Journal of Bioscience and Bioengineering
VOL. xx No. xx, 1–7, 2017

www.elsevier.com/locate/jbiosc

Arabidopsis thaliana cold-regulated 47 gene 5′-untranslated region enables stable high-level expression of transgenes

Shotaro Yamasaki,[1] Yuji Sanada,[1] Ryoji Imase,[1] Hideyuki Matsuura,[2] Daishin Ueno,[1] Taku Demura,[1] and Ko Kato[1,*]

Graduate School of Biological Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan[1] *and Graduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, Japan*[2]

Received 13 June 2017; accepted 15 August 2017
Available online xxx

Transgene expression is regulated through several steps, this study focuses on the mRNA translation step. The expression level of transgenes can be increased by 5′-untranslated region (5′UTR) sequences in certain genes which act as translational enhancers. On the other hand, translation in most mRNA species is repressed by growth, development, and stress events. There is a possibility that transgene mRNA is also repressed in these conditions, despite the use of a translational enhancer. Therefore, a consistently efficient translational enhancer is needed to develop a reliable transgene expression system. Herein we searched for mRNAs translated stably under different growth, development and environmental conditions using data sets of polysome fraction assays and microarray analysis. Correct 5′UTR sequences of candidate genes were determined by cap analysis of gene expression and we tested translational ability of the candidate 5′UTRs by reporter assays. We found the 5′UTR of cold-regulated 47 gene to be an effective translational enhancer, contributing to stable high-level expression under various conditions.

[**Key words:** Plant molecular farming; Expression system; Translational enhancer; Translational repression; Growth and development; Abiotic stress; 5′UTR; Recombinant proteins; Transgenic]

Plant biotechnology has advanced and developed considerably, with many research institutions and companies actively seeking to improve plant function, to utilize plant abilities, as well as to develop productivity of useful materials, through transgene expression. Particularly, plant molecular farming (PMF), which uses plants to produce useful materials, such as human therapeutic proteins, is a growing field of interest. Compared with production systems using other organisms, PMF has several major advantages such as high production speed, low cost, and safety of usage (1). Maintenance cost is low relative to bacterial, yeast, or mammalian cell expression systems. In plants, proteins can be correctly folded and assembled into complexes with similar glycosylation types to that in the human system, different from bacteria and yeast. The risk of contamination by animal pathogens is also low. Many companies have adopted PMF for production of pharmaceuticals such as antibodies and vaccines (1). For example, during the 2014 Ebola virus outbreak in Africa, the drug ZMapp (Mapp Biopharmaceutical, San Diego, CA, USA) was produced from tobacco leaves, then the drug was actually used to cure some patients (2). One of the challenges of transgene expression in plants is optimization of the expression level. The expression level of a transgene is not always high, but increasing the expression level leads to less cost and simpler purification procedure. One of the important steps in the regulation of transgene expression is mRNA translation. We previously reported that the 5′-untranslated regions (5′UTRs) of alcohol dehydrogenase (*ADH*) and arabinogalactan-protein 21 (*AGP21*) genes function as translational enhancers in plant cells (3–5). Using these translational enhancers, we succeeded in improving the expression level by approximately 100-fold compared with existing expression systems. mRNA translation is regulated in response to abiotic stresses such as drought, elevated temperature, and high salinity (6–8). Most mRNA species are translationally repressed under stress conditions, except for a subset of mRNA species. Similarly, mRNA translation is regulated through growth and development processes, with a majority of mRNA species showing translational repression in the developed plant (9). It is difficult to remove all abiotic stresses from cultured cells and plants, meaning that there is a risk of decreased transgene expression by translational repression. Moreover, large developed tissue is mainly used for mass production, but there is still a risk that productivity of transgene is repressed by translational repression in the developed plants. We have successfully improved transgene expression by alleviating translational repression in a cultured cell system (10). Our next objective was to develop a stable high expression system that resists translational repression brought about by abiotic stresses, growth, and development *in planta*. In this study, we searched for candidate genes translated stably across growth stages, leaf development, heat stress, and salt stress using polysome fraction assay and microarray analysis data sets (6,9). In addition, the correct 5′UTR sequences were determined by cap analysis of gene expression (CAGE), and we found the 5′UTR sequence of cold-regulated 47 (*COR47*) gene to be a

* Corresponding author. Tel.: +81 743 72 5464; fax: +81 743 72 5469.
E-mail address: kou@bs.naist.jp (K. Kato).

1389-1723/$ – see front matter
http://dx.doi.org/10.1016/j.jbiosc.2017.08.007

consistent, effective translational enhancer. We confirmed the ability of this translational enhancer and resistance of translational repression in the *COR47* 5′UTR by reporter assays in transiently expressed cells and stable transformant plants. We demonstrate that *COR47* 5′UTR is a powerful tool to enable extremely high expression of the transgenes, independent from translational repression due to the growth, development, and stress.

MATERIALS AND METHODS

Plant materials, culture conditions, and growth conditions *Arabidopsis thaliana* T87 suspension cell (11) was cultured in modified Murashige-Skoog medium with constant agitation at 80 rpm at 22 °C with a 16 h light/8 h dark photoperiod. Eight ml of suspension cultures were transferred into 95 ml of fresh medium per week. To evaluate the effect of heat stress in the suspension cell, 3 day cultured cells were incubated at 22 °C and 37 °C for 10 min, then collected as control sample in suspension cell, and heat stress sample in suspension cell, respectively.

The seeds of *A. thaliana* ecotype Columbia-0 and its transformants (described below) were surface sterilized, sown on germination medium plates, and kept at 4 °C in the dark for 3 days for vernalization. The seeds were placed in growth chambers under long-day conditions (16 h light/8 h dark) at 22 °C. To evaluate the effect of growth stage on mRNA translation, whole plants grown 2 days after germination (DAG) and above ground parts plants aged 21DAG were collected as 2DAG and 21DAG samples, respectively. To evaluate the effect of leaf development, three younger leaves (expanding leaves) and three older, non-cotyledonous leaves (expanded leaves) were collected from 21DAG plants as young leaf and mature leaf samples, respectively. To evaluate the effect of heat stress, whole 4DAG plants incubated at 22 °C and 40 °C in the dark for 1 h were collected as control and heat-stressed samples, respectively. To evaluate the effect of salt stress, seeds were grown on 108 μm open nylon mesh on germination medium plates, with whole 4DAG plants being incubated without NaCl and with 600 mM NaCl for 3 h, corresponding to control and salt-stressed samples, respectively.

Nicotiana benthamiana seeds were surface sterilized, sown on soil, and kept at 4 °C in the dark for 3 days for vernalization. The seeds were grown in the green house under long-day conditions (16 h light/8 h dark) at 25 °C. The 14DAG plants were transplanted 1 seedling/pot. Moderately expanded leaves (5th leaf from the youngest) from 40DAG plants were used for transient expression assay.

Polysome/microarray data analysis We used two previously reported microarray data sets that evaluated the polysome association using polysome analysis. The microarray data for *A. thaliana* T87 suspension cell under stress conditions reported by Matsuura et al. (6) is available in the ArrayExpress database at the European Bioinformatics Institute (EBI) under ID E-DORD-69 (http://www.ebi.ac.uk/arrayexpress/experiments/E-DORD-69/). The microarray data for *A. thaliana* at different growth stages and leaf developmental stages reported by Yamasaki et al. (9) is available in the public Gene Expression Omnibus (GEO) database in the National Center for Biotechnology Information (NCBI) with accession number GSE69070 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69070). The polysome association was evaluated as polysome score (PS) or polysome ratio (PR) values defined by the following formula Eqs. 1 and 2.

$$\mathrm{PS} = \text{accumulated level of polysomal RNA/non-polysomal RNA level} \quad (1)$$

$$\mathrm{PR} = \text{accumulated level of polysomal RNA/whole RNA level} \quad (2)$$

PS values were transformed to PR values using the following formula Eqs. 3 and 4.

$$\mathrm{PR} = \text{accumulated level of polysomal RNA/(polysomal RNA and non-polysomal RNA level)} \quad (3)$$

$$\mathrm{PR} = \mathrm{PS}/(1 + \mathrm{PS}) \quad (4)$$

In addition, PR values were normalized by the following formula Eq. 5 for ranking polysome association among various conditions.

$$\text{Normalized PR} = (\mathrm{PR} - \text{mean of PRs})/\text{standard deviation of PRs} \quad (5)$$

CAGE and data analysis Frozen samples were pulverized with mortar and pestle. Total RNA was isolated from the frozen powder using TRIzol Reagent (Thermo Fisher Scientific, Waltham, MA, USA) followed by purification using RNeasy mini spin columns (Qiagen, Hilden, Germany) with on-column DNase I treatment. Eluted RNA solution was further subjected to precipitation by the addition of LiCl at a final concentration of 1.5 M, followed by centrifugation. We prepared total RNA in two independent biological replicates. Afterwards, nAnT-iCAGE libraries were generated from purified total RNAs using the method previously described by Murata et al. (12). Sequencing with 50-base single read was carried out on an Illumina HiSeq 2500. We obtained 6–26 million raw reads per sample. The MOIRAI compact workflow system for CAGE analysis (13) was used for quality control and mapping. Using the MOIRAI workflow, we performed additional quality control to select tags containing G nucleotide on the 5′end prior to mapping, because correct CAGE tags have the G nucleotide derived from cap structure. Subsequently, the added G nucleotide of the cap was removed from tags, then tags were mapped on the genome. Mapped tags were counted at the 5′end position of the genome as transcription start site (TSS). In addition, we removed tags that were not shown to be reproducible between two replicates in certain TSSs (existing in only one side of replicates). Tag counts were converted to tag per million (TPM) values, and averaged between two replicates. Finally, expressed TSSs within the range of 500 bp upstream of the defined TSS in the reference genome to start codon (ATG) were annotated the gene. To obtain more reliable data, we used limited genes that expressed more than 10 TPM in both replicates. Genome information from TAIR10 (http://www.arabidopsis.org/) was used as a reference for rRNA tag removal, mapping, and annotation. CAGE tags are available in the DDBJ Sequence Read Archive (DRA) database (https://trace.ddbj.nig.ac.jp/DRASearch/) with accession numbers DRR029684, DRR029686, DRR029688, DRR029690, DRR093534–DRR093539.

Polysome fractionation analysis, RNA isolation from sucrose gradients, and quantitative RT-PCR analysis Polysome fractionation analysis, RNA isolation from sucrose gradients, and quantitative RT-PCR (qRT-PCR) analysis were performed according to previously described methods by Yamasaki et al. (9). Cell extracts were layered on a 26.25–71.25% sucrose density gradient and centrifuged. After centrifugation, the gradients were collected into eight fractions using a piston gradient fractionator (BioComp Instruments, Fredericton, NB, Canada), with simultaneous recording of absorbance profiles at 254 nm using a UV monitor. Total RNA was purified from each fraction using an RNeasy kit (Qiagen) with on-column DNase I treatment, and purified RNAs from sucrose gradients were subjected to qRT-PCR using a LightCycler 480 (Roche Applied Science, Penzberg, Upper Bavaria, Germany). We used an external control to correct differences of efficiencies of RNA isolation and RT-PCR among separated eight fractions. The gene-specific primer sets used are summarized in Supplementary Table S1.

Plasmid construction for transduction of plasmid DNA by polyethylene glycol For the transient expression assay in protoplasts by polyethylene glycol (PEG), plasmids harboring the test 5′UTR and firefly luciferase (*F-Luc*) gene under control of the cauliflower mosaic virus 35S promoter (p35S) were constructed. p35S::*F-Luc*::heat shock protein 18.2 terminator (HSPT) sequence was first excised from plasmid Fluc-HSPT (14) by HindIII and EcoRI digestion, then inserted into HindIII-EcoRI gaps of pBluescript II KS (+) (Agilent Technologies, Santa Clara, CA, USA). In plasmid Fluc-HSPT, the fourth Ala (gcc) of F-Luc protein was changed to Val (gtc) in order to introduce AatII site into the coding region of *F-Luc*. In addition, we generated ClaI site upstream of TSS in p35S (ttTCATTTggagagaaca substituted with ttATCGATggagagaaca), yielding plasmid pBlue-p35S-FL-HSPT. The test 5′UTRs were amplified by PCR using gene-specific primers that contain a part of the modified 35S promoter with ClaI site or a part of *F-Luc* coding region with AatII site (Table S2). Each PCR fragment was introduced into plasmid pBlue-p35S-FL-HSPT at the ClaI/AatII sites to generate the plasmids pBlue-p35S-test-5′UTR-FL-HSPT. These plasmids were designed to initiate transcription immediately upstream of a test 5′UTR under the control of a modified 35S promoter (TTCATTATCGATGGAGAGAA::test 5′UTR sequence). The p35S::Renilla luciferase (*R-Luc*)::HSPT sequence was excised from plasmid CaMV35S-Rluc-HSPT (15) by HindIII and EcoRI digestion and inserted into HindIII-EcoRI gaps of pBluescript II KS (+) to generate the plasmid pBlue-p35S-RL-HSPT. Inserted fragments were verified by sequencing.

Plasmid construction for transduction of *in vitro* synthesized RNA by PEG A reporter plasmid for transient expression assay by PEG-mediated protoplast transformation using *in vitro* synthesized RNA was constructed by modifying plasmid pFL-pA (16). The test 5′UTRs were amplified by PCR using gene-specific primers that included a part of the T3 promoter with NcoI site and a part of *F-luc* coding region with AatII site (Table S3). Each PCR fragment was introduced into pFL-pA at the NcoI/AatII sites to generate the plasmids pT3-5′UTR-FL-pA. Inserted DNA fragments were verified by sequencing.

Plasmid construction for transduction of plasmid DNA by agrobacterium For transient expression assay by agro-infiltration in leaves and transformation of plants, p35S::test-5′UTR::*F-Luc*::HSPT sequences were first excised from constructed plasmids pBlue-p35S-test-5′UTR-FL-HSPT by HindIII and EcoRI digestion, then inserted into HindIII-EcoRI gaps of pRI909 (Takara Bio, Kusatsu, Japan), yielding plasmids pRI-p35S-test-5′UTR-FL-HSPT. Similarly, plasmid pRI-p35S-RL-HSPT was generated from plasmid pBlue-p35S-RL-HSPT. In addition, we constructed the p35S::5′UTR of pBI vector:: nopaline synthase terminator (NOST) sequence by substitution of pBI121 vector (Takara Bio USA, Mountain View, CA, USA) from *GUS* to *F-Luc* sequence. The constructed sequence was excised by HindIII and EcoRI digestion, and inserted into HindIII-EcoRI gaps of pRI909, yielding plasmid pRI-p35S-pBI-5′UTR-FL-NOST.

Please cite this article in press as: Yamasaki, S., et al., *Arabidopsis thaliana* cold-regulated 47 gene 5′-untranslated region enables stable high-level expression of transgenes, J. Biosci. Bioeng., (2017), http://dx.doi.org/10.1016/j.jbiosc.2017.08.007