Published online 21 June 2018

Nucleic Acids Research, 2018, Vol. 46, No. 13 ***6909–6919***
doi: 10.1093/nar/gky532

Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga *Chlamydomonas reinhardtii*

Thomas Baier, Julian Wichmann, Olaf Kruse* and Kyle J. Lauersen

Bielefeld University, Faculty of Biology, Center for Biotechnology (CeBiTec), Universitätsstrasse 27, 33615 Bielefeld, Germany

Received March 14, 2018; Revised May 16, 2018; Editorial Decision May 17, 2018; Accepted June 08, 2018

ABSTRACT

Among green freshwater microalgae, *Chlamydomonas reinhardtii* has the most comprehensive and developed molecular toolkit, however, advanced genetic and metabolic engineering driven from the nuclear genome is generally hindered by inherently low transgene expression levels. Progressive strain development and synthetic promoters have improved the capacity of transgene expression; however, the responsible regulatory mechanisms are still not fully understood. Here, we elucidate the sequence specific dynamics of native regulatory element insertion into nuclear transgenes. Systematic insertions of the first intron of the ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit 2 (rbcS2i1) throughout codon-optimized coding sequences (CDS) generates optimized algal transgenes which express reliably in *C. reinhardtii*. The optimal rbcS2i1 insertion site for efficient splicing was systematically determined and improved gene expression rates were shown using a codon-optimized sesquiterpene synthase CDS. Sequential insertions of rbcS2i1 were found to have a step-wise additive effect on all levels of transgene expression, which is likely correlated to a synergy of transcriptional machinery recruitment and mimicking the short average exon lengths natively found in the *C. reinhardtii* genome. We further demonstrate the value of this optimization with five representative transgene examples and provide guidelines for the design of any desired sequence with this strategy.

INTRODUCTION

Photosynthetic microalgae hold potential as sustainable green microbial cell factories and are natively sources of many interesting bio-products with a range of potential biotechnological applications. To meet the demands for economical use as industrial feedstocks, genetic engineering has been proposed as imperative in order to generate novel high-value traits and sophisticated bio-products. Successful and reproducible genetic engineering of microalgae is currently established in a small cohort of hosts (1–5), however, mainstream application has been hampered by low transgene expression rates from these organisms.

The green freshwater microalga *Chlamydomonas reinhardtii* has been used as a model organism for photosynthesis, phototaxis and flagella studies for over 60 years (6,7) and currently has the most comprehensive and developed molecular tools of any alga. Although genetic manipulation is well established in *C. reinhardtii*, the capacity for robust and reliable transgene expression from the nuclear genome, and consequently successful genetic engineering, is limited by its characteristic low transgene expression levels (8–11). Several attempts have been made to overcome this limitation by: cell line improvement (12,13), design of synthetic promoters (14), and transgene codon optimization (15,16). However, most reports of robust genetic engineering in *C. reinhardtii* involve the expression of only reporter proteins with relatively short coding sequences (CDS) (15,17). In addition, the random gene integration of foreign DNA into the nuclear genome of *C. reinhardtii* results in 'position effects' on the transgene expression causing highly variable transgene expression levels depending on the integration site (11,16,18). Previous efforts have employed direct fusion of a gene of interest (GOI) to reporters or selection markers to facilitate high-throughput screening of transformants in order to identify outliers with sufficient transgene expression (17,19). To date, our knowledge regarding the regulation mechanisms of green algal genomes and the factors responsible for (poor) transgene expression are still limited. Understanding and overcoming these limitations will be crucial to enable advanced green algal synthetic biology and metabolic engineering strategies.

*To whom correspondence should be addressed. Tel: +49 521 106 12258; Fax: +49 521 106 12290; Email: olaf.kruse@uni-bielefeld.de
Present address: Bielefeld University, Faculty of Biology, Center for Biotechnology (CeBiTec), Universitätsstrasse 27, 33615 Bielefeld, Germany.

The nuclear genome of *C. reinhardtii* has an elevated GC content of ∼64% (68% in coding regions) and a narrow codon bias (20). Indeed, the most commonly used aminoglycoside-(3′)-phosphotransferase selection markers (*aphVII, aphVIII*) were derived from prokaryotes with naturally high GC contents (21,22) and codon optimization of desired transgenes is now standard practice for this host (15). Although a very recent report indicated that further transgene optimization holds the potential to increase the overall expression rates (16), it has not typically been considered for nuclear transgene expression experiments in *C. reinhardtii*. Relative to other unicellular eukaryotes and land plants, the nuclear genome of *C. reinhardtii* contains a high percentage of endogenous introns: 88% of all genes contain introns, with 7.3 exons per gene, and an average exon length of 240 nucleotides (*Chlamydomonas* genome 4.0, (20)). Related Chlorophyceae such as *Chlorella variabilis*, *Volvox carteri*, and *Monoraphidium neglectum* exhibit average intron densities of 6.1, 6.3 and 4.0 respectively (23,24). Other microalgae, such as *Nannochloropsis gaditana* and *Phaeodactylum tricornutum* contain average intron densities of 1.7 and 0.8 (25,26). Recently, the average splicing motifs of *C. reinhardtii* were analyzed and reported to be comparable to the canonical eukaryotic consensus sequence of MAG/GT…intron…AG/G (16,27). Intensive processing of mRNA is an important part of the gene expression machinery of *C. reinhardtii* and has been proposed to act as an potential 'immune system' against viral infections and transposable elements (16). In addition, maturation of mRNA including splicing can have a favorable effect on the transgene abundance by either efficient nuclear export or transcript stability which also leads to overall higher expression rates in *C. reinhardtii* (28–30). Most sequences derived by codon optimization for heterologous expression are cDNAs longer than the native average exon length. Such constructs do not match the host genome regulatory structures, which likely contributes to the poor transgene expression levels characteristically observed in this unique host.

The most studied *C. reinhardtii* intron sequences are derived from the rbcS2 gene locus (NCBI: X04472.1). When integrated into a codon-optimized phleomycin resistance gene *sh*ble (18) or a *Renilla reniformis* luciferase (*cr*luc) sequence (31), rbcS2 introns were shown to positively impact the respective gene expression. The first intron (rbcS2i1, 145 bp) was subsequently integrated as a permanent addition to several promoter sequences (14,21,32), was shown to positively affect the efficiency of selection markers (18,22,31,33), and served as an artificial framework for an engineered Cas9 guide RNA (34). Although improved transgene expression mediated by the addition of these introns has been known for 20 years (18), their artificial use in heterologous codon-optimized sequences has not yet become systematic or standard practice. Recently, we applied a sequence optimization strategy wherein we repetitively spread the rbcS2i1 throughout the *Pogostemon cablin* patchoulol (*Pc*Ps) and *Abies grandis* (*E*)-α-bisabolene (*Ag*Bs) synthases which enabled robust expression from the nuclear genome of *C. reinhardtii* (35,36). In addition to enabling the first examples of heterologous terpenoid production from this microalga, synthetic repetitive fusion proteins could be expressed of both sequences up to 223.9 and 249.9 kDa, respectively. These were the largest heterologous proteins produced in this alga to date, indicating the power of endogenous intron addition into codon-optimized CDSs to enable transgene expression in the green microalgal host.

In this work, we demonstrate novel insights into the effective splicing mechanism of artificially inserted introns in eukaryotic transgene sequences that enable robust transgene expression from the nuclear genome of *C. reinhardtii*. We systematically analyzed the effect of the rbcS2i1 at different nucleotide insertion positions within coding sequences to identify the optimal insertion motif, rbcS2i1 was found to effectively splice out of codon optimized genes and also affect transgene expression at the levels of transcription as well as relative protein accumulation. Exon size and rbcS2i1 insertion frequency were determined using the 1,662 bp codon-optimized template cDNA of the *Pc*Ps and intron insertion design was validated as a general strategy using five other heterologous sequences.

MATERIALS AND METHODS

Design, cloning, and transformation of gene expression cassettes

The *C. reinhardtii* rbcS2 intron 1 sequence (145 bp, hereafter rbcS2i1, NCBI: X04472.1) was used to determine appropriate nucleotide integration sites for intron positioning. The rbcS2i1 sequence was PCR amplified (Q5® High-Fidelity DNA Polymerase, NEB) including overhangs composed of additional codons at the exon/intron boundary (GC**G**, GC**C**, GC**A** or GC**T**) and intron/exon boundary (**C**TG, **G**GC, **A**TC or **T**TC) along with restriction enzyme recognition sites. Resulting PCR products were separated in 2% (w/v) agarose gels and purified using the peqGOLD Gel Extraction Kit (VWR). Cloning was performed at the N-terminus of the *sh*ble (NCBI: MG052655) resistance gene of the pOptimized vector system (19) using FastDigest restriction enzymes (Thermo Scientific) and the Rapid DNA Dephos & Ligation Kit (Roche) following manufacturer's instructions.

The *Pogostemon cablin* Benth patchoulol synthase amino acid sequence (*Pc*Ps, UniProt: Q49SP3) was used to chemically synthesize (Genscript) an optimized intron-containing algal transgene using the most frequent codons for the nuclear genome of *C. reinhardtii* and the rbcS2i1 sequence artificially introduced into suitable insertion points every ∼500 bp from the start codon (NCBI: KX097887, (35)). From the initial expression vector (pOpt_*Pc*Ps_YFP_Paro, vector iii, (35)), the rbcS2i1 copies were removed stepwise by PCR amplification of previously designed exon elements, adding complementary overhangs to the neighboring exon sequence (Supplementary Figure S1 and Supplementary Table S1 for primers used in this study). Complementary *Pc*Ps exons were assembled into intron-reduced sequences using overlap extension (oe)PCR (37). All resultant *Pc*Ps genes were cloned in frame with the mVenus (YFP) reporter in the pOpt_mVenus_Paro (KM061060.1, (19)) vector as described above.

The actin intron 3 sequence (95 bp, actini3, NCBI: D50838.1) was assembled using two 60 bp long complementary oligonucleotides (Supplementary Table S1) and inserted scar-less into previously defined intron positions

of the *Pc*Ps using complementary overhangs and oePCR. The PsaD promoter sequence (38) was amplified from the pChlamy3 plasmid and cloned between XbaI and NdeI restriction endonuclease recognition sites to replace the HSP70/RBCS2i1 promoter sequence in the vectors A–D. All plasmids were used for heat-shock transformation of chemically competent *Escherichia coli* DH5a cells followed by selection on 300 mg l^{-1} ampicillin containing LB-agar plates. *E. coli* colonies were checked by colony PCR and plasmids were isolated from overnight cultures using the peqGOLD Plasmid Miniprep Kit I (VWR). All sequences were confirmed by Sanger sequencing (Sequencing Core Facility, CeBiTec, Bielefeld University).

C. reinhardtii cultivation, transformation, and mutant screening

Cultivation of *C. reinhardtii* strain UVM4 (12) was conducted under mixotrophic conditions with Tris acetate phosphate (TAP) medium (39) on agar plates or liquid in shake flasks with 150–200 μmol photons m^{-2} s^{-1} light intensity.

Nuclear transformation was carried out by glass beads agitation as previously described (40) using linearized plasmid DNA (Supplementary Figure S2). Transformants were selected on paromomycin or zeocin (10 mg l^{-1}) containing TAP agar plates for 5–7 days. To determine the effect of the rbcS2i1 nucleotide insertion site on *sh*ble expression, obtained colonies were counted and normalized to the molar DNA amount used during transformation: 10 μg linearized total DNA equates to 2.44 pmol for the intron-less control *sh*ble gene of interest cassette and 2.38 pmol for the intron-containing constructs (Supplementary Table S2).

Mutants expressing *Pc*Ps_mVenus fusions were screened by fluorescence microscopy as previously described (35). For each construct, 20 expressing transformants were isolated and cultivated individually in microtiter plates. Prior to further analysis, cultures were pooled normalized according to their respective cell density.

Extraction and quantification of mRNA by RTqPCR

C. reinhardtii cells were harvested in mid-logarithmic growth phase and total RNA was extracted by acid guanidinium thiocyanate-phenol-chloroform method (41) followed by a DNase treatment (RQ1 RNase-Free DNase I, Promega). Samples of 100 ng RNA were subjected to reverse transcription and qPCR amplification using the Hi-ROX SensiFAST™ SYBR One-Step Kit (Bioline). All Primers were designed to amplify intron-spanning parts of the transcripts: the mVenus transcript was amplified including the C. reinhardtii rbcS2 small subunit 2 intron 2 (rbcS2i2), while the 18S rRNA served as a housekeeping gene (mVenus: for 5′-TGCAGGAGCGCACCATCT-3′ and reverse 5′-GGCCCAGGATGTTGCCGTC-3′; 18S: for 5′-ACCTGGTTGATCCTGCCAG-3′ and reverse 5′-TGATCCTTCCGCAGGTTCAC-3′, (42)). SYBR Green fluorescence was recorded by a StepOnePlus™ Real-Time PCR System (Thermo Scientific) and relative mRNA expression levels were determined according to the 2(-Delta Delta C(T)) method (43). Mean relative mRNA abundance was determined from technical triplicates and error bars indicate the standard deviation.

Actinomycin-D treatment

To test the stability of mRNA generated from each expression construct, early logarithmic cultures were treated with 100 μg ml^{-1} actinomycin-D (Caymen Chemicals) and RNA samples were taken in 15 min time intervals for 1.5 h. RNA extraction and quantification was performed as described above and time-related mRNA amounts were plotted for vectors C, D and I, normalized to their respective amount at t0.

Protein extraction and SDS-PAGE

Cells were resuspended in 2× protein sample buffer (60 mM Tris pH 6.8, 4% (w/v) SDS, 20 % (v/v) glycerol, 0,01% (w/v) bromophenol blue) and proteins were separated by Tris-glycine-SDS-PAGE using 12%-PA-Gels (44). Separated proteins were stained using colloidal Coomassie Brilliant Blue G-250 (45) or after subjection to Western blotting on nitrocellulose membranes (Amersham, GE Healthcare) were analyzed by immunodetection using a HRP-linked rabbit-anti-GFP antibody (1:5000, in 1xTBS containing 5% (w/v) BSA and milk powder, Thermo Scientific, A10260) and Pierce™ ECL Western Blotting substrates (Thermo Scientific). Protein quantifications were performed by Lowry DC-Protein Assay (Bio-Rad, CA, USA) using 0.1–1.5 mg ml^{-1} BSA as a reference. A recombinantly expressed mVenus protein (19) was purified from *E. coli* cells by StrepII®-Tag affinity chromatography (Iba Life Science) and served as a protein standard.

Flow cytometry

Single cell fluorescence measurements were performed using a Gallios™ Flow Cytometer (Beckman Coulter) with excitation by a Blue Solid State Diode (488 nm, 22 mW) and FL1 Detector position with a Band Pass Filter of 550 nm. Acquisition was performed for 100 000 counts for each sample and mean fluorescence per construct was determined.

Terpenoid capture and productivity analysis

Two phase cultivation with a dodecane overlay and quantification of terpenoid productivity was performed by GC–MS measurements as previously described (35).

RESULTS AND DISCUSSION

As a member of the Chlorophyceae, *C. reinhardtii* exhibits one of the most intron dense eukaryotic nuclear genomes, more so than higher plants (20). Therefore, long pre-mRNA transcription, recruitment of the spliceosome complex and mRNA processing must play a crucial role in transcriptional regulation in this organism. Although it has been previously shown that the addition of the rbcS2i1 into genetic constructs can improve expression (18,31), the reasons for this are still unclear. In this work, we investigated the effect of intron addition into codon-optimized transgenes on relative expression from the nuclear genome of *C. reinhardtii*.

We analyzed the intron insertion site, exon length, as well as intron frequency within recombinant sequences and determined an appropriate rule-set for their use in an optimized heterologous transgene design for nuclear expression.

Splicing efficiency of the rbcS2i1 insertion sequence

The ability of reliable mRNA splicing of an artificially introduced rbcS2i1 intron sequence was analyzed in modified versions of the *sh*ble gene, where the N-terminus was designed to contain alternating exon/intron/exon boundaries (Figure 1). Using this approach, functional splicing was indicated directly by the survival of transformants in the presence of zeocin as non- or incorrect splicing results in a frameshift and, consequently, loss of antibiotic resistance (Figure 1B). Transformation was conducted in two individual experiments and the mean transformation efficiency is shown relative to the intron-less control vector. This construct resulted in 681 ± 6 transformants pmol DNA^{-1} (Supplementary Table S2) whereas a frameshifted control lacked the ability to confer resistance completely. Due to this distinct phenotype, the number of regenerated colonies can be directly correlated to the frequency of correct mRNA processing. The 16 constructs generated reflect all possible combinations of the four nucleotides: guanine, adenine, thymine/uracil, and cytosine on either side of the exon boundary. The observed transformation efficiency for these constructs indicates diverse efficacy of splicing depending on the boundary site (Figure 1C). A maximum was observed when the introduced intron was located between two guanine nucleotides (896 ± 28 colonies pmol DNA^{-1}, 132% of control). The lowest efficiency was observed when the same intron was placed between cytosine nucleotides (21 ± 1 colonies pmol DNA^{-1}, 3% of control). In general, combinations of adenine and guanine (purines) resulted in similar transformation efficiency compared to the control (between 86 and 106% of control), whereas combinations of thymine/uracil and cytosine (pyrimidines) resulted in strongly reduced numbers of transformants (15–16% of control). Although the splicing frequency is reduced in some of the tested conditions, spliceosome binding at a specific sequence consensus seems to be somewhat flexible in *C. reinhardtii*, as colonies where found for all designed constructs. At the 5′ splice site, all four triplets used coded for the amino acid alanine, to minimize amino acid dependent changes of *sh*ble activity. Although the variant GCC is the most frequently used alanine codon, this sequence specifically induced lower splicing frequencies in all tested conditions of the 3′ boundary. It is very likely that the formation of antibiotic resistance here was dependent only on the splice sites and not influenced by codon frequency. Although the phenotype of frameshifts is absolute, resulting in no colonies, some colonies were observed even in unfavorable rbcS2i1 insertion site combinations. It remains unclear where splicing occurs in these constructs, or if the presence of introns prior to splicing affects pre-mRNA secondary structure as predicted for the lowest energy fold models (Supplementary Figure S3, (46)).

However, the clear differences in relative colony frequencies indicate that choosing a proper nucleotide insertion position for rbcS2i1 addition is vital to appropriate splic-

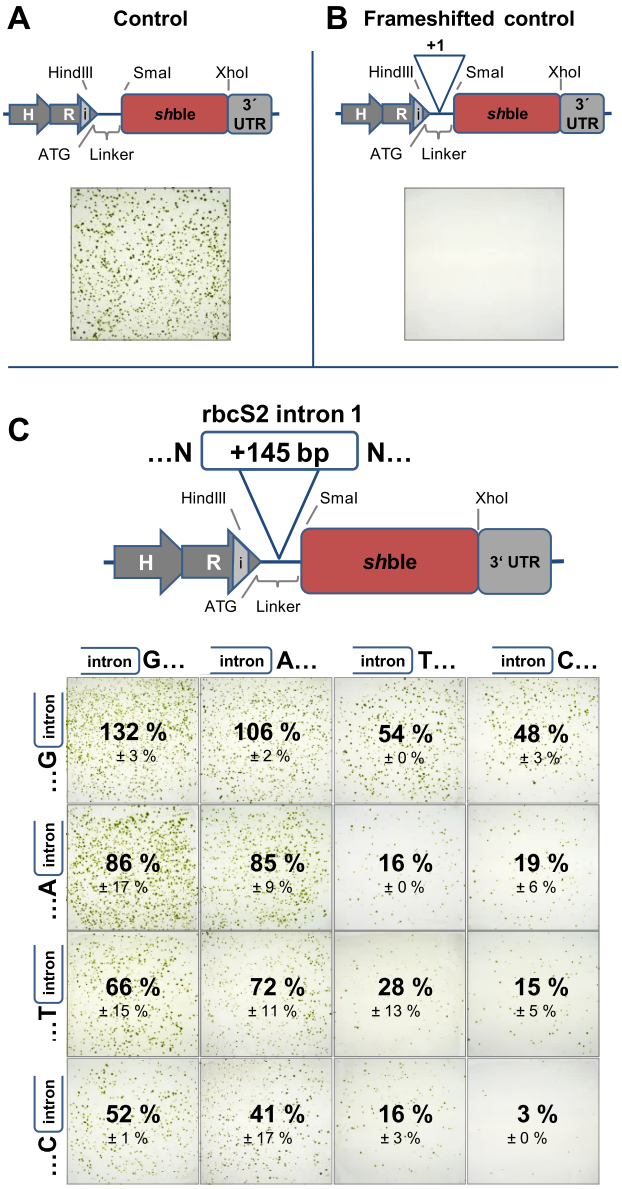

Figure 1. Gene design and splicing efficiency analysis of artificial rbcS2i1 insertion sites in the *sh*ble coding sequence represented by the transformation efficiency of *C. reinhardtii*. (**A**) Antibiotic selection cassette adopted from the pOptimized vector toolkit (19) containing the *sh*ble gene and an N-terminal linker which includes an additional SmaI restriction enzyme cleavage site and a representative TAP agar plate showing regenerated mutants in the presence of 10 mg L^{-1} zeocin. (**B**) A frameshifted selection cassette with a single additional nucleotide results in no viable transformants. (**C**) The *sh*ble selection cassette including an amplified rbcS2i1 (145 bp) with modified exon boundaries. In total 16 combinations of the four nucleotides guanine (G), adenine (A), thymine (T) and cytosine (C) surrounding the intron insertion site were tested for effective splicing by colony survival in the presence of zeocin. The given percentage reflects the mean number of obtained colonies per pmol DNA in relation to the intron-less control. Equal amounts of plasmid DNA were used and linearized prior to transformation (Supplementary Figure S2) H R i – the HSP70/RBCS2i1 promoter, 3′ UTR – 3′ untranslated region of the rbcS2 gene. *shble* – *Streptoalloteichus hindustanus* phleomycin resistance gene, N – denotes nucleotide positions modified in each construct.

ing and enhanced gene expression efficiency. The most favored position correlates with the consensus sequence known for other eukaryotes (27) as well as the recently reported *in silico* analysis of natural spliceosome binding motifs (16,47). The regulation of spliceosome activity in *C. reinhardtii* is not fully elucidated and although the native position of the rbcS2i1 is between the two nucleotides adenine and guanine, the optimal condition for splicing appears to be between two guanine nucleotides. We next analyzed the second neighboring nucleotides of border guanines at the splice site (NG/GN, Supplementary Figure S4) for their additional impact on splicing. Here, no correlative changes in the transformation efficiency were found, indicating that the nucleotides directly at the splice site likely contribute the most to spliceosome activity. Due to the elevated GC content of codon-optimized CDSs used for nuclear gene expression in *C. reinhardtii*, several insertion points (…NG/GN…) can be found in any given CDS. This enables multiple rbcS2i1 insertions throughout potential genes of interest without further modification.

In 14 of the 16 tested conditions, the transformation efficiency was negatively affected by intron addition and only when the rbcS2i1 was inserted in the most suitable insertion point, elevated gene expression was observed. In this experiment, the added rbsc2i1 was inserted only 30 bp downstream of the intron-containing rbcS2 promoter region. We did not observe that the vicinity of these two introns affected their splicing efficiency, since the loss of the intermediate exon region, including the ATG start codon, would induce a loss of protein function. Previously, the positive effect on transgene expression in the *sh*ble gene was reported with two rbcS2i1 copies separated by 169 bp where both copies were inserted between two guanines (18). To fully elucidate the potential for increased nuclear transgene expression, we designed a codon-optimized *Pc*Ps sequence and used it to correlate gene expression levels with intron insertion frequencies and positioning.

*Pc*Ps as a codon-optimized intron-containing algal transgene

In the 1,662 bp long codon-optimized sesquiterpenoid patchoulol synthase (*Pc*Ps) CDS, potential insertion points for artificial intron addition (…NG/GN…) were identified and in total three rbcs2i1 copies were spread *in silico* throughout the sequence. As previous reports had successfully inserted the rbcS2i1 into the nucleotide position between guanidines (18), this site was also used in the HSP70/RBCS2i1 promoter (9), and our investigations of nucleotide insertion position (Figure 1) indicated this was the optimal insertion motif, it was used in our previous construction of the *Pc*Ps intron-containing sequence (35). These copies were interspaced by ∼500 bp exon elements as a compromise for gene-synthesis capacities and to generate exon sizes not longer than twice that of the native average exon length of *C. reinhardtii* genes. Although these modifications led to a gain of construct complexity, increasing the gene length by +25% (435 bp) including repetitive nucleotide segments, the resulting amino acid sequence was not affected. The *Pc*Ps intron-containing optimized sequence was chemically synthesized and cloned into the pOptimized vector backbone to have a C-terminal mVenus (yellow fluorescence protein, YFP) tag for rapid, agar plate-level screening by fluorescence microscopy as previously described (19). A set of six vectors were created by using oePCR (Supplementary Figure S1) to remove individual intron sequences from the *Pc*Ps gene and resulting vectors were used to transform the microalgal host (Figure 2). All variants led to functional gene expression which was detectable by YFP fluorescence, however, at different levels of signal intensity. For each construct, 20 expressing mutants were isolated, individually cultivated in microtiter plates, and pooled prior to further analysis normalized to their respective cell densities (Figure 2A).

The expression levels for each construct were analyzed and compared for all stages of gene expression: relative abundance of mRNA, protein titers determined by immunodetection, and the protein activity reflected by YFP fluorescence as well as patchoulol production (Figure 2B and C, Supplementary Figure S5). An intron-less codon-optimized *Pc*Ps CDS (vector A) was used as a control, reflecting a conventional expression strategy for nuclear transgenes in *C. reinhardtii*. This approach has been previously reported to result in poor expression (3). Although this construct was codon-optimized and transcribed by the strong constitutive HSP70/RBCS2i1 promoter, only minimal expression could be detected also in this experiment. With the stepwise incorporation of the rbcS2i1, the expression level of the *Pc*Ps construct was consistently increased (vectors B and C). The highest expression was observed from the original *Pc*Ps intron-containing design (vector D), which had exons of 583 bp (exon 4) or smaller. We postulate that the intensified optimization step of rbcS2i1 spreading and mimicking exon lengths of native intron-rich genes positively utilizes the algal gene expression regulation machinery, allowing increased gene expression even for long and complex transgenes. Although minimizing exon length is crucial to effective gene expression, it is unclear whether proximity to the 5′ end of the transcript may affect the transcription regulation. Reducing the number of introns from the N-terminus (construct E and F) led to a reduction of expression, suggesting that exon lengths of ∼500 bp or shorter may be optimal as a guideline for ensuring effective transcription. There are variances between constructs containing similar exon length, but at different positions within the transcript (Figure 2, vector C and E). The limited transgene expression generally observed for sequences expressed from the *C. reinhardtii* nuclear genome appears to be due to a lack of transcript abundance, as a striking difference in mRNA amount relating to the intron abundance and exon length was obtained (>100-fold change of relative mRNA abundance comparing vector A and D, Figure 2B). However, the increase in protein content within the cell was more linear, reaching 8-fold greater fluorescence in vector D compared to the intron-less vector A (Figure 2B). It needs to be considered that this comparison is based on quantitative transcript abundance and relative fluorescence in a pooled population. It is likely that the native translation machinery can cope with larger amounts of synthesized transcripts and is not yet a limiting factor in robust gene expression, as an increased transcript level led to increased protein accumulation and higher patchoulol productivity (Figure 2C). This is supported by our recent report where the *Pc*Ps was over-

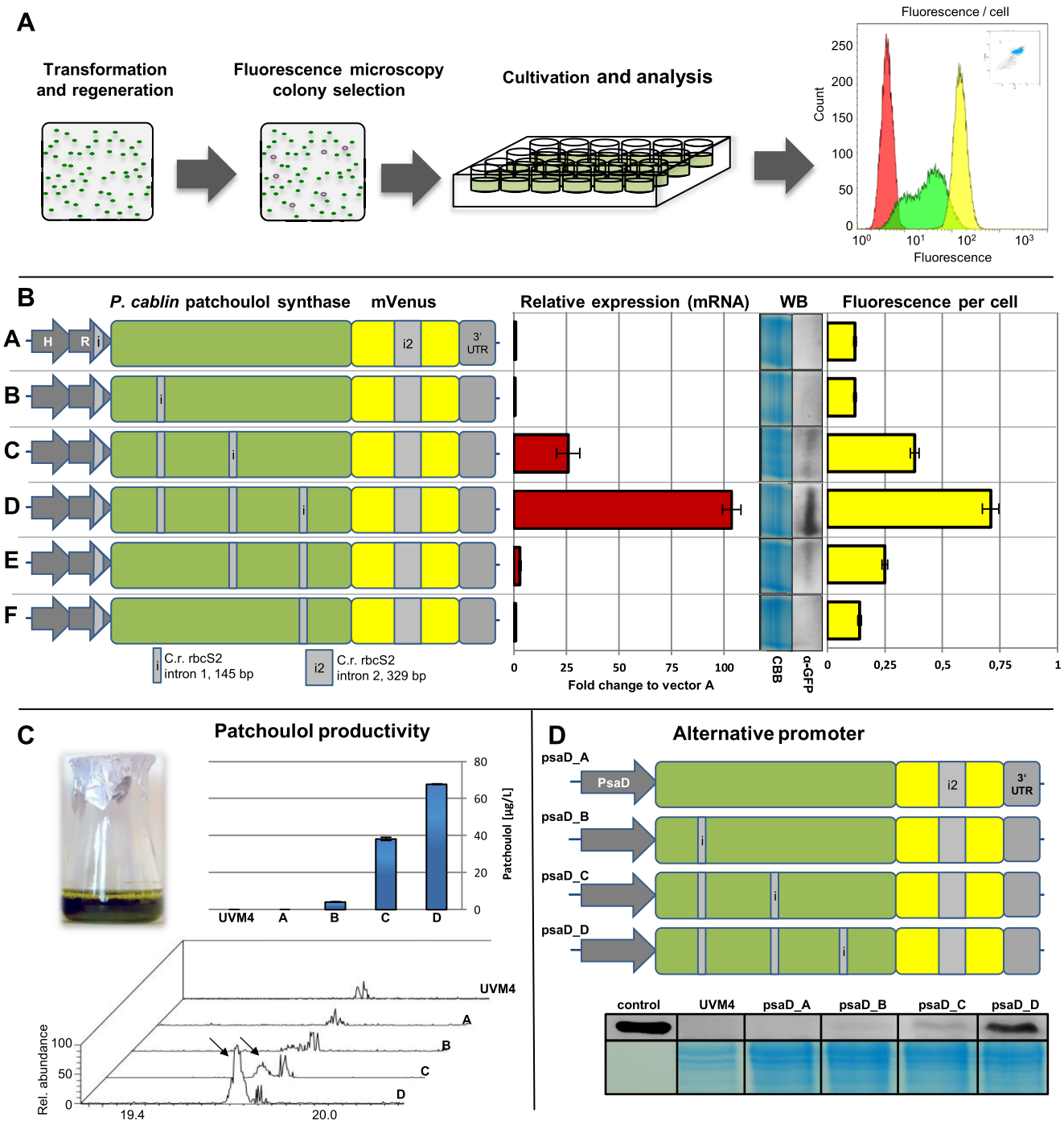

Figure 2. Effects of rbcS2i1 addition on transgene expression levels of the codon-optimized *P. cablin* patchoulol synthase (*Pc*Ps) gene. (**A**) Screening method of regenerated mutants—fluorescence microscopy for the YFP reporter was conducted directly on the transformation plate as previously described (35). Cultivation was conducted in microtiter plates and mutants were pooled according to their respective cell densities prior to analysis e.g. flow cytometry. (**B**) A–F Expression cassettes of *Pc*Ps rbcS2i1 intron variants cloned in the pOpt_mVenus_Paro vector. The relative expression levels for each construct are shown as relative abundance of mRNA (RTqPCR), protein titer (WB – western blot, α-GFP with Coomassie Brilliant Blue (CBB) as loading control) and mean fluorescence cell^{-1} (flow cytometry). (**C**) Patchoulol productivity was quantified by pooled mutant cultivation after 5 days in shake-flasks using 5% (v/v) dodecane overlay followed by GC–MS as previously described (35). (**D**) Vector set with alternative PsaD promoter sequence (38) and respective transformant *Pc*Ps_YFP expression levels analyzed by WB. Error bars represent standard deviations from mean of triplicate measurements for pooled populations, pre-selected by YFP fluorescence. H R i – the HSP70/RBCS2i1 promoter, 3′ UTR – 3′ untranslated region of the rbcS2 gene.

expressed from two vectors by subsequent transformations and screening efforts to increase the relative *Pc*Ps enzyme titer in the algal cytoplasm (35).

Rubisco is considered to be the most abundant protein in plants and microalgae (48) it is likely that complex intracellular regulation is favored towards expression from this gene locus. The rbcS genetic elements (e.g. promoter, rbcS2 introns and 3′ UTR) have long been associated with effective transgene expression elements for *C. reinhardtii* (21,31). A comparable regulation was recently identified in the native *C. reinhardtii*, intron-containing ift25 gene which, when interacting with its corresponding promoter was found to exhibit improved transgene expression levels (30). In our study, the chimeric HSP70/RBCS2i1 promoter was used to drive constitutive transcription. Since both the promoter and intron elements are derived from the same genetic origin, a similar interaction is possible. The endogenous and constitutive PsaD promoter (38) natively drives transcription of an intron-less gene and therefore, represents an effective control as it should not exhibit interactions with the host splicing apparatus. Four expression constructs were made by replacing the HSP70/RBCS2i1 with the PsaD promoter in the intron-containing *Pc*Ps vectors (A–D). The *Pc*Ps_mVenus protein content of transformants was analyzed by immunodetection (Figure 2D). The achieved titers correlate with those generated with the HSP70/RBCS2i1 promoter, indicating that the mechanism affecting transgene expression is not related to the interaction of the rbcS2i1 with its native promoter. This is in line to the findings of studies which used the rbcS2i1 in other promoter sequences (14,32). We did not observe alternative splicing or recombination effects in this, or the previously reported *Ag*Bs optimized algal transgene, even when up to 15 copies of the rbcS2i1 were spread across larger repetitive gene fusions (3x*Pc*Ps (35) and ispA_2x*Ag*Bs (36), respectively), indicating a conserved regulation of transcript processing.

Investigation of intron position effects within the CDS

The *Pc*Ps vector D contained three rbcS2i1 copies spread across the codon-optimized CDS with small exon lengths and exhibited the most robust transgene expression levels. To elucidate whether transgene expression improvements are due to mRNA processing or specific to this intron, an intron of similar composition from the *C. reinhardtii* actin gene locus (intron 3, 95 bp, NCBI: D50838.1) was investigated as an alternative. The actini3 was exchanged with the third rbcS2i1 in the *Pc*Ps gene (vector G) and transgene expression efficiencies were compared to the parental vector (Figure 3). The expression levels were markedly lower than for vector C which had the third rbcS2i1 removed and an exon length of 1057 bp (exon 3), indicating that this intron was not an effective substitute for the rbcS2i1.

We also investigated the *Pc*Ps with up to three copies of the actini3 in the same insertion sites used for the rbcS2i1 (Supplementary Figure S6). Regenerated transformants exhibited construct expression detectable by YFP fluorescence on the initial transformation plate, indicating that correct splicing occurred for all vectors. However, no actini3 copy number-specific increase in the overall expression levels was detected between the constructs (not shown). The results suggest that the effect of transcript processing contributes less to the enhanced transgene expression effect observed for rbcS2i1 than the likely recruitment of the transcriptional machinery and its consequent enhancement of transcript abundance. To further confirm this finding, we investigated mRNA stability of transcripts from different *Pc*Ps gene variants by actinomycin D treatment (Supplementary Figure S7). For all tested sequences, no construct specific increase in transcript abundance was found and mRNA levels were strikingly reduced after ∼45 min of incubation, indicating that there is no difference in mRNA stability relative to rbcS2i1 copy number. It has been previously reported that the rbcS2i1 may contain endogenous enhancer elements, as it has been able to drive higher rates of antibiotic selection even when placed upstream of a promoter region (18). Such an enhancer could function to recruit the elements of the transcription machinery by either direct transcription factor binding or acting on the epigenetic level, making the insertion site more accessible to regulatory protein binding. It was previously reported that knock out of the *met1* gene leads to a globally reduced DNA methylation status, and consequently to higher transgene expression from the nuclear genome of *C. reinhardtii* (13,49). The strain UVM11, which also exhibits improved nuclear transgene expression capacities, reportedly has an reduced Histone 3 occupancy, increased Histone 4 acetylation and an reduced H3 Lysine 9 monomethylation status at the gene of interest site (15). It is possible that similar epigenetic regulations can be induced by rbcS2i1 addition specifically at the gene of interest locus, thereby reducing the formation of transcriptional inactivated heterochromatin. The rbcS2 introns are likely involved in regulating the transcription process as their serial incorporation in CDSs here resulted in step-wise increases in transcript abundance.

To elucidate the correlation between the specific position of introns to each other, and the resulting exon length, a vector was created with the rbcS2i1 at an additional upstream insertion site between the promoter and the first intron of the codon-optimized *Pc*Ps while removing the third position, creating a long downstream exon stretch (vector H, Figure 3). Another version was created leaving this intron in the 3′ end (vector I) so that the *Pc*Ps contained a fourth rbcS2i1. Both *Pc*Ps versions expressed efficiently with higher expression of vector I, indicating that each additional copy of the rbcS2i1 positively affects transgene expression, however, the exon length was a crucial factor in gene expression efficiency. Vector H has a large 3′ exon of 1057 bp (until rbcS2i2 of the mVenus in the pOpt vector) and strains generated with this vector had lower expression than both vectors D and I. The results suggest that there is a synergy of minimal exon length and rbcS2i1 copy number which encourages efficient transgene expression in optimized algal transgenes.

The pOptimized vector reporters were strategically designed for nuclear transgene expression in *C. reinhardtii*, each reporter contains the second intron of rbcS2 in a central position of the reporter CDS (19). When this sequence was changed to an rbcS2 intron 1, the overall expression of the *Pc*Ps was lower than vectors I, D or H, however, it was higher than the intron-less vector A. This result indicates that expression is improved from an interplay of the two

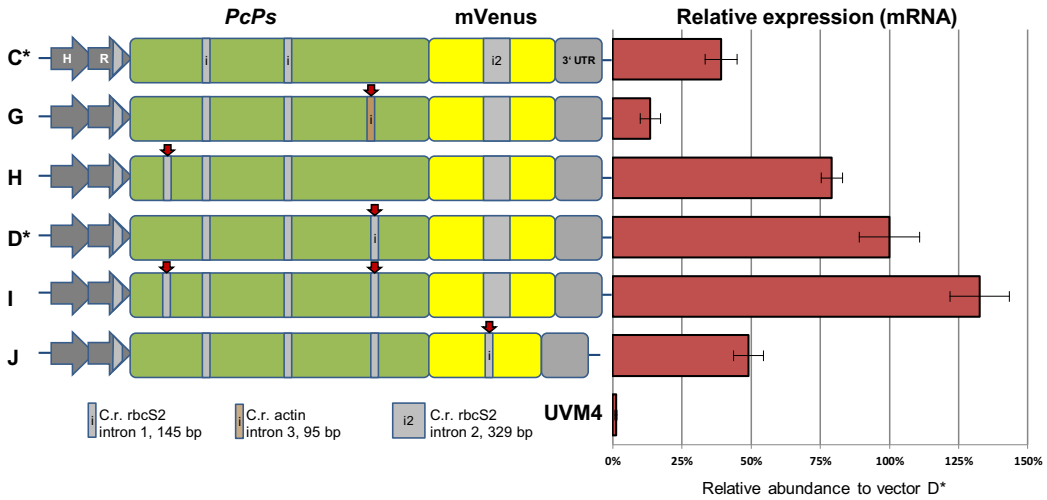

Figure 3. Analysis of position effects of different insertion sites of the rbcS2i1 in the codon-optimized *Pc*Ps CDS. Vector G is a modification of vector C wherein the actin intron 3 (95 bp, NCBI: D50838.1) has been inserted into the empty third position. In vector H the rbcS2i1 was placed upstream of the first insertion site. Vectors C* and D* are shown for construct and expression efficiency comparison. Vector I contains the original third rbcS2i1 position as well as the additional upstream insertion found in vector H. In addition, the rbcS2i2 located in the mVenus reporter was replaced by a rbcS2i1 (vector J). Transformants were generated as above and the relative expression levels per construct were quantified by RTqPCR as the relative abundance of transcript mRNA normalized to vector D. UVM4 parental strain is shown as a negative control. Error bars represent standard deviations from mean of triplicate measurements for pooled populations, pre-selected by YFP fluorescence.

rbcS2 intron sequences (rbcS2i1 and rbcS2i2) spread within a single gene, and that the rbcs2i1 can be used repetitively before the second intron. It was previously reported that the rbcS2 intron 2 may have a positive effect on transgene expression at the posttranscriptional level when placed into the *sh*ble gene (18), and was shown to enhance expression of a recombinant luciferase in synergy with rbcS2i1 (31), further supporting the value of the use of this sequence within the pOptimized vector reporters.

The optimization strategy used here to design the *Pc*Ps expression construct (vector D) was able to generate strains with recombinant protein accumulation of between 10–15 ng/30 μg (0.03–0.05%) total soluble protein (TSP) and 5 ng/30 μg (0.02%) total cellular protein (TP) in early logarithmic growth phase (Supplementary Figure S8). Although the protein accumulation of the *Pc*Ps_YFP is below the maximum reported yield of the YFP reporter alone (∼1% TSP, (15)), the intron-containing design vastly outperforms an intron-less control and achieved a more than 100-fold improvement in mRNA abundance (Figure 2B). The results indicate that there is a potential for increased target protein accumulation mediated by the native translational machinery in the algal cell. We have observed individual transformants with higher *Pc*Ps_YFP expression, however, cells exhibited reduced fitness and survived only when maintained in short generation times on the agar plate level. This is likely due to the intracellular accumulation of the sesquiterpenoid alcohol patchoulol (not shown) which may have affected the TSP levels observed here. Indeed, we recently demonstrated expression of another sesquiterpene synthase in this manner (*Ag*Bs), which produces the protective sesquiterpene (*E*)-α-bisabolene. In fusion with different reporters, it was possible to increase levels of the recombinant *Ag*Bs protein significantly within the algal cell (36).

Transgene design for nuclear expression in *C. reinhardtii*

In this study, we characterized the use of the rbcS2i1 spread throughout the codon-optimized CDS of the sesquiterpenoid synthase *Pc*Ps. Recently, we also applied this strategy to express the large sesquiterpene synthase *Ag*Bs as well as two farnesyl pyrophosphate synthases ERG20 and ispA from the nuclear genome of *C. reinhardtii* (35,36). Since these first demonstrations, we have applied this transgene optimization strategy for the expression of numerous other heterologous codon-optimized transgenes. Selected examples of these are presented in Figure 4 (see Supplementary Data File S1 for complete sequence information of all vector sequences generated in this study). These constructs were codon-optimized and the rbcS2i1 spread throughout their CDSs to ensure exon lengths ∼500 bp as described above. After cloning into the pOpt_YFP_Paro vector for C-terminal YFP reporter protein fusion, the constructs were transformed into *C. reinhardtii* (Figure 4A). For all constructs, expression in fusion with the mVenus reporter resulted in colonies which exhibited fluorescence at the agar plate level, and when selected mutants were analyzed by Western blot, full-length products could be detected at appropriate predicted molecular masses (Figure 4B). The capacity for mutant identification varied greatly between each construct, with larger transgenes such as the *Pc*Ps- or *Ag*Bs_YFP exhibiting far fewer colonies with detectible fluorescence at all levels of expression (Figure 4C). Nevertheless, out of 300 randomly isolated colonies per construct, it was possible to identify clones which expressed robustly. The relative expression frequency for the largest constructs here indicates the value of target protein fusion to a reporter protein for quantification of expression, as the necessity for screening large numbers of transformants would be technically limiting by traditional immunoblotting techniques.

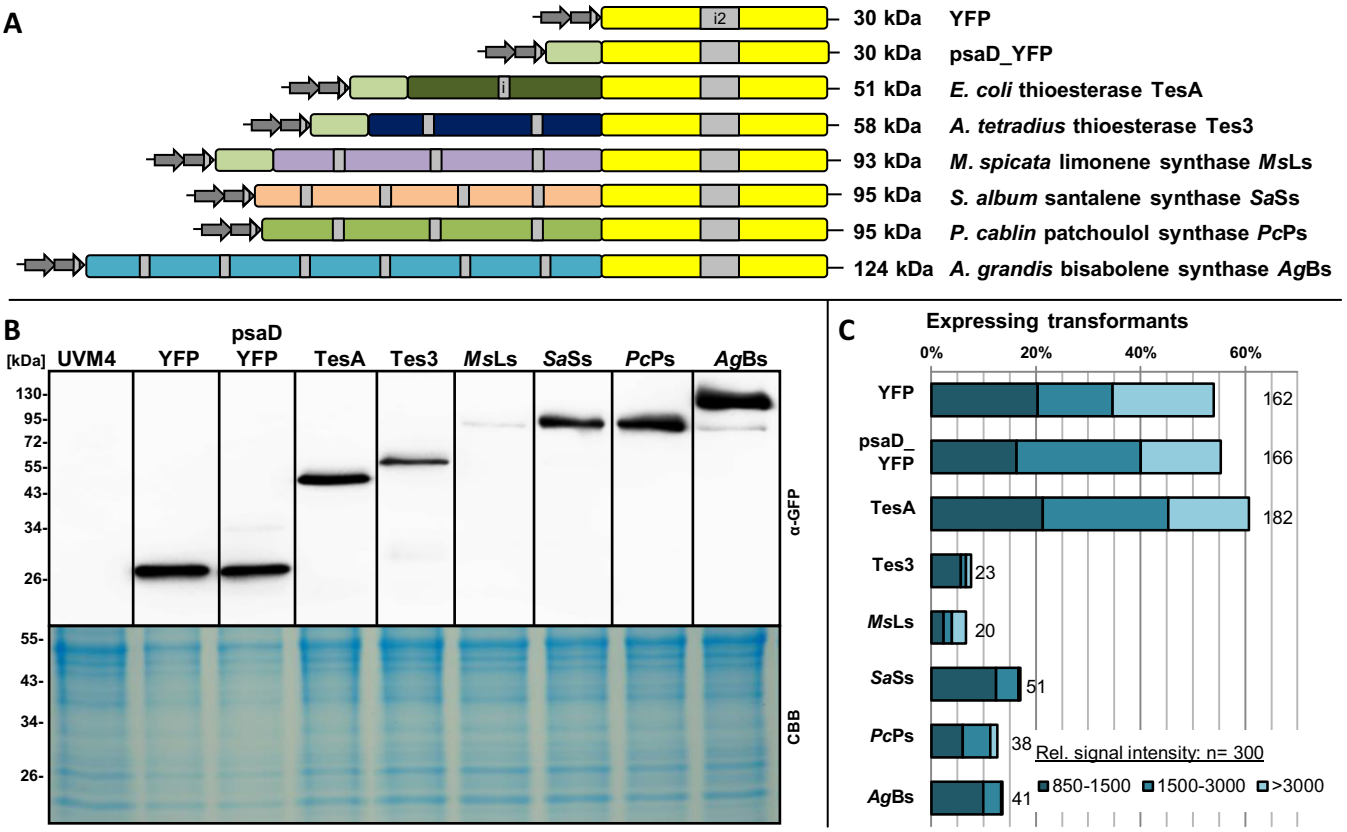

Figure 4. Design of selected gene of interest sequences as optimized algal transgenes for enhanced nuclear expression in *C. reinhardtii*: (**A**) *Escherichia coli* TesA (NCBI accession no: MH004289, (50)), *Anaerococcus tetradius* Tes3 (MH004290, (51)), *Santalum album* santalene synthase *Sa*Ss (MH004288, (52)), *Mentha spicata* limonene synthase *Ms*Ls (MH004287, (53)), *P. cablin* patchoulol synthase *Pc*Ps (KX097887, (54)) and *A. grandis* bisabolene synthase *Ag*Bs (MG052654.1, (55)). Genes were designed *in silico* followed by gene synthesis, cloned into the pOpt_mVenus_Paro vector backbone and expressed in UVM4. (**B**) Protein samples from expressing mutants transformed with the different optimized transgenes were separated by SDS-PAGE prior to western blot analysis (α-GFP). (**C**) Relative numbers of transformants from 300 randomly picked *C. reinhardtii* mutants per construct, which express detectible levels of YFP fusion proteins. Relative expression rates across this population are indicated with different colour intensities. Vector diagrams: light green boxes indicate the 36 amino acid PsaD chloroplast target peptide used on some constructs for subcellular targeting. H R i – the HSP70/RBCS2i1 promoter, 3′ UTR – 3′ untranslated region of the rbcS2 gene. *shble* bleomycin/zeocin resistance gene.

The results indicate that the design strategy, coupling codon optimization and rbcS2i1 intron spreading to minimize exon length, may be a useful tool to enable transgene expression in *C. reinhardtii*. It is possible, however, that not every protein will be accepted in the eukaryotic cell environment and that accumulation of all targets cannot be guaranteed. Nevertheless, the results indicate that stable integration of large transgenes is possible with this optimization. In general, it has enabled the expression of numerous codon-optimized transgene CDSs to a level readily detectable by immunoblotting. Therefore, we propose the following strategy for generation of desired gene expression constructs:

1. Codon optimization (100% most frequent codon for *C. reinhardtii* nuclear genome, Kazuza database: http://www.kazusa.or.jp/codon) of the desired coding sequence.
2. Identification of optimal insertion sites (…NG/GN…) throughout the transgene sequence with respect to maintain exon lengths of ∼500 bp, considering the distance to the promoter and gene fusion partners. Insert the 145 bp rbcS2i1 sequence *in silico*.
3. Gene synthesis and cloning in a suitable expression vector containing an appropriate promoter and 3′UTR. If not using the pOptimized vector and/or reporters, substitute the last intron position for the rbcS2i2.

The mechanism by which the rbcS2i1 is able to reliably enhance transgene expression in the nuclear genome remains unclear, although the results presented here indicate that its inherent enhancer effect in synergy with reduced exon length is likely responsible for increased transcription and higher overall transgene expression rates. How the rbcS2i1 encourages greater transcription and investigation of the chromatin status of surrounding rbcS2i1 containing genes will be of great interest for future studies. Elucidating the transcriptional regulators and DNA–protein binding partners within this sequence will certainly increase our overall understanding of transgene regulation mechanics in *C. reinhardtii* and may have implications for genetic engineering concepts in other green microalgae as well as higher plants. Here, we have presented practical insights into the use of artificial introns as another valuable layer of nuclear transgene optimization and demonstrated their pos-

itive effect on target transgene expression. We have found repetitive spreading of these sequences throughout codon-optimized CDSs enables reliable and robust transgene expression of even larger transgenes from the nuclear genome of *C. reinhardtii* and may be a valuable strategy to overcome some of the nuclear transgene expression limitations from this host.

CONCLUSION

The regulation of gene expression from the nuclear genome of *C. reinhardtii* is a complex process with numerous inherent levels of regulation. Here, we have determined that this regulation can be correlated with the maximal exon length of transformed transgenes and that a novel eukaryotic gene design using the rbcS2 introns to customize transgenes to the nuclear genome environment has a strong positive influence on transgene expression levels. *C. reinhardtii* has a great flexibility of correct mRNA splicing of artificially introduced rbcS2 introns and although the exact mechanisms remain unclear, their enhancing effect on transcript abundances was systematically investigated and correlated with higher overall recombinant protein titers. We provide evidence that the gene expression regulation can effectively be manipulated to express different target sequences designed as intron-containing optimized transgenes and have given clear guidelines for common use of this strategy for the greater microalgal community.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to express thanks to Prof. Dr. Ralph Bock for the strain UVM4 and the Technology Platform at the Center for Biotechnology (CeBiTec) at Bielefeld University for access.

FUNDING

CLIB Graduate Cluster Industrial Biotechnology (CLIB-GC) [to T.B.]; European Union's Horizon 2020 (640720 Photofuel) [to O.K.]. Funding for open access charge: Public funding from Bielefeld University.
Conflict of interest statement. None declared.

REFERENCES

1. Xue,J., Niu,Y.F., Huang,T., Yang,W.D., Liu,J.S. and Li,H.Y. (2015) Genetic improvement of the microalga *Phaeodactylum tricornutum* for boosting neutral lipid accumulation. *Metab. Eng.*, **27**, 1–9.
2. Jeon,S., Lim,J.M., Lee,H.G., Shin,S.E., Kang,N.K., Park,Y. Il, Oh,H.M., Jeong,W.J., Jeong,B.R. and Chang,Y.K. (2017) Current status and perspectives of genome editing technology for microalgae. *Biotechnol. Biofuels*, **10**, 1–18.
3. Mussgnug,J.H. (2015) Genetic tools and techniques for *Chlamydomonas reinhardtii*. *Appl. Microbiol. Biotechnol.*, **99**, 5407–5418.
4. Scaife,M.A. and Smith,A.G. (2016) Towards developing algal synthetic biology. *Biochem. Soc. Trans.*, **44**, 716–722.
5. Jaeger,D., Hübner,W., Huser,T., Mussgnug,J.H. and Kruse,O. (2017) Nuclear transformation and functional gene expression in the oleaginous microalga *Monoraphidium neglectum*. *J. Biotechnol.*, **249**, 10–15.
6. Harris,E.H. (2001) *Chlamydomonas* as a model organism. *Mol. Biol.*, **52**, 363–406.
7. Hegemann,P., Fuhrmann,M. and Kateriya,S. (2001) Algal sensory photoreceptors. *Annu. Rev. Plant Biol.*, **37**, 668–676.
8. 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.
9. Schroda,M., Blöcker,D. and Beck,C.F. (2000) The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. *Plant J.*, **21**, 121–131.
10. Shao,N. and Bock,R. (2008) A codon-optimized luciferase from *Gaussia princeps* facilitates the *in vivo* monitoring of gene expression in the model alga *Chlamydomonas reinhardtii*. *Curr. Genet.*, **53**, 381–388.
11. Cerutti,H. (1997) Epigenetic silencing of a foreign gene in nuclear transformants of *Chlamydomonas*. *Plant Cell Online*, **9**, 925–945.
12. Neupert,J., Karcher,D. and Bock,R. (2009) Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. *Plant J.*, **57**, 1140–1150.
13. Kurniasih,S.D., Yamasaki,T., Kong,F., Okada,S., Widyaningrum,D. and Ohama,T. (2016) UV-mediated *Chlamydomonas* mutants with enhanced nuclear transgene expression by disruption of DNA methylation-dependent and independent silencing systems. *Plant Mol. Biol.*, **92**, 629–641.
14. Scranton,M.A., Ostrand,J.T., Georgianna,D.R., Lofgren,S.M., Li,D., Ellis,R.C., Carruthers,D.N., Dräger,A., Masica,D.L. and Mayfield,S.P. (2016) Synthetic promoters capable of driving robust nuclear gene expression in the green alga *Chlamydomonas reinhardtii*. *Algal Res.*, **15**, 135–142.
15. 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.
16. Weiner,I., Atar,S., Schweitzer,S., Eilenberg,H., Feldman,Y., Avitan,M., Blau,M., Danon,A., Tuller,T. and Yacoby,I. (2018) Enhancing heterologous expression in *Chlamydomonas reinhardtii* by transcript sequence optimization. *Plant J.*, **94**, 22–31.
17. Rasala,B.A., Lee,P.A., Shen,Z., Briggs,S.P., Mendez,M. and Mayfield,S.P. (2012) Robust expression and secretion of xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide. *PLoS One*, **7**, e43349.
18. Lumbreras,V., Stevens,D.R. and Purton,S. (1998) Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. *Plant J.*, **14**, 441–447.
19. Lauersen,K.J., Kruse,O. and Mussgnug,J.H. (2015) Targeted expression of nuclear transgenes in *Chlamydomonas reinhardtii* with a versatile, modular vector toolkit. *Appl. Microbiol. Biotechnol.*, **99**, 3491–3503.
20. Merchant,S.S., Prochnik,S.E., Vallon,O., Harris,E.H., Karpowicz,J., Witman,G.B., Terry,A., Salamov,A., Fritz-laylin,L.K., Maréchal-drouard,L. *et al.* (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*, **318**, 245–250.
21. Sizova,I., Fuhrmann,M. and Hegemann,P. (2001) A *Streptomyces rimosus* aphVIII gene coding for a new type phosphotransferase provides stable antibiotic resistance to *Chlamydomonas reinhardtii*. *Gene*, **277**, 221–229.
22. Berthold,P., Schmitt,R. and Mages,W. (2002) An engineered *Streptomyces hygroscopicus* aph 7″ gene mediates dominant resistance against hygromycin B in *Chlamydomonas reinhardtii*. *Protist*, **153**, 401–412.
23. Jąkalski,M., Takeshita,K., Deblieck,M., Koyanagi,K.O., Makałowska,I., Watanabe,H. and Makałowski,W. (2016) Comparative genomic analysis of retrogene repertoire in two green algae *Volvox carteri* and *Chlamydomonas reinhardtii*. *Biol. Direct*, **11**, 35.
24. Bogen,C., Al-Dilaimi,A., Albersmeier,A., Wichmann,J., Grundmann,M., Rupp,O., Lauersen,K.J., Blifernez-Klassen,O., Kalinowski,J., Goesmann,A. *et al.* (2013) Reconstruction of the lipid metabolism for the microalga *Monoraphidium neglectum* from its

genome sequence reveals characteristics suitable for biofuel production. *BMC Genomics*, **14**, 926.
25. Corteggiani Carpinelli,E., Telatin,A., Vitulo,N., Forcato,C., D'Angelo,M., Schiavon,R., Vezzi,A., Giacometti,G.M., Morosinotto,T. and Valle,G. (2014) Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. *Mol. Plant*, **7**, 323–335.
26. Bowler,C., Allen,A.E., Badger,J.H., Grimwood,J., Jabbari,K., Kuo,A., Maheswari,U., Martens,C., Maumus,F., Otillar,R.P. *et al.* (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*, **456**, 239–244.
27. Mount,S.M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.*, **10**, 459–472.
28. Huang,M.T.F. and Gorman,C.M. (1990) Intervening sequences increase efficiency of RNA 3′ processing and accumulation of cytoplasmic RNA. *Nucleic Acids Res.*, **18**, 937–947.
29. Chapman,B.S., Thayer,R.M., Vincent,K.A. and Haigwood,N.L. (1991) Effect of intron-a from human cytomegalovirus (Towne) immediate-early gene on heterologous expression in mammalian-cells. *Nucleic Acids Res.*, **19**, 3979–3986.
30. Dong,B., Hu,H.H., Li,Z.F., Cheng,R.Q., Meng,D.M., Wang,J. and Fan,Z.C. (2017) A novel bicistronic expression system composed of the intraflagellar transport protein gene ift25 and FMDV 2A sequence directs robust nuclear gene expression in *Chlamydomonas reinhardtii*. *Appl. Microbiol. Biotechnol.*, **101**, 4227–4245.
31. Eichler-Stahlberg,A., Weisheit,W., Ruecker,O. and Heitzer,M. (2009) Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. *Planta*, **229**, 873–883.
32. Ferrante,P., Catalanotti,C., Bonente,G. and Giuliano,G. (2008) An optimized, chemically regulated gene expression system for *Chlamydomonas*. *PLoS One*, **3**, e3200.
33. Bruggeman,A.J., Kuehler,D. and Weeks,D.P. (2014) Evaluation of three herbicide resistance genes for use in genetic transformations and for potential crop protection in algae production. *Plant Biotechnol. J.*, **12**, 894–902.
34. Jiang,W.Z. and Weeks,D.P. (2017) A gene-within-a-gene Cas9/sgRNA hybrid construct enables gene editing and gene replacement strategies in *Chlamydomonas reinhardtii*. *Algal Res.*, **26**, 474–480.
35. Lauersen,K.J., Baier,T., Wichmann,J., Wördenweber,R., Mussgnug,J.H., Hübner,W., Huser,T. and Kruse,O. (2016) Efficient phototrophic production of a high-value sesquiterpenoid from the eukaryotic microalga *Chlamydomonas reinhardtii*. *Metab. Eng.*, **38**, 331–343.
36. Wichmann,J., Baier,T., Wentnagel,E., Lauersen,K.J. and Kruse,O. (2018) Tailored carbon partitioning for phototrophic production of (*E*)-α-bisabolene from the green microalga *Chlamydomonas reinhardtii*. *Metab. Eng.*, **45**, 211–222.
37. Higuchi,R., Krummel,B. and Saiki,R. (1988) A general method of in vitro preparation and specific mutagenesis of dna fragments: Study of protein and DNA interactions. *Nucleic Acids Res.*, **16**, 7351–7367.
38. Fischer,N and R.,J. (2001) The flanking regions of PsaD drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*. *Mol. Genet. Genomics*, **265**, 888–894.
39. Gorman,D.S. and Levine,R.P. (1965) Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardi*. *Proc. Natl. Acad. Sci. U.S.A.*, **54**, 1665–1669.
40. Kindle,K.L. (1990) High frequency nuclear transformation of *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 1228–1232.
41. Chomczynski,P. (1987) Single-step method of RNA isolation by acid guanidinium extraction. *Anal. Biochem.*, **159**, 156–159.
42. Cuvelier,M.L., Ortiz,A., Kim,E., Moehlig,H., Richardson,D.E., Heidelberg,J.F., Archibald,J.M. and Worden,A.Z. (2008) Widespread distribution of a unique marine protistan lineage. *Environ. Microbiol.*, **10**, 1621–1634.
43. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*, **25**, 402–408.
44. Laemmli,U.K., Molbert,E., Showe,M. and Kellenberger,E. (1970) Form- determining function of the genes required for the assembly of the head of bacteriophage T4, Journal of molecular biology. *J. Mol. Biol.*, **49**, 99–113.
45. Dyballa,N. and Metzger,S. (2009) Fast and sensitive colloidal Coomassie G-250 staining for proteins in polyacrylamide gels. *J. Vis. Exp.*, **30**, e1431.
46. Bellaousov,S., Reuter,J.S., Seetin,M.G. and Mathews,D.H. (2013) RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.*, **41**, W471–W474.
47. Labadorf,A., Link,A., Rogers,M.F., Thomas,J., Reddy,A.S. and Ben-Hur,A. (2010) Genome-wide analysis of alternative splicing in *Chlamydomonas reinhardtii*. *BMC Genomics*, **11**, 114.
48. Harris,E.H., Stern,D. and Witman,G. (2009) *The Chlamydomonas Source Book*. Academic Press, Oxford, Vol. **1**.
49. Kong,F., Yamasaki,T., Kurniasih,S.D., Hou,L., Li,X., Ivanova,N., Okada,S. and Ohama,T. (2015) Robust expression of heterologous genes by selection marker fusion system in improved *Chlamydomonas* strains. *J. Biosci. Bioeng.*, **120**, 239–245.
50. Riley,M., Abe,T., Arnaud,M.B., Berlyn,M.K.B., Blattner,F.R., Chaudhuri,R.R., Glasner,J.D., Horiuchi,T., Keseler,I.M., Kosuge,T. *et al.* (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot - 2005. *Nucleic Acids Res.*, **34**, 1–9.
51. Rutter,C.D., Zhang,S. and Rao,C. V. (2015) Engineering *Yarrowia lipolytica* for production of medium-chain fatty acids. *Appl. Microbiol. Biotechnol.*, **99**, 7359–7368.
52. Jones,C.G., Moniodis,J., Zulak,K.G., Scaffidi,A., Plummer,J.A., Ghisalberti,E.L., Barbour,E.L. and Bohlmann,J. (2011) Sandalwood fragrance biosynthesis involves sesquiterpene synthases of both the terpene synthase (TPS)-a and TPS-b subfamilies, including santalene synthases. *J Biol Chem.*, **286**, 17445–17454.
53. Colbys,S.M., Alonsoj,W.R., Katahira,E.J., Mcgarvey,D.J. and Croteaull,R. (1993) 4S-limonene synthase from the oil glands of spearmint (*Mentha spicata*). *J. Biol. Chem.*, **268**, 23016–23024.
54. Deguerry,F., Pastore,L., Wu,S., Clark,A., Chappell,J. and Schalk,M. (2006) The diverse sesquiterpene profile of patchouli, *Pogostemon cablin*, is correlated with a limited number of sesquiterpene synthases. *Arch. Biochem. Biophys.*, **454**, 123–136.
55. Bohlmann,J., Crock,J., Jetter,R. and Croteau,R. (1998) Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (*E*)-alpha-bisabolene synthase from grand fir (Abies grandis). *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 6756–6761.