This is a repository copy of *Novel cis-regulatory elements as synthetic promoters to drive recombinant protein expression from the Chlamydomonas reinhardtii nuclear genome*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/182738/

Version: Published Version

Full length Article

Novel *cis*-regulatory elements as synthetic promoters to drive recombinant protein expression from the *Chlamydomonas reinhardtii* nuclear genome

Josie L. McQuillan [a], Anthony J. Berndt [b], Ashley E. Sproles [b], Stephen P. Mayfield [b], Jagroop Pandhal [a],*

[a] *Department of Chemical and Biological Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK*
[b] *California Center for Algae Biotechnology, Division of Biological Sciences, University of California, San Diego, La Jolla, CA, 92093, United States*

A B S T R A C T

Eukaryotic green microalgae represent a sustainable, photosynthetic biotechnology platform for generating high-value products. The model green alga *Chlamydomonas reinhardtii* has already been used to generate high value bioproducts such as recombinant proteins and terpenoids. However, low, unstable, and variable nuclear transgene expression has limited the ease and speed of metabolic engineering and recombinant protein expression in this system. Here, novel genetic devices for transgene expression in *C. reinhardtii* have been developed by identifying *cis*-regulatory DNA elements capable of driving high transgene expression in *C. reinhardtii* promoters using *de novo* motif discovery informatics approaches. Thirteen putative motifs were synthesized as concatemers, linked to a common minimal basal promoter, and assayed for their activity to drive expression of a yellow fluorescent protein reporter gene. Following transformation of the vectors into *C. reinhardtii* by electroporation, *in vivo* measurements of yellow fluorescent protein expression by flow cytometry revealed that five of the DNA motifs analyzed displayed significantly higher reporter expression compared to the basal promoter control. Two of the concatemerized motifs, despite being much smaller minimal *cis*-regulatory elements, drove reporter expression at levels approaching that of the conventionally-used AR1 promoter. This analysis provides insight into *C. reinhardtii* promoter structure and gene regulation, and provides a new toolbox of *cis*-regulatory elements that can be used to drive transgene expression at a variety of expression levels.

Introduction

Green microalgae are promising biotechnology hosts for the sustainable production of valuable products such as biofuels, omega-3 fatty acids, pigments, bioplastics and recombinant proteins [1,2]. The ability of microalgae to grow photoautotrophically at large scales offers potentially large environmental and economic advantages over yeast and bacterial bioproduction systems. However, the establishment of reliable and diverse genetic tools engineering microalgae to produce desired products at high yields has comparatively lagged behind other microbial systems. The green alga *Chlamydomonas reinhardtii* has become a major model organism for the study of various important cellular processes such as phototaxis, photosynthesis, carbon concentrating mechanisms and the cell cycle [3–6]. As a result, a wealth of -omics data, a well-annotated genome, efficient transformation

techniques and genetic engineering tools are available for *C. reinhardtii* [7].

Advancing the molecular toolkit for recombinant protein expression from the *C. reinhardtii* nuclear genome will be vital for its development as a biotechnological host. Several recent successes in *C. reinhardtii* transgene expression have targeted the chloroplast genome, in part due to its ease of genomic manipulation by homologous recombination and high levels of protein accumulation [8]. Chloroplast genome-encoded transgenes do not appear to be targetable to other parts of the cell outside of the chloroplast (*i.e.* ER-Golgi localization and secretion) and post-translational modifications such as glycosylation that may be critical for the biological activity of a recombinant protein are effectively nonexistent. Moreover, high transgene expression in the chloroplast has generally relied on integration strategies that diminish or abrogate the photosynthetic ability of the alga [9–11]. It is becoming increasingly

---

clear that efficient nuclear genome engineering will be essential for successful metabolic engineering and recombinant protein expression. However, there are significant challenges that have slowed progression of nuclear genome engineering in *C. reinhardtii*, such as epigenetic transgene silencing [12], random integration of transgenes leading to strong positional effects altering transgene expression, and a lack of strong and reliable promoters to drive the expression of nuclear transgenes. The hybrid *Hsp70A-RbcS2* (AR1) and photosystem I protein D (PSAD) are currently the strongest constitutive expression promoters available for *C. reinhardtii* [13–15]; however, they are still susceptible to transgene silencing, and expression levels are highly variable while still being generally lower than what can be achieved from transgenes inserted in the chloroplast genome.

Synthetic promoters have been designed to overcome similar problems in several host cell systems, including bacteria [16], yeast [17], mammalian cells [18] and plants [19]. Non-native promoters can offer several advantages to natively derived promoters, including a reduced propensity for homology-based silencing and the potential to push gene expression to high levels [20]. Recently, novel synthetic promoters for *C. reinhardtii* were designed and tested *in silico* [21]. In this previous study, the promoter regions of highly-expressed genes were analyzed using POWRS motif discovery software [22] to identify common motifs and patterns in highly expressed genes, which were then used to generate 500 bp synthetic promoters. Promoters driving high levels of expression were discovered, as well as one *cis*-regulatory motif that was necessary for transgene expression. This method for developing synthetic promoters was effective, but the individual minimal motifs responsible for eliciting improved expression were not isolated and characterized *in vivo*.

A common method for designing synthetic promoters involves combining known *cis*-regulatory DNA elements (CREs) that are known to recruit transcription factors [20]. This level of understanding is currently lacking for *C. reinhardtii*; very few CREs in *C. reinhardtii* have been identified and characterized, and the ones that have are mostly involved in inducible protein expression under stress conditions, as opposed to constitutive expression [23]. Advancing understanding of individual CREs in *C. reinhardtii* would open up the opportunity to produce bespoke synthetic promoters with interchangeable parts, enabling tailored expression levels by combining high, low, and potentially inducible DNA motifs to optimize nuclear transgene expression. This would additionally increase our understanding of general promoter characteristics in algal systems. Identifying and testing a small set of motifs individually could quickly provide insight into which DNA sequences can be incorporated into synthetic promoters to induce transgene expression. This would enable improved control over synthetic promoter design in *C. reinhardtii* through understanding individual promoter components, facilitating the production of modular promoter 'building blocks' for predictable and more precise protein expression. The aims of this study were: (1) to identify putative *cis*-regulatory elements (pCREs) within the promoter regions of highly expressed genes using previously published transcriptomics data and publicly available motif discovery software, (2) then to screen these motifs *in vivo* for promoter activity: and (3) to assess these *cis*-regulatory elements' suitability as standalone synthetic promoters and as modules for use in future synthetic promoter design for microalgal systems.

Materials and methods

Promoter analysis and motif identification

The workflow applied to identify pCRE sequences is depicted in Fig. 1. Promoter sequences of the top 300 highly-expressed constitutive genes identified from a previously published microarray dataset [24] were selected for analysis (Supp. Table S1). Promoter regions were defined as –1000 bp from the transcription start site (TSS). All *C. reinhardtii* promoter sequences were retrieved in FASTA format from
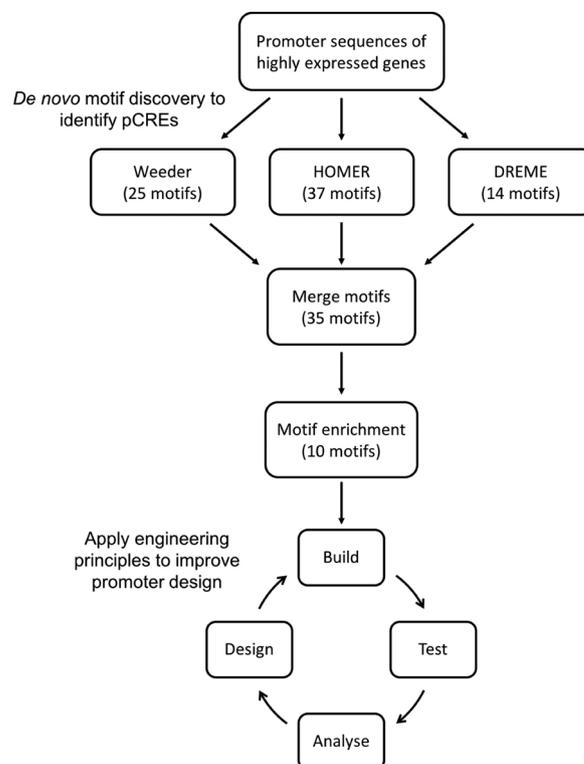

Fig. 1. Putative *cis*-regulatory element (pCRE) discovery and testing workflow.
The promoter analysis pipeline applied to identify and test pCREs. Motifs were discovered then refined computationally before *in vivo* testing.

Phytozome 10 (v5.5, DOE, Joint Genome Institute; http://phytozome.jg i.doe.gov) [25,26] using the BioMart platform [27]. Of the top 300 most highly expressed genes from the microarray dataset, which was based on an earlier version of the *C. reinhardtii* genome, 267 promoter sequences were successfully retrieved from the BioMart interface using v5.5 of the *C. reinhardtii* genome and analysed using WEEDER v2.0 [28,29], HOMER v4.9 [30], DREME v5.1.1 [31,32] and MEME v5.1.1 [33] software. Parameters for discovery software are listed in Supp. Table S2. Motifs were converted into position weight matrices (PWMs) and clustered using RSAT motif clustering software [34] at http://rsat.sb-roscoff .fr/matrix-clustering_form.cgi using default parameters. Motif enrichment was performed using AME v5.1.1 (http://meme-suite.org/tools/ ame) [35] and CentriMo v5.1.1 (http://meme-suite.org/tools/ce ntrimo) [36].

Motif reporter vector design and construction

The core and pCRE promoter fragments were generated from synthetic single stranded DNA (ssDNA) (Supp. Table S3) *via* PCR amplification. All ssDNA and primers were purchased from Life Technologies Corporation, Carlsbad, CA, US. The pOpt_mVenus_Paro [37] vector was used as the backbone to generate the pCore and pCRE reporter vectors; the AR1 promoter region upstream of mVenus was replaced with a synthetic core promoter (Fig. 2). The RbcS2 introns upstream of and within the mVenus gene were retained to enhance gene expression [38]. pOpt_mVenus_Paro was PCR amplified with primers iRbcS2_Amp_F and mVenus_EcoRI_R (Supp. Table S4) to introduce a *ClaI* restriction site upstream of the iRbcS2 intron, generating a DNA fragment containing iRbcS2 and mVenus. To generate the core promoter vector pCore_mVenus, a 50 bp ssDNA template containing the core promoter
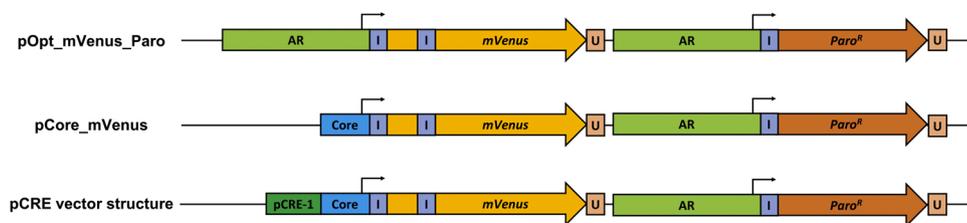
Fig. 2. mVenus reporter vectors to test putative *cis*-regulatory elements *in vivo*. Schematic diagrams of the vectors transformed into *C. reinhardtii*: pOpt_mVenus_Paro, pCore_mVenus, and pCRE-1_mVenus. For each pCRE test vector, the pCRE-1 fragment was replaced with DNA fragments containing repeats of the pCREs listed in Table 3. AR, *Hsp70A/RbcS2* promoter; I, RbcS2 intron; U, RbcS2 3'-untranslated region; *Paro^R*, paromomycin resistance gene.

sequence taken from the 3' end of the SAP-11 synthetic promoter (Supp. Table S3) [21] was PCR amplified using primers pCore_Amp_F and pCore_Amp_R (Supp. Table S4), which introduced *Xba*I and *Cla*I cut sites to the core sequence; the 70 bp core promoter fragment was gel extracted from a 3 % agarose gel. To generate the pCore_mVenus vector, pOpt_mVenus_Paro was digested with XbaI and EcoRI to remove the entire AR1 promoter and mVenus region; the 4950 bp vector backbone was gel extracted from a 1 % agarose gel. The iRbcS2_mVenus PCR fragment was digested with ClaI and EcoRI, and the pCore fragment digested with ClaI and XbaI, for insertion into the pOpt vector backbone. The pCore and iRbcS2_mVenus fragments were ligated together using T4 DNA ligase (New England Biolabs [NEB], Ipswich, MA, US), and 0.3 μL of the ligation mixture was PCR amplified using primers pCore_Amp_F and mVenus_EcoRI_R (Supp. Table S4) to generate the new fragment pCore_iRbcS2_mVenus. pCore_iRbcS2_mVenus was digested with XbaI and EcoRI and purified with a QIAquick PCR Purification kit (Qiagen, Hilden, Germany), then ligated into the cut vector backbone to generate the pCore_mVenus vector. pCore_mVenus was used as the baseline control vector, and as a scaffold for inserting the pCRE motif repeats into the proximal promoter region.

To generate the pCRE vector suite, DNA repeats of each pCRE were inserted upstream of the core promoter in pCore_mVenus. The ssDNA motif templates contained multiple copies of each motif flanked by *Sac*I and *Xba*I restriction sites for their insertion into pCore_mVenus, upstream of the pCore region (Supp. Table S3). These restriction sites were further flanked at the 5' and 3' ends with common extension sequences, enabling amplification of each ssDNA pCRE template by primers pCRE_Amp_F and pCRE_Amp_R (Supp. Table S4). Following amplification, digestion (with enzymes SacI and XbaI), and gel purification of the individual pCRE modules, pCore_mVenus was digested with the same restriction enzymes and ligated to each individual pCRE module, forming 14 pCRE_mVenus vectors (Supp. Table S5). Each pCRE vector contained the common core promoter region upstream of the mVenus reporter gene, and a distinct proximal promoter sequence upstream of the promoter core.

All vectors were propagated in *Escherichia coli* DH5α cells and prepared using a QIAprep spin miniprep kit (Qiagen). All vector sequences were confirmed *via* Sanger sequencing (Core Genomics Facility, University of Sheffield, UK). Gel extractions were performed using a QIAquick Gel Extraction kit (Qiagen). PCRs were performed using Phusion high-fidelity DNA polymerase (NEB) according to the manufacturer's instructions. All restriction enzymes were purchase from NEB. Annotated vector maps of pCore_mVenus and pCRE-1_mVenus as examples of the vector building strategy can be found in the Supplementary Data.

C. reinhardtii cultivation and transformation

The *C. reinhardtii* strain CC-125 was used as the parental strain for all transformations. All strains were grown in mixotrophic conditions on tris-acetate-phosphate (TAP) medium, with 1.5 % agar for static cultures. Cultures were grown at 25 °C with continuous illumination at 150 μmol photons m^2 s^{-1} on an orbital shaker set to 120 rpm. Growth

was monitored by measuring cell number with a Neubauer cell-counting chamber (Sigma-Aldrich, St. Louis, MO, US).

For nuclear transformations, CC-125 cultures were grown to $0.5-1 \times 10^6$ cells mL^{-1} and harvested by centrifugation at $2,000 \times g$, 15 °C, 5 min. Pellets were resuspended in GeneArt Max Efficiency Transformation Reagent (Thermo Fisher Scientific, Waltham, MA, US) to a final concentration of $1-2 \times 10^8$ cells mL^{-1}. Vectors were linearized using the restriction enzyme BsaI (except for the pCRE-8 vector, for which ScaI was used due to a *Bsa*I restriction site in the promoter region), and 2 μg of linearized vector DNA was then mixed with 250 μL cells and incubated on ice for 5–10 min. Cells were then transferred to prechilled 0.4 cm gapped electroporation cuvettes and electroporated using a BioRad GenePulser Xcell electroporator (Hercules, CA, US) using the following parameters: 500 V, 50 μF, 800 O shunt resistance, exponential decay. Cells were then transferred to 60 mM sucrose TAP and shaken overnight (14–18 h) in the dark, 25 °C. The next day, cells were harvested by centrifugation (2,000 $\times g$, 18 °C, 5 min), resuspended in 0.5–1 mL TAP and streaked on to selective TAP-agar plates containing 20 μg mL^{-1} paromomycin. Colonies appeared after incubation at 25 °C with continuous illumination at 150 μmol photons m^2 s^{-1} at ~5 days.

Flow cytometry

CC-125 cells were transformed with each pCRE vector, and the paromomycin-resistant colonies generated from each transformation experiment were scraped into 10 mL liquid TAP. For the negative control, CC-125 cells were electroporated with sterile water and streaked on to non-selective TAP-agar plates; the resulting colonies were scraped into liquid TAP culture. The liquid cultures were incubated overnight (16 h) and filtered prior to analysis. The fluorescence of individual cells was measured by flow cytometry using a BD FACSMelody cell sorter (Franklin Lakes, NJ, US). 50,000 events were measured per run. Chlorophyll fluorescence was measured by excitation with a laser at 488 nm, and detected using an emission filter of 710/45 nm. Forward and side scattering data were used to remove debris and clumps of algal cells. mVenus was measured at excitation and emission wavelengths 488 nm and 513/26 nm, respectively. Scatter plots showing the chlorophyll *vs* mVenus fluorescence signals for CC-125 and cells transformed with the AR1 positive control vector were compared to draw the mVenus gates shown in Supp. Fig. S2. Flow cytometry data analysis was performed using FlowJo software v9. One-way analysis of variance (ANOVA) with Bonferroni's multiple comparisons *post hoc* test were performed using Graphpad Prism software (v8.0) to compare mVenus fluorescence of all populations with that of the AR1 and pCore populations. The null hypothesis of no differences in expression was rejected when $P < 10^{-3}$. This conservative threshold was chosen because the differences in sample size and variance were large between groups being compared. This enabled the identification of promoters that were significantly more effective at driving mVenus expression compared to the basal pCore promoter alone.

Results

Identification of putative cis-regulatory elements (pCREs) in C. reinhardtii high-expression promoters

To identify *cis*-regulatory elements that can be used to design synthetic promoters potentially capable of driving high heterologous gene expression, the promoter regions of *C. reinhardtii* genes exhibiting the highest RNA accumulation in cells grown under conventional laboratory light and temperature conditions were analyzed [24]. The top 300 highly-expressed genes with limited variability during logarithmic growth were identified from a previously published gene expression dataset (Fig. 3) [24]; of these, 267 were present in v5.5 of the *C. reinhardtii* genome. The 1000 bp regions immediately upstream of the top 267 highly expressed genes were considered to contain the promoter sequences, and were therefore analyzed using WEEDER, HOMER and DREME software; 76 unique DNA motifs were found in total, varying in length from 5 to 14 bp (Supp. Table S6).

Many of the motifs were redundant, in that the same short sequences occurred repeatedly within longer motifs; these short sequences are likely to represent true CREs, but needed to be identified to prevent redundant motif testing *in vivo* [28,29]. To condense the motifs to their core sequences, the position weight matrices (PWMs) generated for each motif were phylogenetically compared and aggregated into motif sub-clusters using Regulatory Sequence Analysis Tool (RSAT) matrix-clustering software [34]. The 76 motifs were reduced to 35 sub-clusters, and a new PWM was generated for each merged motif representing the 'root motif' for each sub-cluster using the RSAT program. Consensus sequences of the top 20 root motifs are listed in Table 1. Clusters 1, 2, and 4 were found by all three motif discovery programs, strongly suggesting that these motifs have a role to play in gene structure and/ or regulation.

Motif clusters 1–20 (Table 1) were tested for enrichment within the promoter sequences of the top 267 genes to cross-validate the computationally-generated merged motifs, and to increase the likelihood that the motifs retained their biological relevance. Taking forward only the most enriched sequences for *in vivo* testing narrowed the design space, and increased the likelihood of discovering genuine CREs. Two programs were used to test for motif enrichment: Analysis of Motif
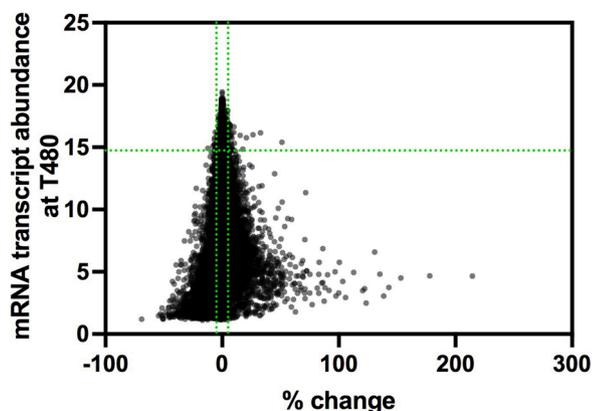

Fig. 3. Selection of genes for promoter analysis.
Scatter plot showing transcript abundance of 11,455 genes after 480 min of growth *vs* percentage change in transcript abundance between 0 min and 480 min timepoints. Data taken from [24]. Vertical green dotted lines represent 5% change in expression threshold. Data points above the horizontal green dotted line were classed as highly expressed and considered for promoter analysis. Data points representing the genes selected from promoter analysis fell between the constraints depicted in green.

Table 1
Consensus sequences of the top 20 clustered motifs.

Motif cluster	Forward sequence	Reverse sequence	Merged motifs
cluster_1	TGCCGTACGA	TCGTACGGCA	DREME_13, Homer_7, Weeder_17, Weeder_8, DREME_3, DREME_7
cluster_2*	GCCCCATKCAGG	CCTGMATGGGGC	Homer_6, Homer_3, DREME_8, Weeder_2, Weeder_3, Weeder_14, Weeder_20, Homer_2, Weeder_10, Weeder_6
cluster_3	CGAGAGVC	GBCTCTCG	Weeder_18, Weeder_21, Weeder_11, Weeder_12, Weeder_9
cluster_4*	GHGAAAGARRGAGA	TCTCYYTCTTTCDC	DREME_10, DREME_2, Homer_29, Homer_1, Weeder_1, Weeder_13, DREME_4, Weeder_4, Weeder_5
cluster_5	CCTSGCC	GGCSAGG	DREME_12, DREME_5
cluster_6	SRGTMCCCC	GGGGKACYS	Homer_36, Weeder_16, Homer_28, Weeder_15, Weeder_7
cluster_7	CTCCAGGKTA	TAMCCTGGAG	DREME_6, Homer_10
cluster_8	TGTAGSCAGG	CCTGSCTACA	Homer_35, Weeder_23, Weeder_25
cluster_9*	TRTGYAGG	CCTRCAYA	DREME_14, DREME_1, DREME_11, Weeder_24
cluster_10*	CTCGGT	ACCGAG	Weeder_22
cluster_11	CRGTWCSGTGTG	CACACSGWACYG	Homer_21, Homer_34
cluster_12*	CCMTCKCGMSCVA	TBGSKCGMGAKGG	Homer_18, Homer_16, Homer_4
cluster_13*	GTATGCHTGCTG	CAGCADGCATAC	Homer_21, Homer_34
cluster_14	CCMTCKCGMSCVA	TBGSKCGMGAKGG	Homer_18, Homer_16, Homer_4
cluster_15	ACGCGGGGTA	TACCCCGCGT	Homer_13
cluster_16	AACCASGGYTAG	CTARCCSTGGTT	Homer_31
cluster_17*	GTCCACCTGG	CCAGGTGGAC	Homer_30
cluster_18	SATSSACCAGGW	WCCTGGTSSATS	Homer_8
cluster_19	GCCCTYCCAAGG	CCTTGGRAGGGC	DREME_9, Homer_9
cluster_20*	CGAGCGTTTTCT	AGAAAACGCTCG	Homer_20

Position weight matrices for all 35 motif clusters are listed in Supp. Table S7. Motifs taken forward for further analysis are starred with an asterisk.

Enrichment (AME) and CentriMo [35,36].

AME identifies user-provided motifs that are relatively enriched in a given set of promoter sequences compared to a control set [35]. For this experiment, the inputted promoter sequences were shuffled to create the control. Seven of the merged motifs were significantly enriched within the highly expressed promoter set relative to the shuffled control (Table 2). CentriMo, similarly to AME, identifies relatively enriched motifs within a sequence set, but additionally determines whether a motif has a particular bias towards a location within a given set of

Table 2

Motif enrichment using AME and CentriMo.

Cluster ID	Consensus	CentriMo		AME
		E-value	Bin center position from TSS	*p*-value
cluster_2	GCCCCATGCARG	9.40E-12	−118.5	1.28E-07
cluster_4	GHGAAAGARRGAGA	6.40E-19	−36.5	2.84E-20
cluster_5	CCTCGCC	–	–	1.74E-10
cluster_9	TRTGYAGG	–	–	4.73E-06
cluster_10	CTCGGT	3.80E+00	−233.5	–
cluster_12	CCMTCKCGMSCVA	–	–	7.65E-12
cluster_13	GTATGCHTGCTG	–	–	7.49E-05
cluster_17	GTCCACCTGG	8.90E+00	−762	–
cluster_18	SATGSACCAGGW	–	–	1.31E-04
cluster_20	CGAGCGTTTCT	7.00E+00	−305	–

Inputted motifs: clusters 1–20. Only motifs with CentriMo *E*-values <10 and AME *p*-values < 0.05 (Fisher's exact test) are shown. Bin centre from transcription start site (TSS) represents the centre of the site at which the motif could be found with the highest probability.

sequences of the same length [36]. Five merged motifs were found to be enriched with a positional bias relative to the TSS by CentriMo, two of which (Clusters 2 and 4) were also found to be enriched using the AME program. Cluster_2 has a strong positional bias around –118 bp from the TSS, whereas Cluster_4 is highly likely to be found –37 bp from the TSS (Table 2; Supp. Fig. S1). Clusters 10, 17 and 20 were found to have statistically significant positional biases further upstream from the TSS within the promoter regions of highly expressed genes (Table 2).

The ten motif clusters found to be significantly enriched within the highly expressed promoter sequences (Table 2) were selected for *in vivo* analysis. In addition to these, the previously discovered 'CCCATGCGA' motif was selected for individual motif analysis, as well as the 'GGGCCCATTC' and 'CGCATGGGGC' motifs discovered by analysing 25 high expression synthetic promoters using Multiple Expectation maximization for Motif Elicitation (MEME) motif discovery (Supp. Table S8) [21,33]. A random 10 bp DNA sequence with a similar GC content to the other motifs was generated as a control; this motif was not significantly enriched in the top promoter sequences (AME *p*-value = 1, no sequence matches). Fig. 4 displays the PWMs of all motifs selected for *in vivo* testing in the form of sequence logos. All motifs were renamed putative *Cis*-Regulatory Element (pCRE) 1–13, plus pCRE-random (RM).

In vivo testing of pCRE modules in C. reinhardtii

To test the selected pCRE motifs for promoter activity *in vivo*, a fluorescent protein reporter system was designed (Fig. 4). Fluorescent proteins such as mVenus have been extensively applied as simple but effective tools for measuring gene expression in *C. reinhardtii*, and are relatively easy to detect without requiring expensive reagents [21, 39–41]. Achieving transgene protein expression levels in *C. reinhardtii* that are high enough to detect by Western blot, let alone to levels high enough to be industrially/commercially viable, can be challenging; this is despite the relative ease of generating drug-resistant transformants that exhibit transgene mRNA expression levels that are detectable *via* real-time quantitative PCR [42,43]. Therefore, the intensity of a fluorescent reporter protein was chosen for use as a more conservative and realistic measure of the potential activity of a *cis*-regulatory element *in vivo*, rather than measuring transcript levels. The pOpt_mVenus_Paro vector was used as a backbone for building the synthetic promoter reporter constructs [37]. The AR1 promoter region of pOpt_mVenus_Paro was removed and replaced with a 50 bp fragment taken from the 3' end

of the previously characterized SAP-11 synthetic promoter to generate a core promoter (pCore), with new restriction enzymes sites included to enable the insertion of proximal promoter elements upstream of the pCore region (Fig. 2 & Supp. Fig. 2) [17]. We determined that this core promoter drove mVenus fluorescent protein expression at low levels when transformed in to *C. reinhardtii*, but above the autofluorescence levels of the un-transformed control; thus, this sequence contains the minimum sequences necessary to recruit basal transcription machinery [17]. The mVenus expression driven by the pCore promoter "empty vector" was therefore adopted as the baseline-control for measuring the activity of other CREs tested in this study.

DNA fragments of ~70 bp, each containing concatemers of the bioinformatically identified motifs shown in Fig. 4 were then synthesized for insertion upstream of the pCore basal promoter in pCore_mVenus to produce individual reporter vectors (Supp. Table S3, S5). The consensus sequence for each pCRE motif was synthesized as ssDNA containing 4–7 copies of each motif, keeping the promoter length within a 70 bp size constraint (Table 3). In yeast, increasing the copy number of motifs has a positive effect on transcription, then tends to saturate after ~4 copies; therefore, at least four repeats of each motif were generated in *C. reinhardtii* [44].

The random integration of transgenes in *C. reinhardtii* results in high variability of transgene expression level due to positional effects. Therefore, several hundred paromomycin resistant colonies were pooled per transformation and examined by flow cytometry to enable the robust characterisation of reporter expression for each transformant population. Increasingly, flow cytometry is being coupled with fluorescence-activated cell sorting to rapidly select for wanted phenotypes in microalgae [45–47], while direct fluorescent protein fusions or co-reporters are being used to select for useful metabolic enzymes and high-value proteins [48–50]. Therefore, practically speaking, considering the full distribution of the transformants' expression levels, while rapidly and efficiently selecting for the top expressers using high-throughput methods, is relevant to the current state of the field [21,51,52].

CC-125 was transformed with vectors containing the motifs listed in Fig. 4. (For list of vectors, see Supp. Table S5). Several hundred transformant colonies for each vector were selected on 20 µg mL^{-1} paromomycin agar plates and scraped into liquid TAP media at shake-flask scale after 5 days. The cells were incubated for 24 h in liquid media to allow for disassociation of colony biomass into single cells, after which flow cytometry was performed. Untransformed CC-125 was used as the "wild type" negative control, and to characterize baseline chlorophyll autofluorescence; this population was electroporated without any DNA and was cultured on TAP media without antibiotic selection. Cells transformed with the pOpt_mVenus_Paro vector containing the original AR1 promoter (AR1 cells) were used as a positive control for mVenus fluorescence detection. Two populations exhibiting different chlorophyll fluorescence intensities were present in each run (Supp. Fig. S2); the lower chlorophyll fluorescence population was excluded from the analysis, as it most likely represented dead or dying biomass due to paromomycin selection [52]. The low chlorophyll population was essentially absent from the CC-125 culture, suggesting that this batch of cells was less stressed and represented a healthier population (Supp. Fig. S2). The mVenus population was pronounced in the AR1 strain as expected (Supp. Fig. S2B), suggesting strong reporter gene expression. The flow cytometry scatter plots generated for untransformed CC-125 and AR1 transformed cells were compared to draw the mVenus gates shown in Supp. Fig. S2. mVenus populations were visibly present within all other transformed strains, indicating reporter gene expression.

The mean fluorescence intensities (FI) of all flow cytometry events captured within the mVenus gate for each population (Supp. Fig. S2) were compared using a One-Way ANOVA with a Bonferroni's *post hoc* test to compare the bioinformatically-identified CREs to the core promoter (negative control) and the AR1 promoter (positive control; F[15, 4784] = 49, $p = 10^{-4}$). Groups were considered to be statistically significantly different when $p < 10^{-3}$ following primary ANOVA and
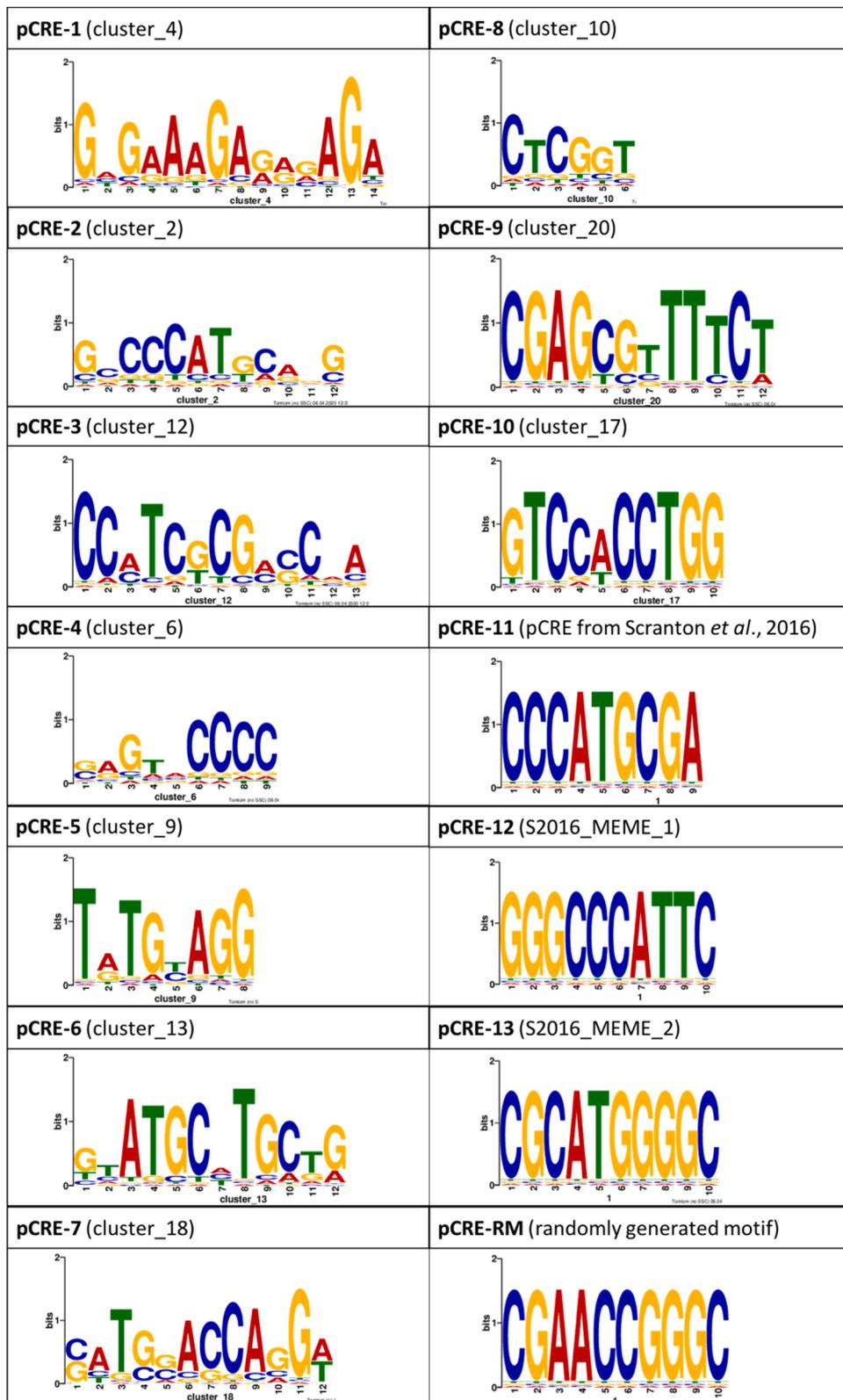
Fig. 4. pCRE motifs selected for *in vivo* analysis.
Position weight matrices for each motif are represented by sequence logos.

Table 3
Motif consensus sequence repeats tested *in vivo.*

Motif	Consensus forward	Motif length (bp)	No. of repeats	Promoter length (bp)
pCRE-1	TCTCTCTCTT	10	6	66
pCRE-2	GCCCCATGAGG	11	5	65
pCRE-3	TTGGTCGCGATGG	13	5	75
pCRE-4	GGGGTACTC	9	7	73
pCRE-5	TATGTAGG	8	7	66
pCRE-6	GCATGCATGCTG	12	5	70
pCRE-7	CATGGACCAGGA	12	5	70
pCRE-8	CTCGGT	6	7	52
pCRE-9	CGAGCGTTTTCT	12	5	70
pCRE-10	GTCCACCTGG	10	6	70
pCRE-11	CCCATGCGA	9	7	73
pCRE-12	GGGCCCATTC	10	6	70
pCRE-13	CGCATGGGGC	10	6	70
pCRE-RM	CGAACCGGGC	10	6	70

The forward DNA consensus sequences were used to construct the test vectors. # repeats = number of motif repeats in the test promoter sequence. The full promoter region used to calculate GC content comprised the motif repeats, core promoter and restriction sites.

Bonferroni's *post hoc* tests; this conservative threshold was applied to account for variability in the sample sizes and a lack of homoscedasticity, and normality of the data, while using a robust statistical test. The AR1 mVenus population exhibited a significantly higher mean FI compared to all other promoter populations (Fig. 5). Comparisons between the pCore mVenus population and all other populations revealed that mVenus expression was significantly higher in the pCRE-5, -6, -9, -12, and -13 promoter populations (Fig. 5A). The randomly generated pCRE-RM was not significantly different from the core promoter negative control. Collectively, this shows that a subset of the pCREs placed upstream of the pCore promoter generated reporter gene expression and that sequence-dependence was likely displayed, since the randomly-
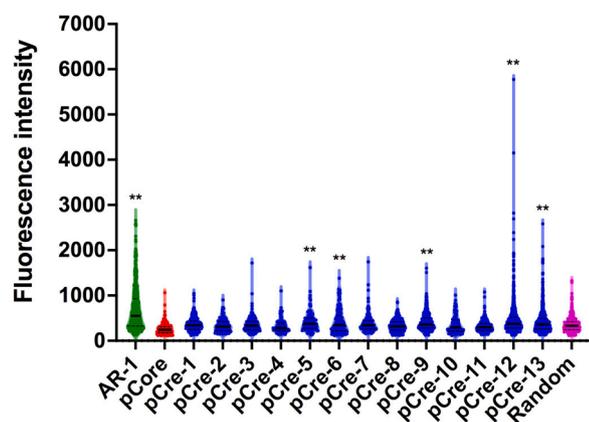

Fig. 5. Violin plot showing the fluorescence intensities (FIs) of all events detected within the mVenus flow cytometry gate.
Black bars represent median values, dashed lines represent upper and lower quartiles. Data points represent the intensity values of each individual mVenus fluorescence event detected. Each population was compared to the pCore population using an ordinary One-Way ANOVA and a Bonferroni's multiple comparisons *post hoc* test. Asterisks represent significant differences in FI from the core promoter population (red). $*p < 1 \times 10^{-3}$, $**p < 1 \times 10^{-4}$. Differences of $p < 1 \times 10^{-3}$ from were deemed to represent significantly stronger expression compared to the core promoter alone. n = 93–1052.

generated sequence and several of the tested pCREs did not generate significantly higher reporter expression.

Since high-throughput screening methods are now consistently used when screening *C. reinhardtii* transformants for recombinant transgene expression, it is reasonable to consider the expression levels of the top 10 % of any given transformant population, such as the top 10 clones when primary transformants are assayed in 96-well microtitre plates or individual cells enriched using fluorescence-activated cell sorting and then characterized in down-selection steps [21,49,52,53]. Therefore, the fluorescence reporter transgene levels of the top 10 % of mVenus positive cells as determined by flow cytometry were then compared for each promoter (Table 4). Compared to the pCore population, pCRE-12 and -13 were again identified as robust drivers of transgene expression; however, the top 10 % of the pCRE-12 and -13 populations still displayed significantly lower mVenus expression compared with that of the much larger AR1 promoter (One-Way ANOVA with Bonferroni's *post hoc* test; F[15, 466] = 27.65, $p < 10^{-4}$, *post hoc* comparison $p < 10^{-3}$ between pCRE-12 or pCRE-13 and AR1). The maximum FI values of pCRE-12 and -13 (5775 and 2582, respectively) were comparable to, or higher than, that of the AR1 promoter (2660; Table 4). By comparison, pCRE-5, -6, and -9 were not consistently strong or robust drivers of reporter expression. While five out of the thirteen pCREs drove some level of reporter activity above the baseline core promoter, only two achieved expression levels that approached those of AR1.

Discussion

In this study, the promoter regions of genes exhibiting high constitutive expression were analyzed *in silico* with the aim of finding *cis*-regulatory elements that can be used to generate high-expression synthetic promoters in the emerging biotechnology host cell, *C. reinhardtii*. The computationally-identified pCREs were tested *in vivo* for their ability to induce the expression of a fluorescent protein using flow cytometry, leading to the identification of active elements in *C. reinhardtii* constitutive promoters. Five pCRE candidates exhibited mVenus expression higher than the core promoter baseline. Two of these promoters (pCRE-12 and pCRE-13) generated expression levels where the upper decile of transformants matched or exceeded the strongest constitutive promoter, AR1, albeit with lower mean overall expression levels when all mVenus expressing transformants were analyzed. Collectively, this suggests that the pCRE-12 and pCRE-13 elements can drive high expression levels; however, they are strongly influenced by other factors such as genome positional effects, while AR1 more consistently generates high expression. Given that the AR1 promoter comprises a 497 bp fusion of two of the strongest native *C. reinhardtii* promoters, *Hsp70A* and *RbcS2*, which contain multiple endogenous CREs, this is unsurprising [54]. CRE-12 and -13 are only 116 bp and 122 bp in size and represent a single putative *cis*-regulatory element concatemerized as 6x repeats (Table 3). Nevertheless, achieving strong reporter expression from concatemerized single motifs suggests that the nucleotide footprint of promoter sequences can be substantially reduced to drive useful recombinant protein expression levels and that these motifs could be further combined to generate even stronger synthetic promoters in future studies. This work has shown that through interrogation and use of open source data and software, genomic data can be mined for regulatory motifs that can effectively drive protein expression.

pCRE-12 produced the highest median fluorescence intensity of the pCRE suite and the highest overall single fluorescence intensity event of all the promoters tested, including AR1. This motif was discovered by scanning 25 active synthetic promoters for the most common sequences with MEME (Supp. Fig. S8) [21]. This provided more evidence for pCRE-12 being a genuine *cis*-regulatory element. pCRE-11 (CCCATGCGA), a motif discovered to be essential for high expression in a strong synthetic promoter, was expected to induce strong mVenus expression, but instead mVenus levels were not significantly higher than the core promoter (Fig. 5) [21]. Previous work suggested that the

Table 4
Descriptive statistics for mVenus (YFP) fluorescence detection *via* flow cytometry.

Promoter	Number of YFP-expressing transformants	% of total events in YFP gate	YFP fluorescence, all events (RFU)		YFP fluorescence, top 10 % of events (RFU)		Max value (RFU)
			Median	Mean	Median	Mean	
No vector (WT)	6	0.01	269	363.7	–	–	871.4
AR	551	1.10	550.4	693	1686	1783	2660
pCore	202	0.40	243.1	262.5	436.5	513.9	1060
pCRE-1	331	0.66	339.2	361.5	629.9	670.6	1044
pCRE-2	105	0.21	316.1	345.8	610.8	648.8	900.7
pCRE-3	188	0.38	333.6	354.6	591.7	686.7	1717
pCRE-4	95	0.19	282	326	632.3	662.4	1100
pCRE-5	100	0.20	374.5	424.8	855.1	924.3	1618
pCRE-6	137	0.27	344.7	427.7	1030	1019	1380
pCRE-7	123	0.25	340.8	390	742.7	894.5	1743
pCRE-8	388	0.78	317.3	333.9	589.4	588.8	856.2
pCRE-9	287	0.57	363	419.5	807	861.8	1602
pCRE-10	93	0.19	293.1	353.1	759.3	790.9	1006
pCRE-11	236	0.47	300.3	322.2	533.4	586.4	1077
pCRE-12	1052	2.10	375.1	458.6	965.8	1177	5775
pCRE-13	537	1.07	360.9	427.5	967.4	1092	2582
pCRE-RM	375	0.75	329.6	356.7	640.2	725.3	1328

YFP, yellow fluorescent protein/mVenus; RFU, relative fluorescence units.

pCRE-11 motif is necessary, but not sufficient, for high expression in the strong synthetic promoter SAP-11 [21]. The pCRE-11 motif was also present in non-functional synthetic promoters [21]. Collectively, this suggests that the modest mVenus expression observed in the flow cytometry analysis for pCRE-11 in this study was probably due to the requirement of this motif to have other DNA elements present to drive strong reporter activity (Fig. 5). Combining pCRE-11 with other CREs could perhaps determine other elements to which this motif can be functionally paired.

The 'CCCAT' motif (present within pCRE-11 and also common to pCRE-2, -12 and -13), was presented as a core motif in other synthetic promoters shown to function in *C. reinhardtii*, but only two of the four pCREs containing this motif (pCRE-12 and pCRE-13) drove high mVenus expression (Fig. 5) [21]. The bases flanking this motif may influence transcription factor binding; pCRE-12 and -13 retain the common sequence GNCCATNC, which could be explored further in variations.

Three algorithmically different *de novo* motif discovery programs were used to mine potential CREs from *C. reinhardtii* promoters. This ensemble consensus approach improved the likelihood of finding statistically enriched motifs, and reduced the chances of finding false positives [55]. In some plant promoter studies, at least 5 motif discovery programs were integrated for motif discovery [19,56]. Although this was in part due to the scale of the studies, in which the promoter sequences of thousands of co-regulated genes were scanned, perhaps including more discovery programs in this pipeline could improve the specificity of the enriched consensus sequences. The most effective motifs (pCRE-12 and pCRE-13) were discovered by scanning the strong computationally-generated synthetic promoters [21]. One reason for this could be that through scanning high-performing synthetic promoters, whose motifs were computationally designed to mimic common consensus sequences but with slight variations, the PWMs for the motif consensus sequences identified in the MEME search in this study were refined and optimized for increased promoter activity.

This study employed a bottom-up approach towards developing synthetic promoters by identifying *in silico* and then testing individual putative promoter elements. Although not all the pCREs tested were effective, increased activation of a basal promoter was achieved by several of the synthetic pCRE concatemers displaying a broad activity range. Moreover, this was attained using minimal promoters of ~70 bp in size, some of which were able to drive gene expression to levels comparable to the AR1 promoter, the latter being a hybrid of two endogenous promoters with a combined length of ~500 bp and derived from two of the highest expressed genes in the *C. reinhardtii* genome. The

use of minimal pCRE promoters reduced the nucleotide footprint of the promoter region, and therefore the size of the transformation cassette.

Including tandem motif repeats in promoters can yield high expression [18,44,57], although this is likely to be more effective when aiming for transactivation by a specific transcription factor. Although the tandem repeat structures did function relatively well as standalone promoters, the motifs probably lacked the synergistic effects of other motif and non-motif elements as well as spatial context, thus limiting the expression levels achievable with this system. The next step would be to examine the reduced promoter parts as modules, by combining the pCREs to construct novel, 'build-your-own' promoters. This has the potential to generate more robust synthetic promoters, given that CREs can function sub-optimally in isolation and may require other CREs to induce expression [17,58]. Furthermore, recent models of *cis*-elements suggest that gene activation in eukaryotes is achieved *via* the cumulative effect of multiple CREs, as well as non-motif DNA elements [59,60]. Examples of these phenomena in *C. reinhardtii* include mutually necessary heat shock elements in the Hsp70A enhancer [61], the 'CCCATGCGA' motif, which is necessary for high expression in some but not other previously-developed synthetic promoters [21]. Combining motifs could highlight synergistic or antagonistic effects of different CRE combinations, as well as provide insight into positional effects of CRE function with respect to the transcription and translation start sites.

Developing a more comprehensive understanding of core promoter elements and structure in *C. reinhardtii* could further facilitate promoter building and optimization, and lead to more promoter parts for tailored promoter design. In other host organisms, viral promoter parts are often used as strong core promoters [62–64]. To date, no known viruses can infect *C. reinhardtii* [65], and the strong plant viral promoter CaMV35 drives poor gene expression [66,67]. Future core promoter design could be performed *in silico*, similar to the core promoter used in this study, but with more in-depth and focused analysis of the core promoter region [21]. Identifying core promoter structural and functional elements in *C. reinhardtii* and testing them in different combinations, positions and copy numbers would be an obvious next step.

This study has provided some leads for building optimized *C. reinhardtii* synthetic promoter modules, although more cycles of design-build-test-learn will need to be implemented to optimize promoter motifs to a competitive standard. This work has expanded the *C. reinhardtii* promoter repertoire, and provided some insight into gene regulation in algae through the discovery of novel CREs.

Availability of data and materials

All data generated or analyzed during this study are included in this article and its supplementary information files.

Funding

Authors' contributions

JM designed and performed the experimental work, analyzed the data and prepared the manuscript; AB prepared, edited and revised the manuscript; AS provided technical advice and assisted in experimental design; SM provided research direction; JP supervised the study, edited and revized the manuscript. All authors have read and approved the final manuscript.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.nbt.2022.01.001.

References

[1] Borowitzka MA. High-value products from microalgae-their development and commercialization. J Appl Phycol 2013;25:743–56. https://doi.org/10.1007/s10811-013-9983-9.

[2] Rasala BA, Mayfield SP. Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. Photosynth Res 2015;123:227–39. https://doi.org/10.1007/s11120-014-9994-7.

[3] Keller LC, Romijn EP, Zamora I, Yates JR, Marshall WF. Proteomic analysis of isolated Chlamydomonas centrioles reveals orthologs of ciliary-disease genes. Curr Biol 2005;15:1090–8. https://doi.org/10.1016/j.cub.2005.05.024.

[4] Kianianmomeni A, Hallmann A. Algal photoreceptors: in vivo functions and potential applications. Planta 2014;239:1–26. https://doi.org/10.1007/s00425-013-1962-5.

[5] Rochaix JD, Lemeille S, Shapiguzov A, Samol I, Fucile G, Willig A, et al. Protein kinases and phosphatases involved in the acclimation of the photosynthetic apparatus to a changing light environment. Philos Trans R Soc B Biol Sci 2012;367:3466–74. https://doi.org/10.1098/rstb.2012.0064.

[6] Umen JG. Evolution of sex and mating loci: an expanded view from Volvocine algae. Curr Opin Microbiol 2011;14:634–41. https://doi.org/10.1016/j.mib.2011.10.005.

[7] Salomé PA, Merchant SS. A series of fortunate events: introducing Chlamydomonas as a reference organism. Plant Cell 2019;31:1682–707. https://doi.org/10.1105/tpc.18.00952.

[8] Rasala BA, Mayfield SP. Photosynthetic biomanufacturing in green algae; Production of recombinant proteins for industrial, nutritional, and medical uses. Photosynth Res 2015;123:227–39. https://doi.org/10.1007/s11120-014-9994-7.

[9] Gimpel JA, Hyun JS, Schoepp NG, Mayfield SP. Production of recombinant proteins in microalgae at pilot greenhouse scale. Biotechnol Bioeng 2015;112:339–45. https://doi.org/10.1002/bit.25357.

[10] Gregory JA, Li F, Tomosada LM, Cox CJ, Topol AB, Vinetz JM, et al. Algae-produced pfs25 elicits antibodies that inhibit malaria transmission. PLoS One 2012;7:1–10. https://doi.org/10.1371/journal.pone.0037179.

[11] Surzycki R, Greenham K, Kitayama K, Dibal F, Wagner R, Rochaix JD, et al. Factors effecting expression of vaccines in microalgae. Biologicals 2009;37:133–8. https://doi.org/10.1016/j.biologicals.2009.02.005.

[12] Cerutti H, Johnson AM, Gillham NW, Boynton JE. Epigenetic silencing of a foreign gene in nuclear transformants of Chlamydomonas. Plant Cell 1997;9:925–45. https://doi.org/10.1105/tpc.9.6.925.

[13] Schroda M, Blöcker D, Beck CF. The HSP70A promoter as a tool for the improved expression of transgenes in Chlamydomonas. Plant J 2000;21:121–31. https://doi.org/10.1046/j.1365-313X.2000.00652.x.

[14] Sizova I, Fuhrmann M, Hegemann P. A Streptomyces rimosus aphVIII gene coding for a new type phosphotransferase provides stable antibiotic resistance to Chlamydomonas reinhardtii. Gene 2001;277:221–9. https://doi.org/10.1016/S0378-1119(01)00616-3.

[15] Fischer N, Rochaix JD. The flanking regions of PsaD drive efficient gene expression in the nucleus of the green alga Chlamydomonas reinhardtii. Mol Genet Genomics 2001;265:888–94. https://doi.org/10.1007/s004380100485.

[16] Johnson AO, Gonzalez-Villanueva M, Tee KL, Wong TS. An engineered constitutive promoter set with broad activity range for Cupriavidus necator H16. ACS Synth Biol 2018;7:1918–28. https://doi.org/10.1021/acssynbio.8b00136.

[17] Gertz J, Siggia ED, Cohen BA. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature 2009;457:215–8. https://doi.org/10.1038/nature07521.

[18] Brown AJ, Sweeney B, Mainwaring DO, James DC. Synthetic promoters for CHO cell engineering. Biotechnol Bioeng 2014;111:1638–47. https://doi.org/10.1002/bit.25227.

[19] Koschmann J, Machens F, Becker M, Niemeyer J, Schulze J, Bülow L, et al. Integration of Bioinformatics and synthetic promoters leads to the discovery of novel elicitor-responsive cis-regulatory sequences in Arabidopsis. Plant Physiol 2012;160:178–91. https://doi.org/10.1104/pp.112.198259.

[20] Venter M. Synthetic promoters: genetic control through cis engineering. Trends Plant Sci 2007;12:118–24. https://doi.org/10.1016/j.tplants.2007.01.002.

[21] Scranton MA, Ostrand JT, Georgianna DR, Lofgren SM, Li D, Ellis RC, et al. Synthetic promoters capable of driving robust nuclear gene expression in the green alga Chlamydomonas reinhardtii. Algal Res 2016;15:135–42. https://doi.org/10.1016/j.algal.2016.02.011.

[22] Davis IW, Benninger C, Benfey PN, Elich T. Powrs: position-sensitive motif discovery. PLoS One 2012;7. https://doi.org/10.1371/journal.pone.0040373.

[23] Scaife MA, Nguyen GTDT, Rico J, Lambert D, Helliwell KE, Smith AG. Establishing Chlamydomonas reinhardtii as an industrial biotechnology host. Plant J 2015;82:532–46. https://doi.org/10.1111/tpj.12781.

[24] Mettler T, Mühlhaus T, Hemme D, Schöttler MA, Rupprecht J, Idoine A, et al. Systems analysis of the response of photosynthesis, metabolism, and growth to an increase in irradiance in the photosynthetic model organism Chlamydomonas reinhardtii. Plant Cell 2014;26:2310–50. https://doi.org/10.1105/tpc.114.124537.

[25] Merchant S. The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science (80-) 2007;318:245–51. https://doi.org/10.1126/science.1143609.

[26] Blaby IK, Blaby-Haas CE, Tourasse N, Hom EFY, Lopez D, Aksoy M, et al. The Chlamydomonas genome project: a decade on. Trends Plant Sci 2014;19:672–80. https://doi.org/10.1016/j.tplants.2014.05.008.

[27] Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, et al. BioMart - biological queries made easy. BMC Genomics 2009;10:1–12. https://doi.org/10.1186/1471-2164-10-22.

[28] Pavesi G, Mereghetti P, Mauri G, Pesole G. Weeder web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. Nucleic Acids Res 2004;32:199–203. https://doi.org/10.1093/nar/gkh465.

[29] Zambelli F, Pesole G, Pavesi G. Using Weeder, Pscan, and PscanChIP for the discovery of enriched transcription factor binding site motifs in nucleotide sequences. Curr Protoc Bioinf 2014;47(2). https://doi.org/10.1002/0471250953.bi0211s47. 11.1-31.

[30] Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010;38:576–89. https://doi.org/10.1016/j.molcel.2010.05.004.

[31] Bailey TL. DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics 2011;27:1653–9. https://doi.org/10.1093/bioinformatics/btr261.

[32] Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res 2015;43:W39–49. https://doi.org/10.1093/nar/gkv416.

[33] Bailey TL, Elkan C. Unsupervised learning of multiple motifs in biopolymers using expectation maximization. Mach Learn 1995;21:51–80. https://doi.org/10.1023/A:1022617714621.

[34] Castro-Mondragon JA, Jaeger S, Thieffry D, Thomas-Chollier M, Van Helden J. RSAT matrix-clustering: dynamic exploration and redundancy reduction of transcription factor binding motif collections. Nucleic Acids Res 2017;45:1–13. https://doi.org/10.1093/nar/gkx314.

[35] McLeay RC, Bailey TL. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics 2010;11. https://doi.org/10.1186/1471-2105-11-165.

[36] Bailey TL, MacHanick P. Inferring direct DNA binding from ChIP-seq. Nucleic Acids Res 2012;40:1–10. https://doi.org/10.1093/nar/gks433.

[37] Lauersen KJ, Kruse O, Mussgnug JH. Targeted expression of nuclear transgenes in Chlamydomonas reinhardtii with a versatile, modular vector toolkit. Appl Microbiol Biotechnol 2015;99:3491–503. https://doi.org/10.1007/s00253-014-6354-7.

[38] Lumbreras V, Stevens DR, Purton S. Efficient foreign gene expression in Chlamydomonas reinhardtii mediated by an endogenous intron. Plant J 1998;14:441–7. https://doi.org/10.1046/j.1365-313X.1998.00145.x.

[39] Fuhrmann M, Oertel W, Hegemann P. A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in Chlamydomonas reinhardtii. Plant J 1999;19:353–61. https://doi.org/10.1046/j.1365-313X.1999.00526.x.

[40] Onishi M, Pringle JR. Robust transgene expression from bicistronic mRNA in the green alga Chlamydomonas reinhardtii. G3 Genes, Genomes. Genet 2016;6: 4115–25. https://doi.org/10.1534/g3.116.033035.

[41] Baier T, Kros D, Feiner RC, Lauersen KJ, Müller KM, Kruse O. Engineered fusion proteins for efficient protein secretion and purification of a human growth factor from the green microalga Chlamydomonas reinhardtii. ACS Synth Biol 2018;7: 2547–57. https://doi.org/10.1021/acssynbio.8b00226.

[42] Baier T, Wichmann J, Kruse O, Lauersen KJ. Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga Chlamydomonas reinhardtii. Nucleic Acids Res 2018;46:6909–19. https://doi.org/10.1093/nar/gky532.

[43] Baier T, Jacobebbinghaus N, Einhaus A, Lauersen KJ, Kruse O. Introns mediate post-Transcriptional enhancement of nuclear gene expression in the green microalga Chlamydomonas reinhardtii. PLoS Genet 2020;16:1–21. https://doi.org/10.1371/journal.pgen.1008944.

[44] Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, et al. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat Biotechnol 2012;30:521–30. https://doi.org/10.1038/nbt.2205.

[45] Fields FJ, Ostrand JT, Tran M, Mayfield SP. Nuclear genome shuffling significantly increases production of chloroplast-based recombinant protein in Chlamydomonas reinhardtii. Algal Res 2019:41. https://doi.org/10.1016/j.algal.2019.101523.

[46] Terashima M, Freeman ES, Jinkerson RE, Jonikas MC. A fluorescence-activated cell sorting-based strategy for rapid isolation of high-lipid Chlamydomonas mutants. Plant J 2015;81:147–59. https://doi.org/10.1111/tpj.12682.

[47] Xie B, Stessman D, Hart JH, Dong H, Wang Y, Wright DA, et al. High-throughput fluorescence-activated cell sorting for lipid hyperaccumulating Chlamydomonas reinhardtii mutants. Plant Biotechnol J 2014;12:872–82. https://doi.org/10.1111/pbi.12190.

[48] Berndt A, Smalley T, Ren B, A B, Sproles A, Fields F, et al. Recombinant production of a functional SARS-CoV-2 spike receptor binding domain in the green algae Chlamydomonas reinhardtii. BioRxiv 2021. https://doi.org/10.1101/2021.01.29.428890.

[49] Lauersen KJ, Baier T, Wichmann J, Wördenweber R, Mussgnug JH, Hübner W, et al. Efficient phototrophic production of a high-value sesquiterpenoid from the eukaryotic microalga Chlamydomonas reinhardtii. Metab Eng 2016;38:331–43. https://doi.org/10.1016/j.ymben.2016.07.013.

[50] Lauersen KJ, Wichmann J, Baier T, Kampranis SC, Pateraki I, Møller BL, et al. Phototrophic production of heterologous diterpenoids and a hydroxy-functionalized derivative from Chlamydomonas reinhardtii. Metab Eng 2018;49: 116–27. https://doi.org/10.1016/j.ymben.2018.07.005.

[51] Rasala BA, Barrera DJ, Ng J, Plucinak TM, Rosenberg JN, Weeks DP, et al. Expanding the spectral palette of fluorescent proteins for the green microalga Chlamydomonas reinhardtii. Plant J 2013;74:545–56. https://doi.org/10.1111/tpj.12165.

[52] Barjona do Nascimento Coutinho P, Friedl C, Buchholz R, Stute SC. Chemical regulation of Fea1 driven transgene expression in Chlamydomonas reinhardtii. Algal Res 2017;26:323–9. https://doi.org/10.1016/j.algal.2017.08.006.

[53] Wichmann J, Baier T, Wentnagel E, Lauersen KJ, Kruse O. Tailored carbon partitioning for phototrophic production of (E)-α-bisabolene from the green microalga Chlamydomonas reinhardtii. Metab Eng 2018;45:211–22. https://doi.org/10.1016/j.ymben.2017.12.010.

[54] Lodha M, Schulz-Raffelt M, Schroda M. NOTE a new assay for promoter analysis in Chlamydomonas reveals roles for heat shock elements and the TATA Box in HSP70A promoter-mediated activation of transgene expression. Eukaryot Cell 2008;7:172–6. https://doi.org/10.1128/EC.00055-07.

[55] D'haeseleer P. How does DNA sequence motif discovery work? Nat Biotechnol 2006;24:959–61. https://doi.org/10.1038/nbt0806-959.

[56] Zou C, Sun K, Mackaluso JD, Seddon AE, Jin R, Thomashow MF, et al. Cis-regulatory code of stress-responsive transcription in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. 2011;108:14992–7. https://doi.org/10.1073/pnas.1103202108.

[57] Baek K, Lee Y, Nam O, Park S, Sim SJ, Jin E. Introducing Dunaliella LIP promoter containing light-inducible motifs improves transgenic expression in Chlamydomonas reinhardtii. Biotechnol J 2016;11:384–92. https://doi.org/10.1002/biot.201500269.

[58] Pilpel Y, Sudarsanam P, Church GM. Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet 2001;29:153–9. https://doi.org/10.1038/ng724.

[59] Zeitlinger J. Seven myths of how transcription factors read the cis-regulatory code. Curr Opin Syst Biol 2020;23:22–31. https://doi.org/10.1016/j.coisb.2020.08.002.

[60] de Boer CG, Vaishnav ED, Sadeh R, Abeyta EL, Friedman N, Regev A. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. Nat Biotechnol 2020;38:56–65. https://doi.org/10.1038/s41587-019-0315-8.

[61] Strenkert D, Schmollinger S, Schroda M. Heat shock factor 1 counteracts epigenetic silencing of nuclear transgenes in Chlamydomonas reinhardtii. Nucleic Acids Res 2013;41:5273–89. https://doi.org/10.1093/nar/gkt224.

[62] Yang Y, Lee JH, Poindexter MR, Shao Y, Liu W, Lenaghan SC, et al. Rational design and testing of abiotic stress-inducible synthetic promoters from poplar cis-regulatory elements. Plant Biotechnol J 2021:1–16. https://doi.org/10.1111/pbi.13550.

[63] Patel YD, Brown AJ, Zhu J, Rosignoli G, Gibson SJ, Hatton D, et al. Control of multigene expression stoichiometry in mammalian cells using synthetic promoters. ACS Synth Biol 2021;10:1155–65. https://doi.org/10.1021/acssynbio.0c00643.

[64] Amack SC, Antunes MS. CaMV35S promoter – a plant biology and biotechnology workhorse in the era of synthetic biology. Curr Plant Biol 2020;24:100179. https://doi.org/10.1016/j.cpb.2020.100179.

[65] Weynberg KD, Allen MJ, Wilson WH. Marine prasinoviruses and their tiny plankton hosts: a review. Viruses 2017;9:1–20. https://doi.org/10.3390/v9030043.

[66] Ruecker O, Zillner K, Groebner-Ferreira R, Heitzer M. Gaussia-luciferase as a sensitive reporter gene for monitoring promoter activity in the nucleus of the green alga Chlamydomonas reinhardtii. Mol Genet Genomics 2008;280:153–62. https://doi.org/10.1007/s00438-008-0352-3.

[67] Díaz-Santos E, de la Vega M, Vila M, Vigara J, León R. Efficiency of different heterologous promoters in the unicellular microalga Chlamydomonas reinhardtii. Biotechnol Prog 2013;29:319–28. https://doi.org/10.1002/btpr.1690.