engineering. We developed Mobius Assembly, a user-friendly and streamlined DNA assembly framework, which combines high cloning capacity with a compact vector toolkit[21,22]. Mobius Assembly is a variant of Golden Gate technology and employs type IIS restriction endonucleases, which cleave DNA outside their recognition sequence allowing the unidirectional fusion of DNA sequences in a one-tube reaction[23,24]. The popularity of the type IIS enzyme-mediated DNA assembly has accelerated the development of new variants and improvements. It is now possible to predict the fidelity of three or four base overhang ligations[25–27]. New methods have been developed to deal with domestication requirements[28], scars at coding sequence borders[29], the polarity of the TU growth[30], backbone modularity, and interoperability of DNA assembly[31,32]. The Golden Gate technology has been adapted for organisms such as cyanobacteria[33,34], fungi[35,36], and Chlamydomonas[34]. Furthermore, the philosophy of 'universality' in DNA assembly has gained popularity[37,38]. Recent developments in the plant field include Loop Assembly[39], an extension of the MoClo toolkit for plants[40] and Joint Modular Cloning[19].

In plant transgenic engineering, binary vectors are used to house transgenes and mediate their delivery and incorporation into the genome of the recipient cells. Although functional, plant binary vectors could be improved further to enhance plant transformation efficiency and vector stability. A typical plant vector for *Agrobacterium*-mediated transformation contains left and right T-DNA borders[41] for the transfer of the DNA construct into the host plant's nuclear DNA genome; plasmid replication functions in *Eschericia coli* and *Agrobacterium*; and markers for the selection and maintenance of the vector in *E. coli and Agrobacterium*. Plasmid instability issues are among the main challenges of recombinant DNA technology, affecting plasmid yield and quality[42]. For plant binary vectors, the challenge is doubled since they should be stable in two different chassis, *E. coli* and *Agrobacterium*. The plasmid size[14], copy number[43], direct repeats[12] and inverted repeats[44], for example, affect structural plasmid stability. Traditionally plant vectors are based on bulky backbones (>6 kb), burdening the size of the final constructs by default. Efforts have been made to generate binary backbones with a smaller size to reduce plasmid instability, such as pGreen (2.5 kb[32]), pLSU (4.6 kb[45]), and pLX (3.3 kb[46]). The pLX vectors transformed binary vector construction as they shifted from the cut-and-paste strategy to a modular design with minimal functional parts and the introduction of stability features.

Here we report the adaptation of Mobius Assembly for synthetic biology engineering of plant systems. Mobius Assembly for Plant Systems (MAPS) is based on Golden Gate DNA assembly technology and allows fast and easy molecular cloning of complex, multigene constructs. The MAPS toolkit includes new, compact plant binary vectors pMAPS that should be compatible with a wide variety of plant cells, tissues, and whole organisms. They are compact in size (3.6 kb), structurally stable and suitable for transformation methods that demand a high amount of plasmid

DNA. MAPS is also equipped with a collection of bioluminescence and fluorescent protein reporters and a selection of new standardized parts of promoters and terminators to drive a gradient of gene expression strength. They are designed small in size to reduce the overall construct size.

The new standard parts were tested in a high-throughput transient expression with *Arabidopsis thaliana* (hereafter Arabidopsis) protoplasts, showing the MAPS promoters can drive a gradient of gene activation levels. We demonstrate that the terminators are also effective in controlling gene expression levels. The strength of promoter and terminator parts was not additive; rather, their combinations determined the gene expression strength, and such interactions extended to the coding sequence selection. We further investigated the mechanism of the combinatorial gene regulation to provide insights into designing context-independent molecular constructs and circuits with more predictive outputs.

MATERIALS AND METHODS

Bacterial strains and growth conditions

E. coli strains DH5α, DH10B, TOP10, JM109 and NEB stable, were used. *E. coli* chemically competent cells were prepared in-house using the TSS preparation described by Chung & Miller[47], except TOP10, which was bought from ThermoFisher. *E. coli* cells were incubated at 37°C (30°C for NEB stable), 200 min^{-1} either in 5 ml (for a high copy) or 10 ml (for a low copy), or 100 ml (for midi prep) in LB growth medium supplemented with antibiotics. Cells bearing the *LhGR-pOp6* and *sXVE-lexA* inducible systems were grown for 24 h instead due to their slower growth rate.

Bacterial transformations and plasmid isolation

For *E. coli* transformation, 5 µl of the plasmid DNA was incubated with 100 µl of the competent cells (25 µl for TOP10) on ice for 30 min, followed by a heat shock at 42°C for 90 sec (30 sec for TOP10) and re-cooled on ice for 5 min. S.O.C medium (400 µl) was added, and after 1 h incubation at 37°C, 100 µl of the cell suspension was plated on LB agar plates with antibiotic selection. Plasmids were isolated using Monarch (NEB) or PureYield (Promega) or GeneJET (ThermoFisher) Plasmid Miniprep Kits. The Promega PureYield Plasmid Midiprep System (Promega) with vacuum manifold and GeneJET Plasmid Midiprep Kit (ThermoFisher) were used for higher yields.

PCR amplification

All PCR amplifications for plasmid construction and cloning were performed using Q5® High-Fidelity DNA Polymerase (NEB), followed by purification with Monarch® PCR & DNA CleanupKit (NEB). Successful DNA assembly was verified first by colony PCR using GoTaq® Green Master Mix (Promega) and then with double restriction digestion with *EcoRI*-HF (NEB) and *PstI*-HF (NEB). The constructs were further verified by Sanger sequencing (GATC Biotech-Eurofins, Edinburgh Genomics and Source BioScience) or by whole-plasmid sequencing (Full Circle Labs).

MAPS vector construction

MAPS Level 1 and Level 2 Acceptor Vectors were built using Gibson Assembly. The Mobius

Assembly cassettes were amplified from the corresponding *E. coli* plasmids in the Mobius Assembly Vector toolkit[20] and fused to the plant binary vectors. The pGreen-based Level 1 vectors were constructed using pGreen0029[48]. The NptI gene was replaced with the spectinomycin resistance gene amplified from the pCR8 vector (ThermoFisher) to generate Level 2 vectors. The pLX-based Level 1 and Level 2 vectors were built using pLX-B3 and pLX-B2[46], respectively.

WKS1 Ori was synthesized (Twist Bioscience) into two parts due to repetitivesequences, and pUC Ori was amplified from pUC19, both flanked by *BsaI* recognition sites. The minimum sequence requirement of pUC for stable replication in *E. coli* was found to includeRNA I/RNA II transcripts on the 5'-side and dnaA/dnaA' boxes on the 3'-side, while co-directional transcription of two different replicons in the same plasmid was shown to increase transformation efficiency and DNA yield[45]. A pLX Level 1 A vector with the construct *NOS:BglR:NOS-UBQ10:nluc:HSP* was amplified outside the BBR1 Ori with primers harbouring *BsaI* recognition sites and fused with pUC and WKS1 Ori. The resulting plasmid was used as the template to amplify the pUC-WKS1 fused Ori, which then replaced BBR1 Ori from the pLX-based Level 1A and Level 2A vectors with Gibson Assembly, resulting in the pMAP Level 1A and Level 2A vectors. The rest of the pMAP vectors were constructed again using isothermal assembly and asa template for the Mobius cloning cassettes, the plasmids from the original Mobius Assemblykit[21].

MethylAble modules were devised with the Gibson Assembly using mUAV as a template. Overlapping primers bearing two outward-facing *BsaI* sites prone to CpG methylation and suitable standard overhangs were used to amplify mUAV into two parts. The resulting parts were purified and digested with *DpnI* (ThermoFisher) to eliminate the template DNA, and they were fused in an isothermal reaction.

Mobius Assembly and standard part library construction

A detailed protocol on Mobius Assembly can be found in[22]. Briefly, the assembly was performed in a one-tube reaction with a total volume of 10 µl, with ~50 ng Acceptor Vectors and double amounts of inserts. Reagents added were 1 µl of 1 mg/ml BSA (diluted from 20 mg/ml - NEB), 1 µl T4 DNA ligase buffer (ThermoFisher/NEB), 0.5 µl *AarI* (ThermoFisher) and 0.2 µl 50x oligos of the *AarI* recognition site for Level 0 and Level 2 cloning or *Eco31I/BsaI-HFv2* (ThermoFisher/NEB) for Level 1 cloning, and 0.5 µl T4 DNA ligase (ThermoFisher/NEB). The reactions were incubated in a thermocycler for 5-10 cycles of 5 min at 37°C and 10 min at 16°C, followed by 5 min digestion at 37°C and 5 min deactivation at 80°C (5 cycles for Level 0 and the first round of Level 1 cloning – 10 cycles for Level 2 and large constructs >10 kb).

For the MethylAble modules, plasmid DNA was extracted by midi-prep. The reaction was carried out in 20 µl total volume according to the NEB protocol. Briefly, up to 2 µg of plasmid DNA was incubated for 4 h at 37°C with 2 µl of CpG Methyltransferase from *M.SssI* (NEB) in the Methyltransferase Reaction Buffer supplemented with 2 µl of diluted SAM (6.4 mM). The reaction was stopped by heating to 65°C for 20 min and was purified by column chromatography.

Promoter and terminator part design and generation

Likely constitutive promoters and terminators were selected from the literature, especially the genes commonly used as positive controls in qRT-PCR[49]. Their sequences were retrieved from the TAIR webpage[50] (http://www.arabidopsis.org/index.jsp) and blasted in NCBI to find the untranslated regions flanking the genes. A 1.5 kb sequence upstream of the start codon was run through the online prediction software and TSSPlant[51] (http://www.softberry.com) to identify TATA and TATA-less promoters or enhancer sites. In the promoter selection, it was also considered, when possible, to include the initiator (INR) elements (YYA(+1)NWYY- TYA(+1)YYN-

TYA(+1)GGG)) and downstream promoter (DPE) element (RGWYV). Potential promoter elements linked to increased gene expression were identified using PlantCare[52] and PLACE[53] software. Terminators were selected with PASPA, a web server for poly(A) site prediction in plantsand algae[54] (http://bmi.xmu.edu.cn/paspa/interface/run_PASPA.php). A sequence 300bp downstream of the stop codon was input into PASPA, and the end of the terminator was setat 10bp after the second polyadenylation site, resulting in a sequence of around 200bp. They were then analysed in RegRNA2.0 to identify other RNA functional motifs[55] (http://regrna2.mbc.nctu.edu.tw).

The parts designed in the first round were short promoters (~300bp) and short terminators (~200bp) from the genes *ACT2*, *FAD2*, *TUB9*, *APT1*, *NDUFA8* and *LEC2*. As the ~300 bp promoters werevery low in activity (Data not shown), we designed a new set of promoters derived from the genes *TUB2*, *UBQ11*, *UBQ4*, *ACT7*, and the longer versions of the previous promoters (~500 kb). Appropriate primers compiling to the Phytobrick standard overhangs were designed for both promoters and terminators for cloning into mUAV.

Protoplast isolation and transformation

Arabidopsis thaliana (Wildtype, Col-0; hereafter Arabidopsis) seeds were sown on the soil. After a 2-day pre-treatmentat 4°C in darkness, they were sawn and grown under long-day conditions (21°C; 16h light / 8h dark cycles; light intensity ~100µmol/m^2s^{-1}; humidity 40-65%) until harvest.

The protoplast isolation and transformation protocol was developed based on Chupeau *et al.*, (2013)[56] and Faraco *et al.*, (2011)[57]. Briefly, leaves from 6–8-week-old plants were treated with MGG cell wall-digestion solution containing Cellulase ONOZUKA R-10 (0.2%), MACEROZYME R-10 (0.06%) (Yakult Pharmaceutical), and Driselase (0.08%) (Sigma-Aldrich) for up to 12 hours. Then, the protoplasts were filtered and washed 3-5 times with MGG without enzymes and resuspended in MMM solution (0.4 M mannitol, 15 mM $MgCl_2$, 0.1% w/v MES, pH 8) to the concentration of $5x10^5$ cells/ml. Protoplasts were transformed into 1 ml 96-well plates. 8 µl DNA (500 ng/µl) was added to 75 µl of the protoplast suspension, followed by the addition of 83 µl PEG (0.4 M mannitol, 0.1 M$Ca(NO3)_2$-4H2O, 40% PEG4000, pH8) and 1 min incubation. Subsequently, the tubes were filled with MGG solution and incubated for 1 h at RT. After the incubation, the supernatant was removed, and the protoplasts were resuspended in 100 µl of fresh MGG and incubated in darkness at RT. For the inducible system, 0.1% EtOH, 2.5 µM DEX (Acros) or 5 µM β-estradiol (LKT laboratories) were added at this point. Gravitational sedimentation is used in all the steps described above, and centrifuge was avoided to prevent damages to the protoplasts.

Luciferase assay

The 96-well plates containing the transformed protoplasts were let to sediment and 60 µl of supernatant was discarded. The protoplasts were resuspended, and 40 µl was transferred to white optical plates in a grid pattern with empty spaces between wellsto reduce luminescence bleed-through. Luciferase activity was assayed in an Omega luminescence plate-reader (Fluostar) with four different gains following the instructions of the Nano Dual-Luciferase® Reporter kit (Promega). A further correction for luminescence bleed-through and background signal was applied using the software developed by Mauri*et al*.[58] Then *NLuc* signal was divided by the *FLuc* signal to normalize for the transformation efficiency.

Fluorescence assay

For fluorescent protein assays, TECAN Spark plate reader was used with either 96-well plates (Greiner) or black Thermo Scientific Nunc F96 MicroWell. Excitation and emission for the super-folder green fluorescent protein (sfGFP) was 485 nm and 520 nm, respectively. For mScarlet-I, excitation and emission wavelengths were 560nm and 620nm. Measurements were taken at two different gains to ensure signals don't overshoot beyond saturation. The sfGFP values were divided by the mScarlet-I values to normalize for the transformation efficiency. Since fluorescence was not always active, unlike luminescence and also because black-walled plates were used to prevent signal bleed-through, deconvolution correction was not applied to the fluorescent samples.[58]

RNA isolation and real-time PCR (qRT-PCR)

For RNA extraction, approximately 1 x 10^6 protoplasts were used per transformation. Isolation and transformation were performed as described above, with reaction and solution volumes upscaled by 26 times. Transformations were performed in 50-ml conical tubes (Falcon) until after the addition of MGG solution following the PEG incubation. Subsequently, each transformation culture was divided into three 14-ml round bottom culture tubes (Greiner) to help the sedimentation.

24 h after transformation, the supernatant was discarded, and sedimented protoplasts were harvested for RNA extraction with GeneJET Plant RNA Purification Kit (ThermoFisher). Extracted RNA was treated with DNA-free DNA Removal Kit (Invitrogen). Then, reverse transcription was performed with UltraScript 2.0 cDNA Synthesis Kit (PCR Biosystems). This cDNA synthesis kit uses optimum amounts of oligo (dT) and random hexamers for unbiased amplification of RNA variants. It is advised to use 2 µM of both oligo (dT) and random hexamers for cDNA synthesis when using a different kit.

qPCR reaction was set up with primers and probe sequences listed in **Supplementary Table S4**. Sso Advanced Universal Probes Supermix (Bio-Rad) was used following the manufacturer's instructions. with the reactions were prepared in MicroAmp Endura 96-well plate and sealed with MicroAmp Optical Adhesive films, and the optical output was measured with StepOnePlus qPCR machine (all Applied Biosystems). A standard curve was included on the plate for each gene target tested; this standard curve was used to calculate cDNA copy number of the samples on the same plate. The *FLuc* or *mScarlet-I* cDNA copy number was used for normalization for the transformation efficiency. The resulting values from all the constructs were then normalized to the *NLuc* cDNA copy number from the *UBQ10:NLUC:NOS* in the same experiment, to account for any batch-to-batch variations.

RNA folding prediction

For mRNA folding predictions, the transcription start site was predicted for each promoter by using TSSPlant[51]. The nucleotide sequence downstream of the TSS to the end of the terminator was used as the mRNA sequence. mRNA folding was predicted by using RNAFold software, where thermodynamic ensemble prediction was used for folding energy predictions and centroid algorithm was used for secondary structure prediction[59].

RESULTS

Mobius Assembly for Plant Systems (MAPS)

MAPS is an extension of our molecular cloning framework Mobius Assembly[21] to enable transformation and expression of transgenes in plant systems. MAPS has a universal acceptor vector (mUAV) at Level 0 to house a standard part in the Phytobrick format. Up to four Level 0 constructs can be combined, for example to form a TU at Level 1. Level 1 constructs are then combined to make Level 2 constructs. MAPS follows a linear cloning strategy until Level 2 and then iterates between two cloning levels (Level 1 and Level 2) for quadruple augmentation of cloning units each time (**Fig. 1a**). Using the rare cutter *AarI*, as opposed to frequently used restriction enzymes that recognize shorter sequences (e.g., *BsmBI* or *BpiI*) minimizes the need for removing internal restriction sites.

Reflecting on user feedback, two specific improvements were made to improve the Mobius Assembly vectors, including the new MAPS vectors (pMAPs). Few users of the original Mobius Assembly kit indicated that the chromoprotein selection has been lost in their clones. Upon investigation, we observed independent events of transposon insertion in the promoter of the chromoprotein genes, which we hypothesized is a response to the stress imposed by chromoprotein production. As a result, we replaced spisPINK chromogenic protein with mScarlet-I (**Supplementary Fig. 1a**). We also noticed some Mobius Assembly clones didn't grow well in the selection media and identified the instability was caused by plasmid dimerization. Therefore, we introduced a 240-bp cer domain, which acts as the substrate for recombination to help keep the plasmids in the monomeric state[60].

pMAP also has two Left Border sequences instead of one. Agrobacterium gene transfer is from the Right Border to the Left Border, and having two Left Borders suppresses backbone transfer to the plant genome[58]. Another feature of pMAP is that terminators flank LB and RB sequences to isolate transgene activity from the plasmid backbone. The MAPS vector architecture can be seen illustrated in **Fig. 1b.** Colourific markers, antibiotics selections, and restriction enzymes used for Mobius Assembly at different levels are summarized in **Fig. 1c**.

MAPS vector toolkit consists of a core set of pMAP cloning/destination vectors (Level 1 and Level 2 Acceptor Vectors, four variations A-Δ for each level), which have a fusion origin of replication to replicate in *E.coli* and Agrobacteria. The mUAV and the seven Auxiliary plasmids are also included, as described in the original Mobius Assembly kit[21] . The MAPS toolkit also contains a selection of plant promoters, terminators, antibiotic resistance genes, and visible reporter genes (bioluminescence and fluorescent proteins). All MAPS plasmids are listed in **Supplementary Table S1** and available through AddGene (https://www.addgene.org/browse/article/28211394/).

Initially, we developed pGreen-based vectors for the transient transformation of Arabidopsis mesophyll protoplasts. however, we had transformation issues with large constructs, consistent with reported instability issues[61,62] . Therefore, we built a new small plant binary vector based on the pLX architecture, called pMAP (**Fig. 1b**). Based on its backbone design, pMAP is suitable for transient expression in protoplasts, cell culture, tissues/organs (e.g., leaves), and whole-plant stable transformation. A new origin of replication for plant binary vectors was devised by fusing pWKS1 and pUC19 Ori. pWKS1 Ori derives from a small, multicopy and cryptic plasmid pWKS1 of *Paracoccus pantotrophus* DSM 11072, which has only replication and mobility function domains[63]. Cryptic plasmids usually encode proteins involved in plasmid replication and mobilization, and they do not have obvious benefits to the host cells that carry them[64,65]. The replication domain consists of an origin of vegetative replication OriV and the gene encoding its cognate binding protein RepA. pWKS1 Ori was found to be functional in Agrobacteria but not in *E. coli* [63]. Thus, the minimal stable pUC Ori from pUC19 was amplified and fused in co-directional transcriptional

orientation with pWKS1 Ori. The MAPS kit also offers pLX (BBR1 and RK2) based destination vectors (Level 1 and Level 2 AcceptorVector As), which are medium and low copy number vectors, respectively, to deal with possible instabilities that occur in large constructs housing repetitive or similar sequences.

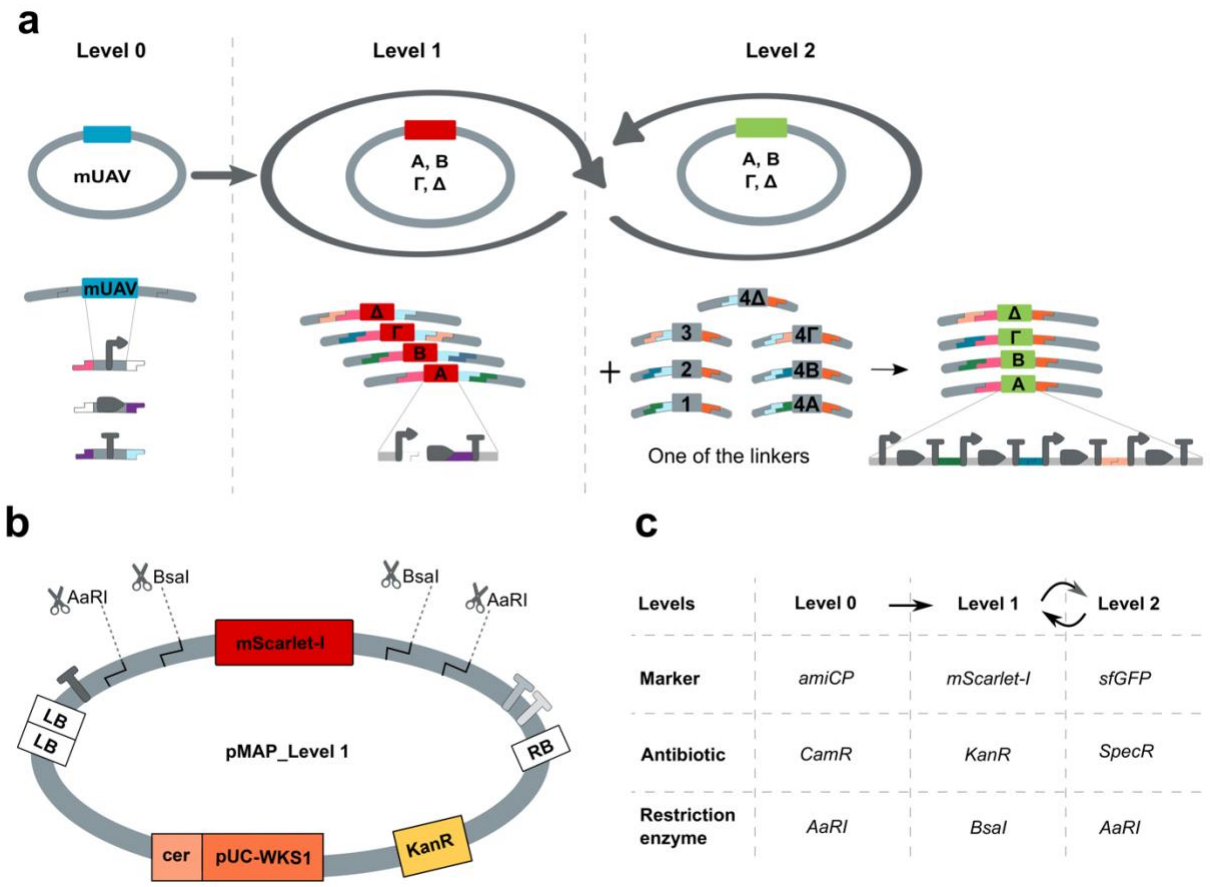


Figure 1. Mobius Assembly for Plant Systems (MAPS) vector structure and workflow.

a Schematics of how Mobius Assembly operates. MAPS is an adaptation of the Mobius Assembly for plant systems based on small binary vectors intended for plant transformation. The core vector toolkit is comprised of one part storage vector in Level 0 (mUAV), four Level 1 Acceptor Vectors, fourLevel 2 Acceptor Vectors, and seven Auxiliary plasmids providing linkers. The vectors release the insert with *BsaI* in Level 1 cloning and the rare cutter *AarI* in Level 0 and Level 2 cloning. **b** Plasmid design of MAPS vectors, showing a Level 1 plasmid as an example. At Level 2, *BsaI* and *AaRI* cut sites swap places. **c.** Summary of the markers, antibiotic resistance and RE enzymes used for Mobius Assembly at different Levels. CamR is a gene conferring the resistance for chloramphenicol, KanR for kanamycin and SpecR for spectinomycin. For simplicity, in the Level 2 construct, overhangs that are carried from Level 0 are not shown.

MethylAble function to aid part-series construct cloning

Combinatorial DNA library is essential for applications such as part characterization and metabolic pathway optimisation[66]. However, the manual generation of a combinatorial DNA construct library takes time, effort, and resources, especially when it involves multiple cloning steps. Therefore, we developed an additional Mobius Assembly feature called 'MethylAble', which allows the introduction of standard part variants (Level 0) to a specific site in Level 1 or 2 constructs (**Fig. 2**). It utilizes the DNA methylation sensitivity of *BsaI* to mask its recognition sites by cytosine methylation during Level 1 cloning. According to Rebase (http://rebase.neb.com), GGTCTC^{m5}/C^{m5}CAGAG methylation prevents *BsaI* digestion.

An *amilCP* expression cassette was designed to carry in each site divergent and convergent *BsaI* recognition sites bordering the four base pair overhangs. The overhangs correspond to the part that the MethylAble replaces. The divergent *BsaI* recognition sites were designed to be prone to CpG methylation (CGGTCTC^{m5}G/GC^{m5}CAGAGC), and consequently, *BsaI* digestion is blocked at these locations, while not at the convergent sites (TGGTCTC^{m5}T/AC^{m5}CAGAGA). Upon CpG methylation, the *amilCP* gene was cloned with intact divergent *BsaI* sites in a Level 1 reaction (**Fig**. **2b**). Therefore, Level 0 parts can directly be inserted in a premade Level 1 TU.. As *amilCP* chromoprotein will be expressed, the correct constructs will have a purple colour until the Level 0 parts replace the *amilCP* expression cassette. The MethylAble feature can be designed for any standard part or combination of parts by setting the appropriate overhangs. It is also possible to use this feature for multi TU constructs, After a Level 2 plasmid is built with a MethylAble part, similar strategy will be used, again with a Level 1 acceptor, making it a Level 3 construct **(Fig.2c).** MethylAble plasmids are built with isothermal assembly using mUAV as a template and overlapping primers to introduce the *BsaI* recognition sites and the selected overhangs.

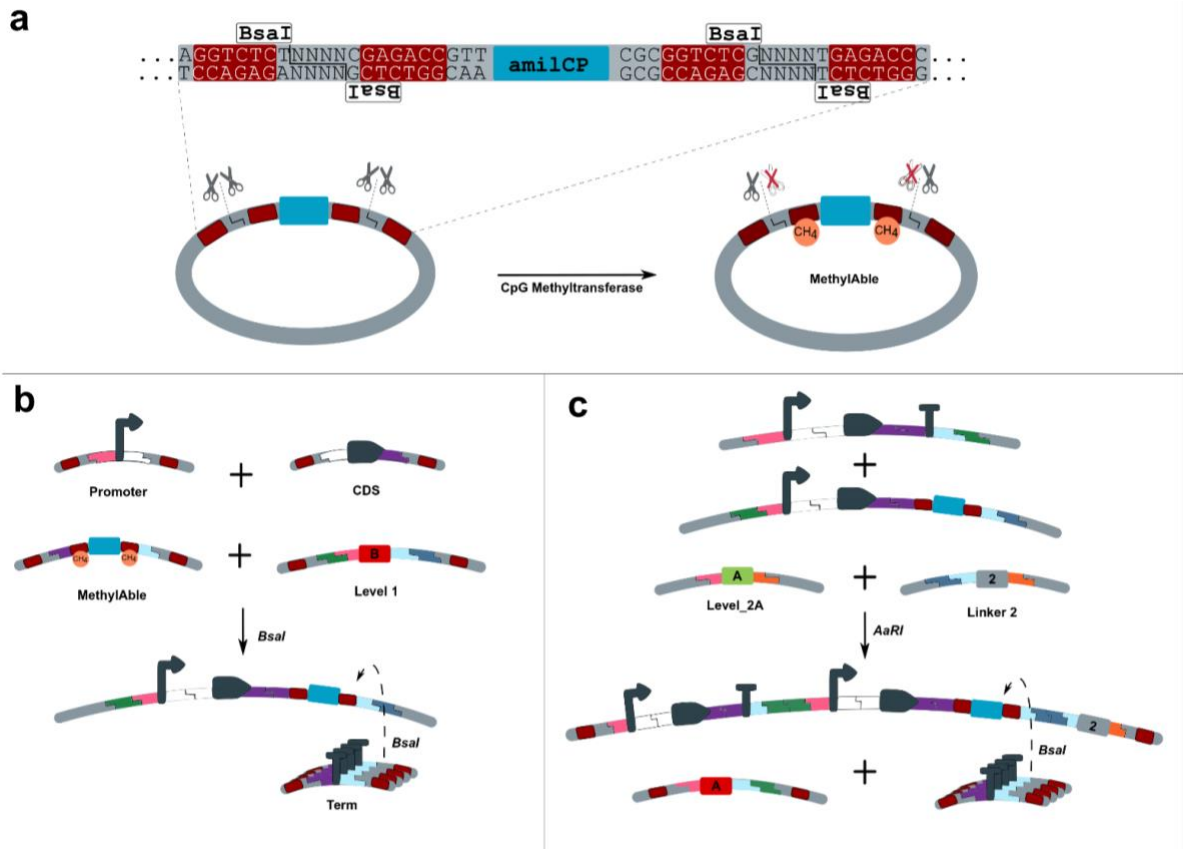


Figure 2. MethylAble allows direct assembly of standard parts at Level 1 or Level 2.

MethyAble is based on partial digestion utilizing the sensitivity of the restriction enzymes to DNA methylation status. **a**. MethylAble plasmid contains an *amilCP* gene flanked by inward and outward-facing *BsaI* recognition sites, which both cleave at the Phytobrick overhangs. The outward-facing *BsaI* sitesare blocked with CpG methylation. **b.** In this example, the MethylAble plasmid is used to create a variation in the terminator part, which is fused with a promoter and a coding sequence in a Level 1 reaction. **c.**Multi-TU constructs can directly receive Phytobricks. A Level 2 construct is made and then digested with *BsaI* to release MethylAble part, along with the multiTU-cassette. Using a Level 1 acceptor vector here demonstrates the MAPS iterative cloning cycle and makes the final construct Level 3. Only *BsaI* restriction sites are shown (*AaRI* sites hidden) for simplicity.

As a proof of concept, the MethylAble protocol was used to build the library of the three inducible promoters, each of which was combined with the 14 terminator parts (see below). The MethylAble cassette was designed to have the terminator overhangs (GCTT-CGCT). In Level 1, the constructs were made as shown in **Supplementary Table S3**. Briefly, a normalizer unit (FLuc) was built in Vector A, and the trans-activator TU in Vector B. The inducible promoter was combined with the MethylAble features in Vector Γ. Without the MethylAble, the inducible promoters should have been combined with all terminators in this step. With MethyAble,in a Level 2 reaction, the normalizer, the trans-activator, and the inducible promoter-containing TU were fused togenerate three constructs in total (three for three different inducible systems). Subsequently, in the second Level 1 reaction, the inducible constructs were combined with Level 0 terminators to give the final 14 constructs per inducible system.

Designing, building, and testing of promoter/terminator standard parts

To select new 'constitutive' promoter and terminator parts, we chose ubiquitously expressed genes that are likely to have strong expression in different tissue types. (**Supplementary Table S4**). The promoters and terminators were characterized via a high-throughput protocol for transient gene expression assay with Arabidopsis mesophyll protoplasts with PEG transformation, which we developed based on[56]. It was adapted to plate reader measurement in 96-well format. The parameters throughout the protocol were optimized to improve the transformation efficiency and reproducibility, simplify the procedure, and reduce the reagents. We took extra care to minimize protoplast stress and did not use any centrifugation.

To evaluate the promoter/terminator activity level, we first opted for bioluminescence for their sensitivity, using the dual luciferase system with nano luciferase (NLuc) and firefly luciferase (FLuc). The highly sensitive NLuc was used as the reporter for the part activity. Protoplast transient assay is prone to varied transformation efficiency from sample to sample; to normalize this variability, we included FLuc gene (*UBQ10:FLuc:UBQ5*) in each construct and calculated the arbitral output of NLuc/FLuc as RLU (Relative Light Unit).

The *NLuc* transcription was driven by 17 different promoters, while being terminated by either the *NDUFA8* or *HSP* terminator (*Promoter:NLuc:HSP/NDUFA8*).*UBQ10* exhibited by far the highest expression activity among the promoters, followed by *MAS* (**Fig. 3a,b**). The *HSP* terminator resulted in elevated gene expression compared to *the NDUFA8 terminator for all the promoters* except *TUB9*. Two of the newly isolated promoters, *UBQ11* and *UBQ4*, were found to have a comparable or even better expression than the commonly used *35S* and *OCS* promoters. Furthermore, the newly isolated promoters *ACT7*, *TUB2*, *TUB9*, *APT1*, *ACT2* and *LEC2* drove stronger expression than the *NOS* promoter. Lastly, the lowest expression was found for *FAD2* and *NDUFA8* promoters.

For terminator evaluation, strong *UBQ10* or weak *NDUFA8* promoter drove the *NLuc* expression, terminated by one of the 14 terminators (*UBQ10/NDUFA8:NLuc:Terminators*). The different terminators resulted in a range of expression spanning 5.3 and 6.3 fold change for *UBQ10* and *NDUFA8*promoters, respectively (**Fig. 3c,d**). For the strong *UBQ10* promoter, *FAD2* terminator resulted in the highest activity with (547.9 RLU) and NOS in the lowest with (103.9 RLU). For the weak *NDUFA8* promoter, the *HSP* terminator led to the highest expression (0.327 RLU) and *APT1* to the lowest (0.052 RLU).

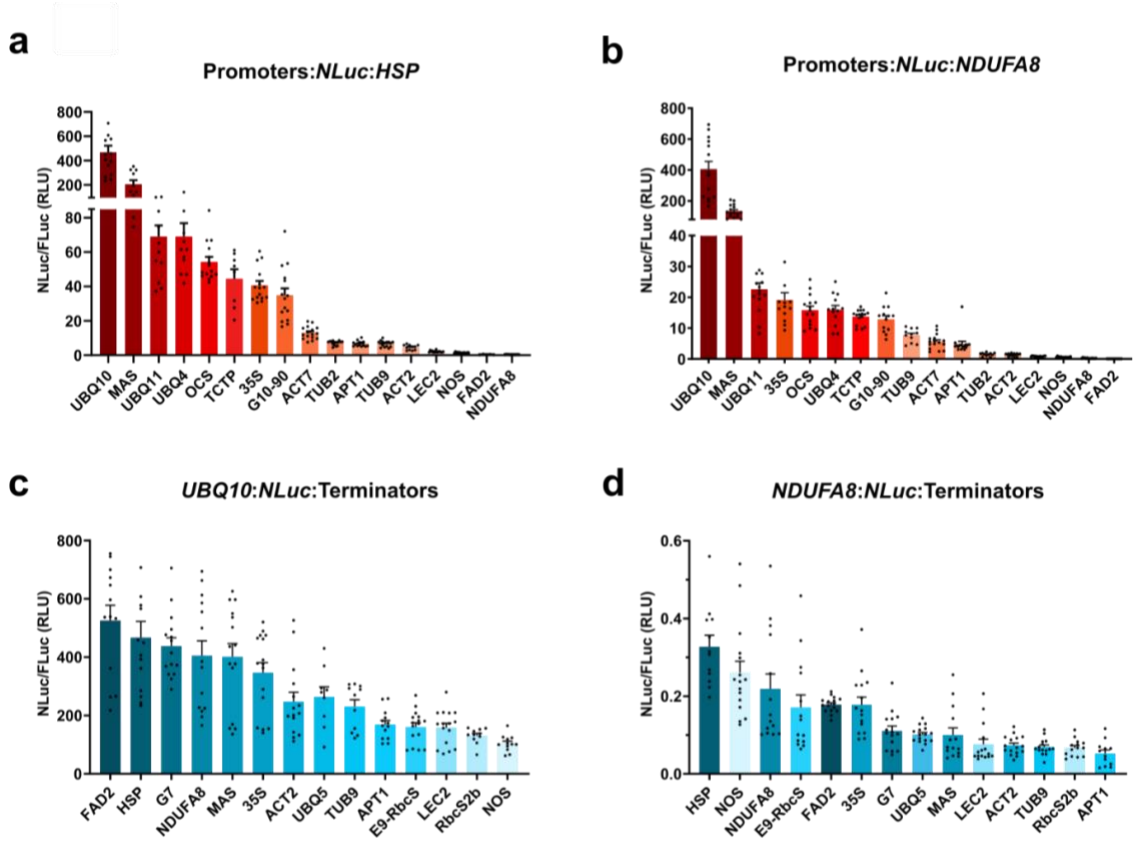


Figure 3. Promoter/terminator part characterization underscores terminator-mediated gene regulation.

MAPS promoter or terminator parts were assessed for their gene expression activity using a transient assay with a dual luciferase system and Arabidopsis leaf mesophyll protoplasts. The promoter library (**a,b**) was characterized either with *HSP* (**a**) or *NDUFA8* (**b**) terminator, with the construct *Promoter:NLuc:HSP/NDUFA8.* The terminator library (**c,d**) was examined either with *UBQ10* (**c**) or *NDUFA8* (**d**) promoter, with the construct. *UBQ10/NDUFA8:NLuc:Terminator.* Each construct also contained *UBQ10:FLuc:UBQ5.* Relative Light Unit (RLU) is the NLuc readout divided by the FLuc readout, normalizing for the variability in the transformation efficiency. The bar graphs show RLU in mean ±SE. The promoter and terminator parts are colour-coded according to the activity strengths in the first characterization series (**a** for promoters; **c** for terminators). The negative control was untransformed protoplasts, whose luciferase values are subtracted from measured sample values during deconvolution, making the negative control values 0. All the data points from three biological replicates are shown on the graph as dots.

Since it was surprising to see a wide range of gene expression levels led by different terminators under the same promoter, we extended terminator characterization with the three chemically inducible systems commonly used in plant sciences: dexamethasone (Dex), estradiol, and ethanol inducible systems[67–69]. They are all based on two-component mechanisms involving at least two transcriptional units, in which the exogenously applied chemical activates the transcription factor that further transactivates the downstream target genes. The target genes are activated by the specific promoters that contain binding sites for the transactivator (*pOp6*, *lexA,* and *alcSynth*), and hence promoters cannot be changed. Terminators, on the other hand, can be swapped.

Interestingly, terminators combined with Dex and estradiol inducible promoters resulted (*pOp6-35S* and *lexA-35S*) in narrow and more uniform *NLuc* expression, spanning at around 2.2 and 2.9 folds, respectively (**Fig. 4a,b**). For the Dex system, RLU was spread between 37.9 and 111.6 when combined with *UBQ5* and *35S* terminators. For the estradiol promoter (*alcSynth-34S*), RLU was between 11.6 and 25.1 with *E9-RbcS* and *35S* terminators. In contrast, the ethanol inducible promoter has shown a much wider range of expression comparable to the constitutive promoters; the expression with *HSP* terminator is 7 times higher than *LEC2* terminator (**Fig. 4c**). Both Dex and estradiol systems showed a basal expression of around 10 RLU (when combined with *HSP* terminator); Dex induced ~11-fold activation, while estradiol had about three fold activation. In contrast, the ethanol system had a high basal expression (i.e., it was leaky), leading to only a 30% increase in luminescence.

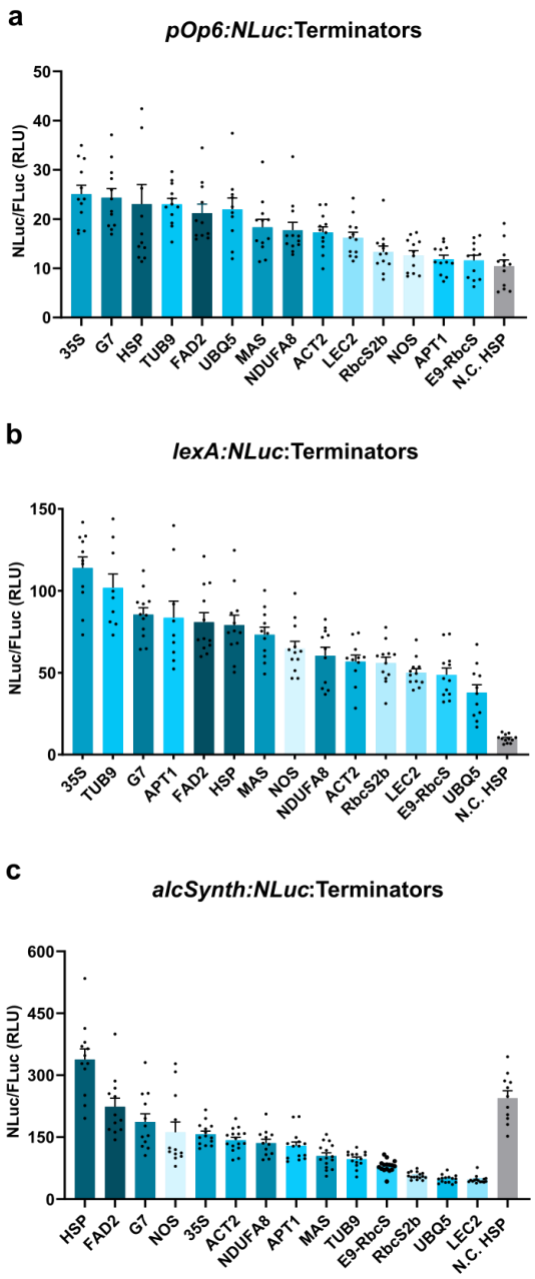


Figure 4. Inducible promoter strength can be controlled by using different terminators.

Fourteen MAPS terminators were characterized in Arabidopsis mesophyll protoplasts as described before, combined with the inducible promoters: (a) Dex (*pOp6-35S*), (b) estradiol (*lexA-35S*), (c) or alcohol (*alcSynth-34S*) inducible promoter. The construct was *UBQ10:fluc:UBQ5-UBQ10:Transactivator:HSP- Promoter:NLuc:HSP*, where 'Transactivator' corresponds to *LhGR* **(a),** *sXVE* **(b)** and *AlcR* **(c).** Chemical induction was achieved with 2.5μM DEX for **(a),** 5μM β-estradiol for **(b),** or 0.1% EtOH for (**c**). Bars show normalized luciferase activity (RLU) in mean ± SE. N.C. = negative control (uninduced with HSP terminator). All the data points from three biological replicates are shown on the graph as dots.

Promoter-coding sequence-terminator interactions in gene regulation

Reflecting on the observed promoter-terminator interactions, we investigated whether changing the coding sequence also influences the promoter/terminator activity. An alternative visible reporter to luciferase systems is fluorescent proteins. However, using fluorescent proteins in plant chassis can be challenging as they emit red and green range autofluorescence from their chloroplasts and blue range autofluorescence from their cytoplasm when stressed[70]. Over the years, fluorescent proteins with improved brightness and expression dynamics have been developed, but their quantitative efficacy hasn't been comprehensively characterized in protoplasts. Therefore, we screened fluorescent proteins from four spectrums (green, red, yellow, and blue) to examine their compatibility with protoplast experimentation with a plate reader or microscopy. Generally, expression of fluorescent proteins was detected 6 hours after transformation and plateaued around 15 hours (**Supplementary Fig. S5**). Informed by this screen, we have chosen the brightest two fluorescent proteins from different spectrums: sfGFP and mScarlet-I. *sfGFP* was the main reporter, while *mScarlet-I* was used as the normalizing gene (similar to *FLuc* above) and expressed under *UBQ10* promoter and *UBQ5* terminator.

Initially promoters were evaluated with *HSP* terminator for *sfGFP* expression. Unlike luciferase reporters, not all the promoters drove strong enough expression that could be detected with a plate reader (**Fig. 5a**). The highest expression was driven by *UBQ10* and *MAS* promoters, followed by *35S* promoter. Readouts from the rest of the promoters could not be distinguished from the background autofluorescence. Subsequently, *UBQ10* and *MAS* promoters were analyzed with all the MAPS terminators **(Fig. 5b,c).** *HSP* terminator was the strongest with both promoters (9.0 RFU with *UBQ10* and 12.1 RFU with *MAS).* The *FAD2* terminator was on the high-expression side for both promoters*,* along with the *UBQ10 promoter-35S terminator* and *MAS promoter-NOS terminator* combinations. *Rbsc2b* terminator gave the lowest expression with both *UBQ10* and *MAS* (promoters. Overall, a 3-fold expression difference was observed by using different terminators with *UBQ10* promoter and the difference was 6-fold with *MAS* promoter.

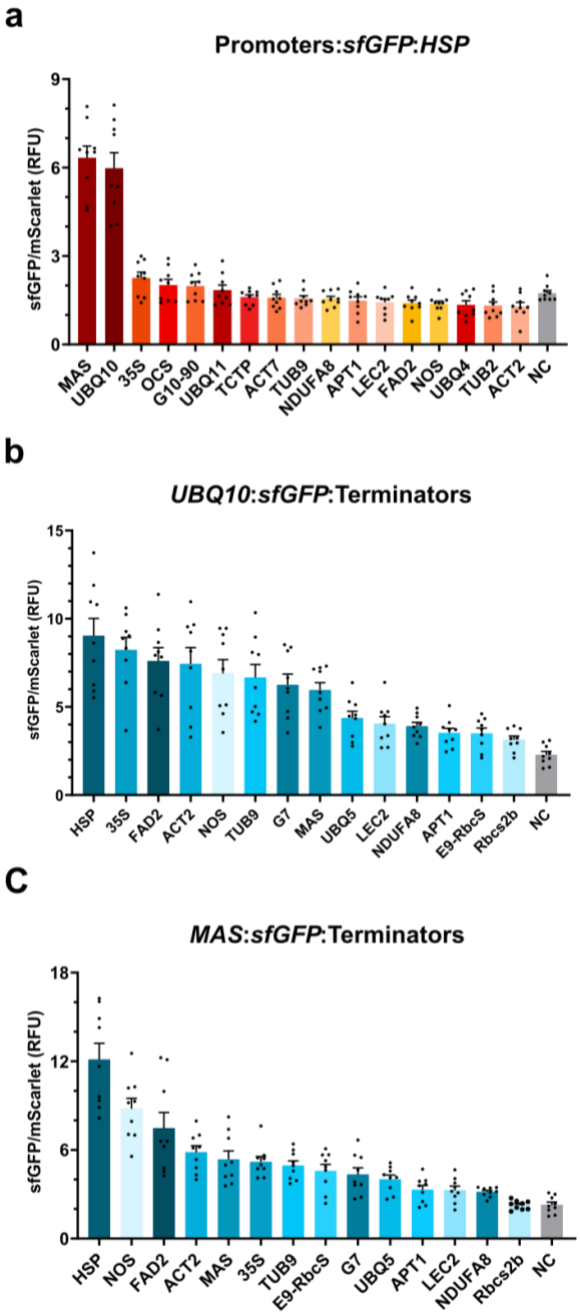


Figure 5. The coding sequence influences the gene expression activity of in combination with promoters and terminators.

MAPS promoter and terminator parts were characterized for sfGFP expression level by PEG-transformed leaf mesophyll protoplast assay. The sfGFP emission signal was divided by the mScarlet-I RFP signal to calculate Relative Fluorescence Unit (RFU). **a** Promoter variation with *HSP* terminator. One-way ANOVA test (Tukey's HSD) shows a significant difference for *MAS* and *UBQ10* promoters from NC ($p < 0.001$) and *35S* promoter from NC ($p < 0.05$)**.** **b** Terminator variation with U*BQ10* promoter. **c** Terminator variation with *MAS* promoter. NC= Negative control (untransformed protoplasts). Error bars represent mean ± SE. All the data points from three biological replicates are shown on the graph as dots.

The part characterization revealed that the gene activation function of a promoter, terminator or coding sequence is not independent or additive, but combinatorial and even synergistic (**Table 1**). A clear example of such all-part interaction is seen with the NOS terminator, which in combination with the *UBQ10* promoter, resulted in weak, *NDUFA8* strong, *lexA* medium, *pOp6* weak and *alcSynth* medium expression in the luciferase system. When we switched to a fluorescence-based reporter, the combination of the *NOS* terminator with *UBQ10* and *MAS* promoters drove medium and strong expression, respectively. Among the 14 terminators characterized, only four exhibited stable behaviours with both coding sequences: *FAD2* and *HSP* were consistently on the strong side, while *LEC2* and *RbcS2b* tended towards the weak side.

		LUCIFERASE					FLUORESCENCE	
	PROMOTERS	***pOp6***	***lexA***	***alcSynth***	***NDUFA8***	***UBQ10***	***UBQ10***	***MAS***
TERMINATORS	***G7***	24.3	85.46	187	0.11	444.8	6.2	4.3
	NOS	12.6	64.7	161.8	0.26	103.9	6.9	8.8
	ACT2	17.3	57.4	142.6	0.073	247	7.4	5.8
	APT1	11.8	83.4	132.2	0.052	171.4	3.5	3.3
	35S	25.1	111.6	157.6	0.177	346.9	8.2	5.2
	LEC2	16.2	50	46.7	0.075	158.2	4	3.3
	TUB9	23	106.4	92.2	0.071	235	6.7	4.9
	MAS	18.3	71.7	104.5	0.098	409.3	5.9	5.4
	E9-RbcS	11.6	48.7	79.7	0.188	160.6	3.5	4.6
	RbcS2b	13.3	55.9	56.7	0.066	132.4	3.1	2.3
	HSP	23	79.1	337.9	0.327	464.2	9	12.1
	UBQ5	20.4	37.9	48.1	0.101	239.6	4.4	4
	NDUFA8	17.7	61.1	135.7	0.216	423.7	3.9	3.2
	FAD2	21.5	80.8	223.4	0.178	547.9	7.6	7.5

Table 1. Combinatorial gene regulation by promoters, coding sequences and terminators

Summary of the reporter expression levels measured for the promoter-conding sequence-terminator combinations tested in the transient expression assay with Arabidopsis protoplasts. The expression level of NLuc normalized by FLuc (RLU) for the luciferase system and sfGFP normalized by mScarlet-I (RFU) for the fluorescent system is shown in different promoter-coding sequence-terminator combinations. The rainbow colour gradient is set independently in each promoter series; The darkest blue dark blue colour shows denotes the lowest expression, light blue mid-level expression, and dark red the highest expression.

Dissecting the mechanism of the combinatorial gene regulation

Next, we wanted to gain insights into how the promoters, coding sequences, and terminators interacted in regulating gene expression. The highly variable and combinatorial relationships observed among the three parts signaled physical interactions at the DNA or RNA level (more in Discussion). Therefore, the interactions may regulate gene expression at the transcriptional or post-transcriptional steps. To distinguish these two possibilities, we performed qPCR to examine how the transcript (mRNA) level correlates with the final protein activity readout. Because transforming enough protoplasts for RNA extraction is laborious, we selected key constructs to test. *HSP* and *FAD2* terminators were chosen for consistently strong expression with all the promoters and coding sequences tested. For consistently weak expression, *NOS* and *Rbcs2b* terminators were selected for *NLuc*, while *Rbcs2b, APT1* and *E9-RbcS* terminators were chosen for *sfGFP*. *NOS* terminator is chosen because their activity shifts between strong and weak expression depending on the partnering promoters or coding sequences; and the *35S* terminator is chosen because it is widely used in plant systems.

The qPCR revealed that the consistently strong *FAD2* and *HSP* terminators have significantly higher mRNA expression (**Fig. 6a**). The weak *E9-Rbcs2b* terminator had significantly lower mRNA than the rest of the terminators (**Fig. 6a**). Generally, *NLuc* displayed a linear correlation between the measured reporter expression levels and mRNA levels (**Fig. 6b**), whereas *sfGFP* exhibited looser correlation (**Fig. 6c**). Surprisingly, the *35S* terminator yielded higher reporter expression in both reporter combinations compared to its mRNA levels (**Fig. 6b,c**). Taken together, the transcript level analysis suggests that the combinatorial regulation is partially explained by transcript abundance and likely to involve post-transcriptional process. The *35S* terminator for both *NLuc* and *sfGFP* has a post-transcriptional enhancer effect, while *NLuc:NOS*, together with *sfGFP:RbcS2b* and *sfGFP:APT1* constructs demonstrated post-transcriptional repression.

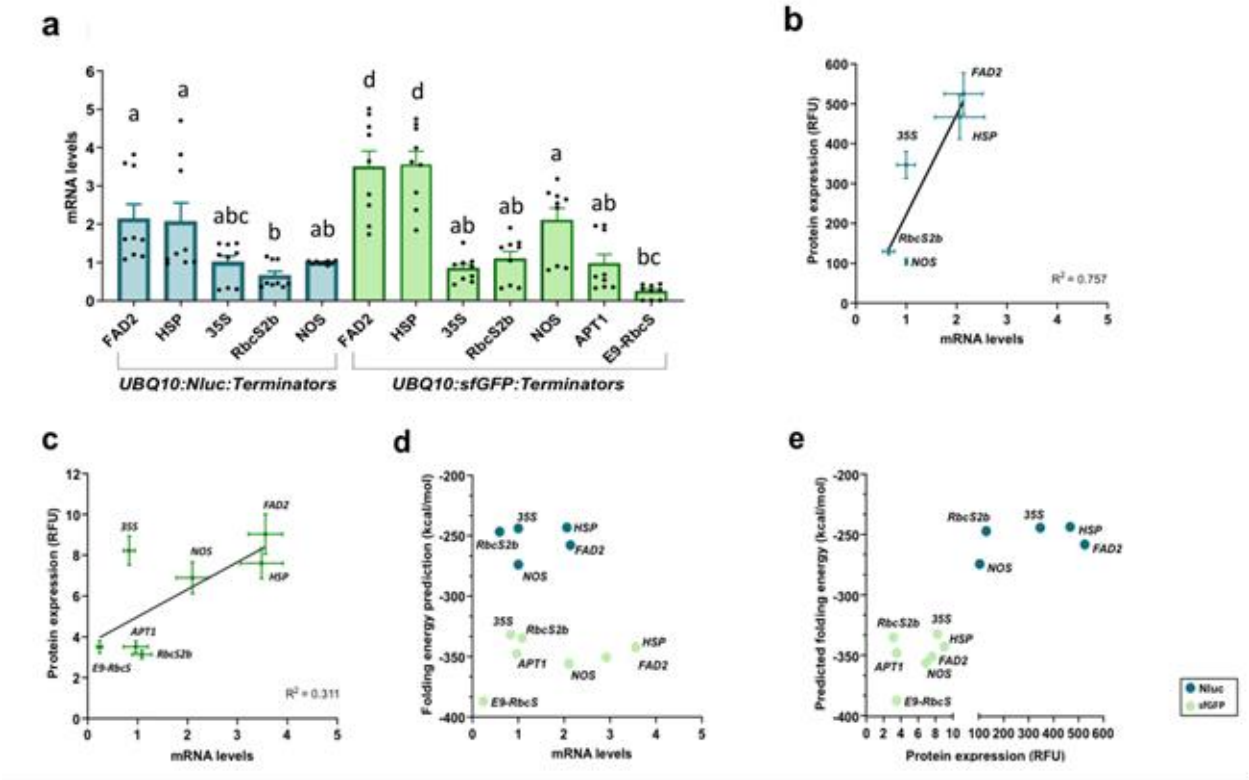


Figure 6. Transcriptional or post-transcriptional regulations contribute to the combinatorial gene expression control

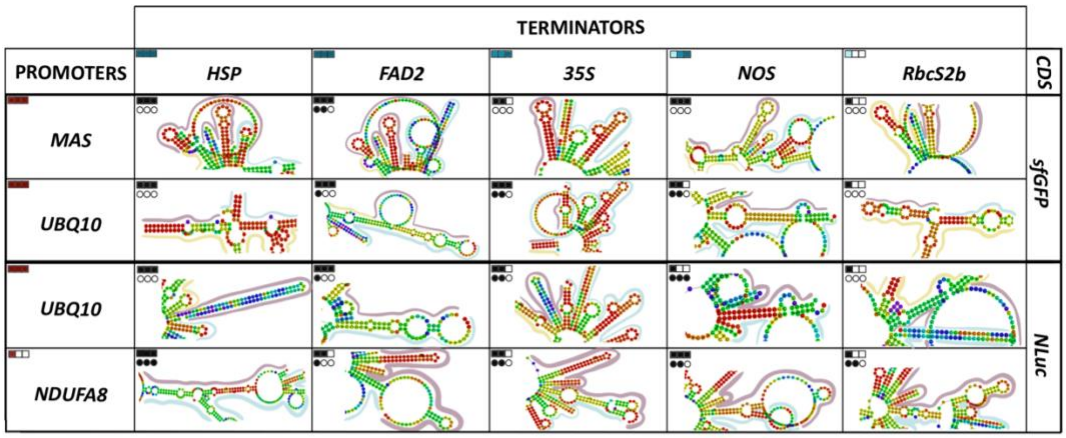
Transcriptional or post-translational gene expression activity was examined with luciferase and fluorescent protein reporters in the transient assay with Arabidopsis leaf mesophyll protoplasts. **a** Transcriptional regulation: mRNA level was quantitated by qPCR for NLuc and sfGFP expression with UBQ10 promoter and different terminators. *UBQ10:FLuc:UBQ5* was used as the normalizing construct for *NLuc*, and *UBQ10:mScarlet-I:UBQ5* for sfGFP. One-way ANOVA was applied for the statistical analysis**.** All the data points from three biological replicates are shown on the graph as dots. (**b,c**) Correlation of the reporter activity to the mRNA levels of different terminator constructs: **b** NLuc and **c** sfGFP. Crossing lines indicates the mean ± SE. (**d,c**) Correlation of predicted RNA folding energy to (**d**) the mRNA levels or (**e**) the reporter protein activity levels. Blue circles represent *NLuc* constructs, and green circles *sfGFP* constructs. Folding energy prediction values are available in **Supplementary Table S5.**

To gain further insights into how the promoter, coding sequence, and terminator interact post-transcriptionally, we studied RNA folding of the transcripts. Using the RNAFold software[59], the folding energy of the whole RNA sequence was calculated for the transcript species we examined with the qPCR analysis above. The transcription start site (TSS) was identified based on the TSSPlant software[51], and the promoter sequence from there was incorporated, along with the protein coding and terminator sequences. However, no direct correlation was seen between the predicted transcript folding energy and mRNA expression levels or between the folding energies and reporter gene expression (**Fig. 6d,e**). We, therefore, proceeded to examine the RNA folding structure and local interactions among the nucleotide sequences.

RNA folding was predicted visually using the ViennaFold2.0 software[59]. The whole transcript (promoter downstream of the TSS, coding sequence, and terminator) was used as the input. Within the transcript, the strong *UBQ10* and *MAS* promoters exhibited little interaction with the other parts: no interaction with *HSP* and *RbcS2b* terminators and minimal to medium interactions with the *FAD2* terminator (**Fig. 7a**). On the contrary, *NOS*, a highly variable terminator, displayed strong interactions with the other parts (e.g., *UBQ10:NLuc* and *pOp6:NLuc*) when it drove weak reporter expression (**Fig. 7a,b**). Interestingly, there was no apparent correlation between inducible promoter-terminator structures and reporter output (**Fig. 7b**; except *pOp6:NLuc:NOS*). We then compared the RNA folding patterns of terminators alone to the folding pattern they adopt within the transcripts. For the strong and consistent terminators like *HSP*, we tended to find loops bigger than 20 bp (**Fig. 7c).** Interestingly, when the *NOS* terminator is paired with strong promoters (*UBQ10* and *MAS*), it also has a large loop structure, which disappeared when combined with other promoters**.**

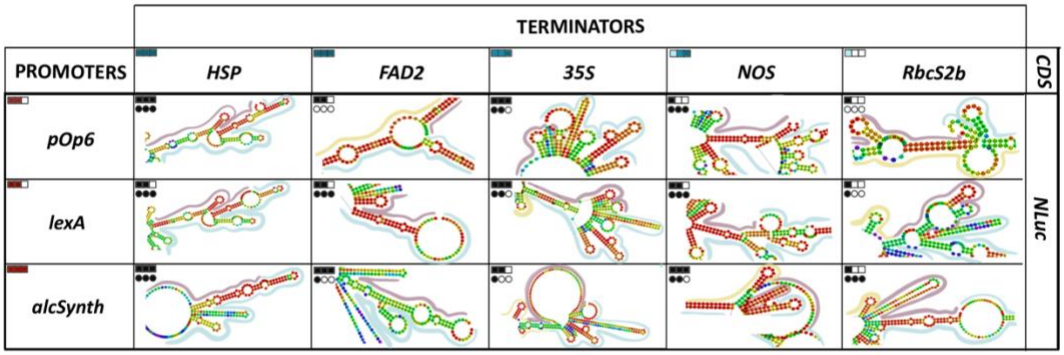
a

Consistent parts are predicted to lack strong interactions when combined with each other



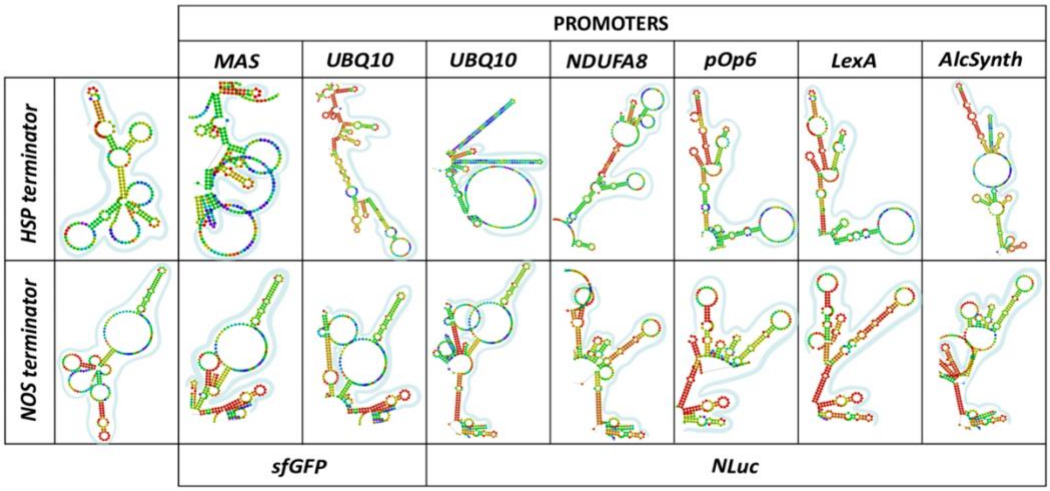
b

Synthetic inducible promoters are not affected by part interactions



c

***NOS* terminator has a conserved domain when paired with strong terminators**



Promoter strength | Terminator strength | Reporter strength | Interaction level

Promoter | Terminator | Coding sequence | Base-pairing probability

Figure 7. Interaction patterns of RNA part structures

RNA folding was visualized to show the likelihood of intra-transcript nucleotide interactions. The terminator strength is represented as a blue grid, where dark blue and full gird represents strong and light blue and empty grid means weak. A full grid with changing colours (such as the one of *NOS*) means the terminator's expression is changing depending on which parts it is combined with. Promoter and reporter expression strength are represented similarly, where full grid means consistently strong and empty grid means weak. Interaction levels are categorized as none (empty circles), weak (one circle full), medium (two circles full) and strong (three circles full). For none, there is no base pairing predicted between promoter and terminator parts. For weak, there is only base pairing with blue and green colours, which means it is <0.5 probable. For medium, there is base pairing in yellow and orange colours, which means 0.75 probability, or 5 bp with red base pairing, indicating ~1.0 probability. Strong interactions are defined as >5 red base pairs.

(**a**) Consistently strong or weak parts don't interact with the other parts. *MAS* (strong), *UBQ10* (strong) *and NDUFA8* (weak) promoters from their predicted transcription start site were combined with the chosen terminators *HSP, FAD2, 35S, NOS* and *RbcS2b. MAS* and *UBQ10* promoters were combined with sfGFP, in addition to NLuc. (**b**) Inducible promoters (*pOp6, lexA and alcSynth)* drive gene expression regardless of RNA fold-based interactions. (**c**) Variable gene expression correlates with loop/stem-loop configuration shift due to the part interactions. A variable activity terminator *NOS* forms a conserved loop domain (>20 bp) when paired with strong promoters like *UBQ10* and *MAS*. The loop domain becomes a stem-loop when it is combined with a weaker promoter like *NDUFA8, pOp6, LexA,* or *AlcSynth.* Meanwhile, a consistently strong HSP terminator retains a loop structure regardless of the partnering parts.

The loop structure encompasses parts of the sequences that is mainly unstructured, and low affinity sequences (e.g., A/T/U-rich) is more prone to the secondary structure. We thus examined the GC content and A/T/U-rich sequences (e.g., the canonical poly-A signal AAUAAA and UGUA motifs) (**Fig. 8)**. Consistently strong terminators have a wide range of GC content *HSP* (29.1%), *FAD2* (36.5%). Followed by *35S* (32%) on the strong side. And weak terminators with *RbcS2b* (30%), *E9-RbcS* (32.8%), *APT1* (35%), *NOS* (38.7%). However, one feature that is predominant in *HSP, FAD2* and *35S* terminators is that the GC content is lowest at a 50 bp window, which is located around their dominant Poly-A signal, and then the GC content goes back up (**Fig. 8a**).

We wanted to gain further insights into how terminators affect gene expression. We generated a deletion series of the strong terminator (*HSP* or *FAD2*). In five terminator variations sequentially deleting the possible regulatory elements, such as Poly-A signals, putative destabilization signals, and Musashi binding elements. Statistical analysis of the results showed only a significant difference between *FAD2* Sequence 5, which is the shortest variation with no Poly-A site, and the intact *FAD2* as well as FAD2 Sequence 1 (200 bp, longest terminator from the deletion series). Concerning the rest of the *HSP* series, as well as the *FAD2* series, there was no statistically significant difference in the reporter expression level in *Arabidopsis* protoplast (**Supplementary Fig.4a,b**). This result suggests a Poly-A site is crucial in enhancing gene expression in some terminators but not in all, and terminator functional elements are as yet unresolved, especially in the plant contexts.

Looking further into consistent terminator strength, we searched for the UGUA motif, which is known to increases the strength of the *35S* terminator, in *HSP* and *FAD2* terminators. The *HSP* terminator has two UGUA motifs at locations 119 bp and 206 bp, while *FAD2* lacks the motif. Interestingly, the XSTREME motif discovery tool identified four motifs that putatively have functional roles, present in *HSP* and *FAD2* terminators (**Fig. 8b**). The most interesting one is the 15-bp motif (CAAAUGUUUUGUGUC) that is found around 145 bp in both *HSP* and *FAD2* terminators, which correspond to the transcript cleavage site. Other three motifs - CUCAUUAUGUUA, UUGUUGUGUUAUGAC, and UUUUUCUAAUAUUA - were found at similar locations in both of the terminators but around 10-20 bp apart. Only the CUCAUUAUGUUA motif is present in *UBQ5* terminator at 167 bp (**Fig. 8b**). Many of these AT/U-rich motifs reside inside the low GC domains as described above.

After investigating consistently strong terminators, focus was then shifted to the terminators that perform stronger or weaker than expected for their transcript levels, which are termed outlier terminators. The GC content of outlier terminators are moderate and varied between 30.5%-35.0%, compared to the average terminator GC content of *A. thaliana* (32.5%) (**Fig. 8c**). The presence of the canonical AAUAAA poly-A signal tends to enhance terminator strength, and from the chosen terminators, *RbcS2b* is the only outlier terminator that doesn't possess the canonical poly-A signal(**Fig. 8c**)

Along with the Poly-A signal, the presence of the UGUA motif around 30-40 bp upstream of the cleavage site is thought to enhance RNA cleavage and thus increase terminator strength[71]. *35S* terminator has three repeats of UGUA motif (x3 starting form 96 bp separated by UU). On the other hand, presence of UUUUU motif is more complicated, where it decreases terminator strength in maize protoplasts, especially if they surround a UGUA motif. In contrast, U-rich sequences increase terminator strength in tobacco leaves[71]. The UUUUU motif is present in *NOS*, *RbcS2b* and *APT1* terminators. All the motifs and their location within the terminators are shown in **Fig. 8c**. No direct correlation emerged between the base-pairing of the AUG start codon and gene expression for the samples (**Fig. 8d**). Base-pairing of the AUG start codon reduces gene expression in bacteria[72]. *35S* terminator is predicted to have base-pairing at AUG codon for *sfGFP;* and *APT1* terminator again with *sfGFP*. Other constructs were predicted to have no base-pairing at AUG (**Fig. 8d**).

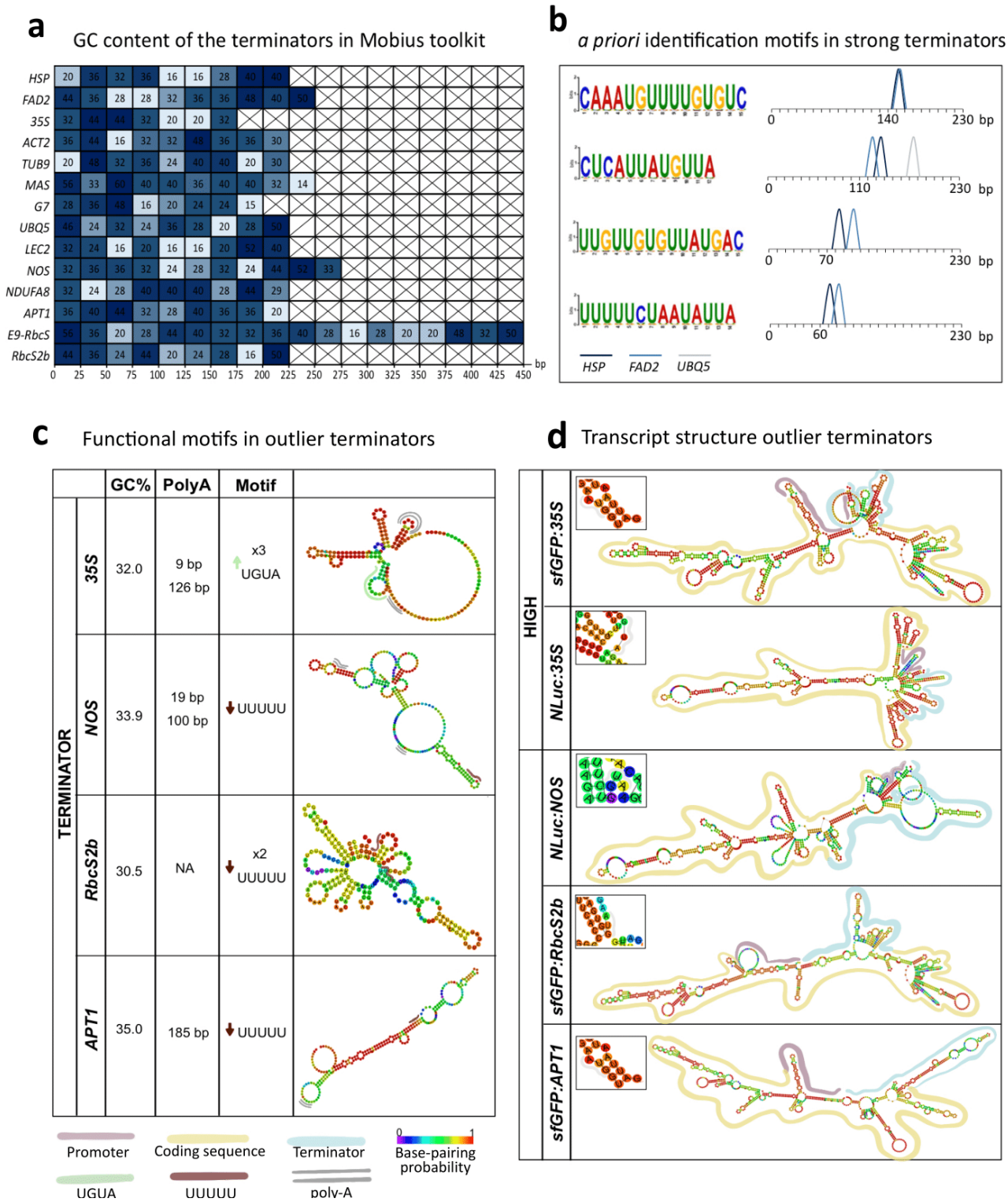


Fig. 8. Terminator structural features that may affect gene expression

(**a**) GC content of the terminators in the toolkit with 25 bp windows. The hue of the blue shade means the GC content (the darker the higher). Consistently strong terminators *HSP* and *FAD2*,

show a local GC drop of 50-bp that correlates with their dominant Poly-A site. (**b)** *A priori* identification of motifs in consistently strong terminators. The XSTREME software identified four motifs that are present only in *HSP* and *FAD2* terminators. First of the identified motifs, CAAAUGUUUGUGUC at 145 bp and 146 bp for *HSP* and *FAD2,* respectively, possibly correlates with high cleave efficiency. (**c**) Functional motifs in the outlier terminators which are terminators that perform stronger or weaker than expected compared to their transcript levels. *RbcS2b* lacks the canonical Poly-A signal, whereas *35S, NOS* and *APT1* do. The enhancer UGUA motif is present in *35S* terminator, while the UUUUU motif that is context dependent is present in *NOS, RbcS2b* and *APT1* terminators. (**d)** Transcript structure of outlier terminators. AUG start codon seems to be base-paired in *sfGFP:35S* and *sfGFP:APT1* constructs, but not for the others. *UBQ10* promoter from its predicted TSS is used as the 'promoter' region.

Unstructured, loop-like secondary structures may also positively regulate gene activation in promoters. *UBQ10* is the strongest promoter in MAPS toolkit, and its TSS structure is predicted to be unstructured and single-stranded (**Fig. 9a,b**), which might be enhancing the promoter's strength. To test this hypothesis, three mutated versions of TSS were created. The first version, Mutated Control (MUTC), involves point mutations that preserve the predicted loop structure. Additionally, two other versions, Mutated 1 (MUT1) and a slightly more structured Mutated 2 (MUT2), were designed with point mutations altering the TSS structure to stem-loops (**Fig. 9a,b**). TSSPlant and Softberry software were used to confirm the mutated versions do not interfere with the identified motifs or introduce new motifs compared to the WT. When the TSS variants were used to drive *sfGFP* with five different terminators (*HSP, FAD2, 35S, NOS,* and *RbcS2b*), the results revealed no statistically significant difference in expression between WT and MUTC, as expected (**Fig. 9c**). However, MUT2 exhibited a reduction in expression by approximately 20% for the *HSP* and *FAD2* terminators. Additionally, MUT1 resulted in a 60% decrease in *35S* expression, while resulting in a 30% increase in gene expression in conjunction with *NOS* terminator. There was no difference between the three mutated versions and the WT for the *RbcS2b* samples, where the WT expression is already weak (**Fig. 9c**).

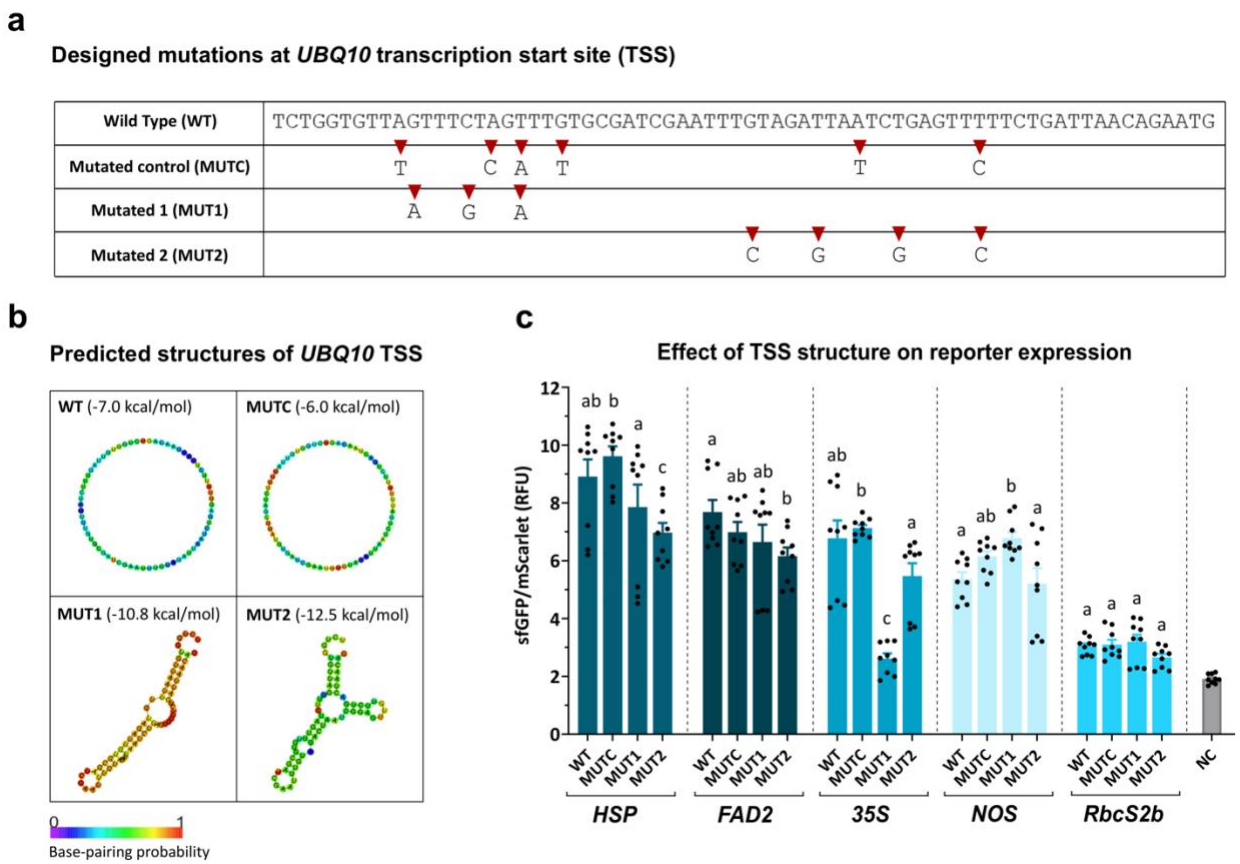


Figure 9. Open/closed structure at TSS affects gene expression.

Strong promoters, such as *UBQ10*, tend to have an open loop structure at the TSS. To test the significance of this structural signature, point mutations were introduced to form closed stem-loop structures instead. (**a** Nucleotide sequences showing the point mutations made for the Wild Type (WT), Mutated control (MUTC), Mutated 1 (MUT1), and Mutated 2 (MUT2) sequences, and their predicted RNA folding structures (**b**). (**c**) Open loop structure at TSS enhances gene expression. sfGFP was expressed under *UBQ10* promoter containing WT, MUTC, MUT1 or MUT2 type TSS. Five different terminators were chosen to assess the effect: *HSP, FAD2, 35S, NOS,* and *RbcS2b*. Normalization construct was *UBQ10:mScarlet-I:UBQ5* for all samples. Bar graphs show mean reporter activity values ± SE. Negative control (NC) was untransformed protoplasts. One-way ANOVA with Tukey's HSD was done individually for each terminator category.

DISCUSSION

MAPS: A new synthetic biology toolkit for engineering with plant systems

In the present work, we established the Mobius Assembly for Plant Systems (MAPS), which is an adaptation of the Mobius Assembly[21] for plant genetic programming. MAPS is a streamlined and user-friendly DNA assembly platform based on small binary vectors, and it comes with a characterized standard part toolkit for plant expression. We have also introduced a new feature for Mobius Assembly to facilitate the generation of a series of constructs with part variations; MethyAble exploits *in vitro* DNA methylation to directly feed Phytobricks in single or multi-TU constructs.

For plant genetic engineering, small binary vectors are vital in enhancing efficiency in cloning and transformation. As we wanted a reliable core vector in our kit forboth cloning and transformation applications, we developed a new binary vector pMAP. It bears a fused replication of origin of pUC with pWKS1 from a *Paracoccus pantotrophus* cryptic plasmid, and it is suitable for experiments that demand high DNA yield, such as the protoplast assay used in this study. Since the high copy number plasmids were at risk of forming plasmid dimers, we have also included *cer* domain for dimer resolution, further improving plasmid stability[60].

Furthermore, we generated and characterized a collection of short promoters (17 constitutive and three inducible) and 14 terminator standard parts for plant expression. Ten of the promoters and six of the terminators were newly isolated, and they have short in length: 300-600 bp (promoters) and 200bp (terminators). The size of the available standard promoter or terminator parts for plant engineering (e.g., MoClo plants, GB plants, and GreenGate) can be as long as 4 kb. Large constructs can reduce plasmid stability in bacteria (*E. coli* and *Agrobacterium*) while also lowering transformation efficiency and causing incomplete/truncated transformation[14,15].

With MAPS, the collective plant synbio toolbox is further equipped with short promoters that can drive a gradient of target gene expression. An *Actin 2* promoter (787bp) was characterized in MoClo and GB2.0 toolboxes, showing a comparable activity to the NOS promoter[73,74]. The shorter version we made in this study *ACT2* (340bp) exhibited similar expression activity (**Fig. 3**). More specifically, in combination with the *HSP* terminator, the *NOS* and *ACT2* promoters had 1.4 and4.6 RLU, respectively, while the corresponding expression from the GB2.0 versions was 2.61 and 4.36 RLU[12]. *UBQ11* and *UBQ4* promoters showed strong expression, and they performed better than the commonly used *35S* and *OCS* in the protoplast context.

Sequence analysis through the online programs PLACE and PlantCARE showed that the *UBQ10* promoter contains 11 CAAT-box and 6 DOFCOREZM, while the *UBQ4* promoter harbors 9 CAAT-box and 7 DOFCOREZM elements (**Supplementary Fig. S2**). DOFCOREZM elements are the binding sites of Dof1 and Dof2 TFs, which were found to enhance gene expression[75]. CCAT-box also influences expression efficiency[76]. The *UBQ10* promoter is a commonly used constitutive promoter, and it exhibited the highest expression in our protoplast assay, ~20 and ~11 times higher than the *35S* promoter, combined with *NDUFA8* and *HSP* terminators, respectively.

Most part characterization to date has been done in either Arabidopsis or *Nicotiana benthamiana,* which are both dicot plants. However, parts can behave differently in monocots;, for example, *35S* and *NOS* promoters drive low activity in monocot *Setaria viridis* [19,77]. Even within dicots, the same part can have different levels of expression. Contrary to our study with Arabidopsis protoplasts, *35S* promoter had the highest activity in *N. benthamiana* leaves[73,74]. Similarly, *OCS* promoter had strong activity in our study but low expression in *N. benthamiana*[73]. Genetic parts can behave

differently in different chassis, and we should not generalize characterization results among different plant species or expression systems (e.g., whole plants, leaves, protoplasts, or cell cultures).

Alternative to constitutive expression, chemically inducible systems are powerful tools to control the timing, localization, and amplitude of transgene expression. Dex, estradiol, and ethanol-inducible systems were developed almost 20 years ago and have been used in many studies since then[67]. In this study, we have shown that the extent of target gene induction can be modulated by exchanging the terminators (**Fig. 4**).

Terminators influence gene expression, in part through interaction with promoters

Comprehensive characterization of the terminators underscored the strong influence of the terminators on gene expression. It is a common practice to rely on promoters to control gene expression level and overlook the impact of terminators. For example, when the domestication of the synthetic plant promoter *G10-90* promoter resulted in loss of expression, researchers replaced it withthe *AtRPL37aC* promoter to rectify[69]. However, they were using the *psE9-RbcS* terminator, which we found to drive low gene expression; the gene expression could have been better restored with a more activating terminator instead. Even though promoters generally cast a great influence on gene expression **(Supplementary Table. S7),** terminators could alter the gene expression 3-8 folds in this study.

Terminators can alter gene expression in multiple ways. 3'UTR participates in quality control inthe post-transcriptional regulatory mechanism for eukaryotic genes through nonsense-mediated mRNA decay[78]. Genome-wide analysis of the transcript degradation profiles in Arabidopsis revealed specific sequence motifs at 3' UTR that stabilize or destabilize mRNA (such as TTGCTT for stabilization and AATTTT destabilization motifs)[79] Several motifs in the 3'UTR might be involved in polyadenylation or in mRNA decay[80]. Strong gene expression is linked with short poly(A) tails, while poly(A)-binding proteins can facilitate both the protection and degradation of mRNA[81]. In Arabidopsis, poly(A) tail was found to block RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) from converting aberrant mRNAs into substrates for degradation[82]. Unpolyadenylated transcripts derived from terminator-less constructs or readthrough mRNAs from transgenes with strong promoters are subjected to RDR6-mediated silencing[83]. Reduction in RDR6-mediated silencing is partially attributed to why double terminators can increase gene expression[84–86].

However, this explanation cannot sufficiently explain the synergistic effects the terminators cast, especially when combined with weak promoters. Meanwhile, direct interactions of the terminators with the promoters and the transcription initiation machinery could. The synergistic regulation between natural promoters and terminators can be explained by the physical interaction of the terminator with the promoter (RNA or DNA/chromatin looping). Gene looping is the structural interaction of the terminator with the promoter within a gene, mediated by DNA-binding proteins[87]. It can influence gene expression through transcriptional memory[88], intron-mediated modulation of transcription[89], transcription directionality[90], reinitiation of transcription[91], and transcription termination[92,93]. It was recently shown that the absence of a terminator caused higher DNA methylation on the promoter region and reduced the transgene expression compared to constructs with a terminator, indicating an alternative role of terminator sequence in epigenetic gene regulation[94].

Promoter-coding sequence-terminator interactions in gene regulation

In addition to the interactions between terminators and promoters, we observed that the protein-coding sequence could modify the strength of promoter-terminator pairs. The coding sequence can impact gene expression through factors such as codon usage bias, gene length, GC content, correct 5' cap, and mRNA folding[95]. Both the *NLuc* (516 bp, GC 52.7%) and *sfGFP* (714 bp, GC 61.3%) have been optimized for plant systems in terms of codon usage. Any factors related to codon usage, gene length, GC content, or correct 5' cap would primarily affect the magnitudes of the reporter readout, but the expression profiles of promoter-terminator pairs would remain the same. However, the relative strengths of promoters with *the HSP* terminator and terminators with the *UBQ10* promoter shifted when reporter proteins were replaced (**Fig. 3, Fig.5**), indicating that the coding sequence interact directly with the promoter and the terminator.

We have observed clear interactions between promoters, coding sequences, and terminators (**Fig. 3-5; Table 1**). Only four out of the 14 terminators exhibited consistent behaviour in both reporters tested: *FAD2* and *HSP* were always strong while *LEC2* and *RbcS2b* were always weak. Therefore, a terminator can have a different impact on gene expression depending on the promoter and coding sequences.

Next, we wanted to understand how the promoter-coding sequence-terminator interactions occur. There are many different mechanisms, but such interactions can be explained at the DNA or RNA level. The qPCR assay indicated that the reporter genes are regulated transcriptionally which changes the stable transcript abundance and also post-transcriptionally, as transcript abundance alone does not always correlate with reporter expression **(Fig. 6).** Besides, it is hard to do *in silico* folding predictions of long strands of double-stranded DNA, while resources for RNA folding are available. Therefore, we focused on the RNA secondary structures.

Analysis of the part interactions predicted the consistent promoter and terminator parts tend to be isolated in terms of secondary structure, displaying a low level of physical interactions from the other parts, regardless of their sequences (**Fig. 7a**). Additionally in the case of highly variable *NOS* terminator, part interactions and therefore the local RNA folding might be affecting the reporter output. It is likely that certain promoters and terminators possess currently unknown regulatory elements that make them strong and physically restrict their interactions with other parts, while the others may require physical interactions with other parts to function effectively. In mammalian systems for instance, terminator elements require specific secondary structures to interact between the polyadenylation complex and mRNA[96]. Similar mechanisms may be at play in plant systems, and the absence of interaction could serve as an enhancing feature for one part while acting as a repressive feature for another. Therefore, using synthetic parts with known features like the *pOp6* and *LexA* inducible promoters are promising, as even though they are predicted to interact with other parts, the interaction does not seem to have a strong effect on the reporter output (**Fig. 7b**).

Recent studies on the roles of RNA structures in gene expression control have revealed intriguing insights. Lower mRNA folding energies, typically associated with greater structural stability, have been found to impact ribosome function and translation in organisms such as *E. coli* and *S. cerevisiae*[97,98]. Beyond global folding, local secondary structures can also play a pivotal role in shaping gene expression. Notably, pseudoknots, hairpins, loops, and the presence of hydrogen bonding patterns upstream of the transcription start site have been shown to reduce ribosome sequestration and translation rates[99,100]. A design-of-experiments study with 224,000 synthetic sequences in *E. coli* has highlighted the importance of secondary structures at the 30 bp region

both upstream and downstream of the start codon (referred to as $STR_{-30:+30}$)[101]. Furthermore, it revealed that even distal regions to the STR can exert a strong influence over transcript and protein abundance by outcompeting initiation-limiting structures[101]. While comprehensive studies in plants are still pending, *in vivo* genome-wide profiling of Arabidopsis RNA secondary structures revealed that unstructured regions upstream of the start codon are enriched in high translation efficiency mRNAs[102]. In the strong promoter *UBQ10*, we found an unstructured TSS (loop) whose structure affected reporter output (**Fig. 9**).

RNA secondary structures also point to a similar regulatory influence in terminator-dependent gene regulation. When the consistently strong terminators *HSP and NOS* were combined with various promoters, loops >20 bp were always present (**Fig. 7a**). This might be a structural feature of strong-acting terminator, since a similar >20 bp loop was predicted when the variable *NOS* terminator was combined with strong promoters (*MAS, UBQ10);* yet the loop re-configured to a stem-loop structure when combined with weaker promoters (**Fig. 7c**). Expanding on these structural themes, terminators from Felippes et al. (2022) that showed increased strength when combined with the first 32 bp of the *HSP* terminator were used as templates. Analysis of the terminator structure alone and within transcripts, with and without the added 32 bp, showed that the addition of the 32 bp led to an increase in the size of the loops >20 bp for *NOS* and *ACS2* terminators, as well as higher base-pairing probability structures for *Rbcs1A* (**Supplementary Fig. S6**).

Sequences and secondary structures contributing to terminator strength

The two consistently high-expression terminators (*HSP* and *FAD2*) exhibited higher mRNA levels compared to the other terminators, which are probably contributing to their consistent high-performance (**Fig. 6a**)**.** The weak *E9-Rbcs2b* terminator, displayed significantly low mRNA level, indicating that it is predominantly regulated at the transcriptional level for *sfGFP* (**Fig. 6a**). However, we have also observed differences between the measured mRNA levels and the reporter protein output. For example, the mRNA levels of *35S* and *NOS* terminators were not significantly different, but the *35S* terminator results in 3 times higher *NLuc* expression. These differences can be explained by the presence of UGUA motif in *35S* terminator, which acts as an enhancer by improving cleavage efficiency; and probably the UUUUU motif functioning as destabilizer for the *NOS* terminator (**Fig. 7c**).

When consistently strong terminators *HSP* and *FAD2* were investigated further, a local drop in GC content of 50 bp was identified, that was absent in other terminators (except the *35S* terminator, which is also on the strong and consistent side). So, it is likely that this low GC region have structural features that aid recognition of transcription termination and cleavage. When *a priori* investigation was done to identify possible motifs that are present in *HSP* and *FAD2*, but lack in other terminators in our toolkit, XSTREME motif discovery tool have identified four motifs. Amongst the four, CAAAUGUUUGUGUC motif at location 145 bp for *HSP*; and 146 bp for *FAD2*, possibly aids with cleavage site recognition and cleavage efficiency contributing to overall terminator strength (**Fig 7b**).

With our deletion series we have seen that only the shortest *FAD2* terminator (70 bp) has significantly reduced strength compared to the full length *FAD2*, while the 35 bp *HSP* terminator's strength is statistically insignificant compared to the full length *HSP* terminator (**Supplementary Fig. S4**). Since Felippes et al., 2023 have also showed that the first 32 bp of the *HSP* terminator is a transferable element that enhances the strength of other terminators, it is likely that the *FAD2*

terminator has similar functionality, but between 70-200 bp regions, which corresponds to the location where the identified motifs reside (**Fig. 7a,b**).

FUNDING

This project was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) High-Value Compound from Plants (HVCfP) Network Proof-of-Concept Award (POC-NOV16-04), Royal Society University Research Fellowship (UF140640 and URF\R\201035), and Schmidt Science Polymath Award to NN, as well as the University of Edinburgh Principal's Career Development PhD Studentship to AIA, the IBioIC CTP PhD Studentship to JN, and Imperial College Bioengineering Studentship to EGK.

REFERENCES

1. Liu, W., Yuan, J. & Stewart, N. Advanced genetic tools for plant biotechnology. *Nature reviews genetics* **14**, 781–793 (2013).
2. Naqvi, S. *et al.* When more is better: multigene engineering in plants. *Trends in Plant Science* **1**, 48–56 (2010).
3. Que, Q. *et al.* Trait stacking in transgenic crops: Challenges and opportunities. *gmcrops* **1**, 220–229 (2010).
4. Townson, J. Recent developments in genome editing for potential use in plants. *Bioscience Horizons* **10**, (2017).
5. Zhu, Q. *et al.* Development of “Purple Endosperm Rice” by Engineering Anthocyanin Biosynthesis in the Endosperm with a High-Efficiency Transgene Stacking System. *Molecular Plant* **10**, 918–929 (2017).
6. Patron, N. Beyond natural: synthetic expansions of botanical form and function. *New Phytologist* **227**, (2020).
7. Mortimer, J. Plant synthetic biology could drive a revolution in biofuels and medicine. **244**, (2019).
8. Wu, G. *et al.* Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nat Biotechnol* **23**, 1013–1017 (2005).
9. Andrianantoandro, E., Basu, S., Karig, D. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology* **2**, (2006).
10. Cai, Y.-M., Lopez, J. C. & Patron, N. Phytobricks: Manual and Automated Assembly of Constructs for Engineering Plants. *Methods in Molecular Biology* **2205**, 179–199 (2020).
11. Felippes, F. *et al.* The key role of terminators on the expression and post-transcriptional gene silencing of transgenes. *The Plant Journal* **104**, 96–112 (2020).
12. Oliveira, P. H., Prather, K. J., Prazeres, D. M. F. & Monteiro, G. A. Analysis of DNA repeats in bacterial plasmids reveals the potential for recurrent instability events. *Appl Microbiol Biotechnol* **87**, 2157–2167 (2010).
13. Peremarti, A. *et al.* Promoter diversity in multigene transformation. *Plant Mol Biol* **73**, 363–378 (2010).
14. Ertl, P. & Thomsen, L. Technical issues in construction of nucleic acid vaccines. *Methods* **31**, 199–206 (2003).
15. Park, S. H., Lee, B.-M., Salas, M. G., Srivatanakul, M. & Smith, R. H. Shorter T-DNA or additional virulence genes improve Agrobactrium-mediated transformation: *Theor Appl Genet* **101**, 1015–1020 (2000).

16. Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature Methods* **10**, 659–664 (2013).
17. Yamanishi, M. *et al.* A Genome-Wide Activity Assessment of Terminator Regions in Saccharomyces cerevisiae Provides a ″Terminatome″ Toolbox. **2**, 337–347 (2013).
18. Niederau, P. A. *et al.* Multifactorial analysis of terminator performance on heterologous gene expression in Physcomitrella. *Plant Cell Rep* **43**, 43 (2024).
19. Chamness, J. C. *et al. An Extensible Vector Toolkit and Parts Library for Advanced Engineering of Plant Genomes*. http://biorxiv.org/lookup/doi/10.1101/2022.10.15.511792 (2022) doi:10.1101/2022.10.15.511792.
20. Tian, C., Zhang, Y., Li, J. & Wang, Y. Benchmarking Intrinsic Promoters and Terminators for Plant Synthetic Biology Research. *BioDesign Research* **2022**, 1–12 (2022).
21. Andreou, A. & Nakayama, N. Mobius Assembly: A versatile Golden-Gate framework towards universal DNA assembly. *PLOS ONE* **13**, (2018).
22. Andreou, A. & Nakayama, N. Mobius Assembly. in *DNA Cloning and Assembly: Methods and Protocols* vol. 1116.
23. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS ONE* **6**, e16765 (2011).
24. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS ONE* **4**, e5553 (2009).
25. Potapov, V. *et al.* A single-molecule sequencing assay for the comprehensive profiling of T4 DNA ligase fidelity and bias during DNA end-joining. *Nucleic Acids Research* **46**, e79–e79 (2018).
26. Potapov, V. *et al.* Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synth. Biol.* **7**, 2665–2674 (2018).
27. Pryor, J. M. *et al.* Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design. *PLoS ONE* **15**, e0238592 (2020).
28. Lin, D. & O'Callaghan, C. A. MetClo: methylase-assisted hierarchical DNA assembly using a single type IIS restriction enzyme. *Nucleic Acids Research* (2018) doi:10.1093/nar/gky596.
29. Taylor, G. M., Mordaka, P. M. & Heap, J. T. Start-Stop Assembly: a functionally scarless DNA assembly system optimized for metabolic engineering. *Nucleic Acids Research* **47**, e17–e17 (2019).
30. van Dolleweerd, C. J. *et al.* MIDAS: A Modular DNA Assembly System for Synthetic Biology. *ACS Synth. Biol.* **7**, 1018–1029 (2018).
31. Valenzuela-Ortega, M. & French, C. Joint universal modular plasmids (JUMP): a flexible vector platform for synthetic biology. *Synthetic Biology* **6**, ysab003 (2021).
32. Damalas, S. G., Batianis, C., Martin-Pascual, M., Lorenzo, V. & Martins dos Santos, V. A. P. SEVA 3.1: enabling interoperability of DNA assembly among the SEVA, BioBricks and Type IIS restriction enzyme standards. *Microb. Biotechnol.* **13**, 1793–1806 (2020).
33. Vasudevan, R. *et al.* CyanoGate: A Modular Cloning Suite for Engineering Cyanobacteria Based on the Plant MoClo Syntax. *Plant Physiol.* **180**, 39–55 (2019).
34. Crozet, P. *et al.* Birth of a Photosynthetic Chassis: A MoClo Toolkit Enabling Synthetic Biology in the Microalga *Chlamydomonas reinhardtii*. *ACS Synth. Biol.* **7**, 2074–2086 (2018).
35. Hernanz-Koers, M. *et al.* FungalBraid: A GoldenBraid-based modular cloning platform for the assembly and exchange of DNA elements tailored to fungal synthetic biology. *Fungal Genetics and Biology* **116**, 51–61 (2018).
36. Egermeier, M., Sauer, M. & Marx, H. Golden Gate-based metabolic engineering strategy for wild-type strains of *Yarrowia lipolytica*. *FEMS Microbiology Letters* **366**, (2019).
37. Pollak, B. *et al.* Universal loop assembly: open, efficient and cross-kingdom DNA

fabrication. *Synthetic Biology* **5**, ysaa001 (2020).
38. Chiasson, D. *et al.* A unified multi-kingdom Golden Gate cloning platform. *Sci Rep* **9**, 10131 (2019).
39. Pollak, B. *et al.* Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytol* **222**, 628–640 (2019).
40. Gantner, J. *et al.* Peripheral infrastructure vectors and an extended set of plant parts for the Modular Cloning system. *PLoS ONE* **13**, e0197185 (2018).
41. Komori, T. *et al.* Current Status of Binary Vectors and Superbinary Vectors. *Plant Physiology* **145**, 1155–1160 (2007).
42. Silva, F., Queiroz, J. A. & Domingues, F. C. Evaluating metabolic stress and plasmid stability in plasmid DNA production by Escherichia coli. *Biotechnology Advances* **30**, 691–708 (2012).
43. Moore, S. J. *et al.* EcoFlex: A Multifunctional MoClo Kit for *E. coli* Synthetic Biology. *ACS Synth. Biol.* **5**, 1059–1069 (2016).
44. Bi, X. & Liut, L. F. DNA rearrangement mediated by inverted repeats. *Proc. Natl. Acad. Sci. USA* (1996).
45. Lee, S., Su, G., Lasserre, E., Aghazadeh, M. A. & Murai, N. Small high-yielding binary Ti vectors pLSU with co-directional replicons for Agrobacterium tumefaciens-mediated transformation of higher plants. *Plant Science* **187**, 49–58 (2012).
46. Pasin, F. *et al.* Multiple T-DNA Delivery to Plants Using Novel Mini Binary Vectors with Compatible Replication Origins. *ACS Synth. Biol.* **6**, 1962–1968 (2017).
47. Chung, C. & Miller, R. Preparation and storage of competent Escherichia coli cells. *Methods in Enzymology* **218**, 621–627 (1993).
48. Hellens, R. P., Edwards, E. A., Leyland, N. R. & Bean, S. pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation.
49. Guenin, S. *et al.* Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. J. *Journal of Experimental Botany* **60**, 487–93 (2009).
50. Huala, E. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Research* **29**, 102–105 (2001).
51. Shahmuradov, I. A., Umarov, R. Kh. & Solovyev, V. V. TSSPlant: a new tool for prediction of plant Pol II promoters. *Nucleic Acids Res* gkw1353 (2017) doi:10.1093/nar/gkw1353.
52. Lescot, M. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research* **30**, 325–327 (2002).
53. Higo, K., Ugawa, Y., Iwamoto, M. & Korenaga, T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Research* **27**, 297–300 (1999).
54. Ji, G. *et al.* PASPA: a web server for mRNA poly(A) site predictions in plants and algae. *Bioinformatics* **31**, 1671–1673 (2015).
55. Chang, T.-H. *et al.* An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. *BMC Bioinformatics* **14**, S4 (2013).
56. Chupeau, M.-C. *et al.* Characterization of the Early Events Leading to Totipotency in an *Arabidopsis* Protoplast Liquid Culture by Temporal Transcript Profiling. *The Plant Cell* **25**, 2444–2463 (2013).
57. Faraco, M., Di Sansebastiano, G. P., Spelt, K., Koes, R. E. & Quattrocchio, F. M. One Protoplast Is Not the Other! *Plant Physiology* **156**, 474–478 (2011).
58. Mauri, M., Vecchione, S. & Fritz, G. Deconvolution of Luminescence Cross-Talk in High-Throughput Gene Expression Profiling. *ACS Synth. Biol.* **8**, 1361–1370 (2019).
59. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**, 26 (2011).

60. Summers, D. & Sherratt, D. Multimerization of high copy number plasmids causes instability: Cole 1 encodes a determinant essential for plasmid monomerization and stability. *Cell* **36**, 1097–1103 (1984).
61. Watson, M. R. *et al.* An Improved Binary Vector and *Escherichia coli* Strain for *Agrobacterium tumefaciens* -Mediated Plant Transformation. *G3 Genes|Genomes|Genetics* **6**, 2195–2201 (2016).
62. Andreou, A. I., Nirkko, J., Ochoa-Villarreal, M. & Nakayama, N. *Mobius Assembly for Plant Systems Highlights Promoter-Terminator Interaction in Gene Regulation*. http://biorxiv.org/lookup/doi/10.1101/2021.03.31.437819 (2021) doi:10.1101/2021.03.31.437819.
63. Bartosik, D., Baj, J., Sochacka, M., Piechucka, E. & Wlodarczyk, M. Molecular characterization of functional modules of plasmid pWKS1 of Paracoccus pantotrophus DSM 11072.
64. Challacombe, J. F., Pillai, S. & Kuske, C. R. Shared features of cryptic plasmids from environmental and pathogenic Francisella species. *PLoS ONE* **12**, e0183554 (2017).
65. Zaleski, P. *et al.* pIGWZ12 – A cryptic plasmid with a modular structure. *Plasmid* **79**, 37–47 (2015).
66. Naseri, G. & Koffas, M. Application of combinatorial optimization strategies in synthetic biology. *Nature communications* **11**, (2020).
67. Borghi, L. Inducible Gene Expression Systems for Plants. *Plant Developmental Biology* **655**, 65–75 (2010).
68. Craft, J. *et al.* New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in Arabidopsis. *The Plant Journal* **41**, 899–918 (2005).
69. Schlücking, K. *et al.* A New β-Estradiol-Inducible Vector Set that Facilitates Easy Construction and Efficient Expression of Transgenes Reveals CBL3-Dependent Cytoplasm to Tonoplast Translocation of CIPK5. *Molecular Plant* **6**, 1814–1829 (2013).
70. Donaldson, L. Autofluorescence in Plants. *MDPI, Molecules* **25**, (2020).
71. Gorjifard, S. *et al. Features That Govern Terminator Strength in Plants*. http://biorxiv.org/lookup/doi/10.1101/2023.06.16.545379 (2023) doi:10.1101/2023.06.16.545379.
72. Chełkowska-Pauszek, A. *et al.* The Role of RNA Secondary Structure in Regulation of Gene Expression in Bacteria. *IJMS* **22**, 7845 (2021).
73. Engler, C. *et al.* A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synth. Biol.* **3**, 839–843 (2014).
74. Sarrion-Perdigones, A. *et al.* GoldenBraid 2.0: A Comprehensive DNA Assembly Framework for Plant Synthetic Biology. *PLANT PHYSIOLOGY* **162**, 1618–1631 (2013).
75. Yanagisawa, S. Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. *The Plant Journal* **21**, 281–288 (2000).
76. Biłas, R., Szafran, K., Hnatuszko-Konka, K. & Kononowicz, A. K. Cis-regulatory elements used to control gene expression in plants. *Plant Cell Tiss Organ Cult* **127**, 269–287 (2016).
77. Wilmink, A., van de Ven, B. C. E. & Dons, J. J. M. Activity of constitutive promoters in various species from the Liliaceae. *Plant Mol Biol* **28**, 949–955 (1995).
78. Karousis, E. D. & Mühlemann, O. Nonsense-Mediated mRNA Decay Begins Where Translation Ends. *Cold Spring Harb Perspect Biol* **11**, a032862 (2019).
79. Narsai, R. *et al.* Genome-Wide Analysis of mRNA Decay Rates and Their Determinants in *Arabidopsis thaliana*. *The Plant Cell* **19**, 3418–3436 (2007).
80. Bernardes, W. S. & Menossi, M. Plant 3' Regulatory Regions From mRNA-Encoding Genes and Their Uses to Modulate Expression. *Front. Plant Sci.* **11**, 1252 (2020).
81. Lima, S. A. *et al.* Short poly(A) tails are a conserved feature of highly expressed genes. *Nat Struct Mol Biol* **24**, 1057–1063 (2017).

82. Baeg, K., Iwakawa, H. & Tomari, Y. The poly(A) tail blocks RDR6 from converting self mRNAs into substrates for gene silencing. *Nature Plants* **3**, 17036 (2017).
83. Luo, Z. & Chen, Z. Improperly Terminated, Unpolyadenylated mRNA of Sense Transgenes Is Targeted by RDR6-Mediated RNA Silencing in *Arabidopsis*. *The Plant Cell* **19**, 943–958 (2007).
84. Beyene, G. *et al.* Unprecedented enhancement of transient gene expression from minimal cassettes using a double terminator. *Plant Cell Rep* **30**, 13–25 (2011).
85. Yamamoto, T. *et al.* Improvement of the transient expression system for production of recombinant proteins in plants. *Sci Rep* **8**, 4755 (2018).
86. Diamos, A. G. & Mason, H. S. Chimeric 3' flanking regions strongly enhance gene expression in plants. *Plant Biotechnol J* **16**, 1971–1982 (2018).
87. Calvo, O. & Manley, J. L. Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev.* **17**, 1321–1327 (2003).
88. Hampsey, M., Singh, B. N., Ansari, A., Lainé, J.-P. & Krishnamurthy, S. Control of eukaryotic gene expression: Gene loops and transcriptional memory. *Advances in Enzyme Regulation* **51**, 118–125 (2011).
89. Moabbi, A. M., Agarwal, N., El Kaderi, B. & Ansari, A. Role for gene looping in intron-mediated enhancement of transcription. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8505–8510 (2012).
90. Tan-Wong, S. M. *et al.* Gene Loops Enhance Transcriptional Directionality. *Science* **338**, 671–675 (2012).
91. Al Husini, N., Kudla, P. & Ansari, A. A Role for CF1A 3′ End Processing Complex in Promoter-Associated Transcription. *PLoS Genet* **9**, e1003722 (2013).
92. Mukundan, B. & Ansari, A. Srb5/Med18-mediated Termination of Transcription Is Dependent on Gene Looping. *Journal of Biological Chemistry* **288**, 11384–11394 (2013).
93. Medler, S. & Ansari, A. Gene looping facilitates TFIIH kinase-mediated termination of transcription. *Sci Rep* **5**, 12586 (2015).
94. Pérez-González, A. & Caro, E. Effect of transcription terminator usage on the establishment of transgene transcriptional gene silencing. *BMC Res Notes* **11**, 511 (2018).
95. Indian Institute of Integrative Medicine, CSIR, Canal Road, Jammu-180001 & Singh, R. A report on DNA sequence determinants in gene expression. *Bioinformation* **16**, 422 (2020).
96. Zarudnaya, M. I. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. *Nucleic Acids Research* **31**, 1375–1386 (2003).
97. Faure, G., Ogurtsov, A. Y., Shabalina, S. A. & Koonin, E. Adaptation of mRNA structure to control protein folding. *RNA Biology* **14**, 1649–1654 (2017).
98. Tuller, T., Waldman, Y., Kupiec, M. & Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *PNAS* **107**, 3645–3650 (2010).
99. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. *Science* **324**, 255–258 (2009).
100. Gu, W., Zhou, T. & Wilke, C. A Universal Trend of Reduced mRNA Stability near the Translation-Initiation Site in Prokaryotes and Eukaryotes. *PLoS Computational Biology* **6**, (2010).
101. Cambray, G., Guimaraes, J. C. & Arkin, A. P. Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. *Nat Biotechnol* **36**, 1005–1015 (2018).
102. Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2014).