UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

STRATEGIES DE PRODUCTION DE METABOLITES A HAUTE VALEUR PAR INGENIERIE METABOLIQUE DE DIATOMEES

THÈSE PRÉSENTÉE COMME EXIGENCE PARTIELLE DU
DOCTORAT EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE

PAR
ANIS MESSAABI

AOUT 2025

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

DOCTORAT EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE

Direction de recherche :

_____

Isabel Desgagné-Penix Directrice de recherche

Jury d'évaluation de la thèse :

_____

Isabel Desgagné-Penix Directrice de recherche

_____

Emmanuel Milot Président de jury

_____

David Charbonneau Évaluateur interne

_____

Muriel Bardor Évaluateur externe

Thèse soutenue le 16 Juillet 2025

La science, mon garçon, est faite d'erreurs, mais d'erreurs qu'il est bon de commettre, car elles mènent peu à peu à la vérité.

Jules Verne

La main qui berce l'enfant est la main qui domine le monde

Curtis Hanson

Je tiens à dédier ce travail à ma mère, qui n'a jamais cessé de croire, de croire en moi. Tu m'as appris à être unique et à toujours persévérer.

A Nadia, le sens de ma vie

Table des matières

List of Figures

List of Tables

Remerciements

Je tiens tout d'abord à exprimer ma gratitude à l'ensemble des membres du jury pour avoir accepté d'évaluer ce travail de thèse. Pour le temps précieux que vous avez consacré à la lecture et à l'analyse de mon manuscrit.

Je tiens également à remercier les Professeurs Hugo Germain, Degbedji Tagnon Missihoun et Marc Germain qui m'ont évalué lors de mes précédents examens. Vos conseils ont été précieux. Un grand merci à vous.

Je remercie tout particulièrement ma directrice Professeure Isabel Desgagné-Penix, je tiens à vous exprimer ma gratitude pour la confiance que vous avez placée en moi dès le début de ce projet. En 2020 en tant que stagiaire, puis en 2021 au doctorat. Malgré les difficultés et les incompréhensions rencontré, votre soutien, dans mes projets scientifiques, m'a permis de m'épanouir pleinement.

À mon équipe, je ne saurais trouver les mots justes pour exprimer ma gratitude autrement qu'en vous disant un immense merci d'être ce que vous êtes. Cet environnement chaleureux et rassurant que nous avons construit ensemble, ces instants précieux de complicité, de rires et de fous rires, resteront gravés dans ma mémoire. Ce sentiment constant de confiance et de soutien m'a accompagnée tout au long de ces années, et je mesure pleinement la chance inestimable que j'ai eue de vous avoir à mes côtés. Petite pensée pour Serge, Vahid, Nicolas, Ayoub et Narimene !

À la Docteure Elisa Ines Fantino, tu as un jour dit que j'étais un magicien, et je t'ai répondu, que c'était toi qui m'avais appris la magie. Merci d'avoir été ma superviseur et ma mentor depuis 2020. Je t'ai toujours appelé « my lab's mom ». Sans toi, je ne serais le scientifique que je suis aujourd'hui. Pour ta douceur enfouie sous une carapace tel un membre des tortues ninja, ta bienveillance, et pour ce temps que nous avons eu la chance de partager ensemble. Les temps n'ont pas été facile entre nous, mais je ne retiens que le positif, sache le !

À la Docteure Natacha Mérindol, ta douceur, ton humour et ton soutien ont été des piliers infaillibles. Tu as toujours été une personne sur laquelle j'ai pu compter, une oreille attentive et infaillible. Tu as été un véritable modèle pour moi, et je te remercie profondément de m'avoir inspirée et montré la voie. Merci pour tout ce que tu as fait pour moi. Tu m'as ouvert les portes de chez toi et je tiens à souligner la famille formidable que vous formez avec ton mari et tes deux enfants.

À la Docteure Fatma Meddeb, sous le ton autoritaire, se cache une femme bienveillante, déterminée et prête à tout pour nous voir réussir. Ton soutien inconditionnel a été un pilier précieux tout au long de ces quatre années. Tu as toujours été une personne sur laquelle j'ai pu compter. Une seconde « maman » pour moi et je te remercie sans aucunes retenues.

À tous ceux qui ont contribué, de près ou de loin, à ces quatre années de thèse, un immense merci. Votre soutien et votre travail ont été essentiels au bon déroulement de ce parcours.

À mes précieux amis rencontrés à la faculté de Sorbonne, Loren, Anaël, Jules et Céline. Qui aurait cru que quatre d'entre nous finissent Docteur(e). Depuis les amphis pour certains, le bus de voyage direction Barcelone pour d'autres. Nous avons vécu des moments inoubliables au quarte coins de la France. Même dispersés, je sais que notre lien reste présent. Que ce soit à Tahiti, au Chili ou ailleurs, je ressens votre soutien et vos encouragements, et je vous en suis infiniment reconnaissant. À nos prochaines soirées ou escape Games.

À tous ceux que j'ai pu oublier, rencontrer chacun d'entre vous a rendu la vie plus douce ! Merci pour tout.

À mes grands-parents, mes oncles et tantes, mes cousins et cousines, vous êtes mes premiers soutiens, et je ressens pleinement l'amour et la fierté que vous me portez. À mon grand-père Abdelaziz, je sais que tu es fier de là où tu te trouves, je n'oublierai jamais ce que tu as pu apporter dans ma vie. Merci infiniment pour tout ce que vous êtes.

Et enfin, à ma mère... À ma maman, Nadia, cette superwoman qui m'a entourée d'un amour et d'un soutien inconditionnels tout au long de ma vie. Tu as été une inspiration pour moi, autant par ton parcours que par ton dévouement personnel. Toujours prête à affronter le monde entier pour moi, ton enfant, tu m'as inspirée par ta force et ton courage, sans relâche. Tu seras, aujourd'hui et pour toujours, la femme de ma vie, un pilier irremplaçable et une source infinie d'amour et de joie. Je sais que, dans l'ombre, tu as toujours travaillé sans relâche pour assurer notre avenir et ma réussite. Merci pour ton amour, pour nos moments de rires aux éclats, pour nos voyages, pour nos disputes et pour toutes choses. C'est grâce à toi que j'ai pu vivre mes rêves. Hâte de continuer à écrire nos histoires, ensemble et avec beaucoup d'amour. Je t'aime profondément maman.

RESUME EN FRANÇAIS

Depuis la récente légalisation du cannabis au Canada en 2018, le marché de la est en forte expansion. Plusieurs études ont permis de démontrer le potentiel médical et thérapeutique des cannabinoïdes, dans le traitement d'une variété de maladies, une action à la fois précise et large dans la fonction et dans l'effet physiologique. Ses cannabinoïdes sont utilisées lors des thérapies neurologiques et dans le traitement de plusieurs syndromes, tels que l'épilepsie, la maladie de Parkinson, d'Alzheimer, les tumeurs cérébrales et les traumatismes crânien. De plus, il a été démontré que chez l'humain, il existerait un système endocannabinoïde central, permettant d'interagir, via des récepteurs (CB1 et CB2), avec le cannabis. Ce système joue un rôle clé dans l'homéostasie générale et régule de nombreux processus physiologiques tels que l'inflammation, la perception de la douleur et l'immunité. Cependant, afin de minimiser l'impact environnemental causé par l'agriculture conventionnelle industrialisée, nous continuons à rechercher des approches modernes qui intègrent des solutions écologiques et durables pour l'avenir.

Les diatomées sont un groupe majeur du phytoplancton responsable de près de 50 % de la production primaire nette dans les océans. Malgré un effort considérable porté sur l'étude biologique de ces organismes, ce groupe de microalgue n'avait pas encore été étudié et utilisé pour la production de cannabinoïdes. La capacité photosynthétique de *Phaeodactylum tricornutum,* les outils moléculaires disponible et son fort potentiel de production de molécules d'intérêts, rendent l'étude et le développement d'un châssis biologique de production de cannabinoïdes inévitable.

Dans cette thèse, deux axes ont été investigués. Le premier visait à mettre en place une stratégie afin de produire l'acide cannabidiolique (CBDA) par la diatomée *P. tricornutum*. Cette stratégie nous a permis d'obtenir des clones exprimant et accumulant la protéine CBDAS, responsable de la production de CBDA. De plus, les essais enzymatiques ont été positif, montrant une activité détectable de cette enzyme. Notre deuxième axe d'investigation a consisté à élucider le clivage *in vivo* de la séquence de la thrombine LVPRGS dans *P. tricornutum,* afin de développer un outil moléculaire permettant l'expression polycistronique de plusieurs gènes, facilitant la production de cannabinoïdes dans *P. tricornutum.* En plus du clivage, nos résultats suggèrent l'existence de

nouvelles protéases de type thrombine chez *P. tricornutum.* De plus, la structure de ses protéases ont été prédit dans cette étude.

ABSTRACT

Since the legalization of cannabis in Canada in 2018, the cannabis market has experienced rapid growth. Numerous studies have demonstrated the medical and therapeutic potential of cannabinoids in treating a wide range of diseases, showcasing both targeted and broad physiological effects. These molecules are widely used in neurological therapies and for the treatment of various syndromes, including epilepsy, Parkinson's disease, Alzheimer's disease, brain tumors, and traumatic brain injuries. Furthermore, research has shown the existence of a central endocannabinoid system in humans, which interacts with cannabis through CB1 and CB2 receptors. This system plays a crucial role in overall homeostasis, regulating several physiological processes such as inflammation, pain perception, and immunity. However, to reduce the environmental impact caused by conventional industrialized agriculture, efforts continue to focus on developing modern approaches that incorporate eco-friendly and sustainable solutions for the future.

Diatoms represent a major group of phytoplankton, contributing nearly 50% of the net primary production in the oceans. Despite extensive biological studies on these organisms, diatoms have not yet been explored or utilized for cannabinoid production. The photosynthetic capacity of *P. tricornutum*, combined with the availability of molecular tools and its strong potential for synthesizing valuable biomolecules, makes the development of a biological platform for cannabinoid production both relevant and inevitable.

This thesis explores two main research directions. The first focuses on establishing a strategy for producing cannabidiolic acid (CBDA) using the diatom *P. tricornutum*. Through this approach, we successfully generated clones expressing and accumulating the Cannabidiolic acid synthase (CBDAS) protein, which is responsible for CBDA biosynthesis. Additionally, *in vitro* enzymatic assays confirmed detectable activity of the enzyme. The second research objective was to investigate the *in vivo* cleavage of the thrombin recognition sequence LVPRGS in *P. tricornutum*.

Our findings suggest not only the successful cleavage of this sequence but also the presence of unidentified thrombin-like proteases in *P. tricornutum*. Furthermore, the structure of these proteases was predicted as part of this study.

CHAPTER I

1 Introduction

1.1 Plant specialized metabolites

Plants synthesize hundreds of thousands of organic compounds. Based on their functions, researchers classify these compounds into three categories: primary metabolites, which are highly conserved and essential for plant growth and development[1]; hormones, which regulate developmental processes and signaling networks involved in plant responses to a wide range of biotic and abiotic stresses[2]; and specialized metabolites, previously named as secondary metabolites, which facilitate plant-environment interactions (Figure 1-1).

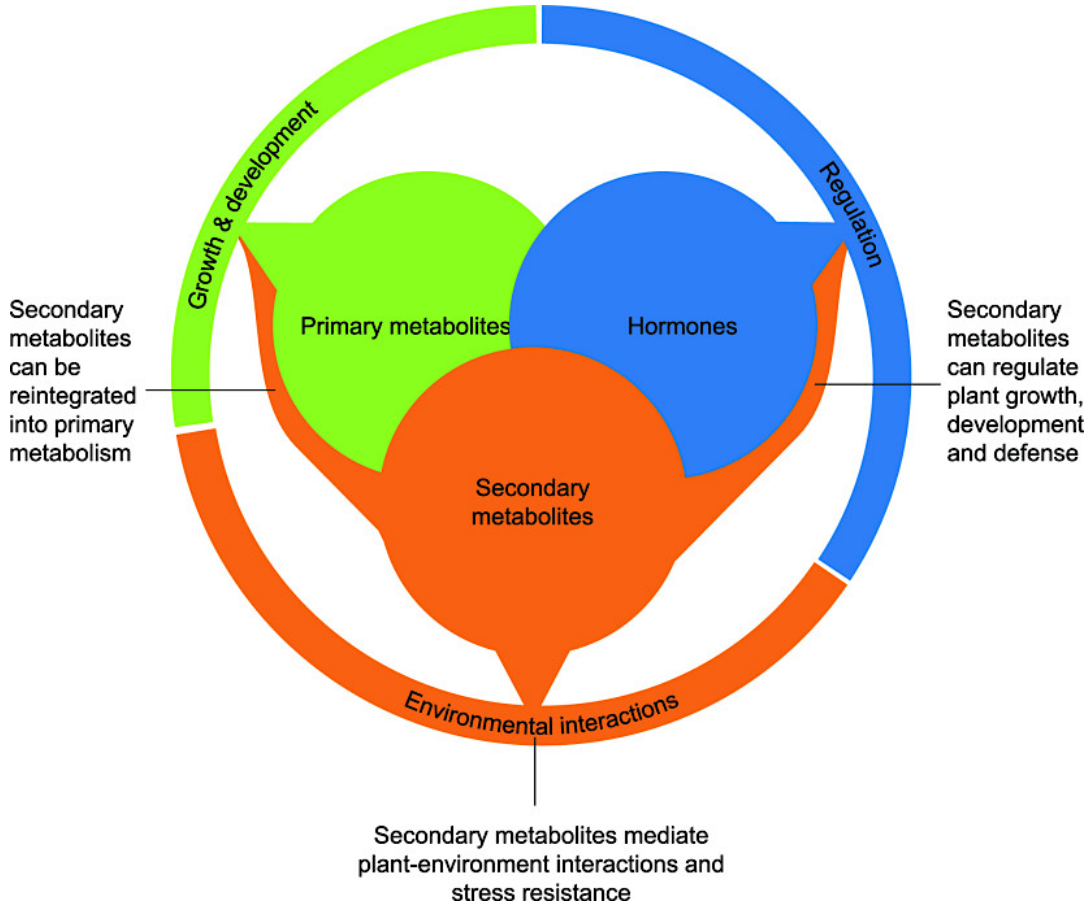


Figure 1-1: ***Organic compounds in plants, classified in three groups;*** *primary metabolites, hormones, or specialized metabolites (secondary metabolites). Source:[3]*

Plant specialized metabolites are organic compounds that are not directly involved in the primary metabolism, but play crucial roles in plant defense, communication, and adaptation to environmental stresses[4]. These compounds are highly diverse and often unique to specific plant species or groups. In *C. sativa*, the cannabinoids are biosynthesized with other various classes of specialized metabolites, including terpenoids, flavonoids, fatty acids, acyl sugars, and other[5, 6]. Interestingly, they are interconnected to the primary carbon metabolism and to each other, sharing pathways (Figure 1-2). Beyond their diverse functions in native organism, cannabinoids have a wide array of industrial applications.

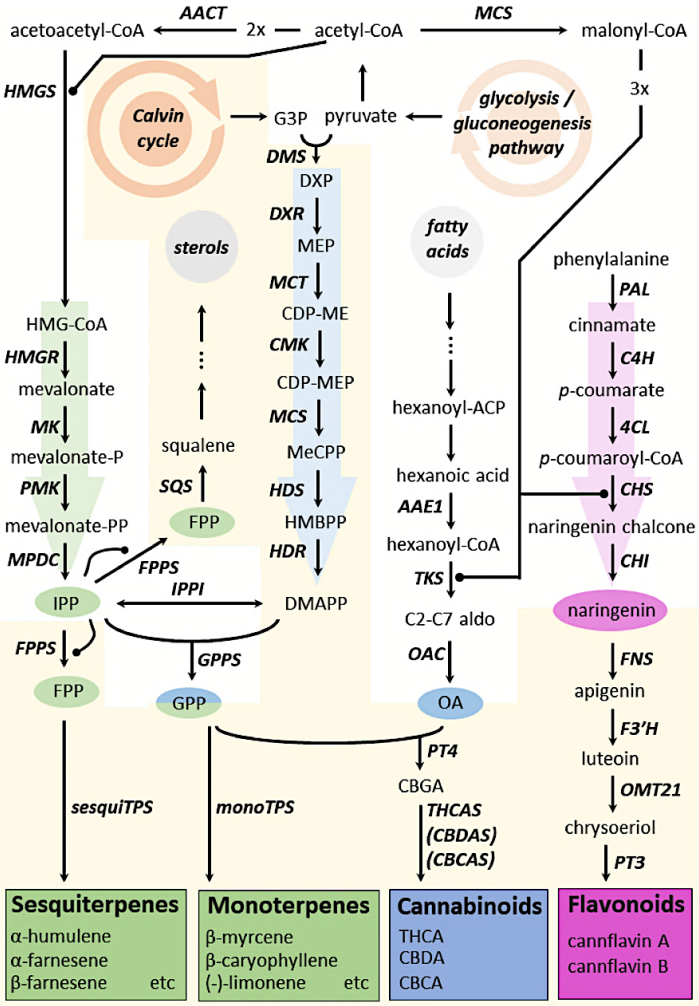


Figure 1-2: ***Biosynthesis pathways of some specialized metabolite in C. sativa, Terpenes (green), Cannabinoids (blue) and flavonoids (pink).***

The MEP pathway, the MVA pathway and the phenylpropanoid pathway are represented by thick blue, green and pink arrows respectively. White background represents cytosolic compartment, while pale orange background represents cellular compartments or the apoplast. Enzymes defined as follow: AACT: acetoacetyl-coenzyme A thiolase; MCS: malonyl-CoA synthetase; DMS: 1-deoxy-d-xylulose 5-phosphate synthase; DXR: 1-deoxy-D-xylulose 5-phosphate reductoisomerase; MCT: 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; HDS: 4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HDR: (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase; HMGS: hydroxymethylglutaryl-CoA synthase; HMGR: 3-hydroxy-3-methyl-

glutaryl-coenzyme A reductase; MK: mevalonate kinase; PMK: phosphomevalonate kinase; MPDC: mevalonate pyrophosphate decarboxylase; IPPI: isopentenyl diphosphate isomerase; GPPS: geranyl diphosphate synthase; AAE1: acyl activating enzyme 1; TKS: 3,5,7-Trioxododecanoyl-CoA synthase (polyketide synthase); OAC: olivetolic acid cyclase; FPPS: farnesyl pyrophosphate synthase; PAL: phenylalanine ammonia-lyase; C4H: cinnamate 4-hydroxylase; 4CL: 4-coumaroyl CoA-ligase; CHS: chalcone synthase; CHI: chalcone isomerase; SQS: squalene synthase; PT4: olivetolate geranyl transferase; THCAS: THCA synthase; CBDAS: CBDA synthase; CBCAS: CBCA synthase; sesquiTPS: all sesquiterpene synthases grouped under a single appellation; monoTPS: all monoterpene synthases grouped under a single appellation; FNS: flavone synthase; F3′H: flavonoid 3′-hydroxylase; OMT21: O-methyltransferase 21; PT3: prenyl transferase 3. Source:[6]

1.2 Cannabinoids

The medicinal properties of the cannabis plant have been recognized for millennia, with its use originating in Central Asia. Also called Marijuana or hemp, it has long been utilized for its purported therapeutic effects. The earliest documented use of cannabis dates back to 2800 BC, when it was included in Emperor Shen Nung's pharmacopoeia. Additionally, cannabis was employed at various points throughout history (Figure 1-3), such by Indian Hindus, Egyptians, Assyrians, Greeks, and Romans[7]. In the 19th century, the medicinal use of cannabis was introduced to Europe.

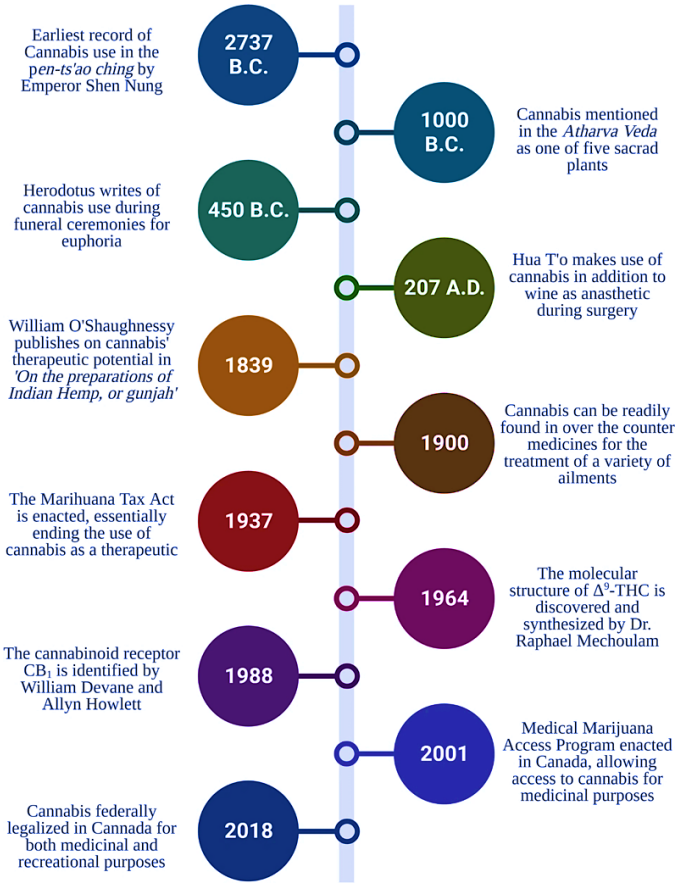


Figure 1-3: ***History of cultural & medical Cannabis.*** *Source:[8]*

In 1940, cannabidiol (CBD) was isolated from cannabis, and in 1964, the structure of tetrahydrocannabinol (THC), the primary psychoactive phytocannabinoid, was identified. To date, over 180 types of cannabinoids have been identified[9], classified into 11 structural families and three types: phytocannabinoids, endocannabinoids and synthetic cannabinoids (SCBs)[10]. Phytocannabinoids are naturally occuring cannabinoids found in the Cannabis plant, with Δ-9-tetrahydrocannabinol, cannabidiol and cannabinol (CBN) being the most prominent. The plant is only capable of synthesizing phytocannabinoids directly in their carboxylated (acidic) form and are predominantly present in this state (THCA, CBDA and CBNA). However, these acids are unstable and are readily converted into their neutral form by heat.

Endocannabinoids are naturally produced by humans and animals. They are able to bind to cannabinoids receptors found almost everywhere in the body (Figure 1-4). They form the endocannabinoid system (ECS). The elucidation of the ECS has also contributed to the growing interest, legalization, and use of cannabis. The two primary endocannabinoids are anandamide (AEA) and 2-arachidonoylglycerol (2-AG). These molecules are able to bind to two main known receptors, CB1 and CB2, part of G protein-coupled receptor family[11]. CB1 receptors are predominantly located in the central nervous system, while CB2 receptors are mainly found in peripheral tissues and the immune system. The activation of these receptors result in regulating many various physiological mechanisms, such as the control of neurotransmitter release and the perception of pain.

The last group are the synthetic cannabinoids, also named “Spice or K2”, are exclusively synthetized and created in labs. Some of them are used to mimic and study the effects of natural cannabinoids. Although, they are becoming a large public health concern due to their unpredictable toxicity and abuse potential, reported in many serious cases such as cardiac ischemia, kidney injury and psychomotor agitation[12].

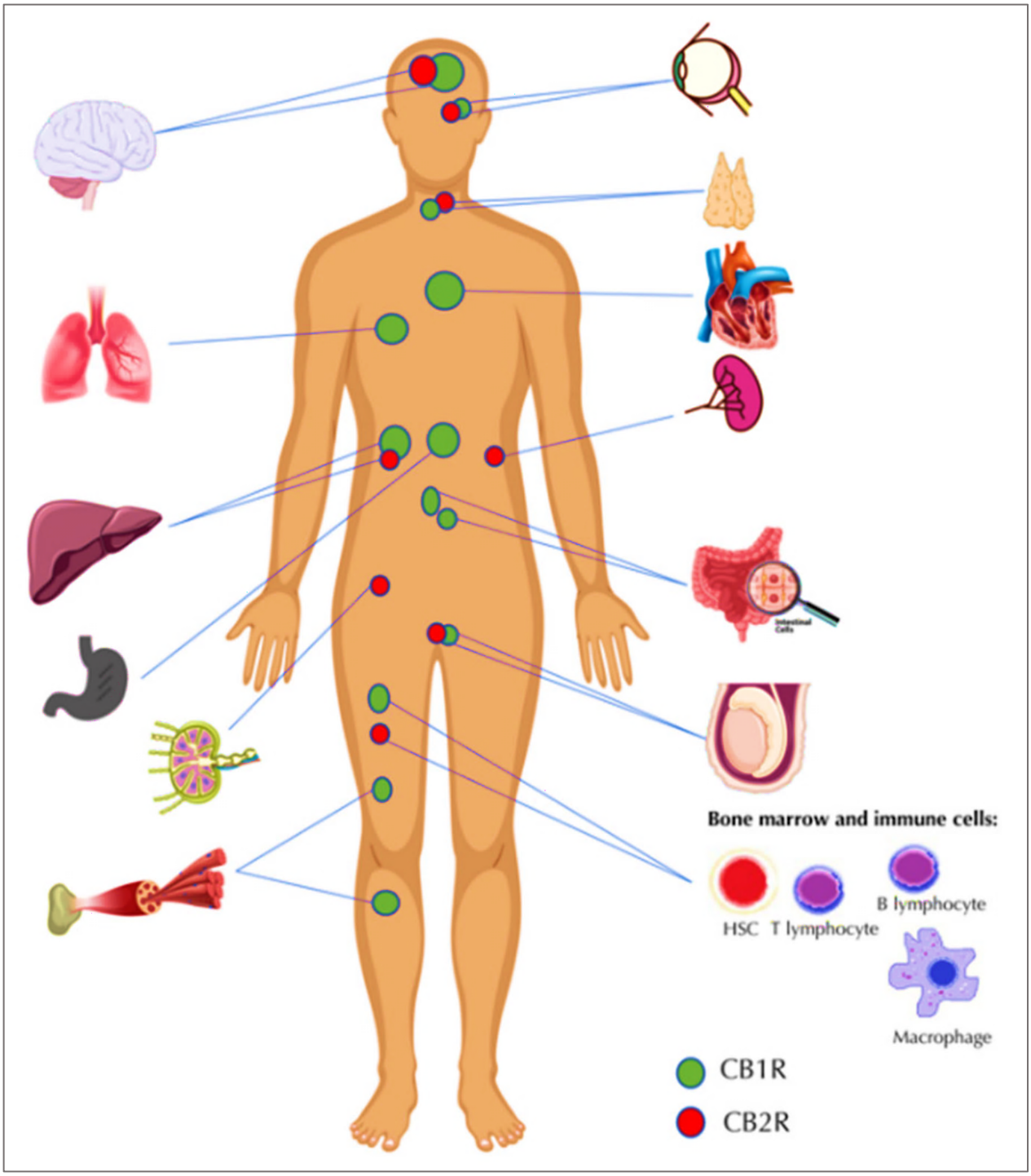


Figure 1-4: ***The distribution of cannabinoid receptors within the human body.*** *CB1R-cannabinoid receptor type 1, CB2R-cannabinoid receptor type 2, HSC-hematopoietic stem cell. Source:[13]*

1.3 Cannabis sativa

The cannabis plant is dicotyledonous annual flowering plant, a member of the Cannabaceae family and a specie of the Cannabis genus. Thanks to genetic analyses, we can distinguish three subspecies, *C. sativa* subsp. *sativa, C. sativa subsp. indica* and *C. sativa* subsp. *ruderalis[14]*. It can reach up to 5-6 meters depending on genetic constitution and environmental factors. The stems are erect, furrowed, and usually branched, with a woody interior, and may be hollow in the internodes[15]. It can be grown in outdoor fields, greenhouses, or indoor environments. Wild *C. sativa* is dioecious in nature, its flowers are male and female in distinct plants and possess 2n=20 chromosomes in which 18 are autosomes and 2 are sex chromosomes such as XX in female and XY in male[16]. Genome size of female and male plant is 1636 and 1683 Mbp respectively, due to a larger size of Y chromosome.

The life cycle of *C. sativa* can be divided in four main stages, germination, seedling, vegetative and flowering respectively. It usually takes 4 to 6 months to complete. Like most plant seeds, cannabis needs the right combination of moisture, oxygen, temperature and light conditions in order to begin germination during the spring. As autumn begins, days begin to shorten, *C. sativa* start to produce flowers, in order to promote pollination and the production of seeds for the next cycle. As the flower reaches the late stages of flower, the trichomes density begins to increase and the plants will become sticky. The pistils, which were initially white, begin to turn into an orange-brown color due to cannabinoids compounds production and accumulation.

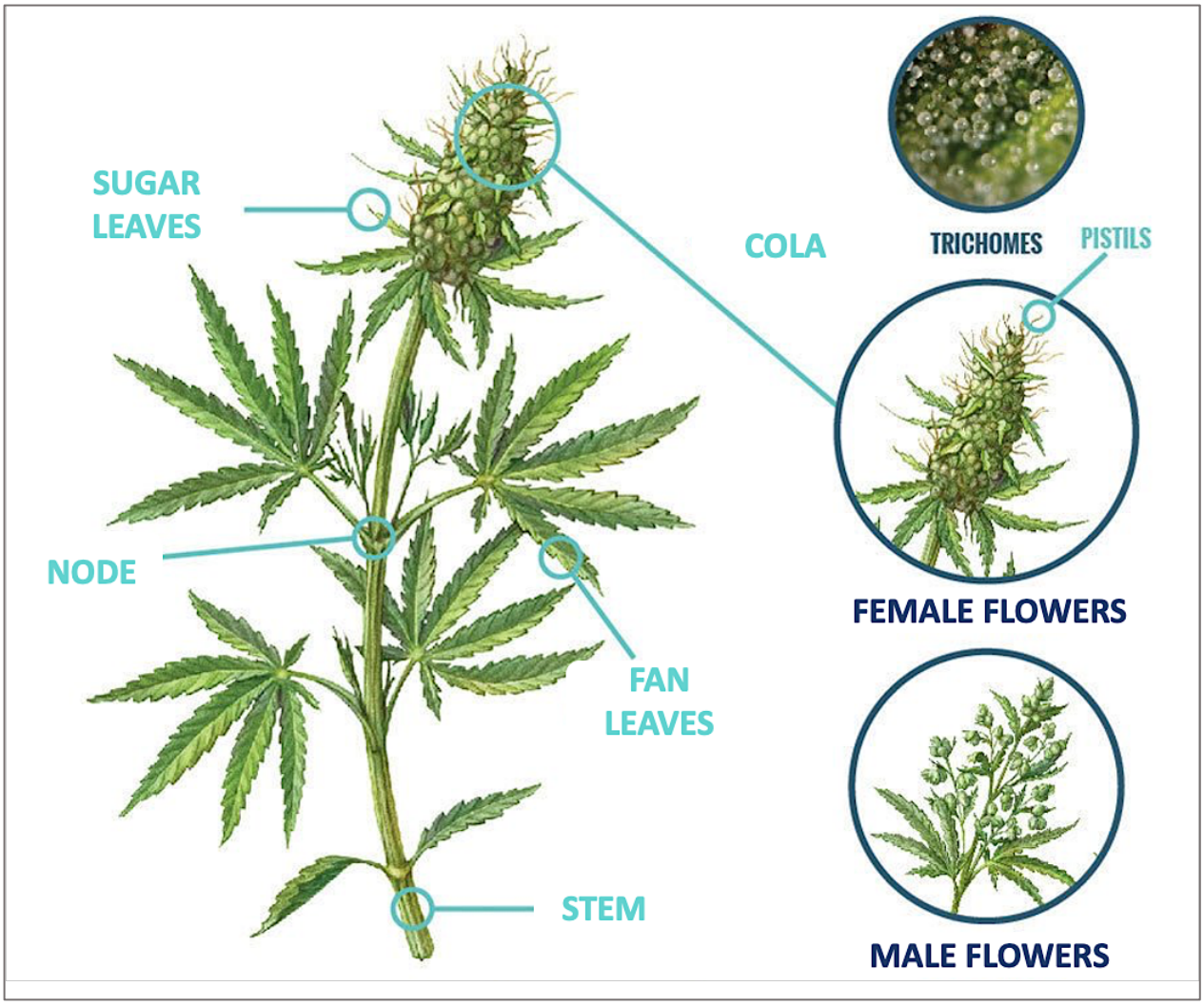


Figure 1-5: ***Cannabis plant representation.*** *Legends translated in English. Source: zamnesia.fr website.*

1.3.1 Cannabinoids biosynthesis pathway in *C. sativa*

THC and CBD, the most well-known cannabinoids, are mainly biosynthesized in glandular trichomes, which are abundant on female flowers (Figure 1-5). These compounds play a crucial role as a chemical weapon against, pathogens, but also play a role in attracting pollinators[17, 18]. At the molecular level, cannabinoids share a common initial pathway (Figure 1-6). The biosynthesis begins with a short-chain fatty acid. Subsequently, two enzymes facilitate the biosynthesis of olivetolic acid (OA): a type III polyketide synthase (PKS or TKS) and an olivetolic acid cyclase (OAC). First, TKS catalyzes the sequential condensation of one hexanoyl-CoA with

three molecules of malonyl-CoA into 3,5,7-trioxododecaneoyl-CoA[19], followed by cyclization into olivetolic acid via OAC[20]. Then, another condensation step occurs through an aromatic prenyltransferase known as cannabigerolic acid synthase (CBGAS) or aromatic prenyltransferase (APT), which combines OA to geranyl pyrophosphate from the chloroplast, specifically the stroma, leading to the formation of the core intermediate, cannabigerolic acid (CBGA)[21]. Ultimately, two different oxydocyclases, tetrahydrocannabinolic acid synthase (THCAS) and cannabidiolic acid synthase (CBDAS) catalyze the conversion of CBGA into the biosynthesis of Δ-9-tetrahydrocannabinolic acid (Δ-9-THCA) and cannabidiolic acid (CBDA)[22, 23]. The neutral form of these cannabinoids results from non-enzymatic decarboxylation.

THCAS and CBDAS enzymes have an estimated molecular mass of approximately 62 kDa and share 84% sequence identity in their amino acids, including the N-terminal secretion signal peptide consisting of 28 amino acids[24, 25]. Furthermore, Structural superimposition of THCAS onto CBDAS highlights the high conservation of the FAD binding region and suggests that FAD reduction serves as a common reaction mechanism for producing cannabinoids[26]. To avoid plant cells toxicity, mitochondrial impermeability and DNA degradation, which ultimately induce apoptosis[27], cannabinoids biosynthesis take place in secretory cells, but proteomic and metabolomic analyses revealed that THCAS and CBDAS are transferred and accumulated to this large sub-cuticular storage cavity surrounding secretory cells ([28]; Figure 1-7).

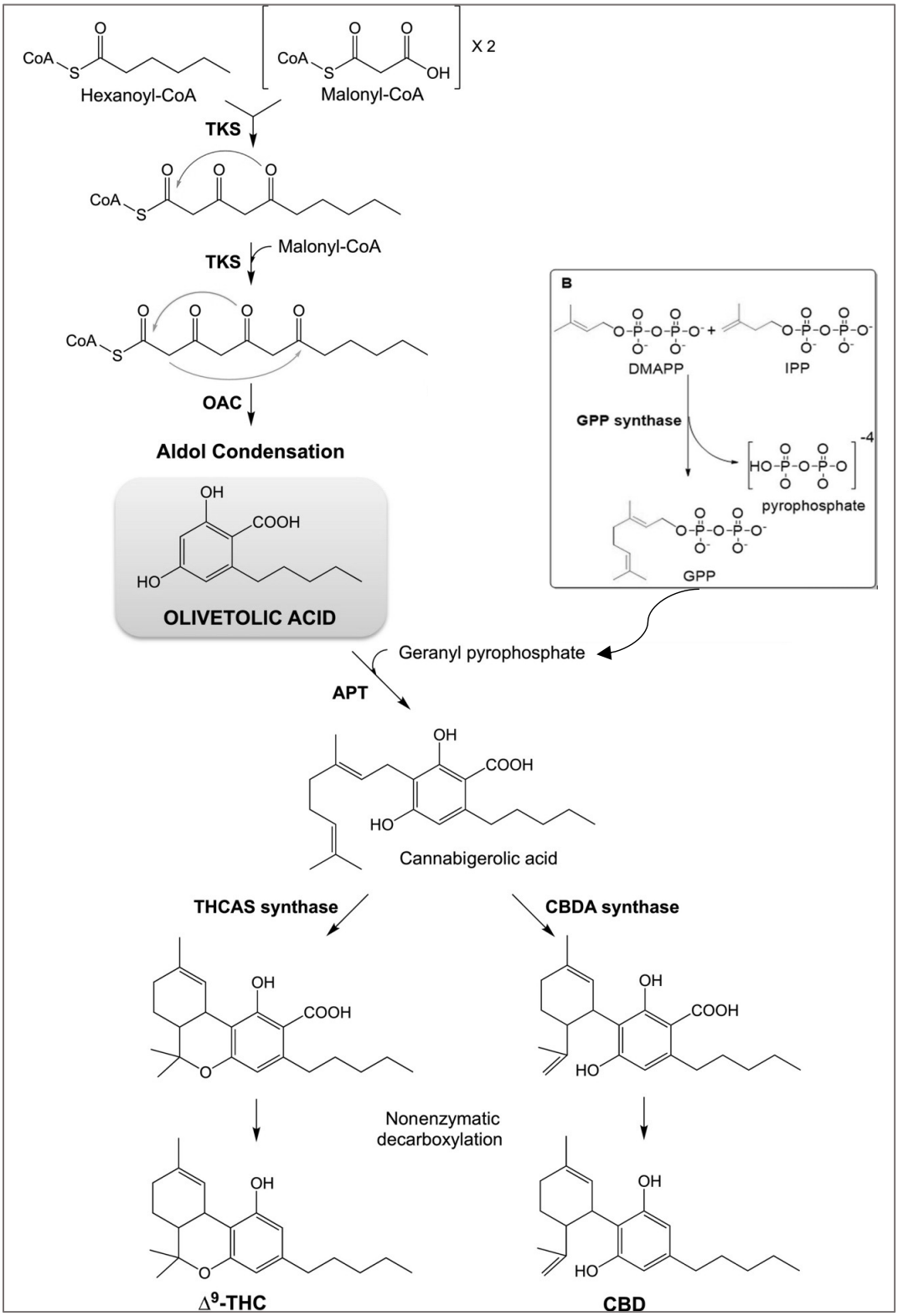


Figure 1-6: ***Biosynthetic pathway for the production of the two main cannabinoids Δ9-THC and CBD in C. sativa.***

The enzymes in the pathway are TKS, OAC, APT, THCAS and CBDAS, respectively. B. Synthesis of geranyl pyrophosphate (GPP) from MEP/MVA pathways. Dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP) catalyzed by geranyl pyrophosphate synthase. Modified figure. Source:[29, 30]

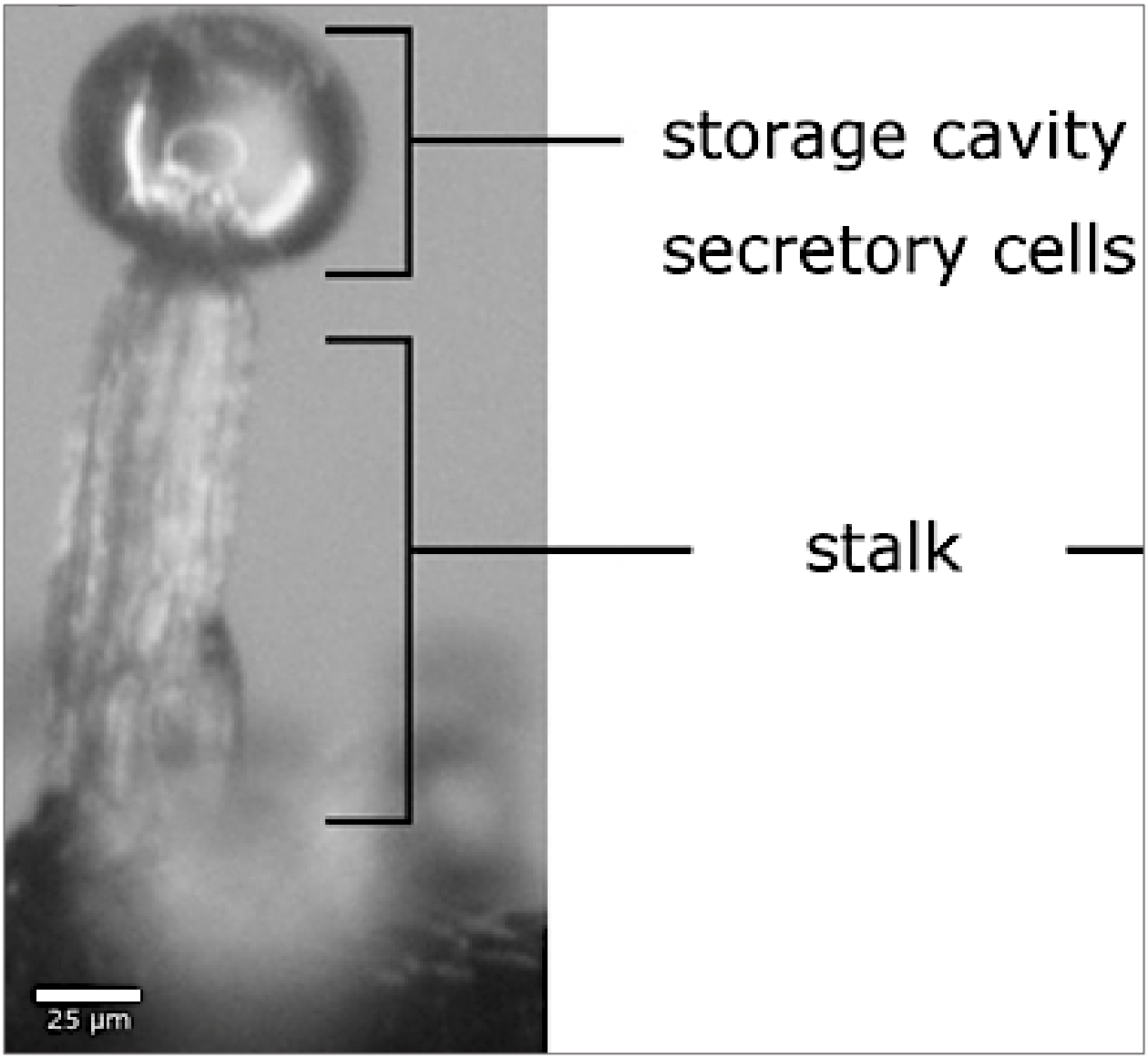


Figure 1-7: ***Glandular trichome at 7th week of the flowering stage from the drug-type Cannabis (Euphoria) strain.*** *Modified figure, Source: [28]*

1.4 Clinical properties of CBD

In recent decades, numerous studies have been conducted to assess the therapeutic potential of cannabinoids. These studies have provided information on the effectiveness of cannabinoids in various indications, some of which are still undergoing validation. These studies have led to significant interest in their therapeutic potential for a variety of medical conditions, such in neurodegenerative and psychiatric disorders[31], inflammation and pain[32], cardiovascular diseases[33], diabetes[34], and other (Figure 1-8).

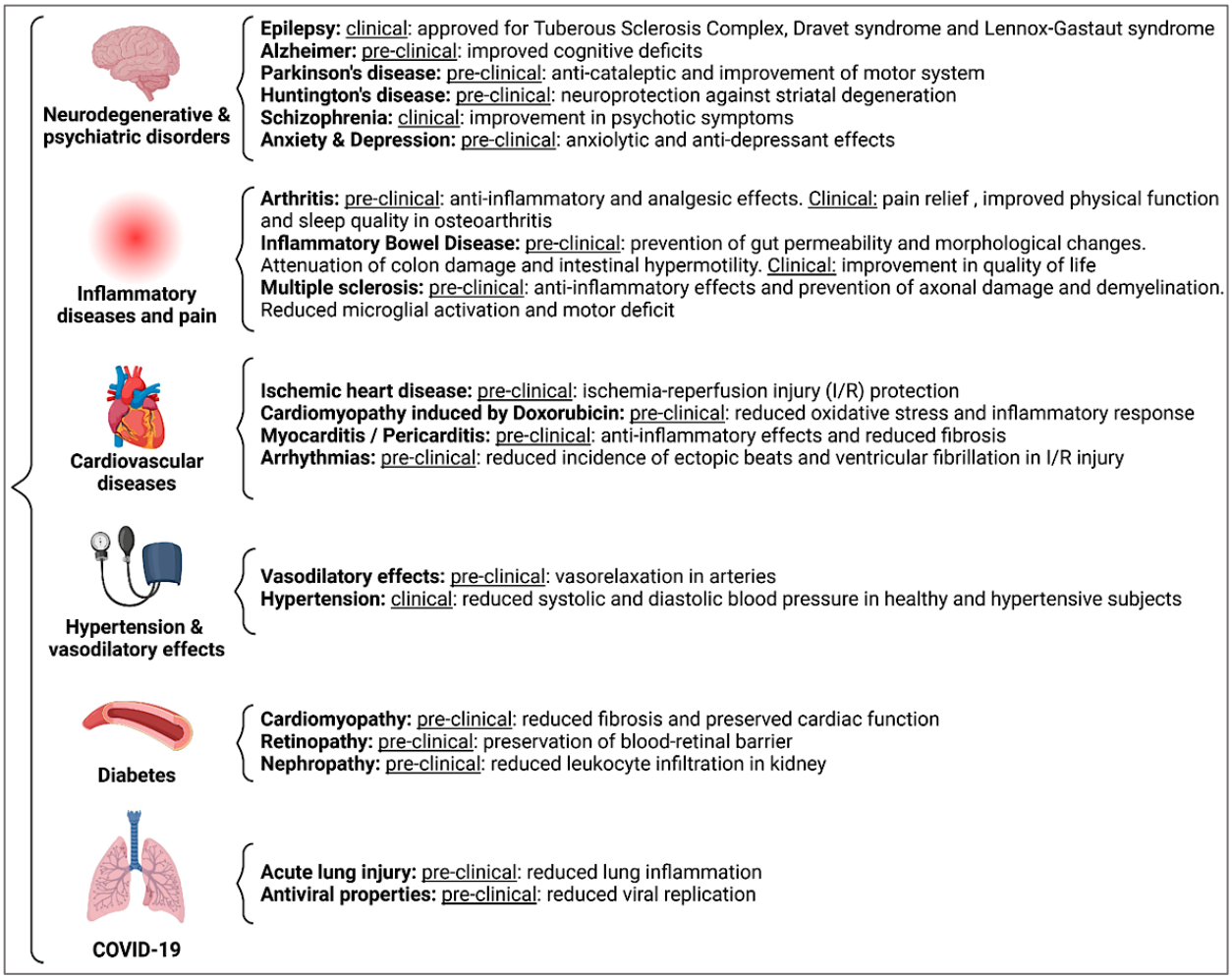
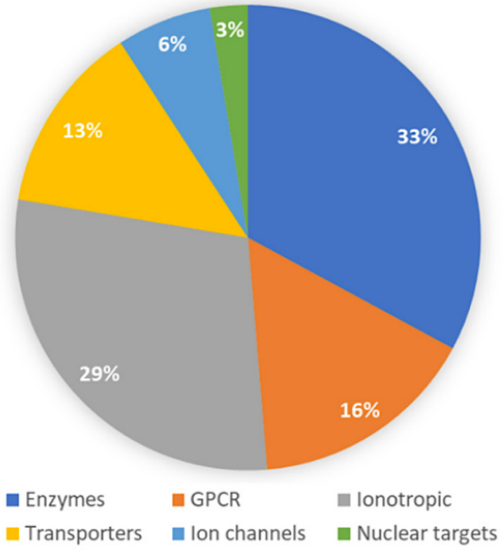


Figure 1-8: ***Summary of the effects of CBD in different organ systems and diseases.***

Such as neurodegenerative conditions, pain management, cerebral ischemia-reperfusion injury, heart diseases, vascular diseases, diabetes, and hypertension. Modified figure, Source: [35]

Cannabinol is a small lipophilic molecule that was initially discovered in 1940. It exert its biological effects primarily on endocannabinoid system through cannabinoids receptors[36]. The interaction between CBD and these receptors influences a wide range of physiological functions, such as anticonvulsant, antipsychotic, anxiolytic and neuroprotective properties[37]. Indeed, recent studies have shown CBD directly interacting with various receptors, ion channels, enzymes and transporter (Figure 1-9). The mechanism of action on G protein-coupled receptors (GPCRs) result on signal transduction, with a downstream activation of several pathways.



	Number of targets	% of all targets
total	76	
Enzymes	25	33
GPCR	12	16
Ionotropic	22	29
Transporters	10	13
Ion channels	5	7
Nuclear targets	2	3
Total GPCR types	12	
Gq	2	17
Gi	7	58
Gs/i	3	25
Ion permeability	27	
cationic	14	52
anionic	13	48
Total interactions	39	
stimulating	10	26
inhibitory	29	74

Figure 1-9: ***Pie chart showing the percentage of specific protein targets of CBD from a pool of all protein targets.***

Table presenting the number of identified CBD targets and percentage of specific Gα proteins coupled to GPCR receptors and ion permeability of both ion channels and ionotropic receptors. Source: [38]

1.5 Commercial and industrial interest of cannabinoids

Cannabinoids have garnered considerable attention across commercial and industrial sectors due to their wide range of biological effects and potential uses. As cannabis legalization progresses, especially in areas where it is permitted for medical and recreational purposes, interest in cannabinoids has surged (Figure 1-10). This growing demand spans various industries, including pharmaceuticals, agriculture, food, and cosmetics.

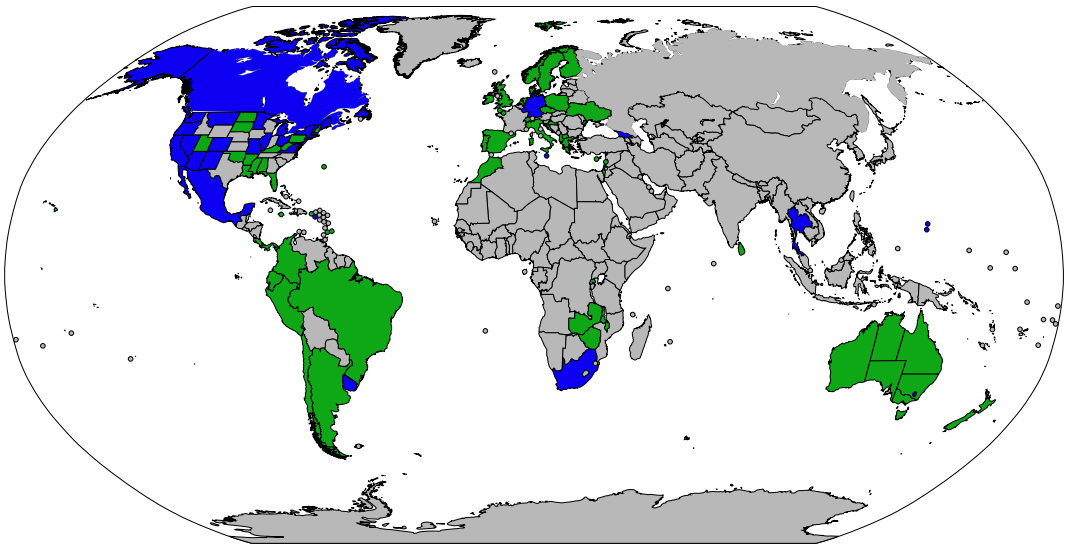


Figure 1-10: ***Map showing actual adult legal status of medical cannabis across the world.*** *Blue: legal for recreational and medical uses, green: legal only for certain medical uses. Source : Wikipedia*

The industrial hemp industry is increasingly focusing on optimizing cannabinoid extraction methods, genetic selection for higher yields, and improving crop management for sustainable farming practices. Only three years after the legalization of recreational cannabis, Canada's cannabis industry is contributing billions to the country's economy and government revenue. Between 2018 and 2021, total sales from both recreational and medical cannabis are estimated at $11 billion, with two-thirds of that amount coming from recreational cannabis (Figure 1-11, upper-panel). Considering direct, indirect, and induced economic activity, cannabis industry has approximately contributed $43.5 billion to Canada's national gross domestic product (GDP) from legalization through 2021. Of this, about $25.2 billion has been in the form of labour income, supporting roughly 98,000 jobs each year (Figure 1-11, middle-panel). Moreover, since 2018, the cannabis sector has generated substantial government tax revenue, estimated to $15.1 billion, especially through indirect taxes, amounting to $7.3 billion (Figure 1-11, lower-panel).

Industry commonly included CBD in skincare products, creams, lotions, shampoos, and even cosmetics like makeup and moisturizers. The growing trend in natural wellness products has driven many companies to innovate in formulating cannabinoid-infused products aimed at enhancing skin health, relaxation, and overall well-being. Today, CBD-infused foods, drinks, gummies, chocolates, and snacks are becoming widely available, capitalizing on the growing consumer interest in functional foods.

Despite the growing commercial and industrial interest, the cannabinoid industry faces several challenges like other forms of industrialized agriculture, particularly related to highly energy-intensive and contributes significantly to carbon and greenhouse gas (GHG) emissions. Indoor cultivation, in particular, requires substantial infrastructure to maintain stable artificial growing conditions, including high-intensity lighting and supplemental carbon dioxide (CO_2) to enhance plant growth (Figure 1-12). Over its life cycle, the production of one kilogram of dried cannabis flower can generate up to 5,184 kg of CO_2. By comparison, beef production emits approximately 99.48 kg of CO_2 per kilogram of food product[39]. To minimize the environmental impact, carbon emissions, and waste in the cannabis sector, the industry can address this challenge by adopting an eco-friendly and sustainable solutions.

Hence, some companies are investigating ways to optimize cannabinoid production in engineered microorganisms, such as yeast or bacteria, as a more sustainable and scalable alternative to traditional plant-based extraction methods. This could lead to more cost-effective and environmentally friendly production processes for cannabinoids.

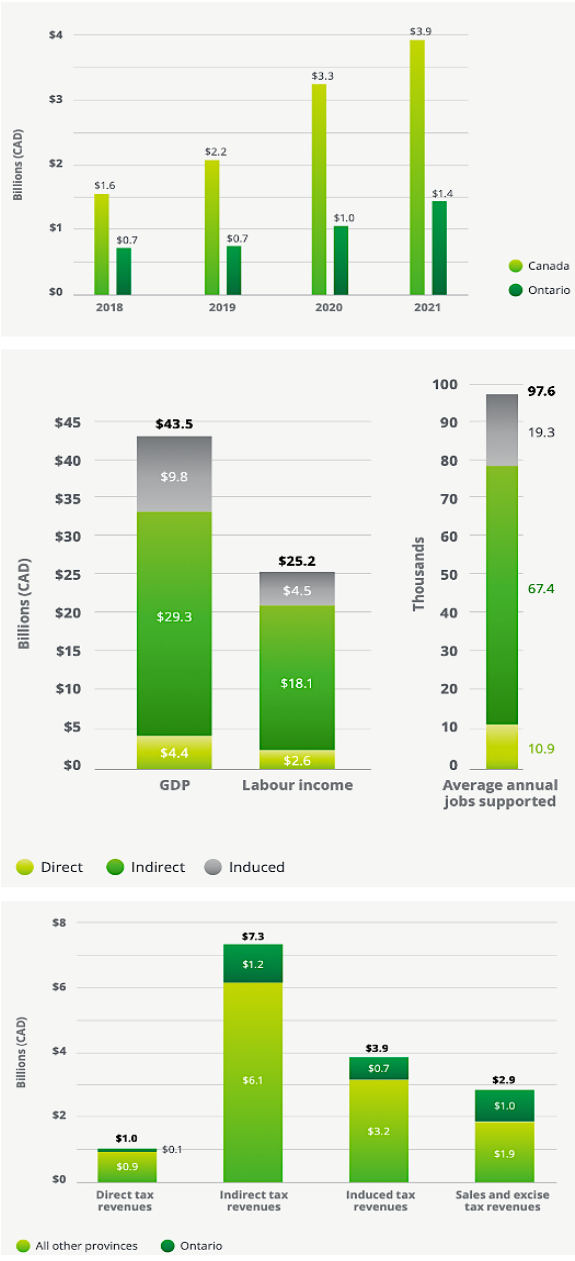


Figure 1-11: ***The economic impact of Canada's cannabis sector between 2018 and 2021.***

Upper-panel: cannabis revenues have been growing in Canada and total sales across the country is estimated at $11 billion , middle-panel: cannabis sector economic contributions, indirect impacts are larger than direct impacts due to significant capital expenditures on construction and retrofitting-related activities , lower-panel: $15.1 billion in government tax revenues, direct and indirect tax revenue refers to direct and indirect economic contributions to government revenues,. Source: Statistic Canada; Deloitte Analysis.

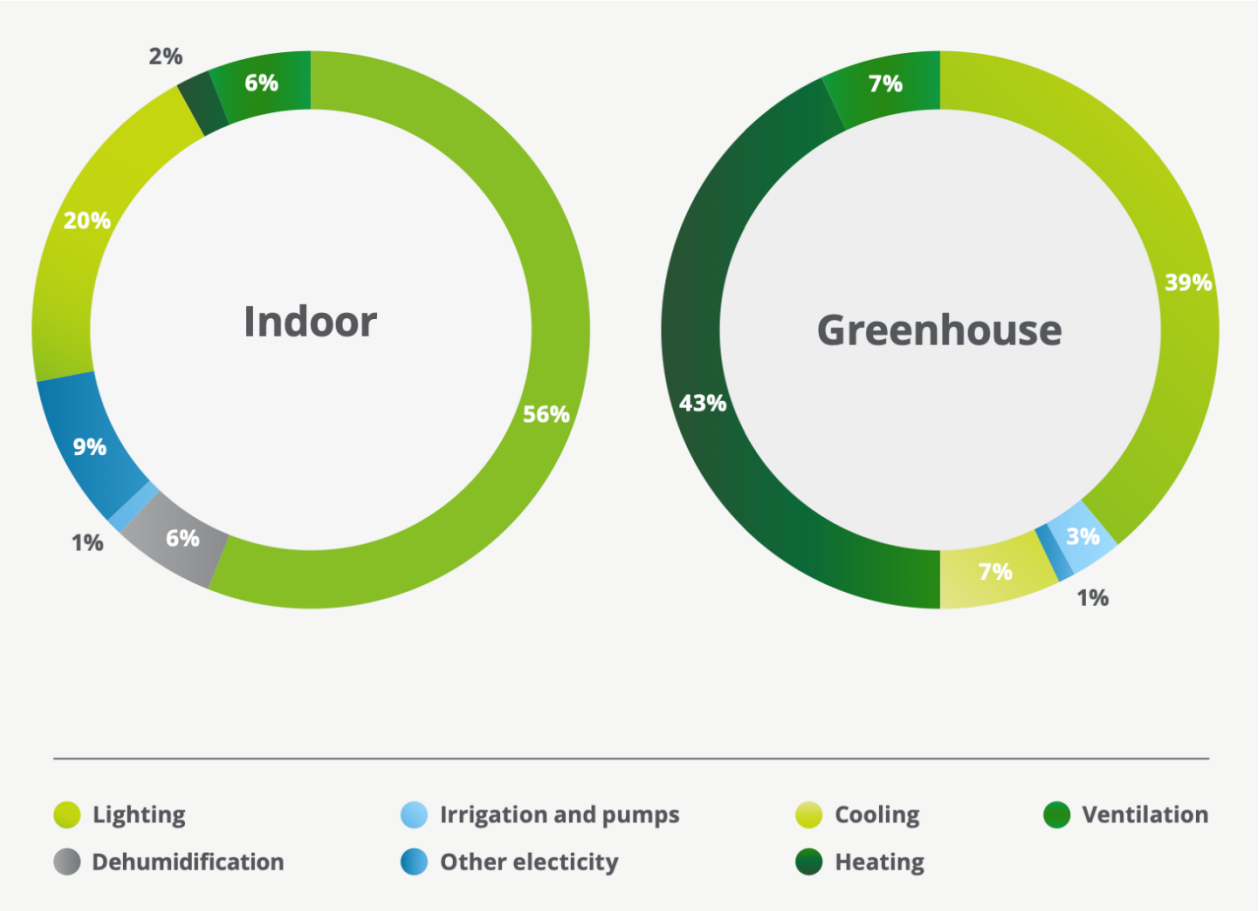


Figure 1-12: ***Indoor and greenhouse energy consumption in Ontario.***

Indoor facilities exceeds greenhouse-based production sites, consuming 30% to 60% more energy than indoor facilities. Source: Energy management best practices for cannabis greenhouses and warehouses.

1.6 Diatoms

Diatoms are photosynthetic eukaryotic unicellular microorganisms, able to convert sunlight into energy through the process of photosynthesis. Because of this, they play a crucial role in marine ecosystems and the global carbon cycle, exhibiting a rich diversity. They seem to have emerged from a secondary endosymbiosis event that led to the acquisition of photosynthesis. In this process, a heterotrophic host engulfed a red alga, which had originally obtained photosynthesis by ingesting a cyanobacterium (Figure 1-13). They belong to the class *Bacillariophyceae* and are the most dominant phytoplankton group, with over 200,000 species showing variability in shape and size[40, 41]. Ranging from just a few micrometers in the smallest species to more than 2 mm in the largest ones. The word 'diatom' comes from the Greek *diatomos*, meaning 'cut in two,'

referring to the two siliceous valves found in this superclass. Antoni Van Leeuwenhoek, a Dutch merchant and scientist renowned for enhancing the microscope by developing advanced lenses, was the first to observe diatoms in 1702. Unlike other phytoplankton communities, diatoms possess a silica cell wall known as a frustule. These frustules feature various geometric patterns that make diatoms visually striking under a microscope (Figure 1-14). This innate ability to absorb silicon from the environment, compensating the low concentrations, has made them an intriguing community. Indeed, silica transporters, known as SITs, have been identified on their membrane and appear to be unique to diatoms[42]. They are a key in many ways, starting with an ecological importance as they are the most important primary producers in aquatic ecosystems, providing energy for a wide range of organisms, from tiny zooplankton to larger fish. Indeed, they are used as a bioindicator in water samples and monitoring water quality.

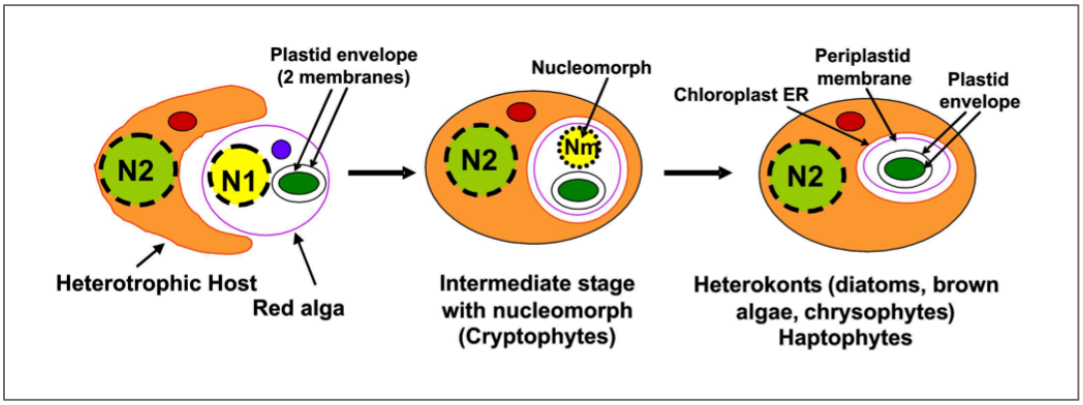


Figure 1-13: ***Origin of chloroplasts by secondary endosymbiosis involving a red algal endosymbiont.***

Following this event, the endosymbiont's nucleus (N1) disappeared after transferring many of its genes to the host nucleus (N2). However, in cryptophyte algae, a remnant nucleus, known as a nucleomorph (Nm), persists between the plastid and two additional membranes—one derived from the red algal plasma membrane (periplastid membrane) and the other from the host's endomembrane system (chloroplast ER). Ribosomes attached to the outer surface of the chloroplast ER, as well as its continuity with the broader ER system, are not depicted. Nuclear-encoded plastid proteins are synthesized on these ribosomes and must pass through four membranes to reach their destination. Additionally, the red algal mitochondrion (small blue circle) was lost during this process. Source: [43]

Regarding reproduction, diatoms can reproduce both sexually and asexually. While *Phaeodactylum tricornutum* is traditionally regarded as a predominantly asexual organism, the identification of functional meiosis-related genes (SPO11) in its genome suggests the potential for sexual reproduction; however, this remains unconfirmed due to the lack of direct observational evidence[44]. When a cell reaches a critical size, a specific cellular mechanism is triggered to

initiate sexual reproduction, restoring its original size[45]. Sexual reproduction varies between pennate and centric diatoms. In centric diatoms, it follows an oogamous process, where a male gametangial cell undergoes a series of specialized divisions to generate a limited number of microspores. These microspores then undergo meiosis, giving rise to flagellated sperm cells. Meanwhile, female gametangial cells produce one or two 'eggs,' which, upon fertilization by a sperm cell, form a zygote that expands by absorbing water. In pennate diatoms, sexual reproduction occurs through either isogamy or anisogamy. Gametangial cells pair up and initiate gametogenesis, with each parental cell producing two gametes. The fusion of an active and a passive gamete results in a zygote that, similar to centric diatoms, absorbs water to enlarge and develop into a new cell, thereby restoring the original cell size. Beyond size, it is also essential for the environmental conditions to be met. In addition, besides their various morphological and biochemical adaptations, diatoms can also transition into a vegetative state by forming dormant spores when environmental conditions become unfavorable for growth. These spores can settle on the ocean floor, where they may persist for several years.

Finally and above all, diatoms offer significant potential for various scientific and industrial applications[46].

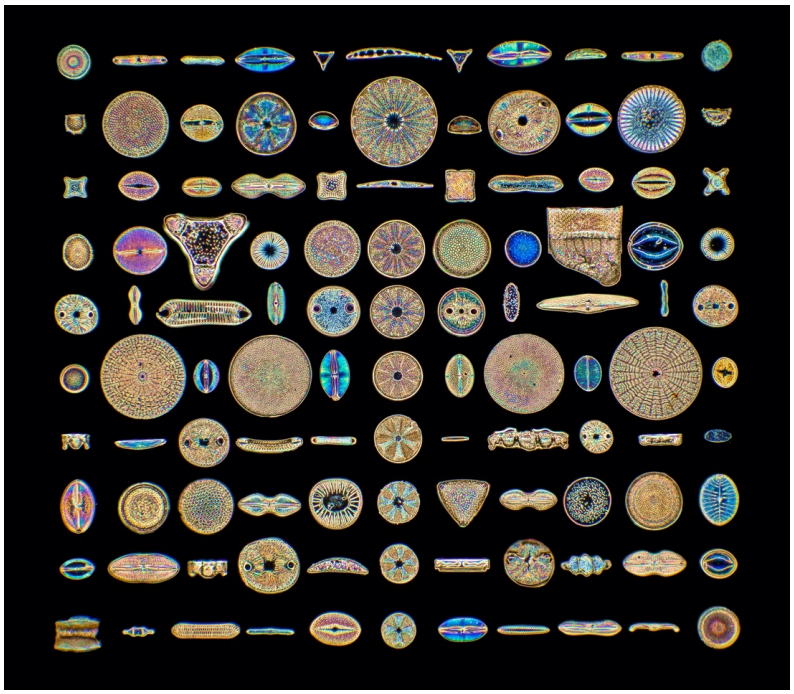


Figure 1-14: ***Microscopic artwork.*** *Source: Scenics and Science / Alamy Stock Photo*

1.6.2 Phaeodactylum tricornutum

The diatom *Phaeodactylum tricornutum (P. tricornutum*) is an unicellular mixotrophic organism capable of feeding both autotrophically, through photosynthesis, and heterotrophically by consuming pre-existing organic compounds in its environment. Unlike other diatoms, the presence of silica in the cell wall of *P. tricornutum* is not necessary for its growth. The absence of a rigid frustule surrounding the cell allows *P. tricornutum* to modify its cell morphology and adapt to environmental conditions[47]. It can adopt different morphotypes (fusiform, oval, triradiate, or round). In a benthic environment, with low agitation, salinity, and temperature, oval-shaped cells are overrepresented. When these conditions persist, the oval cells transform into round cells. This round form has the ability to form biofilms, allowing survival for several months. When conditions become optimal again, for example in a planktonic environment, the cells return to a fusiform shape. Finally, triradiate-shaped cells are produced when optimal conditions are maintained (Figure 1-15).

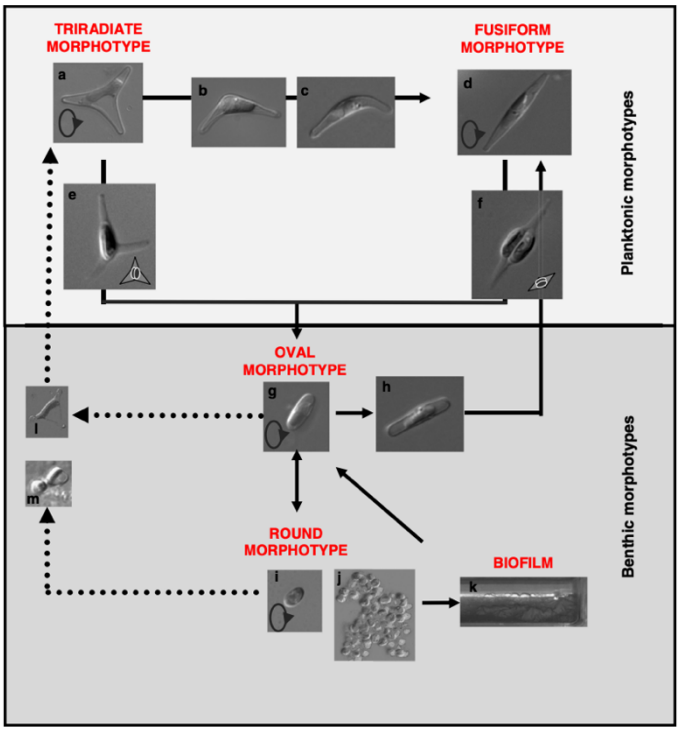


Figure 1-15: ***Diagram illustrating the various cellular morphologies of P. tricornutum.***

Triradiate (a) and fusiform (d) cells are linked to the planktonic morphotype, whereas oval (g) and round (i) cells are classified as benthic morphotypes. Oval and round cells can aggregate (j) and ultimately develop into biofilms (k). Arrows indicate the pathways of morphological transitions between cell types. Photos b, c, e, f, h, l, and m depict intermediate forms. Source: [47]

P. tricornutum was first described in 1897 by Bohlin, K. It is predominantly found in marine environments and is able to survive in nutrient-poor conditions. It was the first diatom to be transformed in 1996[48], and has been studied for its potential in lipid metabolism, in biofuel production and in producing valuable compounds such as omega-3 fatty acids. Long before the availability of its genomic sequences in 2008[49], it became established as a key model organism in diatom biological studies due to its ability to grow under simple culture conditions, low light, and high tolerance to elevated pH[50, 51]. Furthermore, genetic engineering allows researchers to investigate its potential and this has led to the development of a multitude of molecular tools (Figure 1-16).

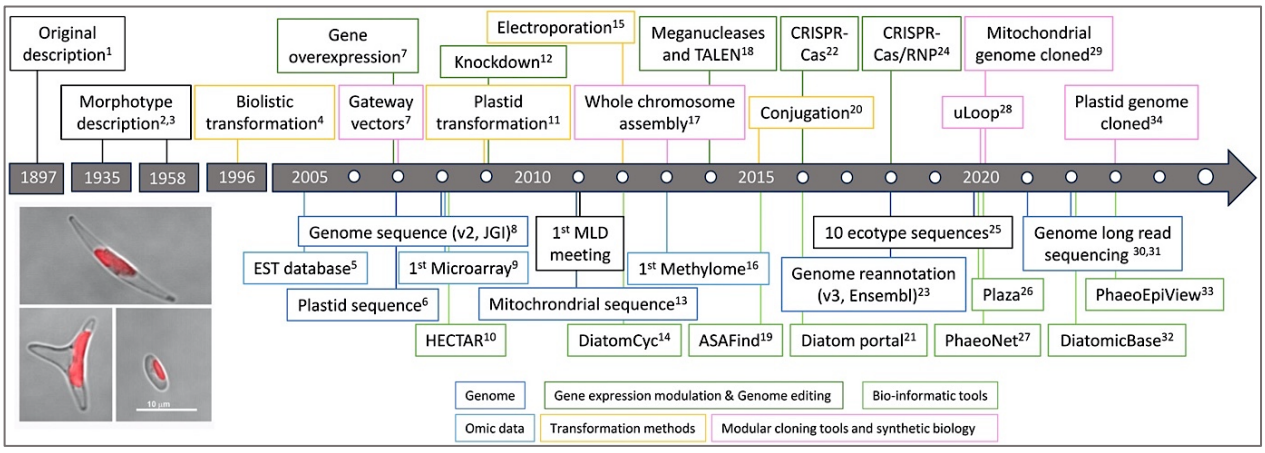


Figure 1-16: ***Milestones in the P. tricornutum molecular research.***

The figure also shows confocal microscopy images of P. tricornutum morphotypes (fusiform, triradiate, and oval), taken merging the bright-field and chlorophyll autofluorescence channels. In red, P. tricornutum plastid. Source: [52]

1.6.3 Synthetic biology, a sustainable alternative

Over the past twenty years, the field of synthetic biology has experienced significant growth, particularly with the establishment of a connection between genetics, epigenetics, and the environment. Synthetic biology presents a promising solution to the challenges posed by the limited availability of natural products from native plants. This approach combines metabolism engineering with molecular biology to create modified living things, like yeast, bacteria and microalgae, that can perform certain tasks. Providing a chance to solve complicated problems of sustainability, and to produce a diverse range of desired natural compounds in a cost- and time-

effective manner. After transformation, theses microorganisms turn into cellular factories for botanical therapeutics. Microalgae constitute a taxonomically diverse group of organisms distributed throughout the eukaryotic tree (Figure 1-17), and have the biochemical machinery to synthesize high-value phytochemicals.

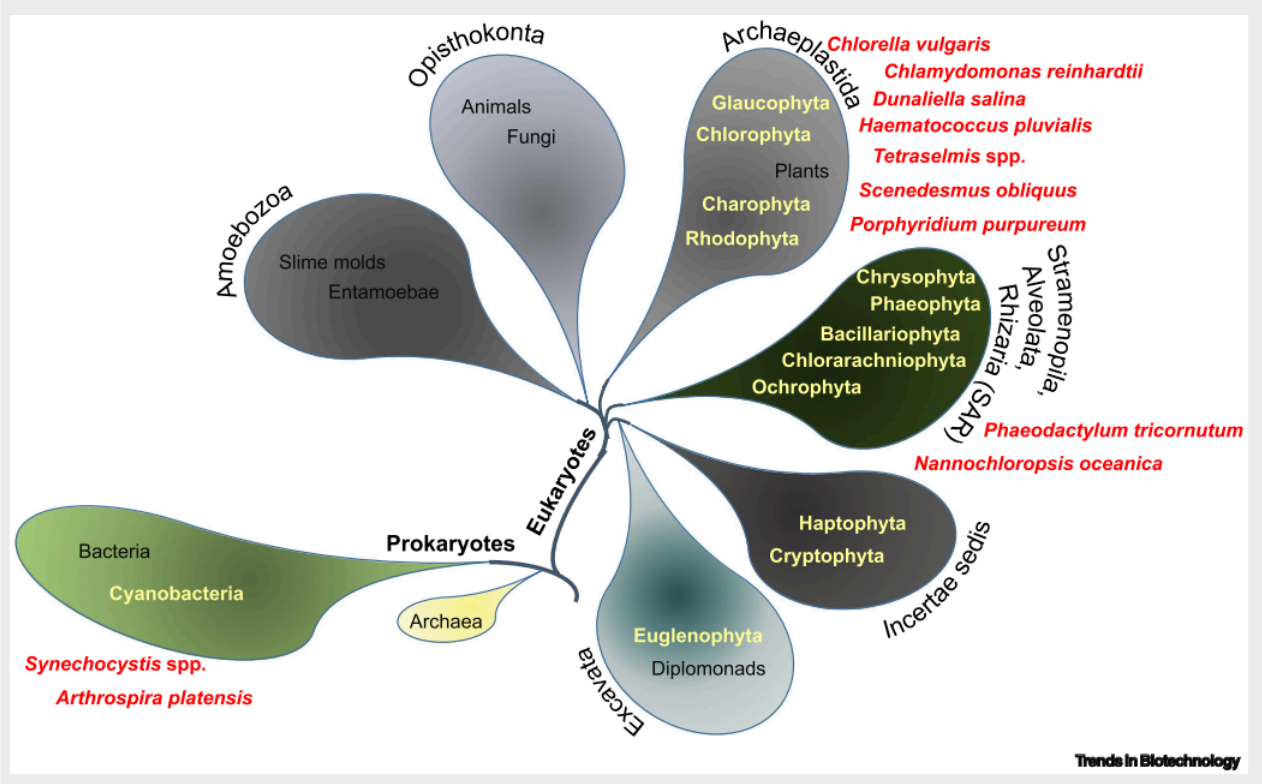


Figure 1-17: ***A Cartoon Representation of the Phylogenetic Diversity of Microalgae Phyla (in Yellow) and Representative Species under Development (in Red).*** *Other representative lineages such as animals and plants are also indicated (in black). Source: [53]*

Unlike plant cells, microalgae exhibit rapid growth, with doubling times typically measured in hours and biomass production occurring within days rather than months or years. They can be cultivated on non-arable land and have the potential to reclaim nutrients from waste resources, such as wastewater and waste CO_2 sources. While various microalgae are being developed as cell factories at different stages of progress, the growing molecular toolkit (Figure 1-18), combined with computational methods, positions the pennate diatom *P. tricornutum* as a highly promising algal system for metabolic engineering[46]. It is known to accumulate a spectrum of marketable products, and is a commercially viable species for large-scale cultivation (Figure 1-19).

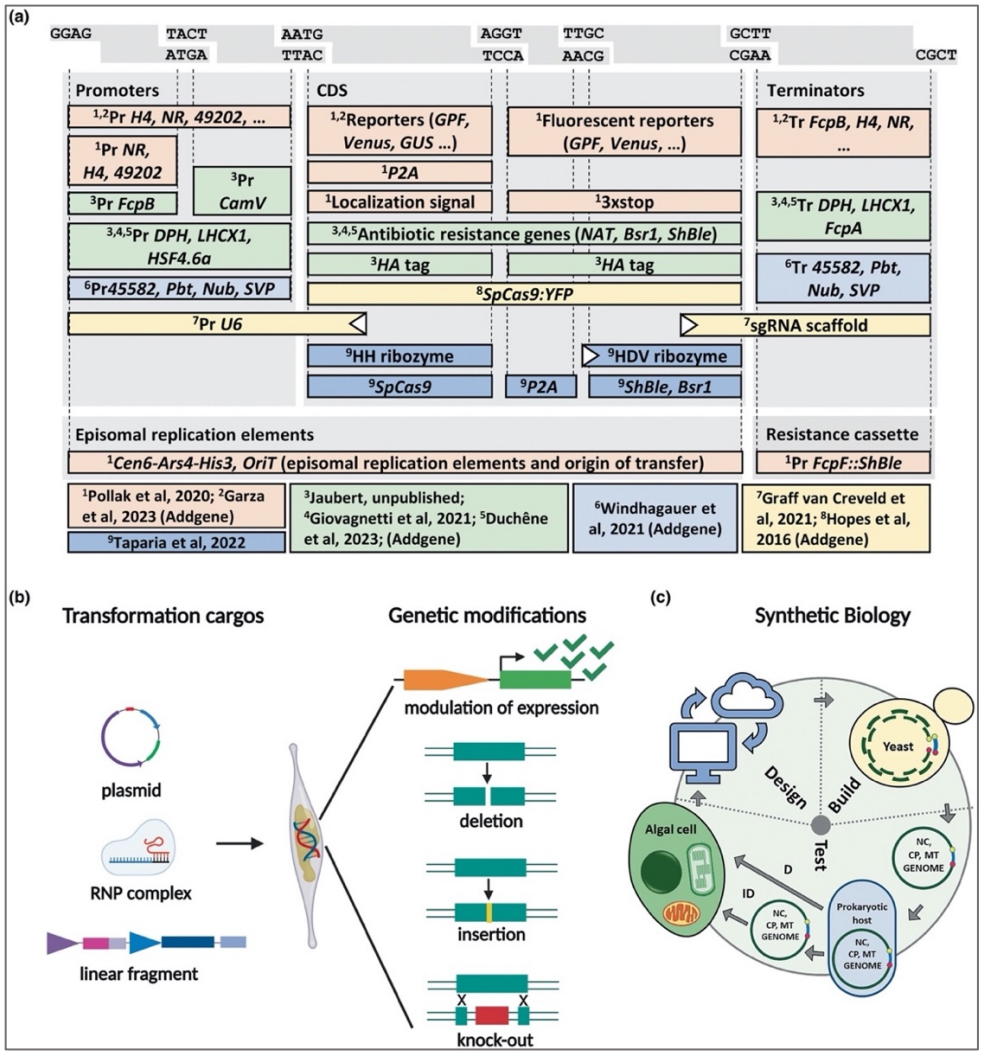


*Figure 1-18: **Tools for genetic engineering and synthetic biology in P. tricornutum.***

(a) Overview of the P. tricornutum basic parts. (b) Graphic representation of the strategies used to obtain engineered strains by targeted genetic modification. (c) The design–build–test cycle for engineering P. tricornutum genomes. Designed synthetic genomes: nuclear (NC), mitochondrial (MT), and chloroplast (CP) will be built in yeast and then transferred to a prokaryotic host to be delivered directly (D), for example, via conjugation, or indirectly (ID), for example, via electroporation, to the appropriate cell compartment to test for designed functions. Source: [52]

Furthermore, *P. tricornutum* have been cultivated in various systems, indoor and outdoor cultivation methods, using photobioreactors (PBRs), including column, tubular and flat-panel designs (Figure 1-20). As well as open raceway ponds, yielding varying levels of biomass and product accumulation. However, challenges such as high energy demands for temperature regulation and biomass loss due to photorespiration remain key obstacles for PBR-based cultivation[54, 55].

Product class	Product	Strain	Product yield/ productivity reported[a]	Functional end-use	Operational conditions
Lipids	EPA	UTEX 640	3% DW; 56 mg/l/d	Nutraceutical; cardiovascular health; precursor for prostaglandin-3, thromboxane-3, and leukotriene-5 eicosanoids	Outdoor chemostat split-cylinder airlift PBR (50–60 l), Almeria, Spain; photoautotrophic/mixotrophic with glycerol
	DHA	CCAP 1055/1 (R) Pt_EI05	0.64% DW	Nutraceutical; primary structural component of the human brain, cerebral cortex, skin, and retina	Indoor horizontal fence PBR (550 l); photoautotrophic
		CCAP 1055/1 (R) Pt_EI05	0.26% DW		Indoor raceway (1250 l); photoautotrophic
		CCAP 1055/1 (R) Pt_MCAT_PtD5b	0.92% DW		Indoor flasks; photoautotrophic
	ARA	CCAP 1055/1 (R) Pt_MCAT_PtD5b	1.89% DW	Nutraceutical; prostaglandin precursor	Indoor flasks; photoautotrophic
	TAG	UTEX 640	58.5 mg/l/day, 45% DW	Biodiesel	Green wall panel III (≤40 l); outdoor; photoautotrophic
	Brassicosterol	CCAP 1055/1 (R) LjLUS-25		Decreased risk of coronary heart disease; anti-inflammatory activities	Lab PBR (≤1 l); photoautotrophic
Carbohydrates	Chrysolaminarin	CAS	14% DW; 94 mg/l/day	Antioxidant	Indoor flat-plate PBR (50 l); photoautotrophic
Terpenoids	Fucoxanthin	CAS	0.7% DW; 4.7 mg/l/day	Antioxidant, anti-obesity, anticancer, anti-inflammatory	Indoor flat-plate PBR (50 l); photoautotrophic
	Lupeol	CCAP 1055/1 (R) LjLUS-25	0.01% DW	Antiprotozoal, antimicrobial, antitumor, chemopreventative	Lab Algem PBR (≤1 l); photoautotrophic
		CCAP 1055/1 (R) AtLUS-6	0.0013% DW		Indoor horizontal fence PBR (550 l); photoautotrophic
	Betulin	CCAP 1055/1 (R) LjLUS-25	Detectable levels	Antitumor	Lab Algem PBR (≤1 l); photoautotrophic
Heterologous compounds and proteins	Polyhydroxybutyrate (PHB)	CCAP 1055/1 (R)	10.6% DW	Bioplastics	Indoor flasks (≤1 l); photoautotrophic
	Human IgGαHBsAg: antibody against hepatitis B virus surface protein	UTEX 646 (R)	0.0021% DW; 8.7% TSP	Monoclonal antibody	Indoor flasks (≤1 l); photoautotrophic
	IgG1/kappa Ab CL4mAb: antibody to hepatitis B virus surface protein lacking the ER retention signal (DDEL) at the C-termini of both antibody chains	UTEX 646 (R)	2.5 mg/l (secreted)	Monoclonal antibody	Indoor flasks (≤1L); photoautotrophic
	Monoclonal IgG antibodies against the nucleoprotein of Marburg virus (close relative of Ebola virus)	UTEX 646 (R)	2 mg/l (secreted)	Monoclonal antibody	Indoor flasks (≤1L); photoautotrophic
Whole cell	Biomass	UTEX 640	25.4 g/l, 1.7 g/l/day	Aquaculture/animal feed	Outdoor split-cylinder airlift PBR (60 l); mixotrophic (0.1 M glycerol)

*Figure 1-19: **Products derived from P. tricornutum cultivation.** **Abbreviations:** PBR, photobioreactor; R, recombinant strain; TSP, total soluble protein. Source: [53]*

Additionally, in 2015, a system for extrachromosomal expression was developed for *P. tricornutum*, enabling a plasmid to function as an episome[56]. In eukaryotic cells, an episome is a circular DNA molecule that replicates independently within the host without integrating into the genome. This allows for the stable propagation of heterologous DNA in *P. tricornutum*, avoiding position effect issues associated with genomic integration. The episome was created by inserting segments of *P. tricornutum* chromosomes into the plasmid p0521s. Upon transforming the diatom with this plasmid containing the *Sh ble* resistance gene, zeocin-resistant colonies were successfully obtained. Further investigations identified the specific sequences responsible for episomal maintenance, highlighting the crucial role of the CEN6-ARSH4-HIS3 sequence. Analysis of CEN6-ARSH4-HIS3 revealed that its ability to function as a centromere in *P. tricornutum* was associated with its low guanine and cytosine (GC) content[57]. Building upon these advancements in episomal expression, our efforts focused on optimizing secretion pathways and regulatory elements to enhance the functional expression of complex heterologous proteins in *P. tricornutum*

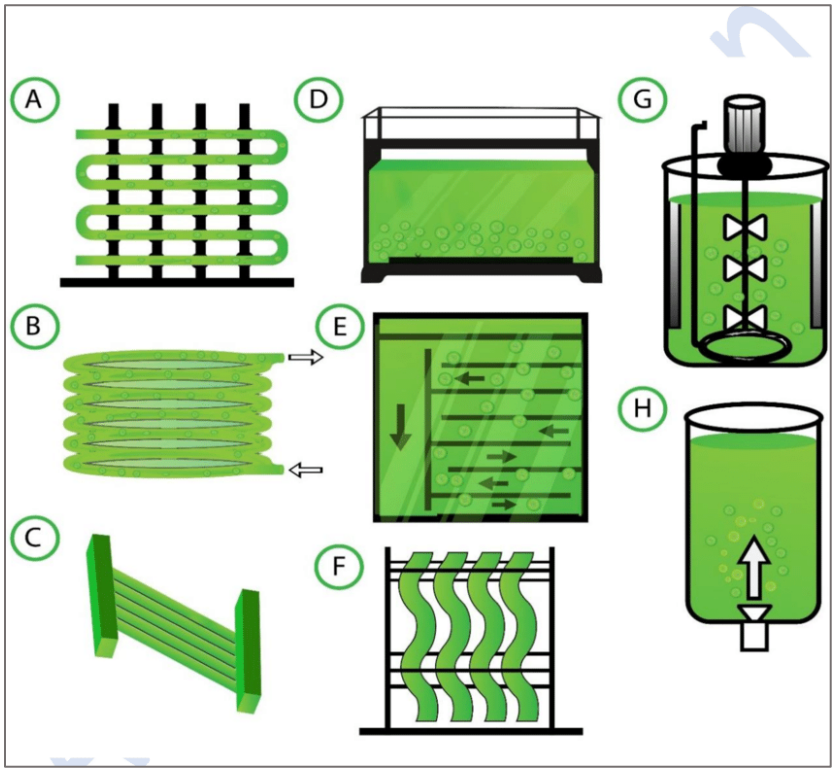


Figure 1-20: ***Schematic representation of different PBRs;*** *(A) fence tubular; (B) helical tubular; (C) horizontal tubular; (D) vertical flat panel; (E) air lift type; (F) accordion type; (G) stirred tank; (H) bubble column. Source: [58]*

Finally, since cannabinoids are produced into the apoplast of glandular trichomes of *C. sativa*, suggesting that CBDAS enzyme is able to pass throw the secretory pathway, and confirmed in heterologous systems and transgenic tobacco[22, 59]. Indeed, previous study highlighted the cannabinoid synthases structural changes during the secretory pathway, processing post-translational modifications. Seven possible Asn N-glycosylation sites have been already confirmed for cannabinoid synthases, and the enzymes were found to be highly glycosylated in heterologous systems[60, 61]. On the other hand, in 2019, a novel endogenous promoter designated 'highly abundant secreted protein 1 (HASP1) and its signal peptide were identified and characterizes in *P. tricornutum*[62]. For instance, using an episome for heterologous DNA propagation provides key advantages, such as preventing integration into low-expression genomic regions and reducing the risk of random gene inactivation in *P. tricornutum*. On top of that, the HASP1 promoter strongly drove protein expression during all growth phases of *P. tricornutum*, and the HASP1 signal peptide was sufficient to enable the secretion of GFP. This makes them a valuable tool for metabolic engineering applications. As a result, in combination with cultivations methods and genetic manipulation, *P. tricornutum* emerges as the most suitable diatom model species for focusing similar research efforts[52].

1.7 Problematic, goal and research objectives

Despite significant advances in understanding the biosynthesis of cannabinoids in *Cannabis*, their industrial production remains limited by sustainability challenges, low yields, and environmental costs. As a result, the search for alternative biotechnological platforms for cannabinoid production has intensified. *P. tricornutum*, an emerging model in synthetic biology, presents a promising solution due to its photosynthetic capacity, sequenced genome, and the availability of advanced molecular tools, including episomal expression systems. Within this context, the central research question addressed in this thesis is: *Can key enzymes of the cannabinoid biosynthetic pathway, such as CBDAS, be functionally expressed in the diatom P. tricornutum to initiate heterologous cannabinoid biosynthesis, leveraging its endogenous secretion pathway and episomal expression potential?* To address this question, the research builds on recent

developments in microalgal metabolic engineering and the production of specialized plant metabolites. It seeks to fill a knowledge gap, as the production of cannabinoids in diatoms remains largely unexplored despite favorable conditions for developing sustainable biotechnological approaches.

The present work forms part of a broader laboratory initiative in collaboration with the company Algae-C, which aims to develop sustainable biotechnological platforms for producing plant-derived pharmaceutical compounds using engineered microorganisms. Cannabinoids, particularly cannabidiol (CBD), represent one such class of therapeutically promising molecules. To date, they are primarily obtained via plant extraction, a process constrained by low yield, seasonal variability, and ecological concerns such as overharvesting and biodiversity loss. These limitations underscore the urgent need for alternative, scalable, and eco-friendly production strategies. In our lab, the cannabinoid biosynthetic pathway has been partially reconstructed in a microalgal host system—*P. tricornutum*—highlighting its potential as an efficient chassis for microbial cannabinoid production. The goal is to use actual molecular tools, to introduce cannabis genes into *P. tricornutum*, allowing the expression of the enzyme CBDAS, and evaluating the production of CBDA/CBD. This work was structured around main objectives:

Objective 1: establishing an heterologous expression platform to express CBDAS in *P. tricornutum*: Generating episomes carrying *CBDAS* variants, incorporating the native signal peptide (CBDAS) or the highly abundant secreted protein 1 secretory signal peptide (SP:CBDAS). *CBDAS* variants were tagged with the yellow fluorescent protein to perform cellular localization using confocal microscopy. And finally, evaluating protein accumulation and CBDA production by western blot and enzymatic assays. While I was investigating objective 1, I came across interesting results, that pushed us to investigate objective 2, studying the presence of sequences potentially encoding thrombin-like proteins.

Objective 2: Performing *in silico* analyses to identify sequences potentially encoding thrombin-like proteins, which are involved in recognizing and cleaving the thrombin sequence LVPRGS in *P. tricornutum* that we used in our constructs. Protein structure predictions, generating recombinant constructs with a thrombin cleavable sequence, evaluating and validating cleavage

efficiency of thrombin-like protease on fusion proteins *in vivo*. Finally, this study may advances bioengineering tools for diatom-based biotechnological applications.

Chapters 2 and 3 consist of published articles corresponding to objectives 1 and 2, respectively, followed by Chapter 4, which provides a general discussion and conclusion. Additionally, appendices I and II contain supplementary data, including additional tables and figures from the two published articles. Finally, appendices 7 and 8 present two coauthored works: one published article and one review. For the article, I contributed by generating the strains used in the study, performing protein analysis (Western blot) and confocal analysis, as well as assisting with writing, reviewing, and editing the manuscript. Regarding the review, I was involved in data collection and analysis, as well as contributing to writing and reviewing. Although these two appendices are not directly related to the core objectives of this thesis, they contribute to broader research efforts aimed at deepening our understanding of *P. tricornutum* and optimizing its use as a chassis for the production of high-value compounds.

CHAPTER II

This chapter explores the production of cannabidiolic acid (CBDA) using a microbial heterologous platform. *Phaedodactylum tricornutum's* clones containing different variants of CBDA synthase (CBDAS), responsible of the conversion of cannabigerolic acid (CBGA) to CBDA were characterized here. This chapter was written in article format and published in the **Algal Research journal.**

Contribution: Anis Messaabi and Elisa Fantino performed plasmid construction, microbial transformation, strain selection, protein extraction and analysis, cellular localization and enzymatic assays experiments. Natacha Merindol helped with flux cytometry experiments and contributed to writing the draft and revision. Fatima Awwad, Nicolas Sene, Kimy-Li Rhéaume and Fatma Meddeb-Mouelhi contributed to review & edit the draft. Sarah-Eve Gélinas and Alexandre Custeau performed LC-MS analysis. Isabel Desgagné-Penix supervised the project, provided funding and material for the project, and assisted with the writing and revising of the paper.

2 Extrachromosomal expression of functional cannabidiolic acid synthase in *Phaedodactylum tricornutum*

Elisa Fantino[1,2†], Anis Messaabi[1†], Natacha Mérindol[1], Fatima Awwad[1], Nicolas Sene[1], Sarah-Eve Gélinas[1], Alexandre Custeau[1], Kimy-Li Rhéaume[1], Fatma Meddeb-Mouelhi[1,2], Isabel Desgagné-Penix[1,2*]

[1]Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada.
[2]Plant Biology Research Group, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada.

[†]These authors contributed equally to the work.
**Correspondance:* Isabel.Desgagne-Penix@uqtr.ca

Received 13 October 2024, Revised 12 December 2024, Accepted 29 December 2024, Available online 31 December 2024, Version of Record 2 January 2025.
https://doi.org/10.1016/j.algal.2024.103889

2.1 Abstract

Cannabis sativa's cannabidiolic acid (CBDA) offers significant therapeutic potential without inducing psychotropic effects but is typically found as part of a complex mixture of metabolites in plant extracts. Using a heterologous expression platform could allow the production of pure CBDA. Here, we propose to express CBDA synthase (CBDAS) in *Phaeodactylum tricornutum.* Episomes carrying *CBDAS* variants, incorporating the native signal peptide (CBDAS) or the highly abundant secreted protein 1 secretory signal peptide (SP:CBDAS) were constructed. *CBDAS* variants were tagged with the yellow fluorescent protein (YFP), introduced into the marine diatom, and screened by fluorescence. Confocal microscopy revealed that CBDAS and SP:CBDAS arranged in aggregated structures indicative of secretory pathway involvement. Western blot assays confirmed whole construct accumulation intracellularly, while soluble YFP was detected extracellularly. Finally, enzymatic assays showed CBDA production by both CBDAS and SP:CBDAS strains, confirming the potential of *P. tricornutum* as a platform for cannabinoid biosynthesis.

2.2 Keywords

Microalgae; cannabinoids; biofactories; cannabidiolic acid synthase; marine platform; episome; secretory pathway.

2.3 Abbreviations

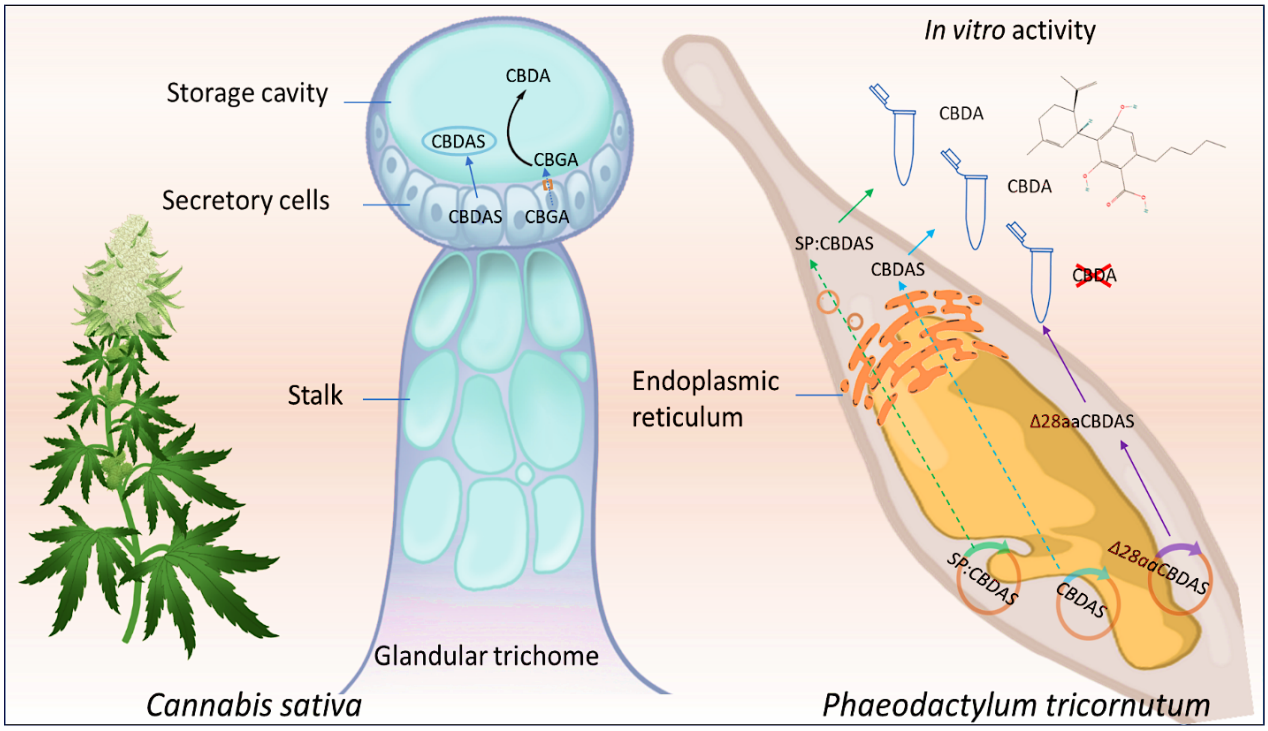
Au, Absorbance units; CBD, Cannabidiol; CBDA, Cannabidiolic acid; CBDAS, Cannabidiolic acid synthase; CBGA, Cannabigerolic acid; EV, Empty Vector; ER, Endoplasmic Reticulum; FAD, Flavin Adenin Dinucleotide; HA, Hemagglutinin, HASP1, highly abundant secreted protein; OD, Optical density; SP, Signal peptide; THC, Δ9-tetrahydrocannabinol; THCA, Δ-9-tetrahydrocannabinolic acid; Thr, Thrombine; YFP, yellow fluorescent protein.

2.4 Highlights

- Microalgae production of cannabidiolic acid (CBDA), a bioactive cannabinoid (CB)
- Promising photosynthetic platform for sustainable biomanufacturing pharmaceuticals

- Cannabis CBDA synthase was successfully heterologously expressed in *Phaeodactylum tricornutum*
- Fluorescent tagging and microscopy used to track CBDA synthase localization
- Secretory signal peptides enhance CBDA synthase targeting and secretion pathways

2.5 Graphical abstract



2.6 Introduction

Cannabis sativa-based preparations are prescribed in the USA and Canada to treat nausea following cancer chemotherapy, appetite loss due to acquired immune deficiency syndrome, and to provide symptomatic relief from neuropathic pain in multiple sclerosis patients[63]. Cannabinoids (CBs) are a group of metabolites long coveted for their psychoactive and medicinal properties[7, 64-67]. Δ^9-Tetrahydrocannabinol (THC) and cannabidiol (CBD), as well as their carboxylated forms Δ^9-tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA), are the potent CBs associated with the therapeutic benefits of *C. sativa (Cs).* CBD, in particular, has been

reported to reduce sustained anxiety and improve sleep while being devoid of psychotropic effect[68-70]. In planta, cannabinoid and terpene biosynthesis pathways occur in secretory cells of the glandular trichomes of *C. sativa.* Both pathways utilize precursors such as geranyl pyrophosphate from the methylerythritol phosphate pathway and intermediate such as acetyl-CoA from fatty acid metabolisms[71]. THCA and CBDA synthesis is catalyzed in the glandular trichome cavity by the THCA synthase (*Cs*THCAS) and the CBDA synthase (*Cs*CBDAS), respectively. Both secreted enzymes are members of the berberine bridge enzyme-like gene family, which contains a 28 amino acids (aa) N-terminal signal peptide and a flavin adenine dinucleotide (FAD) binding domain[72-74]. These FAD-dependent oxygenases catalyze the oxidative cyclization of cannabigerolic acid (CBGA), the last step of the major CB pathway. The structural and functional properties of *Cs*CBDAS are quite similar to those of *Cs*THCAS with 84% identity in the amino acid sequences[72, 75]. Both possess a disulfide bond and several N-glycosylation sites[73]. The main difference between the reactions they catalyze is the proton transfer step, i.e. *Cs*CBDAS extracts a proton from the methyl group at the end of CBGA to yield CBD, while the *Cs*THCAS enzyme extracts a hydroxyl group to produce THC[75, 76].

The isolation of pure CBD, desirable for its specific pharmaceutical properties, from *C. sativa* is challenging, as the extraction results in formulations containing hundreds of plant chemicals and minor cannabinoids[77, 78]. Although there has been progress, its chemical synthesis has not been proven to be a cost-effective method either, mainly due to the complexity of its structure, resulting in high production and environmental costs, and suboptimal yields[79]. For these reasons, there is a need for alternative production platforms that are controlled and sustainable. To this aim, CBs pathway enzymes have been introduced in heterologous prokaryotic and eukaryotic hosts such as *E. coli*[80], yeast [81-85], model plants like *Nicotiana benthamiana[86],* filamentous fungi *like Penicillium chrysogenum*[87], and marine diatoms such as *Phaeodactylum tricornutum[88, 89].* These studies achieved variable yields of CBs and precursors, reaching up to hundreds of mg/L[90], and revealed that CBs are more suitable to be produced in eukaryotic compared to prokaryotic organisms. For instance, expression in *E. coli* cytosol failed to yield an active *Cs*THCAS[82]. The coding sequence of *Cs*THCAS fused to different signal peptides and tagged with the yellow fluorescent protein (YFP) was transiently heterologously expressed in *N.*

benthamiana[86, 91, 92]. Interestingly, *Cs*THCAS enzyme was detected only when targeted to the endoplasmic reticulum (ER); while cytosolic and plastid localization resulted in no detectable protein[92]. Recent studies suggested that glycosylation and other post-translational modifications of *Cs*THCAS occurring in the ER might contribute to the correct folding of the enzyme and influence its stability[84, 92].

Directing *Cs*THCAS or *Cs*CBDAS to specific intracellular compartments, such as the vacuoles, was also key to their function and CBs production in yeast[83-85, 93]. These studies indicated that the expression and translocation of *Cs*THCAS and *Cs*CBDAS to specific intracellular compartments might result in the stabilization of the enzyme and increase their activity, possibly involving the pH-dependency of the reaction. Moreover, the translocation of the enzymes into specific compartments might shield the cells from the accumulation of CBs in the cytosol which was shown to be toxic to C. sativa protoplasts and *P.* tricornutum cell cultures[89, 91]. Thus, an efficient *Cs*CBDAS heterologous platform would ideally allow for eukaryotic post-translational modifications through compartmentalization or secretion of the enzyme, or its product, and contain the necessary pathway precursors.

Microalgae are unicellular photosynthetic organisms emerging as heterologous hosts to produce plant specialized metabolites[94, 95]. Diatoms are especially resistant to diverse environments and play an important role in fixing carbon from the ocean[96]. *P. tricornutum*, in particular, is known for its rapid growth rate and richness in fatty acids [97], and has been successfully used as a platform to produce lipids, terpenoids, and geraniol[98-101]. Recently, we demonstrated that diatom cell cultures could support in vivo production of the cannabinoid precursors olivetolic acid and CBGA following heterologous expression of tetraketide synthase and olivetolic acid cyclase [88, 89]. Based on these characteristics, we hypothesize that *P. tricornutum* could be modified to efficiently biosynthesize CBDA.

This study aims to address these challenges associated with functional CBDAS heterologous expression and explore the suitability of producing CBDA in a sustainable and efficient manner. We engineered diatom strains with episomes containing three versions of the *Cs*CBDAS sequence driven by the *P. tricornutum* endogenous promoter from the highly abundant secreted protein 1 (HASP1)[102, 103]. The cassettes included 1) the native *Cs*CBDAS sequence, 2) a truncated

sequence lacking the 28 aa N-terminal signal peptide (SP), and 3) a hybrid sequence, replacing the native SP with HASP1 18 aa secretory SP to direct *Cs*CBDAS to the extracellular matrix, mimicking enzymes behavior in planta. All sequences were codon-optimized for expression in the diatom and fused to YFP at the C-terminal, to easily follow protein accumulation and subcellular localization. After transformation, *P. tricornutum* strains harboring *Cs*CBDAS complete sequence (CBDAS), truncated (Δ28aaCBDAS), and secreted version (SP:CBDAS) were characterized by flow cytometry, immunoblotting, confocal microscopy, and enzymatic assay. SP:CBDAS transconjugants yielded more clones with higher fluorescence compared to the other two designs. Both CBDAS and SP:CBDAS formed aggregate-like structures between the two plastid lobes, suggesting secretory pathway localization. Enzymatic assays confirmed that enzymes were active in both contexts, yielding CBDA as a product, whereas constructs without signal peptide were not active. The successful heterologous expression of *Cs*CBDAS in *P. tricornutum* represents an important step in the development of microalgal-based biofactories for cannabinoid production. Moreover, our results reinforce the potential of the marine diatom, as a platform for efficient production of complex metabolites with pharmaceutical applications

2.7 Results and Discussion

2.7.1 Engineering P. tricornutum strains to express CsCBDAS

To metabolically engineer the brown algae *P. tricornutum* to biosynthetize CBDA, we designed three episomes harboring different versions of the gene encoding *Cs*CBDAS (Fig. 2-1A). Gene expression was driven from 499-bp of the *HASP1* promoter (HASP1p); followed by 1) a complete codon optimized *CsCBDAS* gene sequence (CBDAS), 2) a truncated version (Δ28aaCBDAS) without the secretory signal peptide (SP, 84 bp), or 3) the truncated version with an additional 54 bp from the 5' coding sequence of *HASP1* (SP:CBDAS). Enzyme sequences were tagged with *YFP* and 3x hemagglutinin (HA) at their C terminal, linked by a thrombin cleavage sequence (LVPRGS)[104]. Three additional episomes were included as controls; the empty vector (EV) containing only the nourseothricin resistance cassette with the *nat* gene, a plasmid with the *YFP* reporter gene (YFP) under HASP1p, and one with *YFP* fused with the HASP1 SP (SP:YFP) at its N-terminal under HASP1p

(Fig. 2-1A). These five constructs were successfully transformed into *P. tricornutum* by *E. coli* conjugation, and characterized for YFP signal, protein detection, and enzymatic assay (Fig. 2-1B).

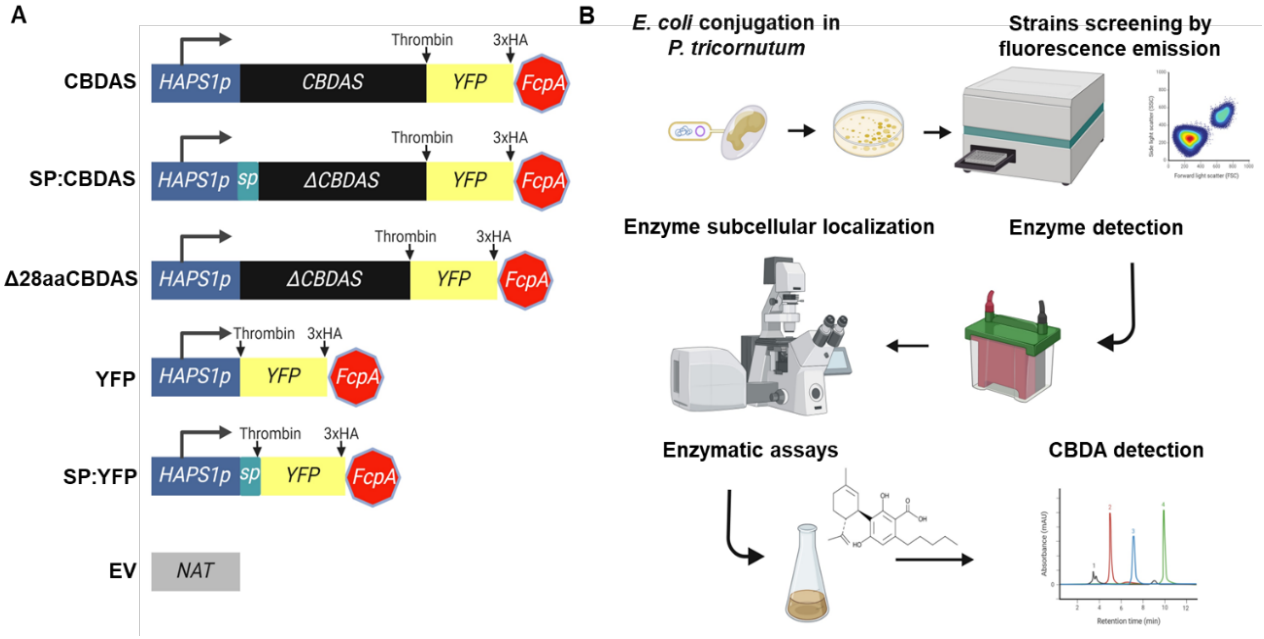


Figure 2-1: : ***Expression system and platform development.***

A. Scheme of the recombinant cassettes expressing cannabidiolic acid synthase (CsCBDAS) driven by the diatom promoter HASP1p and tagged with the yellow fluorescent protein (YFP) reporter gene. Expression cassettes included the complete codon-optimized native sequence of CsCBDAS (CBDAS), a truncated version (Δ28aaCBDAS), and a truncated version plus 54 bp from the 5' coding sequence of HASP1 (SP:CBDAS). All sequences were tagged in the C terminal with YFP and 3xhemagglutinin (HA), linked to the enzyme by a thrombin cleavage sequence (LVPRGS). Moreover, three controls were used, i.e. YFP gene, SP:YFP both driven by HASP1p, and Empty vector (EV) encoding only the nat gene conferring nourseothricin resistance. B. Study workflow. P. tricornutum cells were transformed by E. coli conjugation, then clones were screened and analyzed by fluorescence emission in a microplate reader and flow cytometer, enzyme size was verified by immunoblotting assay and subcellular localization was observed in the confocal microscope. Enzyme activity was confirmed by the detection of CBDA on an HPLC-DAD. The figure was created with BioRender.com [105].

Two weeks post-transformation, 36 colonies of each strain were randomly selected and grown in nourseothricin-containing plates for 10 more days, for a second antibiotic selection round. At that stage, colonies were screened by epifluorescence microscopy. The frequency of YFP$^+$ colonies, *i.e.* with higher fluorescence compared to the EV strain, was computed for each line (Table 2-1).

Table 2-1: ***Transconjugant strains screening by yellow fluorescent protein (YFP) fluorescence.***

Thirty-six randomly selected colonies of each strain, CsCBDAS:YFP (CBDAS), CsCBDAS fused to the HASP1 signal peptide (SP:CBDAS), and a truncated CsCBDAS version without the first 28 aa (Δ28aaCBDAS) were screened under a fluorescence microscope. YFP+ clones selected from the first screen were grown in L1 liquid media for 10 days in 96-well plates, and YFP

fluorescence was analyzed with a microplate reader (second screen). Empty vector (EV) strain was used as a negative control to settle the basal autofluorescence.

Strain	1st screening YFP+ clones (Fluorescence microscopy)	2nd screening YFP+ clones ($Ex_{500}Em_{539}$)/OD_{680} nm
CBDAS	30.6% (n=11/36)	72.7% (n=8/11)
SP:CBDAS	50% (n=18/36)	100% (n=18/18)
Δ28aaCBDAS	19.4% (n=7/36)	100 % (n=7/7)

SP:CBDAS strain presented a higher percentage of YFP+ colonies (50%) compared to CBDAS (30.55%) and Δ28aaCBDAS (19.44%) strains (Table 2-1). Strains that presented fluorescence were grown in liquid culture for 10 days and further analyzed in a second screening round by recording YFP emission normalized on culture OD_{680} nm in a microplate reader, using EV strain as negative control (Fig. 2-2A, B; Fig. S1). Fluorescence was detected in YFP, SP:YFP, CBDAS, SP:CBDAS and Δ28aaCBDAS transconjugants, and SP:CBDAS clones displayed more fluorescence compared to CBDAS (One way Anova, Tukey post-test $p<0.05$, Fig. 2-2A). A threshold equivalent to the mean background fluorescence detected in the EV clones, plus twice the standard deviation of that mean, was used to determine positive clones (Table 2-1, Fig. 2-2B). All three YFP and SP:YFP clones, used as positive controls, were fluorescent (Fig. 2-2A, B). Eight of the 11 clones of the CBDAS strain (72.7%) remained fluorescent (Table 2-1). All tested clones from SP:CBDAS and Δ28aaCBDAS displayed YFP fluorescence above the threshold (Fig. 2-2B).

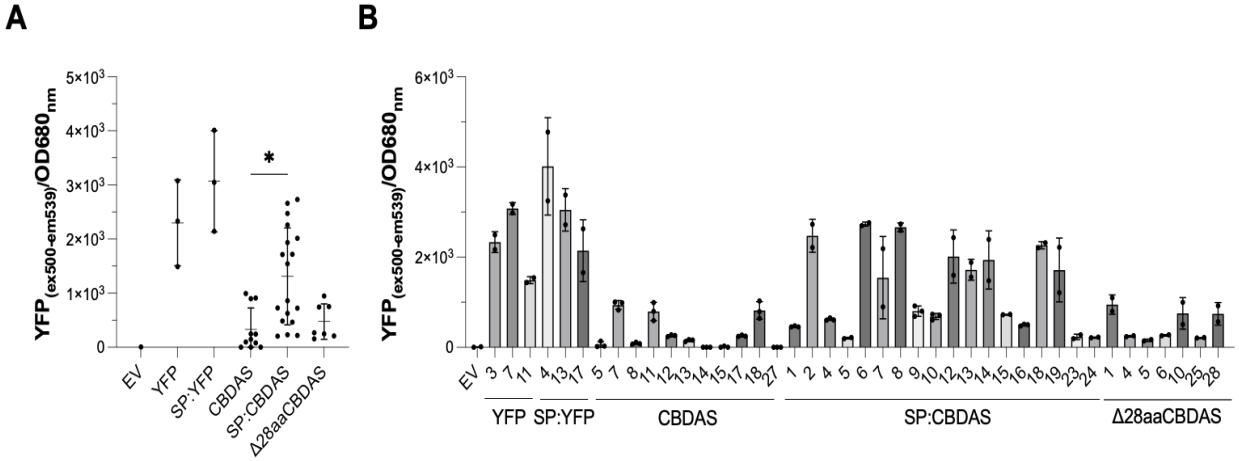


*Figure 2-2: **CBDAS strains screening for YFP fluorescence.***

The strains selected by the fluorescence microscope were transferred to liquid media where the YFP fluorescence intensity was measured by the microplate reader, normalized to the OD680 nm, and compared with EV strain autofluorescence (A). Mean ±

*SD; n=3 biologically independent samples for each strain were plotted. The clones with a YFP fluorescence higher than the EV fluorescence (mean + (SD) x 2) were determined as YFP positive (B). A One way Anova with Tukey post-test was performed (*p<0.05).*

Thus, our results indicated that the constructs were successfully transformed and heterologously expressed by *P. tricornutum*, and that SP:CBDAS transconjugants presented a higher YFP fluorescence compared to CBDAS constructs. To some extent, these results are in contrast with Erdene-Ochir observations, which suggested that the addition of the SP signal to GFP decreased transcript levels compared to GFP alone constructs[102]. The authors suggested that the presence of the SP sequence could affect the steady-state level of GFP mRNA. In our study, the addition of the SP signal had a positive impact at the fluorescence level, but we used different constructs, and episomal expression rather than randomly integrated genomic expression, limiting our ability to compare with Erdene-Ochir *et al.*[102].

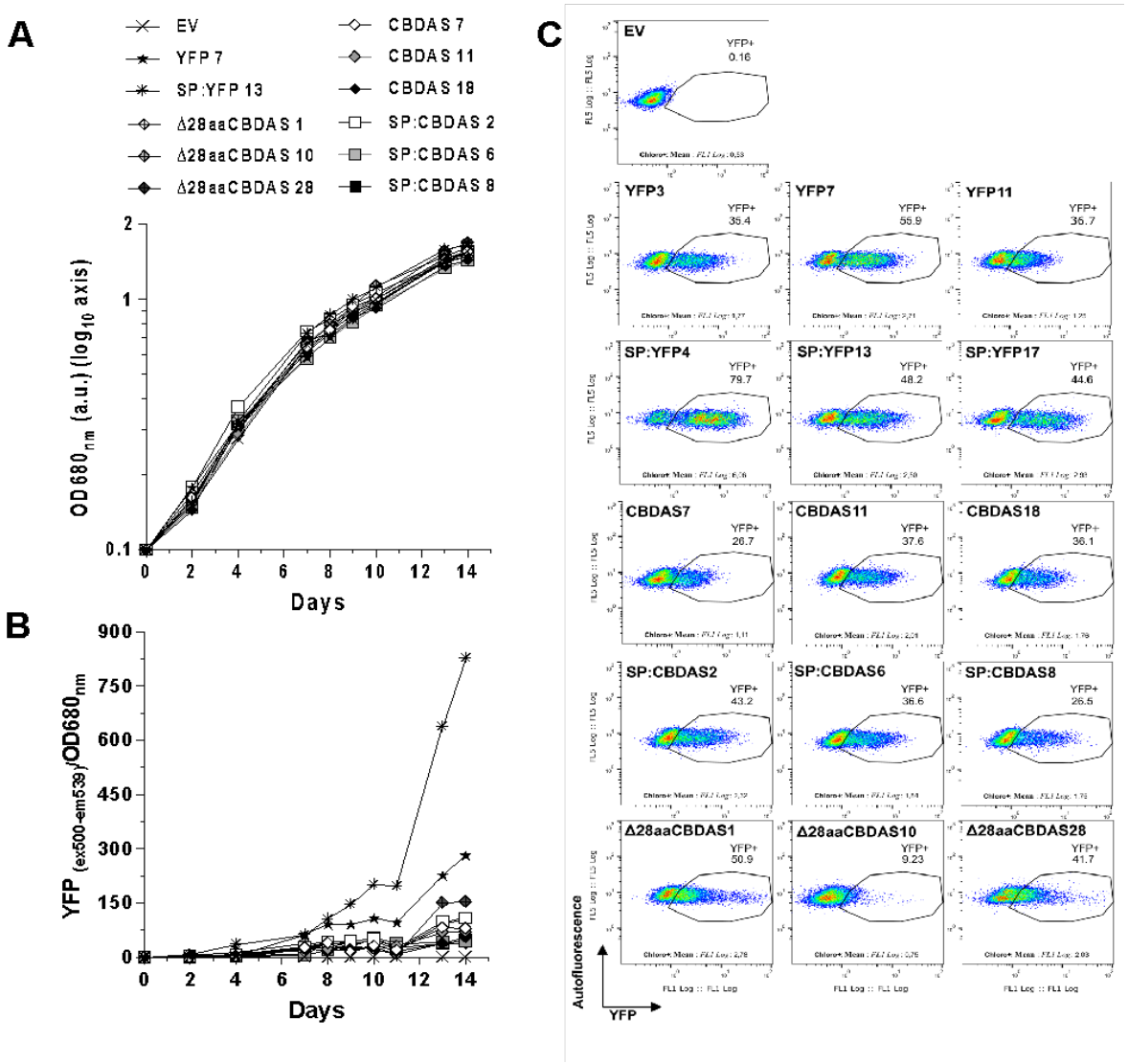


Figure 2-3: ***Characterization of the selected transconjugants.***

A. Growth curves of each strain, the optical density (OD) at 680 nm was followed for 14 days. B. YFP fluorescence emission (excitation 500 nm- emission 539 nm) was measured and normalized to the OD for the 12 cultures monitored in A. C. Representative dot plots of 3 selected clones of each strain. Pseudo-color dot plots of empty vector (EV) and transformants strains; YFP fluorescence was detected at 530 nm on the x-axis, and autofluorescence at 448 nm on the y-axis. Gates and frequencies of total YFP populations were designed according to the autofluorescence of the negative control EV shown as a reference.

2.7.2 Characterization of the selected fluorescent clones

Three clones with the highest fluorescence for each strain were selected for further characterization. The culture growth curves and YFP fluorescence were registered for 14 days and compared to the EV strain (Fig. 2-3 A and B). Strains presented a similar pattern, showing that the presence of the transgene did not negatively affect the growth of *P. tricornutum* cells. The YFP

fluorescence increased from day 7, as shown before[102, 103], with clone SP:YFP 13 showing the highest level at day 14.

Clones were then analyzed on a flow cytometer at day 10 (Fig. 2-3C). The EV strain was used as a negative control. YFP clones 3, 7, and 11 presented 35.4 %, 55.9 %, and 36.7 % of YFP^+ cells, respectively, while SP:YFP clones 4, 13, and 17 showed 79.7 %, 48.2%, and 44.6 % of YFP^+ cells. The selected CBDAS clones (7, 11, and 18) and SP:CBDAS (2, 6, and 8) showed similar frequency of YFP^+ cells; 26.7 %, 37.6 % and 36.1 % *vs.* 43.2 %, 36.6 % and 26.5 %. Regarding Δ28aaCBDAS, clone 10 presented a cell population with a lower percentage (9.23 % of YFP^+ cells), compared to clones 1 and 28, with 50.9 %, and 41.7 %, respectively. These results are consistent with George *et al*, showing that episomal expression is an efficient method to generate *P. tricornutum* transconjugants that express transgenes, requiring less screening compared to random integration methods[101].

2.7.3 Subcellular localization of CBDAS in P. tricornutum transconjugants

To assess the subcellular localization of CBDAS from the three constructs, *i.e.* without SP (Δ28aaCBDAS), with endogenous SP (CBDAS), and with HASP1 SP (SP:CBDAS), transconjugants were visualized on a confocal microscope, acquiring YFP fluorescence (yellow), and chlorophyll autofluorescence (red) using EV, YFP and SP:YFP as controls (Fig. 2-4). YFP fluorescence in the YFP control strain was visible throughout the cytoplasm at the early stationary phase (day 10). As previously observed, the HASP1 SP caused considerable changes in the subcellular localization of YFP, clustering in the chloroplast ER rather than the cytoplasm, as a clustered dotted structure between the two plastid lobes[102, 106-108]. According to Erdene-Ochir *et al*, this suggests that the HASP1 signal peptide leads to YFP entry into the secretory pathway. Interestingly, both CBDAS and SP:CBDAS strains presented a localization pattern consistent with clustered structure between the plastid lobes, similar to SP:YFP (Fig. 2-4, S2 and S3). CBDAS fluorescence pattern also corresponded to the signal of the medial Golgi marker protein XylT[108]. A similar pattern was observed when the potential retromer subunit Vps29, the structure involved in endosomal retrograde transport to the Golgi apparatus, was heterologously expressed in *P. tricornutum*. These results suggest that CBDAS and SP:CBDAS both localized in the secretory pathway.

Unexpectedly, the soluble version of Δ28aaCBDAS enzyme did not behave as a cytosolic protein, although more diffuse than CBDAS and SP:CBDAS, aggregates were observed close to the plastid (Fig. 2-4 and S4), in all three clones. Overall, the localization pattern of SP:CBDAS and CBDAS suggests that they cluster in the chloroplastic ER, where post-translational modifications like N-glycosylation, required for activity, stability, and correct folding of endogenous as well as transgenic enzymes, occur[109].

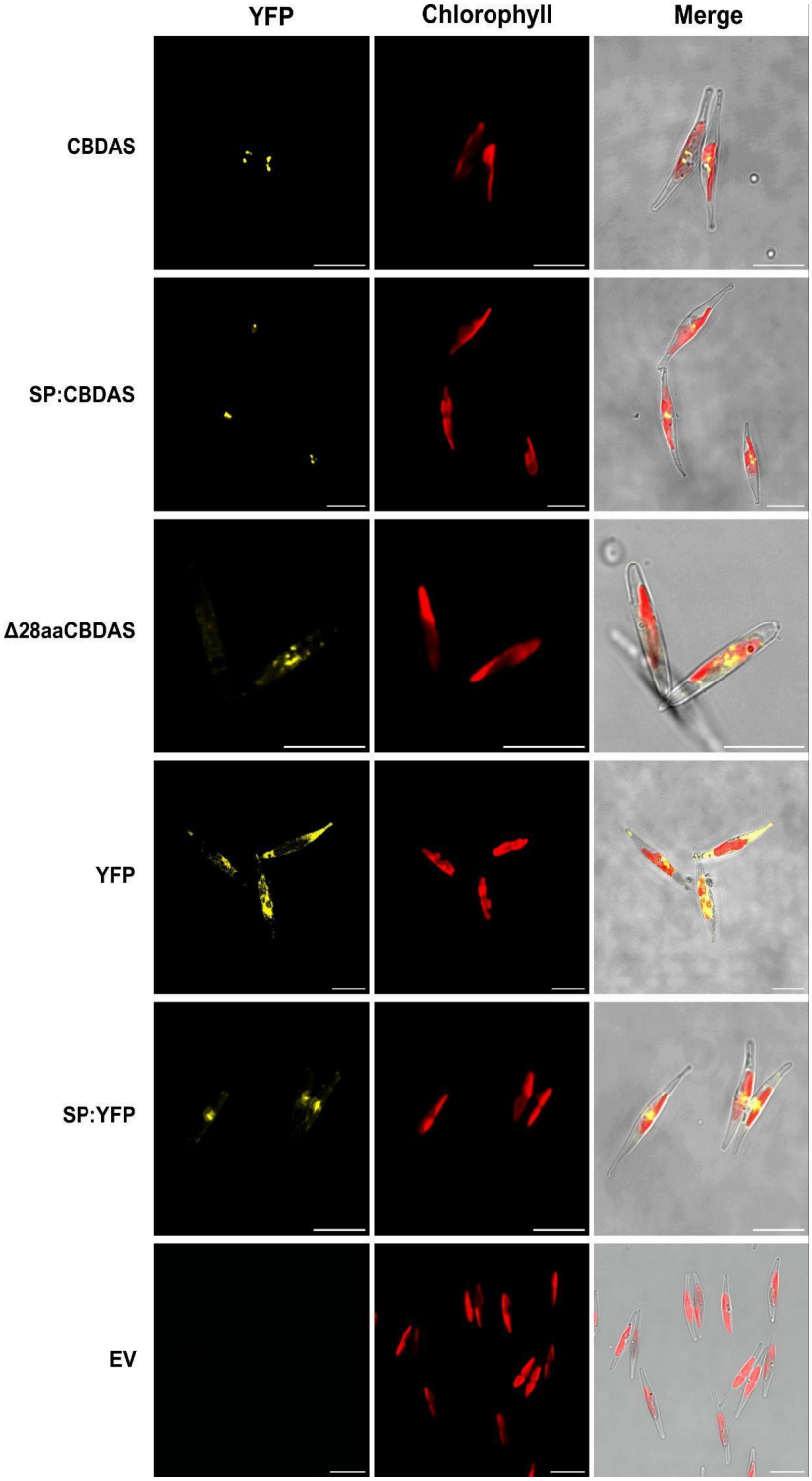


Figure 2-4: ***SP:CBDAS and CBDAS localize at the chloroplastic endoplasmic reticulum in P. tricornutum transconjugants.*** *YFP fluorescence, chlorophyll autofluorescence, and the merging of three fields are shown in transgenic lines producing CBDAS:YFP, SP:CBDAS:YFP, Δ28aaCBDAS:YFP, SP:YFP, and YFP were visualized by confocal laser microscopy. YFP and EV cells were used as positive and negative controls. Scale bars = 10 μm.*

2.7.4 CBDAS accumulation and secretion

To analyze the impact of HASP1 signal peptide on enzyme secretion into the extracellular matrix, we performed a western blot using protein pellets from cell extracts and supernatants from stationary phase cultures (Fig. 2-5-5A & B). As expected, the EV protein extract did not show any bands following immunoblotting with anti-YFP (Fig. 2-5A), while the positive control (10 ng of purified GFP) showed the predicted band at 27 kDa. YFP clones presented the expected band at 32 kDa (Thr:YFP:3HA) and a band below it (∼30 kDa). The YFP clone 11, presented both bands, but the upper band showed higher intensity. In the case of the SP:YFP strains, clones showed the expected band (~34 kDa, SP:Thr:YFP:3HA), with additional smaller bands at high intensity for clone 4. For further assays, clones YFP 7 and SP:YFP13 were used. All cell extracts from CBDAS clones and SP:CBDAS produced proteins of the predicted size of 92-93 kDa, confirming intracellular accumulation. When Slattery *et al.* heterologously expressed the receptor-binding domain of the SARS-CoV-2 spike protein, under a longer version of HASP1p containing the secretory signal (642-bp), only one of the 9 selected clones produced the desired protein [36]. In our study, the fusion of CBDAS to an endogenous and recognized signal peptide by the secretory mechanism could stabilize the enzyme, as proven when targeted to the ER[92]. To verify, if CBDAS and SP:CBDAS were secreted, the extracellular media was concentrated 250 times, and analyzed by western blot (Fig. 2-5B). YFP clone 7 and EV strains supernatants were used as controls and no specific band was detected. However, SP:YFP, CBDAS, and SP:CBDAS displayed bands close to YFP size (∼ 27

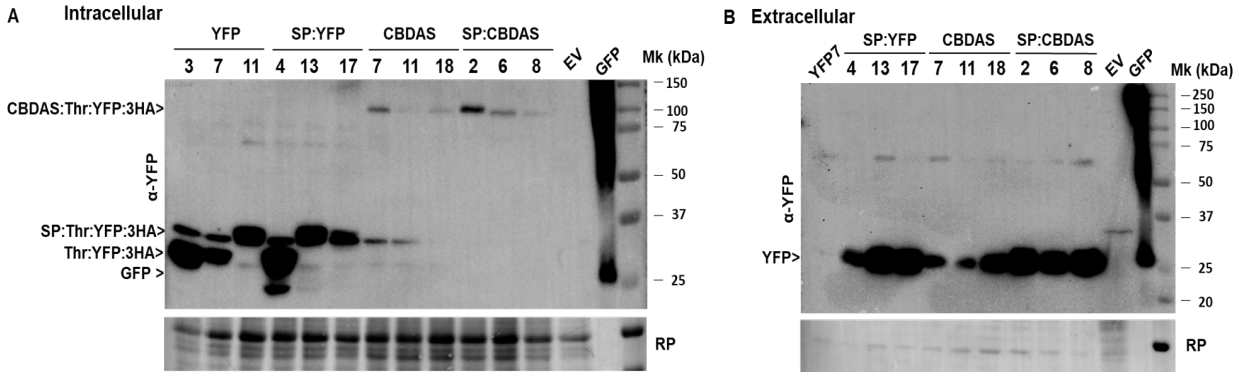


Figure 2-5: ***CBDAS protein detection. Western blot analysis using anti-YFP antibody of total intracellular protein extract from cell lysate and extracellular culture.*** *(A) cell lysates and (B) culture supernatants. YFP (Thr:YFP:3HA, 32 kDa), SP:YFP (SP:Thr:YFP:3HA, 34 kDa), CBDAS (CBDAS:Thr:YFP:3HA, 93 kDa), SP:CBDAS (SP:CBDAS:Thr:YFP:3HA, 92 kDa), EV and purified GFP/YFP (27 kDa). Lower panel, stained blot with red ponceau solution (RP). Full-length Western blots are presented in Supplementary Fig. S5.*

kDa), but CBDAS:YFP was not detected. Uncropped blot and red ponceau stain are shown in Fig. S5.

Thus, we could not detect CBDAS and SP:CBDAS in the extracellular media, only the YFP cleaved portion was visible. The detection of YFP in the extracellular is consistent with Erdene-Ochir *et al.*, which detected secreted GFP from SP:GFP constructs by western blot during the stationary phase[110]. The absence of detection of the enzymes could be due to degradation during the secretion process, and/or cleavage from YFP. Studies of protein production in yeast showed that when cells reached their maximal secretory adaptation capabilities, the subset of high protein accumulator cells trigger protein degradation in the entire cell population, or transiently experience a secretion burnout[111]. Alternatively, a cleaved tag-free version of CBDAS could be present in the extracellular matrix but would remain undetectable. A previous study has suggested that the thrombin cleavage sequence could be recognized and cleaved in *P. tricornutum*[112].

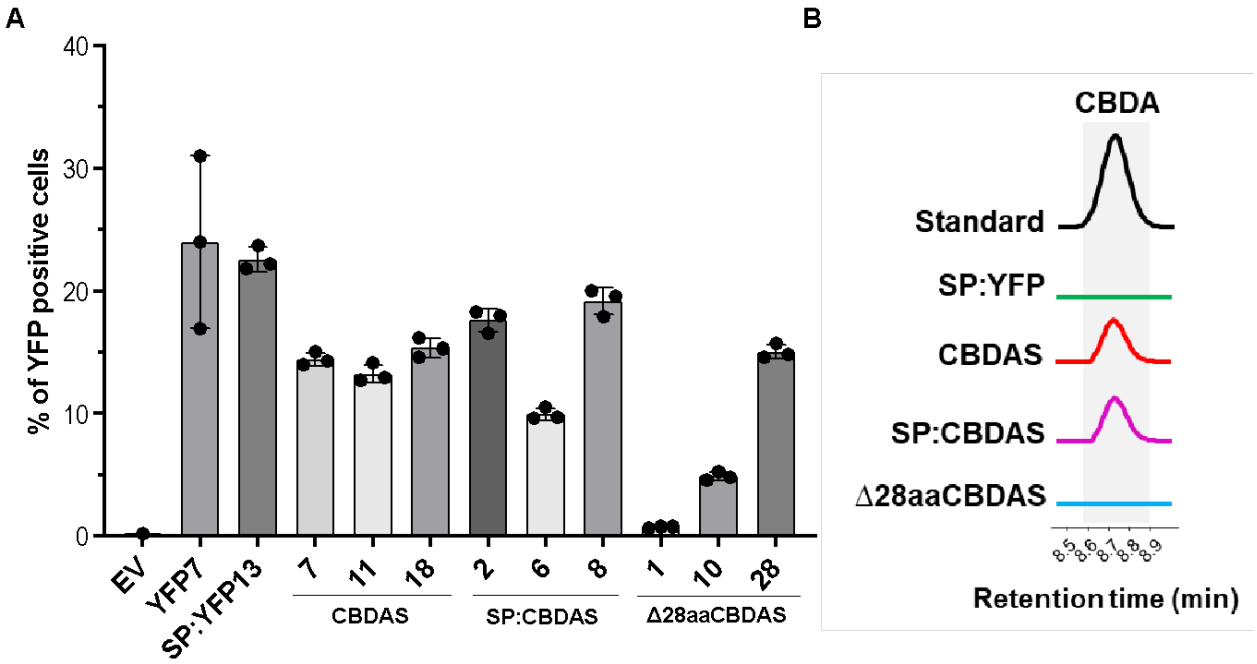


*Figure 2-6: P. **tricornutum transconjugants display CBDAS activity.***

A) Percentage of YFP+ cells of selected transconjugants after 21 days of culture. B) HPLC-DAD chromatograms of in vitro CBDA production. Extracts were analyzed by HPLC-DAD and signals were compared to authentic CBDA standard (Fig. S6, S7, S8, and S9).

2.7.5 CBDAS enzymatic activity

As the enzyme was detected in the cell extracts but not in supernatant, cell lysates were used to assess the activity of CBDAS transconjugants, similarly to Slattery *et al.*[103].

Clones were scaled up from the L1 agar plate colonies obtained after *E. coli* conjugation, to avoid possible mutations generated by continuous transfers[113]. The percentage of YFP^+ cells was measured in cell cultures grown to the stationary phase for 21 days (Fig. 2-6A). Total soluble protein was extracted from fluorescent CBDAS 7, 11, 18, SP:CBDAS 2, 6, 8, Δ28aaCBDAS clones 1, 10, 28, EV, YFP clone 7, and SP:YFP clone 13 transconjugants and used for enzymatic assay without protein purification, to avoid protein degradation. CBGA was used as a substrate, and the products of the reaction were analyzed and identified by HPLC-DAD (Fig. 2-6-6B, S5, S6). Neither CBDA nor any other traces of CBs were detected in cell protein extracts of SP:YFP and Δ28aaCBDAS strains. A signal eluting at 8.827 min corresponding to the CBDA standard was detected when CBDAS and SP:CBDAS transconjugants protein extracts were incubated with CBGA as substrate. The UV spectrum of the *in vitro* assay was also compared with the one of authentic reference standard CBDA (Fig. S7) and then the identification was confirmed by HPLC-MS/MS (Fig. S8). The highest yield corresponding to 0.55 mg/L of CBDA was obtained in SP:CBDAS clone 6 cell protein extract. Thus, both constructions, with endogenous signal peptide (CBDAS) and HASP1 signal peptide (SP:CBDAS) led to the *in vitro* production of CBDA, while removing the signal peptide abolished the enzyme activity. These results add to the growing evidence that the enzyme CBDAS requires to be processed through the secretory pathway. It also adds to our previous studies suggesting that the diatom is a suitable host for CBs production[88, 89].

The first report of CBDAS heterologous production was published in 2018, using wild-type and mutant enzymes heterologously expressed in *Komagataella phaffii*[114]. The mutant strain CBDAS A414V+A46V+T47A was able to produce 0.42 $g.L^{-1}$ CBDA and 0.13 $g.L^{-1}$ of THCA when supplemented with CBGA[84], increasing CBDAS catalytic activity by 3.3 times[84]. In 2019, Luo *et al.* generated strains of *S. cerevisiae* that produced different CBs from galactose. The highest reported titers of CBDA by hexanoic acid supplementation were 4.3 $\mu g.L^{-1}$ CBDA[85]. A second group successfully implemented *de novo* biosynthesis of CBD in *S. cerevisiae*, yielding 6.92 mg/L,

by overexpressing the bile pigment transporter 1, which efficiently transfers CBGA from the cytoplasm to the vacuole, where catalysis takes place[76]. Experiments in *S. cerevisiae* with co-overexpression of WT *Cs*CBDAS and transcription factor HAC1, which activates the unfolded protein response to restore ER homeostasis[115], resulted in increased specific activities of 11-fold, as determined by *in vitro* assays[93]. Moreover, multiple copies of *Cs*CBDAS were integrated using a multicopy integration system, and the resulting strain produced 0.13 mg/L CBDA at pH 4 [93]. Overall, the heterologous expression systems for CB production still require optimization and fine-tuning to be profitable to the industry. However, these studies indicate several avenues for further optimization of *Cs*CBDAS activity in *P. tricornutum.*

Here, we succeeded in engineering the diatom *P. tricornutum* with an active CBDAS, showing the importance of its localization in the ER through the encoded signal peptide. This work highlights the importance of studying linkers and signal peptide behavior in new heterologous platforms, such as the diatom *P. tricornutum*, as well as characterizing the secretory network to construct more efficient tools.

Future research should include the evaluation of medium optimization (phosphate and iron deprivation)[103], directed mutagenesis, the use of smaller tags, different sequence of linkers[112], and increasing the gene copy number on CBDA titers in *P. tricornutum*. In addition, strategies to reduce CBs cytotoxicity in *P. tricornutum* cells[89] such as cell compartmentalization and the use of cannabinoid transporters must be considered.

2.8 Conclusion

We previously demonstrated the use of *P. tricornutum* as an *in vivo* heterologous production platform for olivetolic acid and CBGA, precursors of cannabinoids. In the present study, we report, for the first time, the production of one of the most critical enzymes in the cannabinoid biosynthetic pathway in microalgae. Our results showed that the production of three versions of CBDAS i.e. native, modified with an endogenous secretory signal peptide, and truncated one, can be produced in *P. tricornutum*. The YFP fluorescence-based screening method enabled high-throughput clone selection and facilitated the detection of high fluorescence in the constructs driven by the HASP1 promoter. Enzymatic assays confirmed the activity of CBDAS, with CBDA as

the product in both the native and secretory constructs. Overall, we demonstrated that CBDA production through the heterologous expression of CBDAS in *P. tricornutum,* achieves competitive levels when compared to yeast platforms. Although additional improvements are required, these findings validate the diatom as a platform for cannabinoid biosynthesis, offering a sustainable and scalable alternative to traditional production systems.

2.9 Methods

Plasmids Construction

Plasmids were constructed by Gibson assembly using the NEBuilder® HiFi DNA Assembly Bundle for Large Fragments (New England Biolabs, Canada). Amplicons used for assemblies were amplified by PCR with PrimeSTAR GXL DNA Polymerase (Takara Bio, Japon) following the manufacturer's protocol. Plasmid pPtGE31 is used as a template to construct the expression vectors[116]. The *C. sativa CBDAS* gene was fused in the C terminal to the YFP gene, linked by thrombin cleavage sequence, and a 3xHemagglutinin (HA) tag was introduced in the C terminal of the *YFP* gene. All the sequences were codon optimized for *P. tricornutum* and synthesized by Bio Basic (Markham, Ontario, Canada). These sequences were introduced in the expression cassette containing the 499 bp '*Highly abundant secreted protein 1'* (*HASP)1* promoter[102] and *FcpA* terminator. *N-acetyltransferase* gene (*nat*) conferring resistance to Nourseothricin was used as a selection marker, under *FcpC* promoter and *FcpC* terminator. As a result, the following plasmids were generated; the *CBDAS* complete codon sequence, *HASP1pro:CBDAS:Thr:YFP:3HA* (CBDAS), a truncated *CBDAS* version without 84 bp from the 5', *HASP1pro:Δ84CBDAS:Thr:YFP:3HA* (Δ28aaCBDAS) and the truncated *CBDAS* version with the endogenous HASP1 signal peptide SP, 54 bp in the 5', *HASP1pro:SP54:Δ84CBDAS:Thr:YFP:3HA* (SP:CBDAS). As a control, *Thr:YFP:3HA* sequence (*YFP*) and *SP:Thr:YFP:3HA* sequence with the signal peptide *SP (SP:YFP)*. To enhance protein production, a minimal Kozak sequence was placed directly before each ATG initial codon. A modified version of the pPtGE31 vector[117], harboring *nat* gene cassette, was used as a negative control. All DNA sequences and primers used in this study are listed in Supporting Information Table S1 and S2 respectively.

Microbial Strains, Growth Conditions and Transformation

Escherichia coli (NEB® 10-beta, New England Biolabs, Canada) were grown in Luria Broth (LB) supplemented with appropriate antibiotics (chloramphenicol, 30 mg/L). Plasmids were extracted using a miniprep kit allowing the extraction of large vectors (Biobasic EZ10 miniprep kit, NY, USA), sequenced by CCIB DNA Core (Massachusetts General Hospital, United States of America) through Next-Generation sequencing platform and then amplified in Epi 300 strain containing pTA-MOB plasmid to allow conjugation with diatoms, as described in the literature[117]. Briefly, 1 mL of wild-type *P. tricornutum* was seeded on 0.5X L1, 1% agar plates and grown at 18 °C on a light/dark cycle of 16/8h for 4 days. Before transformation, 1 mL of L1 media was added to each agar plate, cells were scraped and recovered by pipetting in a sterile tube. Cell concentration was then adjusted to 5.0×10^8 cells/mL. A volume of 25 mL *E. coli* culture containing the assembled plasmid and pTA-MOB was grown at 37°C under agitation to OD600 of 0.8, then centrifuged at 3000 x g for 10 min and resuspended in 250 µL of SOC media. Conjugation was initiated by adding 200 µL of *P. tricornutum* to 200 µL of *E. coli* cells. The cell mixture was plated on 0.5X L1, 5% LB, ~1% agar plates, incubated at 30°C for 90 min in the dark, and transferred to 18 °C in the light and grown for 2 days. After the recovery period, 1 mL of L1 media was added to the plates to collect cells by scraping. Then, cells were plated on 0.5 X L1, 1% agar plates supplemented with nourseothricin 200 mg/L for selection and incubated at 18 ° C. Transformed colonies appeared after 2 weeks.

Strains Selection

Thirty-six clones per construct were plated on new selective media supplemented with nourseothricin 200 mg/L. After 10 days colonies were observed with a magnification of 80 to 120x and screened by fluorescence under a Fluorescent Stereo Microscope Leica M165 FC with GFP filter. Then, 200 µL cell culture of the selected strains was analyzed in the Synergy H1 BioTek microplate reader to measure the YFP fluorescence at Ex/Em wavelengths of 500/539 nm (n = 3) and OD680nm. EV and YFP strains were used as negative and positive controls, respectively. Strains with fluorescence higher than the mean ratio ($[Ex_{500}/Em_{539}]/OD_{680}$) + 2 x [SD] of EV were determined as YFP positive and analyzed with a CytoFLEX S flow cytometer (Beckman) equipped with violet (405 nm), blue (488 nm), yellow-green (561 nm) and red (638 nm) lasers. In this case,

75 µL of 10-day-old cultures were filtered and transferred to clear 96-well plates with 200 µL of L1 in each well. Chlorophyll autofluorescence was detected in the PerCP channel (690/50 nm), while YFP fluorescence was detected in the FITC channel (525/40 nm). Figures and statistics were analyzed using BD FlowJo version 10 software (BD Biosciences, La Jolla, CA, USA, 2020).

Protein extraction

Fifty mL of bioengineered strains were grown for 21 days in L1 media containing nourseothricin (200 mg/L) then centrifuged at 3,500 x g for 15 min at 4ºC. The supernatant was kept and concentrated 200x using Amicon Ultra-15 30K Centrifugal Filter Units from Sigma (cat. #UFC903008). Pellets were weighed and protein extraction was performed according to Fantino *et al*.[89]. Concentrated supernatants containing soluble protein fraction were kept at -20°C for further use in western blot. Total protein samples were quantified using RC DC™ Protein Assay Kit I (Bio-Rad cat # 5000121).

Western blot

For each clone, 60 µg of total proteins were loaded in 10% SDS- PAGE. Transfer settings: 80 volts until proteins went through the stacking gel, then at 120 volts for 2 hours were used. Primary antibodies were incubated overnight at 4ºC. Primary anti-GFP/YFP/CFP and anti-HA from Cedarlane (Ontario L7L 5R2 Canada, cat. #CLH106AP) and ThermoFisher Scientific (Illinois 61101 USA, cat. #MA1-21316) respectively. Both were used at a 1:1000 dilution in 3% BSA. After three washes with TBST solution, the blots were incubated for 1 hour in a 1:20000 dilution, of Immun-Star Goat Anti-Mouse (GAM)-HRP Conjugate from Bio-Rad (Ontario L5T 1C9 Canada, cat. #1705047) in 5% milk. Purified GFP (10 ng) and Multiple Tag (10 ng), from GenScript (cat. #M0101) were used as a positive control for YFP and HA Tag. After three washes with TBST solution, protein detection was realized by using Clarity Max Western ECL Substrate-Luminol solution from Bio-Rad (cat #1705062S). Chemiluminescence detection and Red Ponceau stained (Glacial Acetic Acid 5% v/v, Ponceau Red dye 0.1% m/v) of the blots were visualized using ChemiDoc Imaging System with Image Lab Touch Software (Bio-Rad cat # 12003153) and Image Lab™ Software (Bio-Rad cat # 1709690).

Confocal Microscopy

Live cell Images were captured, and protein localization was visualized with a Leica TCS SP8 confocal laser scanning microscope (Leica Microsystems) with a 40×/1.30 oil immersion objective. The YFP excitation wavelength used was 488 nm and the emission of fluorescence signals was detected from 500 to 525 nm. Chlorophyll auto-fluorescence was observed with an excitation wavelength of 552 nm and the emission of fluorescence signals was detected from 630 to 670 nm. The combined images were generated using Leica Las X software.

CBDAS in vitro enzymatic assay

Enzymatic assays were carried out using 500 µg of total protein extraction, from 3 biologically independent samples, in a 200 µL volume reaction. The reaction was adapted from *Valliere et al.* [118], composed of (pH 4.5), 25 mM $MgCl_2$, 25 mM KCl, 250 µM FAD, and 300 uM CBGA, and incubated ON at 30°C. The reactions were then extracted with 3 volumes of methanol. Metabolite extracts were filtered (0.2 µm PTFE, Agilent Technologies, cat. no. 5190-5265), then dried in a SpeedVac concentrator and resuspended in 100 µL methanol, HPLC grade, and stored at -20°C for high-performance liquid chromatography coupled with diode-array detection (HPLC-DAD) analyses. Metabolite detection was confirmed by HPLC coupled with tandem mass spectrometry (MS/MS).

Cannabinoids and precursors standards

Olivetolic acid (OA, CAS 491-72-5) and olivetol (OL, CAS 500-66-3) were purchased from Santa Cruz Biotechnologies (Dallas United States). Delta-9-tetrahydrocannabinol (THC, CAS 1972-08-3), cannabidiol (CBD, CAS 13956-29-1), cannabinol (CBN, CAS 521-35-7), Δ-9-tetrahydrocannabinolic acid (THCA, CAS 23978-85-0), cannabidiolic acid (CBDA, CAS 1244-58-2), cannabigerolic acid (CBGA, CAS 25555-57-1), cannabichromene (CBC, CAS 20675-51-8), cannabigerol (CBG, CAS 25654-31-3), tetrahydrocannabivarin (THCV, CAS 31262-37-0) and cannabidivarin (CBDV, CAS 24274-48-4) were purchased from Agilent Technologies (QC, Canada). Cannabinolic acid (CBNA, CAS 2808-39-1) was purchased from Sigma-Aldrich (ON, Canada).

HPLC-DAD and HPLC-MS/MS analysis

Analyses were conducted using high-performance liquid chromatography (HPLC) with diode-array detection (DAD). Chromatographic separation of analytes was performed using an InfinityLab Poroshell 120 EC-C18 column (4.6 × 100 mm, 2.7 mm; Agilent Technologies, QC, Canada) maintained at 30°C. Ten microliters of the sample were injected.

into the analytical device. Mobile phases used during analysis were made of (A) formic acid 0.1 % v/v in milli-Q water and (B) formic acid 0.1 % v/v in methanol with a flow rate of 1 mL/min. The HPLC gradient program was set as follows: 0 min, 70 % B; 1.0 min, 70 % B; 6.0 min, 77 % B; 15.0 min, 90 % B; 15.1 min, 70 % B and 18.0 min, 70 % B. The total run time per sample was 18.5 min to allow the reconditioning of the column before the next injection. The diode array detector was set to acquire the wavelength range of 190 to 400 nm with a deuterium (D2) lamp. All the analyses were done using a UV wavelength of 220 nm. Compounds were identified by comparing retention time and maximum absorption wavelengths obtained with the ones of reference standards (Table S3). Standard calibration curves were prepared as follows to allow absolute quantification; two working solutions were prepared containing CBGA and CBDA at 10 mg/L and 100 mg/L each in HPLC grade methanol. These solutions were further diluted to prepare calibration solutions with the following concentrations in triplicate: 0.5, 1, 2, 4, 5, 10, 25, 50, and 100 mg/L. These standard solutions were injected into the HPLC-DAD system and used to generate calibration curve regressions. Fig. S1 shows the calibration curves obtained by plotting the area under the curve obtained as a function of the analyte's concentration. This allowed CBGA and CBDA quantification for the enzymatic and supplementation assays. Confirmatory analyses were performed using high-performance liquid chromatography (HPLC) coupled with tandem mass spectrometry (MS/MS) (Agilent, QC, Canada). This system has an Agilent Jet Stream ionization source, a binary pump, an autosampler, and a column compartment. Compound separation was achieved using an InfinityLab Poroshell 120 EC-C18 column (4.6 × 100 mm, 2.7 mm; Agilent Technologies, QC, Canada). *In vitro* and supplementation assays samples were centrifuged for 10 min at 12000 rpm and diluted 10-fold in the mobile phase (i.e., formic acid 0.1 % v/v in milli-Q water and formic acid 0.1 % v/v in methanol (30:70)). Five µL of each sample were injected onto the column that was set at 50°C. A gradient method made of (A) formic acid 0.1 % v/v in milli-Q water and (B) formic

acid 0.1 % v/v in methanol with a flow rate of 0.5 mL/min was used to achieve chromatographic separation. The HPLC elution program was as follows: 0 min, 70 % B; 7.0 min, 100 % B; 10 min, 100 % B; 12.0 min, 70 % B. The total run time was 14 min per sample. The parameters used in the MS/MS source were set as follows: gas flow rate 8 L/min, gas temperature 220°C, nebulizer 55 psi, sheath gas flow 12 L/min, sheat gas temperature 380°C, capillary voltage 4500 V and nozzle voltage 0 V. Agilent MassHunter Data Acquisition (version 1.2) and MassHunter Qualitative Analysis (version 10.0) softwares were used for data acquisition and processing respectively. Samples analyses were carried out in triggered multiple reaction monitoring (tMRM) acquisition mode allowing compound identification using authentic standards. Table S4 shows MRM transitions and MS/MS parameters used for targeted compound identification.

2.10 Statistics and reproducibility

General data analysis (means and standard deviation) was performed primarily by GraphPad Prism 10.0.2. All experiments were performed with three biological replicates and values were expressed as means ± standard errors.

2.11 Conflicts of interest

The authors declare no conflict of interest.

2.12 Acknowledgments

Special thanks to Professor Hugo Germain and MSc. Mélodie B. Plourde from the Université du Québec à Trois -Rivières for providing us with the purified GFP to use as a control for immunoblotting of Fig. 4. We acknowledge that financial support for this research was funded by the Natural Sciences and Engineering Research Council of Canada through the Alliance program Award No ALLRP 570476-2021 to IDP. Additional support in the form of scholarships to EF, AM, FA, NS, K-LR, and AC from Mitacs-Acceleration/Globalink program grants no IT12310, IT16463, and IT19432 to IDP is also acknowledged.

2.13 CRediT authorship contribution statement

Elisa Fantino & Anis Messaabi: Conceptualization, Data curation, Formal analysis, Investigation, Laboratory work, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing; Natacha Mérindol: Data curation, Project administration, Visualization, Writing – review & editing; Fatima Awwad: Formal analysis, Writing – original draft, Writing – review & editing; Nicolas Sene: Formal analysis, Writing – original draft, Writing – review & editing; Sarah-Eve Gélinas: Investigation, Formal analysis, methodology, Writing – review & editing; Alexandre Custeau and Kimy-Li Rhéaume: Investigation, Formal analysis; Writing – review & editing; Fatma Meddeb-Mouelhi: Conceptualization, Supervision, Project administration, Writing – review & editing; Isabel Desgagné-Penix: Conceptualization, Funding acquisition, Resources, Supervision, Project administration, Writing – review & editing.

CHAPTER III

This chapter and this entire study was born after curious results, obtained during previous experiments. It pushed us to study, highlight and discuss the presence of sequences potentially encoding thrombin-like proteins, which are involved in recognizing and cleaving the thrombin sequence LVPRGS in *P. tricornutum*. This chapter was written in article format, and published in the **Applied Microbiology and Biotechnology Journal**, and might not be directly related to our first goal, which is to produce CBDA.

Contributions: Anis Messaabi pointed out these curious results, carried out molecular and microbial experiments, protein extraction and analysis, and writing the manuscript. Lea Bohnenblust performed molecular and microbial experiments, she was an intern, supervised by Anis Messaabi. Natacha Merindol performed *in silico* analysis, writing and editing the manuscript. Fatma Meddeb review the manuscript. Isabel Desgagné-Penix supervised the project, provided funding and material for the project, and assisted with the writing and revising of the paper.

3 In Vivo Thrombin Activity in the Diatom *Phaeodactylum tricornutum* : Biotechnological Insights

Anis Messaabi[1], Natacha Merindol[1], Lea Bohnenblust[1], Elisa Fantino[1,2], Fatma Meddeb-Mouelhi[1,2], Isabel Desgagné-Penix[1,2*]

[1]Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada.

[2]Plant Biology Research Group, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada.

***Correspondence:** Isabel.Desgagne-Penix@uqtr.ca

Received 02 July 2024, Revised 24 September 2024, Accepted 01 October 2024, Available online 08 October 2024, Biotechnologically Relevant Enzymes and Proteins, Volume 108, article number 481, (2024). DOI: 10.1007/s00253-024-13322-z

3.1 Abstract

Diatoms are responsible for 20% of global carbon dioxide fixation and have significant potential in various biotechnological and industrial applications. Recently, the pennate diatom *Phaeodactylum tricornutum* has emerged as a prominent platform organism for metabolic engineering and synthetic biology. The availability of its genome sequence has facilitated the development of new bioengineering tools. In this study, we used *in silico* analyses to identify sequences potentially encoding thrombin-like proteins, which are involved in recognizing and cleaving the thrombin sequence LVPRGS in *P. tricornutum*. Protein structure prediction and docking studies indicated a similar active site and ligand positioning compared to characterized human and bovine thrombin. The evidence and efficiency of the cleavage were determined *in vivo* using two fusion-protein constructs that included YFP to measure expression, protein accumulation, and cleavage. Western blot analysis revealed 50-100% cleavage between YFP and N-terminal fusion proteins. Our findings suggest the existence of a novel thrombin-like protease in *P. tricornutum.* This study advances the application of diatoms for the synthesis and production of complex proteins and enhances our understanding of the functional role of these putative thrombin sequences in diatom physiology.

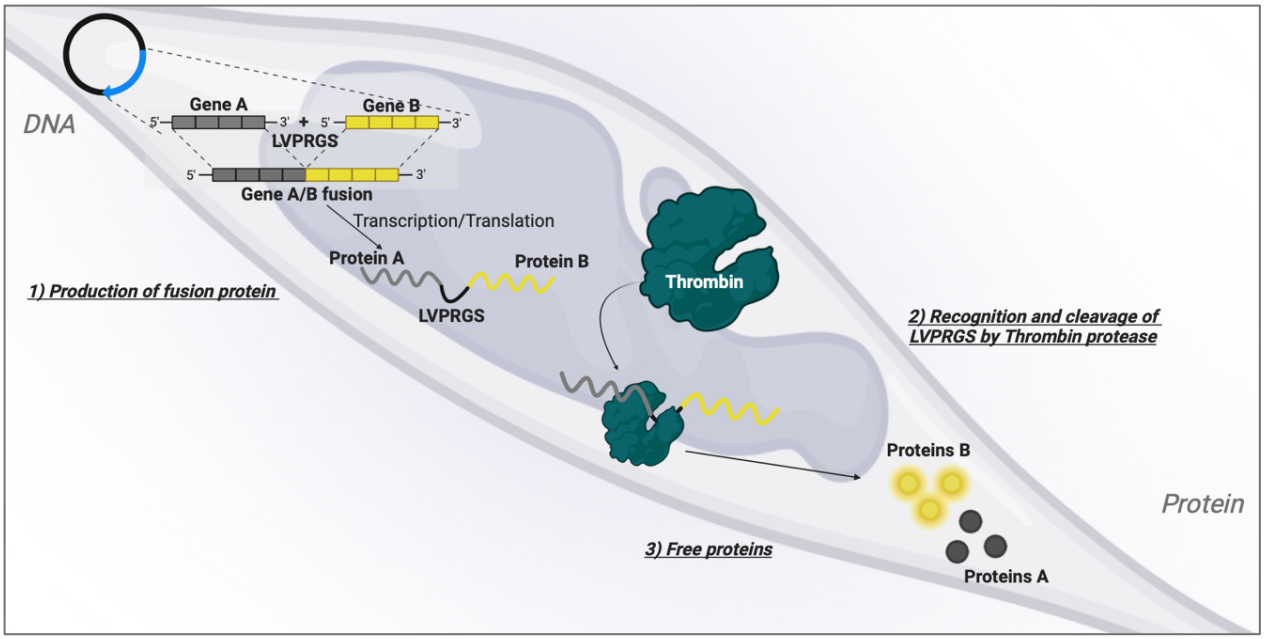
3.2 Keywords

Cleavage peptide, Bioengineering, Synthetic biology, Microalgae, Thrombin-like protease, YFP (Yellow Fluorescent Protein)

3.3 Key points

- Protein structure predictions reveal thrombin-like active sites in *P. tricornutum*.
- Validated cleavage efficiency of thrombin-like protease on fusion proteins *in vivo*.
- Study advances bioengineering tools for diatom-based biotechnological applications.

3.4 Graphical abstract



3.5 Introduction

Synthetic biology relies on efficient heterologous protein production systems, crucial for numerous biotechnology applications involving engineered proteins. Bioengineering combines natural sciences and engineering to address global challenges, such as drug production for disease treatment[119]. Current production platforms include well-characterized hosts like bacteria, yeast, and plants[120]. The pennate diatom *Phaeodactylum tricornutum* is emerging as a promising chassis for producing high-value compounds. It has been successfully utilized to produce heterologous proteins, such as the SARS-CoV-2 spike receptor-binding domain and plant-specialized metabolites such as geraniol and cannabinoids[46, 121-126]. However, the use of this diatom as a bioengineering factory is in its infancy and would benefit from new protein engineering tools.

The production of a metabolite of therapeutical interest often requires the simultaneous expression of multiple genes involved in its biosynthetic pathway. Exogenous episomes delivered to the nucleus via bacterial conjugation have revolutionized metabolic engineering in diatoms[127], enabling the expression of five to six genes from a single episome. Polycistronic

constructs allow the simultaneous expression of several genes like a transgene and a reporter gene. Fusion proteins are invaluable for various biotechnological applications, protein purification[128], expression level monitoring[129, 130], and cellular localization[131]. These fusion proteins have been utilized in vaccine development and are being evaluated in clinical studies[132, 133]. For example, the extracellular domains of two surface proteins from *Streptococcus pneumoniae* have been tested in a mouse model of pneumonia. Similarly, fusion proteins have been used as antigens in vaccines against *Neisseria meningitidis* serogroup B[134, 135].

However, producing recombinant fusion proteins is challenging due to potential issues like misfolding, instability, or poor expression caused by the proximity of fused proteins[136, 137]. Linkers, or spacers, are short amino acid sequences that separate multiple domains in a single protein, improving structural stability and enhancing biological activity[138]. Linkers are classified into three groups: rigid, flexible, and cleavable. Rigid linkers, rich in Proline residues and adopting α-helical conformations, maintain distance between domains to preserve protein stability and activity[139, 140]. Some other rigid linkers are (EAAAK)n[141]. Flexible linkers, composed of small, non-polar (e.g., G) or polar (e.g., S or T) amino acids, allow movement or interaction between proteins[142]. The most widely used flexible linker is (G-G-G-G-S)n[143]. Cleavable linkers, such as those recognized by proteases like thrombin and Factor Xa, are particularly useful in enzyme characterization and allow for tag removal after purification[144]. Thrombin, a trypsin-like serine protease, specifically cleaves peptide bonds on the carboxyl side (P’1) of basic amino acid residues, primarily arginine, making it an essential tool in synthetic biology. Proteases play a crucial role in various biological processes, and their application in synthetic biology and metabolic engineering has been well documented in several systems[145]. The LVPRGS linker, derived from bovine factor XIII and cleaved by thrombin, is frequently used in chemical proteomics workflows and for expressing fusion proteins in *E. coli* without *in vivo* cleavage[104, 146-149].

Despite the advances in using *P. tricornutum* for heterologous protein production, there is limited knowledge regarding the diversity and functionality of endogenous proteases in this diatom. The potential of *P. tricornutum* to harness specific proteolytic activities, such as those resembling

thrombin-like serine proteases, remains unexplored in the literature. Whether the thrombin cleavage sequence can be used in bioengineering strategies of *P. tricornutum* has never been verified. Understanding and leveraging such protease activities could open new avenues for the diatom's application in producing recombinant proteins and associated metabolites.

In this study, we identified candidates of thrombin-like proteases in *P. tricornutum in silico*. We constructed two fusion proteins: mFruits monomeric red fluorescent proteins (mCherry) fused to yellow fluorescent protein (YFP), and cannabidiolic acid synthase (CBDAS) fused to YFP, both separated by the LVPRGS sequence. Both cassettes were expressed under the Highly Abundant Secreted Protein1 promoter (HASP1) and Fucoxanthin chlorophyll a binding protein terminator (FcpA). We introduced the cassettes into *P. tricornutum* by bacterial transkingdom conjugation for extrachromosomal expression. We present evidence of recognition and cleavage of the LVPRGS sequence, confirmed in a bicistronic construct encoding an enzyme fused to the yellow fluorescent protein with the same thrombin cleavage sequence. This study provides a proof of concept that *P. tricornutum* possesses thrombin-like proteases, which could be valuable for projects involving heterologous gene expression in this model diatom.

3.6 Materials and Methods

Plasmid Construction

Plasmids were constructed by Gibson assembly using the NEBuilder® HiFi DNA Assembly Bundle for Large Fragments (New England Biolabs, Canada). Fragments to assemble were amplified by PCR with PrimeSTAR GXL DNA Polymerase (Takara Bio, Japan) following the manufacturer's protocol. Plasmid pPtGE30 was used as a template to construct the expression vectors[150]. Genes encoding mCherry, YFP, and cannabidiol acid synthase (CBDAS) were codon optimized for *P. tricornutum* and synthesized by Bio Basic (Markham, Ontario, Canada). To enhance protein production, a minimal Kozak sequence (ACC) was placed directly before each ATG initial codon. Reporter gene *YFP* was tagged with 3×*HA*. DNA fragments were expressed under an endogenous promoter *'Highly abundant secreted protein 1'* (*HASP1)* and *fucoxanthin chlorophyll a/c binding protein A (FcpA)* terminator[151] or under the *40S ribosomal protein S8* promoter (40SRPS8) and

fucoxanthin chlorophyll a/c binding protein A (FcpA) terminator. *N-acetyl transferase* gene (*NAT*) conferring resistance to nourseothricin (NTC) was included as a selection marker under *fucoxanthin chlorophyll a/c binding protein C* promoter and terminator (*FcpC)* sequences in all constructs. Construct named here *C-T-YFP* includes *HASP1pro* and *CBDAS* is fused to *YFP* with a thrombin cleavage sequence (LVPRGS). Construct *mCherry-T-YFP* includes *HASP1pro* and m*Cherry* is fused to the reporter *YFP* gene with a thrombin cleavage sequence. Construct *mCherry & YFP* includes *40SRPS8pro* and m*Cherry* with a stop codon and a separated expression cassette with *HASP1pro* controlling the expression of *YFP*. A modified version of pPtGE30 vector, harboring *NAT* gene cassette instead of *Sh ble* gene, conferring zeocin resistance, was used as negative control (EV). All DNA sequences, oligonucleotides, and plasmids used to generate these constructs are listed in supplementary data (Table S1, S2, and S3).

Microbial Strains, Growth Conditions and Transformation

Escherichia coli (NEB® 10-beta, New England Biolabs, Canada) was grown in Luria Broth (LB, ThermoFisher, Canada) supplemented with appropriate antibiotics (chloramphenicol, 30 mg.L^{-1}, ThermoFisher, Canada). Plasmids DNA constructs were extracted using a miniprep kit allowing the extraction of large vectors (Biobasic EZ10 miniprep kit, NY, USA), and the integrity of these plasmids DNA was verified by next-generation sequencing in CCIB DNA Core (Massachusetts General Hospital, United States of America). Plasmid constructs were transformed to *E. coli* Epi 300 strain containing pTA-MOB plasmid to allow conjugation with diatoms, as described previously[127]. Epi 300 strain was grown in LB supplemented with chloramphenicol (30 mg. L^{-1}) and gentamicin (40 mg. L^{-1}, ThermoFisher, Canada). Microalgae *P. tricornutum* (Culture Collection of Algae and Protozoa CCAP 1055/1) was grown in L1 media, as described previously(Karas et al. 2015) at 18°C under white fluorescent lights (75 µE m^{-2} s^{-1}) and a photoperiod of 16 h light:8 h dark with an agitation of 130 rpm for liquid cultures. *P. tricornutum* conjugation was performed as in[122]. Transconjugant colonies appeared after 2 weeks of incubation where 36 colonies per construct were plated on new selective media supplemented with nourseothricin (200 µg mL^{-1}). After 7 days, transconjugants were transferred from the agar plates to 96-well plates with to 200

µL of L1 supplemented with 200 µg mL^{-1} in each well and incubated at 18ºC under gentle agitation. After 14 days of liquid growth, putative transconjugants were screened for fluorescence (YFP, l_{ex}/λ_{em}= 513/527 nm and mCherry, l_{ex}/λ_{em}= 587/610 nm) using a plate reader (Synergy H1 BioTek microplate reader (Agilent, Santa Clara, CA, USA)). The selected nourseothricin-resistant colonies were transferred to a 25 mL L1 liquid medium containing nourseothricin in 125 mL flasks for kinetic growth, measurement of the OD at 680 nm, and analysis.

Protein extraction

Bioengineered 10-day-old *P. tricornutum* liquid cultures of 25 mL were centrifuged at 3,500 × g for 15 minutes at 4ºC. Pellets were weighed and resuspended in lysis buffer (50 mM Tris pH 7.4, 500 mM NaCl, 0.1% Tween20, 1 × protease inhibitor cocktail) and kept for 20 min on ice. Sonication was performed 6 times at 35% amplitude, 30 sec on, 30 sec off (3 min total). Cell debris was removed by centrifugation at 12,000 × g for 60 min at 4°C. Total protein samples were quantified using RC DC™ Protein Assay Kit I (Bio-Rad, Hercules, CA, USA). Protein extracts were stored at −80°C until further use.

Western blot

For each clone, a quantity of 60 µg of total protein extract was loaded in a 10% SDS-PAGE. Protein separation was performed at 80 volts until passage through to the stacking gel, followed by 120 volts for the rest of the run. Wet electroblotting was performed at 120 volts was performed for 2 hours to transfer the separated proteins to a 0.2 µm PVDF membrane using the Bio-Rad Mini-PROTEAN Trans-Blot Turbo Transfer System. The blot was equilibrated in 1 × TBS solution and blocked with 5% milk in 1 × TBS for 1 hour at room temperature followed by three washes. Primary antibodies were added for an overnight incubation at 4ºC. Primary anti-YFP (cat no# CLH106AP), anti-HA (cat no# A01244-100), and anti-mCherry (cat no# MA5-32977) were purchased from Cedarlane (Canada) and ThermoFisher Scientific (USA), respectively, and used at a 1:1000 dilution in 3% BSA. After three washes with TBST solution, the blots were incubated for 1 hour with a secondary antibody at a 1:20000 dilution of Immun-Star Goat Anti-Mouse (GAM)-HRP conjugate from Bio-Rad (Canada), in 5% milk. A quantity of 10 ng of multiple Tag (GenScript, USA) and 10 ng of purified recombinant GFP protein (10 ng) from *E. coli* were used as a positive control for HA Tag

and YFP detection, respectively[152]. Two confirmed clones were used as extra controls for the detection of CBDAS. One of them expressing YFP-tagged with a single HA-tag (YFP-1×HAtag) and the other one expressing YFP-tagged with three HA-tag (YFP-3×HAtag). After three washes with TBST solution, protein detection was performed by using Clarity Max Western ECL Substrate-Luminol solution from Bio-Rad. Chemiluminescence detection and Ponceau red staining (Glacial acetic acid 5% v/v, Ponceau red dye 0.1% m/v) of the blots were visualized using the ChemiDoc Imaging System with Image Lab Touch Software (Bio-Rad) and Image Lab™ Software (Bio-Rad).

Flow cytometry (FC)

At least 10 000 events per transconjugant strain were analyzed on a CytoFLEX S flow cytometer (Beckman) equipped with violet (405 nm), blue (488 nm), yellow/green (561 nm) and red (638 nm) lasers. A volume of 200 µL from each culture was filtered and transferred to a clear 96-well plate as described in[122]. Autofluorescence was detected and gated using the PerCP channel (690/50 nm). Cells with high non-specific autofluorescence were excluded based on emission in the PB450 channel (450/45 nm). The fluorescence of YFP and mCherry was analyzed in the FITC channel (525/40 nm) and the ECD/mCherry channel (610/20 nm), respectively. Empty vector (with *NAT*) was used as a negative control to set the gates.

Bioinformatics analysis and phylogenetic tree

To identify *P. tricornutum* thrombin-like proteins, the amino acid sequences of bovine (OX=9913), human (OX=9606), mouse (OX=10090), pig (OX=9823), orangutan (OX=9601) and rat (OX=10116) prothrombins from UniProtKB/Swiss-Prot were blast searched in the *P. tricornutum* protein database version 3. The selected amino acid coding sequences were analyzed using different website applications: PROSITE[153], PEST-find[154], UniProt, ClustalO and Unipro UGENE[155]. Solvent accessibility and secondary structure analysis of putative *P. tricornutum* thrombin-like candidates were performed using Predict protein (https://www.predictprotein.org/). Phylogenetic analyses were carried out after multiple amino acid sequence alignments of 55 prothrombin candidates (Table S4) and 100 non-thrombin serine proteases sequences (Fig. S2) using MEGA11 software[156]. Phylogeny was inferred using the neighbor-joining method and the

evolutionary distances were computed using the maximum composite likelihood method for both trees in this study.

Protein structure prediction and docking study

The models of the 7 putative thrombins of *P. tricornutum* were predicted using ColabFold v1.5.3 (AlphaFold2 using MMseqs2)[157]**.** Structures were visualized and overlaid using Pymol (Shrödinger). MOE2022.09 software (Chemical Computing Group) was used to analyze the conformation of resulting models, and prepare the receptors for docking, as described in[158]. The pocket residues detected by *Site Finder* were superimposed on the crystal structure of *H. sapiens* α-thrombin (PDB:1PPB and 1NRS[159, 160] in MOE and the active site was predicted based on residues interacting with ligands from orthologous crystal structures. Receptors were then prepared using structure preparation, correcting issues, capping, charging termini, selecting appropriate alternatives, and calculating optimal hydrogen positions and charges using Protonate 3D. Structures were energy minimized with 1PPB ligand (D-Phe-Pro-Arg chloromethylketone) inside the pocket.

The folding of the thrombin cleavage sequence LVPRGS was predicted using ColabFold v1.5.3 (AlphaFold2 using MMseqs2) and further prepared in MOE. Two predicted protomers, at pH=7, were included as possible ligands. The MMFF94 × force field was applied. Triangle Matcher was used as the placement method for 200 poses and tethered induced fit as the refinement to perform flexible docking, yielding 10 best poses. Docking poses were analyzed by comparison with crystallized human α-thrombin ligand complex (1PPB with $_{D}$FPR chloromethylketone, 1NRS with hirugen), and the first coherent pose (with the best docking score) consistent with the catalysis of cleavage of LVPR↓GS was selected for each ligand. H-bonds were predicted, and images were further processed using PyMOL (Shrödinger).

3.7 Results

3.7.1 In silico evidence for thrombin-like function

A comprehensive analysis of the *P. tricornutum* proteome using characterized sequences as baits identified seven potential amino acid sequences with features of prothrombin and thrombin-like

proteins (Table S4). Interestingly, over the Bacillariophyta phylum (taxid:2836), 113 hits were obtained using these sequences as baits (Table S5), including 35 in the order Bacillariales, 31 in Thallassiosirales, 14 in Chaetocerotales, 24 in Naviculales, and 9 in Fragilariales[161], suggesting that orthologs of thrombin-like serine proteases are present in diverse diatom species. The seven *P. tricornutum* candidate sequences were aligned with prothrombin and thrombin-like sequences from human, cow, mouse, pig, orangutan, rat, salmon, and snake species. Phylogenetic analysis revealed that the selected *P. tricornutum* sequences formed a distinct cluster (Fig. S1), with only one of the *P. tricornutum* sequences (Phatr3 J45961) clustering closely with the complete and partial prothrombin sequences of human, cow, pig, orangutan, mouse, rat, salmon. Further phylogenetic analysis, which included the candidates, thrombin baits and non-thrombin serine proteases, revealed that even though *P. tricornutum* sequences cluster with mammalian prothrombins, the node bootstrap values were low, and insufficient to support their characterization as thrombin serine proteases (Fig. S2). *Homo sapiens* and *Bos taurus* sequences were selected for a detailed comparison with the *P. tricornutum* thrombin candidates as they are well characterized due to their pharmacological interest[162, 163]. The identity and similarity between *P. tricornutum* candidates and human thrombin ranged from 25 to 32%, and 39 to 47%, respectively. Phatr3 J49772 and J54319 exhibited the highest identity (32% and 30%) and similarity (47% and 43%) to the human sequence. Protein domain predictions using the PROSITE database confirmed that all *P. tricornutum* candidates contained a trypsin domain with the catalytic triad (Histidine (H), aspartate (D), and serine (S)), similar to human and bovine thrombin sequences (Fig. 3-1).

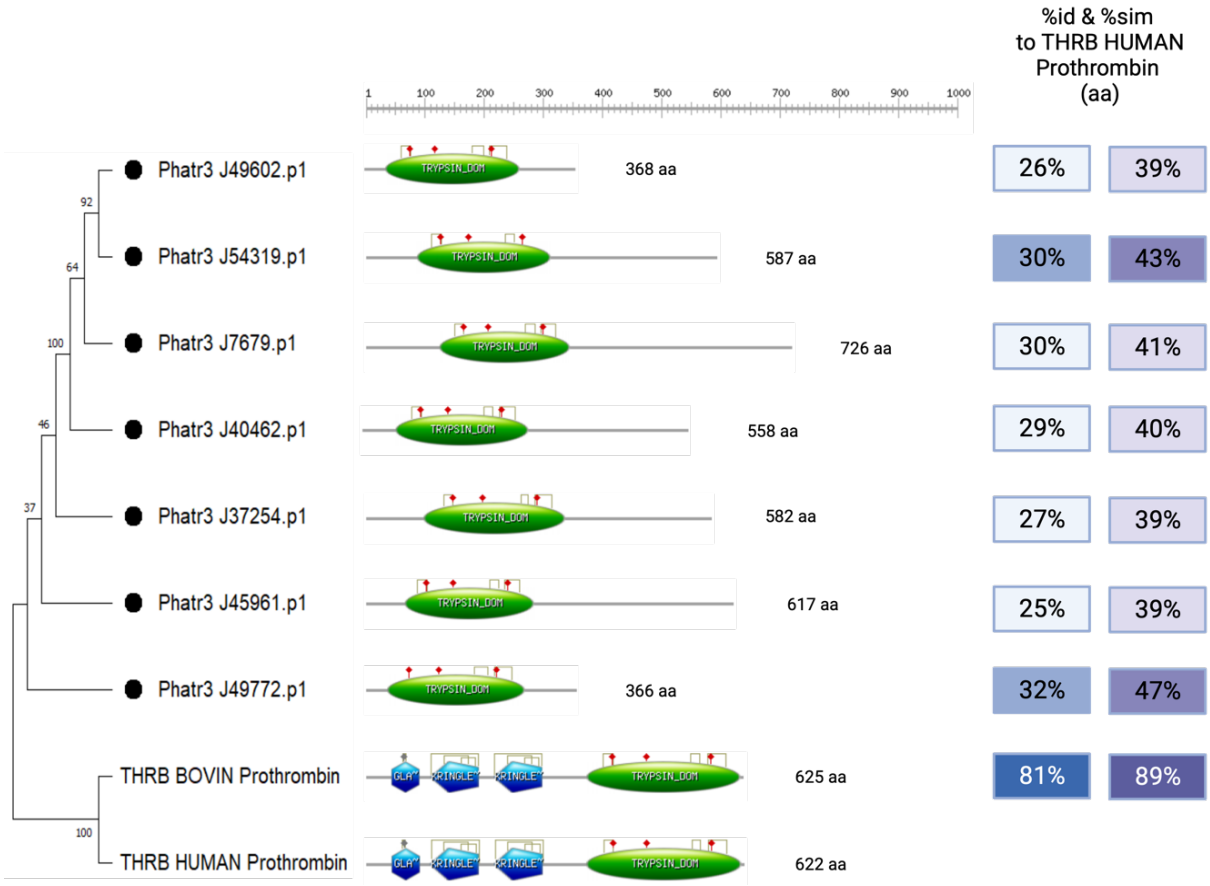


Figure 3-1: : ***A phylogenetic tree of seven potential serine protease thrombin-like sequences in P. tricornutum. Lengths, domains, and percentages of identity (% id) and similarity (% sim) to the Human Thrombin sequence (amino acids) are shown.***

Potential Thrombin-like sequences contain a Trypsin domain (TRYPSIN_DOM) shown in green like the Human and Bovine thrombin sequence. Active sites are represented in red arrow and disulfide bonds in grey. Gamma-carboxyglutamic acid-rich (Gla) and Kringle domains are represented in blue.

An alignment of the trypsin domains showed further similarity, with 26 strictly conserved residues including the catalytic triad (Fig. 3-2). The gamma-carboxyglutamic acid-rich (Gla) domain and the two Kringle domains of human and bovine sequences were absent in *P. tricornutum* candidates. The Gla domain is known to be a membrane-binding motif that interacts with phospholipid membranes in the presence of calcium ions[164]. Kringles are disulfide cross-linked domains found in some serine proteases[165]. They play a role in the regulation of proteolytic activity, but also in binding to membranes and other proteins[166]. Their absence in *P. tricornutum* may be due to their function in blood coagulation by platelet activation, which are not relevant to diatom.

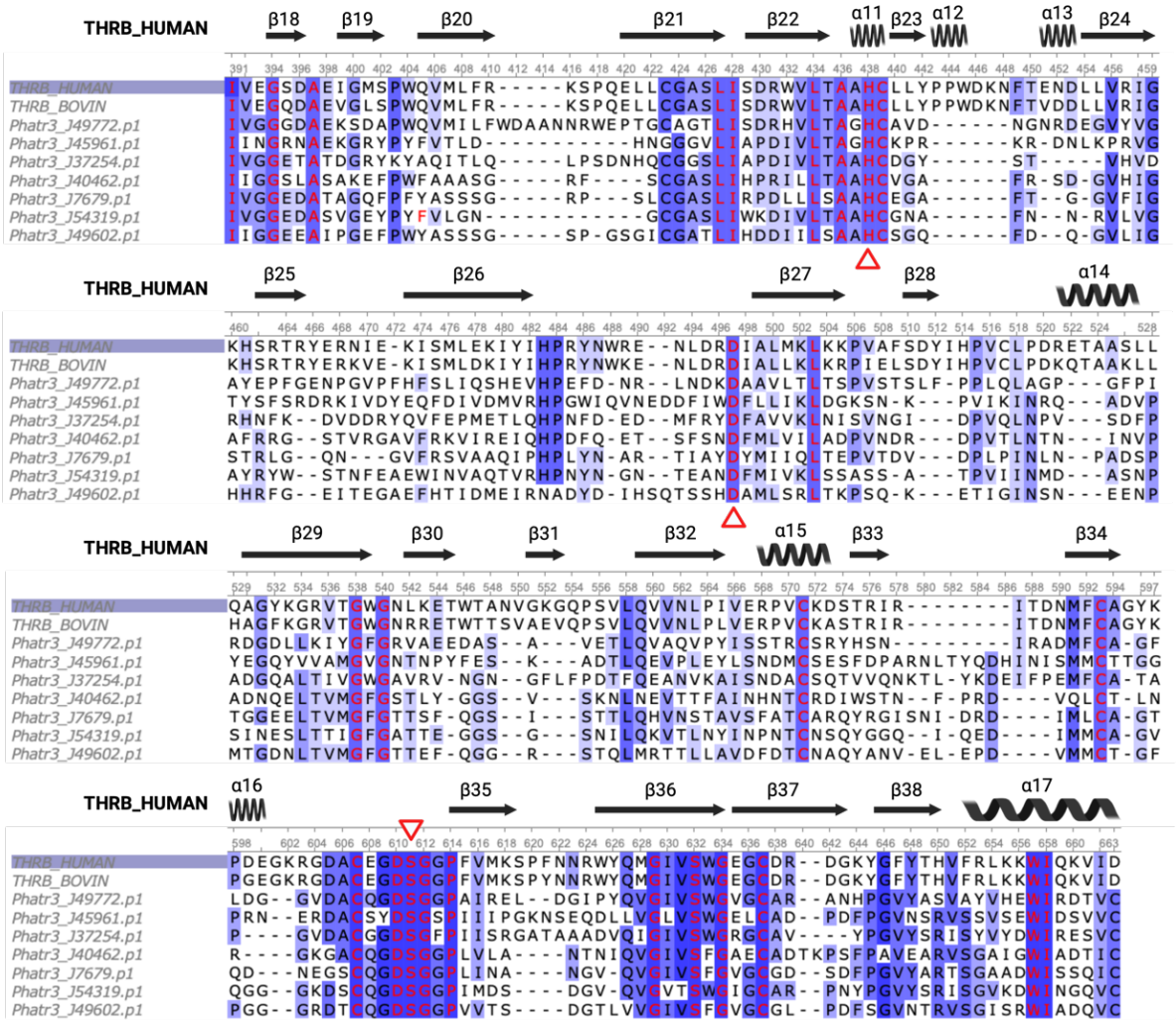


Figure 3-2: ***Amino acid sequence alignment of predicted Trypsin domains from P. tricornutum with two known Human (Accession number: P00734) and Bovin (Accession number: P00735) structures.*** *Identical residues are highlighted in blue and strictly conserved residues are highlighted in blue with red characters. Red triangles represent the active sites.*

3.7.2 In silico evidence for thrombin-like function

The folding of the seven putative thrombins with trypsin domain was predicted using ColabFold v1.5.3 (AlphaFold2 using MMseqs2)[157] and compared with crystallized human and bovine thrombin (Fig. S3a and S3b). All predicted models displayed a trypsin-like domain with two interacting six-stranded β-sheet barrel domains surrounded by helical and loop regions (Fig. S3). The active site residues were predicted and are shown in Table S6. Superposition of the predicted *P. tricornutum* trypsin domains with human β-thrombin showed similar overall folding and orientation of the active site residues (Fig. S4). The conserved catalytic triad (H, D, nd S) is located at the rims of the two interacting barrels, as expected. The oxyanion hole, composed of the

catalytic serine and a near-adjacent glycine, which stabilizes the oxyanion intermediate by hydrogen bonding to the oxygen of the P1 residue, was also observed in all structures[167]. Additionally, equivalent structures to the sodium binding site loop, the 180 loop, the autolysis γ-loop, and a covalent disulfide bond between two cysteines connecting the catalytic serine with the oxyanion hole were identified in the diatom candidates, as seen in human thrombin structures, but not in other serine proteases. However, the exosites, and the 60-loop of human β-thrombin were not detected[168]. Outside of the trypsin domain, *P. tricornutum* predicted structures contained β-barrels (Fig. S3 f, g, h, i) or α-helices (Fig. S3 c, d, g, h, i), suggesting that they might be transmembrane proteins with additional functions.

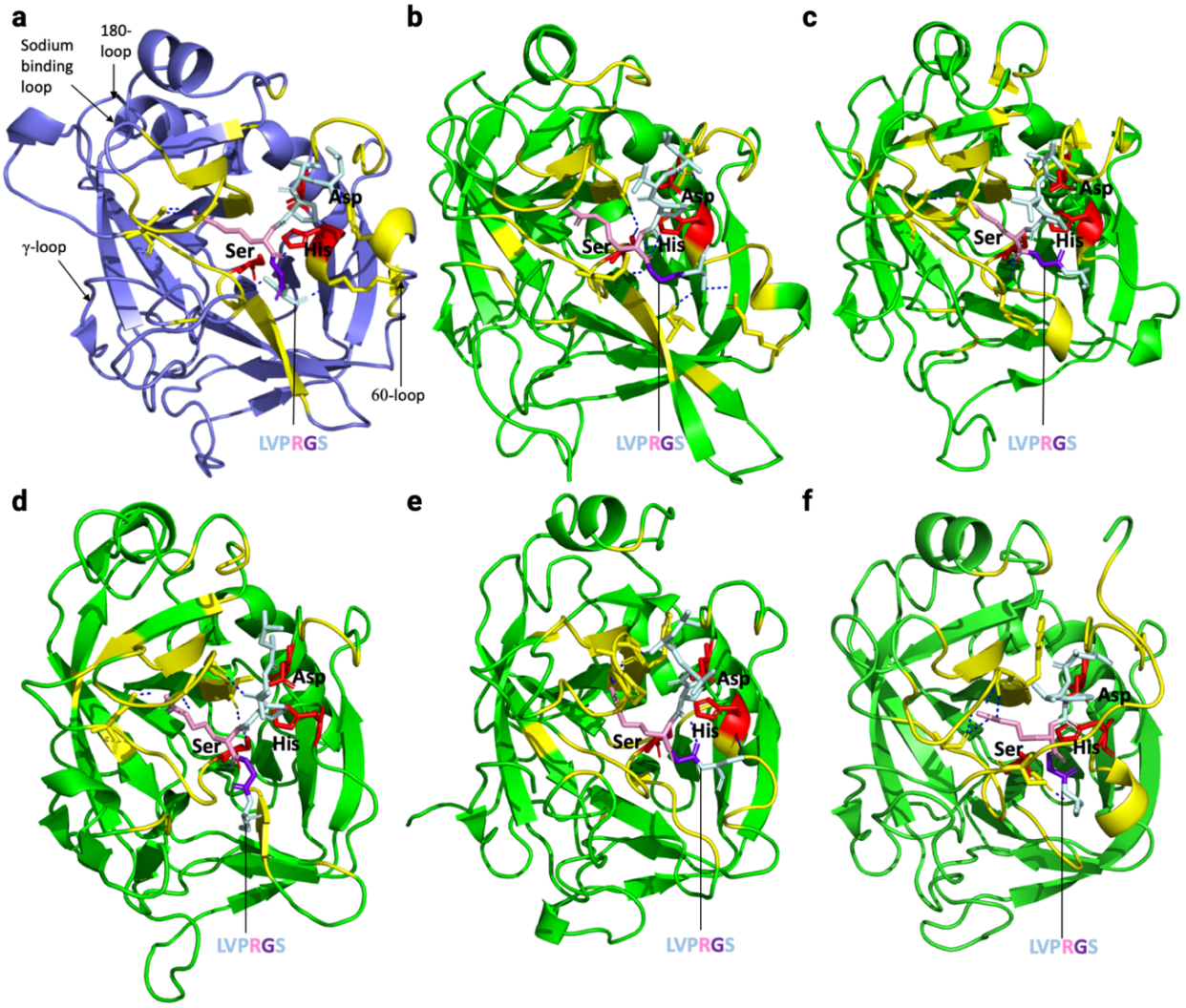


Figure 3-3: ***Cartoon representation of docked thrombin-cleavage peptide in the active site of human (a. 1PPB in purple) and 5 putative thrombins from P. tricornutum:***

(b. Phatr49772, c. Phatr45961, d. Phatr37254, e. Phatr40462, f. Phat54319 in green). The cleavage peptide (LVPRGS) is shown as light blue sticks with pink arginine and violet glycine. The active site is shown in yellow and catalytic residues (His, Asp, and Ser) are displayed as red sticks. Only the trypsin domain of each protein is shown. H-bonds are shown as dashed blue lines.

As the diatom's candidate displayed a predicted trypsin domain that closely matched with the human's, the thrombin cleavage peptide was docked into their active site and the interactions were analyzed (Table S6). LVPRGS docked with a score of −10.87 kcal/mol in human ß-thrombin active site. In the case of Phatr3_J7679 and Phatr3_J49602 candidates, the ligand docked further from the catalytic residues than 1PPB and 1NRS crystallized ligands. Therefore, the docking poses of these structures were rejected. In five of the putative predicted structures, the ligand docked in a position consistent with a cleavage between R and G (Fig. 3-3), with scores ranging from −8.92 to −10.6 kcal/mol (Table S6). Overall, LVPRGS docked very similarly into human and trypsin domains of selected putative *P. tricornutum* thrombin, with the cleavage site positioned in between the catalytic serine and histidine. Several H-bonds were detected between the docked peptide and the active site, including with catalytic residues in the case of 1PPB, J49772, J45961, J37254, and J40462. J54319 showed the highest docking score with a position most similar to the docking configuration in the human protein (Fig. 3-3a to 3f, and Table S6). These results suggest that *P. tricornutum* produces endogenous proteins containing a trypsin domain that could efficiently cleave the thrombin cleavage sequence *in vivo*.

3.7.3 P. tricornutum transformation and screening of transconjugant strains

To analyze potential thrombin activity in *P. tricornutum* cells, we constructed and introduced four expression cassettes into *P. tricornutum* wild type (Fig. 3-4). An empty vector (EV) encoding antibiotic resistance, and a plasmid harboring two genes *(mCherry & YFP*) expressed under two distinct promoters (40SRPS8 and HASP1, respectively) were used as controls (Fig. 3-4b). Additionally, two other plasmids were constructed containing bicistronic cassettes encoding two recombinant proteins (CBDAS or mCherry) fused to YFP by a thrombin cleavable sequence hereafter referred to as C-T-YFP and mCherry-T-YFP (Fig. 3-4a, 4b). Highly Abundant Secreted Protein1 (HASP1p) promoter drove the expression of mCherry-T-YFP and C-T-YFP recombinant proteins. *P. tricornutum* wild type cells grown to stationary phase were transformed by

conjugation[127]. Transconjugants were selected with the appropriate antibiotic and screened for mCherry and YFP fluorescence using a microplate reader and flow cytometer. A total of 108 colonies were screened, with 36 transconjugants per construct. The percentage of positive clones was higher in the mCherry & YFP lines (69% of clones were mCherry^{+}) compared to the mCherry-T-YFP lines (39%) (Table 3-1). Additionally, 25% of the C-T-YFP clones were YFP^{+}. Three to six independent transgenic lines from each construct were selected for further analysis (Table 3-1).

Table 3-1: : ***Transconjugant positive strains screened for mCherry or YFP fluorescence.***

Transconjugants	mCherry^{+} (FC)	YFP^{+} (PR)	No. of selected strains
mCherry & YFP	69% (n=36)	n.t.	6
mCherry-T-YFP	39% (n=36)	n.t.	6
C-T-YFP	n.a.	25% (n=36)	3

n.a.: not applicable, n.t.: not tested. Percentage of mCherry^{+} and YFP^{+}clones, n=36, Fluorescence was detected with a microplate reader (PR) or with a Flow Cytometer (FC).

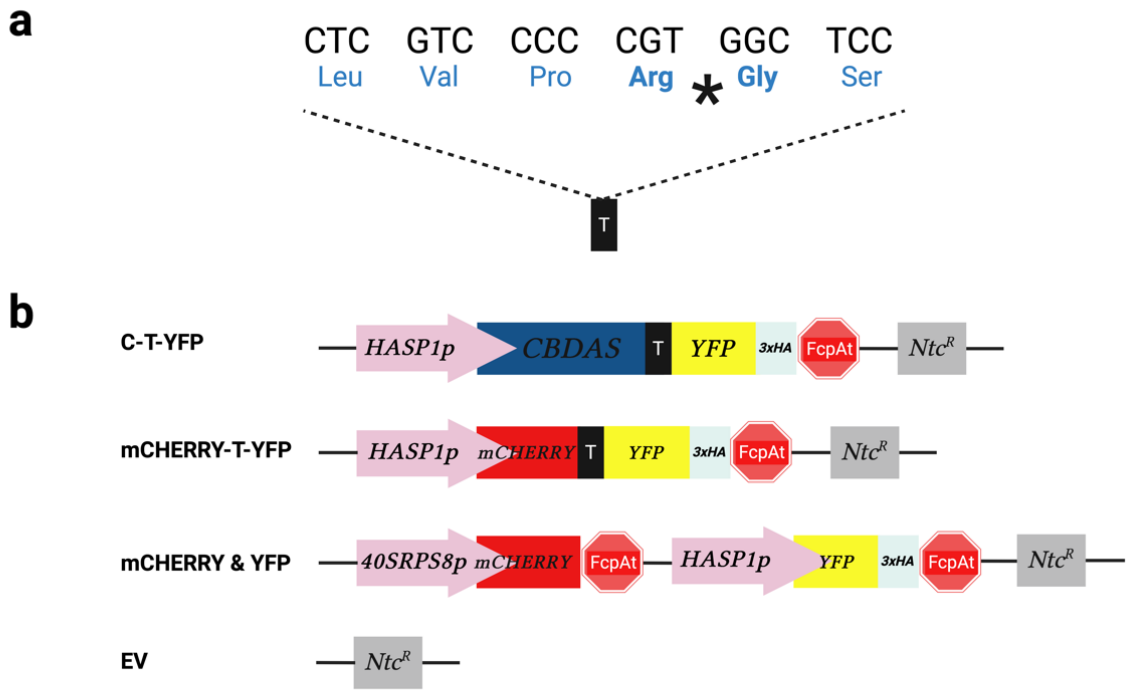


Figure 3-4: ***Expression cassette of recombinant constructs with a thrombin cleavable sequence.***

(a). Thrombin cleavable sequence (LVPRGS) derived from the sequence in bovine factor XIII, cleavage occurs between Arginine (Arg) and Glycine (Gly). (b). Scheme of the expression cassettes, mCherry-Thrombin-YFP-3×HAtag (mCherry-T-YFP) and CBDAS-Thrombin-YFP-3×HAtag (C-T-YFP). The expression of mCherry and CBDAS genes is driven by the endogenous highly abundant secreted protein 1 promoter (HASP1), fused to YFP reporter gene with a thrombin cleavable sequence in between. Construct mCherry & YFP containing mCherry and YFP genes are driven by two different promoters, HASP1 and 40SRPS8, respectively. EV containing only Nourseothricin resistance gene (NtcR) mCherry & YFP were cloned in two different vectors and used as controls.

3.7.4 Different levels of transgene fluorescence in P. tricornutum transconjugants

Transconjugant growth was monitored for 10 days, and all selected transgenic lines exhibited similar results with no significant differences compared to the negative control (EV) (**p*>0.05, one-way ANOVA, Fig. 3-5a, upper panel), indicating that transgenes did not affect the growth of *P. tricornutum* cells. The percentage of mCherry$^+$ and YFP$^+$ cells and relative fluorescence were measured on day 10 by flow cytometry and microplate reader, respectively, and normalized to the autofluorescence of the negative control (EV) (Fig. 3-5a, middle and lower panels). The percentage of mCherry$^+$ cells varied significantly between constructs (**p*<0.05, Student's *t*-test), with a mean of 55% for mCherry & YFP transconjugants compared to 7.5% for mCherry-T-YFP (Fig. 3-5a, middle panel). However, since mCherry was expressed under different promoters in these constructs, the differences in fluorescence may be due to changes in the expression levels of the 40SRPS8 and HASP1 promoters rather than the construction design. Nonetheless, the percentage of YFP$^+$ cells was also significantly higher with mCherry & YFP (**p*<0.05, *t*-test), although the difference was smaller, with a mean of 3.4% for mCherry & YFP transconjugants compared to 1.2% for mCherry-T-YFP transconjugants. C-T-YFP transconjugants did not show a significant difference in YFP$^+$ percentages compared to mCherry & YFP (1.8% of YFP$^+$, **p*>0.05, *t*-test). A similar trend was observed for relative fluorescence measured with the microplate reader (Fig. 3-5a, lower panel). Comparison of mCherry & YFP clones with mCherry-T-YFP clones revealed a significantly higher relative mCherry fluorescence in mCherry & YFP clones (mean of 830 RFU) (**p*<0.05, Student's t-test). In contrast, mCherry-T-YFP clones exhibited a mean relative fluorescence unit (RFU) of 11. Moreover, the relative fluorescence emitted by YFP in mCherry & YFP clones (152 RFU) was significantly higher than in mCherry-T-YFP (29 RFU) (**p*<0.05, Student's *t*-test), while C-T-YFP clones showed no significant changes (mean of 75 RFU). Taken together, these results demonstrate that the fluorescence levels of the transgenes varied, with higher mCherry emission

under the 40SRPS8 promoter than the HASP1 promoter. Furthermore, *YFP* expression, driven by the HASP1 promoter, slightly decreased when fused to the C-terminal of another protein.

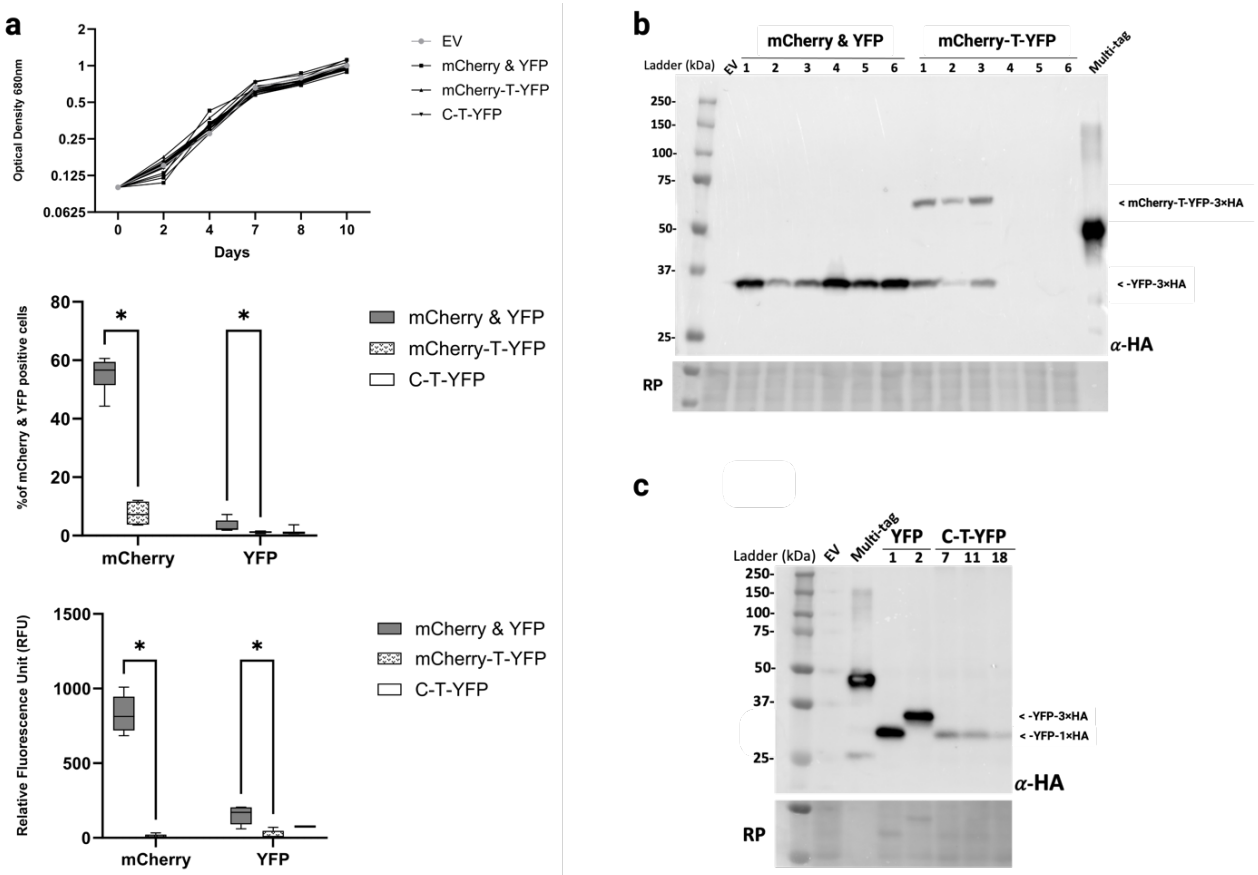


Figure 3-5: ***Characterization of P. tricornutum clones and production of protein with specific cleavage of the thrombin cleavable sequence.***

*a. Upper panel, cell growth curves for each P. tricornutum strain of each construct, the optical density (OD) at 680 nm was monitored for 10 days. Middle panel, the mean percentage of mCherry and YFP positive cells were monitored for each construct by flow cytometer. Lower panel, mean relative fluorescence intensity of mCherry and YFP for each construct measured by plate reader. All strains were grown in L1 culture media supplemented with the appropriate antibiotic and incubated at 18ºC under gentle agitation of 130rpm. b. Western blot analysis with an anti-HA tag antibody of total protein extracts from mCherry & YFP clones (32 kDa), mCherry-T-YFP clones (58 kDa) and Multi-tag (40 kDa). c. Western blot with an anti-HA antibody of total protein extracts from C-T-YFP clones (28,9 kDa), YFP clones used as controls (1: 28,9 kDa and 2: 32 kDa) and Multi-tag (40 kDa). a-HA: anti-Hemagglutinin antibody, RP: Ponceau staining of the membrane is shown as a loading control. n=6 or n=3 number of biologically independent samples for each strain was plotted, t-test (*p<0.05) was performed.*

3.7.5 In vivo evidence of recognition and cleavage of the thrombin cleavable sequence in P. tricornutum

To verify the possible cleavage of the thrombin sequence in *P. tricornutum* transconjugants, we analyzed the transgenes by western blotting using anti-HA antibody (Fig. 3-5b). No bands were

detected in the total protein extract of *P. tricornutum* transconjugants 4, 5 and 6 harboring the *mCherry-T-YFP* cassette consistent with the low fluorescence detected by both flow cytometry and microplate reader. In contrast, two bands corresponding to the cleaved (32 kDa) and uncleaved forms (58 kDa) of mCherry-T-YFP protein from transconjugants 1, 2, and 3 were observed (Fig. 3-5b). The presence of the cleaved protein form was confirmed by using an anti-mCherry antibody (Fig. S5). Cleavage efficiency was determined by relative quantification of band intensity using Image Lab software. In transconjugants (mCherry-T-YFP) 1, 2 and 3, the fusion proteins were partially cleaved with a mean ratio of uncleaved to cleaved forms of approximately 53/47%. Heterogeneity between clones was observed, with transconjugant 1 showing the highest cleaved form (57%) compared to transconjugant 2 and 3, which showed 37% and 46%, respectively. The recognition and cleavage of the thrombin cleavage site *in vivo* by *P. tricornutum* were further confirmed using another construct C-T-YFP, which contained the gene of interest (*C*: *CBDAS*) linked to the *YFP* gene through the thrombin (*T*) sequence (Fig. 3-4b). This construct showed a total cleaved fraction (100%) with a size of 28.9 kDa (Fig. 3-5c). All C-T-YFP transconjugants displayed a corresponding size of YFP with a single HA tag instead of three, explained by the deletion of two HA tags, confirmed by DNA sequencing analysis (Table S7). Additionally, several substitutions were detected in the plasmid DNA sequence introduced into the diatom cells. However, all these substitutions were outside the coding sequence (CDS), and were not expected to affect the results. Altogether, these results confirm the *in vivo* cleavage of the thrombin cleavage site in *P. tricornutum.*

3.8 Discussion

Cloning multiple genes to produce multiple proteins from a single vector has been challenging in the heterologous host *P. tricornutum*. Previous applied studies have developed a multi-gene expression system based on the *Thosea asigna* virus 2A self-cleaving peptide (2A), replacing older constructs reliant on the presence of internal ribosomal entry sites (IRES)[169-171]. The 2A-based multigene expression method relies on ribosomal skipping during translation, though varying cleavage efficiencies among different 2A sequences remain under investigation for future research[149]. Hence, alternative methods are being explored.

In this study, we used several known thrombin and thrombin-like sequences from various species to identify putative thrombin sequences from *P. tricornutum* proteome database. Initially, seven sequences were identified with identities ranging from 25 to 32% and similarities from 39 to 47% compared to human thrombin. All sequences exhibited the characteristic trypsin domain of serine proteases family but the evidence supporting their classification as thrombin-like proteins remains insufficient, based on the current phylogenetic tree analysis. Structural modeling predicted their folding similarity to human and bovine thrombin domains, with conserved catalytic residues (H, D and S). The predicted structures exhibited key features typical of serine proteases, along with specific elements characteristic of thrombin proteases, such as a sodium binding site, an autolysis γ-loop, a 180 loop, and a covalent disulfide bond linking the catalytic serine to the oxyanion hole. The sodium binding site enhances catalytic efficiency, the autolysis γ-loop facilitates substrate and regulator interactions essential for coagulation, the 180 loop is crucial for substrate recognition and specificity, and the disulfide bond stabilizes the active site structure[167, 168]. Docking analysis indicated that the active site conformation was compatible for interaction with the thrombin-cleavage peptide and with potential cleavage between R and G of LVPRGS sequence in five putative *P. tricornutum* thrombin candidates.

While the domains of the *P. tricornutum* candidates displayed structural similarities to characterized thrombin, these findings did not preclude the possibility that these candidates are serine proteases without thrombin-like peptide cleavage activity. To confirm the cleavage capability of the thrombin sequence in *P. tricornutum*, we designed three constructs using two promoters (40SRPS8 and HASP1), with and without the LVPRGS cleavage site between two fluorescent proteins (mCherry and YFP), or between an enzyme (CBDAS) and YFP. Initially, we evaluated fluorescence levels associated with transgene expression. Notably, mCherry fluorescence was brighter under 40SRPS8 promoter (mCherry & YFP) than under HASP1 (mCherry-T-YFP) at day 10, possibly due to higher 40SRPS8 promoter activity in this growth phase rather than protein fusion or cleavage site addition. Although reports suggest detectable HASP1-dependent expression in the exponential phase, our findings indicated specificity to the late stationary phase[151]. Furthermore, we assessed differences in YFP fluorescence (expressed under HASP1 in all constructs). YFP fluorescence levels were slightly lower in mCherry-T-YFP

transconjugants compared to mCherry & YFP transconjugants, suggesting that fusion with mCherry reduced fluorescence intensity. Of note, the sensitivity of the method to measure YFP fluorescence was suboptimal due to significant overlap between YFP's excitation and emission spectra, limiting precise excitation and emission measurements at their peak wavelengths. The flow cytometer is equipped with a 488 nm laser, which excites only 41% of YFP[172], further constrain YFP emission strength.

Western blot analysis revealed that the constructs with the thrombin cleavage sequences in between two genes were cleaved, at a mean ratio of uncleaved/cleaved of ~53/47% for mCherry-T-YFP *P. tricornutum* transconjugants. This finding is consistent with the analysis of YFP fluorescence emission levels in mCherry-T-YFP clones. The fluorescence emitted by the uncleaved fraction may be lower than normal when two fluorescent proteins are fused and separated by a small distance[173]. Interestingly, 100% of CBDAS-T-YFP clones showed cleaved proteins. This homogeneity in CBDAS-T-YFP clones compared to mCherry-T-YFP clones may be due to lower efficiency in detecting larger proteins by Western blot. Further investigations must be done. Further functional studies are required to confirm the thrombin-like activity of these proteins. Also, to elucidate their cleavage efficiency and uncover the elements that could affect the cleavage in *P. tricornutum*.

The physiological role of these thrombin candidates in *P. tricornutum* is unknown. Intriguingly, candidates were also observed in various diatom orders. Mammals' thrombin proteases are involved in the blood clotting process. Aggregation is possible via activation of protease-activated receptors (PARs)[174, 175]. PARs are a subfamily of (GPCRs). These transmembrane proteins are activated by proteolytic cleavage such as performed by thrombin[176]. GPCRs are involved in signal transduction pathways in response to external stimuli in mammalian cells[177]. Interestingly, the availability of complete genome sequence and comparative genomics studies have revealed the presence and the conservation of these receptors in four diatoms, including *P. tricornutum*[178]. While the function of GPCRs in diatoms has not been elucidated, previous findings suggest a role in intracellular signaling pathways in response to environmental changes. For example, Gamma-aminobutyric acid (GABA) in plants plays a role in protecting against abiotic

and biotic stress[179]. In diatoms, it may act in a similar way protecting against salinity or temperature changes[180]. Other GPCRs could also contribute to biofilm formation and surface colonization in *P. tricornutum*[181]. Interestingly, we observed that several thrombin- candidates also encode for transmembrane domains. The presence of transmembrane domains in the predicted thrombin candidates in *P. tricornutum* suggests that they could co-localize with PARs and be involved in intercellular communication.

Future studies should investigate the diatom candidates and uncover their relative efficiency, as well as their specificity for the cleavage peptide to confirm the thrombin-like activity of these proteins. Overall, our results suggest that thrombin cleavage site linkers can be added to the molecular toolkit, allowing the expression of multi-gene cassettes in an episomal system in *P. tricornutum*. In comparison to 2A self-cleavage peptides, the use of thrombin strategy is based on proteolytic activity. Moreover, to enhance thrombin proteolytic activity, kinetic studies must be conducted to determine the best consensus substrate recognition sequence of thrombin in *P. tricornutum*[182]. This tool could be adapted to develop a new approach for assessing simultaneous *in vivo* functional and subcellular localization studies of endogenous or heterologous enzymes linked to a fluorescent protein in *P. tricornutum.*

To conclude, our results highlight for the first time in the microalgae diatom *P. tricornutum* the outcome and the potential of adding a thrombin cleavage sequence for the simultaneous expression of more than one gene under a single promoter. In studies, it will be interesting to unravel the functional role of these putative thrombin sequences in the diatom *P. tricornutum.* Our findings may contribute to the understanding of diatom's biology, which is important for both marine ecology and biotechnology. *P. tricornutum* is a promising bioengineering platform that will benefit from further studies of its endogenous processes.

3.9 Acknowledgments

Authors wish to thank all the staff and students of the lab for their cooperation and kind support. Their contributions have improved the quality of this work.

3.10 Author Contribution Statement

A.M.: Conceptualization, Investigation, Laboratory work, Data curation, Formal analysis, Methodology, Led manuscript drafting and editing; N.M.: Conceptualization, Investigation, Laboratory work, Data curation, Project administration, Writing – review & editing; L.B.: Laboratory work, Formal analysis, Writing – review & editing; E.F.: Laboratory work, Formal analysis, Writing – review & editing; F.M-M.: Project administration, Writing – review & editing; I.D-P.: Conceptualization, Funding acquisition, Resources, Supervision, Project administration, Writing - review & editing, Resources.

3.11 Funding

Authors acknowledge that financial support for this research work was funded by the Canada Research Chair on plant specialized metabolism Award No 950-232164 to I.D-P. Many thanks are extended to the Canadian taxpayers and to the Canadian government for supporting the Canada Research Chairs Program.

3.12 Data availability statement

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files. Should any raw data files be needed in another format they are available from the corresponding author upon reasonable request.

3.13 Ethics approval

This article does not contain any studies with human participants or animals performed by any of the authors.

3.14 Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

4 CHAPTER IV

4.1 Discussion, Conclusion and Perspective

4.1.1 *P. tricornutum* as a heterologous expression system for the biosynthesis of CBDA

P. tricornutum was able to produce two cannabinoid precursors in previous studies from our lab, olivetolic acid and cannabigerolic acid, respectively[88, 126]. The current work presented in this thesis build upon previous studies and successfully demonstrated, for the first time, the feasibility of utilizing the diatom *P. tricornutum* as a heterologous expression system for the biosynthesis of CBDA through the expression of CBDAS. This step is the most critical in the cannabinoid pathway, because of THCA and CBDA cytotoxicity as shown in previous studies[126]. Our enzymatic activity assays confirmed that CBDA production was achieved in *P. tricornutum* using both native and secretory CBDAS constructs, supporting the hypothesis that proper processing through the secretory pathway is necessary for CBDAS activity, as presented in previous studies in heterologous systems and transgenic tobacco[22, 59]. These results are supported by the subcellular localization analysis, indicating that both the native and signal peptide-modified versions of CBDAS clustered in the chloroplastic ER, emphasizing the importance of post-translational modifications such as N-glycosylation, as confirmed for cannabinoid synthases, where seven Asn N-glycosylation sites have been found[60, 61]. Moreover, in a previous study, the structural analyses of *N*-linked glycans demonstrated that *P. tricornutum* proteins carry mainly high mannose type *N*-glycans ranging from Man-5 to Man-9[183]. To conclusively validate these observations in our work, two targeted experimental approaches should be undertaken: a treatment with the fungal toxin Brefeldin A (BFA) to inhibit the secretory pathway by disrupting the Golgi apparatus and inducing intracellular accumulation[184, 185], and an enzymatic digestion with PNGase F (Peptide:N-glycosidase F) or Endo H (Endoglycosidase H), to confirm the presence of N-glycosylation modifications on the expressed proteins[183].

Regarding CBDAS accumulation and secretion, we observed that while the YFP tag was detected in the extracellular medium as observed previously[62], CBDAS and SP:CBDAS were not. This suggests potential degradation during secretion or cleavage from YFP. An alternative explanation

is that a cleaved, tag-free version of CBDAS could be present in the extracellular matrix but remained undetectable. Our second study in chapter 3, suggested the recognition and cleavage of the thrombin cleavage sequence in *P. tricornutum.* Further optimization, including alternative linker sequences and secretion pathway modifications, may enhance extracellular CBDAS detection and accumulation.

The episomal expression approach proved to be an efficient method for generating *P. tricornutum* transconjugants, as it required minimal screening efforts compared to random integration methods[186]. Also, the screening process was greatly facilitated by the use of a fluorescently tagged CBDAS protein. This aligns with findings from previous studies, were reinforcing the reliability of episomal expression in microalgae biotechnology, had previously demonstrated the stability of an episome maintained over four months of cultivation[150]. However, rearrangements and instability in episomes are possible as demonstrated previously in our lab[187, 188]. By introducing the same plasmid into the diatom via bacterial conjugation, two distinct groups of clones emerged: those exhibiting detectable reporter protein fluorescence and those without any visible fluorescence, without detecting the protein. These results corroborate our study, and notably, we achieved detection of CBDAS using Western blot. Finally, the exact mechanism by which the plasmid enters the nucleus of *P. tricornutum* is not fully understood. Following cellular entry, the plasmid must be transported to the nucleus, where it can persist as an extrachromosomal element. it is hypothesized that specific host cell proteins may recognize plasmid-associated signals and mediate its trafficking through the cytoplasm and translocation across the nuclear envelope.

Despite these promising findings, further optimization is required to enhance CBDA production to commercially viable levels. Strategies such as medium optimization (e.g., nitrate, phosphate and iron deprivation), directed mutagenesis, alternative linker sequences, and increasing gene copy number should be explored to improve enzyme efficiency and yield. Additionally, mitigating Cannabinoids cytotoxicity through cell compartmentalization and cannabinoid transporter utilization will be crucial for maximizing production efficiency.

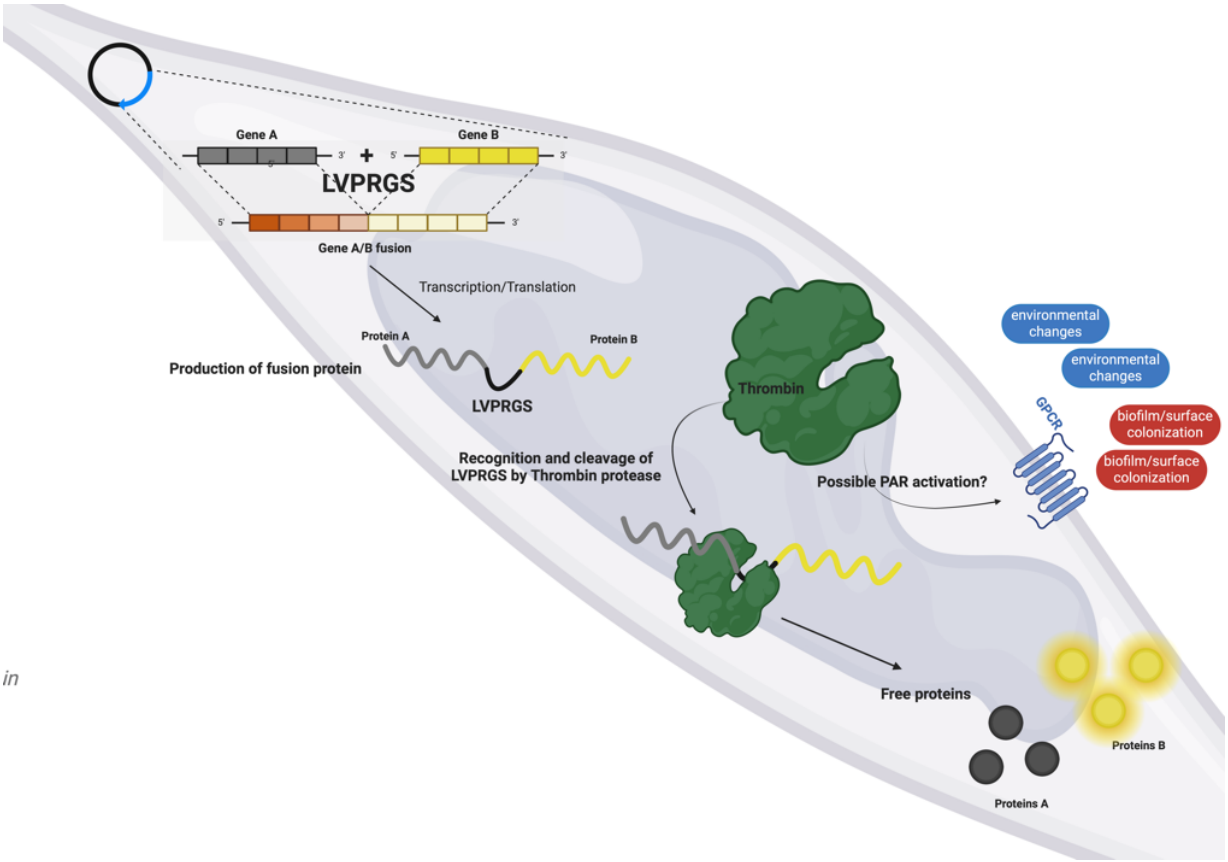
4.1.2 Thrombin cleavage sequence in *P. tricornutum*

This study marks the first demonstration of a thrombin cleavage sequence in *P. tricornutum*. By analyzing the *P. tricornutum* proteome, we identified several putative thrombin-like candidates with sequence and structural similarities to human and bovine thrombin, including conserved catalytic residues and key functional domains. In addition to the catalytic triad, several other conserved residues may contribute to substrate binding. The spatial arrangement and chemical characteristics of these residues are critical in defining interaction specificity, thereby influencing both ligand recognition and binding affinity. Moreover, structural modeling and docking analysis suggested that these candidates may have thrombin-like cleavage capabilities, particularly between the R and G residues in the LVPRGS sequence. However, a substantial portion of these proteins—particularly at the N- and C-terminal regions—appears to be intrinsically disordered, lacking a well-defined globular structure associated with known functional domains. Notably, such terminal regions are often predicted with lower confidence by structural modeling tools like AlphaFold. This observation warrants further discussion to determine whether these segments are genuinely unstructured, potentially contributing to protein–protein interactions or signaling functions, or if the apparent disorder reflects limitations inherent to current predictive modeling approaches.

To assess the functional capability of these sequences, we designed constructs incorporating thrombin cleavage sites between fluorescent proteins and between an enzyme and YFP. Protein analysis confirmed the cleavage of thrombin site-containing constructs, with a cleavage efficiency of approximately 47% and 100%, for two fluorescent proteins and for an enzyme and YFP, respectively. This suggests that thrombin-based cleavage strategies could provide a reliable mechanism for multi-gene expression in *P. tricornutum*, since 2A self-cleaving peptides presented low cleavage efficiencies. Future research should focus on characterizing the efficiency and specificity of these thrombin candidates for cleavage sequences in *P. tricornutum*. Additionally, kinetic studies are necessary to determine the optimal substrate recognition sequence for thrombin activity in diatoms. Expanding the molecular toolkit to include thrombin cleavage

sequences could provide a new strategy for simultaneous in vivo functional and subcellular localization studies of endogenous and heterologous enzymes.

Regarding the thrombin-like proteases physiological role in *P. tricornutum* remains unclear, though their presence in other diatoms suggests they may play a role in cellular signaling. In Mammals, thrombin proteases activate protease-activated receptors, which mediate various biological processes[174, 175]. Interestingly, homologous GPCRs have been identified in diatoms, hinting at their possible role in environmental stress responses, biofilm formation, or intercellular communication[178-181]. Moreover, the presence of transmembrane domains in predicted thrombin candidates strengthens a potential functional link to these signaling pathways. However, their classification as true thrombin-like proteases requires further experimental validation.



in

Figure 4-1: ***A schematic hypothesis illustrating the other potential role of thrombin-like proteases in cellular signaling in P. tricornutum.***

Thrombin (in green) might be implicated in GPCR receptors (in blue) activation/inhibition in response to environmental changes, resulting in biofilm formation and surface colonization.

4.2 Final Conclusion

This study reinforces *P. tricornutum* as a promising heterologous platform for biotechnological applications, particularly in cannabinoid biosynthesis. We successfully produced and characterized different versions of CBDAS in *P. tricornutum*, demonstrating enzymatic activity and CBDA production. Furthermore, our study marks the first demonstration of a thrombin cleavage sequence in *P. tricornutum*, providing a novel strategy and alternative for the simultaneous expression of multiple genes under a single promoter. Also, understanding the functional roles of these proteases in *P. tricornutum* could provide further insights in diatom biology, with implications for both marine ecology and industrial biotechnology. While additional optimizations are needed to enhance protein expression efficiency, yield stability, and secretion pathways, our findings validate *P. tricornutum* as a sustainable and scalable biofactory.

5 Appendix I

5.1 Supplementary Data of Chapter II

Table 5-1: ***Table S1. Sequences used in this study.***

Name	**Nucleotide sequence (codon optimized)**	**Reference**
Highly abundant secreted protein 1 (HASP1) promoter	Catacagtgaatgtaactttcgaattgacagtattagtagtcgtattgacagtgaggcacgcccctcaatgtgcgaggtggaaaatataccagcatgacaatgaatcttggagattcttttgctgtcatcaagattcaccgccaaatcttcaggaacctatcacgtccacaggcgatgttaattcttgagtcgtcaaaacaaagtcctgtcctacctgtagaagttgacagcgagcaattgtatgcaaacttctgactttgttataataacattaaaggtaattaagtatcttcaattaggcattttgtcactgtcagtccgttccgacaatataggtagatttggaatgaatcttttctatgctgctgcgaatcttgtacaccttgaggccgtagattctgtccgacgaagcgataatattgcaaaatacatggactcattattttgattcgatttctttttggtatccgactcgaaaagatccatcacggcgagc	[102]
Kozak sequence	Gccacc	[102]
Highly abundant secreted protein 1 (HASP1) secreted signal	ATGAATCTTCGTTGTATCCTTCCGTTTCTCCTCGCAAGCTTCTCGGCTGGGGCT	[102]
Cannabidolic acid synthase (CBDAS)	ATGAAGTGCTCCACCTTTTCCTTTTGGTTCGTCTGTAAGATTATCTTCTTCTTTTTCTCCTTCAACATCCAGACCTCCATTGCCAACCCCCGTGAAAACTTCCTCAAGTGCTTCTCCCAGTACATTCCCAACAACGCCACCAACCTCAAGCTCGTCTACACCCAGAACAACCCCTTGTACATGTCCGTCCTCAACTCCACCATTCACAACCTCCGTTTCACCTCCGATACTACCCCCAAGCCGCTCGTCATTGTCACCCCCTCCCACGTTTCGCACATTCAGGGAACCATCCTCTGCTCCAAGAAGGTCGGATTGCAGATCCGTACCCGTTCCGGAGGACACGACTCCGAAGGAATGTCCTACATTTCCCAGGTCCCCTTTGTCATTGTCGATCTCCGTAACATGCGTTCCATCAAGATTGACGTCCACTCCCAGACCGCCTGGGTCGAAGCCGGAGCCACCTTGGGAGAAGTTTACTACTGGGTCAACGAAAAGAACGAAAACCTCTCCCTCGCCGCCGGATACTGCCCCACCGTCTGCGCCGGAGGCCACTTCGGTGGAGGAGGTTACGGACCCCTCATGCGCAACTACGGACTCGCCGCCGACAACATTATTGACGCCCACCTCGTCAACGTCCACGGCAAGGTCCTTGACCGTAAGTCGATGGGTGAAGATTTGTTCTGGGCCTTGCGCGGTGGAGGAGCCGAATCCTTCGGCATTATCGTTGCCTGGAAGATTCGCTTGGTCGCCGTCCCCAAATCGACCATGTTCTCGGTCAAGAAGATTATGGAAATTCACGAACTCGTCAAGCTCGTCAACAAGTGGCAGAACATTGCCTACAAGTACGACAAGGACTTGCTCCTCATGACCCACTTCATTACCCGTAACATTACCGACAACCAGGGGAAAGAACAAGACCGCCATTCACACCTACTTCTCCTCTGTTTTTCTCGGTGGTGTCGATTCCCTCGTTGACTTGATGAACAAGTCCTTCCCGGAACTCGGAATCAAAAAGACCGATTGCCGTCAGCTCTCCTGGATTGACACCATTATTTTCTACTCCGGAGTCGTCAACTACGACACCGACAACTTCAACAAGGAAATTCTCTTGGATCGTTCCGCCGGACAGAACGGTGCCTTCAAGATTAAGCTCGACTACGTCAAGAAGCCCATTCCCGAATCCGTCTTTGTCCAGATCTTGGAAAAGCTCTACGAAGAAGATATTGGAGCCGGAATGTACGCCCTCTACCCCTACGGTGGAATTATGGACGAAATTTCCGAATCCGCCATTCCCTTTCCCCACCGCGCCGGCATTCTCTACGAACTCTGGTACATTTGCTCGTGGGAAAAGCAGGAAGATAACGAAAAGCACTTGAACTGGATTCGTAACATCTACAACTTCATGACCCCCTACGTTTCCAAGAACCCCCGTTTGGCCTACCTCAACTACCGTGACCTCGATATTGGAATTAACGACCCCAAGAACCCCAACAACTACACCCAGGCCCGTATTTGGGGAGAAAAGTACTTCGGAAAAAACTTTGACCGTTTGGTCAAGGTCAAGACCCTAGTTGACCCTAATAACTTCTTCCGTAATGAGCAGAGCATCCCCCCTCTCCCACGCCACCGCCAC	Genbank: NP_001384865.1 Synthesized by Biobasic
Thrombin sequence	CTCGTCCCCCGTGGCTCC	pET-28a-c. [112]. Synthesized by Biobasic

Yellow fluorescent protein (*YFP*)	ATGCGTAAGGGAGAAGAACTCTTCACCGGAGTCGTCCCCATTCTCGTCGAACT CGACGGTGACGTCAACGGACACAAGTTCTCCGTTTCCGGAGAAGGAGAAGG CGACGCCACCAACGGAAAGCTCACCTTGAAGTTCATTTGTACCACCGGAAAG CTCCCCGTCCCCTGGCCCACCCTCGTCACCACCCTCACCTACGGAGTCCAGTG CTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCA TGCCCGAAGGATACGTCCAGGAACGTACCATTTCCTTTAAGGACGATGGAACC TACAAGACCCGCGCCGAAGTCAAGTTCGAAGGAGACACCCTCGTCAACCGTA TTGAACTCAAGGGAATTGACTTCAAGGAGGACGGAAACATTCTCGGACACAA GCTCGAATACAACTTCAACTCCCACAACGTCTACATCACCGCCGACAAGCAGA AGAACGGAATTAAGGCCAACTTTAAGATTCGTCACAATGTCGAAGATGGATCG GTCCAGCTCGCCGACCACTACCAGCAGAACACCCCCATTGGTGACGGTCCCG TCCTCCTCCCCGACAACCACTACCTCTCCTACCAGTCCGCCCTATCCAAGGACC CCAACGAAAAGCGTGACCACATGGTCCTCCTCGAATTTGTCACCGCCGCCGG AATTACCCACGGAATGGATGAATTGTACAAGCGTCCTGCCGCTAATGACGAAA ACTACGCCGCCTCCGTC	[189]
Hemagglutinin tag (3xHA)	TACCCCTACGACGTCCCTGACTACGCCGGCTACCCGTACGACGTCCCCGACTA CGCCGGTTCGTACCCCTACGATGTCCCGGACTACGCC	[190] Synthesized by Biobasic
Fucoxanthin chlorophyll a/c binding protein A (*FcpA*) terminator	ccgcaacaactacctcgactttggctgggacactttcagtgaggacaagaagcttcagaagcgtgcta tcgaactcaaccagggacgtgcggcacaaatgggcatccttgctctcatggtgcacgaacagttggg agtctctatccttccttaaaaatttaattttcattagttgcagtcactccgctttggttt	[116]
Fucoxanthin chlorophyll a/c binding protein C (*FcpC*) promoter	gagcacaagaggtgacaaaagccaccggctggatcgcacttctcggaatttccccctactatcaaa caaattcgaattgccaaaggtgaagggactaactgtaaatcctgatcaatcaaggtctcaatcaagta caatgggctacaatgatatttagatgggaacacaatgaaacaaattgaaacttctactgacaggagc gcaattgacttgtgtagcttttcatgagcacttgattgctaccaattgtgaacgggatggggaaagactc gaaaaggtgcatgcttccgataatctactatattttctagaatcaaataatatttaaatgaatgaggtcc tcagcgtacgttaagcctacttatttagaacgagaagtcagaccgaggggtactaaaattctaagggt tgagaggtatcttgattccgggtctatggaagcccatccttgttgaagcttgaacacgatccttgtgaaa ggccgacgttgcgcgaaaaaacagcctgccgatttctttccttctttctcgtctcaacctatatactttcat aatctctgttagagtttaccaacaacacatatatacatttcgacaaa	[116]
nat, nourseothricin resistance gene	ATGACCACTCTTGACGACACGGCTTACCGGTACCGCACCAGTGTCCCGGGGG ACGCCGAGGCCATCGAGGCACTGGATGGGTCCTTCACCACCGACACCGTCTT CCGCGTCACCGCCACCGGGGACGGCTTCACCCTGCGGGAGGTGCCGGTGGA CCCGCCCCTGACCAAGGTGTTCCCCGACGACGAATCGGACGACGAATCGGAC GACGGGGAGGACGGCGACCCGGACTCCCGGACGTTCGTCGCGTACGGGGA CGACGGCGACCTGGCGGGCTTCGTGGTCGTCTCGTACTCCGGCTGGAACCGC CGGCTGACCGTCGAGGACATCGAGGTCGCCCCGGAGCACCGGGGGCACGG GGTCGGGCGCGCGTTGATGGGGCTCGCGACGGAGTTCGCCCGCGAGCGGG GCGCCGGGCACCTCTGGCTGGAGGTCACCAACGTCAACGCACCGGCGATCC ACGCGTACCGGCGGATGGGGTTCACCCTCTGCGGCCTGGACACCGCCCTGTA CGACGGCACCGCCTCGGACGGCGAGCAGGCGCTCTACATGAGCATGCCCTGC CCCTG	[116]
Fucoxanthin chlorophyll a/c binding protein C (*FcpC*) terminator	ttttgttacattgacttcaaggagtcgaggaatcgatactgccgtcgtttccaggatccgaggtttctata gactctctatagactctgttaacctaatagaatcagacatacctctcctgctattttgtttttatgaatttg gcttttgcctctctagtcagatttgaatgttattttccgccaggtgtgttagtcgggctctcgtttgagttac aagagggattgagtggcgaggattcactctaatgtaaatatgactgtgaacaaaactttaaaattact acgcatcttctttgactgtcagatattcgtcggtgacagcagtcaatgcctgcaaattgtcctcctgggtc gcaatttggttttggattgacctggtatgcattatgaagaaaaaaattcgttattagccaactgcctagc gtgcacattgcatggttagacctccttgacgactgtg	[116]

Table 5-2: ***Table S2. Primers used in this study.***

Name	Sequence (5′ to 3′) (Priming sequence for PCR)	Description	Use
FcpCt_HASP1p F	GAGCCTACATCCTTCTGCAACAAGCTG CAATCATACAGTGAATGTAACTTTCGA ATTGAC	Forward for HASP1 promotor with FcpC terminator homology	Gibson assembly of pYFP, pSP:YFP, pCBDAS, pSP:CBDAS
HASP1 p_CBDAS R	GTCAATTCGAAAGTTACATTCACTGTAT GATTGCAGCTTGTTGCAGAAGGATGTA GGCTC	Reverse for HASP1 promotor with CBDAS homology	Gibson assembly of pCBDAS
HASP1p_CBDAS F	TCGAAAAGATCCATCACGGCGAGCGC CACCATGAAGTGCTCCACCTTTTCCTTT TGGTTC	Forward for CBDAS with HASP1 promotor homology	Gibson assembly of pCBDAS
CBDASCt_Thr R	TTCTCCCTTACGGGAGCCACGGGGGA CGAGGTGGCGGTGGCGTGGGAGAGG GGGGATGCT	Reverse for CBDAS with Thrombin homology	Gibson assembly of pCBDAS, pSP:CBDAS and pΔ28aaCBDAS
CBDASCt_Thr F	AGCATCCCCCCTCTCCCACGCCACCGC CACCTCGTCCCCCGTGGCTCCCGTAAG GGAGAA	Forward for Thrombin, YFP and HA tag with CBDAS homology	Gibson assembly of pCBDAS, pSP:CBDAS and pΔ28aaCBDAS
HA_FcpAt R	CGGACCGGCGGTGTTGGTCGGCGTCG GTTATTAGGCGTAGTCCGGGACATCGT AGGGGTA	Reverse for Thrombin, YFP and HA tag with FcpA terminator homology	Gibson assembly of pYFP, pSP:YFP, pCBDAS, pSP:CBDAS
HA_FcpAt F	TACCCCTACGATGTCCCGGACTACGCC TAATAACCGACGCCGACCAACACCGCC GGTCCG	Forward for FcpA terminator with HA tag homology	Gibson assembly of pYFP, pSP:YFP, pCBDAS, pSP:CBDAS
FcpAt-Frg1 R	GGAAGATCTATATTACCCTGTGAAGAC GAGCTAGTGTTATTCCTGACTGTGAAA CCAAAG	Reverse for FcpA terminator with vector (backbones) homology	Gibson assembly of pYFP, pSP:YFP, pCBDAS, pSP:CBDAS
FcpAt-Frg1 F	ATAACACTAGCTCGTCTTCACAGGGTA ATATAGATCTTCCGCTGCATAACCCTGC TTCGG	Forward primer used to amplify half of the backbone with FcpA terminator homology	Gibson assembly of pYFP, pSP:YFP, pCBDAS, pSP:CBDAS
D630R	TTTGCAAACCAAGTTCGACAACTGCGT ACGGCCTGTTCGAAAGATCTACCACCG CTCTGG	Reverse primer used to amplify half of the backbone with the other half of backbones homology	Gibson assembly of pYFP, pSP:YFP, pCBDAS, pSP:CBDAS
D631F	GGGCCGTGCGTGGAGTAAAAAGGTTT GGATCAGGATTTGCGCCTTTGGATGA GGCACTTT	Forward primer used to amplify half of the backbone with the other half of backbones homology	Gibson assembly of pYFP, pSP:YFP, pCBDAS, pSP:CBDAS
D631R	ATGGGCTTCGCCCTGTCGCTCGACTGC GGCGAGCACTACTGGCTGTAAAAGGA CAGACCA	Reverse primer used to amplify half of the backbone with FcpC promoter homology	Gibson assembly of pYFP, pSP:YFP, pCBDAS, pSP:CBDAS

D635F	TGGTCTGTCCTTTTACAGCCAGTAGTGCTCGCCGCAGTCGAGCGACAGGGCGAAGCCCATGAGCACAAGAGGTGACAAAA	Forward for FcpC promotor, Nat and FcpC terminator with half of the backbone homology	Gibson assembly of pYFP, pSP:YFP, pCBDAS, pSP:CBDAS
FcpCt_HASP1p R	GTCAATTCGAAAGTTACATTCACTGTATGATTGCAGCTTGTTGCAGAAGGATGTAGGCTC	Reverse for FcpC promotor, Nat and FcpC terminator with HASP1 promotor homology	Gibson assembly of pYFP, pSP:YFP, pCBDAS, pSP:CBDAS
HASP1SP_CBDAS R	GAAGCACTTGAGGAAGTTTTCACGGGGGTTAGCCCCAGCCGAGAAGCTTGCGAGGAGAAA	Reverse for HASP1 promotor and signal peptide (SP) with CBDAS homology	Gibson assembly of pSP:CBDAS
HASP1SP_CBDAS F	TTTCTCCTCGCAAGCTTCTCGGCTGGGGCTAACCCCCGTGAAAACTTCCTCAAGTGCTTC	Forward for cbdas with signal peptide (SP) homology	Gibson assembly of pSP:CBDAS
HASP1p_Thr R	TTCTCCCTTACGGGAGCCACGGGGGACGAGCATGGTGGCGCTCGCCGTGATGGATCTTTTCGA	Reverse for HASP1 promotor with thrombin and YFP homology	Gibson assembly of pSP:CBDAS
HASP1p_Thr F	TCGAAAAGATCCATCACGGCGAGCGCCACCATGCTCGTCCCCCGTGGCTCCCGTAAGGGAGAA	Forward for thrombin and YFP with HASP1 promotor homology	Gibson assembly of pSP:CBDAS
HASP1 _woSPCBDAS F	ATCCGACTCGAAAAGATCCATCACGGCGAGCGCCACCATGAACCCCCGTGAAAACTTCCT	Forward for Δ28aaCBDAS with HASP1 promotor homology	Gibson assembly of pSP:CBDAS
HASP1_woSPCBDAS R2	GAAGCACTTGAGGAAGTTTTCACGGGGGTTCATGGTGGCGCTCGCCGTGATGGATCTTTT	Reverse for HASP1 promotor with Δ28aaCBDAS homology	Gibson assembly of pSP:CBDAS

Table 5-3: ***Table S3. Analytical parameters used for compounds identification using high-performance liquid chromatography with diode-array detection (HPLC-DAD).***

Abbreviations: olivetolic acid (OA), olivetol (OL), delta-9-tetrahydrocannabinol (THC), cannabidiol (CBD), cannabinol (CBN), delta-9-tetrahydrocannabinolic acid (THCA), cannabidiolic acid (CBDA), cannabinolic acid (CBNA), cannabigerolic acid (CBGA), cannabichromene (CBC), cannabigerol (CBG), tetrahydrocannabivarin (THCV), cannabidivarin (CBDV), retention time (RT), maximum absorption wavelengths (λmax).

Compound	**RT** **(min)**	**λmax** **(nm)**
OA	1.936	214/261/300
OL	1.716	202/276
THC	11.827	199/279/328
CBD	7.811	199/276
CBN	10.773	198/221/283
THCA	15.636	199/223/270/306
CBDA	8.709	198/223/269/307
CBNA	14.547	199/222/261/328
CBGA	10.146	198/222/268/305
CBC	13.517	199/226/280
CBG	8.025	199/274
THCV	8.239	199/278
CBDV	4.942	200/275

Table 5-4: ***Table S4. Optimized instrumental parameters used for HPLC-MS/MS analyses in ESI+.***

References: olivetolic acid (OA), olivetol (OL), delta-9-tetrahydrocannabinol (THC), cannabidiol (CBD), cannabinol (CBN), delta-9-tetrahydrocannabinolic acid (THCA), cannabidiolic acid (CBDA), cannabinolic acid (CBNA), cannabigerolic acid (CBGA), cannabichromene (CBC), cannabigerol (CBG), tetrahydrocannabivarin (THCV), cannabidivarin (CBDV), retention time (RT), collision energy (CE). Quantification MRM transitions are bold while qualifier MRM transitions are not.

Compound	RT (min)	Parent ion (m/z)	Product Ion (m/z)	Fragmentor (V)	CE (V)	Polarity
OA	4.272	**225**	**207**	**75**	**10**	**+**
			123	75	18	+
			189	75	18	+
OL	4.083	**181**	**111**	**75**	**10**	**+**
			71	75	10	+
			93	75	26	+
THC	9.225	**315**	**193**	**80**	**22**	**+**
			123	80	34	+
			135	80	18	+
CBD	7.752	**315**	**193**	**75**	**22**	**+**
			123	75	34	+
			135	75	18	+
CBN	8.825	**311**	**223**	**75**	**22**	**+**
			293	75	14	+
			241	75	18	+
THCA	10.334	**359**	**341**	**80**	**14**	**+**
			219	80	34	+
			285	80	26	+
CBDA	7.935	**359**	**341**	**80**	**10**	**+**
			219	80	30	+
			261	80	26	+
CBNA	9.933	**355**	**337**	**85**	**14**	**+**
			253	85	30	+
			235	85	30	+
CBGA	8.289	**361**	**219**	**80**	**26**	**+**
			343	80	10	+
			237	80	10	+
CBC	9.556	**315**	**193**	**75**	**18**	**+**
			259	75	10	+
			81	75	10	+
CBG	7.655	**317**	**193**	**80**	**14**	**+**
			123	80	34	+
			207	80	10	+
THCV	8.162	**287**	**165**	**75**	**22**	**+**
			123	75	34	+
			231	75	18	+
CBDV	6.642	**287**	**165**	**75**	**22**	**+**
			123	75	34	+
			231	75	14	+

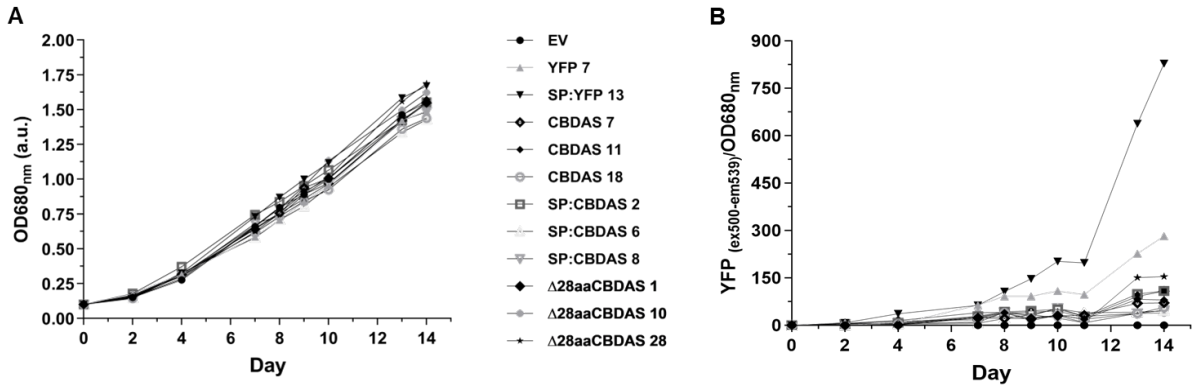


Figure 5-1: ***Fig. S1. CBDAS strains growth curves and YFP fluorescence emission.***

A. Growth curves of each strain, the optical density (OD) at 680 nm was followed for 14 days. B. YFP fluorescence emission (excitation 500 nm- emission 539 nm), from the 12 cultures monitoring in A, was measured and normalized to the OD.

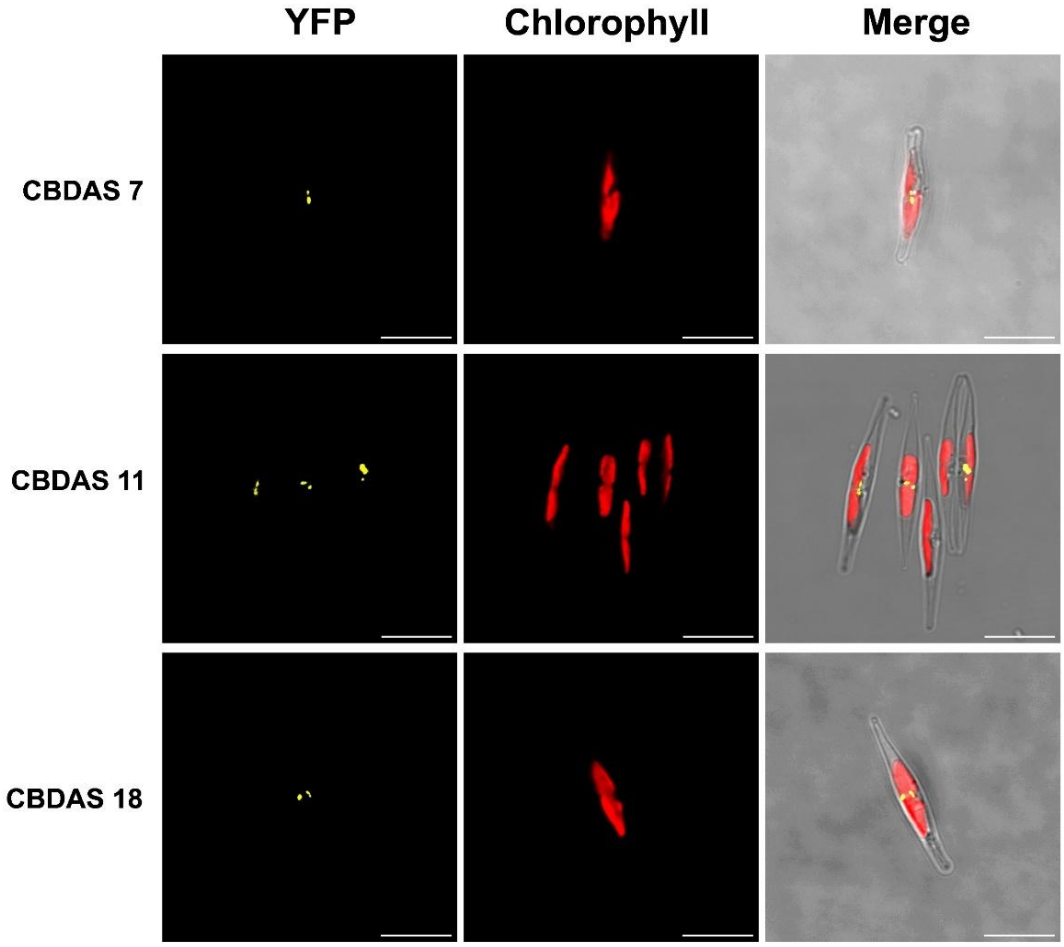


Figure 5-2: ***Fig. S2. Subcellular localization of three CBDAS clones.*** *YFP fluorescence, chlorophyll autofluorescence, and the merging of three fields are shown in transgenic lines producing CBDAS:YFP, visualized by confocal laser microscopy. Scale bars = 10 µm.*

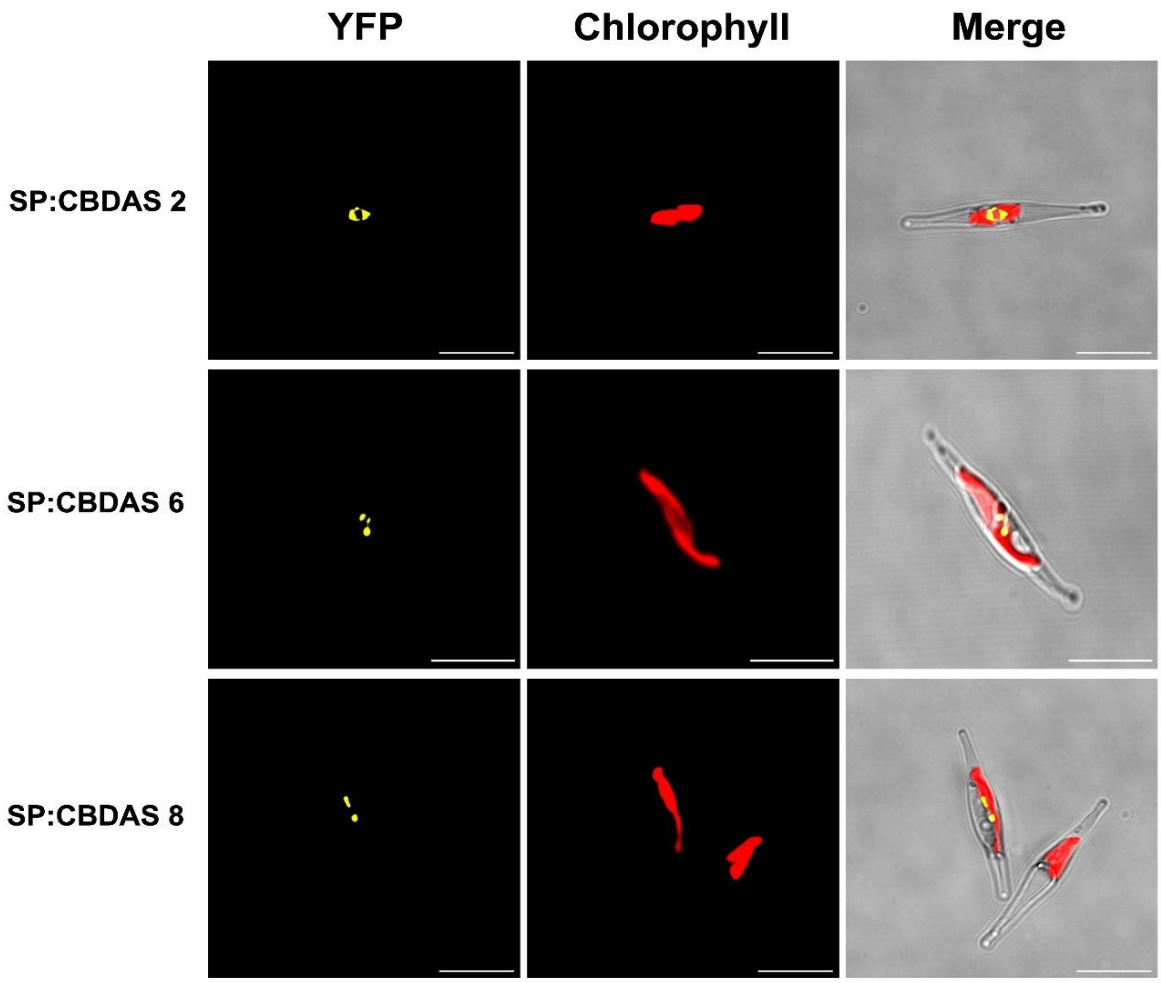


Figure 5-3: ***Fig. S3. Subcellular localization of three SP:CBDAS clones.*** *YFP fluorescence, chlorophyll autofluorescence, and the merging of three fields are shown in transgenic lines producing SP:CBDAS:YFP, visualized by confocal laser microscopy. Scale bars = 10 µm.*

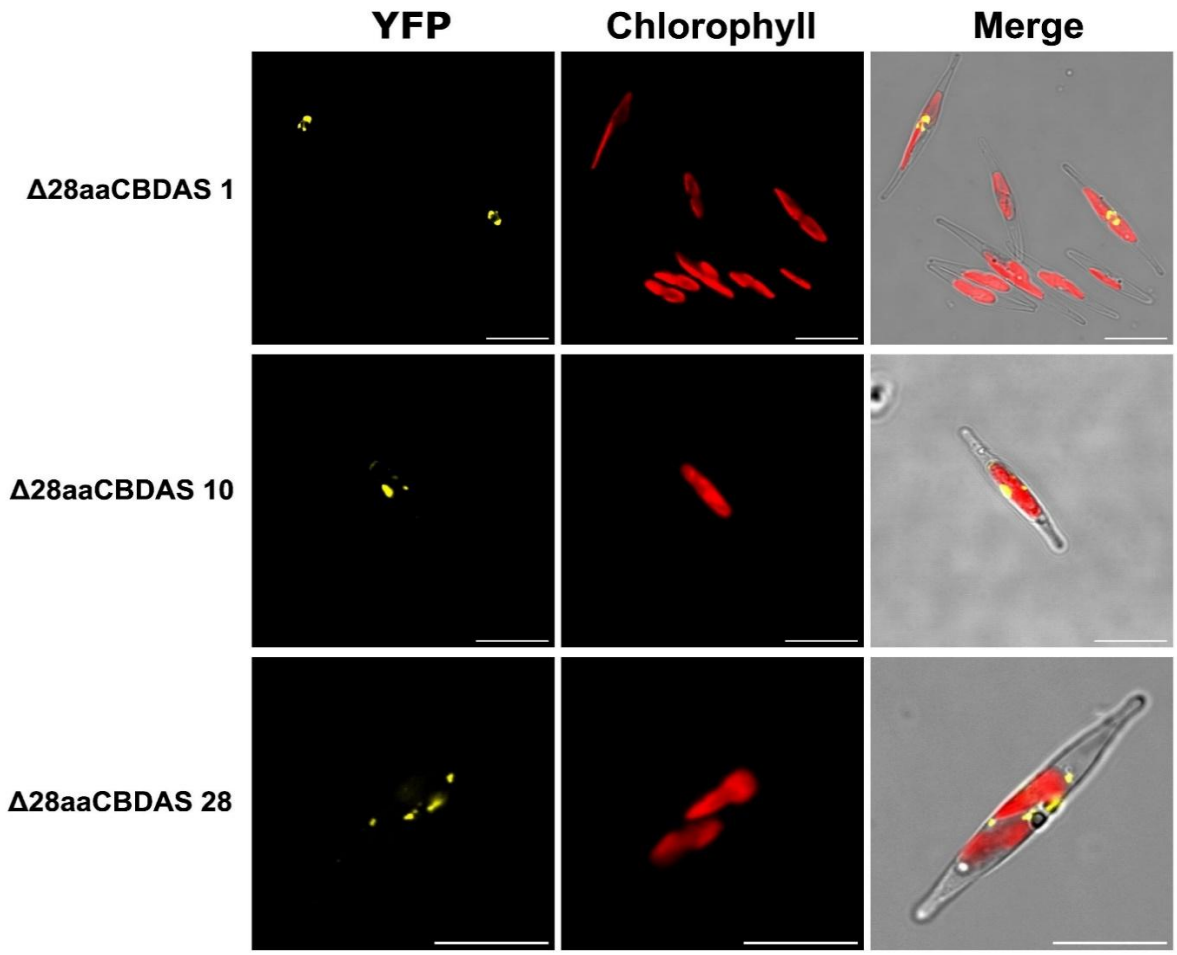


Figure 5-4: ***Fig. S4. Subcellular localization of three Δ28aaCBDAS clones.*** *YFP fluorescence, chlorophyll autofluorescence, and the merging of three fields are shown in transgenic lines producing Δ28aaCBDAS:YFP, visualized by confocal laser microscopy. Scale bars = 10 µm.*

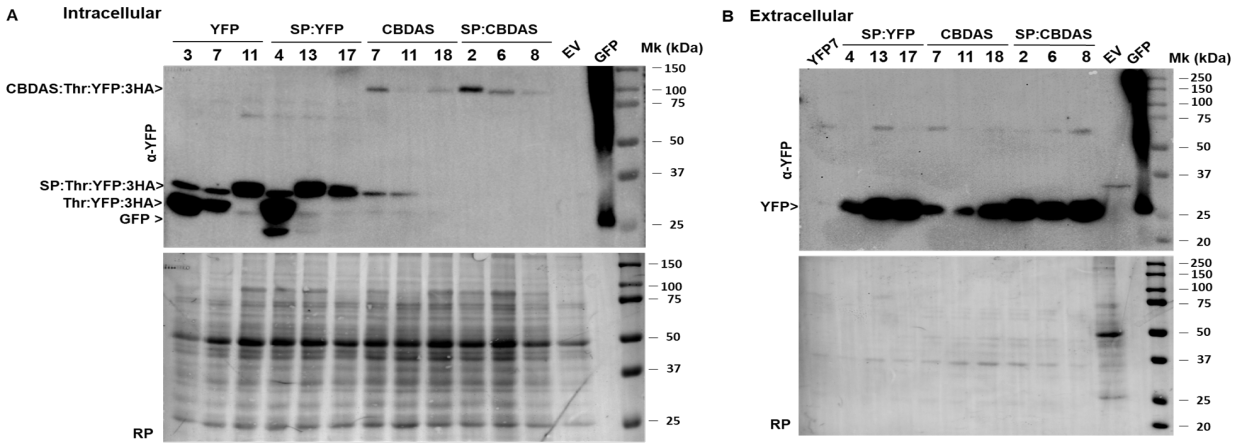


Figure 5-5: ***Fig. S5. Intracellular detection of CBDAS enzyme.***

Uncropped Western blots from Fig. 4A & B, with an anti-GFP/YFP/CFP antibody of total cell extracts from YFP (Thr:YFP:3HA, 32 kDa), SP:YFP (SP:Thr:YFP:3HA, 34 kDa), CBDAS (CBDAS:Thr:YFP:3HA, 93 kDa), SP:CBDAS (SP:CBDAS:Thr:YFP:3HA, 92 kDa), EV and purified GFP (27 kDa). Lower panel, stained blot with red ponceau solution (RP).

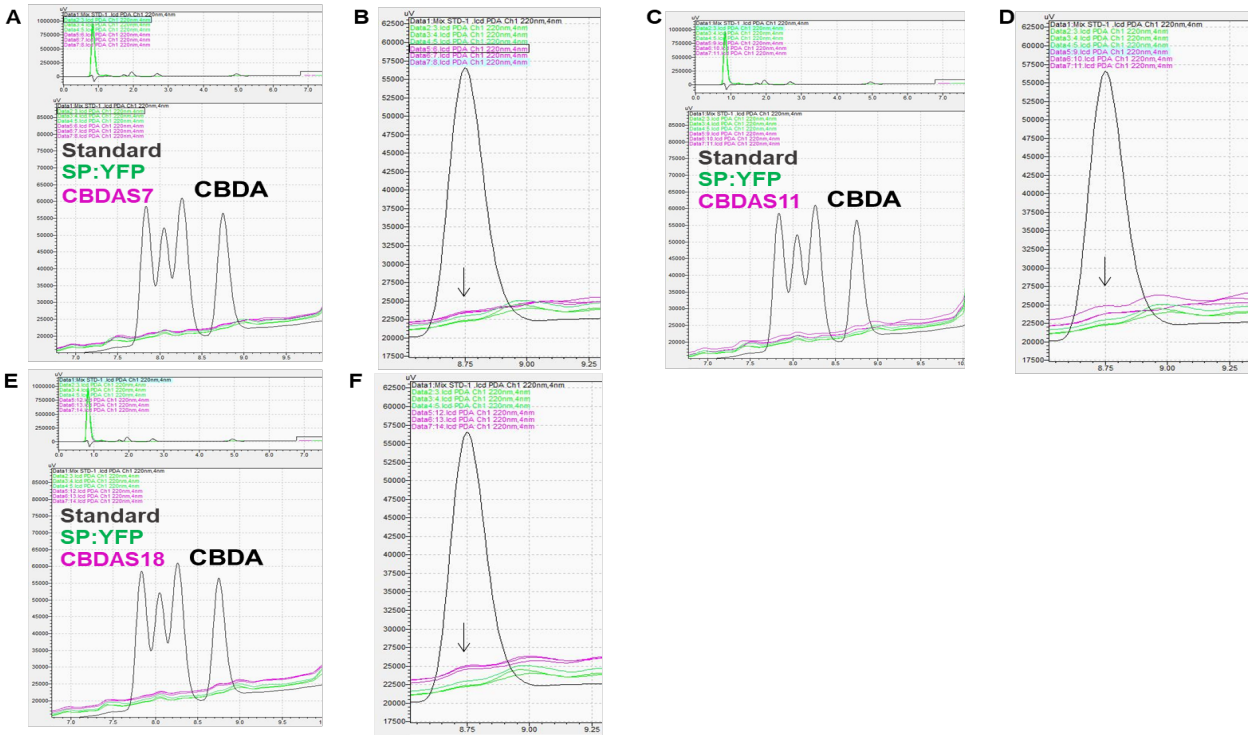


Figure 5-6: ***Fig. S6. CBDA detection by HPLC-DAD at 220 nm.***

A Mix solution with 10 ppm of commercial standards (black), in vitro assays using SP:YFP (green) and CBDAS7 strain (pink) protein extracts in presence of CBGA. B. Zoom in of A, showing CBDA peak with a black arrow. C-D and E-F CBDAS11 and CBDAS18 strains chromatograms respectively.

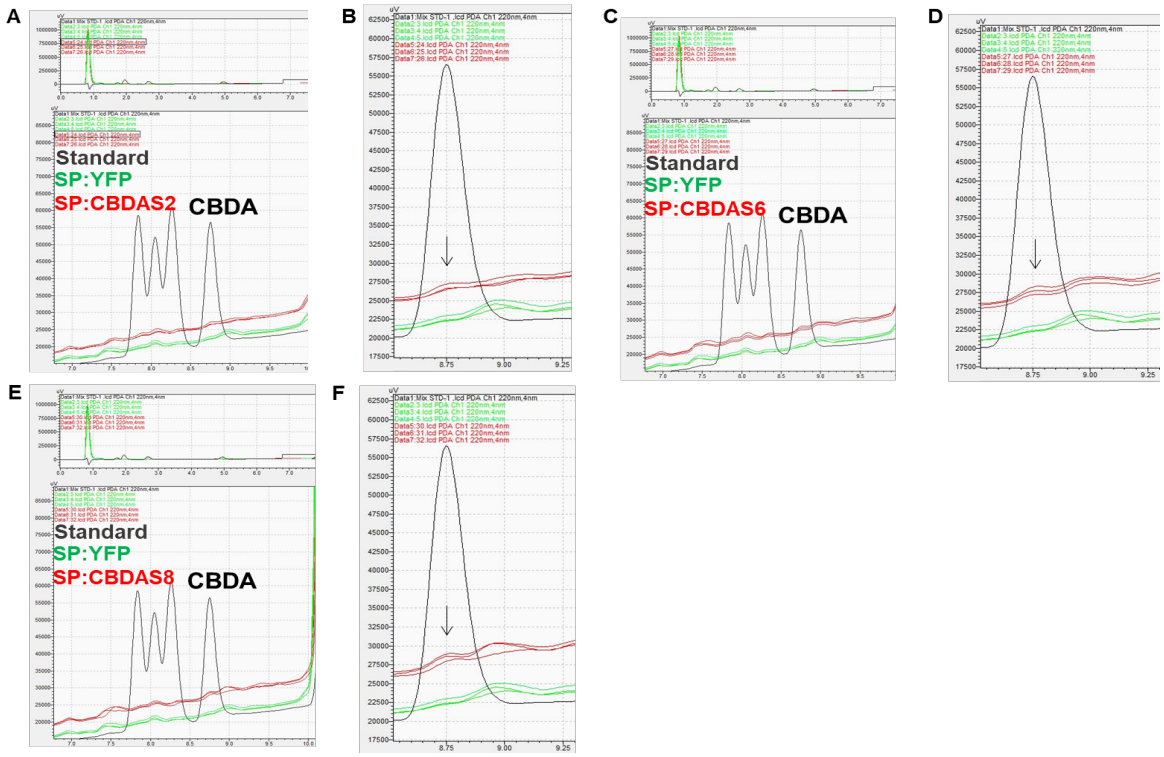


Figure 5-7: ***Fig. S7. CBDA detection by HPLC-DAD at 220 nm.***

A. Mix solution with 10 ppm of commercial standards (black), in vitro assays using SP:YFP (green) and SP:CBDAS2 strain (red) protein extracts in presence of CBGA. B. Zoom in of A, showing CBDA peak with a black arrow. C-D and E-F, SP:CBDAS6 and SP:CBDAS8 strains chromatograms respectively.

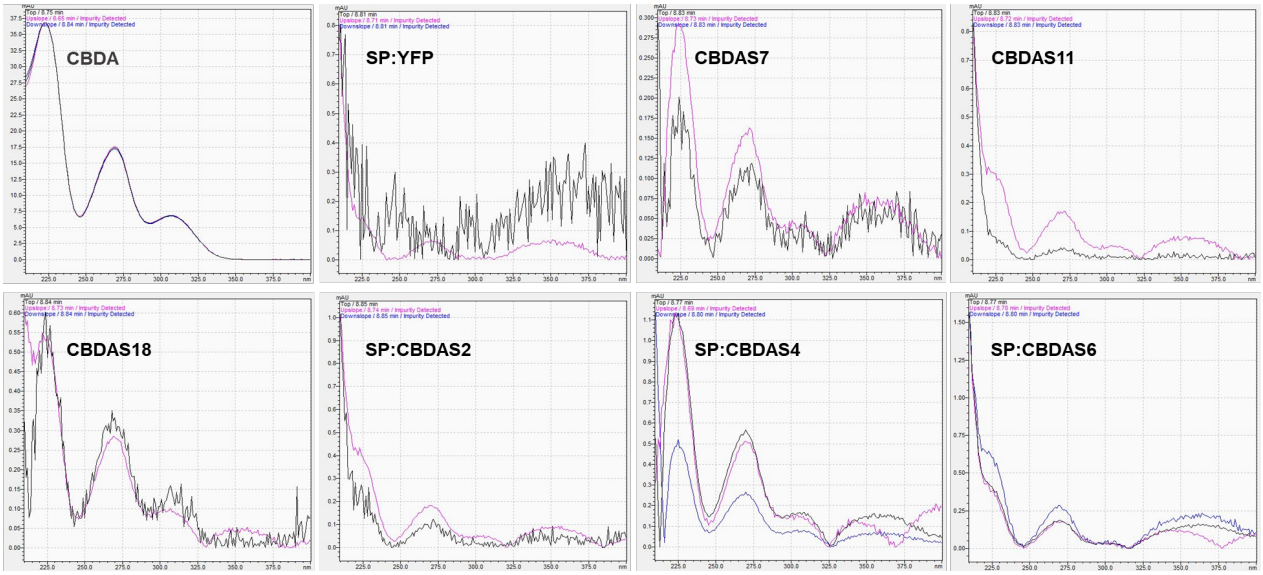


Figure 5-8: ***Fig. S8. UV spectrum (190-400nm) of 10 ppm of Standard CBDA and in vitro assay results from SP:YFP, CBDAS7, CBDAS11, CBDAS18, SP:CBDAS2, SP:CBDAS 4 and SP:CBDAS6 strains (black).***

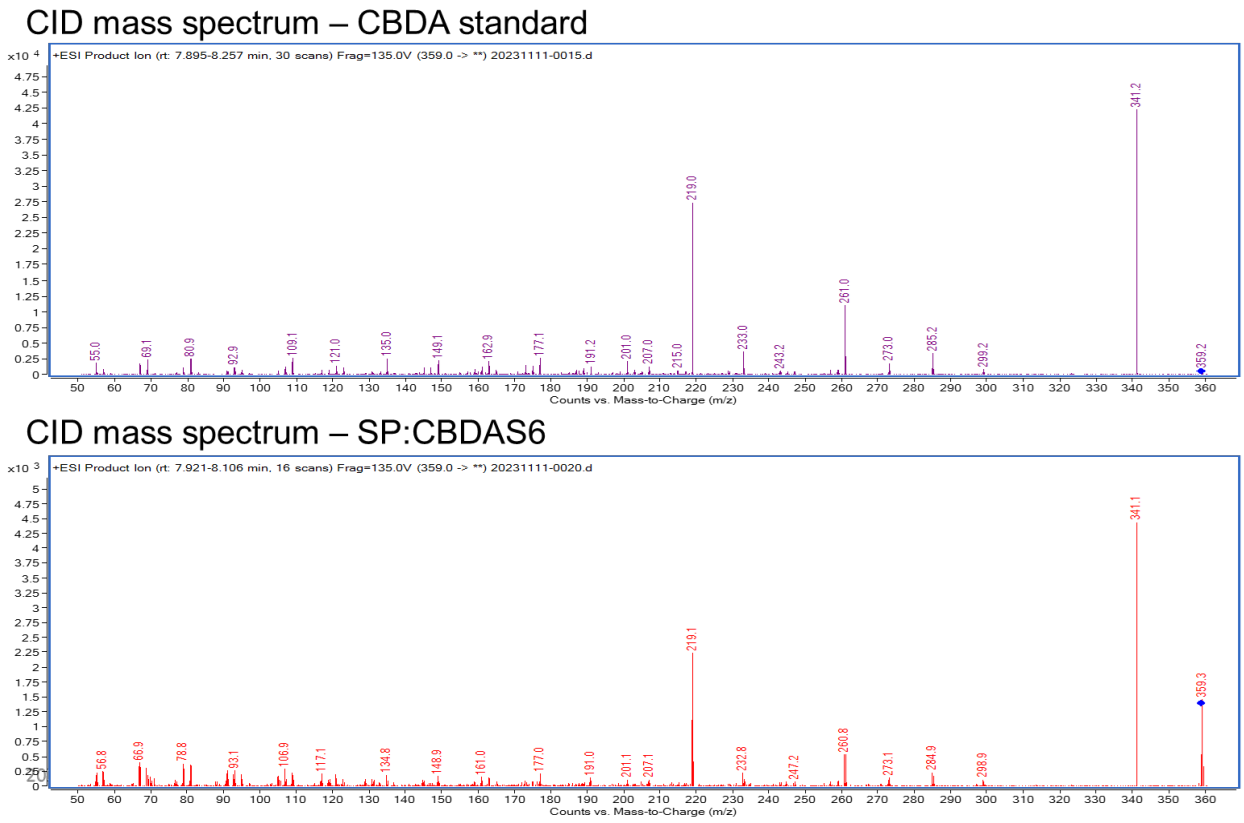


Figure 5-9: ***Fig. S9. CBDA detection and validation by HPLC-MS/MS.*** *MS/MS fragmentation of the in vitro assays confirming identity of CBDA (standard, upper panel) and SP:CBDAS6 (lower panel).*

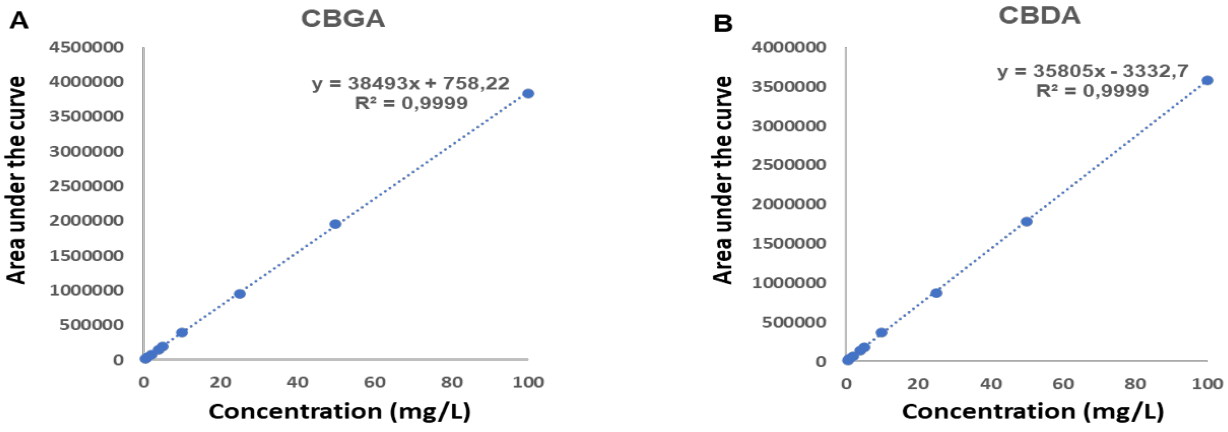


Figure 5-10: ***Fig. S10. Cannabigerolic acid (A) and cannabidiolic acid (B) calibration curves.***

6 Appendix II

6.1 Supplementary Data of Chapter III

*Table 6-1: **Supplementary Table S1**. List of genetic parts used in this study.*

Name	Sequence	Description
HASP1 promoter	catacagtgaatgtaactttcgaattgacagtattagtagtcgtattgacagtgaggcacg ccctcaatgtgcgaggtggaaaatataccagcatgacaatgaatcttggagattcttttg ctgtcatcaagattcaccgccaaatcttcaggaacctatcacgtccacaggcgatgttaa ttcttgagtcgtcaaaacaaagtcctgtcctacctgtagaagttgacagcgagcaattgta tgcaaacttctgactttgttataataacattaaaggtaattaagtatcttcaattaggcattttg tcactgtcagtccgttccgacaatataggtagatttggaatgaatcttttctatgctgctgcg aatcttgtacacctttgaggccgtagattctgtccgacgaagcgataattattgcaaaata catggactcattattttgattcgatttctttttggtatccgactcgaaaagatccatcacggcg agc	Highly Abundant Secreted Protein1 promoter sequence Erdenedolgor et al 2019
40SRPS8 Promoter	ccctgcgatagacctttccaaactcacgcagtccaagaaaacaaaggggtgagaagt atacgcacctttcggtttcggcataattcttaaactcttgtggtcactttcttgtgaagaagcta ggggcactcgttttccctcagagcctgcaaacacaaaattcctgcagtcaattgtcccaa cactcggcaaaccgtatgcgcaagcaacgatgcgcagaaggccgtggatggatggc gactcgcgatatggcttcttgggtcgccagtgtggtacgtccggcgtatgtcaatacgcga attcggacgactggcatctctaggaggaggattccttcttttatgacatgtttatttttatataca ttgatgctttccgacagtcggaagtaataaatgaatttatttcaagactacctatactcctttg acttgttcgactaatcttaccgcttactaaaatctcgaaatcacgcttgacctctcgcacgc aaatttttgctgctggacgctacgcactcggcccaattcttctcggtcctcgtcgtcgcaatt gtcgttgcgttgatcttgcaccgaaggaatcagagaatagaatacc	40S ribosomal protein S8 promoter
Kozak	gccacc	Translation enhancer sequences Erdenedolgor et al 2019
CBDAS	aagtgctcgacgttttcgttttggtttgtctgtaagattatttttttctttttttcctttaacattcaaac atctattgctaatccacgagaaaacttcttgaagtgcttctcgcaatacattccgaacaac	Codon optimized

	gccacaaacctcaagctcgtgtacacgcaaaacaacccactctacatgtcggtgttgaa ctcgaccattcataacttgcgatttacctccgacacaacgcctaagccattggtgattgtta cgccctctcatgtctcgcacattcaaggaaccattttgtgctcgaagaaggttggattgca aattcgtacgcgttcgggaggccacgattcggaaggtatgtcctacatttctcaagtcccc tttgtgattgtggatcttcgaaacatgcgctcgattaaaattgacgtccactcccaaacggc ctgggttgaagctggtgccacattgggagaagtctactattgggtgaacgaaaaaaatg aaaatttgtccttggccgctggttattgtcctacggtgtgcgccggcggtcactttggcgga ggcggatacggtcccttgatgcgaaattacggcttggctgccgacaatattattgacgcc cacttggtgaacgttcacggtaaggtcttggaccgcaagtcgatgggcgaagacttgttc tgggccctccgaggtggaggcgccgaatcgtttggtattattgttgcctggaagattcgctt ggtggccgtccctaagtcgacgatgttttccgtcaagaaaattatggagattcacgaattg gttaagttggtcaacaaatggcaaaacattgcctacaaatacgacaaagacctccttttg atgacccatttcattacgcgaaacattacggataaccagggtaaaaacaaaaccgcca ttcatacgtacttttcgtctgttttcttgggtggagtcgactcgttggtcgacttgatgaacaaa tcctttcccgaattgggaattaagaagacggactgccgacagctctcctggattgacacc attatttttttactctggagtcgtcaactatgacacggacaactttaacaaagagattctgctt gaccgctcggccggacagaacggagcctttaaaattaaattggactatgtcaaaaaac ccattcccgagtcggtctttgtccaaattctcgaaaagttgtacgaagaggacattggagc cggaatgtacgccttgtacccgtacggaggaattatggacgagatttccgaatcggccat tccgtttccgcatcgagctggaattctgtacgaactgtggtacatttgctcgtgggaaaagc aagaagataatgaaaagcacctgaattggattcgcaatatttacaactttatgacgccgt acgtctcgaagaacccgcgcctggcttacctgaactaccgagatctggatattggaatta atgatccgaagaatccgaataattacacccaggctcgaatttggggagaaaagtacttt ggaaagaattttgatcgcctcgtcaaggtcaagacgctcgtcgatccgaataatttttttcg aaacgaacagtccattccgccactccctcgacaccgccac	Cannabidiolic acid synthase
Thrombin cleavage sequence	ctcgtcccccgtggctcc	
Yfp	cgtaagggtgaagaactcttcaccggtgtcgtccccatcctcgtcgaactcgacggaga tgtcaacggacacaagttctccgtctccggtgaaggcgaaggagacgccaccaacgg caagctcaccctcaagtttatttgcaccaccggcaagctccccgtcccctggccgaccct cgttacgaccttgacctacggcgtccagtgcttcgcccgttaccccgaccacatgaagc agcacgacttctttaagtccgccatgcccgaaggctacgtccaggaacgcaccatttcct tcaaggacgacggaacctacaagacccgcgccgaagtcaagtttgaaggtgacacc ctcgtcaaccgtattgaactcaagggcatcgattttaaggaagacggtaacattcttggcc acaagttggaatacaacttcaactcccacaacgtctacattaccgccgataagcagaa gaacggaattaaggccaacttcaagatccgccacaacgtcgaagacggttccgtcca gctcgccgaccactaccagcagaacacccccatcggtgacggacccgtcctcctcccc gacaaccactacttgtcctaccagtcggccctctccaaggacccgaacgaaaagcgtg	Yellow fluorescence protein

	accacatggtcctcctcgaatttgtcaccgccgccggtatcacccacggtatggatgaac tctacaagcgtcccgccgccaacgacgaaaactacgccgcctccgtc	
mCherry	atggtgagcaagggcgaggaggataacatggccatcatcaaggagttcatgcgcttca aggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggc gagggccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtg gccccctgcccttcgcctgggacatcctgtcccctcagttcatgtacggctccaaggccta cgtgaagcaccccgccgacatccccgactacttgaagctgtccttccccgagggcttca agtgggagcgcgtgatgaacttcgaggacggcggcgtggtgaccgtgacccaggact cctccctccaggacggcgagttcatctacaaggtgaagctgcgcggcaccaacttccc ctccgacggccccgtaatgcagaagaaaaccatgggctgggaggcctcctccgagcg gatgtaccccgaggacggcgccctgaagggcgagatcaagcagaggctgaagctga aggacggcggccactacgacgctgaggtcaagaccacctacaaggccaagaagcc cgtgcagctgcccggcgcctacaacgtcaacatcaagttggacatcacctcccacaac gaggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccgg cggcatggacgagctgtacaag	mFruits family of monomeric red fluorescent proteins
3xHA tag	taccctacgacgtccctgactacgccggctacccgtacgacgtccccgactacgccg gttcgtacccctacgatgtcccggactacgcc	
FcpA Terminator	ccgcaacaactacctcgactttggctgggacactttcagtgaggacaagaagcttcaga agcgtgctatcgaactcaaccagggacgtgcggcacaaatgggcatccttgctctcatg gtgcacgaacagttgggagtctctatccttccttaaaaatttaattttcattagttgcagtcac tccgctttggttt	Fucoxanthin chlorophyll a binding protein terminator
FcpC promoter	gagcacaagaggtgacaaaagccaccggctggatcgcacttctcggaatttccccccct actatcaaacaaattcgaattgccaaaggtgaagggactaactgtaaatcctgatcaat caaggtctcaatcaagtacaatgggctacaatgatatttagatgggaacacaatgaaac aaattgaaacttctactgacaggagcgcaattgacttgtgtagcttttcatgagcacttgatt gctaccaattgtgaacgggatggggaaagactcgaaaaggtgcatgcttccgataatct actatattttctagaatcaaataatatttaaatgaatgaggtcctcagcgtacgttaagccta cttatttagaacgagaagtcagaccgaggggtactaaaattctaagggttgagaggtat cttgattccgggtctatggaagcccatccttgttgaagcttgaacacgatccttgtgaaagg ccgacgttgcgcgaaaaaacagcctgccgatttctttccttctttctcgtctcaacctatata ctttcataatctctgttagagtttaccaacaacacatatatacatttcgacaaa	Fucoxanthin chlorophyll c binding protein promoter

Nat	accactcttgacgacacggcttaccggtaccgcaccagtgtcccgggggacgccgag gccatcgaggcactggatgggtccttcaccaccgacaccgtcttccgcgtcaccgccac cggggacggcttcaccctgcgggaggtgccggtggacccgcccctgaccaaggtgttc cccgacgacgaatcggacgacgaatcggacgacggggaggacggcgacccggac tcccggacgttcgtcgcgtacggggacgacggcgacctggcgggcttcgtggtcgtctc gtactccggctggaaccgccggctgaccgtcgaggacatcgaggtcgccccggagca ccgggggcacggggtcgggcgcgcgttgatggggctcgcgacggagttcgcccgcg agcggggcgccgggcacctctggctggaggtcaccaacgtcaacgcaccggcgatc cacgcgtaccggcggatggggttcaccctctgcggcctggacaccgccctgtacgacg gcaccgcctcggacggcgagcaggcgctctacatgagcatgccctgcccctga	Nourseothricin resistance gene
FcpC terminator	ttttgttacattgacttcaaggagtcgaggaatcgatactgccgtcgtttccaggatccgag gtttctatagactctctatagactctgttaacctaatagaatcagacatacctctcctgctattt tgtttttatgaatttggcttttgcctctctagtcagatttgaatgttattttccgccaggtgtgttagt cgggctctcgtttgagttacaagagggattgagtggcgaggattcactctaatgtaaatat gactgtgaacaaaactttaaaattactacgcatcttctttgactgtcagatattcgtcggtga cagcagtcaatgcctgcaaattgtcctcctgggtcgcaatttggttttggattgacctggtat gcattatgaagaaaaaaattcgttattagccaactgcctagcgtgcacattgcatggttag acctccttgacgactgtgagcctacatccttctgcaacaagctgcaat	Fucoxanthin chlorophyll c binding protein terminator
Vector (backbone)	cagggtaatatagatcttccgctgcataaccctgcttcggggtcattatagcgatttttcggt atatccatccttttttcgcacgatatacaggattttgccaaagggttcgtgtagactttccttggt gtatccaacggcgtcagccgggcaggataggtgaagtaggcccacccgcgagcggg tgttccttcttcactgtcccttattcgcacctggcggtgctcaacgggaatcctgctctgcga ggctggccggctaccgccggcgtaacagatgagggcaagcggatggctgatgaaac caagccaaccaggaagggcagcccacctatcaaggtgtactgccttccagacgaacg aagagcgattgaggaaaaggcggcggcggccggcatgagcctgtcggcctacctgct ggccgtcggccagggctacaaaatcacgggcgtcgtggactatgagcacgtccgcga gctggcccgcatcaatggcgacctgggccgcctgggcggcctgctgaaactctggctc accgacgacccgcgcacggcgcggttcggtgatgccacgatcctcgccctgctggcg aagatcgaagagaagcaggacgagcttggcaaggtcatgatgggcgtggtccgccc gagggcagagccatgacttttttagccgctaaaacggccgggggtgcgcgtgattgcc aagcacgtccccatgcgctccatcaagaagaggcacttcgagctgtaagtacatcacc gacgagcaaggcaagacgatccgcgcctgttttattgagaacgttgttcgtgttggcctca atggtagcgatgcgtcattcagcgaagttctggtgctgatgatgtggttcgctttgccactg gtcaatgtggtaagcccgtgtaatgtcagtaacctttttactgatctcagcttgagcacggtc gctgatgagcttatccatggccccacggtaacggatatgatcctctagggcgttgacaaa ttctttgtcggttttcatggggtagacgtcagtgaccagggaacgtctccctacaaaaagtt gcgcatactttgctccgttgtcgacggcaggggtgtcggaccagattgtatctgtggcaac ggcctcattgcgtcaatggacctttaaagccgggaaacgagacttgaaatgtttacgtag	

aggtgcattatagacctcacgcgcatattgagtggttgcaaaaatggttttgcgaacggta
tcactggagacccatgccaggcaaggacggagggcatcataatcatgttcattgcgttg
gacagcatgcttgttacacgtaagtatacggtcgagagcatcgtgacgttgagaattaca
gatggcaacgtgtcggtgcatgtattggccgagattgtcaaatcgtggctcaatatatggc
atgccaggtaagtcatggatgtcggtatgccattcagccttcatgtcgatttcgttgtcgag
aactgatgggtcccagtcaacgtcagaagtaagaacaacatggggaagggaatgaa
gttcatgactggtgggtgcgcgcatatccatgtacggaagaccctgtcgtatgtttagtgg
gatgatgtaatcatccagggtaacaatacgttggttgccgcccacggtacgagaccgat
cttgaacacaattgtggcaatgttcaagttgggcgctggaatggatggttttacctttacca
agatgcgcgtattgatgcatgataacaacaatgggccctcgttgcgattcgacaagacc
ggcagccgtgacaatgtctagattggacaaggtatgttcatttataccggttatgttggctg
aacgtccagttttgtgcagtacggtaacatcactccccgcaagtccaccgttggcaccgc
gatcgacaagtgcagatgtggtattttgtactgtgtgtcgggagacttggtactgtaacaca
tttgactgaagtgacttggtggtaccatgaggacggagggaacccgttgatacggtgttg
cgtgacgctgcaagtaccttacgtatgtcgcctggatccatacgtccaacccgatcagaa
agatgtgctaggagttccgtttctggtgcacaatcatgaaaggtatccacgggagttccat
tgttgtcgtccggggctgcgtcgcgactgccagttgcatgtacatgcgccatcaagggac
gtgccaatccattactaaccgggagaaccttagctggggcagccaatccctgaagtata
gccttagcatcatctgagagctggttccacatatctctagggatataaggtcgttcacggtt
agtgggcgtactatgatgtcggttattagtacaattctccctgcgggcatgcgcattggcttc
atagagtacggaaggtgacaagtcaatattgtagttaacatccggatcagtgtcaaagtc
tgtaggatggtaagaaagatcagtagaatgaatactacgcttttccttgggactacggga
attagagaagttgtttccttattgtagagtgacgctgaagcaagtagaagactaaggtaa
ctctcgtagctaataggattacctcctttggctaggtcaagagtggctgtgatcttcacttga
caaagttccggtacattgtggacagcattctccaaaagactaagacacagttgctttggg
agttgctcagccattggtacagtattgtggtagatacaaaggtggttcttccaatgaaggat
aaatccttctgctgtgcctgtccatgaggatccatatttcgccgtagttaggtaaccaagcg
tggtggctgaactgatcttcgcacttgctgattccgtatagtgtttgacaactttacaaaaca
cttcttgcgcagttcgctctagcctggacagagaagggacaatcttcgtcgttaagactcg
tgcgaatagcaaaagatcacaaaatagcacatcggcaccgaccaacgattatttccaa
ggaaaaaaagaatgcttcactacaagaaattgtgtcatccctatacagagtcttgttactg
tgacagaaaattgatggaagatgtggcggattgcctttacactagccaacttgttcgacta
attgcagcttcttctgagaggcttcaccgagtaacgcgaagaacaccggtgtctcgtaca
tgctcgtcggtgaacgctcgtccaatgacacccccccactttgtatcaatatcccaacttggt
agtgaactggaatgatacatgcaatttcgcgccgcatcaacagccacgggcaccatcg
acgaatagactcggtcgagctggttgccctcgccgctgggctggcggccgtctatggcc
ctgcaaacgcgccagaaacgccgtcgaagccgtgtgcgagacaccgcggccggcc
gccggcgttgtggatacctcgcggaaaacttggccctcactgacagatgaggggcgga
cgttgacacttgaggggccgactcacccggcgcggcgttgacagatgaggggcaggc
tcgatttcggccggcgacgtggagctggccagcctcgcaaatcggcgaaaacgcctg
attttacgcgagtttcccacagatgatgtggacaagcctggggataagtgccctgcggta
ttgacacttgaggggcgcgactactgacagatgaggggcgcgatccttgacacttgag
gggcagagtgctgacagatgaggggcgcacctattgacatttgaggggctgtccacag

gcagaaaatccagcatttgcaagggtttccgcccgtttttcggccaccgctaacctgtctttt
aacctgcttttaaaccaatatttataaaccttgtttttaaccagggctgcgccctgtgcgcgt
gaccgcgcacgccgaaggggggtgccccccccttctcgaaccctcccggtcgagtgag
cgaggaagcaccagggaacagcacttatatattctgcttacacacgatgcctgaaaaa
acttcccttggggttatccacttatccacggggatatttttataattatttttttttatagttttagat
cttctttttttagagcgccttgtaggcctttatccatgctggttctagagaaggtgttgtgacaa
attgccctttcagtgtgacaaatcaccctcaaatgacagtcctgtctgtgacaaattgccct
taaccctgtgacaaattgccctcagaagaagctgttttttcacaaagttatccctgcttattg
actctttttttatttagtgtgacaatctaaaaacttgtcacacttcacatggatctgtcatggcg
gaaacagcggttatcaatcacaagaaacgtaaaaatagcccgcgaatcgtccagtca
aacgacctcactgaggcggcatatagtctctcccgggatcaaaaacgtatgctgtatctg
ttcgttgaccagatcagaaaatctgatggcaccctacaggaacatgacggtatctgcga
gatccatgttgctaaatatgctgaaatattcggattgacctctgcggaagccagtaaggat
atacggcaggcattgaagagtttcgcggggaaggaagtggttttttatcgccctgaagag
gatgccggcgatgaaaaaggctatgaatcttttccttggtttatcaaacgtgcgcacagtc
catccagagggctttacagtgtacatatcaacccatatctcattcccttctttatcgggttac
agaaccggtttacgcagtttcggcttagtgaaacaaaagaaatcaccaatccgtatgcc
atgcgtttatacgaatccctgtgtcagtatcgtaagccggatggctcaggcatcgtctctct
gaaaatcgactggatcatagagcgttaccagctgcctcaaagttaccagcgtatgcctg
acttccgccgccgcttcctgcaggtctgtgttaatgagatcaacagcagaactccaatgc
gcctctcatacattgagaaaaagaaaggccgccagacgactcatatcgtattttccttcc
gcgatatcacttccatgacgacaggatagtctgagggttatctgtcacagatttgagggtg
gttcgtcacatttgttctgacctactgagggtaatttgtcacagttttgctgtttccttcagcctg
catggattttctcatacttttgaactgtaatttttaaggaagccaaatttgagggcagtttgtc
acagttgatttccttctctttcccttcgtcatgtgacctgatatcgggggttagttcgtcatcatt
gatgagggttgattatcacagtttattactctgaattggctatccgcgtgtgtacctctacctg
gagtttttcccacggtggatatttcttcttgcgctgagcgtaagagctatctgacagaacag
ttcttctttgcttcctcgccagttcgctcgctatgctcggttacacggctgcggcgagcatca
cgtgctataaaaataattataatttaaatttttttaatataaatatataaattaaaaatagaaa
gtaaaaaaagaaattaaagaaaaaatagtttttgttttccgaagatgtaaaagactctag
ggggatcgccaacaaatactacctttttaccttgctcttcctgctctcaggtattaatgccga
attgtttcatcttgtctgtgtagaagaccacacacgaaaatcctgtgattttacattttacttat
cgttaatcgaatgtatatctatttaatctgcttttcttgtctaataaatatatatgtaaagtacgct
ttttgttgaaatttttttaaaccttttgtttatttttttttttcttcattccgtaactcttctaccttctttatttact
ttctaaaatccaaatacaaaacataaaaataaataaacacagagtaaattcccaaatta
ttccatcattaaaagatacgaggcgcgtgtaagttacaggcaagcgatcctagtacactc
tatatttttttatgcctcggtaatgattttcatttttttttttccacctagcggatgactctttttttttctta
gcgattggcattatcacataatgaattatacattatataaagtaatgtgatttcttcgaagaa
tatactaaaaaatgagcaggcaagataaacgaaggcaaagatgacagagcagaaa
gccctagtaaagcgtattacaaatgaaaccaagattcagattgcgatctctttaaagggt
ggtcccctagcgatagagcactcgatcttcccagaaaaagaggcagaagcagtagca
gaacaggccacacaatcgcaagtgattaacgtccacacaggtatagggtttctggacc
atatgatacatgctctggccaagcattccggctggtcgctaatcgttgagtgcattggtgac

ttacacatagacgaccatcacaccactgaagactgcgggattgctctcggtcaagctttt
aaagaggccctaggggccgtgcgtggagtaaaaaggtttggatcaggatttgcgccttt
ggatgaggcactttccagagcggtggtagatctttcgaacaggccgtacgcagttgtcg
aacttggtttgcaaagggagaaagtaggagatctctcttgcgagatgatcccgcattttctt
gaaagctttgcagaggctagcagaattaccctccacgttgattgtctgcgaggcaagaa
tgatcatcaccgtagtgagagtgcgttcaaggctcttgcggttgccataagagaagccac
ctcgcccaatggtaccaacgatgttccctccaccaaaggtgttcttatgtagttttacacag
gagtctggacttgacgctagtgataataagtgactgaggtatgtgctcttcttatctccttttgt
agtgttgctcttattttaaacaactttgcggttttttgatgactttgcgattttgttgttgctttgcagt
aaattgcaagatttaataaaaaaacgcaaagcaatgattaaaggatgttcagaatgaa
actcatggaaacacttaaccagtgcataaacgctggtcatgaaatgacgaaggctatc
gccattgcacagtttaatgatgacagcccggaagcgaggaaaataacccggcgctgg
agaataggtgaagcagcggatttagttggggtttcttctcaggctatcagagatgccgag
aaagcagggcgactaccgcacccggatatggaaattcgaggacgggttgagcaacg
tgttggttatacaattgaacaaattaatcatatgcgtgatgtgtttggtacgcgattgcgacg
tgctgaagacgtatttccaccggtgatcggggttgctgcccataaaggtggcgtttacaa
aacctcagtttctgttcatcttgctcaggatctggctctgaaggggctacgtgttttgctcgtg
gaaggtaacgacccccagggaacagcctcaatgtatcacggatgggtaccagatcttc
atattcatgcagaagacactctcctgcctttctatcttggggaaaaggacgatgtcacttat
gcaataaagcccacttgctggccggggcttgacattattccttcctgtctggctctgcaccg
tattgaaactgagttaatgggcaaatttgatgaaggtaaactgcccaccgatccacacct
gatgctccgactggccattgaaactgttgctcatgactatgatgtcatagttattgacagcg
cgcctaacctgggtatcggcacgattaatgtcgtatgtgctgctgatgtgctgattgttccca
cgcctgctgagttgtttgactacacctccgcactgcagttttcgatatgcttcgtgatctgctc
aagaacgttgatcttaaagggttcgagcctgatgtacgtattttgcttaccaaatacagca
atagcaatggctctcagtccccgtggatggaggagcaaattcgggatgcctggggaag
catggttctaaaaaatgttgtacgtgaaacggatgaagttggtaaaggtcagatccggat
gagaactgtttttgaacaggccattgatcaacgctcttcaactggtgcctggagaaatgct
ctttctatttgggaacctgtctgcaatgaaattttcgatcgtctgattaaaccacgctgggag
attagataatgaagcgtgcgcctgttattccaaaacatacgctcaatactcaaccggttga
agatacttcgttatcgacaccagctgccccgatggtggattcgttaattgcgcgcgtagga
gtaatggctcgcggtaatgccattactttgcctgtatgtggtcgggatgtgaagtttactcttg
aagtgctccggggtgatagtgttgagaagacctctcgggtatggtcaggtaatgaacgtg
accaggagctgcttactgaggacgcactggatgatctcatcccttcttttctactgactggt
caacagacaccggcgttcggtcgaagagtatctggtgtcatagaaattgccgatggga
gtcgccgtcgtaaagctgctgcacttaccgaaagtgattatcgtgttctggttggcgagctg
gatgatgagcagatggctgcattatccagattgggtaacgattatcgcccaacaagtgct
tatgaacgtggtcagcgttatgcaagccgattgcagaatgaatttgctggaaatatttctgc
gctggctgatgcggaaaatatttcacgtaagattattacccgctgtatcaacaccgccaa
attgcctaaatcagttgttgctctttttttctcaccccggtgaactatctgcccggtcaggtgat
gcacttcaaaaagcctttacagataaagaggaattacttaagcagcaggcatctaacct
tcatgagcagaaaaaagctggggtgatatttgaagctgaagaagttatcactcttttaact
tctgtgcttaaaacgtcatctgcatcaagaactagtttaagctcacgacatcagtttgctcct

ggagcgacagtattgtataagggcgataaaatggtgcttaacctggacaggtctcgtgtt
ccaactgagtgtatagagaaaattgaggccattcttaaggaacttgaaaagccagcac
cctgatgcgacctcgttttagtctacgtttatctgtctttacttaatgtcctttgttacaggccag
aaagcataactggcctgaatattctctctgggcccactgttccacttgtatcgtcggtctgat
aatcagactgggaccacggtcccactcgtatcgtcggtctgattattagtctgggaccac
ggtcccactcgtatcgtcggtctgattattagtctgggaccacggtcccactcgtatcgtcg
gtctgataatcagactgggaccacggtcccactcgtatcgtcggtctgattattagtctggg
accatggtcccactcgtatcgtcggtctgattattagtctgggaccacggtcccactcgtat
cgtcggtctgattattagtctggaaccacggtcccactcgtatcgtcggtctgattattagtct
gggaccacggtcccactcgtatcgtcggtctgattattagtctgggaccacgatcccactc
gtgttgtcggtctgattatcggtctgggaccacggtcccacttgtattgtcgatcagactatc
agcgtgagactacgattccatcaatgcctgtcaagggcaagtattgacatgtcgtcgtaa
cctgtagaacggagtaacctcggtgtgcggttgtatgcctgctgtggattgctgctgtgtcct
gcttatccacaacattttgcgcacggttatgtggacaaaatacctggttacccaggccgtg
ccggcacgttaaccgggctgcatccgatgcaagtgtgtcgctgtcgacgagctcgcga
gctcggacatgaggttgccccgtattcagtgtcgctgatttgtattgtctgaagttgtttttacg
ttaagttgatgcagatcaattaatacgatacctgcgtcataattgattatttgacgtggtttga
tggcctccacgcacgttgtgatatgtagatgataatcattatcactttacgggtcctttccggt
gatccgacaggttacggggcggcgacctcgcgggttttcgctatttatgaaaattttccggt
ttaaggcgtttccgttcttcttcgtcataacttaatgtttttatttaaaataccctctgaaaagaa
aggaaacgacaggtgctgaaagcgagctttttggcctctgtcgtttcctttctctgtttttgtcc
gtggaatgaacaatggaagtccgagctcatcgctaataacttcgtatagcatacattata
cgaagttatattcgatgcggccgcaaggggttcgcgtcagcgggtgttggcgggtgtcg
gggctggcttaactatgcggcatcagagcagattgtactgagagtgcaccatatgcggt
gtgaaataccacacagatgcgtaaggagaaaataccgcatcaggcgccattcgccatt
cagctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagct
ggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagt
cacgacgttgtaaaacgacggccagtgaattgtaatacgactcactatagggcgaattc
gagctcggtacccggggatcctctagagtcgacctgcaggcatgcaagcttgagtattct
atagtctcacctaaatagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttat
ccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggt
gcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgg
gaaacctgtcgtgccagctgcattaatgaatcggccaacgcgaaccccttgcggccgc
ccgggccgtcgaccaattctcatgtttgacagcttatcatcgaatttctgccattcatccgctt
attatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaa
aaattacgccccgccctgccactcatcgcagtactgttgtaattcattaagcattctgccga
catggaagccatcacaaacggcatgatgaacctgaatcgccagcggcatcagcacct
tgtcgccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatatt
ggccacgtttaaatcaaaactggtgaaactcacccagggattggctgagacgaaaaa
catattctcaataaacccctttagggaaataggccaggttttcaccgtaacacgccacatct
tgcgaatatatgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatgaa
aacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactatcccatatcacc
agctcaccgtctttcattgccatacgaaattccggatgagcattcatcaggcgggcaaga

atgtgaataaaggccggataaaacttgtgcttatttttctttacggtctttaaaaaggccgta
atatccagctgaacggtctggttataggtacattgagcaactgactgaaatgcctcaaaa
tgttctttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttag
cttccttagctcctgaaaatctcgataactcaaaaaatacgcccggtagtgatcttatttcat
tatggtgaaagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagttggcc
cagggcttcccggtatcaacagggacaccaggatttatttattctgcgaagtgatcttccgt
cacaggtatttattcgcgataagctcatggagcggcgtaaccgtcgcacaggaaggac
agagaaagcgcggatctgggaagtgacggacagaacggtcaggacctggattgggg
aggcggttgccgccgctgctgctgacggtgtgacgttctctgttccggtcacaccacatac
gttccgccattcctatgcgatgcacatgctgtatgccggtataccgctgaaagttctgcaa
agcctgatgggacataagtccatcagttcaacggaagtctacacgaaggtttttgcgctg
gatgtggctgcccggcaccgggtgcagtttgcgatgccggagtctgatgcggttgcgatg
ctgaaacaattatcctgagaataaatgccttggcctttatatggaaatgtggaactgagtg
gatatgctgtttttgtctgttaaacagagaagctggctgttatccactgagaagcgaacga
aacagtcgggaaaatctcccattatcgtagagatccgcattattaatctcaggagcctgtg
tagcgtttataggaagtagtgttctgtcatgatgcctgcaagcggtaacgaaaacgatttg
aatatgccttcaggaacaatagaaatcttcgtgcggtgttacgttgaagtggagcggatta
tgtcagcaatggacagaacaacctaatgaacacagaaccatgatgtggtctgtcctttta
cagccagtagtgctcgccgcagtcgagcgacagggcgaagcccat

*Table 6-2: **Supplementary Table S2**. Oligonucleotides used in this study.*

Name	Sequence (5' to 3')	Description	Use for Gibson
FcpCt_HASP1p F	GAGCCTACATCCTTCTGCAACAAG CTGCAATCATACAGTGAATGTAACT TTCGAATTGAC	Forward for HASP1 promoter with FcpC terminator homology	pC-T-YFP
HASP1 p_CBDAS R	GTCAATTCGAAAGTTACATTCACTG TATGATTGCAGCTTGTTGCAGAAG GATGTAGGCTC	Reverse for HASP1 promoter with C homology	pC-T-YFP
HASP1p_CBDAS F	TCGAAAAGATCCATCACGGCGAGC GCCACCATGAAGTGCTCCACCTTT TCCTTTTGGTTC	Forward for C with HASP1 promoter homology	pC-T-YFP
CBDASCt_Thr R	TTCTCCCTTACGGGAGCCACGGGG GACGAGGTGGCGGTGGCGTGGGA GAGGGGGGATGCT	Reverse for C with Thrombin homology	pC-T-YFP
CBDASCt_Thr F	AGCATCCCCCCTCTCCCACGCCAC CGCCACCTCGTCCCCCGTGGCTCC CGTAAGGGAGAA	Forward for Thrombin, yfp and HA tag with C homology	pC-T-YFP
HA_FcpAt R		Reverse for Thrombin, yfp and HA tag with	pC-T-YFP

	CGGACCGGCGGTGTTGGTCGGCG TCGGTTATTAGGCGTAGTCCGGGA CATCGTAGGGGTA	FcpA terminator homology	
HA_FcpAt F	TACCCCTACGATGTCCCGGACTAC GCCTAATAACCGACGCCGACCAAC ACCGCCGGTCCG	Forward for FcpA terminator with HA tag homology	pC-T-YFP
FcpAt-Frg1 R	GGAAGATCTATATTACCCTGTGAA GACGAGCTAGTGTTATTCCTGACT GTGAAACCAAAG	Reverse for FcpA terminator with vector (backbones) homology	pC-T-YFP
FcpAt-Frg1 F	ATAACACTAGCTCGTCTTCACAGG GTAATATAGATCTTCCGCTGCATAA CCCTGCTTCGG	Forward primer used to amplify half of the backbone with FcpA terminator homology	pC-T-YFP
D630R	TTTGCAAACCAAGTTCGACAACTG CGTACGGCCTGTTCGAAAGATCTA CCACCGCTCTGG	Reverse primer used to amplify half of the backbone with the other half of backbones homology	pC-T-YFP, pmCherry-T-YFP, pmCherry & YFP
D631F	GGGCCGTGCGTGGAGTAAAAAGG TTTGGATCAGGATTTGCGCCTTTG GATGAGGCACTTT	Forward primer used to amplify half of the backbone with the other half of backbones homology	pC-T-YFP, pmCheryy-T-YFP, pmCherry & YFP

D631R	ATGGGCTTCGCCCTGTCGCTCGACTGCGGCGAGCACTACTGGCTGTAAAAGGACAGACCA	Reverse primer used to amplify half of the backbone with FcpC promoter homology	pC-T-YFP
D635F	TGGTCTGTCCTTTTACAGCCAGTAGTGCTCGCCGCAGTCGAGCGACAGGGCGAAGCCCATGAGCACAAGAGGTGACAAAA	Forward for FcpC promoter, Nat and FcpC terminator with half of the backbone homology	Gibson assembly of pC-T-YFP
FcpCt_HASP1p R	GTCAATTCGAAAGTTACATTCACTGTATGATTGCAGCTTGTTGCAGAAGGATGTAGGCTC	Reverse for FcpC promoter, Nat and FcpC terminator with HASP1 promoter homology	Gibson assembly of pC-T-YFP
mCherry-40SRPS8p R	TGATGATGGCCATGTTATCCTCCTCGCCCTTGCTCACCATGGTATTCTATTCTCTGATTC	Reverse for 40SRPS8p promoter with mCherry homology	Gibson assembly of pmCherry & YFP
40SRPS8p-mCherry F	GCGTTGATCTTGCACCGAAGGAATCAGAGAATAGAATACCATGGTGAGCAAGGGCGAGGA	Forward for mCherry with 40SRPS8p promoter homology	Gibson assembly of pmCherry & YFP
40SRPS8p-FcpAt R	TGATGATGGCCATGTTATCCTCCTCGCCCTTGCTCACCATGGTATTCTATTCTCTGATTC	Reverse for FcpA terminator with 40SRPS8p promoter homology	Gibson assembly of pmCherry & YFP

FcpAt-40SRPS8p F	CTTTGGTTTCACAGTCAGGAATAACACTAGCTCGTCTTCACCCTGCGATAGACCTTTTCC	Forward for 40SRPS8p promoter with FcpA terminator homology	Gibson assembly ofpmCherry & YFP
mCherry-thr-YFP F	GGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGCTCGTCCCCCGTGGCTCCCG	Forward for thrombin, yfp with mCherry homology	Gibson assembly of pmCherry-T-YFP
mCherry-HASP1p R	TGATGATGGCCATGTTATCCTCCTCGCCCTTGCTCACCATGGTGGCGCTCGCCGTGATGG	Reverse for HASP1p promoter with mCherry homology	Gibson assembly of pmCherry-T-YFP
HASP1p-mCherry F 2	TCGAAAAGATCCATCACGGCGAGCGCCACCATGGTGAGCAAGGGCGAGGAGGATAACATG	Forward for mCherry with HASP1p promoter homology	Gibson assembly of pmCherry-T-YFP
Thr-YFP-mCherry R 2	TTCTCCCTTACGGGAGCCACGGGGGACGAGCTTGTACAGCTCGTCCATGCCGCCGGTGGA	Reverse for mCherry with thrombin, yfp homology	Gibson assembly of pmCherry-T-YFP

Table 6-3: ***Supplementary Table S3.*** *Plasmids used in this study.*

Plasmid	Description	Reference
pPtGE30	*P. tricornutum* expression vector, FcpD promoter and FcpA terminator driving *Sh ble*	(Slattery *et al*., 2018)
pEV	*P. tricornutum* expression vector, FcpC promoter and FcpC terminator driving Ntc resistance gene.	This study
pC-T-YFP	*P. tricornutum* expression vector, HASP1 promoter and FcpA terminator driving yfp containing a 3xHA tag in 3' linked in 5' with a thrombin to the native enzyme cbdas	This study
pmCherry & YFP	*P. tricornutum* expression vector, HASP1 promoter and FcpA terminator driving yfp containing a 3xHA tag in 3'. Also, 40SRPS8 promoter and FcpA terminator drinving mCherry.	This study
pmCherry-T-YFP	*P. tricornutum* expression vector, HASP1 promoter and FcpA terminator driving yfp containing a 3xHA tag in 3' linked in 5' with a thrombin cleavage sequence to mCherry.	This study
HASP1	Plasmid containing HASP1 promoter and FcpA terminator driving yfp containing a 3xHA tag in 3' linked	From our lab

	in 5’ with a thrombin to an endogenous secretion signal peptide (SP)	
DMI5+6	Plasmid that contains the optimized Nat resistance gene sequence	From our lab
CBDAS	Plasmid containing the optimized native CBDAS sequence.	From our lab
pTA-Mob	Conjugative plasmid that contains the genes to produce the proteins that will establish the bacterial conjugation for the transformation. These proteins will be essential for the formation of the physical link for the transfer of the cargo plasmid of interest from *E. coli* to *P. tricornutum* cells.	(Strand *et al.*, 2014)

*Table 6-4: **Supplementary Table S4**. Prothrombin and the thrombin-like enzymes sequence blasted with P.t genome*

Name	References (sequences)
Human	THRB_HUMAN Prothrombin OS=Homo sapiens OX=9606 GN=F2 PE=1 SV=2 [Source:UniProtKB/TrEMBL;Acc:P00734]
Bovin	THRB_BOVIN Prothrombin OS=Bos taurus OX=9913 GN=F2 PE=1 SV=2 [Source:UniProtKB/TrEMBL;Acc:P00735]
Mouse	THRB_MOUSE Prothrombin OS=Mus musculus OX=10090 GN=F2 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P19221]
Pig	THRB_PIG Prothrombin OS=Sus scrofa OX=9823 GN=F2 PE=2 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q19AZ8]
Ponab	THRB_PONAB Prothrombin OS=Pongo abelii OX=9601 GN=F2 PE=2 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q5R537]
Rat	THRB_RAT Prothrombin OS=Rattus norvegicus OX=10116 GN=F2 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P18292]
Salsa	THRB_SALSA Thrombin (Fragments) OS=Salmo salar OX=8030 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P84122]
J49772	Phatr3_J49772.p1-pep-chromosome:ASM15095v2:24:251021:252769:-1 gene:Phatr3_J49772-transcript:Phatr3_J49772.t1-gene_biotype:protein_coding transcript_biotype:protein_coding-description:Predicted-protein [Source:UniProtKB/TrEMBL;Acc:B7GBU8]
J45961	Phatr3_J45961.p1-pep-chromosome: ASM15095v2:8:783171:784660:1 gene:Phatr3_J45961-transcript:Phatr3_J45961.t1-gene_biotype:protein_coding transcript_biotype:protein_coding-description:Predicted-protein [Source:UniProtKB/TrEMBL;Acc:B7FZA8]
J37254	Phatr3_J37254.p1pep-chromosome:ASM15095v2:12:525221:526327:-1 gene:Phatr3_J37254-transcript:Phatr3_J37254.t1-gene_biotype:protein_coding transcript_biotype:protein_coding-description:Predicted-protein [Source:UniProtKB/TrEMBL;Acc:B7G2K1]
J40462	Phatr3_J40462.p1-pep-chromosome:ASM15095v2:23:467199:468962:1 gene:Phatr3_J40462-transcript:Phatr3_J40462.t1-gene_biotype:protein_coding transcript_biotype:protein_coding-description:Predicted-protein [Source:UniProtKB/TrEMBL;Acc:B7GBH8]

J7679	Phatr3_J7679.p1-pep-chromosome:ASM15095v2:23:460233:462413:1 gene:Phatr3_J7679-transcript:Phatr3_J7679.t1-gene_biotype:protein_coding transcript_biotype:protein_coding-description:Predicted-protein [Source:UniProtKB/TrEMBL;Acc:B7GBH6]
J54319	Phatr3_J54319.p1-pep-chromosome:ASM15095v2:5:598945:600798:-1 gene:Phatr3_J54319-transcript:Phatr3_J54319.t1-gene_biotype:protein_coding transcript_biotype:protein_coding-description:Predicted-protein [Source:UniProtKB/TrEMBL;Acc:B7FW25]
J49602	Phatr3_J49602.p1-pep-chromosome:ASM15095v2:23:167354:169030:-1 gene:Phatr3_J49602-transcript:Phatr3_J49602.t1-gene_biotype:protein_coding transcript_biotype:protein_coding-description:Predicted-protein [Source:UniProtKB/TrEMBL;Acc:B7GB65]
Gloha	VAE1_GLOHA Thrombin-like enzyme agkihpin-1 (Fragment) OS=Gloydius halys OX=8714 PE=2 SV=1 [Source:UniProtKB/TrEMBL;Acc: N0AAE6]
	VAE2_GLOHA Thrombin-like enzyme agkihpin-2 (Fragment) OS=Gloydius halys OX=8714 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:N0A5N4]
	VSPI_GLOHA Thrombin-like enzyme AhV_TL-I OS=Gloydius halys OX=8714 PE=1 SV=2 [Source:UniProtKB/TrEMBL;Acc:I4CHP3]
Crodu	VSP13_CRODU Thrombin-like enzyme gyroxin B1.3 OS=Crotalus durissus terrificus OX=8732 PE=2 SV=1 [Source:UniProtKB/TrEMBL;Acc:B0FXM1]
	VSP14_CRODU Thrombin-like enzyme gyroxin B1.4 OS=Crotalus durissus terrificus OX=8732 PE=2 SV=1 [Source:UniProtKB/TrEMBL;Acc:B0FXM2]
	VSP17_CRODU Thrombin-like enzyme gyroxin B1.7 OS=Crotalus durissus terrificus OX=8732 PE=2 SV=1 [Source:UniProtKB/TrEMBL;Acc:B0FXM3]
	VSP21_CRODU Thrombin-like enzyme gyroxin B2.1 OS=Crotalus durissus terrificus OX=8732 PE=2 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q58G94]
Botan	VSP1_BOTAN Thrombin-like enzyme TLBan (Fragments) OS=Bothrocophias andianus OX=1144373 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P0DJG3]
Botjr	VSP1_BOTJR Thrombin-like enzyme BjussuSP-1 OS=Bothrops jararacussu OX=8726 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q2PQJ3]
Botma	VSP1_BOTMA Thrombin-like enzyme TLBm OS=Bothrops marajoensis OX=157554 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P0DJE9]
Crodo	VSP1_CRODO Thrombin-like enzyme collinein-1 OS=Crotalus durissus collilineatus OX=221569 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:A0A0S4FKT4]

Deiac	VSP1_DEIAC Thrombin-like enzyme acutobin OS=Deinagkistrodon acutus OX=36307 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q9I8X2]
	VSPA_DEIAC Thrombin-like enzyme acutin (Fragment) OS=Deinagkistrodon acutus OX=36307 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q9YGS1]
Globl	VSP1_GLOBL Thrombin-like enzyme halystase OS=Gloydius blomhoffii OX=242054 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P81176]
Globr	VSP1_GLOBR Thrombin-like enzyme kangshuanmei OS=Gloydius brevicaudus OX=259325 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P85109]
Glous	VSP1_GLOUS Thrombin-like enzyme calobin-1 OS=Gloydius ussuriensis OX=35671 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q91053]
	VSPE2_GLOUS Thrombin-like enzyme CPI-enzyme 2 OS=Gloydius ussuriensis OX=35671 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:O42207]
Proel	VSP1_PROEL Thrombin-like enzyme elegaxobin-1 OS=Protobothrops elegans OX=88086 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P84788]
	VSP2_PROEL Thrombin-like enzyme elegaxobin-2 OS=Protobothrops elegans OX=88086 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P84787]
Profl	VSP1_PROFL Thrombin-like enzyme flavoxobin OS=Protobothrops flavoviridis OX=88087 GN=TLF1 PE=1 SV=2 [Source:UniProtKB/TrEMBL;Acc:P05620]
Triab	VSP1_TRIAB Thrombin-like enzyme 1 OS=Trimeresurus albolabris OX=8765 PE=2 SV=1 [Source:UniProtKB/TrEMBL;Acc:A7LAC6]
	VSP2_TRIAB Thrombin-like enzyme 2 OS=Trimeresurus albolabris OX=8765 PE=2 SV=1 [Source:UniProtKB/TrEMBL;Acc:A7LAC7]
	VSPC_TRIAB Thrombin-like enzyme chitribrisin OS=Trimeresurus albolabris OX=8765 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P0DJF6]
Agkbi	VSP2_AGKBI Thrombin-like enzyme bilineobin OS=Agkistrodon bilineatus OX=8718 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q9PSN3]
Agkco	VSP2_AGKCO Thrombin-like enzyme contortrixobin OS=Agkistrodon contortrix contortrix OX=8713 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P82981]
Botja	VSP2_BOTJA Thrombin-like enzyme KN-BJ 2 OS=Bothrops jararaca OX=8724 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:O13069]
	VSPA_BOTJA Thrombin-like enzyme bothrombin OS=Bothrops jararaca OX=8724 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P81661]

Cerce	VSPA_CERCE Thrombin-like enzyme cerastotin (Fragment) OS=Cerastes cerastes OX=8697 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P81038]
	VSPP_CERCE Thrombin-like enzyme cerastocytin OS=Cerastes cerastes OX=8697 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q7SYF1]
Botal	VSPBH_BOTAL Thrombin-like enzyme bhalternin OS=Bothrops alternatus OX=64174 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P0CG03]
Croad	VSPCR_CROAD Thrombin-like enzyme crotalase OS=Crotalus adamanteus OX=8729 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:F8S114]
Calrh	VSPF1_CALRH Thrombin-like enzyme ancrod OS=Calloselasma rhodostoma OX=8717 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P26324]
	VSPF2_CALRH Thrombin-like enzyme ancrod-2 OS=Calloselasma rhodostoma OX=8717 PE=2 SV=1 [Source:UniProtKB/TrEMBL;Acc:P47797]
Botat	VSPF_BOTAT Thrombin-like enzyme batroxobin OS=Bothrops atrox OX=8725 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P04971]
Lacmu	VSPF_LACMU Thrombin-like enzyme gyroxin analog OS=Lachesis muta muta OX=8753 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P33589]
Glosh	VSPGL_GLOSH Thrombin-like enzyme gloshedobin OS=Gloydius shedaoensis OX=88083 PE=1 SV=2 [Source:UniProtKB/TrEMBL;Acc:P0C5B4]
Botas	VSPL_BOTAS Thrombin-like enzyme asperase OS=Bothrops asper OX=8722 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q072L6]
Botlc	VSPL_BOTLC Thrombin-like enzyme leucurobin OS=Bothrops leucurus OX=157295 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:P0DJ86]
Trist	VSPST_TRIST Thrombin-like enzyme stejnobin OS=Trimeresurus stejnegeri OX=39682 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q8AY81
Glosa	VSPSX_GLOSA Thrombin-like enzyme saxthrombin OS=Gloydius saxatilis OX=92067 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:Q7SZE1]
Botba	VSP_BOTBA Thrombin-like enzyme barnettobin (Fragment) OS=Bothrops barnetti OX=1051630 PE=1 SV=1 [Source:UniProtKB/TrEMBL;Acc:K4LLQ2]

Table 6-5: ***Supplementary Table S5.*** *BLAST analysis of candidate thrombin-like proteins across diatoms phylum.*

Organism	S	Hits (n)	Species taxonomy*
Bacillariophyta	n.a.	113	n.a.
.Bacillariophycidae	n.a.	59	n.a.
..Naviculales	n.a.	24	n.a.
...Naviculaceae	n.a.	20	n.a.
Seminavis robusta	119	16	Bacillariophyceae; Bacillariophycidae; Naviculales;
Mayamaea pseudoterrestris	107	1	Bacillariophyceae; Bacillariophycidae; Naviculales;
Fistulifera solaris	94.4	3	Bacillariophyceae; Bacillariophycidae; Naviculales
Phaeodactylum tricornutum CCAP 1055/1	104	4	Bacillariophyceae; Bacillariophycidae; Naviculales
Cylindrotheca closterium	111	17	Bacillariophyceae; Bacillariophycidae; Bacillariales;
Fragilariopsis cylindrus CCMP1102	94.4	5	Bacillariophyceae; Bacillariophycidae; Bacillariales;
Nitzschia inconspicua	95.5	11	Bacillariophyceae; Bacillariophycidae; Bacillariales;
Pseudo-nitzschia multistriata	87.8	2	Bacillariophyceae; Bacillariophycidae; Bacillariales
Thalassiosira pseudonana CCMP1335	102	20	Coscinodiscophyceae ; Thalassiosirophycidae ; Thalassiosirales
Chaetoceros tenuissimus	106	14	Coscinodiscophyceae; Chaetocerotophycidae; Chaetocerotales
Fragilaria crotonensis	103	9	Fragilariophyceae; Fragilariophycidae; Fragilariales

Skeletonema marinoi	101	7	Coscinodiscophyceae ; Thalassiosirophycidae ; Thalassiosirales
Thalassiosira oceanica	85.1	4	Coscinodiscophyceae ; Thalassiosirophycidae ; Thalassiosirales

S : score, n.a. non applicable. * from *Schoch CL, et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford). 2020: baaa062*. PubMed: 32761142 PMC: PMC7408187.

*Table 6-6: **Supplementary Table S6**. Trypsin domain catalytic site residues, docking score of LVPRGS and predicted H-bonds with docked ligands.*

	Site 1	Score	H-bonds
Human 1PPB	Arg35, Glu39, Leu40, Leu41, Cys42, His57, Cys58, Tyr60A, Trp60D, Lys60F, Phe60H, Arg73, Leu99, Trp141, Gly142, Asn143, Trp148, Gly150, Gln151, Pro152, Asp189, Ala190, Cys191, Glu192, Gly193, Ser195, Val213, Ser214, Trp215, Gly216, Glu217, Gly219, Cys220, Gly226	−10.87	Asp189 (n=2), Tyr60 (n=2), Lys60, Gly193, His57*, Gly219
J49772	His146, Asn193, Asp196, Asp283, Ala284, Cys285, Gln286, Asp288, Ser289, Val307, Ser308, Trp309, Gly310, Gly312, Cys313, Gly320, Val321	−9.6	Arg154, Thr129, Ser289* (n=2), Ser308, His146*, Gln286, Gly312, Asn193
J45961	His58, Asn59, Gly60, His75, Gln118, Asn120, Glu121, Asp122, Asp123, Phe124, Asn198, Leu199, Thr200, Tyr201, His204, Thr212, Asp221, Ala222, Cys223, Ser224, Tyr225, Asp226, Ser227, Val247, Ser248, Trp249, Gly250, Glu251, Leu252, Cys253, Phe258, Gly260, Val261, Asn262,	−10.11	Tyr225, Ser227*, Asp221(n=2), Leu 252, Asp122*
J37254	Phe169, Asp216, Ala217, Cys218, Gly219, Gly220, Asp221, Ser222, Val242, Ser243, Trp244, Gly245, Gly247, Cys248, Gly253, Val254, Tyr255	−9.55	Gly247 (n=2), Asp216 (n=2), Gly245, Ser222*(n=2), Ser243
J40462	Val204, Val223, Thr225, Thr249, Leu250, Asn251, Lys254, Gly255, Ala256, Cys257, Gln258, Gly259, Asp260, Ser261, Leu265, Val277, Ser278, Phe279, Ala281, Glu282, Cys283, Ala284, Ala292, Val293, Glu294	−8.92	Ala281, Glu282, Phe279 (n=2), Gly280, His125*(n=2), Ser261*

J7679	Gln121, Val122, Ile123, Glu124, Ser125, Lys126, Ile127, Ser149, His166, Ile205, Asp208, Val242, Met243, Gly244, Val261, Asn262, Ser263, Tyr275, Ala289, Gly290, Glu295, Gly296, Ser297, Cys298, Asp301, Ser302, Gly303, Val317, Ser318, Phe319, Gly320, Val321, Gly322, Cys323, Gly324, Asp325, Phe328, Pro329, Gly330, Val331, Tyr332	nd	na
J54319	Pro57, Leu58, Ser59, Pro60, Val61, Ile62, Ala63, Glu64, Ser65, Cys87, His102, Cys103, Gly141, Asn142, Thr143, Glu144, Tyr212, Gln215, Asp232, Ser233, Cys234, Gln235, Gly236, Ser238, Thr253, Ser254, Trp255, Gly256, Ile257, Gly258, Tyr264, Pro265, Gly266, Val267, Tyr268	−10.6	Glu64, Gly256, Trp255, Ser233
J49602	Pro47, Leu48, Thr49, Asn50, Asp51, Ala52, Lys53, Arg54, Val55, Glu56, Ser57, Arg58, Gly83, Ile84, Cys85, His100, Ser102, Gln104, Tyr137, Asp138, Ile139, His140, Ser141, Gln142, Thr143, Ser144, Asp147, Val179, Met180, THR198, Leu200, Gln211, Tyr212, Asn214, Val215, Thr225, Arg232, Asp233, Thr234, Cys235, Gly237, Asp238, Ser239, Gly240, Gly241, Pro242, Val255, Ser256, Phe257, Gly258, Val259, Gly268, Val269, Asn270, Ala316, Glu317, Gln364	nd	na

Na: not applicable, nd: not detected. The score is in kcal/mol. * catalytic residue.

*Table 6-7: **Supplementary Table S7**. Sequencing analysis of the DNA plasmid rescue from P. tricornutum clones.*

Clone	Reference position	Description	Features affected	Effect
	511	C instead of G	pPTGE30 backbone	Unknown
mCherry & YFP	12994	A instead of G	FcpC promoter	The Nat gene will be expressed. The level of expression of FcpC promoter might be affected.
	13013,13082	G instead of A		
	16362	G instead of A	40SRPS8 promoter	The mCherry gene will be expressed. The level of expression of 40SRPS8 promoter might be affected.
	16445	T instead of C		
mCherry-T-YFP	12,994	A instead of G	FcpC promoter	The Nat gene will be expressed. The level of expression of FcpC promoter might be affected.
	13013, 13082	G instead of A		
C-T-YFP	1255	T instead of C	pPTGE30 backbone	Unknown
	1309,5030,7353	A instead of G		
	10850	T instead of C	FcpC promoter	The Nat gene will be expressed. The level of expression of FcpC promoter might be affected.
	13013, 13082	G instead of A		
	13429	T instead of C		
	17590	A instead of C	FcpA terminator	Unknown
	17433-17492	Deletion 2xHA	HA tag	HA tag shorter

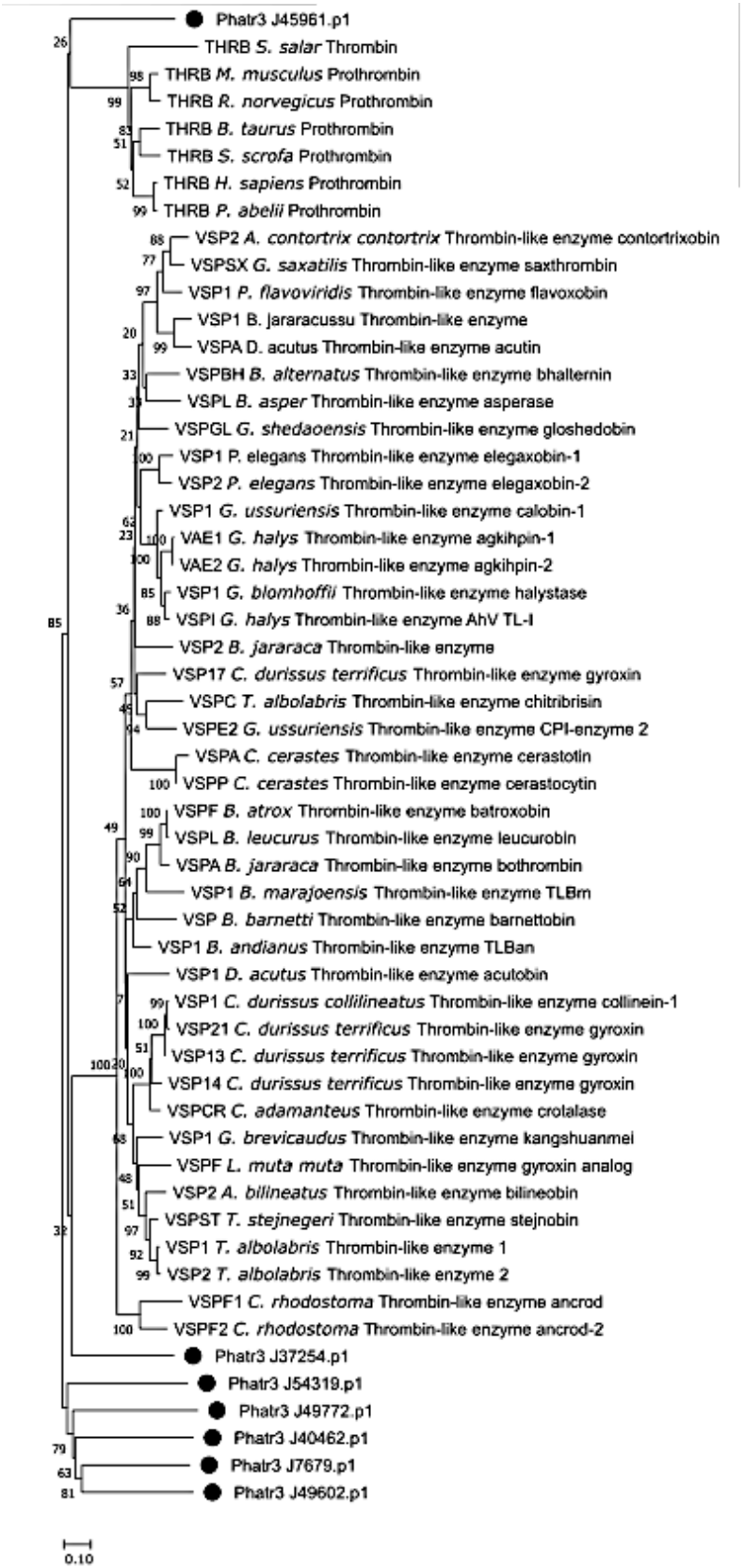


Figure 6-1: ***Supplementary Fig. S1. Phylogenetic tree of several thrombin, thrombin-like and our potential thrombin amino acid sequences found in P. tricornutum sequence database (Phatr).***

Figure 6-2: ***Supplementary Fig. S2. Phylogenetic tree of P. tricornutum candidates (red) with thrombin (blue) and 100 non-thrombin (black) serine proteases.***

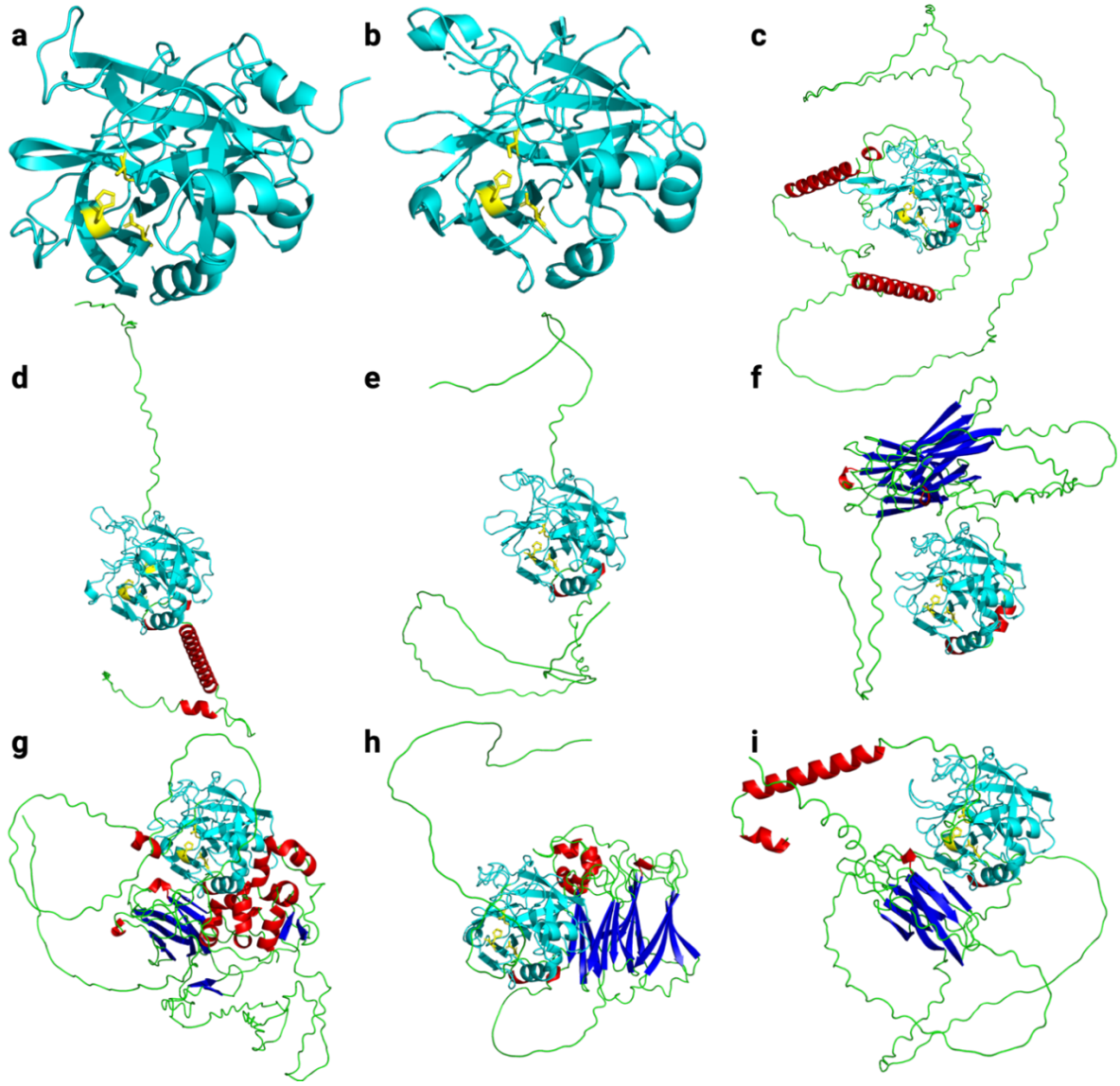


Figure 6-3: ***Supplementary Fig. S3. Predicted structures of the 7 putative thrombins of P. tricornutum.***

The trypsin-domain is colored in cyan blue, b-barrels in blue and α-helices in red. ***a*** *Human Trypsin-domain (1PPB),* ***b*** *Bovin Trypsin-domain (3PMB),* ***c*** *PhatrJ49772,* ***d*** *PhatrJ45961,* ***e*** *PhatrJ37254,* ***f*** *PhatrJ40462,* ***g*** *PhatrJ7679,* ***h*** *PhatrJ54319,* ***i*** *PhatrJ49602.*

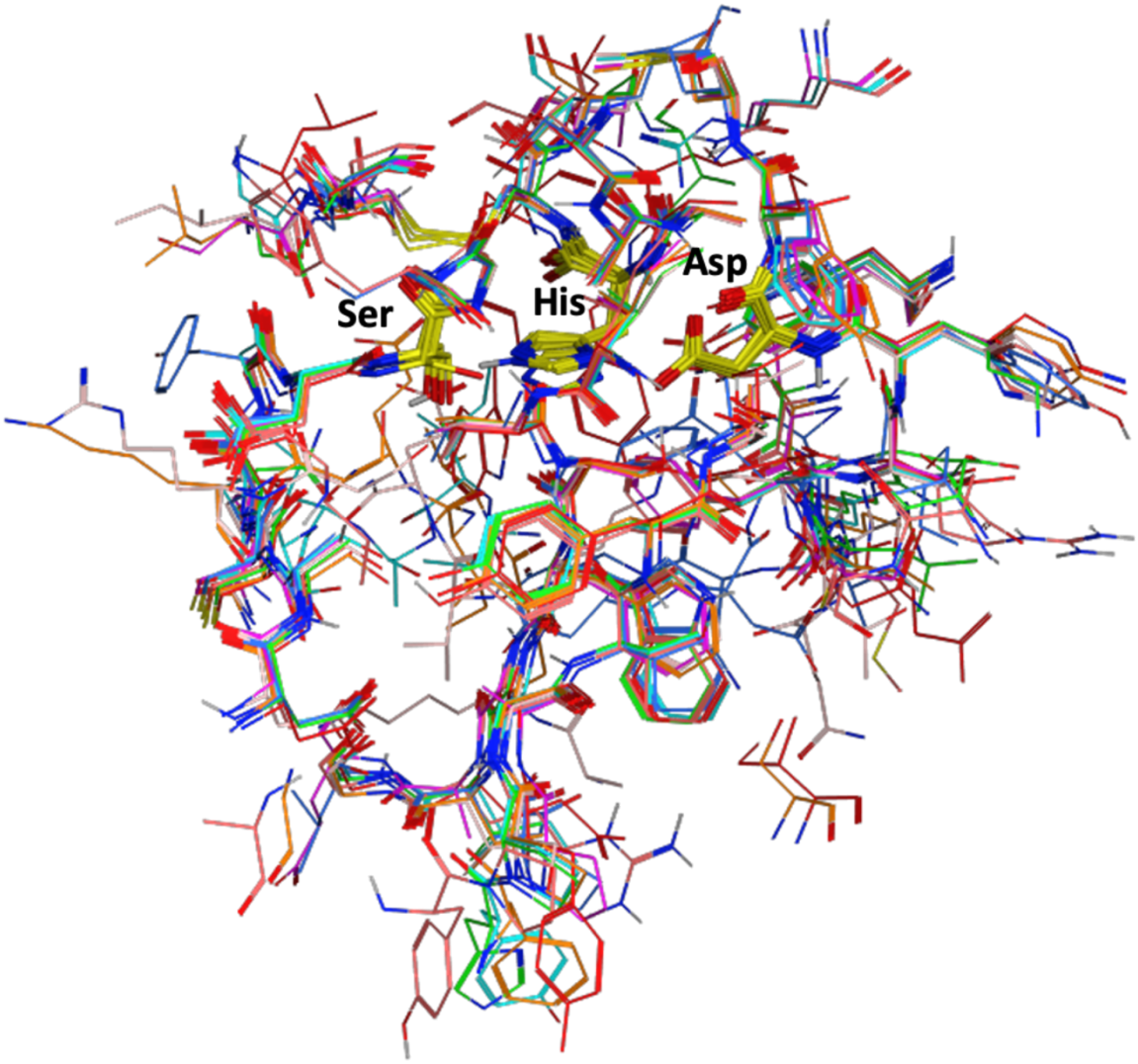


Figure 6-4: ***Supplementary Fig. S4. Superposed active site residues of putative thrombin from P. tricornutum.***

Residues are shown as thin sticks. The catalytic triad His, Asp, Ser is shown in yellow. Human ß-thrombin is red, J49772 is green, J45961 is blue, J37254 is salmon, J40462 is bright pink, J7679 is orange, J54319 is turquoise, J49602 is light pink. The proposed conserved triad of catalytic residues of serine proteases is shown in yellow.

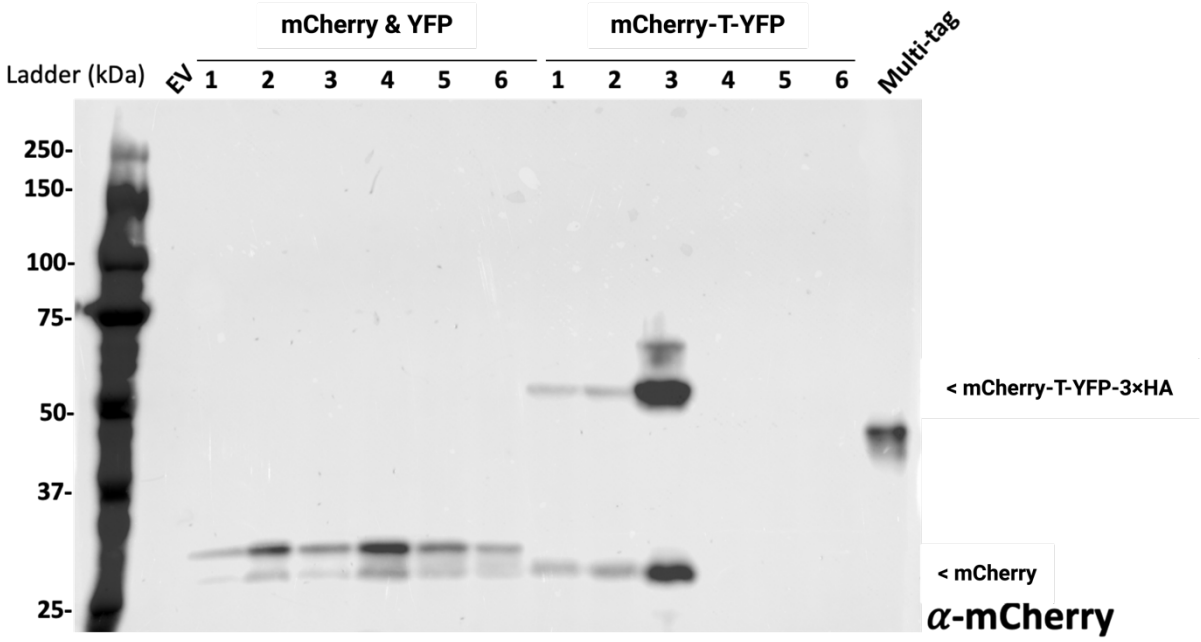


Figure 6-5: ***Supplementary Fig. S5. Western blot analysis using anti-mCherry antibody on total protein extracts.***

Extracts from transcongugants harboring mCherry &YFP and mCherry-T-YFP cassettes. EV: used as negative control, Multi-Tag is purified protein with expected size of 40 kDa used as a positive control. Expected protein sizes from mCherry & YFP and mCherry-T-YFP were 26.5 and 58 kDa respectively.

Appendix III, Research Paper (Co-authorship)

7 Multifactorial interaction and influence of culture conditions on yellow fluorescent protein production in *Phaeodactylum tricornutum*

Arun Augustine [a], Anis Messaabi [a], Elisa Fantino [a b], Natacha Merindol [a], Fatma Meddeb-Mouelhi [a b], Isabel Desgagné-Penix [a b]

[a] Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières, 3351 boulevard des Forges, Trois-Rivières, QC G9A 5H7, Canada
[b] Plant Biology Research Group, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada

Received 28 November 2024, Revised 2 March 2025, Accepted 2 March 2025, Available online 3 March 2025, Version of Record 5 March 2025.

https://doi.org/10.1016/j.biortech.2025.132336

7.1 Highlights

- Bioengineered P. tricornutum expressing yellow fluorescent protein extrachromosomally.
- A model was developed to predict culture parameters on YFP production in P. tricornutum.
- Optimized conditions increased YFP fluorescence and yield by 4.2x and 1.8x, respectively.
- Inhibition of the ubiquitin–proteasome system increased YFP levels by 50–150 % after 4–8 h.
- Cycloheximide treatment significantly reduced YFP stability at 20 h.

7.2 Keywords

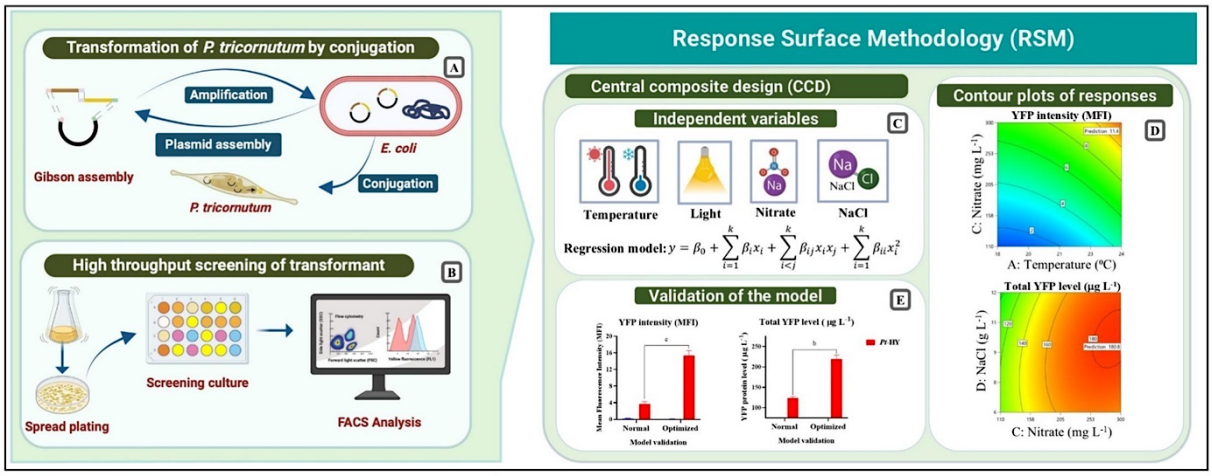
Microalgae; Genetic engineering; Response surface methodology; Diatoms; Protein inhibitors

7.3 Abstract

Phaeodactylum tricornutum is a promising host for light-driven synthesis of heterologous proteins. However, the marine cold-water environment and alkaline-acidic pH shifts in the culture, necessitated by the diatom's growth requirements. In this study, we analyzed the

influence of growth condition on maturation and dynamics of the yellow fluorescent protein (YFP) in episomal-transformant *P. tricornutum*. A mathematical model was developed to detect the parameters that affect biomass and YFP production. Optimized conditions increased YFP mean fluorescence intensity (MFI) per cell by 4.2-fold (3.6 ± 0.6 to 15.4 ± 1.1) and total protein levels in the culture by 1.8-fold (123 ± 4 to 219 ± 9 µg L^{-1}), without affecting biomass. YFP stability studies in P. tricornutum showed that the ubiquitin–proteasome system contributes the degradation of the recombinant protein, whereas newly synthesized YFP remains stable for up to 12 h. This optimization provides insights into the fluorescent protein-based heterologous production in diatoms.

7.4 Graphical abstract



7.5 Introduction

Diatoms are major primary producers of the ocean and are considered natural cell factories for high value metabolites, such as pigments, polyunsaturated fatty acids, and terpenoids (Chen et al., 2022). Recent advances in genome sequencing, annotation (Filloramo et al., 2021), and editing tools have increased our ability to harness their potential. *P. tricornutum* is a cold-water obligate photoautotroph with a maximum growth rate at 20 ± 2°C (De Martino et al., 2011). This diatom has been investigated for diverse biotechnological applications, particularly for the light-driven production of biopharmaceutical specialized metabolites (Fantino et al., 2024). It exhibits a poorly silicified frustules, which reduces the transformation barrier (Karas et al., 2015), and a rapid growth rate under diverse environmental conditions. However, the efficacy of this heterologous production platform is strongly influenced by environmental

factors such as temperature, salinity, light intensity, CO_2 supplementation, and silicate concentration (De Martino et al., 2011, Gómez-Loredo et al., 2016, Liang et al., 2014).

To overcome photoinhibition and achieve maximum biomass production, the light intensity must be optimized based on the cell density, culture volume, and surface area (Sivakaminathan et al., 2018). Nitrogen, as a building block for biomolecules synthesis, is the most important macronutrient for algal growth. Higher levels increase algal growth and metabolism, while nitrate-deficiency leads to reduced biosynthesis and increases recycling of amino acids, proteins, and nucleic acids (Alipanah et al., 2015). Phosphate is also essential to the diatom growth, as a key component of membrane phospholipids. A deficiency causes the downregulation of processes like photosynthesis, ribosome and nucleic acid biosynthesis, and of nitrogen assimilation (Alipanah et al., 2018). The concentration of micronutrients such as vitamins and metal ions in the medium play a crucial role in determining the growth of algae as well (Ouyang et al., 2018).

In addition to environmental and culture factors, recombinant protein turnover and maturation process also depend on the nature of the promoter and the cell division rate (Guerra et al., 2022). The recently reported endogenous highly abundant secreted protein 1 (HASP1) promoter strongly and constitutively expresses heterologous genes of interest in *P. tricornutum* at all growth phases, especially in the stationary phase when the culture reaches maximum biomass (Erdene-Ochir et al., 2019), attracting the interest of researchers (Slattery et al., 2022).

Several fluorescent proteins (FPs) have been used in the diatom to gain insight of various cellular dynamics and processes, facilitate screening, monitor recombinant gene expression, and measure protein turnover (Diamond et al., 2023, Fantino et al., 2024). FPs must undergo an autocatalytic process of maturation, which includes various folding and post-translational modifications that result in the formation of a functional chromophore and fluorescence emission (Guerra et al., 2022). The kinetics of this process are influenced by environmental factors such as molecular oxygen, temperature, pH, and chloride ion (Cl^-), and impacts on the accuracy of dynamic measurements and *in vivo* fluorescence intensity.

Systematic characterization and screening of *P. tricornutum* culture parameters for optimal growth and FP accumulation would be highly valuable to help researchers select the optimum conditions for heterologous protein production To monitor the impact of culture parameters on growth, fluorescence formation, and heterologous protein production in *P. tricornutum*, we

used transconjugants lines carrying the yellow FP (YFP), and applied a strategic Plackett-Burman design (PBD) as well as a central composite design (CCD) (Rajendran et al., 2008). YFP has been frequently used in *P. tricornutum* studies, displays a high pKa, extinction coefficient, and relatively bright fluorescence compared to other FPs (Marter et al., 2020), and mathematical models such as the response surface methodology (RSM) allows for the evaluation of the interactions between multiple factors influencing protein expression.

This study mainly focused on: (1) generating a transgenic line of *P. tricornutum* with an episomal vector carrying the *yfp* gene under the HASP1 promoter and evaluating the expression, protein production, and chromophore formation under varying culture conditions; (2) screening for factors impacting YFP production; (3) optimizing the culture conditions for maximum growth, maturation, and total YFP production; and (4) assessing the stability of YFP *in silico* and in *P. tricornutum* cells using proteasome and protein synthesis inhibitors.

7.6 Materials and methods

Diatom growth and culture conditions

P. tricornutum Bohlin strain (Culture Collection of Algae and Protozoa CCAP 1055/1) was maintained in an algal growth chamber at 18 ± 1°C, 75 µmol m^{-2} s^{-1}, pH 8.0 ± 0.5 with a 16:8h light/dark photoperiod and continuous orbital shaking at 130 rpm (Innova S44i, Eppendorf) in modified L1 media with 150 mg L^{-1} of $NaNO_3$ and 10 mg L^{-1} of $NaH_2PO_4H_2O$ (Slattery et al., 2022). All algal experiments were conducted in 250 mL conical flasks in batch culture mode. The optical density was measured at OD_{680} using a microplate reader (Synergy H1, BioTek), and the dry cell weight of *P. tricornutum* was calculated from the regression graph built against OD_{680} versus the dried algal biomass (40 mL culture) and expressed in terms of g L^{-1}.

Generation of *P. tricornutum* transconjugants

Plasmid DNA constructs were constructed using Gibson assembly as described by Diamond et al. (2023). Empty vector (EV) carrying the antibiotic selection marker gene *N-acetyltransferase* (*nat*) against nourseothricin (NTC) under the FcpC promoter and terminator in pPtGE30 was used as a negative control. The other episome carried *YFP* under the HASP1 promoter and FcpA terminator by replacing the UR3 region of the EV backbone (Fig. 6-1 and (see supplementary material)) tagged with 3HA at the C-terminal.

Transgenic lines of *P. tricornutum* were obtained by conjugation as described by Karas et al. (2015). The positive transformants carrying the recombinant plasmid started to form colonies on the NTC plate after a two-week period. Positive transformants were screened, as described in (Diamond et al., 2023). YFP fluorescence from *P. tricornutum*-positive transformants was assessed using the microplate reader. The cell culture (200 µL) was measured for fluorescence in black 96 well plates at Ex/Em wavelengths of 500/531 nm (n = 3). Strains containing plasmids YFP (16,168 bp) and EV (14,530 bp) are referred to here as *Pt*-HY and *Pt*-EV, respectively.

Fluorescence-activated cell sorting of bioengineered *P. tricornutum*

P. tricornutum transconjugants were sorted by YFP production using a FACSMelody™ (BD bioscience, USA). The screened *P. tricornutum* transconjugants were grown in L1 liquid media with appropriate antibiotics in log phase (14 days) and used for the first sorting. Culture preparation, gating strategy and singlet selection were performed as in Diamond et al. (2023). YFP was analyzed on a 527/32 nm bandpass filter channel. Sorted cells were collected in an Eppendorf tube containing 0.5 mL L1 media without antibiotics, centrifuged at a low speed of 3500 *g*for 10 min. The resulting cell pellet was resuspended in fresh L1 medium supplemented with NTC 100 µg mL^{-1} and chloramphenicol 25 µg mL^{-1}. For the initial screening, 30 positive transconjugants per construct were screened based on antibiotic selection (*Pt*-EV) and YFP fluorescence (*Pt*-HY). *P. tricornutum* transconjugants with higher mean YFP fluorescence intensity (MFI) were used for full-scale experiments. *Pt*-EV transformants were used as negative controls to deduce spillover YFP autofluorescence signals during the screening process. (Diamond et al., 2023).

Whole-cell YFP fluorescence measurements by flow cytometry

Cytometry was performed on a Cytomics FC 500 flow cytometer (Beckman Coulter) equipped with Argon (488 nm) and HeNe (633 nm) lasers. Only the *P. tricornutum* population with

homogenous cell size and chlorophyll autofluorescence level was considered in the analysis to exclude cell debris and other background noise. A least 20,000 gated cells were analyzed for each sample. The inherent properties of the cells and other characteristics of heterologous fluorescence properties (MFI, % of YFP^+ cells, and cell size variations in the culture) were analyzed using FlowJo version 10.8 software (BD Biosciences, USA) (Diamond et al., 2023).

Subcellular localization of YFP

Transconjugant *P. tricornutum* cells were visualized using a confocal microscope under a 40x objective lenses (Leica M165 FC Fluorescent Stereo Microscope). The *in vivo* localization of YFP in the transformed cells was performed in an 8-day-old culture. YFP, chlorophyll, and propidium iodide (PI, Sigma Aldrich) were excited at 488 nm and detected using YFP (520–535 nm), PI (610–620 nm), and RFP (625–720 nm) filters, respectively, as in Diamond et al. (2023).

Total RNA extraction and qPCR-based quantification of *YFP*

Total RNA was isolated from the transformed *P. tricornutum* biomass (15 mL of 8-day-old culture). The algal pellets were collected by centrifugation at 4000 rpm for 5 min at 4 °C and flash frozen in liquid nitrogen, was stored at −80 °C for later extraction using the RNeasy Plant Mini Kit, according to the manufacturer's protocol (QIAGEN). RNA quality and quantity were confirmed using an Implen NanoPhotometer®. Real-time qPCR analysis was performed using the Luna® Universal One-Step RT-qPCR Kit according to the manufacturer's instructions (New England Biolabs). Amplification and relative quantification were performed in triplicate using antibiotic (*nat*) and *YFP* specific primers (see supplementary material), on a CFX-connected Real-Time PCR System (Bio-Rad). Primer specificity was verified by analyzing the melting curves. Data processing and statistical analysis were performed using the CFX Maestro software (Bio-Rad, USA). The relative expression of all genes was calculated by the $2^{-\Delta\Delta CT}$ method (Pfaffl, 2001) using *P. tricornutum Tuba* (Tubulin alpha) and *EF1α* (Elongation factor alpha) as the endogenous reference genes (Siaut et al., 2007).

Protein extraction

P. tricornutum culture was centrifuged (20 mL) at 4000 g for 10 min at 4 °C. The pellet was re-suspended in solubilization buffer at a ratio of 100 mg mL^{-1} (0.75 mM SDS, 10 % glycerol,

51.4 mM Tris HCl pH 8, and 0.02 mM EDTA) of the wet weight and stored samples in −80 °C for further analysis. The pellet was thawed in ice water, and extraction was performed as described in Fantino et al. (2024), and proteins stored at −80 °C for further analysis. Total protein amount was quantified using the RC DC Protein Assay Kit (Bio-Rad) and a linear calibration curve was constructed using Bovine Serum Albumin (BSA) as a protein standard.

Elisa-based YFP quantification

Quantification of recombinant YFP in *P. tricornutum* total protein extract was performed using a GFP sandwich ELISA (# Cat. No. AKR 121, Cell Biolabs, USA), according to the manufacturer's instructions starting from, 100 µL of total protein per well (1 µg) extracted from the transformed *P. tricornutum* cell lysate. The color intensity of the solution in each well was read in the microplate reader (OD_{450}). The quantification of YFP in the total protein samples was determined using a standard curve prepared with GFP standard 0–2 ng mL^{-1} included in the kit.

Western blot detection of YFP in *P. Tricornutum*t ransconjugants

The soluble protein fraction (40 µg) of the cell lysate from transgenic *P. tricornutum* cells was loaded onto 12 % SDS-PAGE for western blot detection of heterologous proteins. The transfer into a polyvinylidene fluoride membrane and immunoblotting were performed as described in Fantino et al. (2024). The BLUelf prestained protein ladder was used as the molecular weight marker for comparison (GeneDireX, Cat. No. PM008-0500), and 10 ng of positive-tag control (protein of GenScript Multiple tag M0101) loaded onto the SDS-PAGE gel as a positive control. In all blotting analyses, two gels with identical samples were run in parallel, one used for western blotting and the other for the loading control after staining with Coomassie brilliant blue.

Plackett-Burman design screening for significant culture conditions for biomass and YFP production

The experiments were performed using L1 medium as the basal medium supplemented with NTC at a concentration of 100 µg·mL^{-1}. Exponentially growing *P. tricornutum* cells were used

as the inoculum with initial OD_{680} adjusted to 0.05. The significant variables affecting biomass (OD_{680}) and YFP production (flow cytometry) in *P. tricornutum* were screened using a PB design with Design-Expert® software V 7.0 (Stat-Ease Inc.). Seven culture variables, including temperature, light, $NaNO_3$, NaH_2PO_4, NaCl, metal solution, and vitamin solution, were tested in their upper (+) and lower limits (−) for their influence on biomass or YFP production. A set of 12 different experiments was carried out in triplicate (see supplementary material), and the mean values of biomass, MFI of YFP, and % of YFP^+ cells were considered as a response for screening the significant culture variables (Plackett & Burman, 1946). Responses were analyzed by ANOVA, and variables with a confidence level above 95 % ($p < 0.05$) were considered to have significant effects on biomass, YFP intensity, and % of YFP^+ cells (Rajendran et al., 2008).

Optimization of culture conditions for biomass and YFP production by response surface Methodology

A 2^4 factorial CCD-based RSM with 30 sets of experiments was used to optimize the level of the significant variables identified from the PB experiments, such as temperature (x_1), light (x_2) nitrate (x_3), and NaCl (x_4) at five different levels. Other parameters, which were less significant, such as phosphate, metal, and vitamin concentrations, were maintained at constant levels as per the standard conditions. The experiments were conducted using exponentially grown diatoms at initial $OD_{680}$of 0.05 in 250 mL Erlenmeyer flasks containing 50 mL of medium prepared according to the design shown in (see supplementary material).

The relationship between independent variables and responses was calculated using the second-order polynomial equation response (1).

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 \quad (1)$$

where *y* represents the predicted response (biomass, YFP MFI, % of YFP^+ cells, and total YFP level), β_0 is the model constant, β_i is the linear coefficient, β_{ii} is the quadratic coefficient, β_{ij} is the interaction coefficient, and *k* is the number of factors. The ANOVA for the experimental data and model coefficients were calculated using Design-Expert® software. To visualize the variable interactions and their impact on the response, 2D contour plots were generated.

Experimental validation of the model and statistical analysis

The statistical model was validated for biomass, YFP MFI, % of YFP^+ cells, and YFP level using the methods described earlier. Experimental values were subsequently compared with model-predicted values and standard culture conditions. All experiments were performed in triplicate (n = 3), and the results are described as mean ± standard deviation (SD), tested with a two-way ANOVA, in which different culture conditions were considered fixed factors and responses as dependent variables. Statistical analyses, such as ANOVA and Student's *t*-test, were performed using GraphPad Prism V8 software. The statistically difference between the control and experiment groups tested using Bonferroni's multiple comparison. A p-value < 0.05 was considered statistically significant, and shown as $p < 0.05$(a), $p < 0.005$(b), and $p < 0.0005$(c).

Protein stability analysis

***In silico* thermal stability analysis**

The thermal stability curve of the YFP was predicted using the protein structure-based melting temperature (*Tm)* prediction tools SCooP v-1.0 (Pucci et al., 2017). The folded structure of YFP was generated using a ColabFold from the amino acid sequence (Mirdita et al., 2022).

***In vivo* stability analysis of YFP in transformed *P. tricorn*utum by protein inhibitors**

Eight-day-old P. tricornutum cultures (100 mL) grown in optimized conditions with OD_{680} = 0.6 in 250 mL conical flasks were used for protein inhibitor treatment, as described by Im et al. (2024). Cultures with equal OD_{680} were treated with 50 µM of the proteasome inhibitor MG132 (MedChemExpress) dissolved in DMSO and/or 100 µM of the protein synthesis inhibitor cycloheximide (CHX) (Sigma-Aldrich) dissolved in water. DMSO was used as solvent control. The samples were harvested at the beginning of the light period (8 am), corresponding to the time of treatment initiation, and then every 4 h for the next 24 h. After the first harvest, the respective inhibitor solutions were added to the cultures (8 am). At each time point, 10 mL of each sample was collected and centrifuged at 4,000 *g* for 10 min at 4 °C and pellets resuspended in 250 µL of solubilization buffer. The pellet storage, protein extraction and western blotting were performed as described in 2.6 Protein extraction, 2.8 Western blot detection of YFP in except that the membrane was incubated at room temperature for 1 h with primary mouse YFP antibody (Cedarlane) 1:1000 dilution in TBS-T with 3 % BSA. For

immunoblotting of housekeeping proteins, *β*-actin antibody conjugated with HRP (BioLegend) at a 1:20,000 dilution in TBS-T was used. The densiometric analysis of immunoblot was done using ImageJ software v1.54 k. The effect of protein inhibitor treatment on cell viability were visualized by confocal microscopy by PI staining, added to the treated culture at a final concentration of 3 µg mL^{-1} and incubated at room temperature for 30 min (Fantino et al., 2024).

7.7 Results and discussion

Growth and YFP production of *P. tricornutum* transconjugants

In this study, we used two *P. tricornutum* transconjugant strains, *Pt*-EV (negative control) and *Pt*-HY strains harboring an episomal vector to produce YFP:3HA (Fig. 6-1A). To maintain high YFP expression, and to mitigate episomal instability and transgene silencing (Diamond et al., 2023), cultures underwent periodic fluorescence-based sorting. Guerra et al. (2022)showed that *in vivo* brightness and fluorescent protein maturation are tightly coupled with cell cycle gene expression, histone-like proteins, and DNA replication in yeast, suggesting that modeling the culture dynamics and quantification of fluorescence of individual cells requires synchronized sampling. As a result, biomass and fluorescence production were assessed at the end of the light period when *P. tricornutum* cells have completed division and exhibit peak fluorescence (Vartanian et al., 2009). We focused on optimizing growth and heterologous protein production by monitoring *P. tricornutum* cultures until they reached the stationary phase.

YFP fluorescence in *Pt*-HY was confirmed by confocal analysis (Fig. 6-1B and C), and western blotting (Fig. 6-1D). A preliminary one-factor-at-a-time analysis (OFAT) analysis assessed the impact of temperature (18, 21, and 24 °C with 75 µmol m^{-2} $s^{-1}$of light), and light intensity (50, 75, and 100 µmol m^{-2} s^{-1} at 18 °C) on growth and fluorescence production (see supplementary material). Cultures reached stationary phase after 15 days, regardless of the temperature or light conditions Temperature significantly affected early biomass accumulation. At 24 °C compared to 18 °C, growth was reduced at day 5 (p = 0.001; 33 % decrease) and day 8 (p = 0.003; 12 % decrease), but difference was no longer significant from day 10 (p = 0.18), suggesting adaptation mechanisms. The MFI, and % of YFP^+ cells in the culture reached a maximum at day 8 under different temperatures and light intensities (see supplementary

material). Then, YFP fluorescence decreased despite higher biomass. This might be due to flow cytometry single cell resolution analysis compared the microplate reader measuring accumulated fluorescence in the well, including proteins secreted in the extracellular media (Erdene-Ochir et al., 2019). Nonetheless, YFP fluorescence was enhanced at 24 °C, with a 13.0 ± 0.7 % increase in the % of YFP^+ cells ($p = 0.01$) and a 3.2 ± 0.1-fold increase YFP MFI ($p = 0.0001$), a variable that is proportional to the amount of YFP fluorescence intensity per cell. At that temperature with 75 µmol m^{-2} s^{-1} of light, more than 60 % of the cell population produced a high level of YFP fluorescence until the 15th day, consistently with reports of enhances YFP folding at higher maturation (Iizuka et al., 2011).

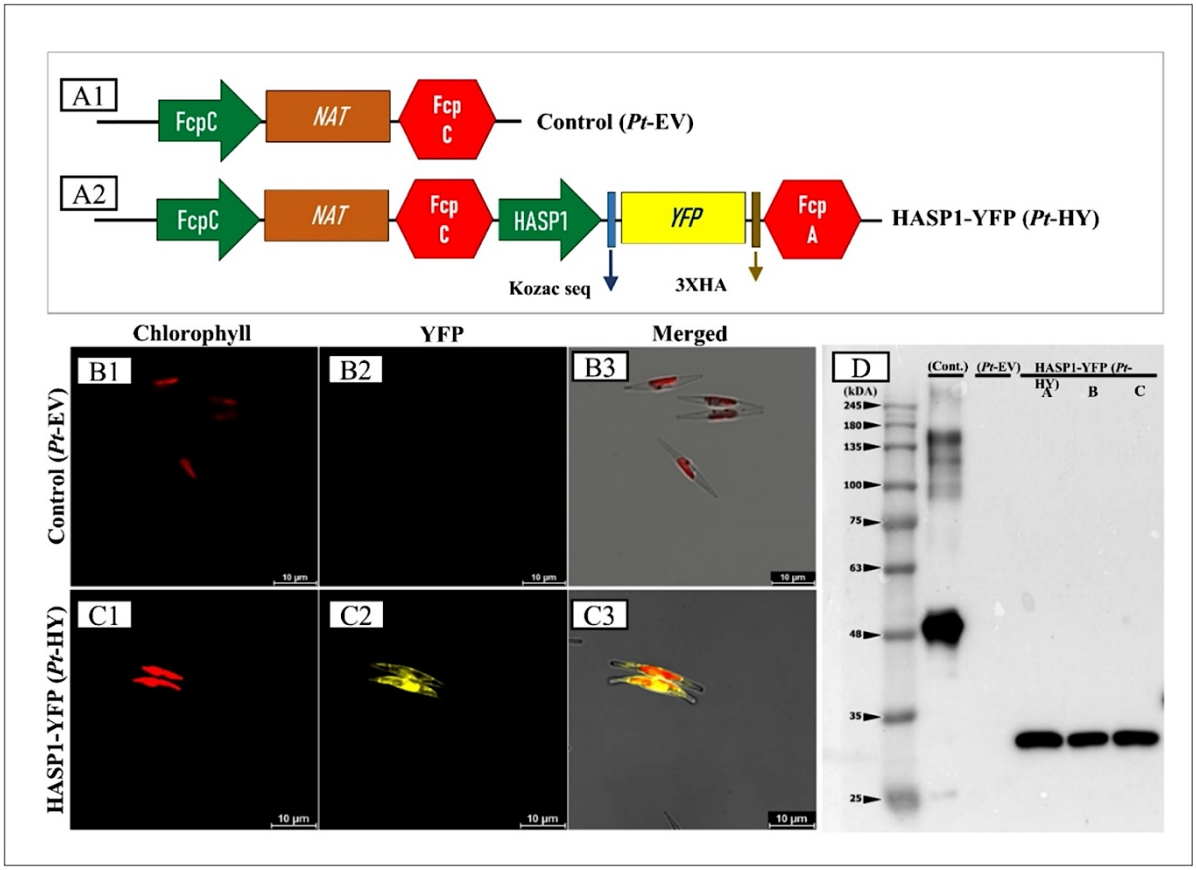


Figure 7-1: ***P. tricornutum episomal transconjugants accumulate yellow fluorescent protein.***

(A) Schematic overview of constitutive vector circuit design. (A1) Synthetic module showing the nat gene under the constitutive FcpC promoter and terminator (Pt-EV). (A2) The construct with the antibiotic-resistant gene and yfp gene under HASP1 constitutive promoter and FcpA terminator (Pt-HY). (B and C) In vivo subcellular localization of YFP in Pt-EV and Pt-HY, respectively, by confocal microscopy. (B1 and C1) Chlorophyll fluorescence is shown as red signal. (B2 and C2) YFP fluorescence is shown as yellow signal. (B3 and C3) The overlay of the two signals. The scale bars represent 10 µm. (D) Western blot analysis showing YFP detection (YFP:3HA; 32 kDa) in the transformed Pt-HY cells using an anti-HA antibody. Pt-EV was used as control.

Light intensity also influenced growth and fluorescence. Cultures grown with 50 compared to 75µmol m^{-2} s^{-1} of light showed a significant reduction in biomass at day 5 ($p = 0.007$; 28 % decrease) and the 8th day ($p = 0.02$; 8.3 % decrease). At day 8, these cultures exhibited showed

a 1.4 ± 0.1-fold increase in both MFI (p = 0.01) and a 7.8 ± 3.0 % increase in % of YFP^+ cells (p = 0.001) (see supplementary material). Overall, among the two physical parameters tested, only higher temperatures had a transient impact on biomass in the initial growth phases. Despite variations in growth, MFI and % of YFP^+ cells peaked at day 8 under all conditions, suggesting a stable fluorescence expression pattern. This time point was selected for further optimization experiments.

Screening for significant factors affecting biomass and YFP production

To optimize YFP production, we used a PB design to evaluate the impact of seven culture factors on biomass, MFI and % of YFP^+cells. Each factor was investigated experimentally at two levels, as per the PBD model (see supplementary material). The R^2(coefficient of determination) value determines how well the model fits the experimental data. The regression model showed a high fit for biomass production (R^2 = 0.97), meaning that up to 97 % of the variation could be explained by the model (Table 6-1). The adjusted R^2 (0.91) was in reasonable agreement with the predicted R^2 of 0.72 (difference < 0.2), and the F-value (17.46, p = 0.008) confirmed the model' significance. Similarly, MFI and % of YFP^+ had R^2 values of 0.97 and 0.93, respectively. The adjusted R^2 (0.93) for MFI agreed reasonably well with the predicted R^2 of 0.75 (Table 6-1). However, the adjusted R^2 (0.79) for the % of YFP^+ cells differed significantly from the predicted R^2 (0.36) (>0.2) (Table 6-1), suggesting that the model was too simplistic for this outcome. The adequate precision value (7.99 > 4) confirmed that the model provided a good signal-to-noise ratio for exploring the design space. The PB factorial screening revealed that biomass increased with higher nitrate, light intensity, and NaCl concentration, and decreasing temperature (see supplementary material). YFP fluorescence (MFI and % of YFP^+ cells) increased with higher temperature and nitrate concentration (see supplementary material). In contrast with the preliminary OFAT analysis, light was a significant factor for biomass but not for MFI or % of YFP^+ cells, which may be due to interactive effects with other medium components.

Table 7-1: ***Regression coefficient and statistical significance of PBD for biomass, MFI, and YFP+ cells in P. tricornutum.***

Regression Coefficients	Biomass (g L^{-1})	YFP intensity (MFI)	YFP^+ cells (%)
Model	[a]	[a]	[a]
Intercept	0.58	5.09	54.06
A-Temperature	−0.05[b]	2.56[b]	10.11[a]
B-Light	0.08[b]	0.14	0.42
C-Nitrate	0.08[b]	2.59[b]	17.16[a]
d-Phosphate	0.01	−0.83	−12.76[a]
E-NaCl	0.07[a]	−0.33	−3.81
F-Metal Solution	−0.04	0.73	−5.92
G-Vitamin	0.02	−0.30	6.48
R^2	0.96	0.97	0.92
Adjusted R^2	0.91	0.92	0.79
Predicted R^2	0.71	0.75	0.36
Adeq Precision	11.94	14.55	7.99

P value = < 0.0005[c], < 0.005[b], < 0.05[a].

In summary, temperature, and nitrate significantly affected YFP, while temperature, light, nitrate, and NaCl influenced biomass. Temperature, nitrate, and phosphate significantly modulated the % of YFP^+ cells. Thus, temperature, light, nitrate, and NaCl with significant effects on biomass and YFP production were selected for further optimization using CCD.

Mathematical modeling and process optimization for biomass and YFP production

RSM was used to evaluate interactions between multiple culture factors, with temperature (A), light (B), nitrate (C), and NaCl (D) chosen based on the PBD responses. In the CCD model, each variable was analyzed at five coded levels (−α, −1, 0, +1, +α) (see supplementary material). Second-order polynomial equations were obtained to describe biomass production, YFP intensity, % of YFP^+ cells, and total YFP level by applying multiple regression analysis to the experimental response data (see supplementary material). The CCD model was highly significant ($p < 0.0001$) for all responses, with a non-significant lack of fit (Table 6-2). The *F*-values for biomass, MFI, % of YFP^+cells, and total YFP were 29.82, 71.18, 88.75, and 16.57, respectively. Among the tested factors, nitrate had the highest regression coefficients (*β)* for

all responses (R1 = 0.08, R2 = 3.10, *R*3 = 11.75, and R4 = 35.73), which represents the mean change of dependent variable given a one-unit shift in an independent variable (Dayana Priyadharshini & Bakthavatsalam, 2016). Temperature had the second highest regression coefficient, while NaCl and light intensity had the lowest *β*-values for all responses (Table 6-2). This confirms that nitrate and temperature are the most influential factors for biomass, and YFP production, while NaCl and light had minimal impact, except for NaCl on total YFP yield. The quadratic terms A, B, C, and D and the interactive terms AB, AC, and BD impacted significantly biomass production (Table 6-2). The model terms A, AC, and BC affected MFI levels, while the % of YFP^+ cells were significantly modulated by B, C, D, AB, and AC. Interestingly, total YFP level measured by ELISA was influenced by C, D, and interactive terms (CD) (Table 6-2). Overall, these findings suggest that biomass and YFP production were significantly influenced by each independent variable, with noticeable interactions between factors further improving the responses.

Table 7-2: ***Regression coefficient and statistical significance of CCD for biomass and YFP production in P. tricornutum.***

Regression Coefficients	**Biomass ($g\ L^{-1}$)**	**YFP intensity (MFI)**	**YFP^+cells (%)**	**Total YFP level ($\mu g\ L^{-1}$)**
Intercept	0.83	5.03	67.53	164.94
A-Temperature	−0.06[c]	1.87[c]	6.99[c]	5.31
B-Light	−0.01	0.02	2.18[c]	−5.51
C-Nitrate	0.08[c]	3.10[c]	11.75[c]	35.73[c]
d-NaCl	−0.01	−0.01	−1.69[b]	−10.37[b]
AB	0.04[b]	0.24	−1.27	−1.75
AC	−0.05[c]	0.35[a]	−5.86[c]	−3.61
AD	0.01	0.02	−1.36[a]	5.18
BC	−0.01	0.37[a]	−0.10	−2.01
BD	−0.03[b]	0.16	−1.01	7.26
CD	0.01	0.25	−0.90	8.55[a]
A^2	−0.05[c]	0.82[c]	−0.94	0.62
B^2	−0.04[c]	−0.03	2.80[c]	1.16
C^2	−0.05[c]	0.15	−5.68[c]	−19.12[c]

Regression Coefficients	Biomass (g L^{-1})	YFP intensity (MFI)	YFP^+cells (%)	Total YFP level (µg L^{-1})
D^2	−0.02[b]	−0.04	2.49[c]	−9.69[b]
R^2	0.97	0.99	0.99	0.94
Adjusted R^2	0.93	0.97	0.98	0.88
Predicted R^2	0.83	0.92	0.93	0.70
Adeq Precision	18.29	31.12	38.4	16.89

P value = < 0.0005[c], < 0.005[b], < 0.05[a].

Mutual interactions between the significant factors

Interactive factors affecting biomass production in *P. tricornutum*

The RSM approach facilitated the identification of the optimal conditions for fluorescence intensity and heterologous protein yield in *P. tricornutum*. The contour plots in Fig. 6-2 were generated from the experimental responses of biomass as per the CCD model using a second-order polynomial Eq. (1). They were constructed using the variables in the experimental range, while the other factors were maintained at their optimum levels as per the standard culture conditions. Significant interactions for biomass production were observed between temperature (A) vs. light (B) (p = 0.001), A vs. nitrate (C) (p = 0.0001), and B vs. NaCl (D) (p = 0.001), (Fig. 6-2 A1, A2, A3). As observed following the PB-based screening, an increase in biomass (from 0.5-0.65 g L^{-1}) was observed with increasing nitrate and increasing NaCl concentration (3–18 g L^{-1}), with other medium components temperature (20.5 °C), light (70 µmol·m^{-2}·s^{-1}) and nitrate (188 mg L^{-1}) at central point (see supplementary material). However, the weight of NaCl on biomass production was not significant at higher temperature (24 °C), light intensity (78.2 µmol·m^{-2}·s^{-1}) and nitrate concentration (300 mg L^{-1}) (Fig.6-2 A3).

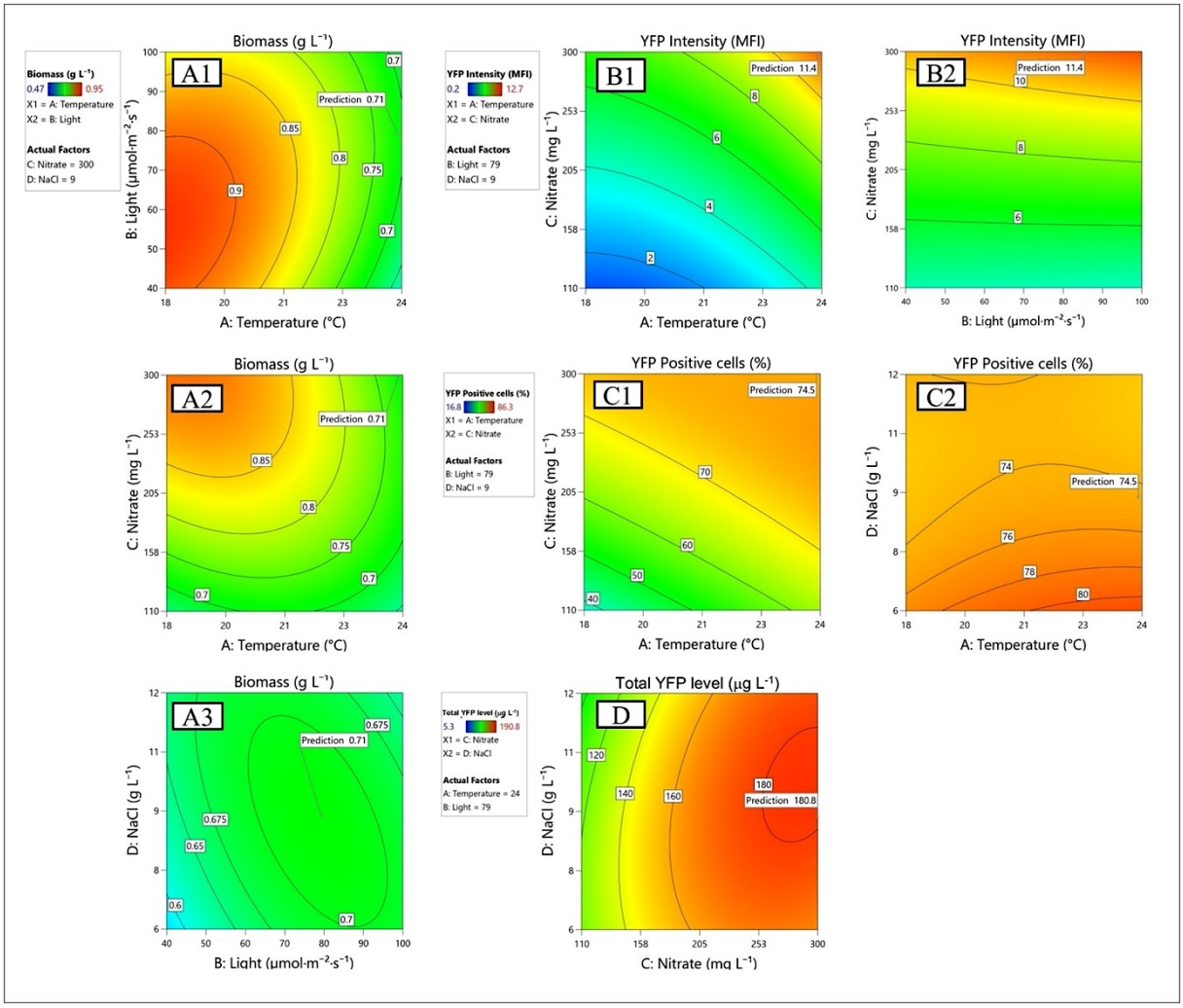


Figure 7-2: ***Contour surface plot of biomass and heterologous yellow fluorescent protein production.***

(A) The interaction of various culture factors on biomass production. (A1) Temperature (°C) vs light intensity (µmol m−2 s−1), (A2) temperature vs nitrate (mg L-1), (A3) light intensity and NaCl (g L-1). (B, C and D) The interaction of various culture conditions on the YFP mean fluorescent intensity (MFI) per cell, the % of YFP+ cells in the culture population using flow cytometry, and total heterologous protein by ELISA, respectively. (B1 and C1) The interaction of temperature vs nitrate, (B2) light vs nitrate, (C2) temperature vs NaCl, and (D) interaction of nitrate and NaCl on heterologous protein (YFP) level in P. tricornutumquantified using ELISA.

In agreement with the present observation, a salinity study by Liang et al. (2014) in *P. tricornutum* showed maximum photochemical efficiency of PSII (Fv/Fm) and maximal relative electron transport rate (rETRmax) at a salinity of 20–40 practical salinity units (PSU). As temperature and light intensity decreased, biomass increased. The present observation agrees with the previous factorial optimization studies of Nur et al. (2018) and Gómez-Loredo et al. (2016) in *P. tricornutum,* which showed that there was a significant interactive effect between temperature and light intensity on the level of major light-harvesting pigments such as chlorophyll *a* and fucoxanthin. The level of these pigments reduced with increasing light intensity above 120 µmol·m^{-2}·s^{-1}. In contrast, Wu et al. (2010) observed an increase in

biomass at higher temperatures and light intensity under high CO_2 levels in the media. This could be due to the downregulation of the carbon concentration mechanism, which serves as an energy sink. As a result, elevated CO_2 levels in the medium reduce non-photochemical quenching and the energy utilized by this process, which ultimately increases the photosynthetic carbon fixation rate. Overall, biomass production was driven by interactions between temperature, light, nitrate, and NaCl, with optimal growth under moderate temperature and light and increased NaCl and nitrate.

Interactive factors affecting YFP production in *P. Tricornutum*

In addition to biomass optimization, the main objective of this study was to maximize the MFI per cell, % of YFP^+ cells, and total YFP level. Temperature and nitrate played a significant role in YFP intensity, and significant interactions were observed between temperature and nitrate (p = 0.03), as well as light and nitrate (p = 0.02). As per the model prediction, YFP MFI could be increased from 2 to 11.4 by optimizing these significant factors, while keeping other factors at optimum levels (light 78.2 $\mu mol \cdot m^{-2} \cdot s^{-1}$ and NaCl 8.5 g L^{-1}), (Fig. 6-2B1 and B2). The interactions between temperature and nitrate (p = 0.0001) and between temperature and NaCl (p = 0.035) were also significant when monitoring for % of YFP^+ cells in the culture. The % of YFP^+cells increased from 40 to 74.5 % by optimizing the most significant factors in the model, such as temperature and nitrate, while keeping other factors at optimum (Fig. 6-2C1 and *C*2). Total YFP level was significantly modulated by nitrate (p = 0.0001) and NaCl (p = 0.003) concentrations, and a significant interaction was observed between nitrate and NaCl (p = 0.034). Total YFP reached a maximum level at high nitrate with NaCl concentration ranging from 8-11 g L^{-1}, while maintaining other factors at optimum level (temperature 24 °C and light 78.2 $\mu mol \cdot m^{-2} \cdot s^{-1}$) (Fig. 6-2D).

In the present study, of the four parameters tested, NaCl had the least significant impact, with the lowest regression coefficients for MFI (β = -0.01) and % of YFP^+ cells (β = -1.69), which may be due to the specific amino acid substitutions F64L, M153T, and V163A of YFP. These substitutions are known to improve maturation without affecting the excitation and emission spectra and display almost equivalent extinction coefficients and fluorescent quantum yields (Nagai et al., 2002). The protonation and Cl^- binding of YFP chromophore are the two main mechanisms that can decrease its absorbance. The above-mentioned mutations decrease the

pH sensitivity and remove the Cl^- sensitivity compared to the original YFP (Nagai et al., 2002). Similar observations on photostability and maturation with regards to temperature and Cl^- were reported in previous study of Guerra et al. (2022). Overall, nitrate, temperature, and light were the most interacting elements for MFI and % YFP^+cells in the culture. Nonetheless, NaCl content also demonstrated a beneficial interaction with nitrate and was a major factor in the total YFP yield.

Experimental validation of the model

Biomass production in *P. tricornutum*

The CCD model-predicted factors and responses for biomass, YFP MFI, % of YFP^+, and YFP production were experimentally validated using the predicted values listed in Table 6-3. Under optimized conditions (see supplementary material), there were no significant differences in the biomass of *Pt*-EVs compared to the standard condition. However, *Pt*-HY biomass production increased slightly but significantly (p = 0.04; Bonferroni's multiple comparisons test) from 0.77 ± 0.03 to 0.82 ± 0.06 g L^{-1}, compared to standard culture conditions (Fig. 6-2A, Table 6-3). This significant difference in biomass may be due to changes in the culture conditions that affect differently *Pt*-HY growth. A biomass production of 0.82 ± 0.06 g L^{-1} in the optimized media aligns with the CCD model predicted biomass of value of 0.71 g L^{-1}, validating the model for biomass production. Furthermore, the biomass yield on day 8 was agreed with the previously reported biomass in *P. tricornutum* by Villanova et al. (2021)0.87 ± 0.08 g L^{-1} (E10 medium) using different media composition. The work also showed that biomass increased by 2.1-fold (up to 1.86 ± 0.10 g L^{-1}) when changing from phototrophic to mixotrophic culture by the addition of glycerol and bicarbonate ($NaHCO_3$) to the E10 media. Enriched E10 medium increased biomass by a factor of 9 (11.55 ± 0.24) in an upscale photobioreactor (Villanova et al., 2021). The present experiment was performed in batch mode in a conical flask using carbon-free L1 media. A further option to boost the biomass and heterologous protein yield in a bioreactor will be to supplement the algal medium with both organic and inorganic carbon sources.

Table 7-3: ***Experimental validation of CCD model predicted conditions and responses.***

Conditions	**Standard conditions**	**Optimized conditions**	
	Standard level	**Coded levels**	**Actual levels**
Temperature (°C)	18	0.96	23.9
Light (µmol·m^{-2}·s^{-1})	75	0.27	78.2
Nitrate (mg L^{-1})	150	1.00	300
NaCl (g L^{-1})	12	−0.16	8.5
Responses	**Experimental value (Std)**	**Predicted Value**	**Experimental Values (Opt)**
Biomass (g L^{-1})	0.77 ± 0.03	0.71	0.82 ± 0.06
YFP intensity (MFI)	3.6 ± 0.6	11.3	15.4 ± 1.1
YFP Positive cells (%)	49.2 ± 3.5	75	65.5 ± 1.4
Total YFP level (µg L^{-1})	123 ± 4	180.2	219 ± 9

$Z = Z^0 – Z^C/ \Delta Z$ Where Z and Z^0 indicate coded and real levels of independent variables respectively ΔZ represents the step change while Zc indicates the actual value at central points. Std- standard conditions, Opt- Optimized conditions.

YFP intensity and percentage of YFP positive cells

In the present study, under a optimized conditions, MFI increased significantly from 3.6 ± 0.6 to 15.4 ± 1.1 ($p > 0.001$; 4.2-fold increase), and the percentage of YFP$^+$ cells in *P. tricornutum* culture increased from 49.2 ± 3.5 to 65.5 ± 1.4 % ($p > 0.001$; 16.3 % increase) compared to the standard culture conditions without affecting biomass accumulation (Fig. 6-3 A, B, C & F and Table 6-3). This is consistent with previous observations showing that increasing temperature accelerates the maturation of GFP variants (Guerra et al., 2022). Under ideal circumstances, the CCD model projected an MFI of 11.3 and a percentage of YFP$^+$ cells of 75. The experimental result for MFI (15.4) was greater than the predicted value, indicating that the model was verified. However, % of YFP$^+$ cells experimental value (65.5) was lower; as a

result, the model cannot be verified for this parameter, possibly due to limitations of excitation and detection of this FP by the flow cytometer.

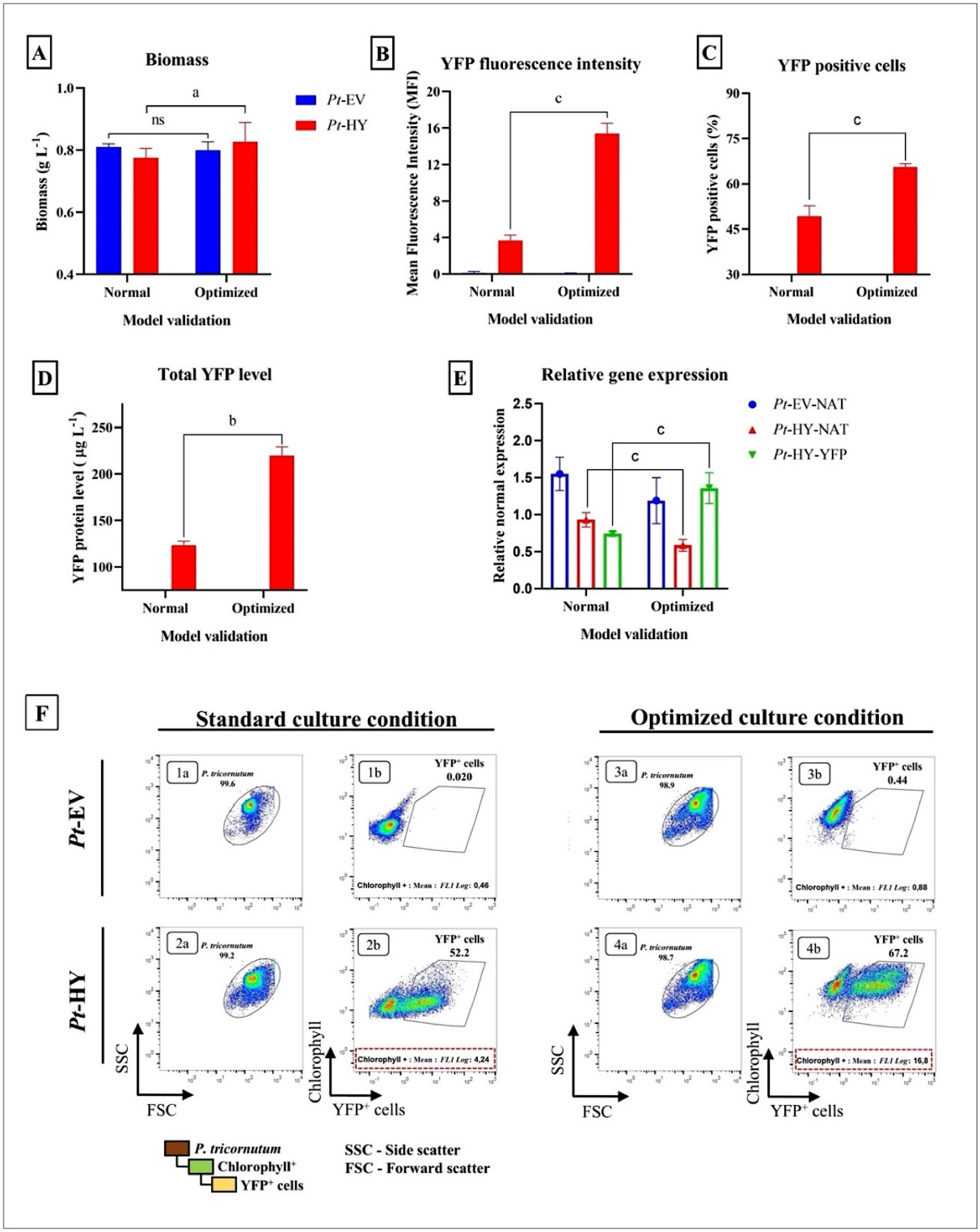


Figure 7-3: ***Experimental validation of the CCD model and compare with standard conditions.***

(A) Biomass accumulation in Pt-EV and Pt-HY (B and C) YFP MFI and % YFP+ cells quantified using flow cytometry (D) YFP level quantified by ELISA. (E) Relative expression of yfp and nat gene in Pt-EV and Pt-HY. The Pt-EV-NAT and Pt-HY-NAT refers to the analysis of natexpression levels in Pt-EV and Pt-HY, while Pt-HY-YFP- refers to yfp gene expression levels in Pt-HY. (F) Pseudo-color dot plots of Pt- EV, and Pt-HY (1 and 2) under standard conditions; (3 and 4) under optimized culture conditions. (1b-4b) Pseudo-color dot plots of Pt-EV and Pt-HY cultures with YFP in the x-axis, autofluorescence of chlorophyll on the y-axis. The gates of the total YFP+ population were designed according to Pt-EV autofluorescence.

Heterologous protein production in *P. tricornutum*

In the optimized condition, heterologous protein (YFP) yield increased significantly from 123 ± 4 to 219 ± 9 µg L^{-1} (p = 0.004; 1.8 factor increase) compared to standard conditions (Fig. 6-3 D and Table 6-3), which corresponded to about 0.2 % of the total soluble protein. As per the CCD model, the independent variables such as increasing nitrate and decreasing NaCl play significant role in improving the total YFP yield. Transcriptome and metabolites studies conducted on *P. tricornutum* by Alipanah et al. (2015) revealed a significant change in gene expression during nitrogen deficiency, reorienting the carbon metabolism toward energy storage mode and increasing the recycling of nitrogen compounds. Erdene-Ochir have reported much higher yield of GFP when its expression was driven by HASP1 promoter (6 mg/L) (Erdene-Ochir et al., 2019). This difference could be due to the relative quantification method based on fluorescence intensity rather than on direct protein quantification compared to ELISA. In agreement with our present observation, Kiefer et al. (2022) reported 11.2 µg L^{-1}SARS-CoV-2 spike protein, Chávez et al. (2016) reported 28.0 ± 4.3 µg L^{-1} of human vascular endothelial growth factor and Eichler-Stahlberg et al. (2009) reported 100 µg L^{-1} for human erythropoietin in *Chlamydomonas*. Other studies Ramos-Martinez et al. (2017), reported significantly higher secretory protein yields in mixotrophic green algae, including 15 mg L^{-1} of mVenus. Compared to heterotrophic hosts such as bacteria and yeast, phototrophic systems typically yield lower amount of heterologous protein due to lower biomass production, strong transgene silencing, and protease activity (Ramos-Martinez et al., 2017). These limitations could be overcome through advances in photobioreactor technology and stabilization of heterologous protein using carrier protein or fusion with protease inhibitor or synthetic glycol modules (Ramos-Martinez et al., 2017, Villanova et al., 2021).

The total experimental accumulation of YFP reached 219 ± 9 µg L^{-1}, which was higher than the predicted value of 180.2 µg L^{-1}. Overall, the experimental values were close or above to the predicted response, thus validating the model for all responses except for the percentage of YFP^{+} cells. The results also confirm that microalgae are suitable platforms for producing functional proteins in higher yield than plants (Sojikul et al., 2003).

Relative quantification of YFP expression

To determine whether the growth conditions modulated *YFP*transcript levels, as shown by Slattery et al. (2022), we measured the expression of the *yfp* gene in *Pt*-HY and *nat* genes in both strains *Pt*-HY and *Pt*-EV. Although the *nat* gene was driven by the FcpC promoter, the *yfp* gene was expressed under the control of the HASP1 promoter. These mRNA levels were evaluated in an 8-day-old culture in agreement with the highest fluorescence intensity and normalized to *Tuba* and *EF1α*expression. Our results showed that *nat* expression increased under standard conditions compared to optimized conditions in both strains (Fig. 6-3E). In agreement with MFI and YFP^+ cells % in the population, the relative gene expression of *YFP* was significantly higher ($p < 0.001$, 1.83-fold increase) under optimized conditions, as shown in Fig. 6-3E. These results highlight the importance of selecting appropriate promoters for multiple genes in the same expression cassette.

In silico* and *in vivo* stability analysis of YFP in transformed *P. Tricornutum.

Structure based thermodynamic stability prediction

We evaluated the intrinsic thermodynamic stability of YFP to understand its folding properties and resistance to denaturation at temperatures used to grow the diatom. Using predictive tools, we assessed key thermodynamic parameters, including *Tm*, which are essential indicators of YFP stability under various environmental conditions (Leuenberger et al., 2017). Protein stability depends on structure, protease activity, host cells, and cellular environment. Dehouck et al. (2008) showed a strong correlation between the T_m value of a protein and the growth temperature of its host. SCooP was recently reported (Pucci et al., 2017) as a protein structure-based protein stability prediction tool, which thermodynamically predicts the protein folding transition by standard free energy $\Delta G(T)$ as a function of temperature (*T)* when all other parameters (ionic strength, pressure, and pH) are constant. The 3D structure of YFP (Fig. 6-4A1) was then used to generate the thermodynamic characteristics using the SCooP portal. The YFP T_m was 63 °C, indicating the temperature at which half of the protein population was unfolded. At 21–25 °C, temperature range optimal for *P. tricornutum* growth, the YFP exhibited the most stable folding structure (Fig. 6-4A2). In agreement with our predicted T_m, Anderson et al. (2023) demonstrated the remarkable thermal stability of mutated YFP variants (Q66E and Q66E + H193Y), showing only a 60 % loss

in initial fluorescence after 1 h of exposure at 60 °C. The thermodynamic characteristics of YFP suggests that the observation of the present study relies on host-induced effects rather than inherent fluorescence stability.

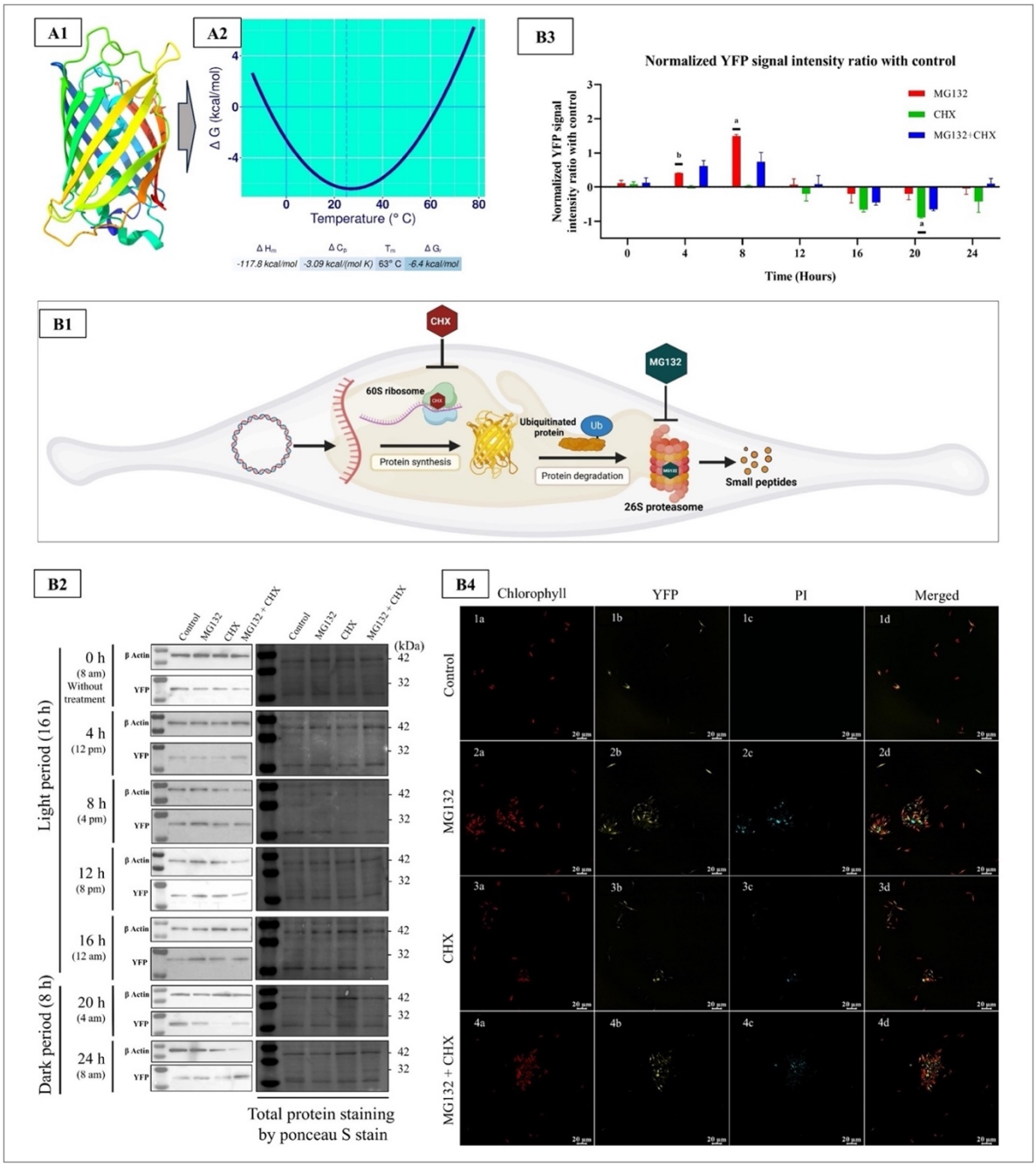


Figure 7-4: ***In silico and in vivo stability analysis of yellow fluorescent protein.***

(A1) 3D structure of YFP generated from Alphafold2. (A2) Thermal characteristics of the protein predicted using SCooP software. (B1) Graphical representation of protein stability analysis using proteasome (MG132) and protein synthesis inhibitors (CHX). (B2) Time courses immunoblot analysis of YFP signal intensity treating with different inhibitors for 24 h period. From left to right in the blots treating with solvent control (control), proteasome inhibitor (MG132), protein biosynthesis inhibitor (CHX), proteasome and protein biosynthesis inhibitor (MG132 + CHX). One representative result of three independent replicates is shown. (B3) Normalized YFP intensity ratio in the inhibitor- vs control treated samples (n = 3) with β actin as housekeeping protein. Mean value is plotted in the graph with standard error. (B4) Viability of P. tricornutum cells after 24 h of treatment following PI staining. (B4.1-4a) chlorophyll fluorescence, (B4.1-4b) YFP fluorescence, (B4.1-4c) fluorescence of PI+ cells, and (B4.1-4d) the merged image of all fluorescence filters.

***In vivo* stability of YFP in transformed *P. tricornutum* by protein inhibitor treatment**

To further elucidate the stability of heterologous proteins in *P. tricornutum*, we studied the turnover and stability of YFP *in vivo*. We investigated the levels and degradation rate of YFP using an immunoblot assay under different treatment conditions. Heterologous protein turnover in cells depends on protein degradation and synthesis. *P. tricornutum* cells were treated with CHX, an antibiotic that inhibits the protein synthesis (Obrig et al., 1971), and MG132, a peptidyl aldehyde that interacts with the 26S proteasome inhibiting ubiquitin-dependent proteasome activity (Lee & Goldberg, 1998)(Fig. 6-4B1). We evaluated the change in the level of YFP in the presence of MG132, CHX, MG132 + CHX, and control conditions every 4 h for 24 h. (Fig. 6-4B2).

YFP level was significantly higher after 4 h of treatment with MG132 compared to the control-treated group ($p = 0.01$; 41 % increase), and the same trend was observed following 8 h ($p = 0.05$; 150 % increase) of treatment (Fig. 6-4B3). The higher accumulation of YFP compared to the control suggests that the recombinant protein was sensitive to 26S proteasomal degradation. MG132 lost its effect on YFP accumulation starting at 12 h and YFP levels were lower than that in the control group after 16 h of treatment. As MG132 is not specifically targeting the recombinant protein, blocking the whole 26S proteasome pathway may lead to interruption of various cellular functions, including cell cycle progression and protein quality (Lee & Goldberg, 1998). Conversely, we observed a steady YFP levels following treatment with CHX for the first 8 h. This suggested that despite the inhibition of protein synthesis, YFP was stable for 8 h in the microalgae. The reduction in the YFP signal compared with the control after 12 h, 16 h, and 20 h of treatment with CHX (significant at 20 h, $p = 0.017$; 88 % decrease) (Fig. 6-4B3) indicated that intracellular accumulated YFP was being degraded at these timepoints, and is consistent with the reduction of YFP signal following MG132 treatment at these timepoints. After 24 h of treatment, the levels of YFP were comparable to those in control group, implying that the inhibitory effects of both MG132 and CHX were reversible. Inhibitor treatment did not cause any visible change in the culture during the light periods (first 16 h); however, the treated cells started to clump together after the dark cycle (24 h). The level of chlorophyll (Fig. 6-4D1-4a), YFP^+ cells (Fig. 6-4D1-4b) and nonviable cells (PI^+ cells) (Fig. 6-4D1-4c) in inhibitor treated (after 24 h) culture was visualized using confocal microscopy. The fluorescence image showed *P. tricornutum* cells clumps formation with few PI^+ cells in all

the treated conditions as compared to homogeneous culture for control (Fig. 6-4D1-4d). This observation was consistent with the previous observations by Im et al. (2024).

The ubiquitin–proteasome system (UPS) based protein degradation exists in most eukaryotic cells but has not yet been verified in diatoms. A genome-wide study by Ma et al. (2023) on *P. tricornutum* showed that 18 ubiquitin-conjugating enzymes play pivotal roles in maintaining cellular homeostasis under changing environmental conditions. A recent study by Im et al. (2024) on *P. tricornutum* using MG132, demonstrated the extent of the UPS systems used in the degradation of Aureochromes (PtAUREO1a) blue-light photoreceptors and transcription factors. These studies are consistent with the present results and suggest that the UPS system is active and participates in protein turnover in *P. tricornutum*. The current results implied that the newly synthesized YFP was stable for at least 8 h under normal conditions and remained resistant to degradation for up to 12 h when protein synthesis was blocked, after which degradation prevailed.

Overall, this analysis suggests that culture-related parameters affected the generation of fluorescent proteins at the transcript, translational, maturation, and stability levels. To understand the mechanism underlying the noticeably greater level of protein yield, MFI, and YFP^+ cells in culture under optimum culture conditions, thorough transcriptomic, proteomic, and metabolic flux analyses will be necessary. The optimization YFP may not fully capture the range of challenges associated with more complex genetic engineering scenario involving multiple proteins. However, the current approach could be adapted or expanded for more diverse protein expression studies in *P. tricornutum* or other microalgal systems. Finaly, scaling up of algal systems faces several challenges, including poor light distribution, uneven nutrient and gas exchange, pH fluctuation, and inefficient harvesting of low-density cultures. Addressing these inherent challenges is essential for establishing microalgae as viable and economical hosts for heterologous production.

7.8 Conclusion

This study uniquely applies RSM to optimize YFP expression in *P. tricornutum*, a model diatom with growing biotechnological relevance. By integrating mathematical modeling and experimental validation, we achieved a 4.2-fold increase in YFP MFI and a 1.8-fold rise in total

YFP levels (219 ± 9 μg L^{-1}) without affecting biomass. We also provided the first detailed assessment of YFP degradation kinetics, confirming ubiquitin–proteasome system involvement. These findings propose novel media composition, offer a predictive model for optimizing culture conditions to enhance recombinant protein stability in phototrophic hosts, advance diatoms as biomanufacturing platforms, and lays the foundation for heterologous production.

CRediT authorship contribution statement

Arun Augustine: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.
Anis Messaabi: Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization.
Elisa Fantino: Writing – review & editing, Formal analysis, Data curation.
Natacha Merindol: Writing – review & editing, Supervision, Data curation.
Fatma Meddeb-Mouelhi: Writing – review & editing, Project administration.
Isabel Desgagné-Penix: Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Isabel Desgagne-Penix reports financial support was provided by Natural Sciences and Engineering Research Council of Canada. Isabel Desgagne-Penix reports financial support was provided by Mitacs Canada. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge that financial support for this research was funded by the Natural Sciences and Engineering Research Council of Canada through the Alliance program Award No ALLRP 570476-2021 to IDP. Additional support in the form of scholarships to AA, AM, and EF, from Mitacs-Acceleration/ program grants no IT16463 and IT19432 to IDP is also acknowledged. AA acknowledge the valuable support from Prof. Tagnon Missihoun, Ms. Mélodie B. Plourde and other lab members from UQTR. The graphical abstract figure was created with BioRender.com. During the preparation of this work, the authors used ChatGPT

4.0 in order to correct grammatical errors and enhance readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

7.9 References

1. L. Alipanah, J. Rohloff, P. Winge, A.M. Bones, T.Brembu
Whole-cell response to nitrogen deprivation in the diatom *Phaeodactylum tricornutum*
J. Exp. Bot., 66 (20) (2015), pp. 6281-6296
View at publisher
CrossrefView in ScopusGoogle Scholar
2. Alipanah et al., 2018
L. Alipanah, P. Winge, J. Rohloff, J. Najafi, T.Brembu, A.M. Bones
Molecular adaptations to phosphorus deprivation and comparison with nitrogen deprivation responses in the diatom *Phaeodactylum tricornutum*
PLoS One, 13 (2) (2018), Article e0193335
View at publisherCrossrefView in ScopusGoogle Scholar
3. Anderson et al., 2023
M.R. Anderson, C.M. Padgett, C.J. Dargatz, C.R.Nichols, K.R. Vittalam, N.M. DeVore
Engineering a yellow thermostable fluorescent protein by rational design
ACS Omega, 8 (1) (2023), pp. 436-443
CrossrefView in ScopusGoogle Scholar
4. Chávez et al., 2016
M.N. Chávez, T.L. Schenck, U. Hopfner, C.Centeno-Cerdas, I. Somlai-Schweiger, C. Schwarz, H.-G.Machens, M. Heikenwalder, M.R. Bono, M.L. Allende, J.Nickelsen, J.T. Egaña
Towards autotrophic tissue engineering: Photosynthetic gene therapy for regeneration
Biomaterials, 75 (2016), pp. 25-36
View PDFView articleView in ScopusGoogle Scholar
5. Chen et al., 2022
J. Chen, Y. Huang, Y. Shu, X. Hu, D. Wu, H. Jiang, K.Wang, W. Liu, W. Fu
Recent progress on systems and synthetic biology of diatoms for improving algal productivity
Front. Bioeng. Biotechnol., 10 (2022), Article 908804
View in ScopusGoogle Scholar
6. Dayana Priyadharshini and Bakthavatsalam, 2016
S. Dayana Priyadharshini, A.K. Bakthavatsalam
Optimization of phenol degradation by the microalga *Chlorella pyrenoidosa* using Plackett-Burman Design and Response Surface Methodology
Bioresour. Technol., 207 (2016), pp. 150-156
View PDFView articleView in ScopusGoogle Scholar
7. De Martino et al., 2011
A. De Martino, A. Bartual, A. Willis, A.Meichenin, B. Villazán, U. Maheswari, C. Bowler
Physiological and molecular evidence that environmental changes elicit morphological interconversion in the model diatom *Phaeodactylum tricornutum*
Protist, 162 (3) (2011), pp. 462-481
View PDFView articleView in ScopusGoogle Scholar
8. Dehouck et al., 2008
Y. Dehouck, B. Folch, M. Rooman
Revisiting the correlation between proteins' thermoresistance and organisms' thermophilicity
PEDS, 21 (4) (2008), pp. 275-278
CrossrefView in ScopusGoogle Scholar
9. Diamond et al., 2023
A. Diamond, A.M. Diaz-Garza, J. Li, S.S. Slattery, N. Merindol, E. Fantino, F. Meddeb-Mouelhi, B.J. Karas, S.Barnabé, I. Desgagné-Penix
Instability of extrachromosomal DNA transformed into the diatom *Phaeodactylum tricornutum*

Algal Res., 70 (2023), Article 102998
View PDFView articleView in ScopusGoogle Scholar
10. Eichler-Stahlberg et al., 2009
A. Eichler-Stahlberg, W. Weisheit, O.Ruecker, M. Heitzer
Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*
Planta, 229 (4) (2009), pp. 873-883
CrossrefView in ScopusGoogle Scholar
11. Erdene-Ochir et al., 2019
E. Erdene-Ochir, B.K. Shin, B. Kwon, C. Jung, C.H. Pan
Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from *Phaeodactylum tricornutum*
Sci. Rep., 9 (1) (2019)
Google Scholar
12. Fantino et al., 2024
E. Fantino, F. Awwad, N. Merindol, A.M. Diaz Garza, S.-E. Gélinas, G.C. Gajón Robles, A. Custeau, F. Meddeb-Mouelhi, I. Desgagné-Penix
Bioengineering *Phaeodactylum tricornutum*, a marine diatom, for cannabinoid biosynthesis
Algal Res., 77 (2024), Article 103379
View PDFView articleView in ScopusGoogle Scholar
13. Filloramo et al., 2021
G.V. Filloramo, B.A. Curtis, E. Blanche, J.M.Archibald
Re-examination of two diatom reference genomes using long-read sequencing
BMC Genomics, 22 (1) (2021), p. 379
View in ScopusGoogle Scholar
14. Gómez-Loredo et al., 2016
A. Gómez-Loredo, J. Benavides, M. Rito-Palomares
Growth kinetics and fucoxanthin production of *Phaeodactylum tricornutum* and *Isochrysis galbana*cultures at different light and agitation conditions
J. Appl. Phycol., 28 (2) (2016), pp. 849-860
CrossrefView in ScopusGoogle Scholar
15. Guerra et al., 2022
P. Guerra, L.A. Vuillemenot, B. Rae, V. Ladyhina, A.Milias-Argeitis
Systematic *in vivo* characterization of fluorescent protein maturation in budding yeast
ACS Synth. Biol., 11 (3) (2022), pp. 1129-1141
View articleCrossrefView in ScopusGoogle Scholar
16. Iizuka et al., 2011
R. Iizuka, M. Yamagishi-Shirasaki, T. Funatsu
Kinetic study of de novo chromophore maturation of fluorescent proteins
Anal. Biochem., 414 (2) (2011), pp. 173-178
View PDFView articleView in ScopusGoogle Scholar
17. Im et al., 2024
S.H. Im, S. Madhuri, B. Lepetit, P.G. Kroth
Functional demonstration of Aureochrome 1a proteasomal degradation after blue light incubation in the diatom *Phaeodactylum tricornutum*
J. Plant Physiol., 292 (2024), Article 154148
View PDFView articleView in ScopusGoogle Scholar
18. Karas et al., 2015
B.J. Karas, R.E. Diner, S.C. Lefebvre, J. McQuaid, A.P.Phillips, C.M. Noddings, J.K. Brunson, R.E. Valas, T.J. Deerinck , J.Jablanovic, J.T. Gillard, K. Beeri, M.H. Ellisman, J.I. Glass, C.A.Hutchison 3rd, H.O. Smith, J.C. Venter, A.E. Allen, C.L. Dupont, P.D. Weyman
Designer diatom episomes delivered by bacterial conjugation
Nat. Commun., 6 (2015), p. 6925
View in ScopusGoogle Scholar
19. Kiefer et al., 2022
A.M. Kiefer, J. Niemeyer, A. Probst, G. Erkel, M.Schroda
Production and secretion of functional SARS-CoV-2 spike protein in *Chlamydomonas reinhardtii*
Front. Plant. Sci., 13 (2022), Article 988870

View in ScopusGoogle Scholar
20. Lee and Goldberg, 1998
D.H. Lee, A.L. Goldberg
Proteasome inhibitors cause induction of heat shock proteins and trehalose, which together confer thermotolerance in *Saccharomyces cerevisiae*
Mol. Cell. Biol., 18 (1) (1998), pp. 30-38
View in ScopusGoogle Scholar
21. Leuenberger et al., 2017
P. Leuenberger, S. Ganscha, A. Kahraman, V.Cappelletti, P.J. Boersema, C. von Mering, M. Claassen, P.Picotti
Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability
Science, 355 (6327) (2017), Article eaai7825
View in ScopusGoogle Scholar
22. Liang et al., 2014
Y. Liang, M. Sun, C. Tian, C. Cao, Z. Li
Effects of salinity stress on the growth and chlorophyll fluorescence of *Phaeodactylum tricornutum* and *Chaetoceros gracilis*(Bacillariophyceae)
Bot. Mar., 57 (6) (2014), pp. 469-476
CrossrefView in ScopusGoogle Scholar
23. Ma et al., 2023
W. Ma, H. Du, S.S.U.H. Kazmi, J. Chen, W. Chen, Y. Fan, Z. Liu, H. Luo, H. Fang, Z. Wang, X. Liu
UBC gene family and their potential functions on the cellular homeostasis under the elevated pCO_2stress in the diatom *Phaeodactylum tricornutum*
Ecol. Indicators, 148 (2023), Article 110106
View PDFView articleView in ScopusGoogle Scholar
24. Marter et al., 2020
P. Marter, S. Schmidt, S. Kiontke, D. Moog
Optimized mRuby3 is a suitable fluorescent protein for in vivo co-localization studies with GFP in the diatom *Phaeodactylum tricornutum*
Protist, 171 (1) (2020), Article 125715
View PDFView articleView in ScopusGoogle Scholar
25. Mirdita et al., 2022
M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S.Ovchinnikov, M. Steinegger
ColabFold: making protein folding accessible to all
Nat. Methods, 19 (6) (2022), pp. 679-682
CrossrefView in ScopusGoogle Scholar
26. Nagai et al., 2002
T. Nagai, K. Ibata, E.S. Park, M. Kubota, K.Mikoshiba, A. Miyawaki
A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications
Nat. Biotechnol., 20 (1) (2002), pp. 87-90
View in ScopusGoogle Scholar
27. Nur et al., 2018
M.M.A. Nur, W. Muizelaar, P. Boelen, A.G.J. Buma
Environmental and nutrient conditions influence fucoxanthin productivity of the marine diatom *Phaeodactylum tricornutum* grown on palm oil mill effluent
J. Appl. Phycol., 31 (1) (2018), pp. 111-122
Google Scholar
28. Obrig et al., 1971
T.G. Obrig, W.J. Culp, W.L. McKeehan, B. Hardesty
The mechanism by which cycloheximide and related glutarimide antibiotics inhibit peptide synthesis on reticulocyte ribosomes
J. Biol. Chem., 246 (1) (1971), pp. 174-181
View PDFView articleView in ScopusGoogle Scholar
29. Ouyang et al., 2018
Z. Ouyang, R. Chen, Q. Liu, L. He, W.J. Cai, K. Yin
Biological regulation of carbonate chemistry during diatom growth under different concentrations of Ca^{2+} and Mg^{2+}
Mar. Chem., 203 (2018), pp. 38-48

View PDFView articleView in ScopusGoogle Scholar
30. Pfaffl, 2001
M.W. Pfaffl
A new mathematical model for relative quantification in real-time RT–PCR
Nucleic Acids Res., 29 (9) (2001), pp. 2002-2007
Google Scholar
31. Plackett and Burman, 1946
R.L. Plackett, J.P. Burman
The design of optimum multifactorial experiments
Biometrika, 33 (4) (1946), pp. 305-325
Google Scholar
32. Pucci et al., 2017
F. Pucci, J.M. Kwasigroch, M. Rooman
SCooP: an accurate and fast predictor of protein stability curves as a function of temperature
Bioinformatics, 33 (21) (2017), pp. 3415-3422
CrossrefView in ScopusGoogle Scholar
33. Rajendran et al., 2008
A. Rajendran, A. Palanisamy, V. Thangavelu
Evaluation of medium components by Plackett-Burman statistical design for lipase production by *Candida rugosa* and kinetic modeling
Chin. J. Biotechnol., 24 (3) (2008), pp. 436-444
View PDFView articleView in ScopusGoogle Scholar
34. Ramos-Martinez et al., 2017
E.M. Ramos-Martinez, L. Fimognari, Y.Sakuragi
High-yield secretion of recombinant proteins from the microalga *Chlamydomonas reinhardtii*
Plant Biotechnol. J., 15 (9) (2017), pp. 1214-1224
CrossrefView in ScopusGoogle Scholar
35. Siaut et al., 2007
M. Siaut, M. Heijde, M. Mangogna, A. Montsant, S.Coesel, A. Allen, A. Manfredonia, A. Falciatore, C. Bowler
Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*
Gene, 406 (1–2) (2007), pp. 23-35
View PDFView articleView in ScopusGoogle Scholar
36. Sivakaminathan et al., 2018
S. Sivakaminathan, B. Hankamer, J.Wolf, J. Yarnold
High-throughput optimisation of light-driven microalgae biotechnologies
Sci. Rep., 8 (1) (2018), p. 11687
View in ScopusGoogle Scholar
37. Slattery et al., 2022
S.S. Slattery, D.J. Giguere, E.E. Stuckless, A.Shrestha, L.K. Briere, A. Galbraith, S. Reaume, X. Boyko, H.H.Say, T.S. Browne, M.I. Frederick, J.T. Lant, I.U. Heinemann, P.O'Donoghue, L. Dsouza, S. Martin, P. Howard, C. Jedeszko, K. Ali, G. Styba, M. Flatley, B.J. Karas, G.B. Gloor, D.R. Edgell
Phosphate-regulated expression of the SARS-CoV-2 receptor-binding domain in the diatom *Phaeodactylum tricornutum* for pandemic diagnostics
Sci. Rep., 12 (1) (2022), p. 7010
View in ScopusGoogle Scholar
38. Sojikul et al., 2003
P. Sojikul, N. Buehner, H.S. Mason
A plant signal peptide–hepatitis B surface antigen fusion protein with enhanced stability and immunogenicity expressed in plant cells
PNAS, 100 (5) (2003), pp. 2209-2214
View in ScopusGoogle Scholar
39. Vartanian et al., 2009
M. Vartanian, J. Desclés, M. Quinet, S. Douady, P.J. Lopez
Plasticity and robustness of pattern formation in the model diatom *Phaeodactylum tricornutum*
New Phytol., 182 (2) (2009), pp. 429-442
CrossrefView in ScopusGoogle Scholar
40. Villanova et al., 2021

V. Villanova, D. Singh, J. Pagliardini, D. Fell, A. Le Monnier, G. Finazzi, M. Poolman
Boosting biomass quantity and auality by improved mixotrophic culture of the diatom *Phaeodactylum tricornutum*
Front. Plant. Sci., 12 (2021), Article 642199
View in ScopusGoogle Scholar

41. Wu et al., 2010

Y. Wu, K. Gao, U. Riebesell
CO_2-induced seawater acidification affects physiological performance of the marine diatom *Phaeodactylum tricornutum*
Biogeosciences, 7 (9) (2010), pp. 2915-2923
CrossrefView in ScopusGoogle Scholar

Appendix IV, Review Paper (Co-authorship)

8 Diatoms Biotechnology: Various Industrial Applications for a Greener Tomorrow

REVIEW article
Front. Mar. Sci. , 23 February 2021
Sec. Marine Biotechnology and Bioproducts
Volume 8 - 2021 | https://doi.org/10.3389/fmars.2021.636613

*Nikunj Sharma[1], Daris Pazhukkunnel Simon[1], Aracely Maribel Diaz-Garza[1], Elisa Fantino[1], Anis Messaabi[1], Fatma Meddeb-Mouelhi[1,2], Hugo Germain[1,2] and Isabel Desgagné-Penix[1,2]**

[1] Department of Chemistry, Biochemistry and Physics, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada, [2] Groupe de Recherche en Biologie Végétale, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada*

The benefits of the complex microscopic and industrially important group of microalgae such as diatoms is not hidden and have lately surprised the scientific community with their industrial potential. The ability to survive in harsh conditions and the presence of different pore structures and defined cell walls have made diatoms ideal cell machinery to produce a variety of industrial products. The prospect of using a diatom cell for industrial application has increased significantly in synch with the advances in microscopy, metabarcoding, analytical and genetic tools. Furthermore, it is well noted that the approach of industry and academia to the use of genetic tools has changed significantly, resulting in a well-defined characterization of various molecular components of diatoms. It is possible to conduct the primary culturing, harvesting, and further downstream processing of diatom culture in a cost-effective manner. Diatoms hold all the qualities to become the alternative raw material for pharmaceutical, nanotechnology, and energy sources leading to a sustainable economy. In this review, an attempt has been made to gather important progress in the different industrial applications of diatoms such as biotechnology, biomedical, nanotechnology, and environmental technologies.

8.1 Introduction

The global trend of economy and society is shifting toward building a greener and more sustainable society to combat climate and health issues. This is a critical issue, which is being

approached with various interdisciplinary strategies to produce a wide range of sustainable products. For instance, biotechnology research has invested a significant number of resources, time, and money in studying microorganisms to exploit them for human consumption in multiple ways. Furthermore, the decades of research and improvisation in cultivation strategies, extraction, and harvesting protocols strongly support a good return on investment in industrial applications of microbes. A pinch of soil and a drop of water contain a diversity of microbes that controls major biogeochemical cycles and subsequently have the potential of producing an abundance of sustainable products. Since the beginning of this century, a high amount of research work has been published on industrial applications of microbes such as bacteria, yeast, and microalgae (Figure 7-1). But, limited attention has been paid to diatoms which have the potential of becoming a robust sustainable industry because diatoms can continuously grow with an average annual yield of 132 MT dry diatoms ha^{-1} over almost 5 years (Wang and Seibert, 2017).

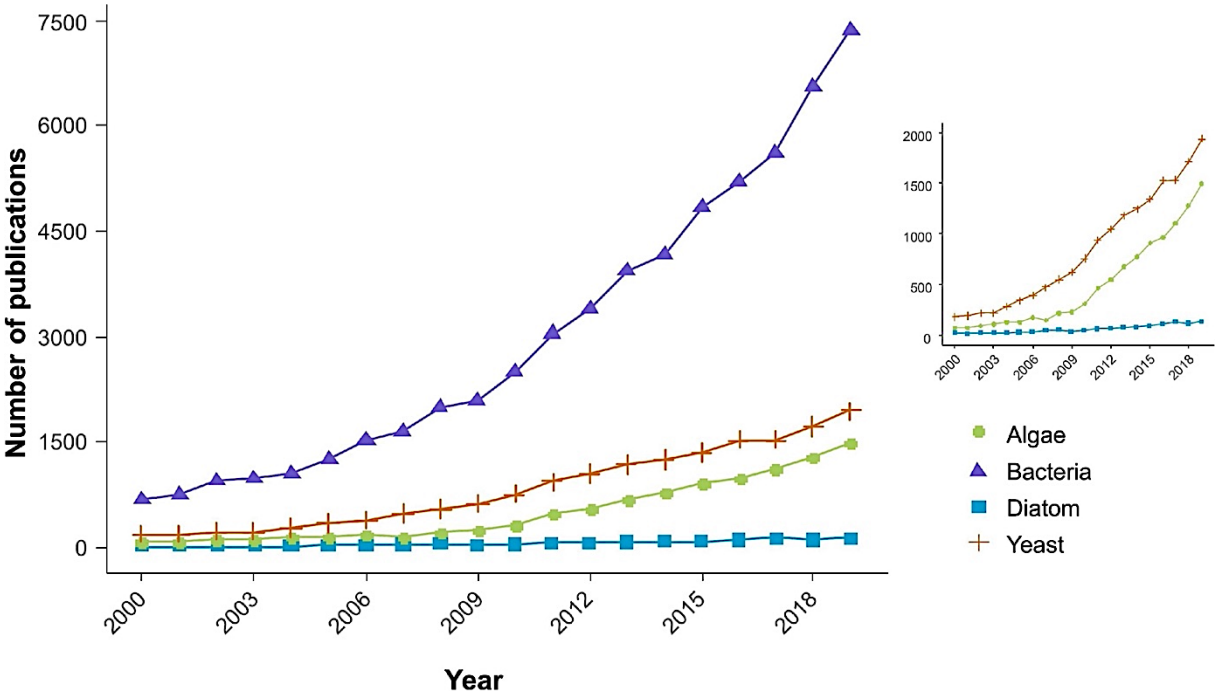


Figure 8-1: ***Approximate number of research articles indexed in Scopus database (September 14, 2020) in the area of industrial application of different microbes (bacteria, yeast, algae, and diatoms).***

Diatoms are dynamic microorganisms with rich diversity and detailed membrane design. They are the most dominating phytoplankton with an overall number of around 200,000 species having complex variability in dimensions and shapes (Round et al., 1990; Smetacek, 1999; Mann and Vanormelingen, 2013). Diatoms' distinctive characteristic compared to the phytoplankton community is their silica cell wall, known as a frustule. This innate ability to

uptake silicon from the environment has made them an interesting community of microbes since the 19th century. Few studies have stated the role of frustule biosilicate as pH buffering material which facilitates shifting of bicarbonate to CO_2 dissolved in cell fluids (the latter is readily metabolized by diatoms) (Milligan and Morel, 2002).

The access to advanced microscopes and modern genetic tools enabled us to study the detailed frustule structure and validate metabolic pathways involved in absorption, transportation, and polymerization of silicon and other biomolecules like lipids (Knight et al., 2016; Zulu et al., 2018). Furthermore, this advanced knowledge of metabolic pathways and validation of diatom structure can be applied to produce a wide range of renewable products such as optoelectronics, biofuels, nutritional supplements, ecology tools, etc. (Marella et al., 2020).

Other common factors that have shaped the evolution of diatoms are their ability to adapt and grow in various natural resources; fresh and marine water, wastewater, rivers, and oceans. Their abundance and adaptability in a wide range of climate and geographical areas make them suitable for different applications (Jin and Agustí, 2018). It was reported that diatoms are responsible to produce yearly, 40% of the organic carbon and 20% of oxygen (Tréguer et al., 1995; Falkowski et al., 1998; Afgan et al., 2016). Besides, these photoautotrophic organisms are involved in biogeochemical cycles, which play a significant role in global carbon fixation, carbon sequestration, and silicon cycle. They are also suitable candidates to capture nitrogen and carbon from various sources, which can be exploited by waste management and the biofuel industry to create carbon-neutral fuels (Singh et al., 2017). Furthermore, these algae are used to produce nutraceutical compounds, such as vegetarian proteins, omega, and other essential fatty acids for pharmaceutical industries (Wen and Chen, 2001a,b).

Multiple epidemiological, clinical, and pre-clinical studies have shown that omega fatty acids such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) are useful in slowing down age-related diseases such as cardiovascular diseases and cancer (Cole et al., 2010; Dyall, 2015; Thomas et al., 2015; Wang and Daggy, 2017). The development of diatoms strains rich in omega fatty acids can replace the dependence on fish as a source of omega oils and reduce the problems associated with seasonal variations and ocean pollution which might affect the biochemical composition of fish oil (Alves Martins et al., 2013). Also, various marine diatoms are considered for the commercial production of antioxidant pigments such as fucoxanthin and other carotenoids. It has been reported that these pigments exhibit various protective effects such as strong antioxidant activities (Xia et al., 2013).

Thus, the flexible and complex nature of diatoms offers immense possibilities to develop a wide range of sustainable products and contributes to carbon neutrality. Because of its dimensions,

pore distributions, and geometries, it is studied to develop tools for nanotechnology and biomedical industry such as nanofabrication techniques, chemo and biosensing, particle sorting, and control of particles in micro- and nano-fluidics (Mishra et al., 2017). Silica and biosilica can be used to develop advanced nanomaterial for electronic and optical technologies which can be employed for ultra-sensitive detection of biological compounds (Dolatabadi et al., 2011). Recent accomplishment in diatoms metabarcoding, a reference database of the global population of diatoms, has advanced its use extensively in studying ecological problems such as climate change, acidification, and eutrophication (Nanjappa et al., 2014). Because of its robust nature and potential to inhabit different photic regions, from the equator to the poles, diatoms offer the potential to develop tools and products for all geographical regions (Medlin, 2016). The technological and infrastructure advancements of diatoms-based applications are at a new level. Besides, it requires different kinds of optimization either in laboratory or large-scale research such as energy utilization for different steps, financial modeling, and collaborating with different industries to make diatom-based products commercially successful. However, the standardization at various levels such as optimization of culture conditions, genetic tools, genome and transcriptome sequencing make diatoms based products commercially viable.

Therefore, this review aims to provide a better understanding of the potential of diatoms research at a laboratory scale. We have tried to provide comprehensive information on a variety of diatoms applications such as energy, biomedical products, and environment monitoring which are being investigated at different levels. All these applications have the potential to contribute toward a greener tomorrow. The purpose of the research is to increase the sustainable economy while reducing the dependence on non-renewable resources. Therefore, recovering and producing various sustainable products like biofuels, feed, bioactive molecules, and services like environment monitoring embedded in diatoms is a promising opportunity to be seized as shown in Figure 7-2.

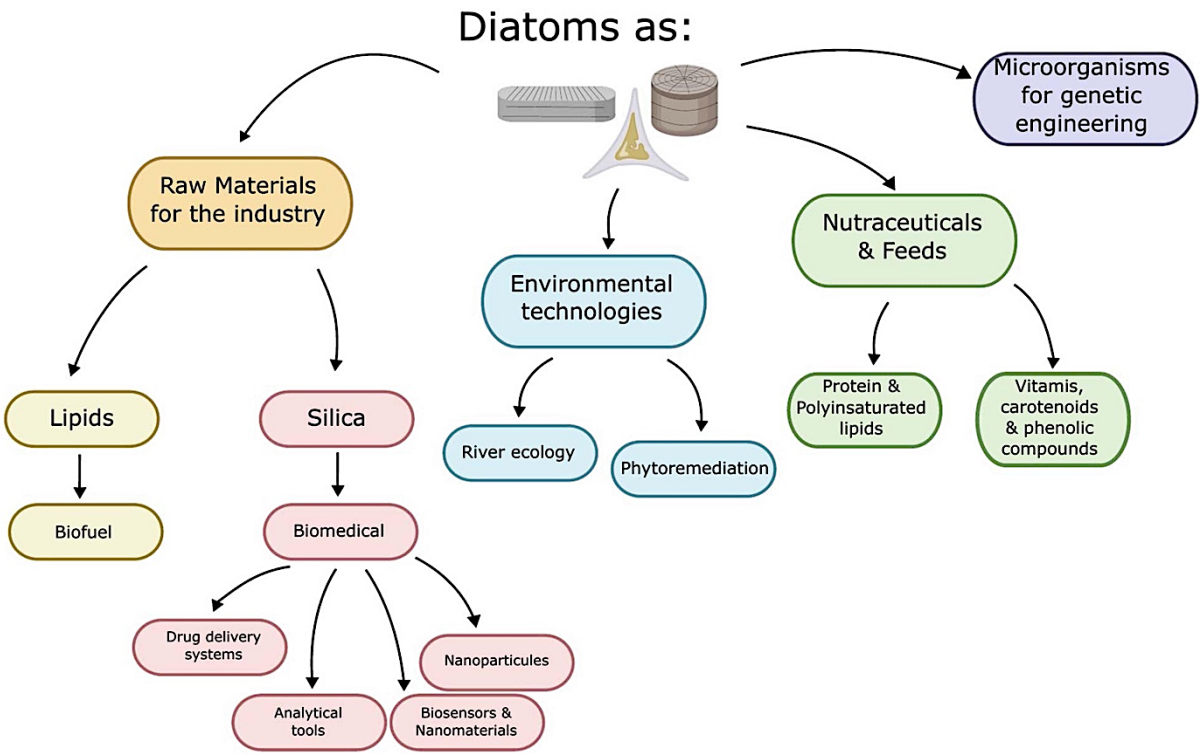


Figure 8-2: ***Scheme of the different uses of diatoms for green industry.***

Biofuel Industry

Fast globalization and industrialization have impacted the ecosystem widely but shutting or slowing down the globalization is not the solution. At the moment, almost 95% of all the transportation industry is based on a non-renewable source of energy (Rodrigue and Notteboom, 2013). Therefore, developing sustainable and carbon-neutral fuels could reduce the existing dependence on fossil fuels and contribute to bringing back harmony in nature without disrupting the existing economic development. Few economic aspects of biofuel production from microalgae such as biodiesel productivity, land use, and oil yield support the use of microalgae for commercial production as compared to corn and other food crops. The oil yield for microalgae with high oil content is almost 15-fold more as compared to corn. Whereas, the land use for corn and maize is 66-fold more as compared to microalgae (Brocks et al., 2003).

The microalgae such as diatoms are the promising feedstock to replace non-renewable sources of energy. It has been proven by geochemists that algal lipids are the major feedstocks of petroleum and these lipids act as the biomarker remaining stable for millions of years (Brocks et al., 2003). The main biomarker for the diatoms is the ratio of C28 and C29 steranes and highly branched isoprenoid alkenes which are found in high-quality oil fields around the globe (Katz et al., 2004).

Moreover, targeting the diatom lipids by manipulating and optimizing the growth and culture conditions such as light, stress, and nutrients can provide an interesting alternative to help meet the existing demands of commercial production of biofuel. Knowing the potential of diatoms to accumulate high lipids and varied compositions of fatty acids, diatoms are an underexploited area of the biofuel industry. The most predominant saturated and unsaturated fatty acids in diatom species are 14:0; 16:0, 16:1, 16:2, 16:3, 18:1, 18:2, 18:3, 20:4, and 20:5 (Dunstan et al., 1993; Sharma et al., 2020). Various reports have been published on different species of diatoms regarding the lipid yield and triacylglycerol accumulation (TAG) under different treatments as shown in Table 7-1.

Table 8-1: ***Lipid content and productivities of different microalgae diatom species (-: no data).***

Microalgae	Culture condition	Lipid (% dry weight)	Lipid productivity mg L^{-1} day^{-1}	TAG productivity μ mol L^{-1} day^{-1}	% of TAG	References
Thalassiosira weissflogii P09	–	29.94 ± 1.17	7.27 ± 0.28	–	51.0 ± 3.2	d'Ippolito et al. (2015)
	Nitrogen limitation	–	–	19 (+20%)	–	d'Ippolito et al. (2015)
Thalassiosira weissflogii CCMP 1010	–	38.84 ± 0.78	4.87 ± 0.10	–	53.0 ± 1.9	d'Ippolito et al. (2015)
Thalassiosira pseudonana CCMP 1335	–	29.33 ± 1.17	1.72 ± 0.07	–	19.0 ± 0.9	d'Ippolito et al. (2015)
	High CO_2 20,000 ppm	–	–	45.5 ± 26 (exponential) (+285%)	–	Jensen et al. (2020)
Cyclotella cryptica CCMP 331	–	41.97 ± 1.26	2.98 ± 0.09	–	55.0 ± 2.1	d'Ippolito et al. (2015)
	Nitrogen limitation	–	–	45 (+20%)		d'Ippolito et al. (2015)
Phaeodactylum tricornutum CCMP 632	–	9.32 ± 0.28	2.09 ± 0.06	–	19.0 ± 0.6	d'Ippolito et al. (2015)
	Tn19745_1 strain + nitrogen limitation	–	–	–	45-fold increase	Daboussi et al. (2014)
	Dark	+2.3-fold	–	–	–	Bai et al. (2016)
	High CO_2	–	–	75.7 ± 9 (stationary) (+50%)	–	Jensen et al. (2020)
textitNavicula pelliculosa (marine)	High CO_2	–	–	158.4 ± 29 (stationary) (+35%)	–	Jensen et al. (2020)

It is possible to improve the quality of biodiesel by optimizing the content of different fatty acids that impacts biodiesel properties; cetane number, level of emissions, cold flow, oxidative stability, viscosity, and lubricity (Knothe, 2005). Fatty acids with chain lengths from C16 to C18 should contribute the maximum amount in the final product (Knothe, 2009). Some researchers have reported that a high percentage of mono-unsaturation is also desirable for biodiesel (Knothe, 2012). Thus, optimizing the fatty acid profile along with increased biomass will significantly enhance their economic value.

Statistical analyses predicted that 100 mt/ha/year biomass of diatoms is required for commercial biofuel production (Gallagher, 2011). Over 10 years, productivity range was observed to be between 29 and 142 mt/ha/year (Sheehan et al., 1998; Huesemann and Benemann, 2009), these values motivate the researchers and industry experts to study diatom cell in-depth for the biofuel industry in both lab-scale and large scale.

Furthermore, the availability of advanced genetic tools can help to achieve the missing targets in developing diatoms cells as biofuel machinery (Radakovits et al., 2010; Tibocha-Bonilla et al., 2018). Based on theoretical calculations about the land area, lipid production, and photosynthetic energy conversion, the biofuel demand of the complete United States population could be met using only 5% of United States land (Levitan et al., 2014). Although various other factors that define the efficacy of biodiesel such as engine performance, that is based on (cylinder pressure, brake mean effective pressure, frictional mean effective pressure, power, torque, brake specific fuel combustion, brake thermal efficiency). The statistical data supports the use of microalgae-based biofuel but there are various limitations at a technological level for large-scale implementation of this project. Therefore, one of the alternatives is to use the blended form of biodiesel. It would be more efficient to make a blended version of petro-diesel and microalgae/diatoms based fuel for large-scale operation. The comparative studies of blended (20% microalgae fuel plus 80% petrodiesel) and 100% petrodiesel have no major performance variations. Furthermore, it was reported that there was a reduction in the CO, unburnt HC, and smoke emissions in blended form as compared to pure diesel (Soni et al., 2020).

Biomedical Industry

Drug Delivery Systems

The cost required to bring a new drug to the market has been estimated by the Tufts Centre for the Study of Drug Development at approximately 2.6 billion dollars (DiMasi et al., 2016). In addition, the current drug delivery systems have limited solubility, poor bio-distribution, lack of selectivity, premature degradation, and unfavorable pharmacokinetics (Aw et al., 2011a,b). Therefore, these limitations have motivated the research and development of alternative drug delivery systems to improve the performance of existing drugs (i.e., increasing bioavailability), while reducing undesirable effects. There is no doubt that existing biomedical technologies have increased the life span but the human society wants to improvise the quality of life further by adopting environment friendly methods. Therefore, we should speed up the process and

conduct in-depth research on using diatom frustules, even other bio-inspired alternatives for biomedical applications.

Among the available drug delivery tools (liposomes, nanogels, carbon nanotubes), the intricate frustule characteristics of diatoms such as specific surface area, thermal stability, biocompatibility, and alterable surface chemistry, have attracted attention for its use in drug and gene delivery. It took million years of evolution for diatoms to manufacture this level of complex and delicate structure to protect from the unwanted conditions like high temperature and variable light fluctuations. 3-D section analyses of diatom frustules have shown the availability of multiple pore patterns that range from nanometer to micrometer (Chandrasekaran et al., 2014; Cicco et al., 2015; Ragni et al., 2017). These characteristics are sufficient to explore alternative and low-priced silica-based materials for the biomedical industry (Mishra et al., 2017; Terracciano et al., 2018). Diatoms' frustule structure changes its homogenous nature, space, and intricate nature according to various environmental factors and silicon uptake efficiency (Knight et al., 2016). This ability can be used to change the frustule shape and pore size, which has multiple applications in the biomedical and nanotechnology industry. The process of biosilicification in diatoms is quite complex, it includes the role of silicic acid transporters, transportation of silica, and polymerization of silica monomers among other processes that have been extensively explained (Martin-Jézéquel et al., 2000; Knight et al., 2016). Moreover, a detailed investigation is being conducted to make the natural 3D porous structure an efficient substitute for delivery systems attributed to its chemical and mechanical features. For instance, some diatom species such as *Coscinodiscus concinnus* sp. (Gnanamoorthy et al., 2014), *Thalassiosira weissflogii* sp. (Aw et al., 2011a) are potential drug carriers candidates due to their amorphous nature and morphology. Additionally, various studies have shown that diatoms microcapsules are effective carriers for poorly soluble and water-soluble drugs, which can be applied in both oral and implant applications (Aw et al., 2011a; Ragni et al., 2017).

The defined structural architecture of diatoms, such as pore volume and controllable particle size, allows the synthesis of biomolecules at the micro- to nano-scale (Losic et al., 2005, 2010; Slowing et al., 2008). The growth of fibroblast and osteoblast has been observed on functionalized frustules supporting the idea of using biosilica from diatoms as smart support for cell growth (Ragni et al., 2017). Regarding modified diatoms, Losic et al. (2010) have designed the magnetically guided drug carrier via a functional surface of diatoms with dopamine-modified iron oxide. This modification has shown the capability of sustained release

of poorly soluble drugs for 2 weeks, presenting an enhanced performance for drug delivery (Losic et al., 2010). Moreover, genetically modified biosilica has been used to selectively deliver anticancer drugs to tumor sites (Delalat et al., 2015). Overall, these findings have opened the doors to novel drug delivery systems using renewable material. Therefore, all properties of diatoms such as uniform pore structure, chemically inert and biocompatible, non-toxic, easy to transport, filtration efficiency, and specific drug delivery make it a potential model for drug delivery tools (Curnow et al., 2012; Milović et al., 2014; Rea et al., 2014; Vasani et al., 2015).

Analytical Tools

The controlled production of nanostructured silica is possible through chemical and mechanical treatment for a wide range of applications. This nanopore structure has a huge potential to attach the desired biomolecule (enzymes, DNA, antibodies) and develop label-free analytical tools or enhance the catalytic properties. It has also been shown that enzymes and DNA (oligonucleotides) can be conjugated to silica (Losic et al., 2005; Zamora et al., 2009). The encapsulation of enzymes in diatom biosilica exhibits improved enzymatic properties as compared to other immobilization technologies (Kato et al., 2020).

Additionally, luminescent nano- and micro-particles have gained the attention of the interdisciplinary scientific community (biology, chemistry, and physics). Current available fluorescent labeling agents are quantum dots, lanthanide-doped compounds, and organic fluorophore-tagged nanobeads, which offer good optical properties and a broad excitation spectrum. However, these agents have limitations in properties such as photobleaching and biocompatibility. For instance, De Stefano et al. (2009) studied diatoms' potential to incorporate fluorophores with increased stability used to study the molecular event of antibody-antigen identification. Moreover, molecular recognition between antibody and antigen was observed in relation to the change in the photoluminescence spectrum of diatoms. Concluding that diatom's frustules, due to their high sensitivity, low-cost, and availability are ideal alternative candidates for lab-on-particle applications (De Stefano et al., 2008, 2009).

There is no concrete evidence of diatoms' presence in land animal bodies. Although, various studies showed the presence of diatoms in the internal organs and circulatory system of alive or dead animals in an aquatic environment (Ludes et al., 1996; Lunetta et al., 1998; Hürlimann et al., 2000; Lunetta and Modell, 2005; Horton et al., 2006; Levkov et al., 2017). The siliceous

cell wall of this organism is resistant to degradation even under high acidic conditions for a long period (Lunetta and Modell, 2005). The investigation on the occurrence of these organisms inside dead bodies of aquatic environment that died from different causalities opened up a new possibility of forensic analysis through the examination of diatoms called 'diatom axiom' or 'diatom test'(Lunetta et al., 1998). The diatom test is based on the hypothesis that the microalgae will not enter the systemic circulation and reach other internal organs and tissues such as bone marrow unless the circulation is functional. A forensic examiner can determine whether the individual was alive when it was entering the water by checking the presence of diatoms in various organs and tissues (Levkov et al., 2017). In addition, since diatoms are highly sensitive to environmental conditions, different water bodies have different diatom species abundance which allows forensics to identify the drowning site (Zhou et al., 2020).

Despite being a distinguishable method, the diatom test has limitations also. One of the major issues is the occurrence of diatoms in a drowning medium. The absence or low presence of diatoms in a water body can lead to a false positive or negative result. The presence of diatoms in different layers (water base, deeper, and surface) of the water body also can be varied (Levkov et al., 2017). Rapid death is another situation where the diatom test can be wrong. Instant death when an animal or human enters the water body for various reasons such as cold shock and cardiac diseases will give a negative result in the diatom test (Smol and Stoermer, 2010). The use of alcohol or drugs is another factor that can mislead in the diatom test (Ago et al., 2011). Recent advances in DNA Barcoding and pyrosequencing opened the possibility of increasing the accuracy of the diatom test by checking the presence of plankton specific genes (e.g., Rubisco gene) in animal tissue (Fang et al., 2019).

Biosensors and Nanomaterials

The advances in biotechnological tools have made it effective to characterize the frustules of diatoms for the fabrication of optoelectronics. The uptake of various elements such as zinc and germanium by diatom like *Stephanodiscus hantzschii*, *Thalassiosira pseudonana, etc.* to change the pore size, shape, and other characteristics which are being studied for a variety of functions such as paleolimnological indicator and photonic device application (Qin et al., 2008; Jaccard et al., 2009). It has been reported a relationship between the amount of Zn/Si (zinc/silicon) and free zinc ions which can be used as a proxy of paleolimnological indicators (Jaccard et al., 2009). The studies have raised intriguing questions about the uptake and the process of various elements which need detailed validations. Although, they have reported that

they could only detect Zn and Fe as chemical elements. The analysis of various trace elements could be used as an environmental indicator which indeed will reduce the total workload needed to monitor large water bodies (Ellwood and Hunter, 2000).

The complex nanobiochemical machinery of diatoms can be exploited to fabricate a wide range of nanostructures with diverse optical and electronic properties (Rorrer et al., 2007). The ability to manufacture different pore size nanostructure molecules has inspired many research groups and industries to use diatoms in biosensing (De Tommasi, 2016). The incorporation of chemical elements such as germanium significantly affects the structure and size of frustule pores. A study tested the possibility of using Si-Germanium composite material in living diatoms in a two-stage photobioreactor cultivation process which reduced the pore size without disturbing the morphology (Rorrer et al., 2007). Another study reported that insertion of germanium in *Nitzschia frustulum* induces the nanocomb structure with blue photoluminescence (Qin et al., 2008). These nanostructure materials exhibit optical properties suitable for use in semiconductors and optoelectronics. Manufacturing of these materials combined with the silica frustule will improve the overall durability and range of applications in nanotechnology industries. These lab-scale scientific discoveries have shown that it is possible to create advanced nanomaterials in living diatoms.

Nanoparticles

The development of well-defined, advanced, and eco-friendly nanoparticles has attracted the attention of many researchers in the area of nanotechnology and its applications. Nanoparticles can be applied to study antimicrobial activity, catalyst, and filtering waste and chemical compounds. Biosynthesis of metallic nanoparticles in photoautotrophic organisms has gained the attention of nanotechnology researchers. Various approaches such as the sol-gel process, atomic layer deposition, chemical bath deposition, and inkjet printing process, have been used to modify the chemical composition of frustules. In this regard, an inexpensive chemical deposition technique was tested to deposit cadmium sulfide (CdS) on the surface on *Pinnularia* sp. without changing its morphology, since CdS has a wide range of applications in photodetectors and solar cells (Gutu et al., 2009).

Recently, it has been reported that diatoms can biosynthesize the nanoparticles such as gold and silver which has shown strong cytotoxicity against harmful microorganisms. Additionally, a highly ductile and malleable metal platinum (Pt) has been introduced in presence of dihydrogen

hexachloroplatinate (IV) hexahydrate (DHH) in the living diatom *Melosira nummuloides*, without interfering the native morphology (Yamazaki et al., 2010). This is due to the platinum's excellent resistance to corrosion and stability at high temperatures, hence having application in a broad spectrum of industries, besides biomedicine. Other various examples of the on-going investigation of diatoms silica-based materials and their applications in biomedicine are shown in Table 7-2.

Table 8-2: ***Biomedical applications of diatom silica-based materials using different diatom species.***

Application	Organism	References
Specific nanoporous biosilica delivery system of chemotherapeutic drug, consisting in the attachment of antibodies and hydrophobic drug molecules, without using cross-linking, to the diatoms biosilica.	*T. pseudonana*	Delalat et al. (2015)
Modified frustule with self-assembled antibacterial aromatic amino acid conjugates Tyr–Zn^{II} as a zinc carrier for its controlled release to bacteria and inhibiting the bacterial growth.	*N. palea*	Singh et al. (2020)
Genetically modified frustule with chimeric fusion proteins: diatom-derived silica targeting peptide Sil3T8 and a small synthetic antibody derivative to detect *Bacillus anthracis*	*T. pseudonana*	Ford et al. (2020)
Rapid and selective detection of typhoid using cross-linked amine-functionalized diatom photoluminescent biosensor.	*Amphora* sp.	Selvaraj et al. (2018)
Nano composite of nanoporous diatom-ZrO_2 selective and highly sensitive sensor for non-enzymatic detection of methyl parathion.	*P. tricornutum*	Gannavarapu et al. (2019)
Biomaterial for negative electrode composed by a 3D-structured diatom biosilica for lithium-ion batteries, showing increased charge capacity compared to graphite.	*P. trainorii*	Nowak et al. (2019)
Improved capacitor performance of *in situ* coating of FeOx on live diatoms as a potential material for super capacitor electrodes.	*P. tricornutum*	Karaman et al. (2019)

We have discussed the major application of diatoms for established industries such as biofuels, nanomaterials, and biomedicine. However, diatoms also have other fascinating applications in environment monitoring, animal feed, and aquaculture, which indeed have a huge potential considering climate change and devastating impacts of globalization on ecology and environment.

Environmental Technologies

River Ecology

Environment monitoring is an important aspect that is considered a necessity to deal with irregular changes or disturbances in our ecosystem. Therefore, researchers are developing tools using biotechnology and informatics to monitor the environment cost-effectively. Water resources are always under the influence of damaging anthropogenic pressures such as plastic waste and industrial sewage, which ultimately change or disturb the biogeochemical cycles and biodiversity. Besides, water is a universal solvent that holds the industries and economies together.

It is a well-established fact that diatoms hold the primary role in maintaining the aquatic ecosystem. Therefore, biodiversity assessment of diatom species in an environmental sample is one of the well-known strategies for biomonitoring. Presently, morphological assessment of the diatoms using microscopy is largely used which is time-consuming and requires special expertise (Larras et al., 2014). However, environmental metabarcoding has opened a quick way of analyzing the microbial DNA diversity in a natural environment such as flora and fauna (Bik et al., 2012; Taberlet et al., 2012). The metabarcoding approach is based on DNA sequencing a specific region (barcode) of the whole DNA extracted from an environmental sample (eDNA). For example, the sequencing data obtained from diatom metabarcoding are then used to assign precise taxonomic identification of the diatoms present in the eDNA sample, which are further compared with the conventional morphological database to confirm the efficacy of metabarcoding results. Diatoms metabarcoding tool has been optimized significantly to quantify the diversity of diatoms at the genus and species level (Vasselon et al., 2017; Kelly et al., 2018).

Currently, this approach is still in development, since various questions have been raised especially when deciding which are the most suitable barcodes. The barcodes that had been used are the ribosomal small subunit, cytochrome c, and the internal transcribed spacer region combined with the 5.8S rRNA gene (Zimmermann et al., 2011; Luddington et al., 2012).

Another main issue is processing the sequencing output data through computing. This method must be consistent with government policies for environmental regulation. For instance, MOTHUR is a comprehensive and efficient platform to study microbial diversity, but there are other bioinformatics software such as R, QIIME2 (Caporaso et al., 2010), LotuS (Hildebrand et al., 2014), and PIPITS (Gweon et al., 2015) that can be used to process a larger amount of data.

Additionally, various other research studies have supported the use of the diatoms metabarcoding approach as an alternative strategy to monitor river ecology on a timely basis. The results provide an estimated number of abundant and scarce species in samples obtained from different locations. Also, they give great insights into the fundamental status of the aquatic ecosystem (Larras et al., 2014). For instance, detailed evidence has been published by the Environmental Agency of the United Kingdom using diatoms indexes for river classification (Kelly et al., 2018). A similar study on detailed information on diatom biodiversity using metabarcoding has been conducted using environmental samples from Mayotte Island, France

(Vasselon et al., 2017). Moreover, a recently published work studied the impact of treated effluents on benthic diatom communities that showed a systematic change in diatom community composition (Chonova et al., 2019). Concluding that detailed information about diatom diversity will give in-depth insights into climate change, micropollutants, and other organic pollutants, to study the disturbing effects of anthropogenic pressure on rivers. The use of metabarcoding for analyzing biodiversity is rapidly increasing and has been adopted by academic institutes and various companies/industries like Spygen (Canada), Naturemetrics (United Kingdom), IGAtech (Italy), Sinsoma (Austria), to name a few. This particular strategy has been adopted by public authorities as well and has shown the potential to be used as an additional screening tool to replace the existing methods, which require excessive infrastructure and human resources. It is indeed possible to make it a primary and permanent tool for river monitoring with advancements in sequencing, big data science, and artificial intelligence tools.

Phytoremediation

Besides the monitoring of river quality, water treatment is one of the major concerns for many countries around the world. In fact, human consumption has undoubtedly increased in the last few decades, subsequently, incrementing waste products presence in aquatic communities (Walker, 1983). Globally, almost 80% of the wastewater generated worldwide is discharged on rivers creating health and environmental hazards. The rise of nutrient accumulation in the aquatic system needs to be neutralized to maintain the balance in the environment. Increasing of pollution is disturbing the basic biogeochemical cycles, killing fish, depleting the dissolved oxygen, and producing different toxins, i.e., neurotoxins (Boyd, 1990). Hence, there is an urgent need to explore new ways and upscale the existing systems to test reports and mitigate pollution from rivers and lakes worldwide.

The use of microalgae for wastewater treatment has been a subject of research for a long period which could be applied in collaboration with small- and large-scale industries. The excess of industrial waste discharged in the aquatic system can be used as nutrient supply by diatoms. Different kinds of wastewater such as brewery (Choi, 2016), aquaculture (Tossavainen et al., 2019), and textile (El-Kassas and Mohamed, 2014) have been studied for phytoremediation capability and have shown interesting results. The published studies have established that diatoms and microalgae can treat the wastewater to an extent, therefore, it would be less damaging to treat the wastewater with microalgae/diatoms before discharging in water bodies. In addition, use the harvested biomass for different industrial products such as biofuel. It is safe

to assume that it is possible to develop small scale business in collaboration with restaurants, breweries, textile industries, to name a few, to treat wastewater, and use the biomass for the production of valuable products such as fertilizers (Suleiman et al., 2020).

Heavy metal pollution is one of the major challenges which comes from the industries working with chemicals and dyes. Diatoms species are desirable organisms to study heavy metal pollution because of the simplicity of metal exposure, absorption, and detoxification of metal ions by single cells. This is a unique detoxification process of diatoms and microalgae due to metal-binding peptides known as phytochelatins (PCs) that protect photosynthetic organisms from heavy metals (Grill et al., 1985). Some intracellular PCs have been characterized in cultures of *P. tricornutum* exposed to different metals such as Cd, Pb, or Zn. Besides, they are used widely in waste degradation considering the unique structure of diatoms and their ability to respond to the changing environment (Glazer and Nikaido, 2007).

A study published in 2015 have reported a novel diatom *Bacillariophyta* sp. (BD1IITG) from petroleum biorefinery wastewater that can degrade phenol in a concentration range of 50−250 mg/L in Fog's media (Das et al., 2016). Another example of the degradation of toxic molecules like phenylalanine hydroxylase into less toxic compounds using simple enzymatic oxidation has been identified in diatoms during the metabolism of phenanthrene and pyrene (Wang and Zhao, 2007). These results are relevant considering that around seven billion kg of phenol is produced for oil refining, pesticide production, and to use in the pharmaceutical industry. Traditional phenol removal techniques involve several steps including the generation of by-products, which increments the cost of the treatment (Senthilvelan et al., 2014). However, there are very few reports available on exploiting the potential of diatoms in biodegrading waste materials. It is interesting to note that the studies have shown interesting results but the field of algae biotechnology requires more entrepreneurs to join the pieces of industrial and academic research to build a successful circular economy. Furthermore, there are some upcoming and growing ventures and companies in microalgae working in diverse applications and producing valuable products such as healthcare, animal feed, water management, chocolates, etc. (Table 7-3).

Table 8-3: ***Different industries producing variety of products from microalgae and diatoms around the world.***

Company	Products/services	Country	Website
Algae Biotechnologia	Wastewater treatment, animal nutrition, carbon dioxide fixation, biofuels, human health	Brazil	http://www.algae.com.br/site/pt/
Algae Farm	Omega3, diatom, water treatment and reuse, nutraceuticals, cosmeceuticals, algae based solar fuels cell, die sensitized solar panel, bioplastics	Canada	https://www.algaefarm.us/
Algorigin	Nutritional supplements	Switzerland	https://algorigin.com/en/
Algaetoomega	Omega 3, astaxanthin, animal feed	United States	https://algae2omega.com/
Algae Control Canada	Pond and lake water management	Canada	https://www.algaecontrol.ca/
The Algae Factory	Chocolate	Netherlands	http://thealgaefactory.com/
Algae Health	Antioxidants	United States	https://www.algaehealthsciences.com/
Swedish Algae Factory	Personal care products	Sweden	https://swedishalgaefactory.com/
Sabrtech	Recombinant proteins, fuel, nutraceuticals, aquaculture, etc.	Canada	https://www.sabrtech.ca/
Pondtech	Astaxanthin, aquaculture	Canada	https://www.pondtech.com/

Diatoms as Nutraceuticals and Feeds

Multiple epidemiological and clinical trials have shown the health benefits of omega fatty acids from fish oils and algae extracts (Cole et al., 2010; Cottin et al., 2011; Thomas et al., 2015; Wang and Daggy, 2017). Besides, there are few publications on cardio-protective and cognitive performance of omega fatty acids which have led to the commercial production of infant foods, infant formula, fortified snack bars, and other dairy products supplemented with omega fatty acids (Arterburn et al., 2007; Cottin et al., 2011).

Diatoms have an immense nutritional value that can be used to produce novel compounds such as antioxidants, vitamins, animal feed, and vegetarian protein supplements. Several photosynthetic pigments have been identified in diatoms including carotenoids such as fucoxanthin (Kuczynska et al., 2015). Additionally, *Nitzschia laevis*, *Nitzschia inconspicua*, *Navicula saprophila*, and *Phaeodactylum tricornutum* extracts have a noticeable amount of EPA and DHA that can be used as a nutritional feed in human diet and animal feed (Kitano et al., 1997; Wen and Chen, 2001a,b; Wah et al., 2015; Tocher et al., 2019).

Moreover, diatoms are known to have diverse defense mechanisms in form of chemical substances for them to be protected against pathogens. For instance, *P. tricornutum* has a high amount of omega-7 monounsaturated fatty acids such as palmitoleic acid (C16:1) and other bioactive compounds that are active against gram-positive pathogens (Desbois et al., 2009). Furthermore, the EPA-rich marine diatom, *Odontella aurita*, used as a dietary supplement has shown antioxidant effects in rats (Haimeur et al., 2012). *O. aurita* has been approved to be commercialized as food in France by following EC regulation 258/97 in 2002 (Pulz and Gross, 2004; Buono et al., 2014).

Increasing the content of these bioactive molecules in diatoms has attracted a large amount of research. Some studies have managed to enhance the production of flavonoid and polyphenol content by culture modifications, for instance, cultivation temperature and nutrient supplementation in *Amphora* sp. (Chtourou et al., 2015). The general tendency when changing the culture temperature is an increase in lipid content in most species, while the chemical composition varied between species (Renaud et al., 2002). For example, the total amount of saturated and monounsaturated fatty acids increases with temperature in *Rhodomonas* sp. (NT15) and *Cryptomonas* sp. (CRFI01). Whereas, there was a comparative decrease in polyunsaturated fatty acids in both *Rhodomonas* sp. (NT15) and *Cryptomonas* sp. (CRFI01) (Renaud et al., 2002).

Genetic Engineering of Diatoms

The debate on using genetically modified microalgae and diatoms is on-going. However, it is a more controlled alternative for the production of recombinant proteins or any precursor molecules, considering the use of bioreactors for their production. The employment of genetic engineering tools in diatoms, to produce or increase the yield of compounds, allows the companies to optimize their use in the applications mentioned above. Therefore, genetic engineering is a promising method and an important branch to be used in the diatoms industry to further enhance the economic value of diatoms. However, it comes with two big challenges, firstly, to redesign the natural metabolic pathways in order to increase the production of desired endogenous compounds, and secondly, producing new heterologous compounds.

In the last 20 years, several projects have shown that these challenges can be solved at lab scale, by optimization of transformation methods, utilization of different gene promoters, expression of recombinant proteins, gene silencing, and genome editing methods; such as targeted mutagenesis techniques using meganucleases, gene knockouts, TALENS, and CRISPR/Cas9. Marketable bioproducts like lipids, pigments, nanomaterials, food supplements, fuel, syntheses of chemicals, drugs, and metabolites have been produced in *P. tricornutum, T. pseudonana*, and other diatoms species. While most of these analyses are related to lipid production for biofuel or bioenergy purposes, other studies showed that diatoms are biological factories that can generate a wide range of products from food to pharmaceutics biomaterial industry (Lauritano et al., 2016; Mishra et al., 2017; Slattery et al., 2018; Dhaouadi et al., 2020; Sharma et al., 2020). In addition, there are few companies such as Algenol Biofuels, Synthetic Genomics, which have reported the use of genetically modified microalgae for the production of biofuels.

Moreover, the approach of synthetic biology along with high throughput sequencing technologies open the doors to understanding the whole genome, the proteins that it encodes, and the regulatory elements of the cell during cellular growth and division (Hildebrand and Lerch, 2015; Huang and Daboussi, 2017). Several sequencing projects have been performed in *P. tricornutum* and *T. pseudonana*strains (Armbrust et al., 2004; Bowler et al., 2008; Koester et al., 2018; Rastogi et al., 2018), generating the transcriptomic and proteomic data sets that make possible precise reconstructions of metabolic networks (Fabris et al., 2012; Levering et al., 2016). Recently, the Synthetic Diatoms Project website has been launched as a platform to provide information to grow, transform, edit, and analyze *P. tricornutum*and *T. pseudonana*[1] These projects have been used as a springboard to facilitate genome annotation for other diatoms species: *T. oceanica, T. weissflogii, Fragilariopsis cylindrus*, *Pseudo-nitzschia multiseries*, *Pseudo-nitzschia multistriata*, *Seminavis robusta, Fistulifera solaris*, *Cyclotella cryptica* (Table 7-4).

Table 8-4: ***Sequence Database of different diatoms species.***

Species	Genome database
Phaeodactylum tricornutum CCAP 1055/1	http://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index
Thalassiosira pseudonana CCMP 1335	https://genome.jgi.doe.gov/Thaps3/Thaps3.home.html
Thalassiosira oceanica CCMP 1005	https://genome.jgi.doe.gov/Thaoce1/Thaoce1.info.html
Thalassiosira weissflogii CCMP1030	https://genome.jgi.doe.gov/portal/
Fragilariopsis cylindrus CCMP 1102	https://genome.jgi.doe.gov/Fracy1/Fracy1.info.html
Pseudo-nitzschia multiseries CLN-47	https://genome.jgi.doe.gov/Psemu1/Psemu1.home.html
Pseudo-nitzschia multistriata B856	http://apollo.tgac.ac.uk/Pseudo-nitzschia_multistriata_V1_4_browser/sequences
Seminavis robusta D6	https://genome.jgi.doe.gov/portal/Semrobnscriptome/Semrobnscriptome.info.html
Fistulifera solaris JPCC DA058	https://trace.ddbj.nig.ac.jp/DRASearch/submission?acc5DRA002403
Cyclotella cryptica CCMP332	http://genomes.mcdb.ucla.edu/Cyclotella/download.html

Diatoms are a robust model for genome editing and cell transformation. Optimized methods of DNA delivery have been developed using biolistic or via electroporation. In both techniques, the transgenes are randomly integrated into the genome, with multiple integration events, variable transgene copy numbers, and chromosomal positions. The biolistic gene transfer method affects genome integrity due to the break and repair of the DNA double-strand by non-homologous end joining (NHEJ) (Zaboikin et al., 2017). However, this method is needed if the aim is to transform the chloroplast genome. An alternative transformation technique is the extrachromosomal-based expression approach that depends on vectors containing a yeast-derived sequence, which can be delivered through bacterial conjugation using *E. coli* (Karas et al., 2015).

An important element for genetic engineering is the promoter. The most commonly used are the light-regulated promoters of the fucoxanthin chlorophyll a/c-binding protein genes

fcpA/B/C/D (LHCF) (Zaslavskaia et al., 2000; Nymark et al., 2013). Alternatively, the elongation factor 2 (EF2) promoter sequence is a constitutive promoter (Seo et al., 2015). Recently, the most abundant secreted protein in *P. tricornutum* was identified, named "highly abundant secreted protein 1" (HASP1), and the activities of its promoter and the signal peptide were characterized using green fluorescent protein (GFP) as a reporter (Erdene-Ochir et al., 2019). A couple of inducible promoters have been reported: like nitrate reductase (NR) and alkaline phosphatase gene promoters in *P. tricornutum*, which are induced under nitrogen or phosphate starvation respectively (Slattery et al., 2018; Fabris et al., 2020) and glutamine synthetase gene promoter, induced by a blue light pulse (De Riso et al., 2009; Erdene-Ochir et al., 2016). In addition, promoter regions containing diatom-infecting viruses (DIVs) mediated a significantly higher level expression of the reporter gene in cells in the stationary phase compared to the exponential phase of growth (Kadono et al., 2015). Other elements needed for genetic engineering are reporter genes and selection markers. Among reporter genes, beta-glucuronidase *uidA* (GUS), fluorescent proteins like GFP/YFP/CFP, chloramphenicol acetyltransferase conferring resistance to chloramphenicol (CAT) and luciferase (LUC) are the most employed, other reporter proteins are listed in Table 7-5. The classic selection markers in diatoms are genes that confer resistance to zeocin, phleomycin, and nourseothricin, as shown in Table 7-5 are the most used. An alternative to using selective markers is the use of auxotrophic strains, such as uracil, histidine, and tryptophan auxotrophs (Sakaguchi et al., 2011; Slattery et al., 2020). Moreover, it is considered that the urease gene, either in an inactive or edited form, is an interesting tool for the selection of *P. tricornutum* and *T. pseudonana* strains (Weyman et al., 2015; Hopes et al., 2016; Slattery et al., 2018). An endogenous selectable marker in diatoms was generated by point mutations at a conserved residue Gly290 to Ser/Arg in the phytoene desaturase (PDS1) gene, which confers resistance to the herbicide norflurazon (Taparia et al., 2019).

Table 8-5: ***Diatoms genetic engineering.***

Species/strain	Genetic and molecular tools				
	Transformation methods and target compartment	**Promoters: (S) strong, (I) inducible and (H) heterologous**	**Reporters (R) and resistance (Re) genes**	**Expression of recombinant proteins**	**Genome editing methods and gene silencing**
Phaeodactylum tricomutum CCAP 1055/1	Biolistic (Cho et al., 2015) Electroporation (Niu et al., 2012) Conjugation (Zaboikin et al., 2017) Nuclear and chloroplast transformation (Xie et al., 2014)	**(S):** Lhcf (Fcp), light responsive (Karas et al., 2015), EF-1a, 40SRPS8, g-tubulin, RBCMT (Erdene-Ochir et al., 2019) and EF2 (Nymark et al., 2013), h4 (Fabris et al., 2020), HASP1 (De Riso et al., 2009). **(I):** rbcL (Xie et al., 2014), NR, low NO3 induce (Schellenberger Costa et al., 2012), V-ATPase C, AP1 low P induce (Lin et al., 2017) Fbp1, Fld, Isi1 iron-responsive (Yoshinaga et al., 2014) ca1, ca2 CO_2-responsive (Harada et al., 2005; Tanaka et al., 2016), U6, RNA polymerase III transcribed (Nymark et al., 2016) **(H):** CdP1, ClP1, ClP2, TnP1, TnP2 (Erdene-Ochir et al., 2016), CMV, RSV-LTR, PCMV, CaMV35S (Sakaue et al., 2008)	**(R):** GUS, GFP (Zhang and Hu, 2014), YFP, CFP (Zaboikin et al., 2017) cat (Karas et al., 2015), LUC (Cho et al., 2015), Aequorin (Falciatore et al., 2000) **(Re):** Zeocin and Phleomycin/sh ble, Nourseothricin/nat, Blasticidin-S/bsr, Streptothricin/sat, Neomycin/nptII (Karas et al., 2015)	Expression of Acyl-ACP thioesterases, increased accumulation of shorter chain (Radakovits et al., 2011). Malic enzyme (Trentacoste et al., 2013). G6PD (Wu et al., 2019), enhanced lipid productivity. Heterologous biosynthesis of the MIAs by CrGES expression under phototrophic conditions (Slattery et al., 2018), Vanillin production (Erdene-Ochir et al., 2019). PHBs for Bioplastics production (Hempel et al., 2011a). Human IgGαHBsAg:(Hempel et al., 2011b) and IgG1/kappa Ab CL4mAb: antibody to hepatitis B virus surface protein against the nucleoprotein of Marburg virus (Hempel and Maier, 2012). Over expression of DXS increased fucoxanthin synthesis (Eilers et al., 2016).	Targeted mutagenesis methods: meganucleases, gene knockouts, TALENS, and CRISPR/Cas9 (Poulsen and Kröger, 2005). Development of auxotrophic strains of *P. tricomutum* by CRISPR/Cas9 (Sakaguchi et al., 2011). A lipid producing strain through the disruption of the UDP-glucose pyrophosphorylase gene (Daboussi et al., 2014).
Thalassiosira pseudonana CCMP 1335	Biolistic (Poulsen et al., 2006) Electroporation (Buggé, 2015) Conjugation (Zaboikin et al., 2017)	**(S):** Lhcf9 **(I):** nr (161) SIT1, Si-starvation inducible (Davis et al., 2017), Thaps3_9619, Si-starvation inducible (Shrestha and Hildebrand, 2017), U6, RNA polymerase III transcribed (Weyman et al., 2015)	**(R):** YFP (Zaboikin et al., 2017) **(Re):** sh ble, nat (Poulsen et al., 2006)	Overexpression a multiple plasmids can be cotransformed; cloning multiple genes of interest Secretion of recombinant proteins has been shown. Localization of SiMat1-GFP (Kotzsch et al., 2016). Expression of the protective HslbpA DR2 antigen for the production of a vaccine against bovine respiratory disease (Davis et al., 2017), scFvTNT scFv and sdAbEA1 to detected *Bacillus anthracis* (Ford et al., 2016).	Targeted mutagenesis methods: meganucleases, TALENS, and CRISPR/Cas9 (Weyman et al., 2015). Gene silencing and gene knockouts are well established (Shrestha and Hildebrand, 2015).
Thalassiosira weissflogii	Biolistics (Cho et al., 2015)	**(S):** Lhcf2 (Cho et al., 2015)	**(R):** GUS (Cho et al., 2015)		
Pseudo-nitzschia multistriata B856	Biolistics (Sabatino et al., 2015)	**(S):** h4 (Sabatino et al., 2015)	**(Re):** sh ble (Sabatino et al., 2015)		
Pseudo-nitzschia arenysensis B858	Biolistics (Sabatino et al., 2015)		**(R):** GUS, GFP (Sabatino et al., 2015)		
Fistulifera solaris JPCC DA058	Biolistics (Muto et al., 2015)	**(S):** Lhcf2 and h4 **(H):** RSV and CaMV35S (Muto et al., 2015)	**(R):** GFP **(Re):** nptII (Muto et al., 2015)	Overexpression of the endogenous *GK* improve lipid productivity (Muto et al., 2015)	
Cylindrotheca fusiformis CCAP 1017/2 –CYL	Biolistics (Kong et al., 2019)	**(I):** nr (Kong et al., 2019)	**(R):** GFP **(Re):** sh ble (Kong et al., 2019)		
Navicula saprophila NAVICI	Biolistics (Dunahay et al., 1995)	**(S):** ACCase (Dunahay et al., 1995)	**(Re):** nptII (Dunahay et al., 1995)		
Chaetoceros gracilis UTEX LB2658	Biolistics (Ifuku et al., 2015)	**(S):** Lhcr5 **(I):** nr (Ifuku et al., 2015)	**(R):** GFP, LUC **(Re):** nat (Ifuku et al., 2015)		

DXS, 1-deoxy-D-xylulose 5-phosphate synthase; 40SRPS8, 40S ribosomal protein S8; ACCase, acetyl-CoA carboxylase; Acyl-ACP thioesterases, acyl–acyl carrier protein thioesterases; AP1, alkaline phosphatase 1; bsr, blasticidin-S resistance gene; Ca1, carbonic anhydrase 1; CaMV35S, cauliflower mosaic virus 35S; CdP, Chaetoceros debilis-infecting DNA virus; ClP, Chaetoceros lorenzianus-infecting DNA virus; cat, chloramphenicol acetyl transferase conferring resistance to chloramphenicol; CRISPR, clustered regularly interspaced short palindromic repeats; CFP, cyan fluorescent protein gene; CMV, cytomegalovirus; Fcp, diatom light-regulated promoters of the fucoxanthin chlorophyll a/c-binding protein genes Lhcf; EF-1α, elongation factor 1 alpha; EF2, elongation factor 2; Fbp1, ferrichrome binding protein1; Fld, flavodoxin; CrGES, Catharanthus roseus geraniol synthase; G6PD, glucose-6-phosphate dehydrogenase; GK, glycerol kinase; GFP, green fluorescent protein gene; HASP1, highly abundant secreted protein 1; h4, histone H4; human IgGαHBsAg, antibody against hepatitis B virus surface IgG1/kappa Ab CL4mAb; HslbpA DR2, IbpA DR2 antigen from Histophilus somni; Isi1, iron-starvation-induced gene 1; MIAs, monoterpenoid indole alkaloids; nptII, neomycin phosphotransferase II; NR, nitrate reductase; nat, nourseothricin acetyl transferase; P, phosphate; PHBs, polyhydroxybutyrate; PCMV, promoter sequences of the cytomegalovirus; psba, PSII reaction center core 2 quinones are associated with D1; Lhcr5, red algal-like LHCRs; RBCMT, ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit N-methyltransferase I; RSV-LTR, Rous sarcoma virus long terminal repeat; rbcL, Rubisco large subunit; SiMat1, silica matrix protein; SIT1, silicon transporter; scFvTNT, single chain antibodies; sdAbEA1, single domain antibodies; U6, small nuclear RNA of the U6 complex; sh ble, Streptoalloteichus hindustanus bleomycin resistance gene; TnP, Thalassionema nitzschioides-infecting DNA virus; TALENs, transcription activator-like effector nucleases; sat, treptothricin acetyl transferase; TAG, triacylglycerol; g-tubulin, tubulin gamma chain; GUS, uidA b-glucuronidase-encoding gene; V-ATPase C, vacuolar H+-ATPase; YFP, yellow fluorescent protein gene.

Concerning heterologous recombinant protein expression, diatom gene codon optimization is required for optimal expression; to avoid silencing expression and better protein translation. Although it has not been reported in diatoms, different projects which were done in green algae, have shown that including introns in the expression cassette can increase transcript abundance (Baier et al., 2018, 2020; Kong et al., 2019). In addition, 5′-UTR and 3′-UTR of nitrate reductase (NR) allow the control of timing and level of transgene expression in *C. fusiformis* (Poulsen and Kröger, 2005). Down-regulation of gene expression can be achieved through silencing by expressing antisense repeat sequences of target genes (Table 7-5).

Industrial processes using diatoms are cost-effective and have performed well in large-scale cultures (Benedetti et al., 2018). This is supported by the plasticity to adapt to extreme environmental conditions of diatoms, making them great candidates for sustainable biofactories (Kung et al., 2012; Cho et al., 2015; d'Ippolito et al., 2015). Altogether, these developments in metabolic pathways and synthesis of heterologous compounds represent promising insights for the improvement of yield, quality of products, and sustainability in the use of diatoms as cell factories.

8.2 Conclusion and Future Perspectives

The documented studies stated the astounding nature and possible all-round use of diatoms. This is one of the approaches to increase human consumption of renewable products and contributes toward reducing carbon emissions. Although the commercial application of diatoms still needs improvements, it is indeed a crucial research area for human wellbeing. For example, developments in diatoms research can lead to innovative products in domains of drug delivery, sensing, and detection parts to build complex biomedical devices and nanoparticles for waste degradation. Moreover, recent advancements in sequencing technology and processing large biological datasets have made it possible to label and store the global biodiversity of diatoms in all geographical locations.

One of the major challenges in diatom-based industries is scaling up the process for large-scale manufacturing which is dependent on many micro and macro factors such as cultivation, harvesting, drying, genetic modification, lack of genomic, proteomic, and metabolic information, etc. However, it is possible to overcome these challenges in near future with advancements in genetic tools, bioreactors, and other infrastructure changes. In general, there are many challenges in bio-based industries at different levels; academic/industrial research, infrastructure, policies, education, and information gaps. The advancements in academic research and discoveries are consistent considering the publications but it requires support from other domains such as the development of infrastructure, reducing the knowledge gaps between scientific researcher and entrepreneurs, changes in the policies at both national and international level. And to conclude, the recent research phenomenon blasted in the last decade, which is diatoms' industrial potential, still leaves many unsolved questions. Major questions will involve studying the extent of genetic or artificial manipulation without compromising its intact structure and delicate silica pattern. The unfolding of various missing links in genetic engineering, cultivation, and harvesting will make it possible to replicate complex plant

pathways in diatoms. These tools have opened the door to study diatoms for eco-friendly processes.

Although the use of silica for food and agriculture has been approved by the FDA and is also labeled/classified as non-carcinogenic by the International Agency for Research on Cancer, this could be a big step toward accelerating its use at the biomedical level. It is not yet approved for biomedicine as it requires long-term evidence (Terracciano et al., 2018). All the biomedical inventions are scrutinized by multiple stakeholders like research leaders, public authorities such as provincial and federal government, before they reach the stage of commercial distribution. It is understandable considering that it will be used directly in the human body. Therefore, an innovative and different approach is required to bring in the academic researchers and bio-entrepreneurs to speed up the innovation rate in biomedical industry without harming the screening process set by public health authorities. The collaboration between entrepreneurs and researchers will allow thorough evaluation of the market for new inventions, manufacturing, investment, and globalization of the product. It seems plausible considering the rapid advancements in the biomedical infrastructure around the world. This has been demonstrated by the quick inventions in response to COVID-19 and should be adopted to be applied in other biotech based industries (Harris et al., 2020). The simultaneous advancements in the use of silica-based support system for drug delivery along with the change in infrastructure in pharmaceutical industries and hospitals to deliver these technologies to the users is possible in the near future. The other requirement is to join the gap of vast and complex scientific information and knowledge between entrepreneurs and academic researchers.

Author Contributions

NS and ID-P conceived, designed, and led the study. NS, DS, AD-G, EF, and AM collected and analyzed the data, and prepared the figures and tables. NS, DS, AD-G, EF, AM, FM-M, HG, and ID-P authored and reviewed the drafts of the manuscript, and approved the final manuscript. All authors contributed to the article and approved the submitted version.

Conflict of Interest

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

Acknowledgments

We acknowledge that financial support for this review was funded by the Canada Research Chair on plant specialized metabolism Award No. 950-232164 to ID-P. Thanks are extended to the Canadian taxpayers and to the Canadian Government for supporting the Canada Research Chairs Program. Additional support in the form of scholarships to NS, DS, AD-G, EF, and AM from Mitacs—Acceleration program grants nos. IT12310 and IT16463 to ID-P is also acknowledged.

Footnotes

https://www.syntheticdiatoms.org/

8.3 References

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Èech, M., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 44, W3–W10.

Google Scholar

Ago, K., Hayashi, T., Ago, M., and Ogata, M. (2011). The number of diatoms recovered from the lungs and other organs in drowning deaths in bathwater. *Leg. Med. (Tokyo)*13, 186–190. doi: 10.1016/j.legalmed.2011.04.002

PubMed Abstract | CrossRef Full Text | Google Scholar

Alves Martins, D., Rocha, F., Castanheira, F., Mendes, A., Pousão-Ferreira, P., Bandarra, N., et al. (2013). Effects of dietary arachidonic acid on cortisol production and gene expression in stress response in *Senegalese sole* (*Solea senegalensis*) post-larvae. *Fish Physiol. Biochem.* 39, 1223–1238. doi: 10.1007/s10695-013-9778-6

PubMed Abstract | CrossRef Full Text | Google Scholar

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., et al. (2004). The genome of the diatom thalassiosira pseudonana : ecology, evolution, and metabolism. *Science* 306, 79–86. doi: 10.1126/science.1101156

PubMed Abstract | CrossRef Full Text | Google Scholar

Arterburn, L. M., Oken, H. A., Hoffman, J. P., Bailey-Hall, E., Chung, G., Rom, D., et al. (2007). Bioequivalence of docosahexaenoic acid from different algal oils in capsules and in a DHA-Fortified food. *Lipids* 42:1011. doi: 10.1007/s11745-007-3098-5

PubMed Abstract | CrossRef Full Text | Google Scholar

Aw, M. S., Simovic, S., Addai-Mensah, J., and Losic, D. (2011a). Polymeric micelles in porous and nanotubular implants as a new system for extended delivery of poorly soluble drugs. *J. Mater. Chem.* 21, 7082–7089. doi: 10.1039/c0jm04307a

CrossRef Full Text | Google Scholar

Aw, M. S., Simovic, S., Addai-Mensah, J., and Losic, D. (2011b). Silica microcapsules from diatoms as new carrier for delivery of therapeutics. *Nanomedicine (Lond)* 6, 1159–1173. doi: 10.2217/nnm.11.29

PubMed Abstract | CrossRef Full Text | Google Scholar

Bai, X., Song, H., Lavoie, M., Zhu, K., Su, Y., Ye, H., et al. (2016). Proteomic analyses bring new insights into the effect of a dark stress on lipid biosynthesis in *Phaeodactylum tricornutum*. *Sci. Rep.* 6:25494.

Google Scholar

Baier, T., Jacobebbinghaus, N., Einhaus, A., Lauersen, K. J., and Kruse, O. (2020). Introns mediate post-transcriptional enhancement of nuclear gene expression in the green microalga *Chlamydomonas reinhardtii*. *PLoS Genet.* 16:e1008944. doi: 10.1371/journal.pgen.1008944

PubMed Abstract | CrossRef Full Text | Google Scholar

Baier, T., Wichmann, J., Kruse, O., and Lauersen, K. J. (2018). Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga *Chlamydomonas reinhardtii*. *Nucleic Acids Res.* 46, 6909–6919. doi: 10.1093/nar/gky532

PubMed Abstract | CrossRef Full Text | Google Scholar

Benedetti, M., Vecchi, V., Barera, S., and Dall'Osto, L. (2018). Biomass from microalgae: the potential of domestication towards sustainable biofactories. *Microbial Cell Fact.* 17:173.

Google Scholar

Bik, H. M., Porazinska, D. L., Creer, S., Caporaso, J. G., Knight, R., and Thomas, W. K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol. Evol.* 27, 233–243. doi: 10.1016/j.tree.2011.11.010

PubMed Abstract | CrossRef Full Text | Google Scholar

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244.

Google Scholar

Boyd, C. E. (1990). *Water Quality in Ponds for Aquaculture.* Auburn StateAL: Alabama Agricultural Experiment Station.

Google Scholar

Brocks, J. J., Buick, R., Logan, G. A., and Summons, R. E. (2003). Composition and syngeneity of molecular fossils from the 2.78 to 2.45 billion-year-old mount bruce supergroup, pilbara craton, Western Australia. *Geochimica et Cosmochimica Acta* 67, 4289–4319. doi: 10.1016/s0016-7037(03)00208-4

CrossRef Full Text | Google Scholar

Buggé, J. A. (2015). *Electroporation-Mediated Transformation and Post-Transcriptional Gene Regulation of Nitrate Reductase in the Marine Diatom Thalassiosira Pseudonana.* Worcester, MA: Clark University.

Google Scholar

Buono, S., Langellotti, A. L., Martello, A., Rinna, F., and Fogliano, V. (2014). Functional ingredients from microalgae. *Food Funct.* 5, 1669–1685. doi: 10.1039/c4fo00125g

PubMed Abstract | CrossRef Full Text | Google Scholar

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336.

Google Scholar

Chandrasekaran, S., Sweetman, M. J., Kant, K., Skinner, W., Losic, D., Nann, T., et al. (2014). Silicon diatom frustules as nanostructured photoelectrodes. *Chem. Commun.* 50, 10441–10444. doi: 10.1039/c4cc04470c

PubMed Abstract | CrossRef Full Text | Google Scholar

Cho, C., Choi, S. Y., Luo, Z. W., and Lee, S. Y. (2015). Recent advances in microbial production of fuels and chemicals using tools and strategies of systems metabolic engineering. *Biotechnol. Adv.* 33, 1455–1466. doi: 10.1016/j.biotechadv.2014.11.006

PubMed Abstract | CrossRef Full Text | Google Scholar

Choi, H.-J. (2016). Parametric study of brewery wastewater effluent treatment using *Chlorella vulgaris* microalgae. *Environ. Eng. Res.* 21, 401–408. doi: 10.4491/eer.2016.024

CrossRef Full Text | Google Scholar

Chonova, T., Kurmayer, R., Rimet, F., Labanowski, J., Vasselon, V., Keck, F., et al. (2019). Benthic diatom communities in an alpine river impacted by waste water treatment effluents as revealed using DNA metabarcoding. *Front. Microbiol.* 10:653. doi: 10.3389/fmicb.2019.00653

PubMed Abstract | CrossRef Full Text | Google Scholar

Chtourou, H., Dahmen, I., Jebali, A., Karray, F., Hassairi, I., Abdelkafi, S., et al. (2015). Characterization of *Amphora sp.*, a newly isolated diatom wild strain, potentially usable for biodiesel production. *Bioprocess Biosystems Eng.* 38, 1381–1392. doi: 10.1007/s00449-015-1379-6

PubMed Abstract | CrossRef Full Text | Google Scholar

Cicco, S. R., Vona, D., De Giglio, E., Cometa, S., Mattioli-Belmonte, M., Palumbo, F., et al. (2015). Chemically modified diatoms biosilica for bone cell growth with combined drug-delivery and antioxidant properties. *ChemPlusChem* 80, 1104–1112. doi: 10.1002/cplu.201402398

PubMed Abstract | CrossRef Full Text | Google Scholar

Cole, G. M., Ma, Q.-L., and Frautschy, S. A. (2010). Dietary fatty acids and the aging brain. *Nutr. Rev.* 68, (Suppl 2), S102–S111.

Google Scholar

Cottin, S. C., Sanders, T. A., and Hall, W. L. (2011). The differential effects of EPA and DHA on cardiovascular risk factors. *Proc. Nutrition Soc.* 70, 215–231. doi: 10.1017/s0029665111000061

PubMed Abstract | CrossRef Full Text | Google Scholar

Curnow, P., Senior, L., Knight, M. J., Thamatrakoln, K., Hildebrand, M., and Booth, P. J. (2012). Expression, purification, and reconstitution of a diatom silicon transporter. *Biochemistry* 51, 3776–3785. doi: 10.1021/bi3000484

PubMed Abstract | CrossRef Full Text | Google Scholar

d'Ippolito, G., Sardo, A., Paris, D., Vella, F. M., Adelfi, M. G., Botte, P., et al. (2015). Potential of lipid metabolism in marine diatoms for biofuel production. *Biotechnol. Biofuels* 8:28. doi: 10.1186/s13068-015-0212-4

PubMed Abstract | CrossRef Full Text | Google Scholar

Daboussi, F., Leduc, S., Maréchal, A., Dubois, G., Guyot, V., Perez-Michaut, C., et al. (2014). Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology. *Nat. Commun.* 5:3831.

Google Scholar

Das, B., Mandal, T. K., and Patra, S. (2016). Biodegradation of phenol by a novel diatom BD1IITG-kinetics and biochemical studies. *Int. J. Environ. Sci. Technol.* 13, 529–542. doi: 10.1007/s13762-015-0857-3

CrossRef Full Text | Google Scholar

Davis, A., Crum, L. T., Corbeil, L. B., and Hildebrand, M. (2017). Expression of *Histophilus somni* IbpA DR2 protective antigen in the diatom *Thalassiosira pseudonana. Appl. Microbiol. Biotechnol.* 101, 5313–5324. doi: 10.1007/s00253-017-8267-8

PubMed Abstract | CrossRef Full Text | Google Scholar

De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C., and Falciatore, A. (2009). Gene silencing in the marine diatom *Phaeodactylum tricornutum. Nucleic Acids Res.*37:e96. doi: 10.1093/nar/gkp448

PubMed Abstract | CrossRef Full Text | Google Scholar

De Stefano, L., Lamberti, A., Rotiroti, L., and De Stefano, M. (2008). Interfacing the nanostructured biosilica microshells of the marine diatom *Coscinodiscus wailesii*with biological matter. *Acta Biomaterialia* 4, 126–130. doi: 10.1016/j.actbio.2007.09.003

PubMed Abstract | CrossRef Full Text | Google Scholar

De Stefano, L., Rotiroti, L., De Stefano, M., Lamberti, A., Lettieri, S., Setaro, A., et al. (2009). Marine diatoms as optical biosensors. *Biosens. Bioelectron.* 24, 1580–1584. doi: 10.1016/j.bios.2008.08.016

PubMed Abstract | CrossRef Full Text | Google Scholar

De Tommasi, E. (2016). Light manipulation by single cells: the case of diatoms. *J. Spectroscopy* 2016:2490128.

Google Scholar

Delalat, B., Sheppard, V. C., Rasi Ghaemi, S., Rao, S., Prestidge, C. A., McPhee, G., et al. (2015). Targeted drug delivery using genetically engineered diatom biosilica. *Nat. Commun.* 6:8791.

Google Scholar

Desbois, A. P., Mearns-Spragg, A., and Smith, V. J. (2009). A fatty acid from the diatom *Phaeodactylum tricornutum* is antibacterial against diverse bacteria including multi-resistant *Staphylococcus aureus* (MRSA). *Mar. Biotechnol.* 11, 45–52. doi: 10.1007/s10126-008-9118-5

PubMed Abstract | CrossRef Full Text | Google Scholar

Dhaouadi, F., Awwad, F., Diamond, A., and Desgagne-Penix, I. (2020). Diatoms' breakthroughs in biotechnology: *Phaeodactylum tricornutum* as a model for producing high-added value molecules. *Am. J. Plant Sci.* 11, 1632–1670. doi: 10.4236/ajps.2020.1110118

CrossRef Full Text | Google Scholar

DiMasi, J. A., Grabowski, H. G., and Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econom.* 47, 20–33. doi: 10.1016/j.jhealeco.2016.01.012

PubMed Abstract | CrossRef Full Text | Google Scholar

Dolatabadi, J. E. N., Mashinchian, O., Ayoubi, B., Jamali, A. A., Mobed, A., Losic, D., et al. (2011). Optical and electrochemical DNA nanobiosensors. *TrAC Trends Anal. Chem.* 30, 459–472. doi: 10.1016/j.trac.2010.11.010

CrossRef Full Text | Google Scholar

Dunahay, T. G., Jarvis, E. E., and Roessler, P. G. (1995). Genetic transformation of the diatoms cyclotella cryptica and navicula saprophila. *J. Phycol.* 31, 1004–1012. doi: 10.1111/j.0022-3646.1995.01004.x

CrossRef Full Text | Google Scholar

Dunstan, G. A., Volkman, J. K., Barrett, S. M., Leroi, J.-M., and Jeffrey, S. W. (1993). Essential polyunsaturated fatty acids from 14 species of diatom (Bacillariophyceae). *Phytochemistry* 35, 155–161. doi: 10.1016/s0031-9422(00)90525-9

CrossRef Full Text | Google Scholar

Dyall, S. C. (2015). Long-chain omega-3 fatty acids and the brain: a review of the independent and shared effects of EPA, DPA and DHA. *Front. Aging Neurosci.* 7:52. doi: 10.3389/fnagi.2015.00052

PubMed Abstract | CrossRef Full Text | Google Scholar

Eilers, U., Bikoulis, A., Breitenbach, J., Büchel, C., and Sandmann, G. (2016). Limitations in the biosynthesis of fucoxanthin as targets for genetic engineering in *Phaeodactylum tricornutum*. *J. Appl. Phycol.* 28, 123–129. doi: 10.1007/s10811-015-0583-8

CrossRef Full Text | Google Scholar

El-Kassas, H. Y., and Mohamed, L. A. (2014). Bioremediation of the textile waste effluent by *Chlorella vulgaris*. *Egyptian J. Aquatic Res.* 40, 301–308. doi: 10.1016/j.ejar.2014.08.003

CrossRef Full Text | Google Scholar

Ellwood, M. J., and Hunter, K. A. (2000). The incorporation of zinc and iron into the frustule of the marine diatom *Thalassiosira pseudonana*. *Limnol. Oceanography* 45, 1517–1524. doi: 10.4319/lo.2000.45.7.1517

CrossRef Full Text | Google Scholar

Erdene-Ochir, E., Shin, B.-K., Huda, M. N., Kim, D. H., Lee, E. H., Song, D.-G., et al. (2016). Cloning of a novel endogenous promoter for foreign gene expression in *Phaeodactylum tricornutum*. *Appl. Biol. Chem.* 59, 861–867. doi: 10.1007/s13765-016-0235-y

CrossRef Full Text | Google Scholar

Erdene-Ochir, E., Shin, B.-K., Kwon, B., Jung, C., and Pan, C.-H. (2019). Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from *Phaeodactylum tricornutum*. *Sci. Rep.* 9:9941.

Google Scholar

Fabris, M., George, J., Kuzhiumparambil, U., Lawson, C. A., Jaramillo-Madrid, A. C., Abbriano, R. M., et al. (2020). Extrachromosomal genetic engineering of the marine diatom phaeodactylum tricornutum enables the heterologous production of monoterpenoids. *ACS Synthetic Biol.* 9, 598–612. doi: 10.1021/acssynbio.9b00455

PubMed Abstract | CrossRef Full Text | Google Scholar

Fabris, M., Matthijs, M., Rombauts, S., Vyverman, W., Goossens, A., and Baart, G. J. E. (2012). The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner–Doudoroff glycolytic pathway. *Plant J.* 70, 1004–1014. doi: 10.1111/j.1365-313x.2012.04941.x

PubMed Abstract | CrossRef Full Text | Google Scholar

Falciatore, A., d'Alcalà, M. R., Croot, P., and Bowler, C. (2000). Perception of environmental signals by a marine diatom. *Science* 288, 2363–2366. doi: 10.1126/science.288.5475.2363

PubMed Abstract | CrossRef Full Text | Google Scholar

Falkowski, P. G., Barber, R. T., and Smetacek, V. V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science* 281, 200–207. doi: 10.1126/science.281.5374.200

PubMed Abstract | CrossRef Full Text | Google Scholar

Fang, T., Liao, S., Chen, X., Zhao, Y., Zhu, Q., Cao, Y., et al. (2019). Forensic drowning site inference employing mixed pyrosequencing profile of DNA barcode gene (rbcL). *Int. J. Legal Med.* 133, 1351–1360. doi: 10.1007/s00414-019-02075-4

PubMed Abstract | CrossRef Full Text | Google Scholar

Ford, N. R., Hecht, K. A., Hu, D., Orr, G., Xiong, Y., Squier, T. C., et al. (2016). Antigen binding and site-directed labeling of biosilica-immobilized fusion proteins expressed in diatoms. *ACS Synth Biol.* 5, 193–199. doi: 10.1021/acssynbio.5b00191

PubMed Abstract | CrossRef Full Text | Google Scholar

Ford, N. R., Xiong, Y., Hecht, K. A., Squier, T. C., Rorrer, G. L., and Roesijadi, G. (2020). Optimizing the design of diatom biosilica-targeted fusion proteins in biosensor construction for bacillus anthracis detection. *Biology* 9:14. doi: 10.3390/biology9010014

PubMed Abstract | CrossRef Full Text | Google Scholar

Gallagher, B. J. (2011). The economics of producing biodiesel from algae. *Renewable Energy* 36, 158–162. doi: 10.1016/j.renene.2010.06.016

CrossRef Full Text | Google Scholar

Gannavarapu, K. P., Ganesh, V., Thakkar, M., Mitra, S., and Dandamudi, R. B. (2019). Nanostructured Diatom-ZrO2 composite as a selective and highly sensitive enzyme free electrochemical sensor for detection of methyl parathion. *Sensors Actuators B: Chem.* 288, 611–617. doi: 10.1016/j.snb.2019.03.036

PubMed Abstract | CrossRef Full Text | Google Scholar

Glazer, A. N., and Nikaido, H. (2007). *Microbial Biotechnology: Fundamentals of Applied Microbiology*. Cambridge: Cambridge University Press.

Google Scholar

Gnanamoorthy, P., Anandhan, S., and Prabu, V. A. (2014). Natural nanoporous silica frustules from marine diatom as a biocarrier for drug delivery. *J. Porous Mater.* 21, 789–796. doi: 10.1007/s10934-014-9827-2

CrossRef Full Text | Google Scholar

Grill, E., Winnacker, E. L., and Zenk, M. H. (1985). Phytochelatins: the principal heavy-metal complexing peptides of higher plants. *Science* 230, 674–676. doi: 10.1126/science.230.4726.674

PubMed Abstract | CrossRef Full Text | Google Scholar

Gutu, T., Gale, D. K., Jeffryes, C., Wang, W., Chang, C.-H., Rorrer, G. L., et al. (2009). Electron microscopy and optical characterization of cadmium sulphide nanocrystals deposited on the patterned surface of diatom biosilica. *J. Nanomater.* 2009:860536.

Google Scholar

Gweon, H. S., Oliver, A., Taylor, J., Booth, T., Gibbs, M., Read, D. S., et al. (2015). PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the illumina sequencing platform. *Methods Ecol. Evol.* 6, 973–980. doi: 10.1111/2041-210x.12399

PubMed Abstract | CrossRef Full Text | Google Scholar

Haimeur, A., Ulmann, L., Mimouni, V., Guéno, F., Pineau-Vincent, F., Meskini, N., et al. (2012). The role of *Odontella aurita*, a marine diatom rich in EPA, as a dietary supplement in dyslipidemia, platelet function and oxidative stress in high-fat fed rats. *Lipids Health Dis.* 11, 147–147. doi: 10.1186/1476-511x-11-147

PubMed Abstract | CrossRef Full Text | Google Scholar

Harada, H., Nakatsuma, D., Ishida, M., and Matsuda, Y. (2005). Regulation of the expression of intracellular β -Carbonic anhydrase in response to CO 2 and light in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiol.* 139, 1041–1050. doi: 10.1104/pp.105.065185

PubMed Abstract | CrossRef Full Text | Google Scholar

Harris, M., Bhatti, Y., Buckley, J., and Sharma, D. (2020). Fast and frugal innovations in response to the COVID-19 pandemic. *Nat. Med.* 26, 814–817. doi: 10.1038/s41591-020-0889-1

PubMed Abstract | CrossRef Full Text | Google Scholar

Hempel, F., and Maier, U. G. (2012). An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microbial Cell Factories* 11:126. doi: 10.1186/1475-2859-11-126

PubMed Abstract | CrossRef Full Text | Google Scholar

Hempel, F., Bozarth, A. S., Lindenkamp, N., Klingl, A., Zauner, S., Linne, U., et al. (2011a). Microalgae as bioreactors for bioplastic production. *Microbial Cell Factories*10:81. doi: 10.1186/1475-2859-10-81

PubMed Abstract | CrossRef Full Text | Google Scholar

Hempel, F., Lau, J., Klingl, A., and Maier, U. G. (2011b). Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PLoS One* 6:e28424. doi: 10.1371/journal.pone.0028424

PubMed Abstract | CrossRef Full Text | Google Scholar

Hildebrand, F., Tadeo, R., Voigt, A. Y., Bork, P., and Raes, J. (2014). LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome* 2:30. doi: 10.1186/2049-2618-2-30

PubMed Abstract | CrossRef Full Text | Google Scholar

Hildebrand, M., and Lerch, S. J. L. (2015). Diatom silica biomineralization: parallel development of approaches and understanding. *Sem. Cell Dev. Biol.* 46, 27–35. doi: 10.1016/j.semcdb.2015.06.007

PubMed Abstract | CrossRef Full Text | Google Scholar

Hopes, A., Nekrasov, V., Kamoun, S., and Mock, T. (2016). Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. *Plant Methods* 12:49.

Google Scholar

Horton, B. P., Boreham, S., and Hillier, C. (2006). The development and application of a diatom-based quantitative reconstruction technique in forensic science. *J. Forensic Sci.* 51, 643–650. doi: 10.1111/j.1556-4029.2006.00120.x

PubMed Abstract | CrossRef Full Text | Google Scholar

Huang, W., and Daboussi, F. (2017). Genetic and metabolic engineering in diatoms. *Philos. Trans. R. Soc. B: Biol. Sci.* 372:20160411. doi: 10.1098/rstb.2016.0411

PubMed Abstract | CrossRef Full Text | Google Scholar

Huesemann, M. H., and Benemann, J. R. (2009). "Biofuels from microalgae: review of products, processes and potential, with special focus on *Dunaliella sp*," in *The Alga Dunaliella: Biodiversity, Physiology, Genomics and Biotechnology*, eds A. Ben-Amotz, J. E. Polle, and D. V. W. Subba Rao (New Hampshire, NH: Science Publishers).

Google Scholar

Hürlimann, J., Feer, P., Elber, F., Niederberger, K., Dirnhofer, R., and Wyler, D. (2000). Diatom detection in the diagnosis of death by drowning. *Int. J. Legal Med.* 114, 6–14. doi: 10.1007/s004149900122

PubMed Abstract | CrossRef Full Text | Google Scholar

Ifuku, K., Yan, D., Miyahara, M., Inoue-Kashino, N., Yamamoto, Y. Y., and Kashino, Y. (2015). A stable and efficient nuclear transformation system for the diatom *Chaetoceros gracilis*. *Photosynth Res.* 123, 203–211. doi: 10.1007/s11120-014-0048-y

PubMed Abstract | CrossRef Full Text | Google Scholar

Jaccard, T., Ariztegui, D., and Wilkinson, K. J. (2009). Incorporation of zinc into the frustule of the freshwater diatom *Stephanodiscus hantzschii*. *Chem. Geol.* 265, 381–386. doi: 10.1016/j.chemgeo.2009.04.016

CrossRef Full Text | Google Scholar

Jensen, E. L., Yangüez, K., Carrière, F., and Gontero, B. (2020). Storage compound accumulation in diatoms as response to elevated CO2 concentration. *Biology* 9:5. doi: 10.3390/biology9010005

PubMed Abstract | CrossRef Full Text | Google Scholar

Jin, P., and Agustí, S. (2018). Fast adaptation of tropical diatoms to increased warming with trade-offs. *Sci. Rep.* 8:17771.

Google Scholar

Kadono, T., Miyagawa-Yamaguchi, A., Kira, N., Tomaru, Y., Okami, T., Yoshimatsu, T., et al. (2015). Characterization of marine diatom-infecting virus promoters in the model diatom *Phaeodactylum tricornutum*. *Sci. Rep.* 5:18708.

Google Scholar

Karaman, E. S., Wang, Z., Di Benedetto, G., Zunino, J. L., Meng, X., and Mitra, S. (2019). Fabrication of supercapacitors and flexible electrodes using biosilica from cultured diatoms. *Mater. Today Energy* 11, 166–173. doi: 10.1016/j.mtener.2018.11.004

CrossRef Full Text | Google Scholar

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P. R., Noddings, C. M., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 6:6925.

Google Scholar

Kato, K., Lee, S., and Nagata, F. (2020). Efficient enzyme encapsulation inside sol-gel silica sheets prepared by poly-L-lysine as a catalyst. *J. Asian Ceramic Soc.* 8, 396–406. doi: 10.1080/21870764.2020.1747167

CrossRef Full Text | Google Scholar

Katz, M. E., Finkel, Z. V., Grzebyk, D., Knoll, A. H., and Falkowski, P. G. (2004). Evolutionary trajectories and biogeochemical impacts of marine eukaryotic phytoplankton. *Annual Rev. Ecol. Evol. Systemat.* 35, 523–556. doi: 10.1146/annurev.ecolsys.35.112202.130137

CrossRef Full Text | Google Scholar

Kelly, M., Boonham, N., Juggins, S., Kille, P., Mann, D., Pass, D., et al. (2018). *A DNA Based Diatom Metabarcoding Approach for Water Framework Directive Classification of Rivers.* Bristol: Environnment Agency.

Google Scholar

Kitano, M., Matsukawa, R., and Karube, I. (1997). Changes in eicosapentaenoic acid content of *Navicula saprophila*, *Rhodomonas salina* and *Nitzschia sp.* under mixotrophic conditions. *J. Appl. Phycol.* 9, 559–563.

Google Scholar

Knight, M. J., Senior, L., Nancolas, B., Ratcliffe, S., and Curnow, P. (2016). Direct evidence of the molecular basis for biological silicon transport. *Nat. Commun.* 7, 11926–11926.

Google Scholar

Knothe, G. (2005). Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. *Fuel Process. Technol.* 86, 1059–1070. doi: 10.1016/j.fuproc.2004.11.002

CrossRef Full Text | Google Scholar

Knothe, G. (2009). Improving biodiesel fuel properties by modifying fatty ester composition. *Energy Environ. Sci.* 2, 759–766. doi: 10.1039/b903941d

CrossRef Full Text | Google Scholar

Knothe, G. (2012). Fuel properties of highly polyunsaturated fatty acid methyl esters. prediction of fuel properties of algal biodiesel. *Energy Fuels* 26, 5265–5273. doi: 10.1021/ef300700v

CrossRef Full Text | Google Scholar

Koester, J. A., Berthiaume, C. T., Hiranuma, N., Parker, M. S., Iverson, V., Morales, R., et al. (2018). Sexual ancestors generated an obligate asexual and globally dispersed clone within the model diatom species *Thalassiosira pseudonana*. *Sci. Rep.* 8:10492.

Google Scholar

Kong, F., Yamaoka, Y., Ohama, T., Lee, Y., and Li-Beisson, Y. (2019). Molecular genetic tools and emerging synthetic biology strategies to increase cellular oil content in *Chlamydomonas reinhardtii*. *Plant Cell Physiol.* 60, 1184–1196. doi: 10.1093/pcp/pcz022

PubMed Abstract | CrossRef Full Text | Google Scholar

Kotzsch, A., Pawolski, D., Milentyev, A., Shevchenko, A., Scheffel, A., Poulsen, N., et al. (2016). Biochemical composition and assembly of biosilica-associated insoluble organic matrices from the diatom *Thalassiosira pseudonana*. *J. Biol. Chem.* 291, 4982–4997. doi: 10.1074/jbc.m115.706440

PubMed Abstract | CrossRef Full Text | Google Scholar

Kuczynska, P., Jemiola-Rzeminska, M., and Strzalka, K. (2015). Photosynthetic pigments in diatoms. *Mar. Drugs* 13, 5847–5881. doi: 10.3390/md13095847

PubMed Abstract | CrossRef Full Text | Google Scholar

Kung, Y., Runguphan, W., and Keasling, J. D. (2012). From fields to fuels: recent advances in the microbial production of biofuels. *ACS Synthetic Biol.* 1, 498–513. doi: 10.1021/sb300074k

PubMed Abstract | CrossRef Full Text | Google Scholar

Larras, F., Keck, F., Montuelle, B., Rimet, F., and Bouchez, A. (2014). Linking diatom sensitivity to herbicides to phylogeny: a step forward for biomonitoring? *Environ. Sci. Technol.* 48, 1921–1930. doi: 10.1021/es4045105

PubMed Abstract | CrossRef Full Text | Google Scholar

Lauritano, C., Andersen, J. H., Hansen, E., Albrigtsen, M., Escalera, L., Esposito, F., et al. (2016). Bioactivity screening of microalgae for antioxidant, anti-inflammatory, anticancer, anti-diabetes, and antibacterial activities. *Front. Mar. Sci.* 3:68. doi: 10.3389/fmars.2016.00068

CrossRef Full Text | Google Scholar

Levering, J., Broddrick, J., Dupont, C. L., Peers, G., Beeri, K., Mayers, J., et al. (2016). Genome-Scale model reveals metabolic basis of biomass partitioning in a model diatom. *PLoS One* 11:e0155038. doi: 10.1371/journal.pone.0155038

PubMed Abstract | CrossRef Full Text | Google Scholar

Levitan, O., Dinamarca, J., Hochman, G., and Falkowski, P. G. (2014). Diatoms: a fossil fuel of the future. *Trends Biotechnol.* 32, 117–124. doi: 10.1016/j.tibtech.2014.01.004

PubMed Abstract | CrossRef Full Text | Google Scholar

Levkov, Z., Williams, D. M., Nikolovska, D., Tofilovska, S., Èakar, Z., Williams, M., et al. (2017). *The Use of Diatoms in Forensic Science: Advantages and limitations of the diatom test in cases of drowning. The Archaeological and Forensic Applications of Microfossils: A Deeper Understanding of Human History.* London: Geological Society of London,Google Scholar

Lin, G., Wang, Y., Guo, L., Ding, H., Hu, Y., Liang, S., et al. (2017). Verification of mutagen function of Zeocin in *Nannochloropsis oceanica* through transcriptome analysis. *J. Ocean University country-regionChina* 16, 501–508. doi: 10.1007/s11802-017-3231-x

CrossRef Full Text | Google Scholar

Losic, D., Mitchell, J. G., and Voelcker, N. H. (2005). Complex gold nanostructures derived by templating from diatom frustules. *Chem. Commun.* 39, 4905–4907. doi: 10.1039/b508733c

PubMed Abstract | CrossRef Full Text | Google Scholar

Losic, D., Yu, Y., Aw, M. S., Simovic, S., Thierry, B., and Addai-Mensah, J. (2010). Surface functionalisation of diatoms with dopamine modified iron-oxide nanoparticles: toward magnetically guided drug microcarriers with biologically derived morphologies. *Chem. Commun.* 46, 6323–6325. doi: 10.1039/c0cc01305f

PubMed Abstract | CrossRef Full Text | Google Scholar

Luddington, I. A., Kaczmarska, I., and Lovejoy, C. (2012). Distance and character-based evaluation of the V4 region of the 18S rRNA Gene for the identification of diatoms (*Bacillariophyceae*). *PLoS One* 7:e45664. doi: 10.1371/journal.pone.0045664

PubMed Abstract | CrossRef Full Text | Google Scholar

Ludes, B., Coste, M., Tracqui, A., and Mangin, P. (1996). Continuous river monitoring of the diatoms in the diagnosis of drowning. *J. Forensic Sci.* 41, 425–428.

Google Scholar

Lunetta, P., and Modell, J. H. (2005). "Macroscopical, microscopical, and laboratory findings in drowning victims," in *Forensic Pathology Reviews. Forensic Pathology Reviews*, ed. M. Tsokos (Totowa, NJ: Humana Press).

Google Scholar

Lunetta, P., Penttilä, A., and Hällfors, G. (1998). Scanning and transmission electron microscopical evidence of the capacity of diatoms to penetrate the alveolo-capillary barrier in drowning. *Int. J. Legal Med.* 111, 229–237. doi: 10.1007/s004140050159

PubMed Abstract | CrossRef Full Text | Google Scholar

Mann, D. G., and Vanormelingen, P. (2013). An inordinate fondness? the number, distributions, and origins of diatom species. *J. Eukaryotic Microbiol.* 60, 414–420. doi: 10.1111/jeu.12047

PubMed Abstract | CrossRef Full Text | Google Scholar

Marella, T. K., López-Pacheco, I. Y., Parra-Saldívar, R., Dixit, S., and Tiwari, A. (2020). Wealth from waste: diatoms as tools for phycoremediation of wastewater and for obtaining value from the biomass. *Sci. Total Environ.* 724:137960. doi: 10.1016/j.scitotenv.2020.137960

PubMed Abstract | CrossRef Full Text | Google Scholar

Martin-Jézéquel, V., Hildebrand, M., and Brzezinski, M. A. (2000). Silicon metabolism in diatoms: implications for growth. *J. Phycol.* 36, 821–840. doi: 10.1046/j.1529-8817.2000.00019.x

CrossRef Full Text | Google Scholar

Medlin, L. K. (2016). Evolution of the diatoms: major steps in their evolution and a review of the supporting molecular and morphological evidence. *Phycologia* 55, 79–103. doi: 10.2216/15-105.1

CrossRef Full Text | Google Scholar

Milligan, A. J., and Morel, F. M. M. (2002). A proton buffering role for silica in diatoms. *Science* 297, 1848–1850. doi: 10.1126/science.1074958

PubMed Abstract | CrossRef Full Text | Google Scholar

Milović, M., Simović, S., Lošić, D., Dashevskiy, A., and Ibrić, S. (2014). Solid self-emulsifying phospholipid suspension (SSEPS) with diatom as a drug carrier. *Eur. J. Pharm. Sci.* 63, 226–232. doi: 10.1016/j.ejps.2014.07.010

PubMed Abstract | CrossRef Full Text | Google Scholar

Mishra, M., Arukha, A. P., Bashir, T., Yadav, D., and Prasad, G. B. K. S. (2017). All new faces of diatoms: potential source of nanomaterials and beyond. *Front. Microbiol.*8:1239. doi: 10.3389/fmicb.2017.01239

PubMed Abstract | CrossRef Full Text | Google Scholar

Muto, M., Tanaka, M., Liang, Y., Yoshino, T., Matsumoto, M., and Tanaka, T. (2015). Enhancement of glycerol metabolism in the oleaginous marine diatom *Fistulifera solaris* JPCC DA0580 to improve triacylglycerol productivity. *Biotechnol. Biofuels*8:4. doi: 10.1186/s13068-014-0184-9

PubMed Abstract | CrossRef Full Text | Google Scholar

Nanjappa, D., Audic, S., Romac, S., Kooistra, W. H. C. F., and Zingone, A. (2014). Assessment of species diversity and distribution of an ancient diatom lineage using a DNA metabarcoding approach. *PLoS One* 9:e103810. doi: 10.1371/journal.pone.0103810

PubMed Abstract | CrossRef Full Text | Google Scholar

Niu, Y. F., Yang, Z. K., Zhang, M. H., Zhu, C. C., Yang, W. D., Liu, J. S., et al. (2012). Transformation of diatom Phaeodactylum tricornutum by electroporation and establishment of inducible selection marker. *Biotechniques* 52, 1–3.

Google Scholar

Nowak, A. P., Sprynskyy, M., Brzozowska, W., and Lisowska-Oleksiak, A. (2019). Electrochemical behavior of a composite material containing 3D-structured diatom biosilica. *Algal Res.* 41:101538. doi: 10.1016/j.algal.2019.101538

CrossRef Full Text | Google Scholar

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951.

Google Scholar

Nymark, M., Valle, K. C., Hancke, K., Winge, P., Andresen, K., Johnsen, G., et al. (2013). Molecular and photosynthetic responses to prolonged darkness and subsequent acclimation to re-illumination in the diatom *Phaeodactylum tricornutum*. *PLoS One*8:e58722. doi: 10.1371/journal.pone.0058722

PubMed Abstract | CrossRef Full Text | Google Scholar

Poulsen, N., and Kröger, N. (2005). A new molecular tool for transgenic diatoms. *FEBS J.* 272, 3413–3423. doi: 10.1111/j.1742-4658.2005.04760.x

PubMed Abstract | CrossRef Full Text | Google Scholar

Poulsen, N., Chesley, P. M., and Kröger, N. (2006). Molecular genetic manipulation of the diatom thalassiosira *Pseudonana* (*Bacillariophyceae*)1. *J. Phycol.* 42, 1059–1065. doi: 10.1111/j.1529-8817.2006.00269.x

CrossRef Full Text | Google Scholar

Pulz, O., and Gross, W. (2004). Valuable products from biotechnology of microalgae. *Appl. Microbiol. Biotechnol.* 65, 635–648. doi: 10.1007/s00253-004-1647-x

PubMed Abstract | CrossRef Full Text | Google Scholar

Qin, T., Gutu, T., Jiao, J., Chang, C.-H., and Rorrer, G. L. (2008). Biological fabrication of photoluminescent nanocomb structures by metabolic incorporation of germanium into the biosilica of the diatom *Nitzschia frustulum*. *ACS Nano* 2, 1296–1304. doi: 10.1021/nn800114q

PubMed Abstract | CrossRef Full Text | Google Scholar

Radakovits, R., Eduafo, P. M., and Posewitz, M. C. (2011). Genetic engineering of fatty acid chain length in *Phaeodactylum tricornutum*. *Metab. Eng.* 13, 89–95. doi: 10.1016/j.ymben.2010.10.003

PubMed Abstract | CrossRef Full Text | Google Scholar

Radakovits, R., Jinkerson, R. E., Darzins, A., and Posewitz, M. C. (2010). Genetic engineering of algae for enhanced biofuel production. *Eukaryot. Cell* 9, 486–501. doi: 10.1128/ec.00364-09

PubMed Abstract | CrossRef Full Text | Google Scholar

Ragni, R., Cicco, S., Vona, D., Leone, G., and Farinola, G. M. (2017). Biosilica from diatoms microalgae: smart materials from bio-medicine to photonics. *J. Mater. Res.*32, 279–291. doi: 10.1557/jmr.2016.459

CrossRef Full Text | Google Scholar

Rastogi, A., Maheswari, U., Dorrell, R. G., Vieira, F. R. J., Maumus, F., Kustka, A., et al. (2018). Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci. Rep.* 8:4834.

Google Scholar

Rea, I., Martucci, N. M., De Stefano, L., Ruggiero, I., Terracciano, M., Dardano, P., et al. (2014). Diatomite biosilica nanocarriers for siRNA transport inside cancer cells. *Biochim. Biophys. Acta* 1840, 3393–3403. doi: 10.1016/j.bbagen.2014.09.009

PubMed Abstract | CrossRef Full Text | Google Scholar

Renaud, S. M., Thinh, L.-V., Lambrinidis, G., and Parry, D. L. (2002). Effect of temperature on growth, chemical composition and fatty acid composition of tropical Australian microalgae grown in batch cultures. *Aquaculture* 211, 195–214. doi: 10.1016/s0044-8486(01)00875-4

CrossRef Full Text | Google Scholar

Rodrigue, J.-P., and Notteboom, T. (2013). The geography of cruises: itineraries, not destinations. *Appl. Geography* 38, 31–42. doi: 10.1016/j.apgeog.2012.11.011

CrossRef Full Text | Google Scholar

Rorrer, G., Jeffryes, C., Chang, C.-H., Lee, D.-H., Gutu, T., Jiao, J., et al. (2007). *Biological Fabrication of Nanostructured Silicon-germanium Photonic Crystals Possessing Unique Photoluminescent and Electroluminescent Properties.*Bellingham, DC: SPIE.

Google Scholar

Round, F. E., Crawford, R. M., Mann, D. G., and Press, C. U. (1990). *Diatoms: Biology and Morphology of the Genera.* Cambridge: Cambridge University Press.

Google Scholar

Sabatino, V., Russo, M. T., Patil, S., d'Ippolito, G., Fontana, A., and Ferrante, M. I. (2015). Establishment of genetic transformation in the sexually reproducing diatoms pseudo-nitzschia multistriata and pseudo-nitzschia arenysensis and inheritance of the transgene. *Mar. Biotechnol. (NY)* 17, 452–462. doi: 10.1007/s10126-015-9633-0

PubMed Abstract | CrossRef Full Text | Google Scholar

Sakaguchi, T., Nakajima, K., and Matsuda, Y. (2011). Identification of the UMP synthase gene by establishment of uracil auxotrophic mutants and the phenotypic complementation system in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiol.* 156, 78–89. doi: 10.1104/pp.110.169631

PubMed Abstract | CrossRef Full Text | Google Scholar

Sakaue, K., Harada, H., and Matsuda, Y. (2008). Development of gene expression system in a marine diatom using viral promoters of a wide variety of origin. *Physiol. Plant.* 133, 59–67. doi: 10.1111/j.1399-3054.2008.01089.x

PubMed Abstract | CrossRef Full Text | Google Scholar

Schellenberger Costa, B., Jungandreas, A., Jakob, T., Weisheit, W., Mittag, M., and Wilhelm, C. (2012). Blue light is essential for high light acclimation and photoprotection in the diatom *Phaeodactylum tricornutum*. *J. Exp. Bot.* 64, 483–493. doi: 10.1093/jxb/ers340

PubMed Abstract | CrossRef Full Text | Google Scholar

Selvaraj, V., Muthukumar, A., Nagamony, P., and Chinnuswamy, V. (2018). Detection of typhoid fever by diatom-based optical biosensor. *Environ. Sci. Poll. Res.* 25, 20385–20390. doi: 10.1007/s11356-017-9362-1

PubMed Abstract | CrossRef Full Text | Google Scholar

Senthilvelan, T., Kanagaraj, J., Panda, R. C., and Mandal, A. B. (2014). Biodegradation of phenol by mixed microbial culture: an eco-friendly approach for the pollution reduction. *Clean Technol. Environ. Pol.* 16, 113–126. doi: 10.1007/s10098-013-0598-2

CrossRef Full Text | Google Scholar

Seo, S., Jeon, H., Hwang, S., Jin, E., and Chang, K. S. (2015). Development of a new constitutive expression system for the transformation of the diatom *Phaeodactylum tricornutum*. *Algal Res.* 11, 50–54. doi: 10.1016/j.algal.2015.05.012

CrossRef Full Text | Google Scholar

Sharma, N., Fleurent, G., Awwad, F., Cheng, M., Meddeb-Mouelhi, F., Budge, S., et al. (2020). Red light shift an effective alternative to boost biomass and lipids in *Phaeodactylum tricornutum*. *Appl. Sci.* 10:2531. doi: 10.3390/app10072531

CrossRef Full Text | Google Scholar

Sheehan, J., Dunahay, T., Benemann, J., and Roessler, P. (1998). *Look Back at the U.S. Department of Energy's Aquatic Species Program: Biodiesel from Algae*. Golden, CO: National Renewable Energy Laboratory.

Google Scholar

Shrestha, R. P., and Hildebrand, M. (2015). Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. *Eukaryot. Cell* 14, 29–40. doi: 10.1128/ec.00209-14

PubMed Abstract | CrossRef Full Text | Google Scholar

Shrestha, R. P., and Hildebrand, M. (2017). Development of a silicon limitation inducible expression system for recombinant protein production in the centric diatoms *Thalassiosira pseudonana* and *Cyclotella cryptica*. *Microbial Cell Factories*16:145.

Google Scholar

Singh, A. K., Sharma, N., Farooqi, H., Abdin, M. Z., Mock, T., and Kumar, S. (2017). Phycoremediation of municipal wastewater by microalgae to produce biofuel. *Int. J. Phytoremediation* 19, 805–812. doi: 10.1080/15226514.2017.1284758

PubMed Abstract | CrossRef Full Text | Google Scholar

Singh, R., Khan, M. J., Rane, J., Gajbhiye, A., Vinayak, V., and Joshi, K. B. (2020). Biofabrication of diatom surface by tyrosine-metal complexes:smart microcontainers to inhibit bacterial growth. *ChemistrySelect* 5, 3091–3097. doi: 10.1002/slct.201904248

CrossRef Full Text | Google Scholar

Slattery, S. S., Diamond, A., Wang, H., Therrien, J. A., Lant, J. T., Jazey, T., et al. (2018). An expanded plasmid-based genetic toolbox enables cas9 genome editing and stable maintenance of synthetic pathways in *Phaeodactylum tricornutum*. *ACS Synthetic Biol.* 7, 328–338. doi: 10.1021/acssynbio.7b00191

PubMed Abstract | CrossRef Full Text | Google Scholar

Slattery, S. S., Wang, H., Giguere, D. J., Kocsis, C., Urquhart, B. L., Karas, B. J., et al. (2020). Plasmid-based complementation of large deletions in *Phaeodactylum tricornutum* biosynthetic genes generated by Cas9 editing. *Sci. Rep.* 10:13879.

Google Scholar

Slowing, I. I., Vivero-Escoto, J. L., Wu, C.-W., and Lin, V. S. Y. (2008). Mesoporous silica nanoparticles as controlled release drug delivery and gene transfection carriers. *Adv. Drug Deliv. Rev.* 60, 1278–1288. doi: 10.1016/j.addr.2008.03.012

PubMed Abstract | CrossRef Full Text | Google Scholar

Smetacek, V. (1999). Diatoms and the ocean carbon cycle. *Protist* 150, 25–32. doi: 10.1016/s1434-4610(99)70006-4

CrossRef Full Text | Google Scholar

Smol, J. P., and Stoermer, E. F. (2010). *The Diatoms: Applications for the Environmental and Earth Sciences.* Cambridge: Cambridge Press.

Google Scholar

Soni, A. K., Kumar, S., and Pandey, M. (2020). Performance comparison of microalgae biodiesel blends with petro–diesel on variable compression ratio engine. *J. Instit. Eng. (India): Series E.* https://doi.org/10.1007/s40034-020-00183-0

Google Scholar

Suleiman, A. K. A., Lourenço, K. S., Clark, C., Luz, R. L., da Silva, G. H. R., Vet, L., et al. (2020). From toilet to agriculture: fertilization with microalgal biomass from wastewater impacts the soil and rhizosphere active microbiomes, greenhouse gas emissions and plant growth. *Resources Conserv. Recycl.* 161:104924. doi: 10.1016/j.resconrec.2020.104924

CrossRef Full Text | Google Scholar

Taberlet, P., Coissac, E., Hajibabaei, M., and Rieseberg, L. H. Environmental DNA. *Mol. Ecol.* 21, 1789–1793.

Tanaka, A., Ohno, N., Nakajima, K., and Matsuda, Y. (2016). Light and CO 2 /cAMP signal cross talk on the promoter elements of chloroplastic β -carbonic anhydrase genes in the marine diatom *Phaeodactylum tricornutum. Plant Physiol.* 170, 1105–1116. doi: 10.1104/pp.15.01738

PubMed Abstract | CrossRef Full Text | Google Scholar

Taparia, Y., Zarka, A., Leu, S., Zarivach, R., Boussiba, S., and Khozin-Goldberg, I. (2019). A novel endogenous selection marker for the diatom *Phaeodactylum tricornutum* based on a unique mutation in phytoene desaturase 1. *Sci. Rep.* 9:8217.

Google Scholar

Terracciano, M., De Stefano, L., and Rea, I. (2018). Diatoms green nanotechnology for biosilica-based drug delivery systems. *Pharmaceutics* 10:242. doi: 10.3390/pharmaceutics10040242

PubMed Abstract | CrossRef Full Text | Google Scholar

Thomas, J., Thomas, C. J., Radcliffe, J., and Itsiopoulos, C. (2015). Omega-3 fatty acids in early prevention of inflammatory neurodegenerative disease: a focus on Alzheimer's Disease. *BioMed. Res. Int.* 2015:172801.

Google Scholar

Tibocha-Bonilla, J. D., Zuñiga, C., Godoy-Silva, R. D., and Zengler, K. (2018). Advances in metabolic modeling of oleaginous microalgae. *Biotechnol. Biofuels*11:241.

Google Scholar

Tocher, D. R., Betancor, M. B., Sprague, M., Olsen, R. E., and Napier, J. A. (2019). Omega-3 long-chain polyunsaturated fatty acids, EPA and DHA: bridging the gap between supply and demand. *Nutrients* 11:89. doi: 10.3390/nu11010089

PubMed Abstract | CrossRef Full Text | Google Scholar

Tossavainen, M., Lahti, K., Edelmann, M., Eskola, R., Lampi, A.-M., Piironen, V., et al. (2019). Integrated utilization of microalgae cultured in aquaculture wastewater: wastewater treatment and production of valuable fatty acids and tocopherols. *J. Appl. Phycol.* 31, 1753–1763. doi: 10.1007/s10811-018-1689-6

CrossRef Full Text | Google Scholar

Tréguer, P., Nelson, D. M., Van Bennekom, A. J., DeMaster, D. J., Leynaert, A., and Quéguiner, B. (1995). The silica balance in the world ocean: a reestimate. *Science*268, 375–379. doi: 10.1126/science.268.5209.375

PubMed Abstract | CrossRef Full Text | Google Scholar

Trentacoste, E. M., Shrestha, R. P., Smith, S. R., Glé, C., Hartmann, A. C., Hildebrand, M., et al. (2013). Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *Proc. Natl. Acad. Sci. U S A.* 110, 19748–19753. doi: 10.1073/pnas.1309299110

PubMed Abstract | CrossRef Full Text | Google Scholar

Vasani, R. B., Losic, D., Cavallaro, A., and Voelcker, N. H. (2015). Fabrication of stimulus-responsive diatom biosilica microcapsules for antibiotic drug delivery. *J. Mater. Chem. B* 3, 4325–4329. doi: 10.1039/c5tb00648a

PubMed Abstract | CrossRef Full Text | Google Scholar

Vasselon, V., Rimet, F., Tapolczai, K., and Bouchez, A. (2017). Assessing ecological status with diatoms DNA metabarcoding: scaling-up on a WFD monitoring network (Mayotte island, France). *Ecol. Indicat.* 82, 1–12. doi: 10.1016/j.ecolind.2017.06.024

CrossRef Full Text | Google Scholar

Wah, N. B., Ahmad, A. L. B., Chieh, D. C. J., and Hwai, A. T. S. (2015). Changes in lipid profiles of a tropical benthic diatom in different cultivation temperature. *Asian J. Appl. Sci. Eng.* 4, 91–101.

Google Scholar

Walker, W. W. Jr. (1983). Significance of eutrophication in water supply reservoirs. *J. AWWA* 75, 38–42. doi: 10.1002/j.1551-8833.1983.tb05056.x

CrossRef Full Text | Google Scholar

Wang, H., and Daggy, B. P. (2017). The role of fish oil in inflammatory eye diseases. *Biomed. Hub.* 2, 1–12. doi: 10.1159/000455818

PubMed Abstract | CrossRef Full Text | Google Scholar

Wang, J.-K., and Seibert, M. (2017). Prospects for commercial production of diatoms. *Biotechnol. Biofuels* 10:16.

Google Scholar

Wang, X. C., and Zhao, H. M. (2007). Uptake and biodegradation of polycyclic aromatic hydrocarbons by marine seaweed. *J. Coastal Res.* SI50, 1056–1061.

Google Scholar

Wen, Z. Y., and Chen, F. (2001a). A perfusion–cell bleeding culture strategy for enhancing the productivity of eicosapentaenoic acid by *Nitzschia laevis*. *Appl. Microbiol. Biotechnol.* 57, 316–322. doi: 10.1007/s002530100786

PubMed Abstract | CrossRef Full Text | Google Scholar

Wen, Z.-Y., and Chen, F. (2001b). Application of statistically-based experimental designs for the optimization of eicosapentaenoic acid production by the diatom *Nitzschia laevis*. *Biotechnol. Bioeng.* 75, 159–169. doi: 10.1002/bit.1175

PubMed Abstract | CrossRef Full Text | Google Scholar

Weyman, P. D., Beeri, K., Lefebvre, S. C., Rivera, J., McCarthy, J. K., Heuberger, A. L., et al. (2015). Inactivation of *Phaeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol. J.* 13, 460–470. doi: 10.1111/pbi.12254

PubMed Abstract | CrossRef Full Text | Google Scholar

Wu, S., Gu, W., Huang, A., Li, Y., Kumar, M., Lim, P. E., et al. (2019). Elevated CO2 improves both lipid accumulation and growth rate in the glucose-6-phosphate dehydrogenase engineered *Phaeodactylum tricornutum*. *Microbial Cell Factories*18:161.

Google Scholar

Xia, S., Wang, K., Wan, L., Li, A., Hu, Q., and Zhang, C. (2013). Production, characterization, and antioxidant activity of fucoxanthin from the marine diatom *Odontella aurita*. *Mar. Drugs* 11, 2667–2681. doi: 10.3390/md11072667

PubMed Abstract | CrossRef Full Text | Google Scholar

Xie, W. H., Zhu, C. C., Zhang, N. S., Li, D. W., Yang, W. D., Liu, J. S., et al. (2014). Construction of novel chloroplast expression vector and development of an efficient transformation system for the diatom *Phaeodactylum tricornutum*. *Mar. Biotechnol. (NY)* 16, 538–546. doi: 10.1007/s10126-014-9570-3

PubMed Abstract | CrossRef Full Text | Google Scholar

Yamazaki, T., Sasanuma, H., Mayama, S., and Umemura, K. (2010). Cultivation of melosira nummuloides cells in the presence of platinum: preparation of metal-containing frustules. *Phys. Status Solidi c* 7, 2759–2762. doi: 10.1002/pssc.200983808

CrossRef Full Text | Google Scholar

Yoshinaga, R., Niwa-Kubota, M., Matsui, H., and Matsuda, Y. (2014). Characterization of iron-responsive promoters in the marine diatom *Phaeodactylum tricornutum*. *Mar. Genom.* 16, 55–62. doi: 10.1016/j.margen.2014.01.005

PubMed Abstract | CrossRef Full Text | Google Scholar

Zaboikin, M., Zaboikina, T., Freter, C., and Srinivasakumar, N. (2017). Non-Homologous end joining and homology directed DNA repair frequency of double-stranded breaks introduced by genome editing reagents. *PLoS One* 12:e0169931. doi: 10.1371/journal.pone.0169931

PubMed Abstract | CrossRef Full Text | Google Scholar

Zamora, P., Narváez, A., and Domínguez, E. (2009). Enzyme-modified nanoparticles using biomimetically synthesized silica. *Bioelectrochemistry* 76, 100–106. doi: 10.1016/j.bioelechem.2009.05.006

PubMed Abstract | CrossRef Full Text | Google Scholar

Zaslavskaia, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R., and Apt, K. E. (2000). Transformation of the diatom *Phaeodactylum tricornutum* (*Bacillariophyceae*) with a variety of selectable marker and reporter genes. *J. Phycol.* 36, 379–386. doi: 10.1046/j.1529-8817.2000.99164.x

CrossRef Full Text | Google Scholar

Zhang, C., and Hu, H. (2014). High-efficiency nuclear transformation of the diatom *Phaeodactylum tricornutum* by electroporation. *Mar. Genomics* 16, 63–66. doi: 10.1016/j.margen.2013.10.003

PubMed Abstract | CrossRef Full Text | Google Scholar

Zhou, Y., Cao, Y., Huang, J., Deng, K., Ma, K., Zhang, T., et al. (2020). Research advances in forensic diatom testing. *Forensic Sci. Res.* 5, 98–105. doi: 10.1080/20961790.2020.1718901

PubMed Abstract | CrossRef Full Text | Google Scholar

Zimmermann, J., Jahn, R., and Gemeinholzer, B. (2011). Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Organ. Divers. Evol.* 11:173. doi: 10.1007/s13127-011-0050-6

CrossRef Full Text | Google Scholar

Zulu, N. N., Zienkiewicz, K., Vollheyde, K., and Feussner, I. (2018). Current trends to comprehend lipid metabolism in diatoms. *Progress Lipid Res.* 70, 1–16. doi: 10.1016/j.plipres.2018.03.001

PubMed Abstract | CrossRef Full Text | Google Scholar

9 References

1. Zaynab M, Fatima M, Sharif Y, Zafar MH, Ali H, Khan KA: **Role of primary metabolites in plant defense against pathogens**. *Microb Pathog* 2019, **137**:103728.
2. Waadt R, Seller CA, Hsu PK, Takahashi Y, Munemasa S, Schroeder JI: **Plant hormone regulation of abiotic stress responses**. *Nat Rev Mol Cell Biol* 2022, **23**(10):680-694.
3. Erb M, Kliebenstein DJ: **Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy**. *Plant Physiol* 2020, **184**(1):39-52.
4. Bansal SPaMBaNTaSPaYK: **Secondary Metabolites of Plants and their Role: Overview**. *Current Trends in Biotechnology and Pharmacy* 2015, **9**:293-304.
5. Huchelmann A, Boutry M, Hachez C: **Plant Glandular Trichomes: Natural Cell Factories of High Biotechnological Interest**. *Plant Physiol* 2017, **175**(1):6-22.
6. Wiles D, Shanbhag BK, O'Brien M, Doblin MS, Bacic A, Beddoe T: **Heterologous production of Cannabis sativa-derived specialised metabolites of medicinal significance- Insights into engineering strategies**. *Phytochemistry* 2022, **203**:113380.
7. Crocq MA: **History of cannabis and the endocannabinoid system** . *Dialogues Clin Neurosci* 2020, **22**(3):223-228.
8. Preteroti M, Wilson ET, Eidelman DH, Baglole CJ: **Modulation of pulmonary immune function by inhaled cannabis products and consequences for lung disease**. *Respir Res* 2023, **24**(1):95.
9. Mechoulam YGaR: **Isolation, Structure, and Partial Synthesis of an Active Constituent of Hashish**
. *Journal of the American Chemical Society* 1964, **86**:1646-1647.
10. Zirpel B, Kayser O, Stehle F: **Elucidation of structure-function relationship of THCA and CBDA synthase from Cannabis sativaL**. *J Biotechnol* 2018, **284**:17-26.
11. Lin YF: **Potassium channels as molecular targets of endocannabinoids**. *Channels (Austin)* 2021, **15**(1):408-423.
12. Mills B, Yepes A, Nugent K: **Synthetic Cannabinoids**. *Am J Med Sci* 2015, **350**(1):59-62.
13. Charytoniuk T, Zywno H, Berk K, Bzdega W, Kolakowski A, Chabowski A, Konstantynowicz-Nowicka K: **The Endocannabinoid System and Physical Activity-A Robust Duo in the Novel Therapeutic Approach against Metabolic Disorders**. *Int J Mol Sci* 2022, **23**(6).
14. Zhang Q, Chen X, Guo H, Trindade LM, Salentijn EMJ, Guo R, Guo M, Xu Y, Yang M: **Latitudinal Adaptation and Genetic Insights Into the Origins of Cannabis sativa L.** *Front Plant Sci* 2018, **9**:1876.
15. Small E, Pocock T, Cavers PB: **The biology of Canadian weeds. 119. Cannabis sativa L.** *Canadian Journal of Plant Science* 2003, **83**(1):217-237.
16. McPartland JM, Small E: **A classification of endangered high-THC cannabis (Cannabis sativa subsp. indica) domesticates and their wild relatives**. *PhytoKeys* 2020, **144**.
17. Filella I, Primante C, Llusià J, Martín González AM, Seco R, Farré-Armengol G, Rodrigo A, Bosch J, Peñuelas J: **Floral advertisement scent in a changing plant-pollinators market**. *Scientific Reports* 2013, **3**(1):3434.
18. Kessler A, Baldwin IT: **Defensive Function of Herbivore-Induced Plant Volatile Emissions in Nature**. *Science* 2001, **291**(5511):2141-2144.

19. Taura F, Tanaka S, Taguchi C, Fukamizu T, Tanaka H, Shoyama Y, Morimoto S: Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. *FEBS Letters* 2009, **583**(12):2061-2066.
20. Gagne SJ, Stout JM, Liu E, Boubakir Z, Clark SM, Page JE: Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides. *Proceedings of the National Academy of Sciences* 2012, **109**(31):12811-12816.
21. Fellermeier M, Zenk MH: Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. *FEBS Letters* 1998, **427**(2):283-285.
22. Taura F, Morimoto S, Shoyama Y: Purification and Characterization of Cannabidiolic-acid Synthase from Cannabis sativa L.: BIOCHEMICAL ANALYSIS OF A NOVEL ENZYME THAT CATALYZES THE OXIDOCYCLIZATION OF CANNABIGEROLIC ACID TO CANNABIDIOLIC ACID *. *Journal of Biological Chemistry* 1996, **271**(29):17411-17416.
23. Taura F, Dono E, Sirikantaramas S, Yoshimura K, Shoyama Y, Morimoto S: Production of Δ1-tetrahydrocannabinolic acid by the biosynthetic enzyme secreted from transgenic Pichia pastoris. *Biochemical and Biophysical Research Communications* 2007, **361**(3):675-680.
24. Taura F, Sirikantaramas S, Shoyama Y, Yoshikai K, Shoyama Y, Morimoto S: Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa. *FEBS Letters* 2007, **581**(16):2929-2934.
25. Onofri C, de Meijer EPM, Mandolino G: Sequence heterogeneity of cannabidiolic- and tetrahydrocannabinolic acid-synthase in Cannabis sativa L. and its relationship with chemical phenotype. *Phytochemistry* 2015, **116**:57-68.
26. Romero P, Peris A, Vergara K, Matus JT: Comprehending and improving cannabis specialized metabolism in the systems biology era. *Plant Science* 2020, **298**:110571.
27. Morimoto S, Tanaka Y, Sasaki K, Tanaka H, Fukamizu T, Shoyama Y, Shoyama Y, Taura F: Identification and Characterization of Cannabinoids That Induce Cell Death through Mitochondrial Permeability Transition in Cannabis Leaf Cells *. *Journal of Biological Chemistry* 2007, **282**(28):20739-20751.
28. Rodziewicz P, Loroch S, Marczak Ł, Sickmann A, Kayser O: Cannabinoid synthases and osmoprotective metabolites accumulate in the exudates of Cannabis sativa L. glandular trichomes. *Plant Science* 2019, **284**:108-116.
29. Tahir MN, Shahbazi F, Rondeau-Gagné S, Trant JF: The biosynthesis of the cannabinoids. *Journal of Cannabis Research* 2021, **3**(1):7.
30. Kearsey LJ, Prandi N, Karuppiah V, Yan C, Leys D, Toogood H, Takano E, Scrutton NS: Structure of the Cannabis sativa olivetol-producing enzyme reveals cyclization plasticity in type III polyketide synthases. *The FEBS Journal* 2020, **287**(8):1511-1524.
31. Bhunia S, Kolishetti N, Arias AY, Vashist A, Nair M: Cannabidiol for neurodegenerative disorders: A comprehensive review. *Frontiers in Pharmacology* 2022, **13**.
32. Lowin T, Tingting R, Zurmahr J, Classen T, Schneider M, Pongratz G: Cannabidiol (CBD): a killer for inflammatory rheumatoid arthritis synovial fibroblasts. *Cell Death & Disease* 2020, **11**(8):714.
33. Naya NM, Kelly J, Hogwood A, Abbate A, Toldo S: Therapeutic potential of cannabidiol (CBD) in the treatment of cardiovascular diseases. *Expert Opinion on Investigational Drugs* 2024, **33**(7):699-712.

34. Mattes RG, Espinosa ML, Oh SS, Anatrella EM, Urteaga EM: **Cannabidiol (CBD) Use in Type 2 Diabetes: A Case Report**. *Diabetes Spectrum* 2021, **34**(2):198-201.
35. Martinez Naya N, Kelly J, Corna G, Golino M, Polizio AH, Abbate A, Toldo S, Mezzaroma E: **An Overview of Cannabidiol as a Multifunctional Drug: Pharmacokinetics and Cellular Effects**. *Molecules* 2024, **29**(2).
36. Tham M, Yilmaz O, Alaverdashvili M, Kelly MEM, Denovan-Wright EM, Laprairie RB: **Allosteric and orthosteric pharmacology of cannabidiol and cannabidiol-dimethylheptyl at the type 1 and type 2 cannabinoid receptors**. *British Journal of Pharmacology* 2019, **176**(10):1455-1469.
37. Crippa JA, Guimarães FS, Campos AC, Zuardi AW: **Translational Investigation of the Therapeutic Potential of Cannabidiol (CBD): Toward a New Age**. *Frontiers in Immunology* 2018, **9**.
38. Mlost J, Bryk M, Starowicz K: **Cannabidiol for Pain Treatment: Focus on Pharmacology and Mechanism of Action**. *Int J Mol Sci* 2020, **21**(22).
39. Poore J, Nemecek T: **Reducing food's environmental impacts through producers and consumers**. *Science* 2018, **360**(6392):987-992.
40. Smetacek V: **Diatoms and the Ocean Carbon Cycle**. *Protist* 1999, **150**(1):25-32.
41. Mann DG, Vanormelingen P: **An Inordinate Fondness? The Number, Distributions, and Origins of Diatom Species**. *Journal of Eukaryotic Microbiology* 2013, **60**(4):414-420.
42. Thamatrakoln K, Hildebrand M: **Silicon Uptake in Diatoms Revisited: A Model for Saturable and Nonsaturable Uptake Kinetics and the Role of Silicon Transporters**. *Plant Physiology* 2007, **146**(3):1397-1407.
43. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M *et al*: **The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism**. *Science* 2004, **306**(5693):79-86.
44. Mao Y, Guo L, Luo Y, Tang Z, Li W, Dong W: **Sexual reproduction potential implied by functional analysis of SPO11 in Phaeodactylum tricornutum**. *Gene* 2020, **757**:144929.
45. Von Dassow P, Chepurnov VA, Armbrust EV: **RELATIONSHIPS BETWEEN GROWTH RATE, CELL SIZE, AND INDUCTION OF SPERMATOGENESIS IN THE CENTRIC DIATOM THALASSIOSIRA WEISSFLOGII (BACILLARIOPHYTA)**. *Journal of Phycology* 2006, **42**(4):887-899.
46. Sharma N, Simon DP, Diaz-Garza AM, Fantino E, Messaabi A, Meddeb-Mouelhi F, Germain H, Desgagné-Penix I: **Diatoms Biotechnology: Various Industrial Applications for a Greener Tomorrow**. *Frontiers in Marine Science* 2021, **8**.
47. De Martino A, Bartual A, Willis A, Meichenin A, Villazán B, Maheswari U, Bowler C: **Physiological and Molecular Evidence that Environmental Changes Elicit Morphological Interconversion in the Model Diatom Phaeodactylum tricornutum**. *Protist* 2011, **162**(3):462-481.
48. Apt KE, Grossman AR, Kroth-Pancic PG: **Stable nuclear transformation of the diatomPhaeodactylum tricornutum**. *Molecular and General Genetics MGG* 1996, **252**(5):572-579.
49. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP *et al*: **The Phaeodactylum genome reveals the evolutionary history of diatom genomes**. *Nature* 2008, **456**(7219):239-244.

50. Spilling K, Brynjólfsdóttir Á, Enss D, Rischer H, Svavarsson HG: **The effect of high pH on structural lipids in diatoms**. *Journal of Applied Phycology* 2013, **25**(5):1435-1439.
51. Remmers IM, Martens DE, Wijffels RH, Lamers PP: **Dynamics of triacylglycerol and EPA production in Phaeodactylum tricornutum under nitrogen starvation at different light intensities**. *PLoS One* 2017, **12**(4):e0175630.
52. Russo MT, Rogato A, Jaubert M, Karas BJ, Falciatore A: **Phaeodactylum tricornutum: An established model species for diatom molecular research and an emerging chassis for algal synthetic biology**. *Journal of Phycology* 2023, **59**(6):1114-1122.
53. Butler T, Kapoore RV, Vaidyanathan S: **Phaeodactylum tricornutum: A Diatom Cell Factory**. *Trends in Biotechnology* 2020, **38**(6):606-622.
54. Tredici MR, Bassi N, Prussi M, Biondi N, Rodolfi L, Chini Zittelli G, Sampietro G: **Energy balance of algal biomass production in a 1-ha "Green Wall Panel" plant: How to produce algal biomass in a closed reactor achieving a high Net Energy Ratio**. *Applied Energy* 2015, **154**:1103-1111.
55. Meiser A, Schmid-Staiger U, Trösch W: **Optimization of eicosapentaenoic acid production byPhaeodactylum tricornutumin the flat panel airlift (FPA) reactor**. *Journal of Applied Phycology* 2004, **16**(3):215-225.
56. Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips APR, Noddings CM, Brunson JK, Valas RE, Deerinck TJ, Jablanovic J *et al*: **Designer diatom episomes delivered by bacterial conjugation**. *Nature Communications* 2015, **6**(1):6925.
57. Diner RE, Noddings CM, Lian NC, Kang AK, McQuaid JB, Jablanovic J, Espinoza JL, Nguyen NA, Anzelmatti MA, Jansson J *et al*: **Diatom centromeres suggest a mechanism for nuclear DNA acquisition**. *Proceedings of the National Academy of Sciences* 2017, **114**(29):E6015-E6024.
58. Ahmed SF, Mofijur M, Nahrin M, Chowdhury SN, Nuzhat S, Alherek M, Rafa N, Ong HC, Nghiem LD, Mahlia TMI: **Biohydrogen production from wastewater-based microalgae: Progresses and challenges**. *International Journal of Hydrogen Energy* 2022, **47**(88):37321-37342.
59. Sirikantaramas S, Taura F, Tanaka Y, Ishikawa Y, Morimoto S, Shoyama Y: **Tetrahydrocannabinolic Acid Synthase, the Enzyme Controlling Marijuana Psychoactivity, is Secreted into the Storage Cavity of the Glandular Trichomes**. *Plant and Cell Physiology* 2005, **46**(9):1578-1582.
60. Zirpel B, Stehle F, Kayser O: **Production of Δ9-tetrahydrocannabinolic acid from cannabigerolic acid by whole cells of Pichia (Komagataella) pastoris expressing Δ9-tetrahydrocannabinolic acid synthase from Cannabis sativa l**. *Biotechnology Letters* 2015, **37**(9):1869-1875.
61. Sirikantaramas S, Morimoto S, Shoyama Y, Ishikawa Y, Wada Y, Shoyama Y, Taura F: **The Gene Controlling Marijuana Psychoactivity: MOLECULAR CLONING AND HETEROLOGOUS EXPRESSION OF Δ¹-TETRAHYDROCANNABINOLIC ACID SYNTHASE FROM CANNABIS SATIVA L. ***. *Journal of Biological Chemistry* 2004, **279**(38):39767-39774.
62. Erdene-Ochir E, Shin B-K, Kwon B, Jung C, Pan C-H: **Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from Phaeodactylum tricornutum**. *Scientific Reports* 2019, **9**(1):9941.
63. Castillo-Arellano J, Canseco-Alba A, Cutler SJ, Leon F: **The Polypharmacological Effects of Cannabidiol**. *Molecules* 2023, **28**(7).

64. Stella N: **THC and CBD: Similarities and differences between siblings**. *Neuron* 2023, **111**(3):302-327.
65. Wang B, Kovalchuk A, Li D, Rodriguez-Juarez R, Ilnytskyy Y, Kovalchuk I, Kovalchuk O, Kovalchuk A, Kovalchuk I, Kovalchuk O *et al*: **In search of preventive strategies: novel high-CBD Cannabis sativa extracts modulate ACE2 expression in COVID-19 gateway tissues**. In: *Aging.* vol. 12; 2020: 22425-22440.
66. Walsh KB, McKinney AE, Holmes AE: **Minor Cannabinoids: Biosynthesis, Molecular Pharmacology and Potential Therapeutic Uses**. In: *Frontiers in Pharmacology.* vol. 12; 2021: 1-18.
67. Quinones R, Moreno S, Smythers AL, Sullins C, Pijor H, Brown G, Trouten A, Richards-Waugh LL, Siddig A: **Quantification of Cannabis in Infused Consumer Products and Their Residues on Skin**. *ACS Pharmacol Transl Sci* 2022, **5**(8):642-651.
68. Cristino L, Bisogno T, Di Marzo V: **Cannabinoids and the expanded endocannabinoid system in neurological disorders**. *Nat Rev Neurol* 2020, **16**(1):9-29.
69. Devinsky O, Marsh E, Friedman D: **Cannabidiol in patients with treatment-resistant epilepsy- Authors' reply**. *Lancet Neurol* 2016, **15**(6):545-546.
70. Devinsky O, Cross JH, Laux L, Marsh E, Miller I, Nabbout R, Scheffer IE, Thiele EA, Wright S, Cannabidiol in Dravet Syndrome Study G: **Trial of Cannabidiol for Drug-Resistant Seizures in the Dravet Syndrome**. *N Engl J Med* 2017, **376**(21):2011-2020.
71. Desaulniers Brousseau V, Wu BS, MacPherson S, Morello V, Lefsrud M: **Cannabinoids and Terpenes: How Production of Photo-Protectants Can Be Manipulated to Enhance Cannabis sativa L. Phytochemistry**. *Front Plant Sci* 2021, **12**:620021.
72. Taura F, Morimoto S, Shoyama Y, Acid C, Cannabidiolic TO, Taura F, Morimoto S, Shoyama Y: **Purification and characterization of cannabidiolic-acid synthase from Cannabis sativa L. Biochemical analysis of a novel enzyme that catalyzes the oxidocyclization of cannabigerolic acid to cannabidiolic acid**. In: *Journal of Biological Chemistry.* vol. 271: Â© 1996 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology.; 1996: 17411-17416.
73. Sirikantaramas S, Morimoto S, Shoyama Y, Ishikawa Y, Wada Y, Shoyama Y, Taura F: **The gene controlling marijuana psychoactivity. Molecular cloning and heterologous expression of Δ1-tetrahydrocannabinolic acid synthase from Cannabis sativa L.** In: *Journal of Biological Chemistry.* vol. 279: Â© 2004 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology.; 2004: 39767-39774.
74. Lim KJH, Lim YP, Hartono YD, Go MK, Fan H, Yew WS: **Biosynthesis of nature-inspired unnatural cannabinoids**. In: *Molecules.* vol. 26; 2021.
75. Andre CM, Hausman J-f, Guerriero G: **Cannabis sativa : The Plant of the Thousand and One Molecules**. In., vol. 7; 2016: 1-17.
76. Dai LN, T.; Luo, R.; Zhang, L.; Zhang, S.; Kang, Y.; Chi, J.; Feng, X.; Shi, J.; Tian, Y.; Gao, B.; Li, Z.: **Improvement of cannabidiolic acid synthetase activity through molecular docking and site-directed mutagenesis**. *Industrial Crops and Products* 2024.
77. de A. Leite J, de Oliveira MVL, Conti R, de SBW, Rosa TR, Filgueiras PR, Lacerda V, Jr., Romao W, Neto AC: **Extraction and isolation of cannabinoids from marijuana seizures and characterization by (1)H NMR allied to chemometric tools**. *Sci Justice* 2018, **58**(5):355-365.

78. Ohtsuki T, Friesen JB, Chen S-N, McAlpine JB, Pauli GF: **Selective Preparation and High Dynamic-Range Analysis of Cannabinoids in "CBD Oil" and Other Cannabis sativa Preparations.** In: *Journal of natural products.* vol. 85; 2022: 634-646.
79. Chiurchiù ES, S.; Allegrini, P.; Ciceri, D.; Ballini, R.; Palmieri A.: **A Novel and Practical Continuous Flow Chemical Synthesis of Cannabidiol (CBD) and its CBDV and CBDB Analogues**. *European Journal of Organic Chemistry* 2021.
80. Kearsey LJ, Prandi N, Karuppiah V, Yan C, Leys D, Toogood H, Takano E, Scrutton NS: **Structure of the Cannabis sativa olivetol-producing enzyme reveals cyclization plasticity in type III polyketide synthases**. In: *FEBS Journal.* 2019: 1-14.
81. Shoyama Y, Tamada T, Kurihara K, Takeuchi A, Taura F, Arai S, Blaber M, Shoyama Y, Morimoto S, Kuroki R: **Structure and Function of Δ 1-Tetrahydrocannabinolic Acid (THCA) Synthase , the Enzyme Controlling the Psychoactivity of Cannabis sativa**. In: *Journal of Molecular Biology.* vol. 423: Elsevier Ltd; 2012: 96-105.
82. Zirpel B, Stehle F, Kayser O: **Production of Delta9-tetrahydrocannabinolic acid from cannabigerolic acid by whole cells of Pichia (Komagataella) pastoris expressing Delta9-tetrahydrocannabinolic acid synthase from Cannabis sativa L.** In: *Biotechnology letters.* vol. 37; 2015: 1869-1875.
83. Zirpel B, Degenhardt F, Martin C, Kayser O, Stehle F: **Engineering yeasts as platform organisms for cannabinoid biosynthesis**. In: *Journal of Biotechnology.* vol. 259: Elsevier; 2017: 204-212.
84. Zirpel B, Degenhardt F, Zammarelli C, Wibberg D, Kalinowski J, Stehle F, Kayser O: **Optimization of Δ9-tetrahydrocannabinolic acid synthase production in Komagataella phaffii via post-translational bottleneck identification**. In: *Journal of Biotechnology.* vol. 272-273: Elsevier; 2018: 40-47.
85. Luo X, Reiter MA, d'Espaux L, Wong J, Denby CM, Lechner A, Zhang Y, Grzybowski AT, Harth S, Lin W *et al*: **Complete biosynthesis of cannabinoids and their unnatural analogues in yeast**. In: *Nature.* vol. 567: Springer US; 2019: 123-126.
86. Gülck T, Booth JK, Carvalho, Khakimov B, Crocoll C, Motawia MS, Møller BL, Bohlmann J, Gallage NJ: **Synthetic Biology of Cannabinoids and Cannabinoid Glucosides in Nicotiana benthamiana and Saccharomyces cerevisiae**. In: *Journal of Natural Products.* vol. 83; 2020: 2877-2893.
87. Kosalková KB, C.; Sánchez-Orejas, I.C.; Cueto, L.; García-Estrada, C.: **Biotechnological Fungal Platforms for the Production of Biosynthetic Cannabinoids**. *Journal of Fungi* 2023.
88. Awwad F, Fantino EI, Héneault M, Diaz-Garza AM, Merindol N, Custeau A, Gélinas SE, Meddeb-Mouelhi F, Li J, Lemay JF *et al*: **Bioengineering of the Marine Diatom Phaeodactylum tricornutum with Cannabis Genes Enables the Production of the Cannabinoid Precursor, Olivetolic Acid**. *Int J Mol Sci* 2023, **24**(23).
89. Fantino EA, F.; Merindol, N.; Diaz Garza, A.M.; Gelinas, S.E; Gajon Robles, G.C.; Custeau, A.; Meddeb-Mouelhi, F.; Desgagne-Penix, I.: **Bioengineering Phaeodactylum tricornutum, a marine diatom, for cannabinoid biosynthesis**. *Algal Research* 2024.
90. Blatt-janmaat K, Qu Y: **The Biochemistry of Phytocannabinoids and Metabolic Engineering of Their Production in Heterologous Systems**. In.; 2021.
91. Sirikantaramas S, Taura F, Tanaka Y, Ishikawa Y, Morimoto S: **Tetrahydrocannabinolic Acid Synthase , the Enzyme Controlling Marijuana Psychoactivity , is Secreted into the Storage Cavity of the Glandular Trichomes**. In., vol. 46; 2005: 1578-1582.

92. Geissler M, Volk J, Stehle F, Kayser O, Warzecha H: **Subcellular localization defines modification and production of D 9-tetrahydrocannabinolic acid synthase in transiently transformed Nicotiana benthamiana**. In: *Biotechnology Letters.* vol. 40: Springer Netherlands; 2018: 913-919.
93. Schmidt CA, M.; Kayser, O.: **Engineering cannabinoid production in Saccharomyces cerevisiae**. *Biotechnology journal* 2024.
94. Hu JM, W.; Su, Y.; Qian, C.; Fu, W.: **Emerging technologies for advancing microalgal photosynthesis and metabolism toward sustainable production**. *Frontiers* 2023.
95. Einhaus A, Baier T, Kruse O: **Molecular design of microalgae as sustainable cell factories**. *Trends Biotechnol* 2024, **42**(6):728-738.
96. Sethi D, Butler TO, Shuhaili F, Vaidyanathan S: **Diatoms for Carbon Sequestration and Bio-Based Manufacturing**. *Biology (Basel)* 2020, **9**(8).
97. Ova Ozcan D, Ovez B: **Evaluation of the interaction of temperature and light intensity on the growth of Phaeodactylum tricornutum: Kinetic modeling and optimization**. In: *Biochemical Engineering Journal.* vol. 154: Elsevier; 2020: 107456.
98. Cui YT-H, S.R.; Chua, E.T.; Schenk, P.M.: **Development of high-level omega-3 eicosapentaenoic acid (EPA) production from Phaeodactylum tricornutum**. *Journal of Phycology* 2021.
99. Jaramillo-Madrid AC, Abbriano R, Ashworth J, Fabris M, Pernice M, Ralph PJ: **Overexpression of key sterol pathway enzymes in two model marine diatoms alters sterol profiles in phaeodactylum tricornutum**. In: *Pharmaceuticals.* vol. 13; 2020: 1-19.
100. Fabris M, George J, Kuzhiumparambil U, Lawson CA, Jaramillo-Madrid AC, Abbriano RM, Vickers CE, Ralph P: **Extrachromosomal Genetic Engineering of the Marine Diatom Phaeodactylum tricornutum Enables the Heterologous Production of Monoterpenoids**. In: *ACS Synthetic Biology.* vol. 9; 2020: 598-612.
101. George J, Kahlke T, Abbriano RM, Kuzhiumparambil U, Ralph PJ, Fabris M: **Metabolic Engineering Strategies in Diatoms Reveal Unique Phenotypes and Genetic Configurations With Implications for Algal Genetics and Synthetic Biology**. In: *Frontiers in Bioengineering and Biotechnology.* vol. 8; 2020: 1-19.
102. Erdene-ochir E, Shin B-k, Kwon B, Jung C: **Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from Phaeodactylum tricornutum**. In: *Scientific Reports.* Springer US; 2019: 1-10.
103. Slattery SS, Giguere DJ, Stuckless EE, Shrestha A, Briere LAK, Galbraith A, Reaume S, Boyko X, Say HH, Browne TS *et al*: **Phosphate - regulated expression of the SARS - CoV - 2 receptor - binding domain in the diatom Phaeodactylum tricornutum for pandemic diagnostics**. In: *Scientific Reports.* Nature Publishing Group UK; 2022: 1-15.
104. Jenny RJ, Mann KG, Lundblad RL: **A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa**. *Protein Expr Purif* 2003, **31**(1):1-11.
105. BioRender: **BioRender**. In.; 2023.
106. Apt KE, Zaslavkaia L, Lippmeier JC, Lang M, Kilian O, Wetherbee R, Grossman AR, Kroth PG: **In vivo characterization of diatom multipartite plastid targeting signals**. *J Cell Sci* 2002, **115**(Pt 21):4061-4069.
107. Kilian O, Kroth PG: **Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids.** In: *The Plant journal : for cell and molecular biology.* vol. 41; 2005: 175-183.

108. Liu X, Hempel F, Stork S, Bolte K, Moog D, Heimerl T, Maier UG, Zauner S: Addressing various compartments of the diatom model organism Phaeodactylum tricornutum via sub-cellular marker proteins. In: *Algal Research.* vol. 20: Elsevier B.V.; 2016: 249-257.
109. Lingg N, Zhang P, Song Z, Bardor M: The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis. *Biotechnol J* 2012, **7**(12):1462-1472.
110. Erdene-Ochir E, Shin BK, Huda MN, Kim DH, Lee EH, Song DG, Kim YM, Kim SM, Pan CH: Cloning of a novel endogenous promoter for foreign gene expression in Phaeodactylum tricornutum. In: *Applied Biological Chemistry.* vol. 59: Springer Netherlands; 2016: 861-867.
111. Sosa-Carrillo S, Galez H, Napolitano S, Bertaux F, Batt G: Maximizing protein production by keeping cells at optimal secretory stress levels using real-time control approaches. *Nat Commun* 2023, **14**(1):3028.
112. Messaabi A, Merindol N, Bohnenblust L, Fantino E, Meddeb-Mouelhi F, Desgagne-Penix I: In vivo thrombin activity in the diatom Phaeodactylum tricornutum: biotechnological insights. *Appl Microbiol Biotechnol* 2024, **108**(1):481.
113. Diamond A, Diaz-garza AM, Li J, Slattery SS, Merindol N, Fantino E, Meddeb-mouelhi F, Karas BJ, Barnab S, Desgagn I: Instability of extrachromosomal DNA transformed into the diatom Phaeodactylum tricornutum. In., vol. 70; 2023.
114. Zirpel B, Kayser O, Stehle F: Elucidation of structure-function relationship of THCA and CBDA synthase from Cannabis sativa L . In: *Journal of Biotechnology.* vol. 284: Elsevier; 2018: 17-26.
115. Cox JSW, P.: A Novel Mechanism for Regulating Activity of a Transcription Factor That Controls the Unfolded Protein Response. 1996.
116. Slattery SS, Diamond A, Wang H, Therrien JA, Lant JT, Jazey T, Lee K, Klassen Z, Desgagne-Penix I, Karas BJ *et al*: An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum. In: *ACS synthetic biology.* vol. 7; 2018: 328-338.
117. Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips APR, Noddings CM, Brunson JK, Valas RE, Deerinck TJ, Jablanovic J *et al*: Designer diatom episomes delivered by bacterial conjugation. In: *Nature Communications.* vol. 6: Nature Publishing Group; 2015: 1-10.
118. Valliere MA, Korman TP, Woodall NB, Khitrov GA, Taylor RE, Baker D, Bowie JU: A cell-free platform for the prenylation of natural products and application to cannabinoid production. In: *Nature Communications.* vol. 10: Springer US; 2019: 1-9.
119. Javaid M, Haleem A, Singh RP, Suman R: Sustaining the healthcare systems through the conceptual of biomedical engineering: A study with recent and future potentials. *Biomedical Technology* 2023, **1**:39-47.
120. Demain, AL., Vaishnav., P.: Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* 2009, **27**(3):297-306.
121. Dhaouadi F, Awwad F, Diamond A, Desgagné-Penix I: Diatoms' Breakthroughs in Biotechnology: <i>Phaeodactylum tricornutum</i> as a Model for Producing High-Added Value Molecules. *American Journal of Plant Sciences* 2020, **11**(10):1632-1670.
122. Awwad F, Fantino EI, Héneault M, Diaz-Garza AM, Merindol N, Custeau A, Gélinas S-E, Meddeb-Mouelhi F, Li J, Lemay J-F *et al*: Bioengineering of the Marine Diatom

Phaeodactylum tricornutum with Cannabis Genes Enables the Production of the Cannabinoid Precursor, Olivetolic Acid. *International Journal of Molecular Sciences* 2023.
123. Butler T, Kapoore RV, Vaidyanathan S: Phaeodactylum tricornutum: A Diatom Cell Factory. *Trends Biotechnol* 2020, 38(6):606-622.
124. Slattery SS, Giguere DJ, Stuckless EE, Shrestha A, Briere LK, Galbraith A, Reaume S, Boyko X, Say HH, Browne TS *et al*: Phosphate-regulated expression of the SARS-CoV-2 receptor-binding domain in the diatom Phaeodactylum tricornutum for pandemic diagnostics. *Sci Rep* 2022, 12(1):7010.
125. Fabris M, George J, Kuzhiumparambil U, Lawson CA, Jaramillo-Madrid AC, Abbriano RM, Vickers CE, Ralph P: Extrachromosomal Genetic Engineering of the Marine Diatom Phaeodactylum tricornutum Enables the Heterologous Production of Monoterpenoids. *ACS Synth Biol* 2020, 9(3):598-612.
126. Fantino E, Awwad F, Merindol N, Diaz Garza AM, Gélinas S-E, Gajón Robles GC, Custeau A, Meddeb-Mouelhi F, Desgagné-Penix I: Bioengineering Phaeodactylum tricornutum, a marine diatom, for cannabinoid biosynthesis. *Algal Research* 2024, 77:103379.
127. Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips AP, Noddings CM, Brunson JK, Valas RE, Deerinck TJ, Jablanovic J *et al*: Designer diatom episomes delivered by bacterial conjugation. *Nat Commun* 2015, 6:6925.
128. Kimple ME, Brill AL, Pasker RL: Overview of affinity tags for protein purification. *Curr Protoc Protein Sci* 2013, 73:9 9 1-9 9 23.
129. Smale ST: Beta-galactosidase assay. *Cold Spring Harb Protoc* 2010, 2010(5):pdb prot5423.
130. Smale ST: Luciferase assay. *Cold Spring Harb Protoc* 2010, 2010(5):pdb prot5421.
131. Belardinelli JM, Jackson M: Green Fluorescent Protein as a protein localization and topological reporter in mycobacteria. *Tuberculosis (Edinb)* 2017, 105:13-17.
132. Wang S, Ju X, Heuler J, Zhang K, Duan Z, Patabendige HMLW, Zhao S, Sun X: Recombinant Fusion Protein Vaccine Containing Clostridioides difficile FliC and FliD Protects Mice against C. difficile Infection. *American Society for Microbiology* 2023.
133. Ellis RW, Rappuoli R, Ahmed S: Technologies for making new vaccines. *Vaccines (Sixth Edition)* 2013:1182–1199.
134. Dos Santos TW, Goncalves PA, Rodriguez D, Pereira JA, Martinez CAR, Leite LCC, Ferraz LFC, Converso TR, Darrieux M: A fusion protein comprising pneumococcal surface protein A and a pneumolysin derivate confers protection in a murine model of pneumococcal pneumonia. *PLoS One* 2022, 17(12):e0277304.
135. Snape MD, Dawson T, Oster P, Evans A, John TM, Ohene-Kena B, Findlow J, Yu L-M, Borrow R, Ypma E *et al*: Immunogenicity of Two Investigational Serogroup B Meningococcal Vaccines in the First Year of Life. *Pediatric Infectious Disease Journal* 2010, 29(11):e71-e79.
136. Zhao HL, Yao XQ, Xue C, Wang Y, Xiong XH, Liu ZM: Increasing the homogeneity, stability and activity of human serum albumin and interferon-alpha2b fusion protein by linker engineering. *Protein Expr Purif* 2008, 61(1):73-77.
137. Amet N, Lee HF, Shen WC: Insertion of the designed helical linker led to increased expression of tf-based fusion proteins. *Pharm Res* 2009, 26(3):523-528.
138. Bouabe H, Fassler R, Heesemann J: Improvement of reporter activity by IRES-mediated polycistronic reporter system. *Nucleic Acids Res* 2008, 36(5):e28.

139. McCormick AL, Thomas MS, Heath AW: **Immunization with an Interferon-g–gp120 Fusion Protein Induces Enhanced Immune Responses to Human Immunodeficiency Virus gp120**. *The Journal of Infectious Diseases* 2001.
140. George RA, Heringa J: **An analysis of protein domain linkers: their classification and role in protein folding**. *Protein Eng* 2002, **15**(11):871-879.
141. Takamatsu N, Watanabe Y, Yanagi H, Meshi T, Shiba T, Okada Y: **Production of enkephalin in tobacco protoplasts using tobacco mosaic virus RNA vector**. *FEBS Letters* 1990, **269**(1):73-76.
142. Argos P: **An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion**. *J Mol Biol* 1990, **211**(4):943-958.
143. Huston JS, Levinson D, Mudgett-Hunter M, Tai MS, Novotny J, Margolies MN, Ridge RJ, Bruccoleri RE, Haber E, Crea R *et al*: **Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli**. *Proc Natl Acad Sci U S A* 1988, **85**(16):5879-5883.
144. Araujo AP, Oliva G, Henrique-Silva F, Garratt RC, Caceres O, Beltramini LM: **Influence of the histidine tail on the structure and activity of recombinant chlorocatechol 1,2-dioxygenase**. *Biochem Biophys Res Commun* 2000, **272**(2):480-484.
145. Rao MB, Tanksale AM, Ghatge MS, Deshpande VV: **Molecular and biotechnological aspects of microbial proteases**. *Microbiol Mol Biol Rev* 1998, **62**(3):597-635.
146. Danckwardt S, Hentze MW, Kulozik AE: **Pathologies at the nexus of blood coagulation and inflammation: thrombin in hemostasis, cancer, and beyond**. *J Mol Med (Berl)* 2013, **91**(11):1257-1271.
147. Lundblad RL, Bradshaw RA, Gabriel D, Ortel TL, Lawson J, Mann KG: **A review of the therapeutic uses of thrombin**. *Thromb Haemost* 2004, **91**(5):851-860.
148. Takagi T, Doolittle RF: **Amino acid sequence studies on factor XIII and the peptide released during its activation by thrombin**. *Biochemistry* 1974, **13**(4):750-756.
149. Liu Z, Chen O, Wall JBJ, Zheng M, Zhou Y, Wang L, Vaseghi HR, Qian L, Liu J: **Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector**. *Sci Rep* 2017, **7**(1):2193.
150. Slattery SS, Diamond A, Wang H, Therrien JA, Lant JT, Jazey T, Lee K, Klassen Z, Desgagne-Penix I, Karas BJ *et al*: **An Expanded Plasmid-Based Genetic Toolbox Enables Cas9 Genome Editing and Stable Maintenance of Synthetic Pathways in Phaeodactylum tricornutum**. *ACS Synth Biol* 2018, **7**(2):328-338.
151. Erdene-Ochir E, Shin BK, Kwon B, Jung C, Pan CH: **Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from Phaeodactylum tricornutum**. *Sci Rep* 2019, **9**(1):9941.
152. Diamond A, Diaz-Garza AM, Li J, Slattery SS, Merindol N, Fantino E, Meddeb-Mouelhi F, Karas BJ, Barnabé S, Desgagné-Penix I: **Instability of extrachromosomal DNA transformed into the diatom Phaeodactylum tricornutum**. *Algal Research* 2023, **70**.
153. Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I: **New and continuing developments at PROSITE**. *Nucleic Acids Res* 2013, **41**(Database issue):D344-347.
154. Rechsteiner M, Rogers SW: **PEST sequences and regulation by proteolysis**. *Trends Biochem Sci* 1996, **21**(7):267-271.
155. Okonechnikov K, Golosova O, Fursov M, team U: **Unipro UGENE: a unified bioinformatics toolkit**. *Bioinformatics* 2012, **28**(8):1166-1167.

156. Tamura K, Stecher G, Kumar S, Battistuzzi FU: **MEGA11: Molecular Evolutionary Genetics Analysis Version 11**. *Molecular Biology and Evolution* 2021, **38**(7):3022-3027.
157. Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M: **ColabFold: making protein folding accessible to all**. *Nat Methods* 2022, **19**(6):679-682.
158. Majhi BB, Gélinas S-E, Mérindol N, Desgagné-Penix I: **Characterization of norbelladine synthase and noroxomaritidine/norcraugsodine reductase reveals a novel catalytic route for the biosynthesis of Amaryllidaceae alkaloids including the Alzheimer's drug galanthamine**. *bioRxiv* 2022:2022.2007. 2030.502154.
159. Bode W, Mayr I, Baumann U, Huber R, Stone SR, Hofsteenge J: **The refined 1.9 A crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment**. *EMBO J* 1989, **8**(11):3467-3475.
160. Mathews, II, Padmanabhan KP, Ganesh V, Tulinsky A, Ishii M, Chen J, Turck CW, Coughlin SR, Fenton JW, 2nd: **Crystallographic structures of thrombin complexed with thrombin receptor peptides: existence of expected and novel binding modes**. *Biochemistry* 1994, **33**(11):3266-3279.
161. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, McVeigh R, O'Neill K, Robbertse B *et al*: **NCBI Taxonomy: a comprehensive update on curation, resources and tools**. *Database (Oxford)* 2020, **2020**.
162. Cheng CM, Meyer-Massetti C, Kayser SR: **A Review of Three Stand-Alone Topical Thrombins for Surgical Hemostasis**. *Clinical Therapeutics* 2009, **31**:1.
163. Bhandari M, Ofosu FA, Mackman N, Jackson C, Doria C, Humphries JE, Babu SC, Ortel TL, Hoffman Van Thiel D, Walenga JM *et al*: **Safety and efficacy of thrombin-JMI: a multidisciplinary expert group consensus**. *Clin Appl Thromb Hemost* 2011, **17**(1):39-45.
164. Freedman SJ, Blostein MD, Baleja JD, Jacobs M, Furie BC, Furie B: **Identification of the phospholipid binding site in the vitamin K-dependent blood coagulation protein factor IX**. *J Biol Chem* 1996, **271**(27):16227-16236.
165. Castellino FJ, McCance SG: **The kringle domains of human plasminogen**. *Ciba Found Symp* 1997, **212**:46-60; discussion 60-45.
166. Wei Wu J, Bancroft D, Suttie JW: **Structural features of the kringle domain determine theintracellular degradation of under-gcarboxylated prothrombin: Studies of chimeric rat/human prothrombin**. *CELL BIOLOGY* 1997, **94**.
167. Schechter I, Berger A: **On the size of the active site in proteases. I. Papain. 1967**. *Biochem Biophys Res Commun* 2012, **425**(3):497-502.
168. Johnson DJ, Adams TE, Li W, Huntington JA: **Crystal structure of wild-type human thrombin in the Na+-free state**. *Biochem J* 2005, **392**(Pt 1):21-28.
169. Tang X, Liu X, Tao G, Qin M, Yin G, Suo J, Suo X: **"Self-cleaving" 2A peptide from porcine teschovirus-1 mediates cleavage of dual fluorescent proteins in transgenic Eimeria tenella**. *Vet Res* 2016, **47**(1):68.
170. Wang Y, Wang F, Wang R, Zhao P, Xia Q: **2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori**. *Sci Rep* 2015, **5**:16273.
171. Sonenberg N, Svitkin Y, Siddiqui N: **Internal Ribosome Entry Site-Mediated Translation**. In: *Encyclopedia of Cell Biology.* 2016: 307-316.
172. **Spectrum [YFP]** [https://www.aatbio.com/fluorescence-excitation-emission-spectrum-graph-viewer/yfp]

173. Zhuang. X, Ha. T, Kim. HD, Centner. T, Labeit. S, Chu. S: **Fluorescence quenching a tool for single-molecule protein-folding study**. *BIOPHYSICS AND COMPUTATIONAL BIOLOGY* 2000.
174. Di Cera E: **Thrombin**. *Mol Aspects Med* 2008, **29**(4):203-254.
175. Coughlin SR: **Thrombin signalling and protease-activated receptors**. *Nature* 2000, **407**:258-264.
176. Lindahl T, Bjerke M, Ramström S, Vretenbrant K: **Platelet activation via PAR4 is involved in the initiation of thrombin generation and in clot elasticity development**. *Thrombosis and Haemostasis* 2017, **97**(03):417-424.
177. Tuteja N: **Signaling through G protein coupled receptors**. *Plant Signal Behav* 2009, **4**(10):942-947.
178. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP *et al*: **The Phaeodactylum genome reveals the evolutionary history of diatom genomes**. *Nature* 2008, **456**(7219):239-244.
179. Li L, Dou N, Zhang H, Wu C: **The versatile GABA in plants**. *Plant Signal Behav* 2021, **16**(3):1862565.
180. Allen AE, Vardi A, Bowler C: **An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms**. *Curr Opin Plant Biol* 2006, **9**(3):264-273.
181. Fu W, Chaiboonchoe A, Dohai B, Sultana M, Baffour K, Alzahmi A, Weston J, Al Khairy D, Daakour S, Jaiswal A *et al*: **GPCR Genes as Activators of Surface Colonization Pathways in a Model Marine Diatom**. *iScience* 2020, **23**(8):101424.
182. Gallwitz M, Enoksson M, Thorpe M, Hellman L: **The extended cleavage specificity of human thrombin**. *PLoS One* 2012, **7**(2):e31756.
183. Baïet B, Burel C, Saint-Jean B, Louvet R, Menu-Bouaouiche L, Kiefer-Meyer MC, Mathieu-Rivet E, Lefebvre T, Castel H, Carlier A *et al*: **N-glycans of Phaeodactylum tricornutum diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme**. *J Biol Chem* 2011, **286**(8):6152-6164.
184. Driouich A, Zhang GF, Staehelin LA: **Effect of brefeldin A on the structure of the Golgi apparatus and on the synthesis and secretion of proteins and polysaccharides in sycamore maple (Acer pseudoplatanus) suspension-cultured cells**. *Plant Physiol* 1993, **101**(4):1363-1373.
185. Langhans M, Hawes C, Hillmer S, Hummel E, Robinson DG: **Golgi regeneration after brefeldin A treatment in BY-2 cells entails stack enlargement and cisternal growth followed by division**. *Plant Physiol* 2007, **145**(2):527-538.
186. George J, Kahlke T, Abbriano RM, Kuzhiumparambil U, Ralph PJ, Fabris M: **Metabolic Engineering Strategies in Diatoms Reveal Unique Phenotypes and Genetic Configurations With Implications for Algal Genetics and Synthetic Biology**. *Front Bioeng Biotechnol* 2020, **8**:513.
187. Diaz-Garza AM, Merindol N, dos Santos KCG, Lavoie-Marchand F, Ingalls B, Desgagné-Penix I: **No two clones are alike: characterization of heterologous subpopulations in a transgenic cell line of the model diatom Phaeodactylum tricornutum**. *Microbial Cell Factories* 2024, **23**(1):286.
188. Diamond A, Diaz-Garza AM, Li J, Slattery SS, Merindol N, Fantino E, Meddeb-Mouelhi F, Karas BJ, Barnabé S, Desgagné-Penix I: **Instability of extrachromosomal DNA transformed into the diatom Phaeodactylum tricornutum**. *Algal Research* 2023, **70**:102998.

189. Sharma N, Fantino EI, Awwad F, Mérindol N, Augustine A, Meddeb F, Desgagné-penix I, Meddeb-Mouelhi F, Desgagné-penix I: **Impact of different light characteristics on the growth and lipid content of diatom Phaeodactylum tricornutum transconjugant strains**. In: *American Journal of Plant Sciences.* 2023: 41-63.
190. Lyska D, Engelmann K, Meierhoff K, Westhoff P: **pAUL: a gateway-based vector system for adaptive expression and flexible tagging of proteins in Arabidopsis**. *PLoS One* 2013, **8**(1):e53787.