This is a repository copy of *Challenges and advances towards the rational design of microalgal synthetic promoters in Chlamydomonas reinhardtii*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/203776/

Version: Accepted Version

Article:

Milito, A. orcid.org/0000-0002-5541-6113, Aschern, M. orcid.org/0000-0003-1474-7562, McQuillan, J.L. orcid.org/0000-0001-8953-9528 et al. (1 more author) (2023) Challenges and advances towards the rational design of microalgal synthetic promoters in Chlamydomonas reinhardtii. Journal of Experimental Botany, 74 (13). pp. 3833-3850. ISSN 0022-0957

https://doi.org/10.1093/jxb/erad100

This is a pre-copyedited, author-produced version of an article accepted for publication in Journal of Experimental Botany following peer review. The version of record, Alfonsina Milito, Moritz Aschern, Josie L McQuillan, Jae-Seong Yang, Challenges and advances towards the rational design of microalgal synthetic promoters in Chlamydomonas reinhardtii, Journal of Experimental Botany, Volume 74, Issue 13, 18 July 2023, Pages 3833–3850, is available online at: https://doi.org/10.1093/jxb/erad100

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Challenges and advances towards the rational design of microalgal synthetic promoters in *Chlamydomonas reinhardtii*

Alfonsina Milito[1*], Moritz Aschern[1*], Josie L. McQuillan[2], Jae-Seong Yang[1#]

[1] Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, Barcelona, Spain

[2] Department of Chemical and Biological Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK

*These authors equally contributed to this work.

#Corresponding author: Jae-Seong Yang

(Jae-Seong Yang)

Mailing address : Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, Barcelona, Spain

E-mail: jaeseong.yang@cragenomica.es

© The Author(s) 2023. Published by Oxford University Press on behalf of the Society for Experimental Biology. All rights reserved. For permissions, please email: journals.permissions@oup.com

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Abstract

Microalgae hold the enormous potential to provide a safe and sustainable source of high-value compounds, acting as carbon-fixing biofactories that could help to mitigate rapidly progressing climate change. Bioengineering microalgal strains will be key to optimizing and modifying their metabolic outputs, and to render them competitive with established industrial biotechnology hosts, such as bacteria or yeast. To achieve this, precise and tunable control over transgene expression will be essential, towards which a key strategy is the development and rational design of synthetic promoters. Among green microalgae, *Chlamydomonas reinhardtii* represents the reference species for bioengineering and synthetic biology; however, the repertoire of functional synthetic promoters for this species, and for microalgae generally, is limited in comparison to other commercial chassis, emphasizing the need to expand the current microalgal gene expression toolbox. Here, we discuss state-of-the-art promoter analyses and highlight areas of research required to advance synthetic promoter development in *C. reinhardtii*. In particular, we exemplify high-throughput studies performed in other model systems that could be applicable to microalgae and propose novel approaches to interrogating algal promoters. We lastly outline the major limitations hindering microalgal promoter development, while providing novel suggestions and perspectives for how to overcome them.

Keywords

bioengineering, bioinformatics, *Chlamydomonas*, nuclear gene expression, high-throughput approaches, microalgae, promoter, synthetic biology, transcriptional regulation

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Highlight

Here we provide an overview on synthetic promoter studies in *Chlamydomonas reinhardtii*, a reference microalgal species for bioengineering purposes and synthetic biology, outlining key obstacles and ways to overcome them.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

1. Introduction

Microalgae represent one of the most ancient and diverse groups of organisms on Earth, whose extraordinary genetic and chemical biodiversity confer them the ability to adapt to and survive within a wide variety of environments (Dorrell and Smith, 2011). Thanks to their metabolic versatility, they are natural sources of many high-value compounds and are considered key players in the green revolution, i.e. the increasingly urgent transition from fossil fuels to sustainable and renewable plant-based energy sources (Mehariya et al., 2021). However, the quantities produced by wild-type strains are typically too low for large-scale production, necessitating metabolic engineering to boost product yields (Hamilton et al., 2014).

Chlamydomonas reinhardtii (hereafter referred to as *Chlamydomonas*) is an established model organism for the study of various cellular processes including photosynthesis, flagella structure/function, metabolic engineering and more recently as a promising synthetic biology (SynBio) chassis (Scaife et al., 2015). The chloroplast, mitochondrial and nuclear genomes have been sequenced and annotated, with efficient transformation techniques available for each (Merchant et al., 2007; Scaife et al., 2015; Craig et al., 2022). Such techniques allowed the establishment of the chloroplast of *Chlamydomonas* as a successful target for recombinant protein expression due to its homologous recombination machinery enabling targeted gene insertion, a lack of transgene silencing, and high protein product yields (Doron et al., 2016). Moreover, genome editing and SynBio tools are available for *Chlamydomonas*, and their number is vertiginously increasing, thus placing this species in an advanced position as a metabolic engineering chassis compared to other, less characterized, microalgal species (Crozet et al., 2018; Ghribi et al., 2020; Li et al., 2016; Emrich-Mills et al., 2021). In addition, this microalga carries GRAS (generally recognized as safe) status since it does not produce any endotoxin and/or infectious agent, making the purification process much simpler, less expensive, and safer compared to hosts not certified as GRAS. Also, cell wall-deficient *Chlamydomonas* strains can be used, allowing even easier product recovery using GRAS solvents such as ethyl acetate or mineral oil (Perozeni et al., 2020).

Despite the advantages of *Chlamydomonas* as an industrial host and the relevant progress reached in relatively recent years, bioengineering this organism is still inefficient compared to other expression models, such as *E. coli* and yeast. Indeed, although the chloroplast offers a powerful target for transgene expression, potentially crucial post-translational modifications such as glycosylation, as well as organellar and secretory protein targeting mechanisms, are not attainable with chloroplast transformation (Mayfield et al., 2007). This reduces the range of possible applications such as complex, multi-enzyme metabolic engineering, and costly downstream processing is required for recombinant protein recovery. Furthermore, transgene integration and high protein production in the chloroplast can also dramatically reduce or abrogate algal photosynthetic activity (Gimpel et al. 2015;

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Gregory et al. 2012; Surzycki et al. 2009). These issues can be circumvented by expressing transgenes in the nuclear genome, from which glycosylation and secretion are achievable (Lauersen et al., 2015; Mathieu-Rivet et al., 2013). Several native promoters, both constitutive and inducible, have been employed for nuclear transgene expression in *Chlamydomonas* (Scaife et al., 2015), the strongest being a fusion of two native promoters, heat shock protein 70A (*HSP70A*) and small subunit 2 of the ribulose bisphosphate carboxylase (*RBCS2*), henceforth referred to as the AR promoter (Schroda et al., 2000).

However, nuclear transgene expression using the AR promoter in *Chlamydomonas* is still inferior and cumbersome compared to other established unicellular eukaryotic expression systems, partially due to positional effects and gene silencing mechanisms (Schroda, 2019). This exposes the need for powerful genetic elements that can enhance gene expression beyond the possibilities of nature. This lack of strong, diverse, and reliable regulatory devices in *Chlamydomonas* not only limits simple transgene expression, but also more sophisticated SynBio applications, such as the construction of artificial metabolic networks, the implementation of Boolean logic circuits, and the design of molecular feedback controllers, thus highlighting the necessity of precisely tunable and tightly controllable expression tools (Filo et al., 2022; Mori and Shirai 2018; Tas et al. 2021).

In this review article, we aim to provide an overview of synthetic promoter studies for improving gene expression in *Chlamydomonas.* In particular, we will list currently available constitutive and inducible synthetic promoters, as well as the general rules and strategies for their rational design. We will also describe the native promoters used in *Chlamydomonas*, and their *cis*-regulatory elements (CREs), which can be used as templates to design synthetic promoters. We will finally discuss the potential of high-throughput approaches guided by next-generation sequencing (NGS) technologies and computational pipelines for promoter studies, and how studies performed in other species can be applied to microalgae to bolster their commercial value and competitiveness with other biotechnological expression platforms.

2. Synthetic promoters

A synthetic promoter is a DNA regulatory sequence that does not exist in nature, and which has been specifically designed to obtain robust and predictable expression of a target gene. Synthetic promoters can either be assembled entirely artificially or from existing parts in a non-natural way, always including a core promoter region, containing DNA elements to recruit the basic transcriptional machinery, and an upstream proximal promoter region, containing targets for additional transcription factors (TFs) (**Figure 1).** The type, copy number, and spacing of specific *cis*-regulatory elements (CREs) define the specific regulatory output (Venter, 2007). The main advantage of synthetic promoters compared to their native counterparts is represented by the possibility to design sequences

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

with minimal footprints allowing the reduction of the expression cassette size, thus simplifying typical SynBio applications such as the *de novo* construction of metabolic pathways. Moreover, synthetic promoters can offer broader sequence diversities and activity ranges compared to native promoters and are more resistant to homology-dependent gene silencing (Kumar et al., 2013; Einhaus et al., 2021).

In the following subchapters we will describe the main strategies used for synthetic promoter construction and the currently available synthetic promoters for *Chlamydomonas*.

2.1 Top-down and bottom-up strategies for the development of synthetic promoters

Synthetic promoters can be created by two fundamentally different strategies: top-down and bottom-up approaches (**Figure 2**). Top-down design takes advantage of well-characterized native promoters, using them as templates and modifying them accordingly. Modifications can be simple, such as point mutations to modify CREs, or promoter truncations to isolate the minimal functional sequence, removing irrelevant components (**Figure 2A**). Furthermore, introducing full novel motifs or multiplying pre-existing ones, as well as altering the spacing between motifs, can change transcriptional properties (Einhaus et al., 2021). A simple, yet effective approach is to hybridize whole, or parts of, native promoters to generate chimeric constructs with novel characteristics (Schroda et al., 2000). Finally, error-prone PCR can randomize sequences and lead to large-scale promoter libraries (Gilman and Love, 2016).

In contrast, bottom-up promoter design involves constructing promoters entirely from scratch, generating completely new assemblies (**Figure 2C**). Thereby it can start from collections of standardized parts, which can be assembled with known CREs to a minimal core promoter, containing at least the TATA box insulated by nucleotide stretches (Cai et al.; 2020; McQuillan et al., 2022). Moreover, these design workflows can employ randomized approaches like shuffling DNA sequences upstream of the transcription starting site (TSS) with the help of degenerate oligonucleotides (Jensen and Hammer, 1998). Another bottom-up way to generate synthetic promoters is computational design followed by gene synthesis (Kotopka and Smolke, 2020).

Since the construction of synthetic promoters is only constrained by the creativity of their designer, a virtually infinite variety of transcriptional outputs can be obtained, achieving a precise constitutive, inducible, or spatiotemporal control of gene expression (Liu et al., 2013). This is a great advantage of synthetic promoters compared to their native counterparts. Indeed, although strong promoters are often required to obtain the maximum yield of the desired product, fine-tuned gene expression is necessary for more nuanced applications, such as engineering metabolic pathways and genetic circuits controlling cellular behavior. Thus, the development of minimal synthetic control elements with defined strengths is helpful to facilitate the generation of consistent and expectable transcriptional

Accepted Manuscript
Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

outputs with reduced noise compared to long endogenous promoters, which might harbor unknown TF binding sites that counteract the experimental goal (**Figure 2B**).

2.2 Currently available synthetic promoters for *Chlamydomonas*

The synthetic assembly of native promoters can be considered the first generation of synthetic promoters. For instance, the fusion of *HSP70A* promoter with other high-expression promoters, i.e. RBCS2, *TUB2* (β-tubulin 2) and HSP70B, represents the first successful example of synthetic promoter design in *Chlamydomonas*. Indeed, when placed upstream to other promoters, *HSP70A* works as a transcriptional activator, leading to strong constitutive expression under physiological conditions, and conferring inducibility under high light or heat-shock conditions (Schroda et al., 2000). The *HSP70A-RBCS2* (AR) hybrid promoter is particularly effective and has been further characterized, revealing the key CREs conferring inducibility (Schroda et al., 2002). AR is currently used as the strong promoter in most laboratories working with *Chlamydomonas* and is used as the reference strong promoter when comparing the strength of newly designed synthetic promoters.

Later it became evident that to fully exploit the potential of synthetic promoter generation, a more sophisticated assembly of genetic elements responsible for constitutive and/or inducible regulation would have been necessary.

Scranton et al. (2016) generated 25 new synthetic promoters via *de novo* assembly of a synthetic core promoter sequence joined to *in silico* identified CREs. In particular they assembled CREs from native promoters in minimal promoters, i.e. 500 bp long, mimicking the -450 bp upstream and 50 bp downstream the TSS where important motifs for transcriptional activity in *Chlamydomonas* are typically found (Scranton et al., 2016). The synthetic promoters were placed upstream of an mCherry fluorescent reporter, then screened for promoter activity by fluorescence activated cell sorting (FACS) of live *Chlamydomonas* cells. The synthetic promoters exhibited a wide activity range compared to the AR promoter, seven of which drove mCherry expression to levels >2-fold higher than those regulated by AR (Scranton et al., 2016). One of the strongest promoters, sap-11 (synthetic algal promoter-11), was later repurposed as a core promoter to test the activity of putative CREs identified in *Chlamydomonas*, leading to the identification of novel CREs capable of driving high gene expression, and which can be incorporated into rational synthetic promoter designs (McQuillan et al., 2022). These constitutive synthetic promoters have the potential to drive stable gene expression at high levels.

Along with these constitutive synthetic promoters, various inducible synthetic promoters have been generated. Indeed, the control of expression under a specific circumstance is often preferable over constitutive expression to avoid any eventual toxic effect of the produced protein or compound in the bioengineered cells. Inducible promoters can be broadly classified as chemically regulated (induced

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

by a chemical e.g. a hormone or metal) or physically regulated (induced by physical stimuli such as changes in temperature or light) (Ali and Kim, 2019). An example of an inducible synthetic promoter is the fusion of the *TUB2* core promoter with the zygotic response element (ZYRE), which succeeded in driving zygotic gene expression in *Chlamydomonas* (Hamaji et al., 2016). Also, the light-inducible protein (LIP) promoter of *Dunaliella* was successfully modified for use in *Chlamydomonas*, by generating a synthetic truncated version containing several copies of a putative light inducible motif (Baek et al., 2016).

The main synthetic promoters developed so far for *Chlamydomonas*, as well as known CREs responsible for their function and/or induction, are listed in **Table 1**.

3. Advances and opportunities towards the rational design of microalgal synthetic promoters

Despite recent advances in developing synthetic promoters in microalgae, only a small set of these regulatory genetic sequences have actually been utilized for transgene expression in *Chlamydomonas* compared to their development and usage in other organisms (Redden et al., 2015; Romanova and Noll, 2017). In this chapter, we will summarize current knowledge regarding CREs identified from *Chlamydomonas* endogenous promoters that can be used for synthetic promoter construction and describe the methods available for their development. We will explore the state-of-the-art and limitations of experimental and computational methods for promoter studies, as well as high-throughput approaches necessary to interrogate these regulatory regions.

Finally, we will propose tools and infrastructures that will help to advance the field of SynBio with regard to transcriptional regulation in *Chlamydomonas* and provide an outlook on synthetic TFs.

3.1 Advanced knowledge about endogenous natural promoters

Synthetic promoter design requires a deep understanding of the endogenous machinery of transcriptional regulators, especially within the promoter regions. Although this is still lacking, a considerable number of endogenous promoters and their CREs have been identified and tested. These findings will be of great help in designing synthetic promoters.

3.1.1 Main endogenous constitutive and inducible promoters

Endogenous constitutive promoters from *Chlamydomonas* that have been utilized for transgene expression include *TUB2* and *RBCS2* promoters (Davies et al., 1992; Stevens et al., 1996). The *PSAD* (photosystem I subunit D) promoter is also commonly used to regulate transgene expression (Fischer and Rochaix, 2001). However, the most frequently used endogenous element is the *HSP70A* promoter, which activates gene expression when placed upstream of other promoters (Schroda et al.,

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Accepted Manuscript

2000). A relatively recent discovery is the *RPL23* (ribosomal protein L23) regulatory region which showed expression capacities similar to those of other commonly used elements (López-Paz et al., 2017). Constitutive expression elements like the ones previously mentioned can be used to extract features to create synthetic promoters of various strengths.

Several inducible promoters have also been characterized in *Chlamydomonas*. Examples include the nickel and cobalt ion-activated *CYC6* (cytochrome c6) promoter (Quinn et al., 2003), *NIT1* (nitrate reductase), which is induced by ammonium deficiency (Loppes et al., 1999), and *FEA1* (Fe-assimilation protein), induced by iron deficiency (Allen et al., 2007). Although most inducible systems are positively regulated, negatively inducible systems also exist in *Chlamydomonas*, such as the *METE* (cobalamin-independent methionine synthase) promoter, which is repressible by the presence of cobalamin in the growth medium (Helliwell et al., 2014). Other elements enable control *via* physical factors such as the *CAH1* (carbonic anhydrase) promoter, which is induced by light and low levels of CO2 (Villand et al., 1997). Synthetic inducible promoters can be constructed from motifs of such regulatory regions that are responsible for driving condition-specific expression. Endogenous promoters and their characteristics are listed and described in **Table 2 and 3**.

3.1.2 Native *cis*-regulatory elements (CREs)

Genomic sequences that govern transcript generation from DNA can be partitioned into CREs which contain binding motifs for TFs (Wittkopp and Kalay, 2012). The vast majority of information about CREs in *Chlamydomonas* has been acquired by dissecting existing promoters through deletion analyses and mutational experiments.

In general, basal transcription in eukaryotes begins with the recruitment of the pre-initiation complex by core promoter elements (**Figure 1**). In *Chlamydomonas*, the regulatory region surrounding the TSS often contains an initiator element (Inr) and/or a TATA box to recruit the pre-initiation complex (Fischer et al., 2009). Additional core CREs such as CAAT- and CCAAT-boxes, MYB binding sites and G-boxes can also be present (Li et al., 2022; Kropat et al., 1995; Sawyer et al., 2015). *In silico* analyses showed that TC-rich motifs are enriched around the TSS of many highly-transcribed genes in *Chlamydomonas* (Scranton et al., 2016), which is particularly interesting, because a TC-rich motif positioned close to the TSS in plants functions similarly to the TATA box (Bernard et al., 2010).

Generally, *Chlamydomonas* regulatory sequences tend to differ from those of higher plants. For example, plant promoters commonly possess Y patches (pyrimidine-rich regions), which are missing from *Chlamydomonas* regulatory regions (Yamamoto et al., 2007). Moreover, the regions upstream of

Accepted Manuscript
Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

highly-expressed genes in microalgae are AT-rich near to the TSS, whereas in plants GC enrichment adjacent to the TSS is instead observed (Fujimori et al., 2005; Calistri et al., 2011; Scranton et al., 2016).

However, it is reasonable to assume that key core promoter motifs remain to be discovered.

Besides these basic elements, several inducible CREs have been characterized in detail. This may facilitate the construction of inducible synthetic promoters, a hitherto underexplored area of applied transcriptional research in microalgae. CREs of particular interest include the heat shock elements (HSE) of the *HSP70A* promoter, which function both as inducible and constitutive expression elements, by attracting heat shock factors (Lodha et al., 2008). Other CREs identified from endogenous promoters of *Chlamydomonas* are listed in **Table 2 and 3**. These may be optimized and combined for synthetic promoter design, and future discovery thereof deserves special attention.

3.2 High-throughput experimental approaches for promoter studies

The regulation of gene expression constitutes a complex network of interactions between various regulatory elements, acting in a *cis*- and *trans*-manner, to precisely determine the spatio-termporal synthesis of a certain amount of transcript. Sequence motifs in promoters act as binding sites for TFs, thus playing a crucial role in modulating transcription. Considering this, we still know little regarding how the location, orientation and affinity of CREs quantitatively influence gene expression. To complicate this picture, the motifs do not act alone, but their function is the result of their co-evolution with coding and non-coding regions of the DNA regulatory structure (Zrimec et al., 2020). Therefore, low-throughput methods like promoter shuffling and individual characterization cannot solve this complex problem, highlighting the necessity to involve high-throughput approaches.

High-throughput techniques are generally based on modern biotechnological tools, such as DNA synthesis technology, FACS, and NGS, which enable the study of gene expression activity of a massive amount of promoter variants simultaneously (**Figure 3**). In several prokaryotic model organisms, these approaches facilitated the *de novo* detection of key motifs that determine transcription, thereby paving the way for synthetic promoter design (Kinney et al., 2010; Levy et al., 2017; Yus et al., 2017). In unicellular eukaryotic systems, high-throughput methods were successfully used to investigate the activities of promoter libraries of vast size, giving mechanistic insights into gene regulation processes (Kotopka and Smolke, 2020; Lubliner et al., 2015). These deep mutational screenings pave the way towards building accurate computational models for predicting gene expression from DNA sequences and to design synthetic sequences for the desired rate of gene expression.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

However, no such high-throughput promoter study has been conducted in microalgae so far. Here, we will present several trailblazing methods to both measure promoter strength and identify TF binding sites in a massively parallel manner.

Synthetic saturation mutagenesis of promoters can be conducted as a reporter gene-free method, which allowed exploring the effects of mutagenized bacterial and mammalian promoters on *in vitro* transcription efficiency (Patwardhan et al., 2009). Sharon et al. (2012) set up a high-throughput method to obtain gene expression levels from thousands of synthetic promoters *in vivo* in yeast cells, by using fluorescent markers and barcoding, with a high accuracy (**Figure 3A**). The relatively novel technique termed ELM-seq (expression level monitoring by DNA methylation) is based on bacterial Dam protein (DNA adenine methylase) and NGS technology (Yus et al., 2017) (**Figure 3F**). In particular, promoter (or 5'UTR) libraries are constructed, placing the Dam encoding gene downstream the promoter region (or the 5'UTR region). Thus, the methylation level of close GATC sequences, the targets of Dam activity, will depend on the promoter (or 5'UTR) strength regulating Dam expression. DNA extraction followed by restriction digestion with methylation-sensitive enzymes, and NGS, allows the determination of a Dam ratio and to the identification of the regulatory sequences responsible for strong/weak transcription (or translation). This method has the advantages of providing a high dynamic range, minimal toxicity, and no need for fluorescent reporters. Moreover, it has the potential, with some optimization, to be applied to any prokaryotic and eukaryotic system to unravel sequence determinants for gene expression (Yus et al., 2017). STARR-seq (self-transcribing active regulatory region sequencing) is a massively parallel reporter assay originally developed for the identification of transcriptional enhancers (Arnold et al., 2013), which was later adapted for application in plants (Jores et al., 2020) (**Figure 3B**). By driving the expression of a barcoded reporter gene and subsequent identification of barcodes by RNA-seq, the experimental scope was recently expanded to allow systematic measurement of core promoter activity (Jores et al., 2021). This adapted technique has the capacity to measure the strength of several tens of thousands of regulatory elements simultaneously. As a result, it can identify key determinants of transcription in plants and likewise, holds great promise for future applications in microalgae.

Although several CREs have been experimentally validated and are available for the construction of synthetic promoters, a systems level understanding of the cistrome of *Chlamydomonas* is currently missing. The application of high-throughput methods to discover TF binding sites, such as DNA affinity purification sequencing (DAP-seq), could systematically reveal the binding relationship between TFs and their corresponding motifs, thereby helping to create new promoters with desirable characteristics (O'Malley et al., 2016) (**Figure 3C**). Furthermore, large-scale chromatin immunoprecipitation (ChIP-seq) and deep cap analysis gene expression (CAGE) are interesting tools in this regard (**Figure 3D and 3E**). While the former technique is based on the enrichment of target DNA sites through TF-specific immune reagents (Johnson et al., 2007), the latter reverse-transcribes

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023
Accepted Manuscript

capped mRNA and physically traps the resulting polynucleotide in order to sequence a short tag corresponding to the 5' of the mRNA (Kawaji et al., 2014). In particular, CAGE coupled with high-throughput sequencing allows the identification of known and novel promoters, their TF-binding motifs, as well as the quantification of the expression levels of the downstream genes (de Hoon and Hayashizaki, 2018).

3.3 Computationally-aided search of transcription determinants

Even though the previously discussed high-throughput techniques are auspicious and indispensable tools to study the fundamental factors of promoter strength, their application is usually challenging, laborious, and expensive. Bioinformatic prediction tools offer help in this regard: progress in the computational field has allowed the establishment of machine learning methods with the potential to distinguish between regulatory and non-regulatory regions and thus identify promoters, as well as to predict their activity performance (Wang et al., 2020).

Within the multitude of available machine learning methods, Artificial Neural Networks (ANNs) are becoming increasingly popular in biology, especially Convolutional Neural Networks (CNNs), a subtype thereof (Yang et al., 2020). Although being a powerful method for feature extraction from continuous signals, like DNA sequences, this network type has several limitations, including high computational cost and output of difficult interpretability (Wang et al., 2020). For an easier interpretation of DNA sequences, k-mer (or k-tuple) approaches are preferable over CNNs even though having lower feature extraction capability. A good strategy is combining both k-mers and CNNs, a method successfully used to generate prediction models of transcription binding sites (Shen et al., 2018). Using a combination of CNNs and FastText N-grams, a Facebook library for text representation and classification, Le et al. (2019) identified promoters and relative strengths with high accuracy. Notably, other types of ANNs architectures beyond CNNs have been successfully used to predict gene expression regulation, such as Deep Neural Networks and Recurrent Neural Networks (Min et al., 2016). Known promoter features, like TATA box, Inr motif, and GC-rich regions, have been used to build predictive models for the identification of promoters. Some pioneering examples are Promoter2.0 (Knudsen, 1999), able to distinguish vertebrate promoters from non-promoter sequences and available online, and the time-delay neural network model created by Reese (2001), having two feature layers (TATA and Inr) but extendible with other features like the downstream promoter element (DPE), CAAT, and GC boxes. Other recently released machine learning tools for promoter prediction from sequence information are iProEP (Lai et al., 2019), CNNProm (Umarov and Solovyev, 2017), DeeReCT-PromID (Umarov et al., 2019), DeePromoter (Oubounyt et al., 2019), and Depicter (Zhu et al., 2021), all four based on deep learning algorithms.

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Accepted Manuscript

Different methods use spatial information between base pairs, i.e. the genetic context of promoters rather than their exact location, like PromoterInspector (Scherf et al., 2000), and MCPromoter1.1 (Ohler et al., 1999). Abeel et al. (2008) built an Easy Promoter Prediction Program (EP3), which uses GC content and many other structural features of DNA, like bendability, duplex-free energy, and DNA denaturation, proved to be highly informative for promoter prediction (Gan et al., 2012), and resulted in a powerful software with no training needed and applicable to any eukaryotic genome.

The release of genome data for *Chlamydomonas* and the development of a computational approach called MERCED (modeling evolution rate across species for cis-regulatory element discovery) allowed the identification of 317 CREs (Ding et al., 2012). With a completely different approach, López García de Lomana et al. (2015) used the cMonkey algorithm on an available RNA-seq dataset and detected more than 400 CREs regulated by nitrogen starvation and summarized in the Chlamy Network Portal. This semi-supervised biclustering algorithm can identify co-regulated genes within a gene expression dataset and use this information for the *de novo* detection of motifs within the gene groups (Reiss et al., 2006). Besides the findings of these two studies, the usage of computational methods to unravel mechanisms of transcription in microalgae is a rather underexploited field, leaving great room for future research harnessing the available machine learning methods.

3.4 Synthetic biology (SynBio) tools and infrastructures

In order to advance the development of synthetic promoters in *Chlamydomonas*, it is important not only to have a deep knowledge of the transcriptional regulatory landscape, but also to have efficient sets of genetic engineering tools for reliable and reproducible genetic transformation, as well as for large-scale DNA construction and screening of synthetic promoter activities. SynBio provides useful DNA assembly toolkits that draw from libraries of existing parts and thus streamline engineering efforts. Currently, there is a wide array of available methods and standards for the purpose of systematic DNA assembly, such as BioBrick, modular cloning (MoClo) or GoldenBraid (Shetty et al., 2008; Vazquez-Vilar et al., 2020, Weber et al., 2011). The underlying methodologies thereby include restriction and ligation cloning, long-overlap-based assembly and Golden Gate Cloning. Due to the advantages of enabling scar-free assembly with a defined workflow, relative ease and high success rates, Golden Gate cloning has become one of the most popular methodologies for DNA assembly (Casini et al., 2015).

A complete golden gate-compatible set of standardized genetic parts has been specifically designed to endow *Chlamydomonas* with new molecular tools (Crozet et al., 2018). This toolkit was successfully used for sustainable recombinant polyamine synthesis and production and secretion of the SARS-CoV-2 spike protein in *Chlamydomonas* (Freudenberg et al., 2021; Kiefer et al., 2022). At present, the number of available standard parts is still considerably low if compared to other model organisms, but

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Accepted Manuscript

recent efforts are contributing to increase the number of available MoClo-compatible parts, including promoters (Einhaus et al., 2021; Niemeyer et al., 2021). These efforts will accelerate the emergence of SynBio applications based on *Chlamydomonas*. In that respect, a golden gate-compatible collection of CREs would enable the effortless assembly of synthetic promoters for *Chlamydomonas* and lead to a substantial advance in microalgal bioengineering.

In recent years, the emergence of molecular biology automation platforms and other high-performance technologies has led to the establishment of biofoundries worldwide, pursuing a highly centralized and specialized approach. These entities provide powerful SynBio infrastructures facilitating the engineering of biological systems according to the Design-Build-Test-Learn cycle. They are open to the research community and organized in the non-profit Global Biofoundries Alliance (GBA) (Hillson et al., 2019). By combining the hardware of robotic liquid-handling devices and high-throughput analytical instruments with softwares, data analysis tools and trained specialists to operate the system, GBA puts forth an integrated molecular biology facility (Holowko et al., 2021). These biofoundries evidently outperform non-automatized approaches, thereby offering a means to harness the power of microalgal SynBio for future challenges such as the global climate crisis.

3.5 Synthetic transcription factors (TFs) and other gene expression regulators

A strategy that has not yet been exploited in *Chlamydomonas* but holds future promise, due to encouraging results in plant species such as *Arabidopsis thaliana* and other eukaryotic organisms, are synthetic TFs. These would facilitate the development of completely novel promoter sequences with high levels of predictable regulation, which would be independent of or have minimal interference from endogenous TFs, but depend on the synthetically introduced ones. Synthetic TFs are usually chimeric proteins consisting of activation or repression domains fused to the DNA binding domains of programmable TFs. Initially, zinc-finger (ZF) TFs or transcription activator-like effector (TALE) TFs were utilized for that purpose (Liu et al., 2013). However, the application of TF-based methods to different model systems and different target genes can be quite demanding, necessitating the synthesis and optimization of proteins.

The discovery of endonucleases based on the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas9 system that can be easily programmable to bind virtually any DNA sequence provided the framework for tools whose development is much faster than those based on ZF or TALE TFs (Bhardwaj and Nain, 2021). By rendering the endonuclease Cas9 catalytically dead (dCas9) while maintaining its high sequence binding specificity, it can be fused to effector proteins or domains thereof to regulate gene transcription in a controllable way (Liu and Stewart, 2016). However, a previous attempt to constitutively express dCas9 in *Chlamydomonas* indicates some degree of toxicity that might complicate its application as a modulator (Jiang et al., 2014). In addition

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Accepted Manuscript

to dCas9, other types of endonucleases, like inactivated meganucleases, have proven to be an auspicious tool for modulating transcription in recent years, thus amplifying the toolbox of retrofittable transcriptional regulators (Suzuki et al., 2020). Overall, the choice of effector domains fused to the inactivated nuclease of choice is not limited to simple activators or repressors. Enzymatically active protein segments fused to deactivated nucleases can trigger local DNA methylation, histone modification and chromatin loop reorganization in the targeted locus (Pandelakis et al., 2020). The control of such epigenetic modifications is especially interesting for future applications in *Chlamydomonas*, as the main reasons for its poor transgene expression capabilities are transcriptional gene silencing mechanisms based on histone alterations and repressive chromatin structures (Schroda, 2019).

4. Challenges and current obstacles and ways to overcome them

Despite the number of molecular tools available for *Chlamydomonas*, the discovery of transcriptional determinants and subsequently, synthetic promoter design and testing efforts, are complicated by several organism-specific bottlenecks. In this chapter we will address several limitations of *Chlamydomonas* as a model system for the development of synthetic promoters, and at the same time we will outline strategies to resolve them.

4.1 Incomplete knowledge on promoter annotation and regulatory networks

As previously mentioned, a deep understanding of transcription regulation in *Chlamydomonas* is still missing, especially with regard to promoter and TSS annotation, as well as TF-DNA binding specificities and TF-TF interactions (**Figure 4A**). In eukaryotic transcription, a certain degree of variability can commonly be observed with regard to the TSS position, resulting in a cluster of sites called transcription starting region (TSR) within a promoter. Moreover, one gene can harbor multiple TSRs, leading to alternative transcription initiation to regulate gene expression at specific times, like during zebrafish development (Haberle et al., 2014), or in response to environmental factors, such as blue light in plants (Kurihara et al., 2018). Due to the importance of TSS selection for gene expression regulation, a plethora of methods have been developed for genome-wide TSS identification (Policastro and Zentner, 2021). The previously mentioned CAGE-seq method can be used to identify TSS activity at nucleotide resolution and, consequently, promoters and enhancers (Morioka et al., 2020). CAGE-seq was successfully applied in plants to demonstrate the occurrence of rare promoter upstream transcripts, a result of the RNA polymerase II activity on the reverse strand of gene promoters (Thieffry et al., 2020). TSS mapping data generated in plants by using CAGE (Tokizawa et al., 2017; Ushijima et al., 2017), deep CAGE (Cumbie et al., 2015), oligo capping, and PEAT (paired-end analysis of transcription start sites, Morton et al., 2014) technologies has enabled the annotation of more than 20,000 promoters in *A. thaliana* (https://epd.epfl.ch/arabidopsis/arabidopsis_

Accepted Manuscript
Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

database.php). The annotation of TSS selection mechanisms for *Chlamydomonas* and the development of TSS prediction databases would provide a great advance for precise promoter mapping and help to identify key regulatory elements.

Despite the abundance of available information about native CREs, we still lack a comprehensive understanding of their role in transcriptional regulatory networks (TRN) in *Chlamydomonas*, which comprise the interactions between TFs and CREs in the target gene promoter regions. TRN in microalgae could be inferred using high-throughput methods like the aforementioned CHIP-seq. However, this approach is limited by the availability of antibodies for specific TFs, and the high and constant level at which the target TFs must be expressed *in vivo* (Mendoza-Parra et al., 2016). Yeast two- (Y2H) and three-hybrid (Y3H) screenings facilitate the identification of protein-protein and protein-RNA interactions, thereby overcoming such limitations. By coupling these techniques with the power of modern sequencing technologies it is possible to generate comprehensive datasets with minimum outlay (Yu et al., 2011; Weinmann et al., 2013). The least laborious of these methods, while remaining highly precise, is recombination-YnH (rec-YnH), which provides a new tool to determine many-by-many interactions in protein libraries. It uses the yeast recombinase system to fuse prey and bait library counterparts, whose combined protein coding sequences are used for paired-end NGS to provide information about interacting pairs (Yang et al., 2018). In the future, techniques like rec-YnH could be harnessed to determine the TF-promoter interactions in microalgae on a systems level and construct TRN databases from this information (**Figure 4A**).

4.2 Position effects

When transforming transgenes into *Chlamydomonas*, commonly less than 1% of transformed cells yield clones that express these genes at a reasonably high degree (Sproles et al., 2022). Hence, vast numbers of colonies need to be screened after each experiment to find transformants with stable protein expression capabilities. Besides gene silencing mechanisms, this phenomenon can be largely attributed to the random nature of integration of heterologous DNA in the nucleus of this organism *via* non-homologous recombination (Zhang et al., 2014). This consequently results in a multitude of possible different transcriptional activities by the same construct (**Figure 4B**), being primarily dependent on the genomic context of the insertion site and complicating the usage of deep mutational screenings to systematically explore transcription determinants. These types of screenings typically involve thousands of different promoters or variants thereof (Kinney et al., 2010; Levy et al., 2017). To test a promoter library of such size, the level of reporter gene expression is usually determined, which should correspond to the strength of a particular promoter variant. However, when position effects are strong as in *Chlamydomonas*, reporter gene expression is not only determined by the promoter variant strength but also by the specific locus of integration. Assuming an exemplary library

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Accepted Manuscript

size of 1,000 promoters, according to the findings of Lodha et al. (2008), at least 240,000 individual transformants should be analyzed to equalize the position effects for each promoter.

Thus, one solution to equalize the positional effects when testing the activity of promoter libraries in *Chlamydomonas* is to achieve a very high transformation efficiency, such to have a large coverage of the mutation variants and be able to screen massive numbers of colonies. The use of innovative electroporation technologies is a promising approach in that respect. Square wave electroporation systems, in contrast to traditional glass bead transformation or exponential decay wave electroporators, have proven high efficiency in both standard transformation procedures and sophisticated gene-editing protocols (Greiner et al., 2017; Yamano et al., 2013). Additionally, continuous-flow devices based on microfluidic droplet electroporation hold great promise due to their efficiency and adaptability for high-throughput protocols (Im et al., 2015; Yoo et al., 2018). Another approach to minimize position effects is the usage of specific strains for testing promoter collections. For example, UVM4 and UVM11, two mutant strains both carrying a defective histamine deacetylase gene reported to be responsible for suppressed gene silencing mechanisms (Neupert et al., 2009; Neupert et al., 2020), can offer a solution in this respect. On the other hand, partial position effects have still been detected in these strains (Hammel et al., 2020; Niemeyer et al., 2021). Thus, their suitability for large-scale promoter library testing remains to be verified.

Direct targeting of the specific insertion location is a third strategy to prevent variation of expression levels in transformants carrying the same DNA cassette. This can be achieved by imposing homologous recombination during transformation via ssDNA-based nuclear gene targeting or a Cas9-assisted technique (Angstenberger et al., 2020; Zorin et al., 2005). Anchoring landing pads in the *Chlamydomonas* genome for site-specific insertion of DNA also holds great promise for this purpose and proved to be effective in other biotechnologically important eukaryotic expression systems (Bourgeois et al., 2018; Gaidukov et al., 2018). However, the efficiency of these targeted insertion approaches needs to be explored.

Lastly, instead of mitigating position effects, such effects could be accounted for in large-scale promoter screenings by mapping the insertion sites of the transforming DNA cassette on the genome. Techniques such as *Chlamydomonas* MmeI-based insertion site sequencing (ChlaMmeSeq) and linear and exponential amplification of insertion site sequence coupled with paired-end sequencing (LEAP-Seq) enable the insertion sites of randomly integrated DNA cassettes to be mapped in the *Chlamydomonas* nuclear genome (Li et al., 2016; Zhang et al., 2014).

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

4.3 Interference of endogenous nucleases

Besides the inconvenience resulting from position effects, the cleavage of expression cassettes by endogenous nucleases released during transformation complicates heterologous gene expression efforts further (Zhang et al., 2014). This results in complex insertion events where DNA is rearranged, truncated, or concatenated, which ultimately leads to an inherent degree of inconsistency when carrying out genetic modifications in the *Chlamydomonas* system (Li et al., 2016). Thus, these unexpected events make it difficult to analyse the experimental outputs, particularly in high-throughput experiments for measuring the activities of many promoter variants. To solve this problem, Sproles et al. (2022) have developed a vector system tailored to dual antibiotic selection, where two selective markers enclose the transgene of interest. They showed it can have a significantly positive effect on the anticipated phenotype, but it has the disadvantage of increasing the size of the transforming cassette as well as requiring double antibiotic selection. Advances in this respect may be provided by implementing 'AND' logic gates which, per definition, only result in an output if both input elements are present (**Figure 4C**). For example, protein-fragment complementation systems could be developed, composed of short interacting parts that facilitate selection only if brought together (Jillette et al., 2019).

4.4 Crosstalk between promoter and other genetic elements

A fundamental problem at the intersection between nature and engineering is the lack of predictability. Comprehensive analyses of various promoter-terminator combinations have shown synergistic interactions between them (**Figure 4D**), which in certain cases result in substantially altered promoter activity when different terminator elements are used in plants (Andreou et al., 2021). In *Chlamydomonas*, recent studies yielded somewhat contrasting information, thereby highlighting the necessity of further research to clarify the extent to which gene expression under the control of a certain promoter can be influenced by different terminators. Geisler et al. (2021) investigated the influence of nine terminators of three different size classes, tested the constructs in three commonly used strains of *Chlamydomonas* and observed only minor differences in the expression of a reporter gene. Other findings, however, showed massive differences in the performance of the same promoter with different terminator elements and support the hypothesis that the behavior of a given genetic element can be highly context dependent *Chlamydomonas* as in other systems (Crozet et al., 2018; Einhaus et al., 2021; Kumar et al., 2013; López-Paz et al., 2017). Beyond promoters and terminators, other regulatory elements and/or expression systems, like the expression plasmid or the genetic background of the microalgal strain, can be expected to lead to crosstalk with the rest of the entity. It is furthermore known from various other eukaryotic taxa that promoters interact with neighbouring promoters (Hampf and Gossen, 2007), introns (Agarwal and Ansari, 2016), enhancers or other distal elements (Bulger and Groudine, 2011; Matharu and Ahituv, 2015; Sanyal et al., 2012), and thereby

Accepted Manuscript
Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

vastly influence the transcriptional output. To what extent these molecular mechanisms take effect in *Chlamydomonas* however, remains an open question. Therefore, a community-wide effort is needed to not only construct uniformly performing components with minimal biological interactions, but also to test existing elements in a wide variety of contexts. In this regard, the MoClo system will be of great help by facilitating straightforward assembly of a plethora of different gene regulative components combinations.

5. Conclusions

In this review, we covered several aspects regarding the challenges and advances in promoter engineering and key strategies for strain optimization and biotechnology in microalgae. In this context, synthetic promoters will enable predictable transcriptional output and a wide activity range required for precise and tuneable gene regulation in microalgal bioengineering. We focused on synthetic promoters developed in *Chlamydomonas* due to its current standing as a model microalgal system, but the tools described here for interrogating promoter architecture can be applied as a framework for developing synthetic promoters in other, less established microalgal species.

Although synthetic promoters are currently being generated for *Chlamydomonas*, future efforts will be needed to improve this field, aided by computational strategies, such as prediction models linking sequence information to expression level, as well as high-throughput methods allowing fast screening of large and complex libraries, and NGS technologies. Moreover, to achieve real advancements in the biotechnological use of microalgae, essential knowledge gaps should be filled. These include a deep understanding of transcriptional regulation and crosstalk between different genetic elements, such as promoters and terminators, as well as developing effective ways to evade position effects in transgene expression and interference by endogenous endonucleases. In addition, although this review is mainly focused on promoters as gene regulative components, the picture is further complicated by additional genetic elements that were not discussed in this article, but that have or might have an important role in determining the transcriptional output when engineering *Chlamydomonas*, such as enhancers/silencers and insulators. Enhancers/silencers may be located remotely from the TSS but helped by chromatin remodelling, can come to interact with TFs and enhance/repress the expression of distantly located genes, with insulators working as moderators of such interactions (Riethoven, 2010). Their function is relatively well understood in plants, with insulators being thought to neutralize position effects and prevent gene silencing by heterochromatin (Bilas et al., 2016). The occurrence and function of these gene regulatory elements, as well as the understanding of how they interact with each other in microalgae, may be key to boosting their potential for biotechnology and synthetic biology.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Conflicts of interest

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

Funding

We acknowledge financial support from the Spanish Ministry of Science and Innovation-State Research Agency (AEI), through the CEX2019-000902-S, PID2020-117772GA-I00, and RYC2020-028880-I funded by MCIN/AEI/10.13039/501100011033. Also, we note that this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 945043.

Acknowledgements

We are grateful to Prof. Michael Schroda and Dr. Jongrae Kim for useful comments on an earlier version of this paper.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

References

Abeel T, Saeys Y, Bonnet E, Rouzé P, Van de Peer Y. 2008. Generic eukaryotic core promoter prediction using structural features of DNA. Genome Research 18, 310–323.

Agarwal N, Ansari A. 2016. Enhancement of Transcription by a Splicing-Competent Intron Is Dependent on Promoter Directionality. PLOS Genetics 12, e1006047.

Ali S, Kim W-C. 2019. A Fruitful Decade Using Synthetic Promoters in the Improvement of Transgenic Plants. Frontiers in Plant Science 10, 1433.

Allen MD, del Campo JA, Kropat J, Merchant SS. 2007. *FEA1* , *FEA2* , and *FRE1* , Encoding Two Homologous Secreted Proteins and a Candidate Ferrireductase, Are Expressed Coordinately with *FOX1* and *FTR1* in Iron-Deficient *Chlamydomonas reinhardtii*. Eukaryotic Cell 6, 1841–1852.

Andreou AI, Nirkko J, Ochoa-Villarreal M, Nakayama N. 2021. Mobius Assembly for Plant Systems highlights promoter-terminator interaction in gene regulation. bioRxiv doi: 10.1101/2021.03.31.437819. [Preprint].

Angstenberger M, de Signori F, Vecchi V, Dall'Osto L, Bassi R. 2020. Cell Synchronization Enhances Nuclear Transformation and Genome Editing *via* Cas9 Enabling Homologous Recombination in *Chlamydomonas reinhardtii*. ACS Synthetic Biology 9, 2840–2850.

Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. 2013. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. Science 339, 1074–1077.

Baek K, Lee Y, Nam O, Park S, Sim SJ, Jin E. 2016. Introducing *Dunaliella LIP* promoter containing light-inducible motifs improves transgenic expression in *Chlamydomonas reinhardtii*. Biotechnology Journal 11, 384–392.

Bandziulis RJ, Rosenbaum JL. 1988. Novel control elements in the alpha-1 tubulin gene promoter from *Chlamydomonas reinhardii*. Molecular and General Genetics MGG 214, 204–212.

Barjona do Nascimento Coutinho P, Friedl C, Heilmann M, Buchholz R, Stute SC. 2019. Validated Nuclear-Based Transgene Expression Regulated by the Fea1 Iron-Responsive Promoter in the Green Alga *Chlamydomonas reinhardtii*. Molecular Biotechnology 61, 305–316.

Baurain D, Dinant M, Coosemans N, Matagne RF. 2003. Regulation of the Alternative Oxidase *Aox1* Gene in *Chlamydomonas reinhardtii*. Role of the Nitrogen Source on the Expression of a Reporter Gene under the Control of the *Aox1* Promoter. Plant Physiology 131, 1418–1430.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Beltran-Aguilar AG, Peraza-Echeverria S, López-Ochoa LA, Borges-Argáez IC, Herrera-Valencia VA. 2019. A novel salt-inducible CrGPDH3 promoter of the microalga *Chlamydomonas reinhardtii* for transgene overexpression. Applied Microbiology and Biotechnology 103, 3487–3499.

Bernard V, Brunaud V, Lecharny A. 2010. TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation. BMC Genomics 11, 166.

Bhardwaj A, Nain V. 2021. TALENs—an indispensable tool in the era of CRISPR: a mini review. Journal of Genetic Engineering and Biotechnology 19, 125.

Biłas R, Szafran K, Hnatuszko-Konka K, Kononowicz AK. 2016. Cis-regulatory elements used to control gene expression in plants. Plant Cell, Tissue and Organ Culture (PCTOC) 127, 269–287.

Blankenship JE, Kindle KL. 1992. Expression of chimeric genes by the light-regulated cabII-1 promoter in *Chlamydomonas reinhardtii*: a cabII-1/nit1 gene functions as a dominant selectable marker in a nit1- nit2- strain. Molecular and Cellular Biology 12, 5268–5279.

Bourgeois L, Pyne ME, Martin VJJ. 2018. A Highly Characterized Synthetic Landing Pad System for Precise Multicopy Gene Integration in Yeast. ACS Synthetic Biology 7, 2675–2685.

Bulger M, Groudine M. 2011. Functional and Mechanistic Diversity of Distal Transcription Enhancers. Cell 144, 327–339.

Brunke KJ, Anthony JG, Sternberg EJ, Weeks DP. 1984. Repeated consensus sequence and pseudopromoters in the four coordinately regulated tubulin genes of *Chlamydomonas reinhardi*. Molecular and Cellular Biology 4, 1115–1124.

Calistri E, Livi R, Buiatti M. 2011. Evolutionary trends of GC/AT distribution patterns in promoters. Molecular Phylogenetics and Evolution 60, 228–235.

Casini A, Storch M, Baldwin GS, Ellis T. 2015. Bricks and blueprints: methods and standards for DNA assembly. Nature Reviews Molecular Cell Biology 16, 568–576.

Craig RJ, Gallaher SD, Shu S, *et al.* 2022. The Chlamydomonas Genome Project, version 6: reference assemblies for mating type plus and minus strains reveal extensive structural mutation in the laboratory. bioRxiv doi: 10.1101/2022.06.16.496473. [Preprint].

Crozet P, Navarro FJ, Willmund F, *et al.* 2018. Birth of a Photosynthetic Chassis: A MoClo Toolkit Enabling Synthetic Biology in the Microalga *Chlamydomonas reinhardtii*. ACS Synthetic Biology 7, 2074–2086.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Cumbie JS, Ivanchenko MG, Megraw M. 2015. NanoCAGE-XL and CapFilter: an approach to genome wide identification of high confidence transcription start sites. BMC Genomics 16, 597.

Davies JP, Grossman AR. 1994. Sequences Controlling Transcription of the *Chlamydomonas reinhardtii* β2-Tubulin Gene after Deflagellation and during the Cell Cycle. Molecular and cellular biology 14, 5165–5174.

Davies JP, Weeks DP, Grossman AR. 1992. Expression of the arylsulfatase gene from the β_2 -tubulin promoter in *Chlamydomonas reinhardtii*. Nucleic Acids Research 20, 2959–2965.

de Hoon M, Hayashizaki Y. 2008. Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference. BioTechniques 44, 627–632.

Deng X, Eriksson M. 2007. Two Iron-Responsive Promoter Elements Control Expression of *FOX1* in *Chlamydomonas reinhardtii*. Eukaryotic Cell 6, 2163–2167.

Ding J, Li X, Hu H. 2012. Systematic Prediction of cis-Regulatory Elements in the *Chlamydomonas reinhardtii* Genome Using Comparative Genomics. Plant Physiology 160, 613–623.

Doron L, Segal N, Shapira M. 2016. Transgene Expression in Microalgae—From Tools to Applications. Frontiers in Plant Science 7, 505.

Dorrell RG, Smith AG. 2011. Do Red and Green Make Brown?: Perspectives on Plastid Acquisitions within Chromalveolates. Eukaryotic Cell 10, 856–868.

Einhaus A, Baier T, Rosenstengel M, Freudenberg RA, Kruse O. 2021. Rational Promoter Engineering Enables Robust Terpene Production in Microalgae. ACS Synthetic Biology 10, 847–856.

Emrich-Mills TZ, Yates G, Barrett J, *et al.* 2021. A recombineering pipeline to clone large and complex genes in *Chlamydomonas*. The Plant Cell 33, 1161–1181.

Fei X, Deng X. 2007. A Novel Fe Deficiency-Responsive Element (FeRE) Regulates the Expression of atx1 in *Chlamydomonas reinharditii*. Plant and Cell Physiology 48, 1496–1503.

Fei X, Eriksson M, Li Y, Deng X. 2010. A Novel Negative Fe-Deficiency-Responsive Element and a TGGCA-Type-Like FeRE Control the Expression of *FTR1* in *Chlamydomonas reinhardtii*. Journal of Biomedicine and Biotechnology 2010, 1–9.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Fei X, Eriksson M, Yang J, Deng X. 2009. An Fe Deficiency Responsive Element with a Core Sequence of TGGCA Regulates the Expression of FEA1 in *Chlamydomonas reinharditii*. Journal of Biochemistry 146, 157–166.

Filo M, Kumar S, Khammash M. 2022. A hierarchy of biomolecular proportional-integral-derivative feedback controllers for robust perfect adaptation and dynamic performance. Nature Communications 13, 2119.

Fischer BB, Dayer R, Schwarzenbach Y, Lemaire SD, Behra R, Liedtke A, Eggen RIL. 2009. Function and regulation of the glutathione peroxidase homologous gene GPXH/GPX5 in *Chlamydomonas reinhardtii*. Plant Molecular Biology 71, 569–583.

Fischer N, Rochaix J-D. 2001. The flanking regions of PsaD drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*. Molecular Genetics and Genomics 265, 888–894.

Freudenberg RA, Baier T, Einhaus A, Wobbe L, Kruse O. 2021. High cell density cultivation enables efficient and sustainable recombinant polyamine production in the microalga *Chlamydomonas reinhardtii*. Bioresource Technology 323, 124542.

Fujimori S, Washio T, Tomita M. 2005. GC-compositional strand bias around transcription start sites in plants and fungi. BMC Genomics 6, 26.

Gaidukov L, Wroblewska L, Teague B, *et al.* 2018. A multi-landing pad DNA integration platform for mammalian cell engineering. Nucleic Acids Research 46, 4072–4086.

Gan Y, Guan J, Zhou S. 2012. A comparison study on feature selection of DNA structural properties for promoter prediction. BMC Bioinformatics 13, 4.

Geisler K, Scaife MA, Mordaka PM, Holzer A, Tomsett EV, Mehrshahi P, Mendoza Ochoa GI, Smith AG. 2021. Exploring the Impact of Terminators on Transgene Expression in *Chlamydomonas reinhardtii* with a Synthetic Biology Approach. Life 11, 964.

Ghribi M, Nouemssi SB, Meddeb-Mouelhi F, Desgagné-Penix I. 2020. Genome Editing by CRISPR-Cas: A Game Change in the Genetic Manipulation of *Chlamydomonas*. Life 10, 295.

Gilman J, Love J. 2016. Synthetic promoter design for new microbial chassis. Biochemical Society Transactions 44, 731–737.

Gimpel JA, Hyun JS, Schoepp NG, Mayfield SP. 2015. Production of recombinant proteins in microalgae at pilot greenhouse scale. Biotechnology and Bioengineering 112, 339–345.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Gregory JA, Li F, Tomosada LM, Cox CJ, Topol AB, Vinetz JM, Mayfield S. 2012. Algae-Produced Pfs25 Elicits Antibodies That Inhibit Malaria Transmission (L Hviid, Ed.). PLoS ONE 7, e37179.

Greiner A, Kelterborn S, Evers H, Kreimer G, Sizova I, Hegemann P. 2017. Targeting of Photoreceptor Genes in *Chlamydomonas reinhardtii* via Zinc-Finger Nucleases and CRISPR/Cas9. The Plant Cell 29, 2498–2518.

Haberle V, Li N, Hadzhiev Y, *et al.* 2014. Two independent transcription initiation codes overlap on vertebrate core promoters. Nature 507, 381–385.

Hamaji T, Lopez D, Pellegrini M, Umen J. 2016. Identification and Characterization of a *cis* -Regulatory Element for Zygotic Gene Expression in *Chlamydomonas reinhardtii*. G3 Genes|Genomes|Genetics 6, 1541–1548.

Hamilton ML, Haslam RP, Napier JA, Sayanova O. 2014. Metabolic engineering of Phaeodactylum tricornutum for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids. Metabolic Engineering 22, 3–9.

Hammel A, Sommer F, Zimmer D, Stitt M, Mühlhaus T, Schroda M. 2020. Overexpression of Sedoheptulose-1,7-Bisphosphatase Enhances Photosynthesis in *Chlamydomonas reinhardtii* and Has No Effect on the Abundance of Other Calvin-Benson Cycle Enzymes. Frontiers in Plant Science 11, 868.

Hampf M, Gossen M. 2007. Promoter Crosstalk Effects on Gene Expression. Journal of Molecular Biology 365, 911–920.

Helliwell KE, Scaife MA, Sasso S, Araujo APU, Purton S, Smith AG. 2014. Unraveling Vitamin B $_{12}$ -Responsive Gene Regulation in Algae. Plant Physiology 165, 388–397.

Hillson N, Caddick M, Cai Y, *et al.* 2019. Building a global alliance of biofoundries. Nature Communications 10, 2040.

Holowko MB, Frow EK, Reid JC, Rourke M, Vickers CE. 2021. Building a biofoundry. Synthetic Biology 6, ysaa026.

Im DJ, Jeong S-N, Yoo BS, Kim B, Kim D-P, Jeong W-J, Kang IS. 2015. Digital Microfluidic Approach for Efficient Electroporation with High Productivity: Transgene Expression of Microalgae without Cell Wall Removal. Analytical Chemistry 87, 6592–6599.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Iwai M, Ikeda K, Shimojima M, Ohta H. 2014. Enhancement of extraplastidic oil synthesis in *C hlamydomonas reinhardtii* using a type-2 diacylglycerol acyltransferase with a phosphorus starvation–inducible promoter. Plant Biotechnology Journal 12, 808–819.

Jensen PR, Hammer K. 1998. The Sequence of Spacers between the Consensus Sequences Modulates the Strength of Prokaryotic Promoters. Applied and Environmental Microbiology 64, 82–87.

Jiang W, Brueggeman AJ, Horken KM, Plucinak TM, Weeks DP. 2014. Successful Transient Expression of Cas9 and Single Guide RNA Genes in *Chlamydomonas reinhardtii*. Eukaryotic Cell 13, 1465–1469.

Jillette N, Du M, Zhu JJ, Cardoz P, Cheng AW. 2019. Split selectable markers. Nature Communications 10, 4968.

Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. Science 316, 1497–1502.

Jores T, Tonnies J, Dorrity MW, Cuperus JT, Fields S, Queitsch C. 2020. Identification of Plant Enhancers and Their Constituent Elements by STARR-seq in Tobacco Leaves. The Plant Cell 32, 2120–2131.

Jores T, Tonnies J, Wrightsman T, Buckler ES, Cuperus JT, Fields S, Queitsch C. 2021. Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. Nature Plants 7, 842–855.

Kawaji H, Lizio M, Itoh M, *et al.* 2014. Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. Genome Research 24, 708–717.

Kiefer AM, Niemeyer J, Probst A, Erkel G, Schroda M. 2022. Production and secretion of functional SARS-CoV-2 spike protein in *Chlamydomonas reinhardtii*. Frontiers in Plant Science 13, 988870.

Kindle KL. 1987. Expression of a gene for a light-harvesting chlorophyll a/b-binding protein in *Chlamydomonas reinhardtii*: effect of light and acetate. Plant Molecular Biology 9, 547–563.

Kinney JB, Murugan A, Callan CG, Cox EC. 2010. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. Proceedings of the National Academy of Sciences 107, 9158–9163.

Knudsen S. 1999. Promoter2.0: for the recognition of PolII promoter sequences. Bioinformatics 15, 356–361.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Kotopka BJ, Smolke CD. 2020. Model-driven generation of artificial yeast promoters. Nature Communications 11, 2113.

Kropat J, von Gromoff ED, Müller FW, Beck CF. 1995. Heat shock and light activation of a *Chlamydomonas* HSP70 gene are mediated by independent regulatory pathways. Molecular and General Genetics 248, 727–734.

Kucho K, Ohyama K, Fukuzawa H. 1999. CO2-Responsive Transcriptional Regulation of *CAH1* Encoding Carbonic Anhydrase Is Mediated by Enhancer and Silencer Regions in *Chlamydomonas reinhardtii*. Plant Physiology 121, 1329–1337.

Kumar A, Falcao VR, Sayre RT. 2013. Evaluating nuclear transgene expression systems in *Chlamydomonas reinhardtii*. Algal Research 2, 321–332.

Kurihara Y, Makita Y, Kawashima M, Fujita T, Iwasaki S, Matsui M. 2018. Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in *Arabidopsis*. Proceedings of the National Academy of Sciences 115, 7831–7836.

Lai H-Y, Zhang Z-Y, Su Z-D, Su W, Ding H, Chen W, Lin H. 2019. iProEP: A Computational Predictor for Predicting Promoter. Molecular Therapy - Nucleic Acids 17, 337–346.

Lauersen KJ, Huber I, Wichmann J, *et al.* 2015. Investigating the dynamics of recombinant protein secretion from a microalgal host. Journal of Biotechnology 215, 62–71.

Le NQK, Yapp EKY, Nagasundaram N, Yeh H-Y. 2019. Classifying Promoters by Interpreting the Hidden Information of DNA Sequences via Deep Learning and Combination of Continuous FastText N-Grams. Frontiers in Bioengineering and Biotechnology 7, 305.

Lee S, Lee YJ, Choi S, Park S-B, Tran Q-G, Heo J, Kim H-S. 2018. Development of an alcohol-inducible gene expression system for recombinant protein expression in *Chlamydomonas reinhardtii*. Journal of Applied Phycology 30, 2297–2304.

Lemaire S, Keryer E, Stein M, Schepens I, Issakidis-Bourguet E, Gérard-Hirne C, Miginiac-Maslow M, Jacquot J-P. 1999. Heavy-Metal Regulation of Thioredoxin Gene Expression in *Chlamydomonas reinhardtii*. Plant Physiology 120, 773–778.

Levy L, Anavy L, Solomon O, Cohen R, Brunwasser-Meirom M, Ohayon S, Atar O, Goldberg S, Yakhini Z, Amit R. 2017. A Synthetic Oligo Library and Sequencing Approach Reveals an Insulation Mechanism Encoded within Bacterial σ 54 Promoters. Cell Reports 21, 845–858.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Li X, Li X, Yang X, Lan C, Huang Y, Jia B. 2022. Identification and Characterization of ATP-Binding Cassette Transporters in *Chlamydomonas reinhardtii*. Marine Drugs 20, 603.

Li X, Zhang R, Patena W, *et al.* 2016. An Indexed, Mapped Mutant Library Enables Reverse Genetics Studies of Biological Processes in *Chlamydomonas reinhardtii*. The Plant Cell 28, 367–387.

Liu W, Stewart CN. 2016. Plant synthetic promoters and transcription factors. Current Opinion in Biotechnology 37, 36–44.

Liu W, Yuan JS, Stewart Jr CN. 2013. Advanced genetic tools for plant biotechnology. Nature Reviews Genetics 14, 781–793.

Lodha M, Schroda M. 2005. Analysis of Chromatin Structure in the Control Regions of the *Chlamydomonas* HSP70A and RBCS2 Genes. Plant Molecular Biology 59, 501–513.

Lodha M, Schulz-Raffelt M, Schroda M. 2008. 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. Eukaryotic Cell 7, 172–176.

López García de Lomana A, Schäuble S, Valenzuela J, *et al.* 2015. Transcriptional program for nitrogen starvation-induced lipid accumulation in *Chlamydomonas reinhardtii*. Biotechnology for Biofuels 8, 207.

López-Paz C, Liu D, Geng S, Umen JG. 2017. Identification of *Chlamydomonas reinhardtii* endogenous genic flanking sequences for improved transgene expression. The Plant Journal 92, 1232–1244.

Loppes R, Radoux M, Ohresser MCP, Matagne RF. 1999. Transcriptional regulation of the Nia1 gene encoding nitrate reductase in *Chlamydomonas reinhardtii*: effects of various environmental factors on the expression of a reporter gene under the control of the Nia1 promoter. Plant Molecular Biology 41, 701–711.

Loppes R, Radoux M. 2001. Identification of short promoter regions involved in the transcriptional expression of the nitrate reductase gene in *Chlamydomonas reinhardtii*. Plant Molecular Biology 45, 215–227.

Lubliner S, Regev I, Lotan-Pompan M, Edelheit S, Weinberger A, Segal E. 2015. Core promoter sequence in yeast is a major determinant of expression level. Genome Research 25, 1008–1017.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Matharu N, Ahituv N. 2015. Minor Loops in Major Folds: Enhancer–Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. PLOS Genetics 11, e1005640.

Mathieu-Rivet E, Scholz M, Arias C, *et al.* 2013. Exploring the N-glycosylation Pathway in *Chlamydomonas reinhardtii* Unravels Novel Complex Structures. Molecular & Cellular Proteomics 12, 3160–3183.

Mayfield SP, Manuell AL, Chen S, Wu J, Tran M, Siefker D, Muto M, Marin-Navarro J. 2007. Chlamydomonas reinhardtii chloroplasts as protein factories. Current Opinion in Biotechnology 18, 126–133.

McQuillan JL, Berndt AJ, Sproles AE, Mayfield SP, Pandhal J. 2022. Novel cis-regulatory elements as synthetic promoters to drive recombinant protein expression from the *Chlamydomonas reinhardtii* nuclear genome. New Biotechnology 68, 9–18.

Mehariya S, Goswami RK, Karthikeysan OP, Verma P. 2021. Microalgae for high-value products: A way towards green nutraceutical and pharmaceutical compounds. Chemosphere 280, 130553.

Mendoza-Parra M-A, Saravaki V, Cholley P-E, Blum M, Billoré B, Gronemeyer H. 2016. Antibody performance in ChIP-sequencing assays: From quality scores of public data sets to quantitative certification. F1000Research 5, 54.

Merchant SS, Prochnik SE, Vallon O, *et al.* 2007. The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions. Science 318, 245–250.

Min S, Lee B, Yoon S. 2016. Deep learning in bioinformatics. Briefings in Bioinformatics 18, 851-869.

Mori Y, Shirai T. 2018. Designing artificial metabolic pathways, construction of target enzymes, and analysis of their function. Current Opinion in Biotechnology 54, 41–44.

Morioka MS, Kawaji H, Nishiyori-Sueki H, Murata M, Kojima-Ishiyama M, Carninci P, Itoh M. 2020. Cap Analysis of Gene Expression (CAGE): A Quantitative and Genome-Wide Assay of Transcription Start Sites. In: Boegel S, ed. Bioinformatics for Cancer Immunotherapy. New York, NY: Springer US, 277–301.

Morton T, Petricka J, Corcoran DL, Li S, Winter CM, Carda A, Benfey PN, Ohler U, Megraw M. 2014. Paired-End Analysis of Transcription Start Sites in *Arabidopsis* Reveals Plant-Specific Promoter Signatures. The Plant Cell 26, 2746–2760.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Neupert J, Gallaher SD, Lu Y, Strenkert D, Segal N, Barahimipour R, Fitz-Gibbon ST, Schroda M, Merchant SS, Bock R. 2020. An epigenetic gene silencing pathway selectively acting on transgenic DNA in the green alga *Chlamydomonas*. Nature Communications 11, 6269.

Neupert J, Karcher D, Bock R. 2009. Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. The Plant Journal 57, 1140–1150.

Niemeyer J, Scheuring D, Oestreicher J, Morgan B, Schroda M. 2021. Real-time monitoring of subcellular H2O2 distribution in *Chlamydomonas reinhardtii*. The Plant Cell 33, 2935–2949.

O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. 2016. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165, 1280–1292.

Ohler U, Harbeck S, Niemann H, Noth E, Reese MG. 1999. Interpolated markov chains for eukaryotic promoter recognition. Bioinformatics 15, 362–369.

Oubounyt M, Louadi Z, Tayara H, Chong KT. 2019. DeePromoter: Robust Promoter Predictor Using Deep Learning. Frontiers in Genetics 10, 286.

Pandelakis M, Delgado E, Ebrahimkhani MR. 2020. CRISPR-Based Synthetic Transcription Factors *In Vivo*: The Future of Therapeutic Cellular Programming. Cell Systems 10, 1–14.

Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. 2009. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nature Biotechnology 27, 1173–1175.

Perozeni F, Cazzaniga S, Baier T, Zanoni F, Zoccatelli G, Lauersen KJ, Wobbe L, Ballottari M. 2020. Turning a green alga red: engineering astaxanthin biosynthesis by intragenic pseudogene revival in *Chlamydomonas reinhardtii*. Plant Biotechnology Journal 18, 2053–2067.

Policastro RA, Zentner GE. 2021. Global approaches for profiling transcription initiation. Cell Reports Methods 1, 100081.

Quinn J, Merchant S. 1995. Two copper-responsive elements associated with the *Chlamydomonas* Cyc6 gene function as targets for transcriptional activators. Plant Cell 7, 623–628.

Quinn JM, Barraco P, Eriksson M, Merchant S. 2000. Coordinate Copper- and Oxygen-responsive Cyc6 and Cpx1 Expression in *Chlamydomonas* Is Mediated by the Same Element. Journal of Biological Chemistry 275, 6080–6089.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Quinn JM, Kropat J, Merchant S. 2003. Copper Response Element and Crr1-Dependent Ni^{2+}-Responsive Promoter for Induced, Reversible Gene Expression in *Chlamydomonas reinhardtii*. Eukaryotic Cell 2, 995–1002.

Redden H, Morse N, Alper HS. 2014. The synthetic biology toolbox for tuning gene expression in yeast. FEMS Yeast Research 15, 1-10.

Reese MG. 2001. Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Computers & Chemistry 26, 51–56.

Reiss DJ, Baliga NS, Bonneau R. 2006. Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. BMC Bioinformatics 7, 280.

Riethoven J-JM. 2010. Regulatory Regions in DNA: Promoters, Enhancers, Silencers, and Insulators. In: Ladunga I, ed. Computational Biology of Transcription Factor Binding. Totowa, NJ: Humana Press, 33–42.

Romanova N, Noll T. 2017. Engineered and Natural Promoters and Chromatin-Modifying Elements for Recombinant Protein Expression in CHO Cells. Biotechnology Journal 13, 1700232.

Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interaction landscape of gene promoters. Nature 489, 109–113.

Sawyer AL, Hankamer BD, Ross IL. 2015. Sulphur responsiveness of the *Chlamydomonas reinhardtii* LHCBM9 promoter. Planta 241, 1287–1302.

Scaife MA, Nguyen GTDT, Rico J, Lambert D, Helliwell KE, Smith AG. 2015. Establishing *Chlamydomonas reinhardtii* as an industrial biotechnology host. The Plant Journal 82, 532–546.

Scherf M, Klingenhoff A, Werner T. 2000. Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. Journal of Molecular Biology 297, 599–606.

Schroda M, Beck CF, Vallon O. 2002. Sequence elements within an *HSP70* promoter counteract transcriptional transgene silencing in *Chlamydomonas*: Counteracting transgene silencing in *Chlamydomonas*. The Plant Journal 31, 445–455.

Schroda M, Blocker D, Beck CF. 2000. The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. The Plant Journal 21, 121–131.

Schroda M. 2019. Good News for Nuclear Transgene Expression in *Chlamydomonas*. Cells 8, 1534.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Scranton MA, Ostrand JT, Georgianna DR, Lofgren SM, Li D, Ellis RC, Carruthers DN, Dräger A, Masica DL, Mayfield SP. 2016. Synthetic promoters capable of driving robust nuclear gene expression in the green alga *Chlamydomonas reinhardtii*. Algal Research 15, 135–142.

Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, Zeevi D, Keren L, Yakhini Z, Weinberger A, Segal E. 2012. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nature Biotechnology 30, 521–530.

Shen Z, Bao W, Huang D-S. 2018. Recurrent Neural Network for Predicting Transcription Factor Binding Sites. Scientific Reports 8, 15270.

Shetty RP, Endy D, Knight TF. 2008. Engineering BioBrick vectors from BioBrick parts. Journal of Biological Engineering 2, 5.

Specht EA, Nour-Eldin HH, Hoang KTD, Mayfield SP. 2015. An improved ARS2-derived nuclear reporter enhances the efficiency and ease of genetic engineering in *Chlamydomonas*. Biotechnology Journal 10, 473–479.

Sproles AE, Berndt A, Fields FJ, Mayfield SP. 2022. Improved high-throughput screening technique to rapidly isolate *Chlamydomonas* transformants expressing recombinant proteins. Applied Microbiology and Biotechnology 106, 1677–1689.

Stevens DR, Roehaix J-D, Purton S. 1996. The bacterial phleomycin resistance gene ble as a dominant selectable marker in *Chlamydomonas*. Mol Gen Genet 251, 23–30.

Surzycki R, Greenham K, Kitayama K, Dibal F, Wagner R, Rochaix J-D, Ajam T, Surzycki S. 2009. Factors effecting expression of vaccines in microalgae. Biologicals 37, 133–138.

Suzuki S, Ohta K, Nakajima Y, Shigeto H, Abe H, Kawai A, Miura R, Kazuki Y, Oshimura M, Miki T. 2020. Meganuclease-Based Artificial Transcription Factors. ACS Synthetic Biology 9, 2679–2691.

Tas H, Grozinger L, Stoof R, de Lorenzo V, Goñi-Moreno Á. 2021. Contextual dependencies expand the re-usability of genetic inverters. Nature Communications 12, 355.

Thieffry A, Vigh ML, Bornholdt J, Ivanov M, Brodersen P, Sandelin A. 2020. Characterization of *Arabidopsis thaliana* Promoter Bidirectionality and Antisense RNAs by Inactivation of Nuclear RNA Decay Pathways. The Plant Cell 32, 1845–1867.

Tokizawa M, Kusunoki K, Koyama H, Kurotani A, Sakurai T, Suzuki Y, Sakamoto T, Kurata T, Yamamoto YY. 2017. Identification of Arabidopsis genic and non-genic promoters by paired-end sequencing of TSS tags. The Plant Journal 90, 587–605.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Traewachiwiphak S, Yokthongwattana C, Ves-Urai P, Charoensawan V, Yokthongwattana K. 2018. Gene expression and promoter characterization of heat-shock protein 90B gene (HSP90B) in the model unicellular green alga *Chlamydomonas reinhardtii*. Plant Science 272, 107–116.

Umarov R, Kuwahara H, Li Y, Gao X, Solovyev V. 2019. Promoter analysis and prediction in the human genome using sequence-based deep learning models (J Hancock, Ed.). Bioinformatics 35, 2730–2737.

Umarov RKh, Solovyev VV. 2017. Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks (IB Rogozin, Ed.). PLOS ONE 12, e0171410.

Ushijima T, Hanada K, Gotoh E, *et al.* 2017. Light Controls Protein Localization through Phytochrome-Mediated Alternative Promoter Selection. Cell 171, 1316-1325.e12.

Vazquez-Vilar M, Garcia-Carpintero V, Selma S, *et al.* 2020. Edition of complex gene families in tobacco with GoldenBraid 4.0, a multipurpose web-based platform for plant genome engineering. bioRxiv doi: 10.1101/2020.10.06.327841. [Preprint]. Venter M. 2007. Synthetic promoters: genetic control through cis engineering. Trends in Plant Science 12, 118–124.

Villand P, Eriksson M, Samuelsson G. 1997. Carbon dioxide and light regulation of promoters controlling the expression of mitochondrial carbonic anhydrase in *Chlamydomonas reinhardtii*. Biochemical Journal 327, 51–57.

Wang H, Cimen E, Singh N, Buckler E. 2020. Deep learning for plant genomics and crop improvement. Current Opinion in Plant Biology 54, 34–41.

Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. 2011. A Modular Cloning System for Standardized Assembly of Multigene Constructs (J Peccoud, Ed.). PLoS ONE 6, e16765.

Weimann M, Grossmann A, Woodsmith J, *et al.* 2013. A Y2H-seq approach defines the human protein methyltransferase interactome. Nature Methods 10, 339–342.

Wittkopp PJ, Kalay G. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature Reviews Genetics 13, 59–69.

Yamamoto YY, Ichida H, Abe T, Suzuki Y, Sugano S, Obokata J. 2007. Differentiation of core promoter architecture between plants and mammals revealed by LDSS analysis. Nucleic Acids Research 35, 6219–6226.

Yamano T, Iguchi H, Fukuzawa H. 2013. Rapid transformation of *Chlamydomonas reinhardtii* without cell-wall removal. Journal of Bioscience and Bioengineering 115, 691–694.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Yang A, Zhang W, Wang J, Yang K, Han Y, Zhang L. 2020. Review on the Application of Machine Learning Algorithms in the Sequence Data Mining of DNA. Frontiers in Bioengineering and Biotechnology 8, 1032.

Yang J-S, Garriga-Canut M, Link N, Carolis C, Broadbent K, Beltran-Sastre V, Serrano L, Maurer SP. 2018. rec-YnH enables simultaneous many-by-many detection of direct protein–protein and protein–RNA interactions. Nature Communications 9, 3747.

Yoo BS, Im DJ, Ahn MM, Park SJ, Kim YH, Um TW, Kang IS. 2018. A continuous droplet electroporation system for high throughput processing. The Analyst 143, 5785–5791.

Yu H, Tardivo L, Tam S, *et al.* 2011. Next-generation sequencing to generate interactome datasets. Nature Methods 8, 478–480.

Yus E, Yang J-S, Sogues A, Serrano L. 2017. A reporter system coupled with high-throughput sequencing unveils key bacterial transcription and translation determinants. Nature Communications 8, 368.

Zhang R, Patena W, Armbruster U, Gang SS, Blum SR, Jonikas MC. 2014. High-Throughput Genotyping of Green Algal Mutants Reveals Random Distribution of Mutagenic Insertion Sites and Endonucleolytic Cleavage of Transforming DNA. The Plant Cell 26, 1398–1409.

Zhu Y, Li F, Xiang D, Akutsu T, Song J, Jia C. 2021. Computational identification of eukaryotic promoters based on cascaded deep capsule neural networks. Briefings in Bioinformatics 22, bbaa299.

Zorin B, Hegemann P, Sizova I. 2005. Nuclear-Gene Targeting by Using Single-Stranded DNA Avoids Illegitimate DNA Integration in *Chlamydomonas reinhardtii*. Eukaryotic Cell 4, 1264–1272.

Zrimec J, Börlin CS, Buric F, Muhammad AS, Chen R, Siewers V, Verendel V, Nielsen J, Töpel M, Zelezniak A. 2020. Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure. Nature Communications 11, 6141.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Table 1. Constitutive and inducible synthetic promoters and putative CREs. The main features of the synthetic promoters developed in *C. reinhardtii* are summarized, including name, origin (species/construction strategy), size, CREs, inducibility, and responsiveness to environmental conditions.

Name	Origin	Size (bp)	CREs	Description	References
AlcR-P_{alcA}	AlcR-P*alcA* system *from Aspergillus nidulans*	246	Three AlcR binding sites	Alcohol inducible; Co-expression of *alc*R under psaD promoter is required	Lee et al., 2018
AR	Hybrid of *HSP70A* and *RBCS2* promoter from *C. reinhardtii*	461	see table 2 for CREs of the promoter parts.	Constitutive; Light and heat-shock increase transcription	Schroda et al., 2000, 2002
A*β*SAP(i)	Hybrid of *HSP70A* and a synthetically optimized version of *TUB2* promoter from *C. reinhardtii*	473 (+ 268 of intronized 5' UTR)	ATANTT, CCCATGCA	Constitutive; Light and heat-shock increase transcription	Einhaus et al., 2021
pCRE-12	*In silico* identified CREs for *C. reinhardtii*	116	GGGCCCATTC; (6 motif repeats, connected to a 50 bp core promoter)	Constitutive	McQuillan et al., 2022
RIA3/PromC	Hybrid of core promoter and a salt-responsive element of the *GPDH3* promoter from *C. reinhardtii*	1230 (+ 285 of 5' UTR)	see table 3 for CREs of the promoter parts.	Salt inducible; Concentrations of above 5 mM NaCl or 100 mM KCl are necessary	Beltran-Aguilar et al., 2019

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

sap 11	Computationally generated from CREs of *C. reinhardtii*	500	CCCATGCA	Constitutive	Scranton et al., 2016
SORLIPX2	Hybrid of core promoter and two light-responsive elements of the *LIP* promoter of *Dunaliella* sp.	139	GGGCCAC	Light inducible; Strong induction through intensities above 300 μmol photons m^{-2} s^{-1}	Baek et al., 2016

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Table 2. Endogenous constitutive promoters and putative CREs. The main features of the natural promoters identified in *C. reinhardtii* are summarized, including name, motifs, and responsiveness to environmental conditions.

Name	Motifs and elements	Description	References
ARG7	Unknown	Constitutive Strength reported as high as that of the AR promoter	Specht et al., 2015
FDX1 (also known as PETF)	Unknown	Constitutive	López-Paz et al., 2017
HSP70A	5 HSE[C] 4 CCAAT boxes (1 regular, 3 inverted)[C] TGAAG (-146 to -141, probably involved in light regulation)[A]	Constitutive; Transcription increases following: - Heat shock - Light	[A] Kropat et al., 1995 [B] Schroda et al., 2000 [C] Lodha and Schroda, 2005
IDA5	Unknown	Constitutive	Kumar et al., 2013
PSAD	Unknown	Constitutive	Fischer and Rochaix, 2001
RBCS2	1 Heat-shock element (HSE) (3 nGAAn repeats in alternating orientations to make one functional HSE)[B]	Constitutive	[A] Schroda et al., 2000 [B] Lodha and

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

	3 CCAAT (1 regular, 2 inverted)[B]		Schroda, 2005
RPL23	Unknown	Constitutive	López-Paz et al., 2017
RPL35a	Unknown	Constitutive	López-Paz et al., 2017
TUB2	GC rich region 10-11 bp long, between TATA box and TSS[A] 7 copies of "Tub box" (GCTC(G/C)AAGGC) ~150 bp upstream of the TSS[A, D]	Constitutive; Activity increases: - After deflagellation - In the mitotic phases of the cell cycle	[A] Brunke et al., 1984 [B] Bandziulis and Rosenbaum, 1988 [C] Davies et al., 1992 [D] Davies and Grossman, 1994

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Table 3. Endogenous inducible promoters and putative CREs. The main features of the natural inducible promoters identified in *C. reinhardtii* are summarized, including name, motifs, inducible factors, and responsiveness to environmental conditions.

Name	Motifs and elements	Inducible factors	References
AOX1	Unknown	- Nitrate - Sodium azide - Nitrate and copper in combination	Baurain et al., 2003
ATX1	GTCGCACTGGCATGT (–529 to –515) GCAGCGATGGCATTT (–300 to –286) Consensus sequence: GNNGCNNTGGCATNT	- Iron deficiency	Fei and Deng, 2007
CAH1 (Ca1)	GGGTTGAANTCCC (−553 to −541)[A] CGCGCC (−319 to −313)[A] AACCCCNGNTGCA (−157 to −145)[A] CAAT box[A]	- Low CO2 - Light	[A] Kucho et al., 1999 [B] Villand et al., 1997
CPX1	GTAC (copper responsive element)	- Copper deficiency - Oxygen deficiency - Nickel - Cobalt	[A] Quinn et al., 2000 [B] Quinn et al., 2003
CYC6	GTAC (-124 to -121, copper responsive element)	- Copper deficiency - Nickel - Cobalt-Oxygen deficiency	[A] Quinn and Merchant, 1995 [B] Quinn et al., 2000 [C] Quinn et al., 2003
FEA1	CTGCGGTGGCAAAGT (-273 to -259)[C] CCGCCGCNNNTGGCACCAGCCT (-106 to -85)[C] Core motif: TGGCAC	Repressed by: - Iron deficiency	[A] Allen et al., 2007 [B] Barjona do Nascimento Coutinho et al., 2019 [C] Fei et al., 2009 [D] Kumar et al., 2013
FOX1	CACACG (-87 to -82) CACGCG (-65 to -60)	- Iron deficiency	Deng and Eriksson, 2007
FTR1	ATGCAGGCT (-287 to -279) AAGCGATTGCCAGAGCGC (-253 to	- Iron deficiency	Fei et al., 2010

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

	-236)		
GPDH 3	11 putative CAAT boxes Salt-responsive motif unknown	- NaCl above 100 mM - KCl above 100 mM	Beltran-Aguilar et al., 2019
GPX5	16 bp palindrome (GCGCCAACGTTGACGC) CRE/AP-1 binding site (TGACGCCA) GC-box or Sp1 element (CCGCCC) two CAAT boxes (CAAT and ATTG)	- Rose bengal (RB) - High light	Fischer et al., 2009
HSP90 B	6 CAAT boxes 17 GC boxes 3 HSE	- Heat - ER stress	Traewachiwiphak et al., 2018
LHCB M6	Unknown	- Light - Heat shock	[A] Blankenship and Kindle, 1992 [B] Kindle, 1987
LHCB M9	Unknown	- Sulfur deficiency - Combined anaerobiosis and sulfur deficiency - Light	Sawyer et al. 2015
METE	Unknown	Repressed by: - Cobalamin	Helliwell et al., 2014
NIA1 (NIT1)	No TATA box[B] 2 repeated GGA/TAGGGT (-231 to -219, -76 to -65 upstream TSS)[B] CGAACTT (-51 to -42) GGCCCCGGG (-33 to -8) ATGGTAGGGT (−232 to −223) AGGGAAGGGT (−80 to −69) HSEs (CNNGAANNTTC, −131 to −46) GA/TAG (−231 to −219 and −76 to −65, light induction)	- Light - Ammonium deficiency - Nitrate - Nitrite	[A] Loppes et al., 1999 [B] Loppes and Radoux, 2001
SQD2	Unknown	- Phosphorus deficiency	Iwai et al., 2014
TRXH1	TCACG…TCACT (-384 to -368) TCACG…TGGCG (-360 to -344) Consensus sequence: TGACG (activation sequence-1 element)	- Mercury - Cadmium	Lemaire et al., 1999

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Figure 1: A schematic microalgal promoter. The core promoter encompasses the region surrounding the transcription start site (TSS), directly upstream of the gene of interest. The pre-initiation complex (PIC), which forms the basic transcriptional machinery and contains RNA polymerase II (RNApolII), is recruited to the core promoter to initiate transcription, guided by regulatory DNA elements such as the TATA box. Transcription factors (TFs) bind to specific DNA sequence motifs, or cis-regulatory elements (CREs), within the proximal promoter region, where they interact with the PIC to modulate transcription. Distal promoter regions can also influence transcription by stabilizing or disrupting the PIC, mediated by TFs.

Figure 2: Design strategies for synthetic genetic control elements. (A) Summary of top-down approaches. Synthetic promoters are generated by modifying native ones either by motif mutation, modification (insertion, deletion, reorganization), promoter truncation or hybridization. CREs and TSSs are represented by geometric shapes and arrows, respectively. (B) Schematic overview of key features of measured transcriptional output from synthetic promoters and their native counterpart. (C) Summary of bottom-up approaches. Synthetic promoters are generated *de novo* by computational design (model-driven strategies), assembly from motif collections, or nucleotide randomization.

Figure 3. High-throughput experimental techniques to interrogate endogenous and synthetic promoters. (A) Workflow for fluorescence-activated cell sorting (FACS) coupled to promoter sequencing (FACS-seq). Constructs containing a promoter library and a fluorescent protein reporter gene are transformed into cells; cells are then isolated according to *in vivo* fluorescence intensity using FACS, and the barcoded promoters sequenced. (B) Overview of self-transcribing active regulatory region sequencing (STARR-seq). Constructs comprising a promoter library and barcoded reporter gene are transformed into cells; mRNA is then extracted, analyzed and quantified by NGS. (C) Workflow for DNA affinity purification sequencing (DAP-seq). Genomic DNA (gDNA) is extracted and fragmented, followed by adapter ligation; the gDNA library is affinity purified using immobilized transcription factors (TFs) fused to affinity tags. Captured DNA containing transcription factor binding sites (TFBSs) are eluted and subjected to NGS. (D) Overview of chromatin immunoprecipitation sequencing (ChIP-seq). Proteins and gDNA are chemically cross-linked, extracted and fragmented. TF-bound DNA fragments are immunoprecipitated using TF-specific antibodies and sequenced. (E) Overview of cap analysis of gene expression sequencing (CAGE-seq). Extracted mRNA is reverse transcribed and the cap regions biotinylated, enabling 'cap-trapping' of cognate single-stranded DNA (ssDNA), which is sequenced and matched to the reference genome. (F) Overview of expression level monitoring by DNA methylation sequencing (ELM-seq). Constructs comprising a promoter library, 4x GATC motifs, and DNA methylase reporter gene (*dam*) are

Accepted Manuscript
Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

transformed into the host; gDNA is then extracted and digested with two restriction enzymes that discriminate between methylated and unmethylated GATC motifs; the ratio between methylated and unmethylated constructs, determined via sequencing, reveals promoter efficiency. (G) Typical outputs and more related techniques of the high-throughput experiments for promoter studies: Gene expression quantification (A, B, E, F), motif discovery and TFBS identification (A, B, C, D, E, F), promoter and TSS determination (A, E)

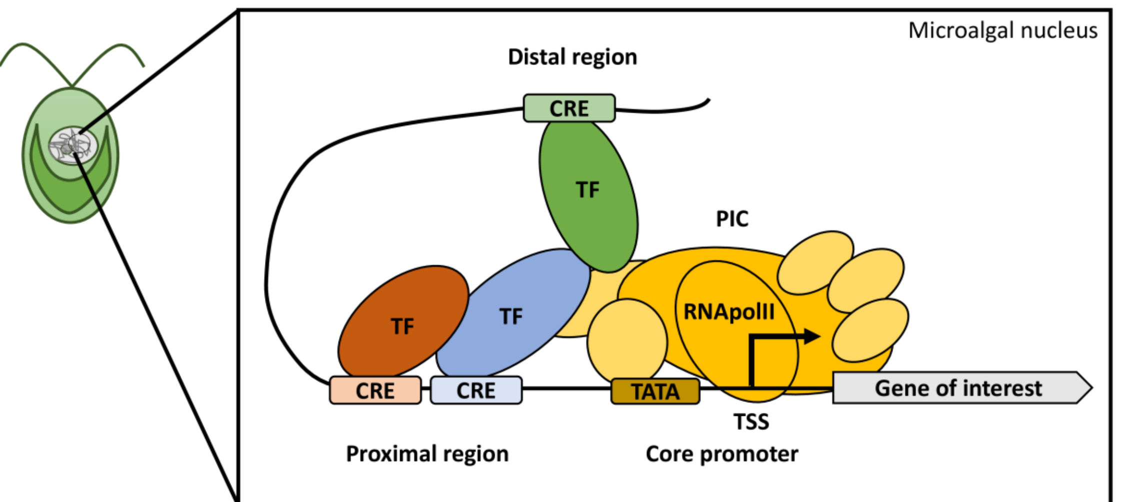
Figure 4. Challenges and current obstacles of *Chlamydomonas* promoter study. (A) Incomplete knowledge on promoter annotation and regulatory networks. TSS and TSR represent the transcription start site and transcription starting region, respectively. Conceptional physical interaction network of TFs (transcription factors) and CREs (cis-regulatory elements) are shown here with black lines if TFs bind to specific CREs. Also, TFs are connected with each other with blue lines if they have physical interaction. (B) Positional effect. Schematic figure showing the nucleus and transgene incorporated in a various location of the genome. Arbitrary gene expression level is represented by color. (C) Interference of endogenous nucleases. GOI represents the gene of interest. The input and output of 'AND' logic gate is represented by a table. (D) Crosstalk between promoter and terminator. Schematic representation of the transcription cycle of RNA polymerase II. Nascent mRNA is represented as the red curved line. The transcription reinitiation is started by recycling RNA polymerase after releasing terminated RNA.

Accepted Manuscript

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

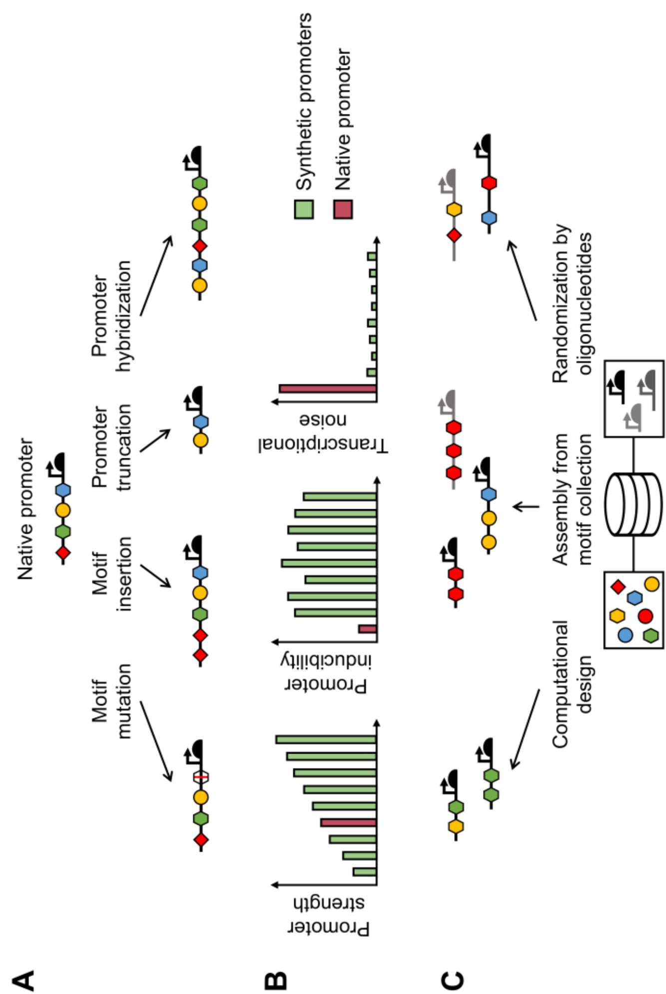
Figure 1



Accepted Manuscript

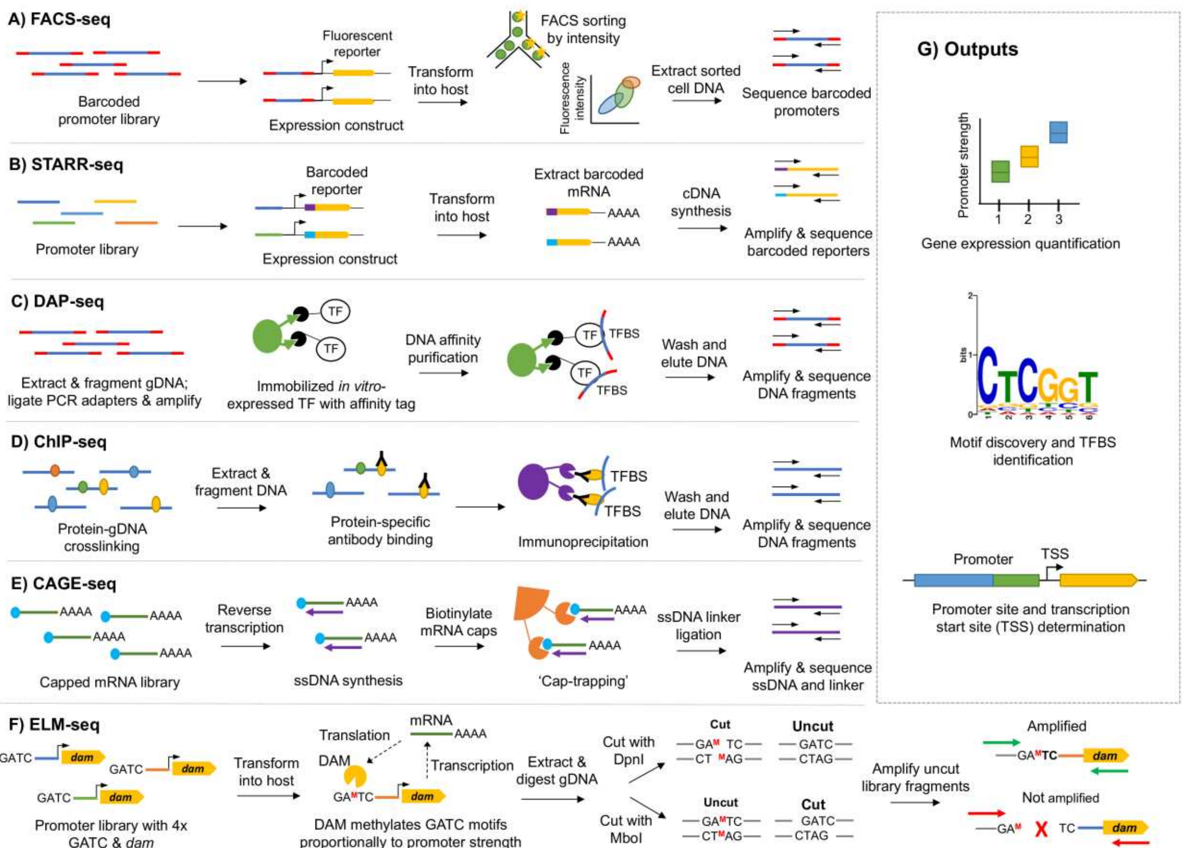
Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Figure 2



Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Figure 3



Downloaded from https://academic.oup.com/jxb/advance-article/doi/10.1093/jxb/erad100/7110540 by University of Sheffield user on 11 April 2023

Figure 4