



OPEN ACCESS

Edited by:

Henrik Toft Simonsen,
Technical University of Denmark,
Denmark

Reviewed by:

Kashmir Singh,
Panjab University, India
Fabien Nogué,
INRA Centre Versailles-Grignon,
France

***Correspondence:**

Stefan A. Rensing
stefan.rensing@biologie.uni-
marburg.de

† Present address:

Lucas Schneider,
Institute for Transfusion Medicine
and Immunohematology, Johann
Wolfgang Goethe University, German
Red Cross Blood Donor Service,
Frankfurt, Germany
Aikaterini Symeonidi,
Institute for Research in Biomedicine
(IRB Barcelona), Barcelona, Spain
Kristian K. Ullrich,
Max Planck Institute for Evolutionary
Biology, Plön, Germany
Mareike Schallenberg-Rüdinger,
Institut für Zelluläre und Molekulare
Botanik, Abteilung Molekulare
Evolution, Universität Bonn, Bonn,
Germany

Specialty section:

This article was submitted to
Plant Biotechnology,
a section of the journal
Frontiers in Plant Science

Received: 19 August 2017

Accepted: 10 October 2017

Published: 31 October 2017

Citation:

Hiss M, Schneider L, Grosche C,
Barth MA, Neu C, Symeonidi A,
Ullrich KK, Perroud P-F,
Schallenberg-Rüdinger M and
Rensing SA (2017) Combination
of the Endogenous *lhcsr1* Promoter
and Codon Usage Optimization
Boosts Protein Expression
in the Moss *Physcomitrella patens*.
Front. Plant Sci. 8:1842.
doi: 10.3389/fpls.2017.01842

Combination of the Endogenous *lhcsr1* Promoter and Codon Usage Optimization Boosts Protein Expression in the Moss *Physcomitrella patens*

Manuel Hiss¹, Lucas Schneider^{1†}, Christopher Grosche¹, Melanie A. Barth¹,
Christina Neu¹, Aikaterini Symeonidi^{1†}, Kristian K. Ullrich^{1†}, Pierre-François Perroud¹,
Mareike Schallenberg-Rüdinger^{1†} and Stefan A. Rensing^{1,2*}

¹ Plant Cell Biology, Faculty of Biology, University of Marburg, Marburg, Germany, ² BIOS Centre for Biological Signaling Studies, University of Freiburg, Freiburg im Breisgau, Germany

The moss *Physcomitrella patens* is used both as an evo-devo model and biotechnological production system for metabolites and pharmaceuticals. Strong *in vivo* expression of genes of interest is important for production of recombinant proteins, e.g., selectable markers, fluorescent proteins, or enzymes. In this regard, the choice of the promoter sequence as well as codon usage optimization are two important inside factors to consider in order to obtain optimum protein accumulation level. To reliably quantify fluorescence, we transfected protoplasts with promoter:GFP fusion constructs and measured fluorescence intensity of living protoplasts in a plate reader system. We used the red fluorescent protein mCherry under 2x 35S promoter control as second reporter to normalize for different transfection efficiencies. We derived a novel endogenous promoter and compared deletion variants with exogenous promoters. We used different codon-adapted green fluorescent protein (GFP) genes to evaluate the influence of promoter choice and codon optimization on protein accumulation in *P. patens*, and show that the promoter of the gene of *P. patens* chlorophyll a/b binding protein *lhcsr1* drives expression of GFP in protoplasts significantly (more than twofold) better than the commonly used 2x 35S promoter or the rice *actin1* promoter. We identified a shortened 677 bp version of the *lhcsr1* promoter that retains full activity in protoplasts. The codon optimized GFP yields significantly (more than twofold) stronger fluorescence signals and thus demonstrates that adjusting codon usage in *P. patens* can increase expression strength. In combination, new promoter and codon optimized GFP conveyed sixfold increased fluorescence signal.

Keywords: *Physcomitrella patens*, codon usage, chlorophyll a/b binding protein, promoter, codon bias, green fluorescent protein (GFP), *lhcsr1*, fluorescence normalization

INTRODUCTION

The strength of protein expression can be influenced by many factors including outside factors such as culture conditions, or inside factors as codon usage or transcription/translation system (Ullrich et al., 2015). Both promoter sequence and coding sequence can be optimized to improve final protein accumulation. The constructs of the first stable mutant lines in *Physcomitrella patens* contained resistance cassettes controlled by the cauliflower mosaic virus (CaMV) 19S and 35S promoters and the *Rhizobium radobacter* (previously *Agrobacterium tumefaciens*) nopaline synthase gene (*nos*) promoter (Schaefer et al., 1991). As *P. patens* was further developed as a plant model, classical strong angiosperm promoters such as the *Oryza sativa* (rice) *actin1* promoter (McElroy et al., 1990) or the *Zea mays* (maize) *ubiquitin* promoter (Christensen et al., 1992) were used successfully to drive protein accumulation in moss (Bezanilla et al., 2003; Horstmann et al., 2004). Inducible expression systems have also been established in *P. patens*, such as the beta-estradiol inducible one (Kubo et al., 2013) or the induction by elevated temperature using a *Glycine max* (soybean) heat shock protein promoter (Saidi et al., 2005). In a previous study, the activity of several promoters was studied by transient *P. patens* protoplast transfection of promoter:luciferase fusion constructs; here, the *actin1* promoter showed 10 times the expression level of the single CaMV 35S promoter and 1.6 times the level of the 2x CaMV 35S promoter (Horstmann et al., 2004). The same study also included endogenous 5' sequences of the genes α 1,3-fucosyltransferase and β 1,2-xylosyltransferase (*fuc-t*, *xyl-t*) that were further characterized *via* deletion constructs, with the 5'-*fuc-t* showing almost double activity as compared to the single CaMV 35S promoter. To confer strong expression, other endogenous promoters were used, e.g., different *tubulin* (Jost et al., 2005) or *actin* promoters (Weise et al., 2006), all of which showed stronger expression than the CaMV 35S promoter, that was later shown to yield only mediocre expression in *P. patens*, especially in the dark (Saidi et al., 2009).

Chlorophyll a/b binding (CAB) proteins are part of the light harvesting complex (LHC) of photosynthetic eukaryotes. *Cab* promoter sequences have been used as strong endogenous and exogenous promoters, e.g., in the charophyte alga *Closterium peracerosum-strigosum-littorale* complex, where an endogenous promoter of a *cab* gene has been used to drive expression of fluorescence marker genes (Abe et al., 2008). The rice *cab1R* gene promoter was used for transient expression of β -glucuronidase (*gus*) in *Nicotiana tabacum*, *Z. mays* and *O. sativa* leaves (Luan and Bogorad, 1992). In addition to the common LHC gene set, a LHC-like protein called LHCSR (or Li818) is present in association with LHC in phylogenetically diverse algae as *Chlamydomonas reinhardtii* (Li et al., 2000) and *Ectocarpus siliculosus*, but not in seed plants (Koziol et al., 2007; Dittami et al., 2010). LHC-like protein expression is regulated by light and stress conditions. In *P. patens* two *lhcsr* gene copies have been identified (Gerotto et al., 2011). *lhcsr1* is induced by high light ($450 \mu\text{mol}/\text{m}^{-2} \text{ s}^{-2}$), whereas *lhcsr2* is expressed in low temperature and low light conditions (Gerotto et al., 2011). Together with the protein PSBS, the LHCSR proteins are

responsible for the non-photochemical quenching in *P. patens* (Alboresi et al., 2010). The mechanisms of photoprotection of *lhcsr1* *via* two dissipative states has recently been revealed (Kondo et al., 2017), as well as its modulation *via* zeaxanthin binding and low pH (Pinnola et al., 2017).

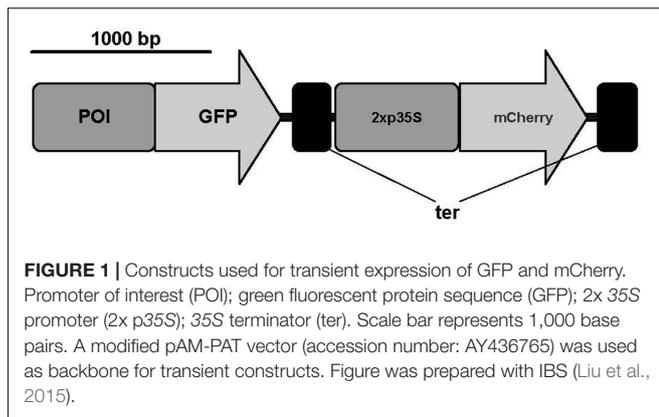
A second important inside factor influencing protein production is the codon usage in the RNA sequence (Quax et al., 2015). Due to the redundancy of the genetic triplet code, almost all amino acids are encoded by more than one codon. Which codon is used to which degree depends, e.g., on the species and on availability of tRNAs (Komar, 2016). Codon usage can be influenced by mutations and affects expression speed or accuracy. When trying to express a gene sequence from one species in another species, the codon usage often needs to be adjusted to fit the target species' codon frequencies. For example, the codon usage of the GFP, that is used in many organisms, e.g., to localize proteins by tagging, has been optimized for different organisms like *Saccharomyces cerevisiae* (yeast) (Cormack et al., 1997) or the alga *C. reinhardtii* (Fuhrmann et al., 1999), leading to stronger GFP signals. For plants a soluble modified GFP (smGFP) was created by site-directed mutagenesis that shows stronger GFP accumulation in *Arabidopsis thaliana* and therefore a stronger signal than the wild-type GFP (Castillo-Davis et al., 2002). Codon usage can vary not only between species but also within a species. In a subgroup of genes from one species, a bias for certain codons can be found, e.g., in highly expressed genes of *Caenorhabditis elegans*, *Drosophila melanogaster*, and *A. thaliana* codon usage differs from more weakly expressed genes of the same species (Duret and Mouchiroud, 1999). In *P. patens*, this codon usage bias seems to be driven by a combination of weak natural selection and the predominant mutational biases (Szovenyi et al., 2017).

Here we used the endogenous promoter sequence of *lhcsr1* to drive gene expression of the GFP in *P. patens* and compared it to the double CaMV 35S (2x p35S) and the rice *actin1* promoter (McElroy et al., 1990). In parallel, we designed two GFP versions with different codon usage and evaluated their GFP accumulation to the smGFP using a novel bimodal fluorescence readout system that allows to normalize the signal of interest in order to account for fluctuations in transfection efficiency. We find that the *lhcsr1* gene promoter increases signal intensity 1.7-fold as compared to the 2x 35S promoter. In combination with the codon-optimized GFP the signal even increases 5.7-fold as compared to the 2x 35S promoter.

RESULTS

Fluorescence Readout System

We designed a novel bimodal readout system that can be used to measure reporter protein fluorescence *in vivo*. We opted to use a microplate reader to measure fluorescence to allow high throughput measurement. For all fluorescence measurements, we used living protoplasts in transfection regeneration medium (Hohe et al., 2004). Initial tests showed that a minimum number of 2,000 protoplasts are necessary to get a signal above background fluorescence (Supplementary Figure 1).



The constructs used for transient transfection contained the promoter:GFP fusion and additionally a normalization cassette consisting of the 2x 35S promoter, the mCherry gene sequence and a 35S terminator (**Figure 1**). We selected the 2x 35S promoter because it is widely used in the *P. patens* community and shows a medium gene expression in *P. patens*, which makes it a good candidate for comparisons with new promoters. The fluorescence from this second reporter was used to normalize the GFP fluorescence values, which show high between-experiment variation due to different transformation efficiencies (Supplementary Figure 2 and Supplementary Table 1). By integrating 2xp35S:mCherry into the same plasmid as the GFP fusion of interest, we also account for uptake of multiple plasmids during transformation. To test the feasibility of our plate reader measurement system, we performed protoplastation of protonemal tissue from stable mutant lines that express GFP or mCherry (Perroud et al., 2011), mixed the protoplasts and found that we can measure both signals in parallel from the same well. Additionally, we prepared a dilution series and found a linear relationship between number of protoplasts and fluorescence intensity for both reporters (Supplementary Figure 3).

Highly Expressed Genes Show Codon Usage Bias in *P. patens*

We calculated the codon frequencies for all v1.6 genes of *P. patens* and grouped them by expression strength based on a broad range of transcriptome microarray experiments (Hiss et al., 2014). The top 233 highly expressed genes (0.9%) are enriched in the Gene Ontology (GO) terms translation, gene expression, and protein synthesis (Supplementary Figure 4). By comparing the codon usage of these highly expressed genes with the codon usage of all genes, we find a significant codon usage bias (Fisher's Exact Test, FDR adjusted $p < 0.05$) for nine amino acids, namely Glu, Phe, Tyr, Cys, His, Gln, Ile, Asn, and Lys (**Figure 2A** and Supplementary File 1). These codons are preferred not only in *P. patens* highly expressed genes but also in other organisms [**Figure 2B**; *Arabidopsis thaliana* (Wright et al., 2004; Morton and Wright, 2007), *Saccharomyces cerevisiae* (dos Reis and Wernisch, 2009), *Schizosaccharomyces pombe* (Hiraoka et al., 2009), and *Homo sapiens* (van Hemert and Berkhout, 1995)]. Using microarray data (Hiss et al., 2014), we can confidently

detect the first biased codon usage (Lys) already within the top 1,967 highly expressed genes (7.3% of all genes measured by the array), and the codon usage bias for all nine amino acids can be detected based on the top 291 highly expressed genes (1.1%, Supplementary File 2). Thus, by using the biased codon usage as an indicator, we detect around 7% of the *P. patens* genes as highly expressed. These genes also show high expression in microarray-based transcriptome studies [Supplementary Figure 5 (Hiss et al., 2014; Ortiz-Ramirez et al., 2015)]. For the highly expressed genes, we detect a GC bias toward higher GC content and a lower effective number of codons (ENC) as compared to the overall gene set (Supplementary Figure 6).

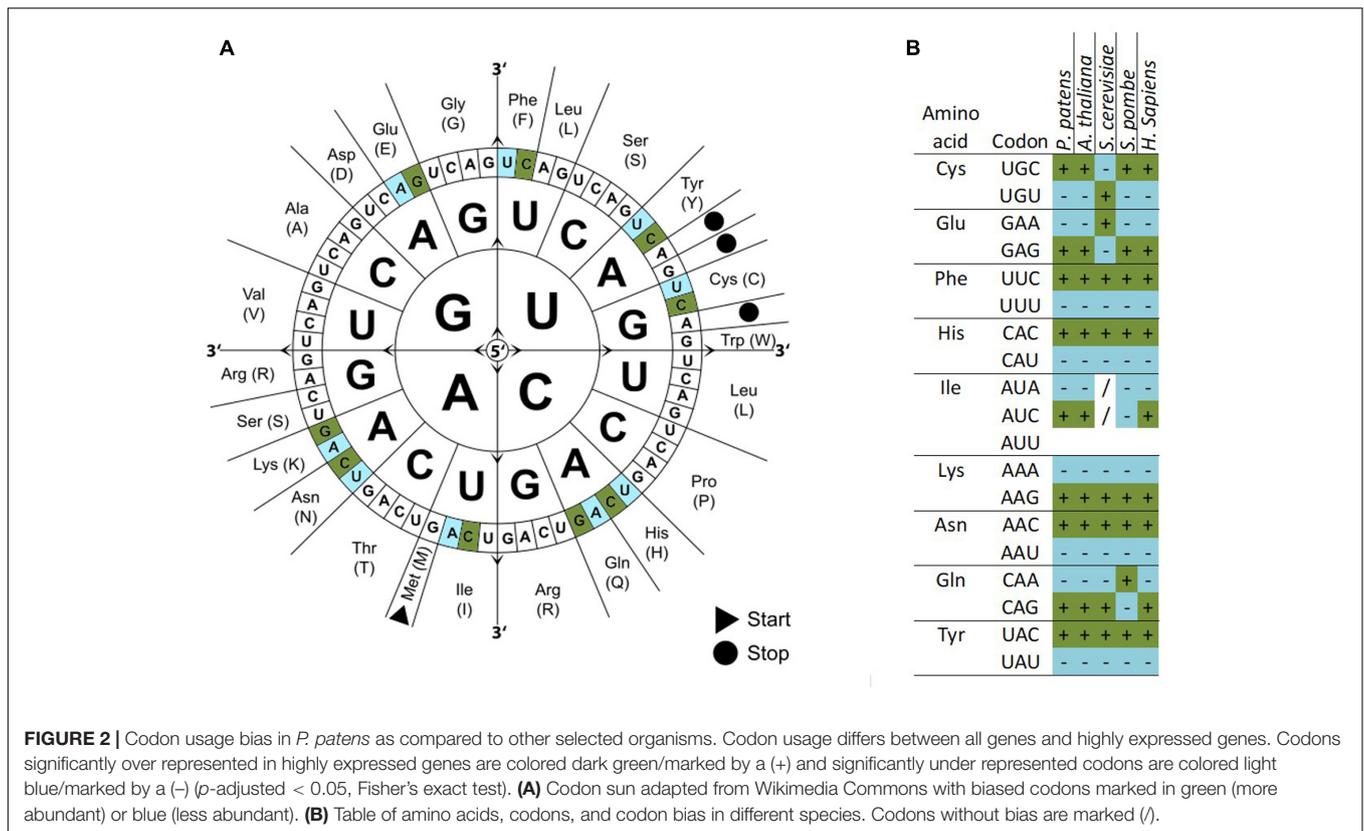
To determine whether the codon bias we have seen in highly expressed genes in *P. patens* can be used to enhance protein expression, we prepared GFP variants with adapted codon usage. Starting from the smGFP (Davis and Vierstra, 1998), we changed all triplets either to the one preferred in highly expressed genes, or to the one not preferred. This led to 39 changes in the coding sequence for “GFPhigh” and 58 changes for “GFPflow,” respectively (Supplementary Figure 7 and Supplementary Table 2). The GC content of the smGFP was 43.7% as compared to 49.1% (GFPhigh) and 35.6% (GFPflow). In both adapted GFP versions, the resulting amino acid sequence was not changed.

All three GFP versions were combined with different promoters 2xp35S, *pactin1*, and *plhcsr1* (1,956 bp upstream region of the *Physcomitrella patens lhcsr1* gene) and the mCherry normalization cassette. These constructs were transfected transiently into *P. patens* protoplasts and the fluorescence signal of GFP and mCherry measured *in vivo* with a microplate reader system.

The measurements were normalized and the 2x 35S promoter in combination with the smGFP signal set to one. The rice *actin1* promoter:GFPhigh fusion is at 1.0 ± 0.6 -fold in our system and therefore not significantly different from the 2x 35S promoter with the same GFP version ($p = 0.15$, Student's *t*-test). We see a shift in signal intensity for the 2x 35S promoter constructs with different GFP versions, with the GFPhigh giving the strongest signal at 1.8 ± 0.5 -fold and the GFPflow the weakest at 0.4 ± 0.3 -fold. Different signal intensities can also be seen with the different *plhcsr1*:GFP combinations starting with GFPflow at 0.9 ± 0.8 -fold to smGFP at 1.7 ± 0.3 -fold up to GFPhigh at 5.7 ± 0.4 -fold (**Figure 3**).

lhcsr1 Promoter Shows Strong Expression

We selected a strong and constitutively expressed gene from our transcriptome microarray data which cover several developmental stages and perturbations (Hiss et al., 2014). To find a suitable endogenous promoter we calculated the coefficient of variation (c. v.) of the expression values of all genes represented on the Combimatrix microarray (Wolf et al., 2010) and selected genes with high expression values and low coefficient of variation, leading to candidate genes with a high expression across many developmental stages and perturbations. Additionally we filtered for genes whose 3 kbp upstream region does not overlap with

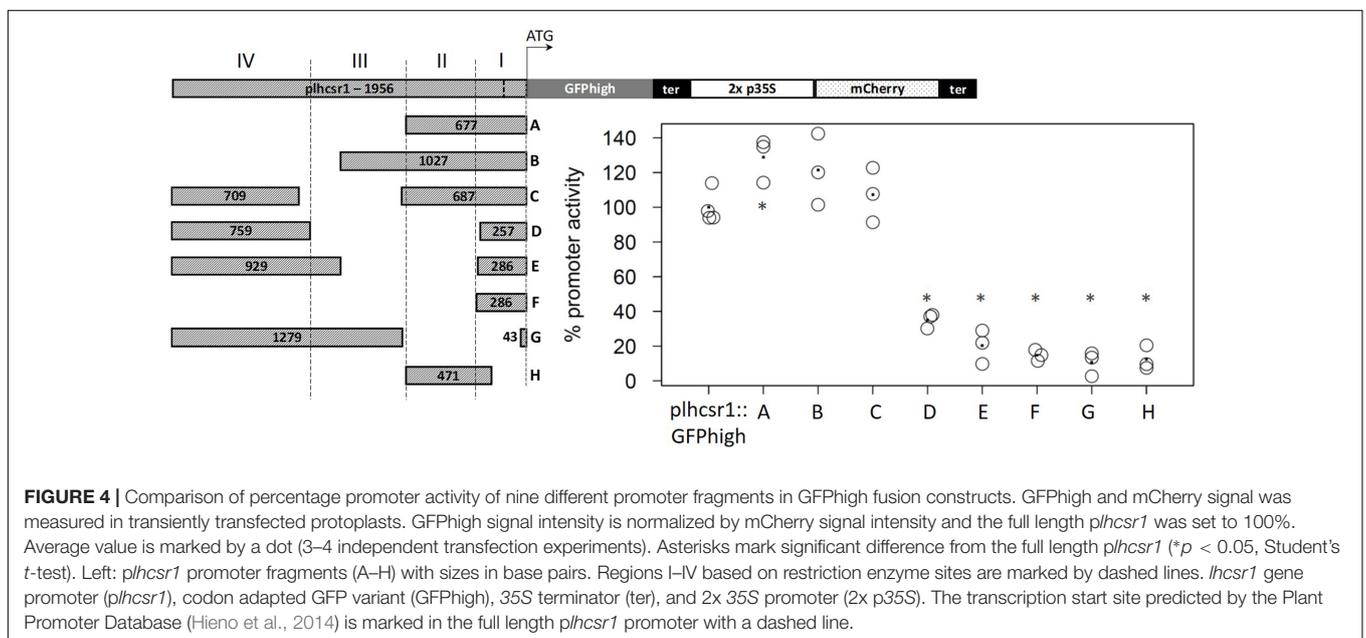
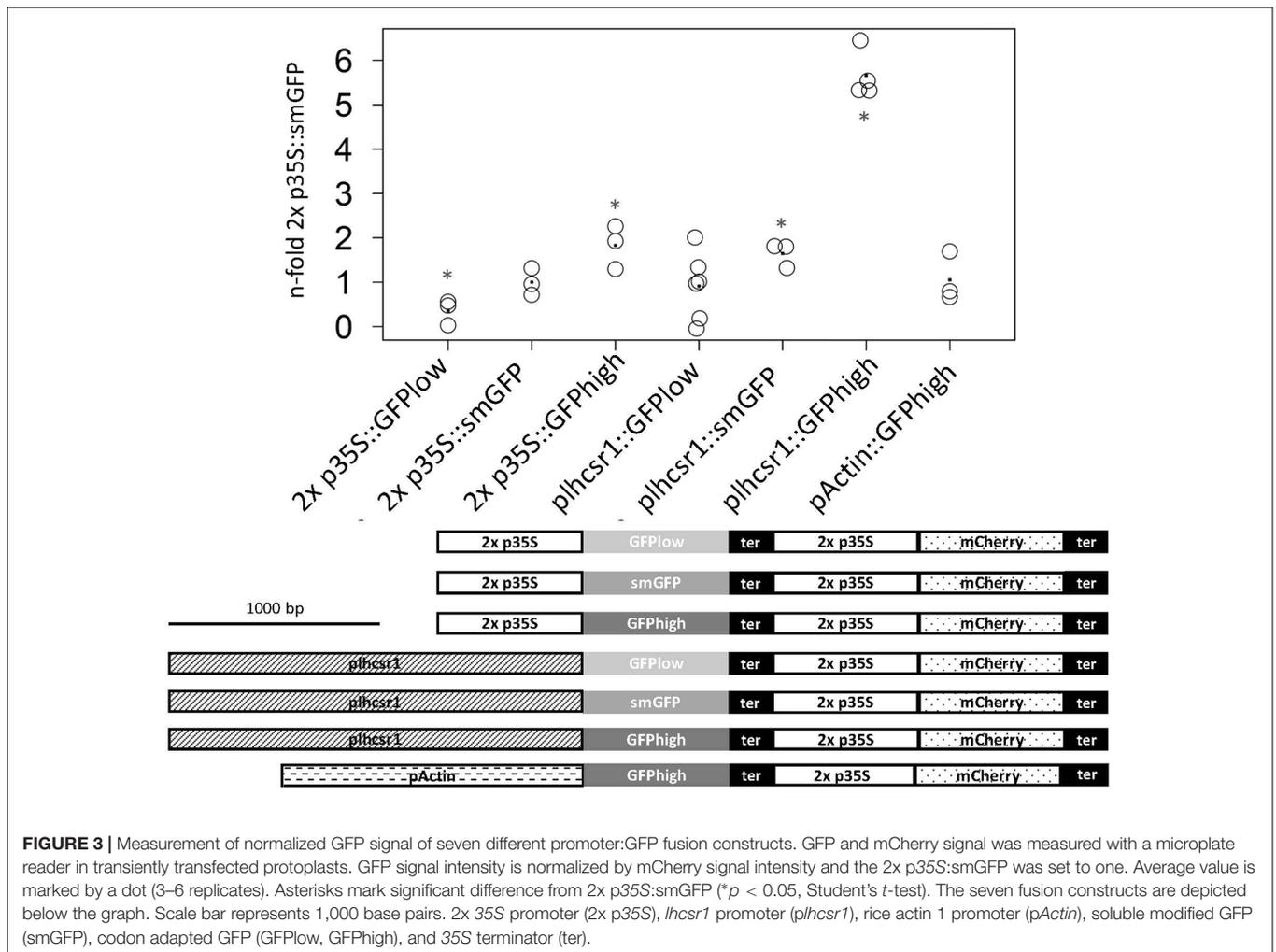


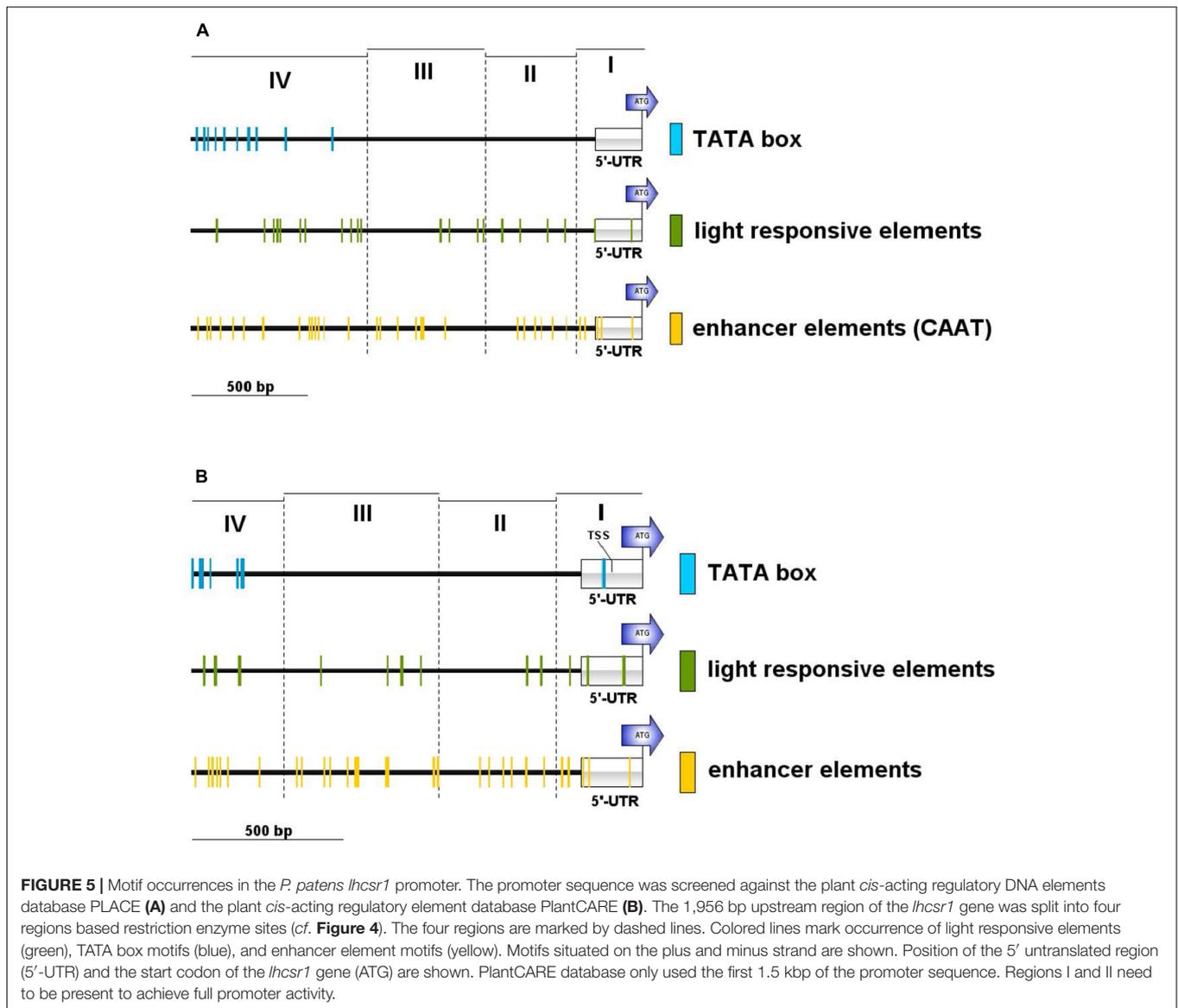
another gene or transposable element (TE). The 3 kbp gene-free upstream region was chosen in order to prevent the regulatory sequences of the gene of interest to fall into the sequence of another gene, and in order not to select atypically short genes (Zimmer et al., 2013). The 3kbp upstream region further should not contain TEs because they are known to be silenced *via* methylation of the corresponding genomic region (Zemach et al., 2010; Widiez et al., 2014). We chose the *cab* protein LHCSR1 (Phypa_169593, Pp1s213_80V6.1, Pp3c9_3440V3.1) since the promoters of *cab* genes of plants have successfully been used as strong endogenous and exogenous promoters before (Abe et al., 2008). We used about 2,000 bp upstream of the coding sequence (Chr09:1,975,316...1,977,272) as putative promoter for cloning and subsequent transfection (expression profile of the *lhcsr1* gene, see Supplementary Figure 8). The *lhcsr1* gene does not show the bias in codon usage we generally observe in the top highly expressed genes (Supplementary Table 3). The constructs containing the full *plhcsr1* promoter show higher signals than the corresponding 2x p35S constructs (Figure 3). We prepared eight shortened promoter sequences by restriction digestion of the original 1,956 bp sequence (Figure 4, constructs A–H). All promoter versions were fused to GFP_{high}:35S-terminator and the mCherry normalization cassette, transfected transiently into moss protoplasts and the GFP and mCherry signal measured *in vivo* with the microplate reader system.

Based on restriction enzyme sites, we divided the 1,956 bp promoter sequence into four regions (Figure 4). Region I contains the 202 bp 5'-UTR and 84 bp of the upstream region.

Region II consists of a 391-bp fragment, Region III of a 520-bp fragment, and Region IV of the 759 bp at the 5' end of the chosen promoter sequence. The removal of the regions IV and III without modifying the rest of the promoter does not strongly affect promoter activity, and the remaining fragments display an activity between 121 and 129% compared to the full length (1,956 bp) promoter (Figure 4, constructs A, B, and C). None of the other partial deletions abolish totally the promoter activity, most of the regions I and II deletions display 12–20% total *plhcsr1* value (Figure 4, constructs D–H). The only exception is the construct D that does not contain the activating region II, but still retains 35% promoter activity (Figure 4, construct D). The region IV appears to contain a domain able to activate transcription once it is fused to the 5'-UTR. This is especially interesting in contrast to construct E which has longer fragments for regions I and IV but a lower signal intensity.

We searched the potential promoter sequence for motifs using Signal Scan and the plant *cis*-acting regulatory DNA elements database [PLACE, (Higo et al., 1999)] as well as the Plant *cis*-acting regulatory element (PlantCARE) database. PLACE finds several known *cis*-acting elements that are associated with light-regulated genes, e.g., GT1-sites (Green et al., 1988) or I-boxes (Giuliano et al., 1988), whereas many TATA boxes are found in region IV, further away from the ATG but not in the 5'-UTR or the regions II and III (Figure 5A). PlantCARE predicts a TATA box core promoter element for both *lhcsr* genes within the 5'-UTR at around -130 bp of the translation start (Figure 5B). The transcription start site



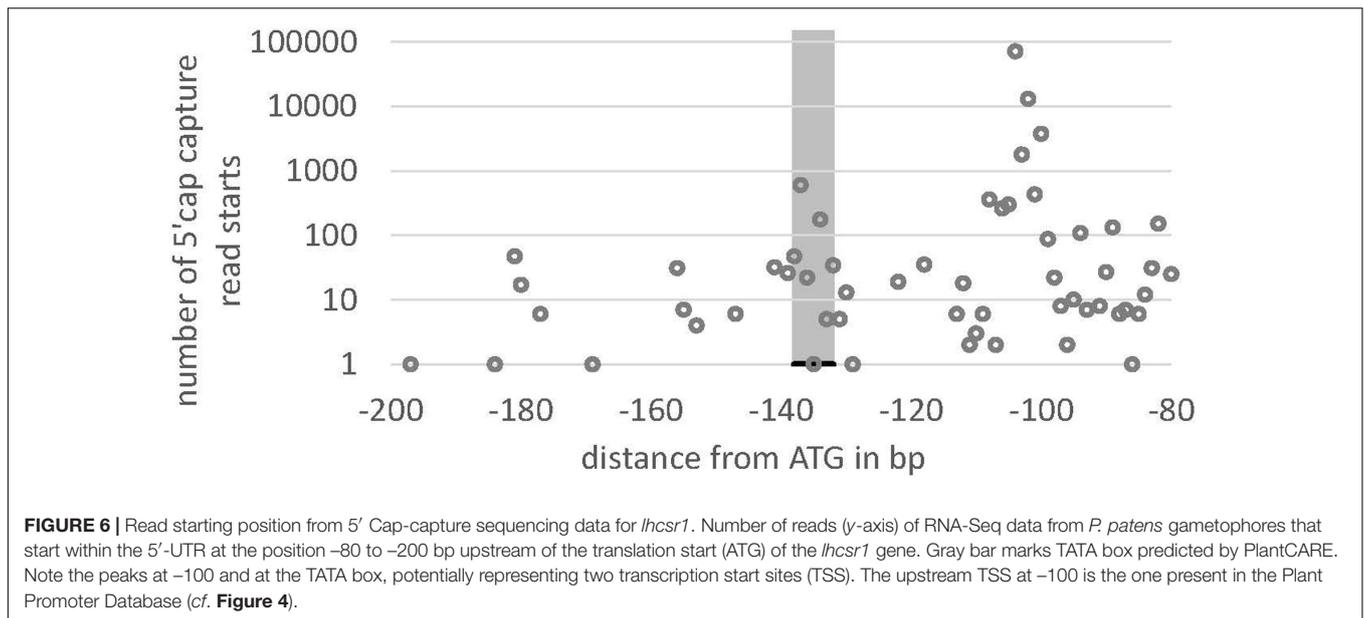


(TSS) data implemented in the Plant Promoter database (Hieno et al., 2014) support this TATA box based on 5' end sequencing of the *lhcsr1* gene with a TSS found 102 bases upstream of the start codon (Supplementary Figure 9). This TATA box is further supported by a possible TSS present in 5' Cap-capture sequencing data available at the *P. patens* CoGe representation (**Figure 6**).

DISCUSSION

Promoter:reporter gene constructs can be used to compare expression strength between different promoters. The signal can be read out after cell lysis (Horstmann et al., 2004) or directly from living cells (Thevenin et al., 2012). The readout system established in this study measures the fluorescence signal of transiently transfected protoplasts expressing the

reporter genes GFP and mCherry. We use a comparatively low-priced multi-well, multi-wavelength plate reader to measure the fluorescence signal of living protoplasts, while a recent study used a FACS system for measurements to investigate transcription factor activities in living cells (Thevenin et al., 2012). Since transformation efficiencies vary a lot between different transformations and to account for the uptake of varying numbers of plasmids into one cell (Supplementary Figure 2), we introduced a mCherry normalization cassette into all of our constructs. This novel readout system reduces the variation we see in our measurements and thereby helps to differentiate between different promoter strength. This system also allows to operate on lower numbers of protoplasts and biological replicates. We assume that detectable fluorescence correlated in linear fashion with protein amount, an assumption supported by the measurement of protoplasts of stable expressing lines (Supplementary Figure 3).



Codon usage of *P. patens* has been calculated and published based on expressed sequence tag (EST) data that were created before the genome sequence was published (Rensing et al., 2005; Stenoien, 2005) and recently based on the gene models predicted from genome sequencing (Szovenyi et al., 2017). In our analysis of codon usage bias based on v1.6 gene models in *P. patens*, we detect a bias in highly expressed genes to prefer certain codons. The same preferences can be seen in other organisms (Figure 2B). We can detect a bias for the first codons already within the top 7.3% expressed genes (Supplementary File 1). These highly expressed genes can be confirmed using microarray data (Supplementary Figure 5), and show a bias toward higher GC content and lower ENC (Supplementary Figure 6). In *C. reinhardtii*, codon usage, rather than GC content, was found as key determinant of gene expression efficiency. However, the nucleotide composition was found to feed back to the chromatin state of the promoter region (Barahimipour et al., 2015). To show that the detected codon bias, acting on translation, can be used for biotechnological applications, we created two GFP versions with adapted codon usage but identical amino acid sequence. Starting from the soluble-modified GFP version that gives high signals in particle bombardment transformation assays in *P. patens* (Cho et al., 1999), we exchanged codons to create a GFP that contains codons biased toward high protein expression (GFP high) and a GFP version that contains the opposite codons (GFP low). With the GFP version using preferred codons, we measure higher GFP signals, similar to studies in maize using optimal human codons (Chiu et al., 1996). The effect of our codon changes on protein expression can already be measured when using the 2x 35S promoter that gives medium expression strength in *P. patens*, and is more pronounced under the control of the stronger *lhcsr1* promoter (Figure 3).

Since the 2x 35S promoter is not the best choice for strong gene expression in biotechnological protein production in *P. patens*, we selected a candidate promoter based on microarray

analysis covering important tissues, several stress conditions, and hormone treatments. This includes typical production conditions like protonemal tissue and protoplasts in light. We selected the promoter of *lhcsr1*, a gene coding for a cab protein. LHCSR1 and LHCSR2 are both part of the light harvesting complex and are activated by light. Even though they are mainly expressed during conditions with photosynthetically active tissue, we see a medium RNA expression level under dark conditions in our microarray data (Supplementary Figure 8). This medium expression in darkness should allow the use of this promoter in selection cassettes where diurnal expression is preferred to allow for the upkeep of resistance. In contrast, the single 35S promoter was found to be inactive in darkness (Saidi et al., 2009). Compared to the 2x 35S and the *actin1* promoter we see a 2- to 6-fold higher expression when using the *lhcsr1* promoter, based on transcriptional activation. In our expression and readout system, the 2x 35S and *actin1* promoters do not show significant difference in expression strength, contrary to previous reports in which the *actin1* promoter reached 1.6 times the expression level of the 2x 35S promoter in a firefly luciferase reporter system (Horstmann et al., 2004). This difference could be due to the different normalization method that uses a second, separate plasmid as transfection control. This resulted in higher variation of the relative luciferase activity between different transfections with a median of coefficients of variation at 0.42 (Supplementary Table 4). Our GFP reporter system with the normalization cassette included into the same plasmid as the promoter of interest provides less variance with the median of coefficients of variation at 0.22 (Supplementary Table 5).

As putative *lhcsr1* promoter sequence, we selected the 1,956 bp upstream of the ATG of the *lhcsr1* gene. Since biotechnological applications will benefit from a shorter promoter sequence, we created shortened versions of the original sequence. We divided the original fragment into four regions based on restriction sites and were able to identify a promoter fragment of 677 bp that

shows full promoter activity. Shorter fragments and fragments that lack the region II (Figure 4) show a weaker promoter activity. Removing the 5'-UTR from the 677 bp fragment also results in a low promoter activity, suggesting a regulatory role for the 5'-UTR. Analyzing the promoter fragments for binding sites, PlantCARE detects a TATA box in the 5'-UTR as well as sites known to enhance RNA expression and several binding sites for light regulated transcription factors. With our fully active 677-bp promoter fragment showing at least twofold expression strength of the 2x 35S promoter, we provide a useful alternative for strong protein expression in *P. patens*. In combination with our codon optimized GFP_{high} the 2x 35S promoter shows a twofold expression strength increase as compared to the 2x 35S promoter with non-optimized GFP. However, the combination of *lhcsr1* promoter and GFP_{high} led to the highest observed activity with up to sixfold expression rate relative to the before mentioned standard.

Our study shows that a significant increase in protein production can be achieved by using suitable combinations of promoter and codon optimization, tackling transcription as well as translation efficiency. In contrast, combining even a strong promoter with genes not codon-optimized for high expression leads to low expression (Figure 3). These findings will be especially helpful in biotechnological and proteomics applications producing proteins in the moss *P. patens*, but also to drive, e.g., fluorescence tags and selectable markers. For these future applications, the generation of stable mutant lines with the expression cassettes integrated into the genome will be necessary to test whether the strong expression can also be seen in the genomic context. This will also allow to evaluate the expression strength in different tissues and conditions.

MATERIALS AND METHODS

Plant Material

Physcomitrella patens Gransden (Rensing et al., 2008) was cultivated on solidified [1% (w/v) agar] mineral medium, also known as modified (Reski and Abel, 1985) Knop's (1868) medium, on 9-cm petri dishes enclosed by laboratory film at 22°C with a 16-h-light/8-h-dark regime under 70 $\mu\text{mol m}^{-2} \text{s}^{-1}$ white light (long-day conditions).

Codon Usage

To assess codon usage bias, the coding sequences of v1.2 gene models on the Combimatrix microarray experiments from Hiss et al. (2014) that show expression values above the detection limit were analyzed (26,856 genes). Genes were sorted according to their expression level and genes with a normalized fluorescence intensity above 200,000 (567 genes, 2.1%) were termed strongly expressed genes whereas a value of 450,000 (233 genes, 0.87%) put them into the group of highly expressed genes. The codon frequencies for each group were calculated with *R* and afterward a Fisher's Exact Test was used for each of the 64 codons to find significant changes in codon frequencies between the groups. To account for the high number of statistical tests, multiple testing adjusted *p*-values were calculated with the *R*

function *p.adjust* and the significance level was set at adjusted $p < 0.05$.

The assessment at which expression level a codon usage bias can be detected was based on the coding sequences of v1.6 gene models with data based on the v1.2 Combimatrix microarray experiments. Custom software used to calculate codon usage and adjusted *p*-values can be found in the Github repository <https://github.com/kullrich/bio-scripts/tree/master/codonusage>.

Generation of Transient Constructs

Plasmids for transient transfection were assembled by Golden Gate Cloning using the *SapI* enzyme. A modified pAM-PAT vector (accession number: AY436765) was used for cloning of 2x p35S constructs. The multiple cloning site (MCS) was removed, and together with a chloramphenicol resistance gene and a *ccdb* kill cassette, two *SapI* restriction sites were introduced to create the vector pAM-PAT-*SapI*. To accommodate the insertion of other promoters, the pAM-PAT-*SapI* vector was modified by restriction with *XhoI* and *SallI*, removing the 2x 35S promoter. For cloning reactions promoter and reporter sequences were amplified with primers containing the *SapI* restriction site with matching overhang. *SapI*-cut vector, amplified promoter and amplified reporter were incubated with *SapI* and T4 ligase in ligase buffer and after 1 h incubation at room temperature transformed into TOP10 *E. coli* cells. The mCherry gene was amplified from a modified p123 vector (kindly provided by Michael Bölker) with *SapI* restriction sites as overhang (mCherry-*SapI*_fwd, mCherry-*SapI*_rev) and inserted into the pAM-PAT vector behind the 2x p35S sequence. The 2xp35S:mCherry fusion was amplified with *NotI* restriction sites as overhang (*NotI*-p35SS_fwd, *NotI*-mCherry) and ligated into the pAM-PAT vector *via* the *NotI* site. Although insertion was not directed by specific restriction sites the clones obtained always contained the 2x p35S:mCherry in the same direction as the promoter:GFP insert. To test whether a read-through from the *lhcsr1* promoter could lead to increased expression of mCherry, we removed the 2x 35S promoter in front of the mCherry. We could not detect a mCherry signal above background when using these constructs (Supplementary Table 6). Plasmid DNA was extracted by the Bibdo protocol (Birnbom and Doly, 1979) or with the NucleoBond Xtra Midi Kit (Macherey-Nagel, Germany).

Modified GFP versions were synthesized by Genent (Germany) and delivered in the pMA-T vector, inserted *via* the *SfiI* restriction site. The GFP versions were amplified with primers containing *SapI* restriction sites as overhang (GFP_high_for_SAP, GFP_high_rev, GFP_low_for, GFP_high_rev). During primer design, the second codon of the smGFP was changed from serine (S) to valine (V). To test whether this has an effect on expression, we compared the GFP/mCherry ratio of the 2x p35S:S-V_smGFP to a correct 2x p35S:smGFP. The plate reader measurements did not show a difference between the S and V versions of the smGFP (Supplementary Figure 10). 1,956 bp of genomic sequence upstream of the coding sequence for the cab protein (Phypa_169593) were amplified by PCR with primers containing the *SapI* restriction site as overhang (p169593_for, p169593_rev). Shortened versions of the *lhcsr1* gene promoter sequence were created by restriction digest and blunt/compatible

end ligation with *EcoRV* + *SalI* (A), *EcoRV* + *BclI* (B), *DraII* (C), *StuI* (D), *BclI* + *BglII* (E), *EcoRV* + *BglII* (F), and *SalI* + *XhoI* (G). The *EcoRV* + *SalI* construct without 5'-UTR (H) was amplified from the full length construct with *SapI* restriction sites as overhang (sh_pCAB_for, sh_pCAB_rev). The rice *actin1* (McElroy et al., 1990) promoter sequence was amplified by PCR from the PIG-AN vector (Schallenberg-Rudinger et al., 2017) with primers containing the *SapI* restriction site as overhang (pActin_Sap_for, pActin_Sap_rev).

Cloning success was tested by selection on ampicillin and either by colony PCR or test digestion. Positive candidate plasmids were sent for Sanger sequencing to GATC (Konstanz, Germany) or Macrogen (Amsterdam, Netherlands). The primers used for plasmid construction are shown in Supplementary Table 6.

Moss Protoplast Transfection

Transfection protocol was adapted from (Hohe et al., 2004). Regularly disrupted protonemal tissue in a 200 mL liquid culture, pH 5.8 was adjusted to 60 mg/L dry weight and transferred to 200 mL liquid Knop medium pH 4.5. After 5–6 days, the culture was harvested by sieving (100 μ m sieve). Protonemal tissue was equilibrated in 12 mL 0.51M Mannitol (pH 5.6–5.8) for 30 min, 4 mL Driselase solution (4%) was added and incubated for 1–2 h on a slowly tumbling shaker. The protoplast solution was sieved first on a 100 μ m, then on a 50 μ m sieve and afterward centrifuged 10 min at 50 g. Supernatant was removed, protoplasts resuspended in 10 mL 0.51 M Mannitol and centrifuged 10 min at 50 g. Supernatant was removed, protoplasts resuspended in 10 mL 0.51 M Mannitol and protoplast number counted on a Fuchs-Rosenthal counting chamber. Protoplast suspension was centrifuged for 10 min at 50 g, supernatant removed and a concentration of 1.2×10^6 protoplasts per mL adjusted with MMM medium [MMM medium, 0.51 M Mannitol, 15 mM $MgCl_2$, 0.1% w/v 2-(N-morpholino)ethanesulfonic acid, pH 5.6]. For transfection, 100 μ L DNA in 0.1 M $Ca(NO_3)_2$, 250 μ L protoplast suspension, and 350 μ L PEG solution (40% PEG 400 in MMM medium) were mixed during a 30 min incubation time. To slowly dilute the transfection solution, first 1 mL of MMM medium is added and mixed, next 2, 3, 4, and 5 mL are added and mixed every 5 min. Suspension was centrifuged for 10 min at 50 g and protoplasts resuspended in regeneration medium (0.28 M glucose and 0.24 M mannitol in Knop medium, pH 5.8). For transient transfections, circular plasmid was used. DNA amounts used for transient transfections were 10–50 μ g. After transfection, protoplasts were left to regenerate in 1 mL of regeneration medium.

Plate Reader Measurements

Fluorescence intensity measurements were performed in a FLUOstar microplate reader (BMG Labtech, Germany) with transfected *P. patens* protoplasts. Sample volumes of 100 μ L (up to 30,000 protoplasts) were placed into black 96-well microplates with transparent bottom (Greiner Bio-one, Austria). The samples were detected using the bottom optic, orbital averaging with 2 mm diameter and 15 flashes per well. For GFP and mCherry fluorescence, emission filters at 485 and

584 nm as well as excitation filters at 520 and 620 nm were used with 10 nm bandpass width. Regeneration medium was used as the blank value and non-transfected protoplasts as background control. Blank values were subtracted and the ratio of GFP and mCherry signals was calculated for each well to normalize for the transfection efficiency. Measurements were done 6–7 days after transfection since time course experiments found the highest fluorescence signal intensity after this time period (Supplementary Figure 11).

Gene Ontology (GO) Analyses and Visualization

The GO bias analyses used Fisher's Exact Test to calculate *p*-values. Multiple testing corrected (Benjamini and Hochberg, 1995) *q*-values were calculated in R with the function *p.adjust* (R Development Core Team, 2008). Word cloud visualizations were created using the online tool wordle¹. Word size is proportional to the $-\log_{10}(q\text{-value})$ and over-represented GO terms were colored dark green if $q \leq 0.0001$ and light green if $q > 0.0001$. Under-represented GO terms were colored dark red if $q \leq 0.0001$ and light red if $q > 0.0001$.

AUTHOR CONTRIBUTIONS

CG, CN, LS, MB, and MH prepared the constructs, transfected protoplasts, and analyzed data. AS and KU calculated codon frequencies. SR conceived of the work. MS-R, P-FP, and SR supervised the project. MH, MS-R, P-FP, and SR designed the experiments. MH and SR wrote the manuscript with contributions by all authors.

FUNDING

This work was supported by the German Federal Ministry of Education and Research (Freiburg Initiative for Systems Biology, 0313921 to SR).

ACKNOWLEDGMENTS

We want to thank the students Christian Peikert, Christian Volk, Stefan Ost, and Anne Genau who helped with cloning, transfection, and bioinformatics analysis. We want to thank Stefanie Pilz, Marco Göttig, and Faezeh Donges for their excellent technical assistance. We thank Michael Bölker for providing the p123 vector.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.01842/full#supplementary-material>

¹<http://www.wordle.net/>

REFERENCES

- Abe, J., Hiwatashi, Y., Ito, M., Hasebe, M., and Sekimoto, H. (2008). Expression of exogenous genes under the control of endogenous HSP70 and CAB promoters in the *Closterium peracerosum-strigosum-littorale* complex. *Plant Cell Physiol.* 49, 625–632.
- Alboresi, A., Gerotto, C., Giacometti, G. M., Bassi, R., and Morosinotto, T. (2010). *Physcomitrella patens* mutants affected on heat dissipation clarify the evolution of photoprotection mechanisms upon land colonization. *Proc. Natl. Acad. Sci. U.S.A.* 15, 11128–11133. doi: 10.1073/pnas.1002873107
- Barahimipour, R., Strenkert, D., Neupert, J., Schroda, M., Merchant, S. S., and Bock, R. (2015). Dissecting the contributions of GC content and codon usage to gene expression in the model alga *Chlamydomonas reinhardtii*. *Plant J.* 84, 704–717. doi: 10.1111/tpj.13033
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Methodol.* 57, 289–300.
- Bezanilla, M., Pan, A., and Quatrano, R. S. (2003). RNA interference in the moss *Physcomitrella patens*. *Plant Physiol.* 133, 470–474. doi: 10.1104/pp.103.024901
- Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 24, 1513–1523. doi: 10.1093/nar/7.6.1513
- Castillo-Davis, C. L., Mekhedov, S. L., Hartl, D. L., Koonin, E. V., and Kondrashov, F. A. (2002). Selection for short introns in highly expressed genes. *Nat. Genet.* 31, 415–418. doi: 10.1038/ng940
- Chiu, W., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H., and Sheen, J. (1996). Engineered GFP as a vital reporter in plants. *Curr. Biol.* 6, 325–330. doi: 10.1016/S0960-9822(02)00483-9
- Cho, S. H., Chung, Y. S., Cho, S. K., Rim, Y. W., and Shin, J. S. (1999). Particle bombardment mediated transformation and GFP expression in the moss *Physcomitrella patens*. *Mol. Cells* 28, 14–19.
- Christensen, A. H., Sharrock, R. A., and Quail, P. H. (1992). Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. *Plant Mol. Biol.* 18, 675–689. doi: 10.1007/BF00020010
- Cormack, B. P., Bertram, G., Egerton, M., Gow, N. A., Falkow, S., and Brown, A. J. (1997). Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in *Candida albicans*. *Microbiology* 143(Pt 2), 303–311. doi: 10.1099/00221287-143-2-303
- Davis, S. J., and Vierstra, R. D. (1998). Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. *Plant Mol. Biol.* 36, 521–528. doi: 10.1023/A:1005991617182
- Dittami, S. M., Michel, G., Collen, J., Boyen, C., and Tontonoz, T. (2010). Chlorophyll-binding proteins revisited—a multigenic family of light-harvesting and stress proteins from a brown algal perspective. *BMC Evol. Biol.* 10:365. doi: 10.1186/1471-2148-10-365
- dos Reis, M., and Wernisch, L. (2009). Estimating translational selection in eukaryotic genomes. *Mol. Biol. Evol.* 26, 451–461. doi: 10.1093/molbev/msn272
- Duret, L., and Mouchiroud, D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 13, 4482–4487. doi: 10.1073/pnas.96.8.4482
- Fuhrmann, M., Oertel, W., and Hegemann, P. (1999). A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*. *Plant J.* 19, 353–361. doi: 10.1046/j.1365-313X.1999.00526.x
- Gerotto, C., Alboresi, A., Giacometti, G. M., Bassi, R., and Morosinotto, T. (2011). Role of PSBS and LHCSR in *Physcomitrella patens* acclimation to high light and low temperature. *Plant Cell Environ.* 34, 922–932. doi: 10.1111/j.1365-3040.2011.02294.x
- Giuliano, G., Pichersky, E., Malik, V. S., Timko, M. P., Scolnik, P. A., and Cashmore, A. R. (1988). An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. *Proc. Natl. Acad. Sci. U.S.A.* 85, 7089–7093. doi: 10.1073/pnas.85.19.7089
- Green, P. J., Yong, M. H., Cuozzo, M., Kano-Murakami, Y., Silverstein, P., and Chua, N. H. (1988). Binding site requirements for pea nuclear protein factor GT-1 correlate with sequences required for light-dependent transcriptional activation of the *rbcS-3A* gene. *EMBO J.* 20, 4035–4044.
- Hieno, A., Naznin, H. A., Hyakumachi, M., Sakurai, T., Tokizawa, M., Koyama, H., et al. (2014). ppdb: plant promoter database version 3.0. *Nucleic Acids Res.* 42, D1188–D1192. doi: 10.1093/nar/gkt1027
- Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 27, 297–300. doi: 10.1093/nar/27.1.297
- Hiraoka, Y., Kawamata, K., Haraguchi, T., and Chikashige, Y. (2009). Codon usage bias is correlated with gene expression levels in the fission yeast *Schizosaccharomyces pombe*. *Genes Cells* 14, 499–509. doi: 10.1111/j.1365-2443.2009.01284.x
- Hiss, M., Laule, O., Meskauskiene, R. M., Arif, M. A., Decker, E. L., Erxleben, A., et al. (2014). Large-scale gene expression profiling data for the model moss *Physcomitrella patens* aid understanding of developmental progression, culture and stress conditions. *Plant J.* 79, 530–539. doi: 10.1111/tpj.12572
- Hohe, A., Egener, T., Lucht, J. M., Holtorf, H., Reinhard, C., Schween, G., et al. (2004). An improved and highly standardised transformation procedure allows efficient production of single and multiple targeted gene-knockouts in a moss, *Physcomitrella patens*. *Curr. Genet.* 44, 339–347. doi: 10.1007/s00294-003-0458-4
- Horstmann, V., Huether, C. M., Jost, W., Reski, R., and Decker, E. L. (2004). Quantitative promoter analysis in *Physcomitrella patens*: a set of plant vectors activating gene expression within three orders of magnitude. *BMC Biotechnol.* 4:13. doi: 10.1186/1472-6750-4-13
- Jost, W., Link, S., Horstmann, V., Decker, E. L., Reski, R., and Gorr, G. (2005). Isolation and characterisation of three moss-derived beta-tubulin promoters suitable for recombinant expression. *Curr. Genet.* 47, 111–120. doi: 10.1007/s00294-004-0555-z
- Knop, W. (1868). *Der Kreislauf des Stoffs: Lehrbuch der Agricultur-Chemie*. Leipzig: H. Haessel.
- Komar, A. A. (2016). The Yin and Yang of codon usage. *Hum. Mol. Genet.* 25, R77–R85.
- Kondo, T., Pinnola, A., Chen, W. J., Dall'Osto, L., Bassi, R., and Schla-Cohen, G. S. (2017). Single-molecule spectroscopy of LHCSR1 protein dynamics identifies two distinct states responsible for multi-timescale photosynthetic photoprotection. *Nat. Chem.* 9, 772–778. doi: 10.1038/nchem.2818
- Kozioł, A. G., Borza, T., Ishida, K., Keeling, P., Lee, R. W., and Durnford, D. G. (2007). Tracing the evolution of the light-harvesting antennae in chlorophyll a/b-containing organisms. *Plant Physiol.* 143, 1802–1816. doi: 10.1104/pp.106.092536
- Kubo, M., Imai, A., Nishiyama, T., Ishikawa, M., Sato, Y., Kurata, T., et al. (2013). System for stable beta-estradiol-inducible gene expression in the moss *Physcomitrella patens*. *PLOS ONE* 8:e77356. doi: 10.1371/journal.pone.0077356
- Li, X. P., Bjorkman, O., Shih, C., Grossman, A. R., Rosenquist, M., Jansson, S., et al. (2000). A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* 27, 391–395. doi: 10.1038/35000131
- Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., et al. (2015). IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* 15, 3359–3361. doi: 10.1093/bioinformatics/btv362
- Luan, S., and Bogorad, L. (1992). A rice cab gene promoter contains separate cis-acting elements that regulate expression in dicot and monocot plants. *Plant Cell* 4, 971–981. doi: 10.1105/tpc.4.8.971
- McElroy, D., Zhang, W., Cao, J., and Wu, R. (1990). Isolation of an efficient actin promoter for use in rice transformation. *Plant Cell* 2, 163–171. doi: 10.1105/tpc.2.2.163
- Morton, B. R., and Wright, S. I. (2007). Selective constraints on codon usage of nuclear genes from *Arabidopsis thaliana*. *Mol. Biol. Evol.* 24, 122–129. doi: 10.1093/molbev/msl139
- Ortiz-Ramirez, C., Hernandez-Coronado, M., Thamm, A., Catarino, B., Wang, M., Dolan, L., et al. (2015). A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants. *Mol. Plant* 11, 205–220. doi: 10.1016/j.molp.2015.12.002
- Perroud, P. F., Cove, D. J., Quatrano, R. S., and McDaniel, S. F. (2011). An experimental method to facilitate the identification of hybrid sporophytes in the moss *Physcomitrella patens* using fluorescent tagged lines. *New Phytol.* 191, 301–306. doi: 10.1111/j.1469-8137.2011.03668.x
- Pinnola, A., Ballottari, M., Bargigia, I., Alcocer, M., D'Andrea, C., Cerullo, G., et al. (2017). Functional modulation of LHCSR1 protein from *Physcomitrella patens*

- by zeaxanthin binding and low pH. *Sci. Rep.* 11, 11158. doi: 10.1038/s41598-017-11101-7
- Quax, T. E., Claassens, N. J., Soll, D., and van der Oost, J. (2015). Codon bias as a means to fine-tune gene expression. *Mol. Cell* 16, 149–161. doi: 10.1016/j.molcel.2015.05.035
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: The R Foundation for Statistical Computing.
- Rensing, S. A., Fritzwosky, D., Lang, D., and Reski, R. (2005). Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss, *Physcomitrella patens*. *BMC Genomics* 6:43.
- Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319, 64–69. doi: 10.1126/science.1150646
- Reski, R., and Abel, W. O. (1985). Induction of budding on chloronemata and caulonemata of the moss, *physcomitrella-patens*, using isopentenyladenine. *Planta* 165, 354–358. doi: 10.1007/BF00392232
- Saidi, Y., Finka, A., Chakhporanian, M., Zryd, J. P., Schaefer, D. G., and Goloubinoff, P. (2005). Controlled expression of recombinant proteins in *Physcomitrella patens* by a conditional heat-shock promoter: a tool for plant research and biotechnology. *Plant Mol. Biol.* 59, 697–711. doi: 10.1007/s11103-005-0889-z
- Saidi, Y., Schaefer, D. G., Goloubinoff, P., Zryd, J. P., and Finka, A. (2009). The CaMV 35S promoter has a weak expression activity in dark grown tissues of moss *Physcomitrella patens*. *Plant Signal. Behav.* 4, 457–459. doi: 10.4161/psb.4.5.8541
- Schaefer, D., Zryd, J. P., Knight, C. D., and Cove, D. J. (1991). Stable transformation of the moss *Physcomitrella patens*. *Mol. Gen. Genet.* 226, 418–424. doi: 10.1007/BF00260654
- Schallenberg-Rüdinger, M., Oldenkott, B., Hiss, M., Trinh, P. L., Knoop, V., and Rensing, S. A. (2017). A single-target mitochondrial RNA editing factor of *Funaria hygrometrica* can fully reconstitute RNA Editing at two sites in *Physcomitrella patens*. *Plant Cell Physiol.* 01, 496–507. doi: 10.1093/pcp/pcw229
- Stenoien, H. K. (2005). Adaptive basis of codon usage in the haploid moss *Physcomitrella patens*. *Heredity* 94, 87–93. doi: 10.1038/sj.hdy.6800547
- Szovenyi, P., Ullrich, K. K., Rensing, S. A., Lang, D., van Gessel, N., Stenoien, H. K., et al. (2017). Selfing in haploid plants and efficacy of selection: codon usage bias in the model moss *Physcomitrella patens*. *Genome Biol. Evol.* 9, 1528–1546.
- Thevenin, J., Dubos, C., Xu, W., Le Gourrierrec, J., Kelemen, Z., Charlot, F., et al. (2012). A new system for fast and quantitative analysis of heterologous gene expression in plants. *New Phytol.* 193, 504–512. doi: 10.1111/j.1469-8137.2011.03936.x
- Ullrich, K. K., Hiss, M., and Rensing, S. A. (2015). Means to optimize protein expression in transgenic plants. *Curr. Opin. Biotechnol.* 32, 61–67. doi: 10.1016/j.copbio.2014.11.011
- van Hemert, F. J., and Berkhout, B. (1995). The tendency of lentiviral open reading frames to become A-rich: constraints imposed by viral genome organization and cellular tRNA availability. *J. Mol. Evol.* 41, 132–140. doi: 10.1007/BF00170664
- Weise, A., Rodriguez-Franco, M., Timm, B., Hermann, M., Link, S., Jost, W., et al. (2006). Use of *Physcomitrella patens* actin 5' regions for high transgene expression: importance of 5' introns. *Appl. Microbiol. Biotechnol.* 70, 337–345. doi: 10.1007/s00253-005-0087-6
- Widiez, T., Symeonidi, A., Luo, C., Lam, E., Lawton, M., and Rensing, S. A. (2014). The chromatin landscape of the moss *Physcomitrella patens* and its dynamics during development and drought stress. *Plant J.* 79, 67–81. doi: 10.1111/tpl.12542
- Wolf, L., Rizzini, L., Stracke, R., Ulm, R., and Rensing, S. A. (2010). The molecular and physiological responses of *Physcomitrella patens* to ultraviolet-B radiation. *Plant Physiol.* 153, 1123–1134. doi: 10.1104/pp.110.154658
- Wright, S. I., Yau, C. B., Looseley, M., and Meyers, B. C. (2004). Effects of gene expression on molecular evolution in *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Mol. Biol. Evol.* 21, 1719–1726. doi: 10.1093/molbev/msh191
- Zemach, A., McDaniel, I. E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 14, 916–919. doi: 10.1126/science.1186366
- Zimmer, A. D., Lang, D., Buchta, K., Rombauts, S., Nishiyama, T., Hasebe, M., et al. (2013). Reannotation and extended community resources for the genome of the non-seed plant *Physcomitrella patens* provide insights into the evolution of plant gene structures and functions. *BMC Genomics* 14:498. doi: 10.1186/1471-2164-14-498

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Hiss, Schneider, Grosche, Barth, Neu, Symeonidi, Ullrich, Perroud, Schallenberg-Rüdinger and Rensing. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.