Algal Research 95 (2026) 104670

Contents lists available at ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

Stabilizing heterologous protein production in *Phaeodactylum tricornutum* through cystatin overexpression☆

Elisa Fantino [a], Natacha Mérindol [a], Carolina Gabriela Gajón Robles [a], Fatima Awwad [a], Karen Cristine Gonçalves dos Santos [a], Kimy-Li Rhéaume [a], Marie-Claire Goulet [b,c], Dominique Michaud [b,c], Isabel Desgagné-Penix [a,c,*]

[a] Département de biochimie, chimie, physique et science forensique, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada
[b] Department de Phytologie, Faculté des Sciences de l'Agriculture et de l'Alimentation, Université Laval, Québec, QC, Canada
[c] Centre de Recherche et d'Innovation sur les Végétaux (CRIV), Université Laval, Québec, QC, Canada

ARTICLE INFO

Keywords:
Marine diatoms
Cystatin
Biofactories
Metabolic engineering
Mean fluorescence intensity

ABSTRACT

The marine diatom *Phaeodactylum tricornutum* is an emerging platform for metabolic engineering and recombinant protein production. However, heterologous protein accumulation often shows strong cell-to-cell variability and decreases over successive generations in bioengineered diatoms. In this study, we investigated whether overexpression of an endogenous cysteine protease inhibitor, cystatin (*Pt*CYS), could stabilize heterologous protein production in *P. tricornutum*. We generated extrachromosomal expression strains co-expressing the *yellow fluorescent protein* (*yfp*) gene with *Pt*CYS. The anti-protease activity of *Pt*CYS was verified *in vitro*. Co-expression of YFP with the full-length *Pt*CYS coding sequence resulted in higher YFP accumulation (~1.6-fold). Increased heterologous protein accumulation was impaired by fusion with *Pt*CYS and dependent on the presence of an endogenous signal peptide. The reintroduction of *Pt*CYS into three independent YFP-only lines with low initial fluorescence led to a marked increase, by ~2-fold, in YFP signal in one genetic background, highlighting the potential of this strategy as well as strain-dependent effects. Importantly, changes in YFP accumulation were not associated with increased *yfp* transgene copy number. Together, these results demonstrate that cystatin-mediated modulation of proteolytic activity can influence heterologous protein stability in *P. tricornutum*, while also revealing strong context- and genetic background–dependent limitations. This work provides critical design considerations for improving recombinant protein production in diatom-based biofactories.

1. Introduction

The global economic movement strives to build a greener and more sustainable society. In this context, significant advancements have been made in biotechnology research, particularly in the industrial applications of *Phaeodactylum tricornutum*. The marine diatom has emerged as a valuable platform for metabolic engineering and molecular farming [1]. *P. tricornutum* has the potential to produce an annual yield of 132 metric tons of dry biomass, demonstrating its scalability for industrial applications as a prospective high-value protein production platform. Significant progress has been made in engineering this diatom for recombinant protein expression (*e.g.*, SARS-CoV-2 proteins; [2]), therapeutic metabolite production (*e.g.*, [3–11]), and omega-3 fatty acid [12]. Key advantages of using microalgae for recombinant protein production are its N-linked glycosylation pathway, which resembles that of mammals [13] and its ability to fold complex heteromeric proteins, allowing for accurate post-translational modifications [14]. By comparison, bacterial expression systems are unable to perform glycosylation, and yeast production systems produce rich, heterogeneous hypermannosyl structures that differ from the more complex mammalian glycan structures and can lead to antigenicity [15,16]. Recombinant proteins produced in insect and mammal cells are expensive and time-consuming to use [17,18]; and mammalian cell cultures are susceptible to viral contamination [18,19]. Plants are ecologically interesting hosts, but their genetic manipulation is time-consuming and requires fertile land [20]. Taken together, diatoms combine the advantages of

☆ This article is part of a Special issue entitled: 'Emerging Aspects in Microalgal Biotechnology' published in Algal Research.

* Corresponding author at: Département de biochimie, chimie, physique et science forensique, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada.
E-mail address: Isabel.Desgagne-Penix@uqtr.ca (I. Desgagné-Penix).

https://doi.org/10.1016/j.algal.2026.104670
Received 11 July 2025; Received in revised form 27 January 2026; Accepted 20 March 2026
Available online 24 March 2026

mammalian and plant systems (as photosynthetic organisms) for recombinant proteins, biopharmaceuticals and vaccines production including high resource efficiency, easy scale-up, complex protein folding. Since the sequencing and annotation of its genome in 2008, *P. tricornutum* has been recognized as a highly suitable organism for large-scale culture due to its capacity to grow in seawater. Nevertheless, several challenges remain. *Diamond* et al. reported that episomal DNA sequences may undergo alterations during transformation and prolonged culture [21]. Similarly, *Diaz-Garza* et al. observed phenotypic heterogeneity [22], and *Awwad* et al. documented a gradual decline in metabolite production despite conserved DNA sequences [6], suggesting the presence of post-transcriptional regulatory mechanisms. *Messaabi* et al., showed endogenous thrombin activity in *P. tricornutum* cells, resulting in the *in vivo* splitting of two proteins linked by the thrombin cleavage sequence [23]. All these findings indicate that even though *P. tricornutum* is an efficient protein production platform, further molecular engineering, optimization, and manipulation are needed to enhance its potential as a model species.

One of the main factors limiting heterologous protein yield is proteolysis mediated by endogenous proteases, which play a crucial role in many physiological processes, including cell division, growth, reproduction, development, photosynthesis, senescence, protein turnover, programmed cell death, and stress responses [24–27]. Papain-like peptidases, a class of cysteine (Cys) proteases, are particularly well-studied for their roles in various biological processes [28,29]. In *P. tricornutum*, proteases are enriched in the ovoid morphotype, associated with stress [28]. Given the critical role of proteases in protein accumulation, different strategies have been proposed to reduce their effects in other heterologous hosts. Competitive inhibitors, antisense or RNA silencing targeting proteases have been successfully employed to increase the accumulation of recombinant proteins in plants [30–32]. For instance, the co-expression of a tomato cathepsin D inhibitor (*Sl*CDI) and of bovine aprotinin in potato plants decreased the levels of cathepsin D-like protease during the protein extraction process [33] and led to *in situ* protection of proteins in *Solanum tuberosum* leaves [34]. Likewise, the co-secretion of an accessory protease inhibitor reduced the degradation of IgG monoclonal antibodies secreted from tobacco roots [35], while silencing *cysteine protease* increased the accumulation of recombinant human interleukin-10 in *Nicotiana tabacum* [36]. Cystatins interact with Cys proteases *via* three key regions: two conserved glycine residues in the N-terminal region, a highly conserved Gln-X-Val-X-Gly motif in the first β-hairpin loop, and a dipeptide motif (ProTrp or LeuTrp) in the second β-hairpin loop at the C-terminal [37–39]. Co-expression of the tomato cystatin *Sl*CYS8 increased target recombinant protein yields in *Nicotiana benthamiana* by acting as a pseudo-substrate, blocking the protease active site and thereby inhibiting proteolytic activity. Specifically, the co-expression of *Sl*CYS8, alongside a murine IgG in *N. benthamiana*, increased antibody yield by nearly 40% [40]. *Sl*CYS8 had positive effects on both the amount of fully assembled IgG antibody purified from leaf tissue and the stability of active antibody fragments containing the heavy chain Fc domain in *N. benthamiana* extracts [41]. Similarly, transgenic tobacco plants constitutively expressing the rice cystatin oryzacystatin-I led to decreased cysteine protease activity, which allowed for higher accumulation and activity of the *E. coli* glutathione reductase in the host plant [42]. Interestingly, the transient expression of *Sl*CYS8 fused to the human alpha-1-antichymotrypsin (α1ACT) in *N. benthamiana* increased α1ACT accumulation by mechanisms independent of protease inhibition, *i.e.* through improved translation and overall stabilization of its tertiary structure preventing proteolytic degradation and/or polymerization [43].

In the present study, we evaluated the overexpression of a novel endogenous *P. tricornutum* cystatin (*Pt*CYS) in the diatom as a strategy to inhibit Cys proteases and enhance the stability and accumulation of heterologous proteins. We cloned a *Pt*CYS candidate identified in the diatom's proteomic and transcriptomic databases and confirmed accumulation and activity *in vitro*. The yellow fluorescent protein (YFP) was used as a marker of protein stability and co-expressed with the immature and mature (lacking the secretory signal peptide) versions of *Pt*CYS. Co-expression with the immature form of *Pt*CYS led to a significant increase in YFP fluorescence in *P. tricornutum* cells, demonstrating enhanced accumulation. These results provide proof-of-concept for the application of *Pt*CYS as an accessory protease inhibitor to boost recombinant protein production in *P. tricornutum*. These findings also offer insights into the functional role of CYSs in diatoms and open new avenues for metabolic engineering and molecular farming.

2. Materials and methods

2.1. Cystatin in silico analysis and molecular phylogenetic tree

The Phatr3_EG01374.p1 amino acid sequence was selected from the most recent genome annotation data [44]. Blastp alignment searches were performed against the UniProt database, confirming that Phatr3_EG01374 has a cystatin domain. A phylogenetic tree was constructed using cystatin domain amino acid sequences from twenty-three plant species, three green microalgae, two cyanobacteria putative cystatins, one diatom and marine microalgae (Table S1). These sequences were selected by performing a blastp search (BLAST+ version 2.14 [45]) with a maximum E-value threshold of 1×10^{-5}, using the partial multicystatin cystatin sequence from *Solanum lycopersicum* (AAF23128; [46]) as the query. The cystatin domains were aligned with Clustal Omega using the R package msa 1.34.0, with default parameters. Distance was then computed using the model Blosum62 (phangorn 2.12.1) and the tree was constructed with the Neighbour Joining method (ape 5.8–1), with 1000 bootstrap replicates. The consensus tree was obtained with MEGA 12 [47]. The signal peptide was predicted using the webserver of SignalP version 6.0 with default settings [48]. Amino acid sequence was analyzed using InterProScan [49] and protein-targeting sequence was predicted with PHOBIUS, version 5.6, using the default parameters [50].

2.2. Cystatin modelization and Cys protease docking simulations

The protein structure of *P. tricornutum* cystatin (Phatr3_EG01374.p1, *Pt*CYS) was predicted using AlphaFold3 [51] and superimposed with *Colocasia esculenta* tarocystatin (*Ce*Cys extracted from the crystal structure 3IMA [52], as a template for cystatins) with the 'super' command using PyMOL default parameters, version 2.6.0 (Schrodinger). Enzyme–inhibitor interaction simulation was performed using *Pt*CYS and the predicted *P. tricornutum* peptidase (a C1A papain C-terminal domain-containing protein, Phatr3_J4936.t1), as target protease, using the amino acid sequences of both candidates as queries for a multimer prediction in AlphaFold3. Predicted folding and H-bonds were analyzed and displayed using PyMOL.

2.3. Plasmid constructions

All plasmid constructs were done by Gibson assembly using the NEBuilder® HiFi DNA Assembly Bundle for Large Fragments (New England Biolabs, Canada). The fragments used for the assemblies were amplified by PCR with PrimeSTAR GXL DNA Polymerase (Takara Bio, Japan) following the manufacturer's protocol. Sh ble episome harbors *Sh ble*, Zeocin resistance, gene driven under the *fucoxanthin chlorophyll-binding protein C* (*fcpC*) promoter [7]. The YFP episome was cloned by excluding the TKS-OAC CDS from the *Pt*OA3 vector from Awwad et al. [6] and assembling the backbone. The *Pt*CYS fused to YFP episomes (SP*Pt*CYS-YFP, *Pt*CYS-YFP, and SP-YFP) were designed by cloning: the complete cystatin coding sequence, the mature version, without the 66 bp from the 5′, and the cystatin's secretory Signal Peptide (SP), 66 bp from the 5′, downstream the 40SRPS8 promoter sequence and fused to YFP gene with a (GGGGS)3 peptide linker, respectively. The episomes containing *Pt*CYS versions and YFP (SP*Pt*CYS_YFP and *Pt*CYS_YFP) in different expression cassettes were generated by cloning the coding

sequence downstream of the *elongation factor alpha 2* (*EF2*) promoter from TKK031 plasmid [53], tagged with 3xHemagglutinin (3xHA) in C-terminal, and introducing it upstream of the YFP expression cassette of the vector. The SP*Pt*CYS_nat episome was designed by cloning the SP*Pt*CYS expression cassette (*EF2prom::SPPtCys:3xHA::FcpA*) in the *N-acetyltransferase* gene (*nat*) episome which confers resistance to Nourseothricin [14]. All DNA and primer sequences used in this study are listed in Tables S2 and S3 respectively.

2.4. Microbial strains, growth conditions and transformation

Escherichia coli (NEB® 10-beta, New England Biolabs, Canada) was grown in Luria-Bertani broth (LB) supplemented with appropriate antibiotics (chloramphenicol, 30 mg/L). Plasmids were extracted using a miniprep kit (Biobasic EZ10 miniprep kit, NY, USA), sequenced by CCIB DNA Core (Massachusetts General Hospital, United States of America) to confirm the assemblies and then amplified in Epi 300 strain containing pTA-MOB plasmid to allow conjugation with diatoms, as described earlier [54]. Briefly, 1 mL of wild-type *P. tricornutum* was seeded on 0.5 X L1, 1% (w/v) agar plates and grown at 18 °C on a light/dark cycle of 16/8 h 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 of *E. coli* cultures containing the assembled plasmid and pTA-MOB was grown at 37 °C under agitation to an OD_{600} 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. 5× L1, 5% (w/v) LB, ~1% (w/v) 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% (w/v) agar plates supplemented with Zeocin 50 mg/L or Zeocin plus Nourseothricin 200 mg/L for selection and incubated at 18 °C. Transformed colonies appeared after 2 weeks.

2.5. YFP intensity measurement

Seventy-two *P. tricornutum* clones per *Pt*CYS episomes were randomly selected and plated on new selective media supplemented with Zeocin 50 mg/L. After 10 days, 31 colonies per strain were transferred into 200 uL of liquid L1 plus Zeocin into 96 well plates and re-transferred to a new plate every 6 days. After five transfer rounds, the clones ($n = 31$) were analyzed in the CytoFLEX S flow cytometer (Beckman-Coulter) equipped with violet (405 nm), blue (488 nm), yellow green (561 nm), and red (638 nm) lasers. Seventy-five μL of 4-day-old cultures were filtered and transferred to clear 96-well plates with 200 μL of L1 in each well. Chlorophyll autofluorescence was detected on the PerCP channel (690/50 nm), while YFP fluorescence was detected on the FITC channel (525/40 nm). The percentage of YFP fluorescence cells and mean fluorescence intensities (MFI) of 10,000 cells from three biological repeats were registered. Images and statistics were analyzed using the BD FlowJo software, version 10 (BD Biosciences, La Jolla, CA, USA). The Sh ble strain was used as a negative control.

2.6. Confocal microscopy

Live cell images of three independent lines from each strain (YFP, SP*Pt*CYS-YFP, *Pt*CYS-YFP, and SP-YFP) 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, according to Awwad et al. [6]. 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. Combined images were generated using Leica Las X and analyzed by Fiji software. The Sh ble strain was used as a negative control.

2.7. PtCYS:3xHA detection

Twenty-five mL of SP*Pt*CYS_YFP clones 10, 40, and 64, YFP clone 4, and the Sh ble strain were grown for 6 days in L1 media containing Zeocin (50 mg/L) and then centrifuged at 3500 x*g* for 10 min at 4 °C. The supernatant was discarded, and pellets were resuspended in 50 mM Tris-HCl, pH 7, containing 10% (*v*/v) glycerol, and sonicated 6 times at 35% amplitude, with a 30s pulse on, and 30s off for 6 min total using FisherbrandTM Model 505 Sonic Dismembrator (Thermo Fisher Scientific). Protein extracts were centrifuged at 20,000 x*g* for 30 min at 4 °C. Supernatants containing the total soluble protein fractions were kept at 4 °C to be used for western blot and cystatin activity assay. Proteins were quantified with the RC DC™ Protein Assay Kit I (Bio-Rad, Canada).

To detect *Pt*CYS:3xHA, 25 μg of total protein were loaded and resolved by 15% (w/v) SDS-PAGE. The proteins were then transferred to a 0.2 μm PVDF membrane, using the Trans-Blot® Turbo™ Transfer System 1,704,150, under the following settings: 1.3 mA constant and 15 V for 20 min. The resulting blot was blocked in 5% (w/v) milk and incubated overnight at 4 °C in a 1:1000 dilution of anti-Hemagglutinin (α-HA) from ThermoFisher Scientific (Illinois USA, cat. #MA1-21316). After three washes with Tris-buffered saline with 0.1% (v/v) Tween 20 (TBST) solution, the blot was incubated for 1 h in a 1:20,000 dilution in 5% (w/v) milk of Immun-Star Goat Anti-Mouse (GAM)-HRP conjugate from Bio-Rad (Ontario Canada, cat. #1705047). The membrane was again washed three times using the TBST solution. Protein detection was done using Clarity Max Western ECL Substrate-Luminol solution (Bio-Rad, Canada). Chemiluminescence detection and Ponceau S staining (Glacial Acetic Acid 5% v/v, Ponceau Red dye 0.1% m/v) were visualized using the ChemiDoc Imaging System and the Image Lab™ Software (Bio-Rad, USA).

2.8. P. tricornutum protease assays

Cystatin enzymatic assay was performed and adapted from Tremblay et al. [55]. Briefly, 7 μg of total protein were mixed with 10 mM L-cysteine in 50 mM Tris-HCl, pH 7, and 14 μM of the synthetic peptide substrate Z-Phe-Arg-methylcoumarin for cathepsin L-like activities, in a final volume of 100 μL. Substrate hydrolysis was monitored for 10 min at 25 °C using a BioTek Synergy H1 fluorimeter (Agilent Technologies, Canada), with excitation and emission filters of 360 nm and 450 nm, respectively.

2.9. YFP fluorescence rescue

To check whether overexpression of the cystatin could rescue YFP intensity, six YFP-alone clones (1, 11, 32, 36, 41, 68) presenting low YFP intensity were selected. The presence of aYFP cassette was first confirmed by colony PCR [6], using 10 uL of boiled *P. tricornutum* cultures as templates, Taq DNA Polymerase (New England Biolabs, Canada) and the primers FcpCt PCR F and pPtGE30 bb R, listed in Supplementary Table 3. The thermocycling conditions were as follows: initial denaturation at 95 °C for 30 s, followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 51 °C for 30 s, and extension at 68 °C for 1 min 45 s, with a final extension of 5 min at 68 °C. Three independent colony PCR positive clones: YFP-alone 1, 32, and 36, were transformed with the SP*Pt*CYS_nat or nat episomes, as described in Section 2.4. After transformation, 36 clones of each strain (YFP_1_SP*Pt*CYS_nat, YFP_32_SP*Pt*-CYS_nat, YFP_36_SP*Pt*CYS_nat, YFP_1_nat, YFP_32_nat and YFP_36_nat) were randomly selected and plated on selective media supplemented with 50 mg/L Zeocin and 200 mg/L Nourseothricin. After 10 days, 31 colonies per strain were transferred into 200 μL of liquid L1 plus Zeocin and Nourseothricin into 96 well plates and re-transferred to a new plate every 6 days. After five transfer rounds, three biological repeats of 4-

day-old cultures ($n = 31$ per strain) were analyzed in the flow cytometer as described before. The Sh ble_nat strain was used as a negative control.

2.10. DNA extraction and copy number quantification

To compare the YFP plasmid copy number between the YFP36_SP*Pt*CYS_nat (clones 2, 5, 7, 10, 26), YFP36_nat (clones; 1, 8, 9, 15, 20) and sh ble_nat strains, DNA was extracted according to [56]. Briefly, 15 mL of 4-day-old algal cultures were centrifuged at 17,000 x*g* for 1 min. The cell pellet was solubilized in 200 µL extraction buffer (2% [w/v] cetyltrimethylammonium bromide (CTAB), 100 mM Tris-HCl, pH 8.0, 20 mM EDTA, 1.4 M NaCl, containing 2% (v/v) freshly added β-mercaptoethanol) by vortexing, followed by the addition of 200 µL of chloroform/isoamyl alcohol (24:1) and incubation at 65 °C for 20 min under shaking at 1200 x*g*. After centrifugation at 17,000 x*g* for 10 min at 10 °C, DNA in the aqueous phase was transferred to a new microcentrifuge tube. DNA was precipitated by adding 0.7 volume of isopropanol and incubated at room temperature for 10 min, followed by centrifugation at 17,000 x*g* for 10 min at 4 °C. The pellet was washed with 1 mL 70% (v/v) ethanol, and the air-dried pellet resuspended in 50 µL of DNase-free water supplemented with 1 µL RNase A (10 mg/mL) and incubated for 15 min at 37 °C. DNA was quantified using a Nanophotometer (IMPLEN).

Relative copy numbers of the *YFP* gene were measured by qPCR using the Luna Universal Master Mix (New England Biolabs, Canada), according to the manufacturer's protocol. The primers to amplify YFP and the reference genes (*Tubulin α* and *Elongation factor 1 α*) were extracted from *Augustine* et al. [57], while the oligonucleotides to detect *Sh ble* were taken from *Diaz-Garza* et al. [22]. The YFP vector was used to determine copy numbers. Dilutions ranging from 5 to 1000 pg of vector were used to perform the standard curve. Plasmid copy numbers were estimated following this formula: (YFP vector (ng) * 6,022,1023) / (YFP (MW) * 1109) according to Integrated DNA Technologies calculations converting from nanograms to copy number. *YFP* gene copy numbers for each strain were determined from 100 ng of DNA using the equation of the standard curve $y = a*\log(x) + b$, where $y = Ct$ and $x =$ plasmid copy number (Fig. S2). Each sample copy number was then determined with the formula $10^{((Ctsample-b)/a)}$, as in Beauchemin et al. [58], normalized to the DNA weight (ng). All primers used in the qPCR are listed in Supplementary Table 3.

3. Results and discussion

3.1. PtCYS, a cystatin candidate from P. tricornutum, exhibits the core structural features of plant cystatins

To select and identify a *P. tricornutum* cystatin, to use it for further assays, two sequence described as "cystatin domain" (Phatr3_EG00714 and Phatr3_EG01374) were extracted for the most recent genome annotation [44]. Only the sequence Phatr3_EG01374 was validated by blastp analysis performed against the UniProt database and showed a significant match to a UniProt entry annotated as a *cystatin/cystatin inhibitor*. The second protein, Phatr3_EG00714, lacked any blastp match and was therefore excluded from further analysis. Moreover, the partial multicystatin cystatin sequence from *Solanum lycopersicum* (AAF23128; [46]) was used as a query against the Phatr3 annotation database, available in the Ensembl archive (protists.ensembl.org; [44]), to detect any previously unannotated cystatin domain–containing sequences; however, no additional sequences were identified.

A phylogenetic tree was constructed with the Phatr3_EG01374 and the cystatin domain sequences of 23 plant species, two putative cystatins from cyanobacteria, one cystatin sequence from the diatom *Pseudo-nitzschia multistriata*, three from green microalgae and one from the marine microalga *Emiliania huxleyi* (Supplementary data_Table 1). The phylogenetic analysis suggested that cystatins were divided into three major clusters and *Pt*CYS set comprise two subgroups. The candidate was closer to a cystatin domain–like protein from *Hibiscus syriacus* (*Hs*CYS-like), showing 85% sequence coverage and 34.78% sequence identity, and the cystatin domain of the green microalga *Trebouxia* sp. A1-2 (*Ta*CYS), with 31.58% identity (Fig. 1A). In addition, *Pt*CYS aligned with the terrestrial cyanobacterium *Crinalium epipsammum* (*Ce*CYS), exhibiting 69% sequence coverage and 28% identity. Finally, two green microalgal species within the same clade, *Raphidocelis subcapitata* (*Rs*CYS) and *Klebsormidium nitens* (*Kn*CYS), showed identity rates of 26.60% and 34.78%, with sequence coverages of 87% and 39%, respectively, relative to *Pt*CYS (Fig. 1A). Overall, *Pt*CYS clustered predominantly with cystatin sequences from microalgae and cyanobacteria rather than with higher plant cystatins, with the exception of *Hs*CYS-like and a cucumber cystatin sequence (*Cucumis sativus*; *Cs*CYS), which showed a low identity percentage (19.57%). The putative sequence encoded a 163 amino acid protein that included a predicted, 22-amino acid secretory signal peptide, similar to several members of the cystatin family [59] (Fig. 1B, upper panel). The complete (immature) sequence presented a molecular mass of 18.8 kDa, whereas the mature version, lacking the secretory signal peptide, has a molecular mass of 11.7 kDa. These sizes are consistent with those reported for plant cystatins, as most plant cystatins have molecular masses in the range of 7.8–16 kDa [56,60], and a few have molecular masses of approximately 23 kDa due to a C-terminal extension, which contributes to the inhibition of legumain-like protease activity [61,62].The encoded protein displayed the characteristic structural motifs of the plant cystatin family members (Fig. 1B), including a conserved Gly-Gly motif in the N-terminal region (or N-trunk), a Ser-**Ala**-Ala-**Phe**-**Ala**-**Val**-Ser-Lys-Leu-Ser-Glu aa string similar to the conserved α-helix of plant cystatins, a Gln-X-Val-X-Gly motif for protease inhibition in the hairpin loop of the central region, and a tryptophan residue at position 135 matching the conserved Pro- (or X)-Trp motif characteristic of the second C-terminal inhibitory loop (Fig. 1B) [63]. Moreover, *Pt*CYS exhibited an extended sequence between the conserved α-helix and the Gln-X-Val-X-Gly motif, which is uncommon in other cystatins. Structurally, *Pt*CYS included the four typical canonical antiparallel β-strands wrapped around the central α-helix of plant cystatins [63]. The N-terminal trunk and two inhibitory loops of the protein that interact with Cys proteases were also observed (Fig. 1C). *In silico* structure predictions suggested that *Pt*CYS contained unique secondary structural elements, including two extra α-helices: one positioned adjacent to the conserved XX-Ala-X-Phe-Ala-Val motif and one found in the C-terminal region. When superimposed with tarocystatin *Ce*CYS, conserved regions aligned closely while differences in structure, particularly in the C-terminal region, became evident. These observations suggested that while *Pt*CYS retained the core structural features of plant cystatins, it also exhibited unique regions, notably in its C-terminal loop, which might influence its inhibitory activity against Cys proteases.

3.2. The cystatin candidate interacts with a papain candidate active site in silico

To further validate the potential functionality of *Pt*CYS as a cystatin candidate, we studied its interaction with a papain-like Cys protease, as predicted *in silico*. Papain-like proteases are main targets of cystatins [64]. From the *P. tricornutum* transcriptome data, ten predicted papain-like Cys protease candidates were identified using the EnsemblProtist database [51] (Table S4). The first candidate was selected for further analysis and modeled using AlphaFold3. The resulting structure showed high similarity to the known structure of papain [29], featuring two lobes of similar size, separated by an active site cleft spanning the whole protein's width (Fig. 1D). This cleft contains the catalytic residues typically found in papain, namely Gln22, Cys31, His202, Asn224, and Trp226, with Cys25 and His159 forming the catalytic dyad and supporting residues Gln19, Asn175, and Trp177 maintaining the nucleophilic character essential for proteolytic activity [29].

Interaction between *Pt*CYS and the predicted papain-like protease

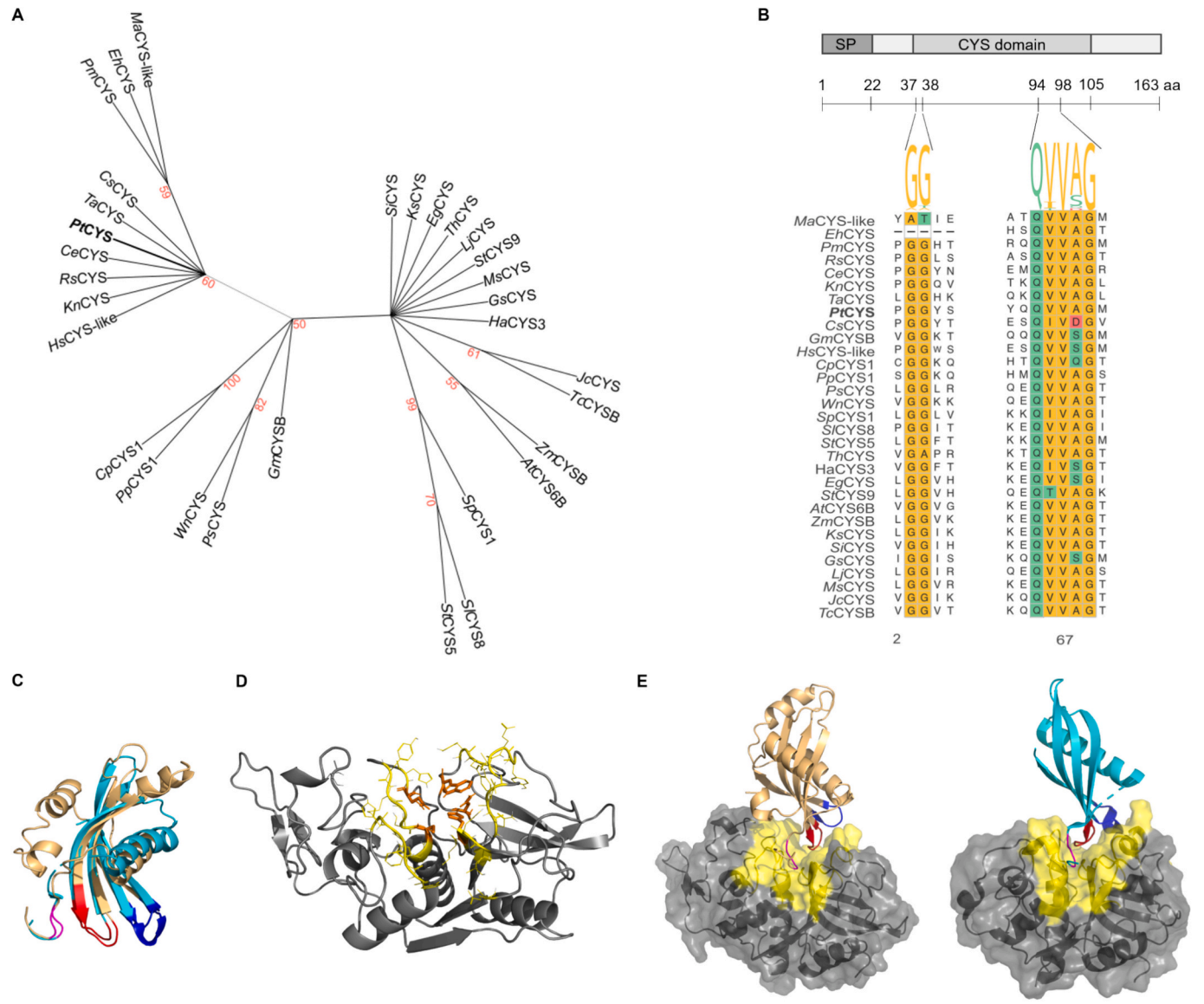


Fig. 1. Identification of *P. tricornutum* cystatin *Pt*CYS. A. Amino acid sequence-based phylogenetic tree was constructed for *Pt*CYS (Phatr3_EG01374.p1) along with 23 plant species (*At*, *Arabidopsis thaliana*; *Cp*, *Ceratodon purpureus*; *Cs*, *Cucumis sativus*; *Eg*, *Elaeis guineensis*; *Gm*, *Glycine* max; *Gs*, *Glycine soja*; *Ha*, *Helianthus annuus*; *Hs*, *Hibiscus syriacus*; *Jc*, *Jatropha curcas*; *Ks*, *Knorringia sibirica*; *Lj*, *Lotus japonicus*; *Ms*, *Medicago sativa*; *Ps*, *Picea sitchensis*; *Pp*, *Physcomitrella patens*; *Si*, *Sesamum indicum*; *Sl*, *Solanum lycopersicum*; *Sp*, *Solanum pennellii*; *St*, *Solanum tuberosum*; *Th*, *Tarenaya hassleriana*; *Tc*, *Theobroma cacao*; *Wn*, *Wollemia nobilis*; *Zm*, *Zea mays*), two putative cystatins domains from cyanobacteria (*Ce*, *Crinalium epipsammum* and *Ma*, *Microcystis aeruginosa*), one cystatin sequence from the diatom *Pseudo-nitzschia multistriata* (*Pm*), three sequences from green microalgae (*Kn*, *Klebsormidium nitens*; *Rs*, *Raphidocelis subcapitata* and *Ta*, *Trebouxia sp. A1-2*) and one from the marine microalga *Emiliania huxleyi* (*Eh*). The evolutionary history, inferred using the neighbour-joining method, was analyzed in MEGA 12. The *Pt*CYS sequence (in bold) clustered with the cystatin domain from *H. syriacus* (*Hs*CYS-like), *C. sativus* (*Cs*CYS), three green microalgal species: *T. sp. A1-2* (*Ta*CYS), *R. subcapitata* (*Rs*CYS), and *K. nitens* (*Kn*CYS), as well as the terrestrial cyanobacterium *C. epipsammum* (*Ce*CYS). The tree was constructed with the Neighbour Joining method, with 1000 bootstrap iterations. Only bootstraps equal and higher than 50% are shown. B. The *Pt*CYS protein primary structure is represented, in the N-terminal, 22 aa-long secretory signal peptide (SP) and the mature cystatin domain (positions 37–115) highlighted (upper part of the panel). The amino acid residues in conserved motifs of the proteins are also highlighted, including the Gly-Gly (GG) motif of the N-terminal region and the Gln-X-Val-X-Gly (QxVxG) motif of the first inhibitory loop (bottom part). The orange background identifies aliphatic amino acids, the green and the red backgrounds identified amide and acidic amino acids, respectively. C. Cartoon representation of *Pt*CYS (*Pt*CYS, EG01374) (in gold) superimposed with wild taro, *Colocasia esculenta*, cystatin *Ce*CYS (in turquoise; PDB 3IMA: [52]). The N-terminal Gly-Gly motif is shown in pink, the 1st inhibitory loop (Gln-X-Val-X-Gly) in red, and the 2nd inhibitory loop in blue. D. Cartoon representation of the predicted folding of *P. tricornutum* papain-like protease candidate (*Pt*Pap, J4936.t1) with active site residues in yellow and catalytic residues Gln22, Cys31, His202, Asn224, and Trp226 shown as orange sticks. E. Predicted folding and interaction of *Pt*CYS (gold cartoon representation) with *Pt*Pap (grey surface with yellow active site) (left panel), and crystal structure of 3IMA *Ce*CYS (turquoise ribbon) interacting with *Carica papaya*, papain (PDB: 3IMA, grey surface) (right panel). *P. tricornutum* candidates' structure and interaction were modeled and docked with Alphafold3. Gly-Gly motif in the N-terminal loop is shown in pink, the Gln-X-Val-X-Gly motif-bearing hairpin loop in red, and the C terminal hairpin loop in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

was modeled using AlphaFold3 [65] (Fig. 1E, left panel), and compared to the crystallized structure of the tarocystatin (*Ce*CYS)-papain complex (PDB 3IMA) (Fig. 1E, right panel). In the reference crystal structure, tarocystatin interacts tightly with papain, forming a tripartite wedge composed of two loops that insert into the active site and the N-terminal trunk that interacts with amino acid residues at the surface of the enzyme (Fig. 1E) [66]. Docking analysis revealed *Pt*CYS bound similarly to the papain-like protease candidate, with the N-terminal trunk and two hairpin loops of *Pt*CYS interacting with the target enzyme's active site cleft. The inferred interaction involved key residues in the papain-like protease, including Asn21-His23, Tyr27-Trp32, Gly69, Ala70-His78, Glu108, Thr177-Tyr186, Thr200-Ile203, Asn224, Trp226-Gln231, and Glu260. These residues encompassed both the catalytic and supporting active site residues (Gln22, Cys31, His202, Asn224, and Trp226), suggesting conserved interactions. This docking analysis strongly suggested that *Pt*CYS could effectively interact with a papain-like protease, reinforcing its resemblance to other cystatins and suggesting its potential to inhibit Cys proteases through similar mechanisms.

3.3. *PtCYS subcellular localization and impact on YFP accumulation*

Building on the *in silico* analysis above, which suggested *Pt*CYS to be a plausible candidate for Cys protease inhibition, we aimed to experimentally assess the impact of overexpressing this inhibitory protein on the accumulation of a heterologous protein. YFP was used as a reporter protein, under the control of the 40SRP8 promoter, assuming the intensity of YFP fluorescence to reflect protein accumulation. *Pt*CYS, a mature version with no signal peptide, and the *Pt*CYS signal peptide (SP) alone were fused at the N terminus of YFP to generate strains SP*Pt*CYS-YFP, *Pt*CYS-YFP, and SP-YFP, respectively (Fig. 2A). To assess the subcellular localization of *Pt*CYS, extrachromosomal expression lines were visualized using a confocal microscope to acquire YFP fluorescence (yellow) and chlorophyll autofluorescence (shown in blue) (Fig. 2B). The YFP strain, producing soluble YFP, served as a positive control,

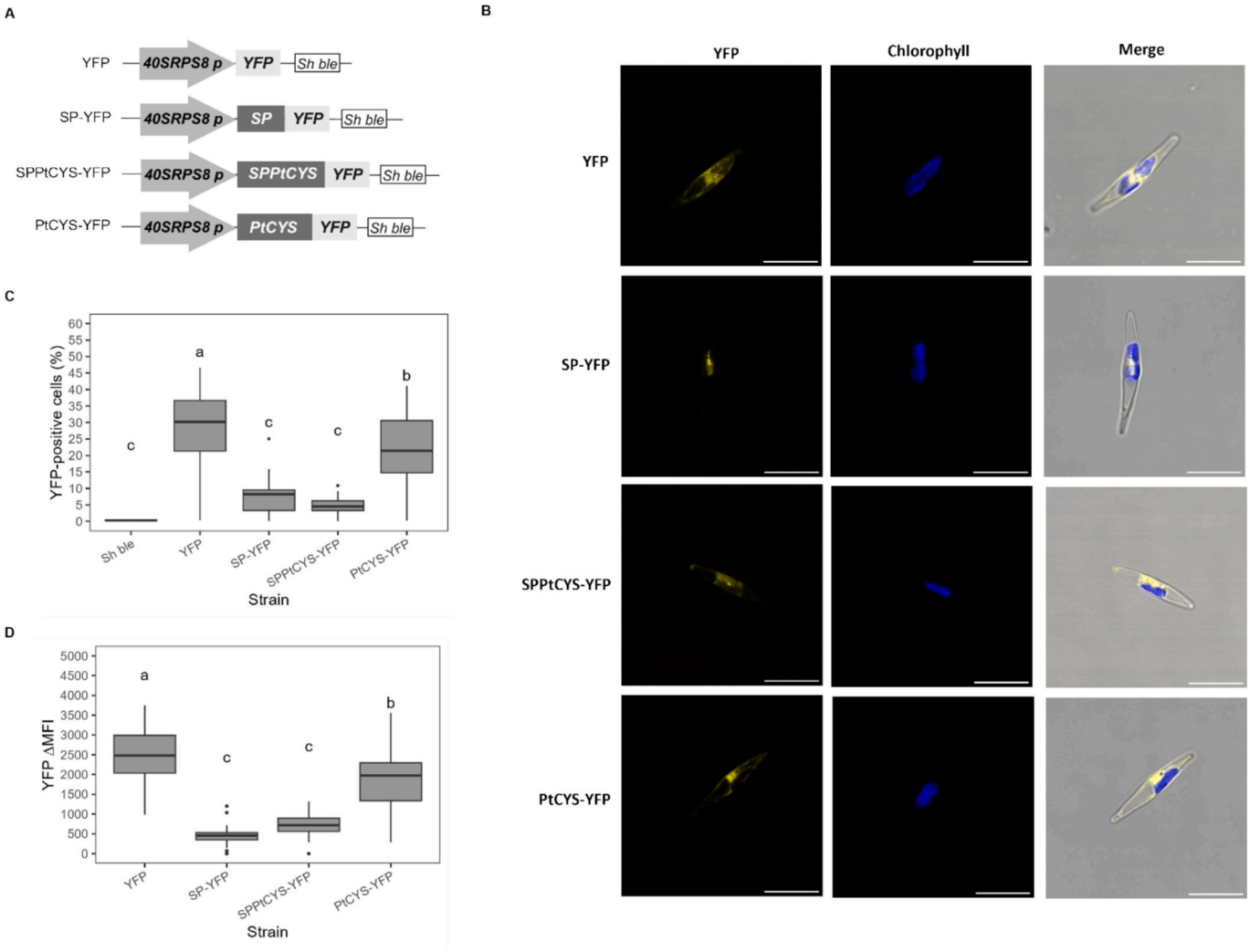


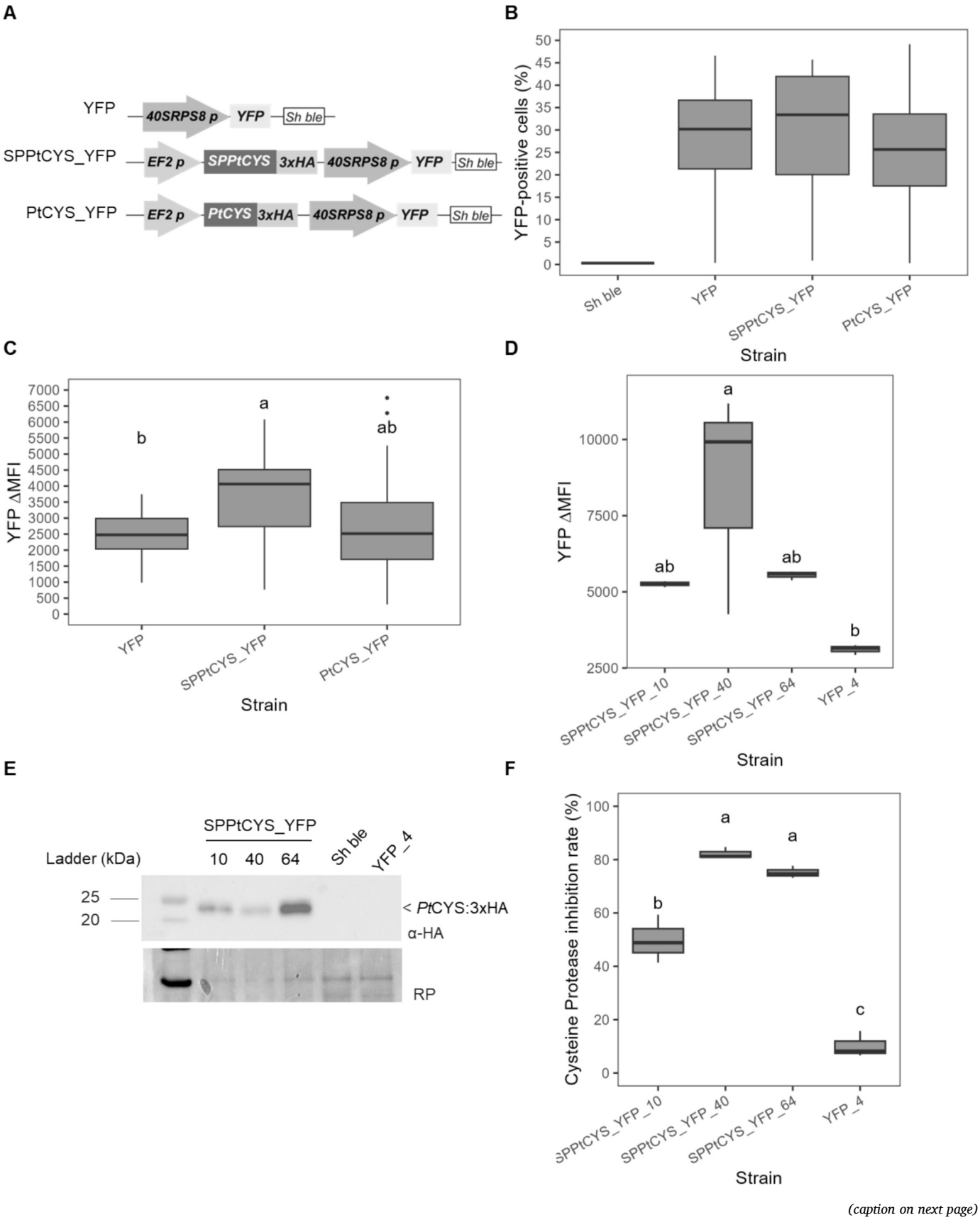
Fig. 2. *P. tricornutum* *Pt*CYS overexpressing strains. A. Schematic representation of the recombinant cassettes expressing YFP with different versions of the *Pt*CYS sequence fused at the N terminus: SP-YFP, SP*Pt*CYS-YFP, and *Pt*CYS-YFP driven by the *40SRPS8* promoter and harboring the *Sh ble* gene for Zeocin resistance, driven under the *fucoxanthin chlorophyll-binding protein C* (*fcpC*) promoter. This figure was created with BioRender.com [72]. B. SP*Pt*CYS-YFP, SP-YFP, and *Pt*CYS-YFP subcellular localization. YFP fluorescence, chlorophyll autofluorescence, and the merging of three fields are shown in transgenic lines producing SP*Pt*CYS-YFP, SP-YFP, *Pt*CYS-YFP, and soluble YFP were visualized by confocal laser microscopy. The Sh ble control strain was used to set parameters for avoiding cell autofluorescence. Scale bars = 10 μm. C. Box plot of the percentage of YFP positive cells was monitored for 31 independent lines of each construction, and all YFP positive cultures were normalized with the Sh ble autofluorescence strain. The fusion of SP*Pt*CYS or *Pt*CYS to YFP did not increase the percentage of YFP cells. D. Box plot of the mean fluorescence intensity (MFI) of YFP subtracting the background fluorescence from the Sh ble strain (ΔMFI) for each construction. The fusion of SP *Pt*CYS or *Pt*CYS to YFP did not increase YFP ΔMFI in the cells. Statistical comparisons were made using one-way ANOVA and Tukey's test. The same letters indicate no statistical differences between samples (p value < 0.01).

while the *sh ble* strain harboring the gene conferring Zeocin resistance, was employed to set reading parameters and minimize cell autofluorescence. YFP fluorescence in the YFP control strain was visible throughout the cytoplasm at day 4. Meanwhile, SP-YFP and SP*Pt*CYS-YFP are localized between the two plastid lobes in a blob-like structure near the chloroplast compartment, tightly possibly consisting of smaller reticular structures, a fluorescence pattern that has been reported in *P. tricornutum* for proteins that are translocated into the secretory pathway [8,67–70]. The soluble version of the enzyme, *Pt*CYS-YFP, behaved as a cytosolic protein, although it showed a more diffuse pattern than YFP-alone. Similar observations were made in mammalian cells transfected with a nonfused GFP plasmid, yielding larger amounts of GFP protein on Western blot, compared to GFP fusion with the HIV-1 envelope protein [71]. This difference could be attributed to protein misfolding, reduced stability, or cellular toxicity associated with the fusion protein. Overall, the signal peptide of PtCYS translocates YFP, in the SP-YFP cells, to the blob-like structures. This finding suggests that the *Pt*CYS SP leads to YFP entry into the secretory pathway (42), whereas the full-length cystatin fusion, SP*Pt*CYS-YFP accumulated YFP in both the endoplasmic reticulum and the cytosol. This localization pattern suggests that *Pt*CYS can be translocated to the secretory pathway, although this trafficking may be partial or inefficient under the conditions tested. To test whether *Pt*CYS could help increase YFP yield or accumulation, the cultures of thirty-one independent clones per construction were analyzed by flow cytometry, measuring both the percentage of YFP^+ cells (Fig. 2C) and the mean fluorescence intensity (MFI), which is relative to the amount of accumulated fluorescent YFP per cell (Fig. 2D). These results were compared to a control construct expressing YFP-alone (YFP) strains, and the background fluorescence from control strain Sh ble was subtracted from each construction (ΔMFI).

Interestingly, when YFP was fused to the *Pt*CYS signal peptide (SP-YFP) or to the immature preprotein (SP*Pt*CYS-YFP), there was a marked decrease in the percentage of YFP^+ cells, specifically, 7.3 ± 5.6% and 4.7 ± 2.9% of the cells in the respective populations exhibited fluorescence, compared with YFP-alone with 28.7 ± 12.4% ($p < 0.01$). The ΔMFI reached 369.9 ± 316.3 in the SP-YFP cell population and 648.7 ± 338.5 in the SP*Pt*CYS-YFP cells, ~4.7 and ~ 2.7-fold less intensity than the YFP-alone ΔMFI (2557 ± 820, $p < 0.01$) (Fig. 2C and D). There was no significant difference in the percentage of YFP+ cells between SP-YFP and SP*Pt*CYS-YFP. These results might suggest that the signal peptide destabilized secreted YFP during transit through the cell secretory pathway. This has been reported for GFP and other fluorescence proteins [43,73]. Although a correlation exists between fluorescence intensity and fluorescent protein accumulation, YFP fluorescence levels may be impacted by factors other than expression alone, which could for instance be triggered in the resulting translational fusion construct [48]. Fusing the mature version of *Pt*CYS to the N-terminus of YFP (*Pt*CYS-YFP) significantly increased YFP fluorescence (21.4 ± 13.4%; ΔMFI = 1748 ± 736) compared to SP-YFP and immature SP*Pt*CYS-YFP ($p < 0.01$; Fig. 2C and D), reinforcing the idea that the SP could contribute to reduced intracellular expression of YFP. However, the overall accumulation of YFP (ΔMFI) remained significantly lower than YFP alone ($p < 0.01$), supporting the observations from the confocal images (Fig. 2B). These results are consistent with studies showing that fluorescent proteins accumulate at higher levels when expressed alone than when fused to a target protein [71,74]. Likewise, when *Vanillia planifolia* vanillin synthase (*Vp*VAN) tagged with eGFP, was heterologous expressed in different *P. triconutum*'s compartments, *Vp*VAN recombinant protein was not detected, even though fluorescence was detected allowing confirmed subcellular localization. This suggested that the fusion protein was produced at a lower amount than the detection levels [75].

Overall, these findings indicated that fusing *Pt*CYS to YFP did not enhance heterologous protein accumulation in *P. tricornutum* cells, in contrast with results obtained using plant cystatins, such as tomato cystatin *Sl*CYS8 used as a fusion partner to promote human α1ACT yield in agroinfiltrated leaves of *N. benthamiana* [43]. As previously observed in *P. tricornutum* [2,8], the presence of a SP at the N-terminus of YFP negatively impacted fluorescence levels in transfected cells, possibly due to a deficient stability or folding of the reporter protein in the cell secretory pathway. Similarly, previous studies in diatoms have associated the cellular secretion of heterologous proteins with low protein yields [2,67] and protein degradation [8] compared to intracellular accumulation. The higher fluorescence observed for the YFP-alone construct suggested that fusing *Pt*CYS to YFP compromised its fluorescence. Based on these findings, in a second set of experiments, we expressed *Pt*CYS and YFP separately to mitigate the effect of the SP and allow the recombinant cystatin to act on a more stable YFP protein, potentially enhancing fluorescence. The complete and truncated coding sequences of *Pt*CYS were tagged with three Hemagglutinin (3xHA) in the C-terminal under the control of the *elongation factor 2* (*EF2*) promoter, while *yfp* expression was driven by the *40SRPS8* promoter sequence (Fig. 3A). The two constructs, SP*Pt*CYS_YFP (including the SP sequence) and *Pt*CYS_YFP (without SP), were tested, along with the YFP-alone construct serving as a positive control. In this setup, no significant difference was observed in the percentage of YFP^+ cells across the three conditions (30 ± 3.7% for YFP control population, 35 ± 2.9% for SP*Pt*CYS_YFP, and 25 ± 4.1% for PtCYS_YFP) (Fig. 3B). On the other hand, there was a significant increase in the ΔMFI of YFP in the SP*Pt*CYS_YFP population compared to the YFP population (ΔMFI = 2500 ± 289.8 for YFP *vs.* 4000 ± 377.3 for SP*Pt*CYS_YFP, $p < 0.01$), supporting the hypothesis that direct fusion to YFP might impair its fluorescence or accumulation, and that *Pt*CYS sent into the secretory pathway could stabilize YFP retained in the cytosol (Fig. 3C). The results suggested that *Pt*CYS can enhance YFP accumulation in *P. tricornutum*, but only when expressed independently and with its SP present. This contrasts with tomato cystatin *Sl*CYS8, which increased heterologous protein accumulation as a translational fusion partner but not when co-expressed separately in the cytosol [43]. The previous study, however, also highlighting the importance of the linker's sequence and length on cystatin activity, suggesting that future studies could help improve the effect of these elements of *Pt*CYS used as a fusion partner to the expressed protein. Finally, sequence and structural characteristics of *Pt*CYS could explain, in part, the differences observed compared to higher plant cystatins, in terms of inhibitory range and efficiency. For instance, slight variations in the N-terminal trunk and C-terminal may impact on inhibitory activity, efficiency, and specificity towards Cys proteases [55,76,77]. Indeed, site-directed mutagenesis in the C-terminal extension has demonstrated that the Asn is essential in the inhibition of legumain-like Cys proteases [61]. Further studies will be warranted to clarify the role of this important structural element in the mode of action of *Pt*CYS towards target Cys proteases.

To determine whether *Pt*CYS overexpression in the diatom cells could confer Cys protease inhibitory activity *in vivo*, 4-day old cultures of three independent SP*Pt*CYS_YFP strains showing a high YFP ΔMFI were selected (Fig. 3D) were compared with a YFP-expressing strain. The presence of *Pt*CYS:3xHA (23 kDa) was confirmed by immunoblotting in all SP*Pt*CYS_YFP strains (Fig. 3E). The protein extracts of the three SP*Pt*CYS_YFP strains were tested for the inhibition of *P. tricornutum* endogenous Cys proteases during the hydrolysis of the Z-Phe-Arg-methylcoumarin. As a result, all three clones showed higher inhibition of Cys protease activity; 50.2%, 82.1% and 75.1% for clones 10, 40 and 64, respectively compared with the YFP-alone strain (10.4%; $p < 0.01$) (Fig. 3F). Together, these observations confirmed that *Pt*CYS identified as a candidate cystatin in the *P. tricornutum* proteome database is indeed a functional cystatin. Although further analyses using purified protein are required to determine the kinetic parameters, these results confirm the potential of this cystatin as an effective accessory inhibitor to enhance recombinant protein accumulation in this organism, consistent with the higher Cys protease inhibitory activity observed in the *Pt*CYS-expressing strains.



(caption on next page)

Fig. 3. The co-expression of YFP with SP*Pt*CYS increases the mean fluorescence intensity (MFI) of YFP and *Pt*CYS is an active Cys protease inhibitor. A. Schematic representation of recombinant protein cassettes designed to express YFP under the *40SRPS8* promoter, along with two versions of *Pt*CYS under the expression of *elongation factor alpha 2* (*EF2*) and tagged in the C-terminal with three hemagglutinin sequences (3xHA). *Pt*CYS constructs either included a sequence for SP*Pt*CYS:3xHA (with the natural signal peptide), or for *Pt*CYS:3xHA (with no signal peptide). All constructs harbored the *Sh ble* gene driven under the *fucoxanthin chlorophyll-binding protein C* (*fcpC*) promoter, which confers Zeocin resistance. The figure was created with BioRender.com [72]. B. Box plot for the percentage of YFP positive cells monitored in 31 independent lines for each construct. All YFP positive cultures were normalized to the Sh ble autofluorescence strain. C. Box plot for the mean fluorescence intensity (MFI) of YFP, after subtracting the background fluorescence from the Sh ble strain (ΔMFI). The assays were reproduced three times, with 31 independent lines per gene construct. Different letters denote statistical differences between the strains, at a $p < 0.01$ (Post-ANOVA Tukey's test). D. Three selected CYS_YFP strains exhibiting a high YFP ΔMFI value. E. SP*Pt*CYS:3xHA (23 kDa) protein detected by western blotting (upper panel) using anti-hemagglutinin (α-HA) and or stained with Red Ponceau (RP) (lower panel). F. Three strains showing higher Cys protease inhibitory activity compared to the strain expressing YFP-alone. Data were normalized to the Sh ble strain. Data presented are the mean of three biological replicates ± SD. Different letters (a, b, c) indicate statistically different values between the strains at $p < 0.01$ (post-ANOVA Tukey's test).

3.4. *PtCYS overexpression rescues YFP fluorescence in strains with low YFP$^+$ cell percentage*

To test whether *Pt*CYS overexpression could rescue the YFP production in transconjugants exhibiting low YFP levels, three independent YFP clones showing a low YFP percentage (YFP_1, YFP_32 and YFP_36, ranges 0.19%, 3.56% and 0.13% of YFP$^+$ cells, respectively) and with the complete YFP expression cassette (*40SRPS8prom::YFP::FcpAter*, Fig. S1), were selected. The three clones were subjected to a new transformation round by conjugation with a second plasmid harboring SP*PtCYS:3xHA* and the *nat* gene conferring Nourseothricin resistance (Fig. 4A, upper panel), or with the empty vector (nat, Fig. 4A, lower panel). Then, 31 independent clones for each of the six *E. coli* conjugations: YFP_1_nat, YFP_1_SP*Pt*CYS_nat, YFP_32_nat, YFP_32_SP-*Pt*CYS_nat, YFP_36_nat and YFP_36_SP-*Pt*CYS_nat were analyzed by flow cytometry to measure the percentage of YFP$^+$ cells and the ΔMFI (Fig. 4B and C). The strain Sh ble_nat was used as a negative control to subtract the autofluorescence. Even though clones derived from YFP_1_SP*Pt*CYS_nat and YFP_32_SP*Pt*CYS_nat clones showed a slightly higher percentage of YFP$^+$ cells compared with their controls (YFP_1_nat and YFP_32_nat, respectively), the observed differences were not significant (Fig. 4B, $p > 0.01$). By comparison, the 31 clones derived from

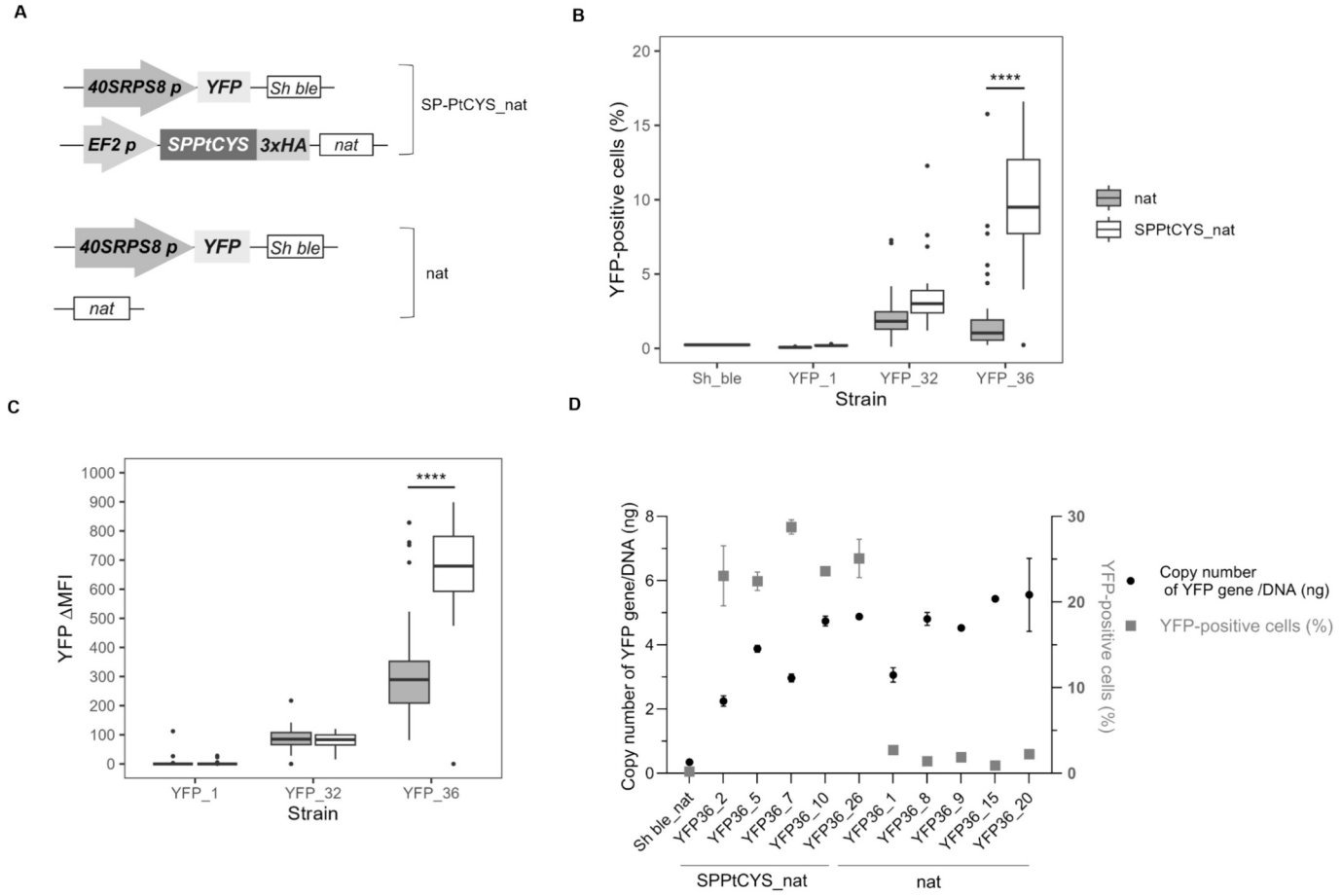


Fig. 4. YFP fluorescence recovery in cells with low YFP Δ mean fluorescence intensity (ΔMFI), by the introduction of SP*Pt*CYS:3xHA, SP*Pt*CYS tagged in the C-terminal with three hemagglutinin sequences. A. Schematic representation of the recombinant cassettes expressing YFP, and the introduction of the SP*PtCYS:3xHA* in a second episome harboring the *nat* gene for Nourseothricin resistance, driven under the *fucoxanthin chlorophyll-binding protein C* (*fcpC*) promoter, (SP*Pt*CYS_nat, upper panel) or the empty vector (nat, lower panel). The figure was created with BioRender.com [45]. B. Box plot for the YFP positive cells percentage of YFP strains 1, 32, 36 transformed with the nat (grey) or SP*Pt*CYS_nat (white) episomes. All cultures ($n = 31$) were normalized to the Sh ble_nat autofluorescence strain. C. Box plot of the mean fluorescence intensity (ΔMFI) of YFP. D. YFP copy numbers in transconjugants. Five YFP36_SP*Pt*CYS_nat clones with high YFP+ cell percentage and five YFP36_nat with low YFP+ cell percentage (grey boxes) were selected to check for the YFP gene copy number (black circles), normalized to total DNA. **** indicates significant difference between nat and CYS_nat clones (p-value < 0.01, two-way ANOVA followed by Tukey's test).

YFP_36_SP*Pt*CYS_nat showed a markedly increased percentage of YFP$^+$ cells compared with the 31 YFP_36_nat clones not expressing the cystatin (~4-fold increase, means of 9 ± 3% *vs* 2 ± 3% of YFP$^+$ cells respectively, $p < 0.0001$). YFP accumulation per cell was also significantly increased in YFP_36_SP*Pt*CYS_nat compared to the YFP_36_nat clones (~ 2-fold increase, MFI of 670 ± 170 *vs* 370 ± 248, respectively, $p < 0.0001$), following the same pattern (Fig. 4C). In these transconjugants, the results were highly reproducible, as 30 derived clones, out of 31, gained in YFP abundance at the cellular (MFI) and population (%) levels upon *Pt*CYS:3xHA overexpression.

To determine whether the differences in YFP percentages and ΔMFI between groups YFP_36_SP*Pt*CYS_nat and YFP_36_nat were due to the YFP gene copy number in the cells, five clones with a high percentage of YFP$^+$ cells (clone 2, 5, 7, 10 and 26) and five YFP_36_nat with a low percentage (clone 1, 8, 9, 15, 20) were selected and analyzed (Fig. 4D, right y-axis, shown in grey). The YFP gene copy number of the ten clones and the Sh ble_nat strain was normalized to the total DNA. As a result, neither YFP percentage nor intensity depended on the *YFP* gene copy number (Fig. 4D, left y-axis, shown in black), as the copy number did not correlate with the YFP percentage and intensity. Similar results were reported by *Díaz-Garza* et al., it was found no significant difference in plasmid copy number between populations with medium and low GFP fluorescence, whereas clones with high GFP fluorescence harbored higher plasmid copy numbers [22].

Overall, the observed increases of YFP intensity and percentage of YFP$^+$ cells in all strains derived from one of the three transconjugants suggest that the overexpression of *Pt*CYS could increase YFP accumulation in a genome background-dependent manner. YFP and other fluorescent proteins are not specific substrates of Cys proteases; however, indirectly, inhibition of endogenous Cys protease activity may reduce the degradation of misfolded or unstable recombinant proteins, thus increasing YFP accumulation. Several genetic and physiological factors could contribute to the variability between strains. First, independent transconjugants may differ in their basal protease repertoire and activation state, including papain-like cysteine proteases whose abundance is enriched in stress-associated morphotypes in *P. tricornutum* [28]. In such cases, *Pt*CYS would be expected to provide a greater benefit when Cys-protease-mediated degradation is a dominant limitation. Second, clone-to-clone differences in episome integrity and gene expression output, as well as post-transcriptional silencing mechanisms reported for episomal systems [21,22], could generate low-fluorescence phenotypes that are not primarily driven by proteolysis and therefore would not be rescued by cystatin expression. Third, variability in proteostasis capacity (*i.e.* folding, unfolded protein response, or autophagy) may shift recombinant proteins away from Cys proteases. Finally, compartmentalization is likely important: the requirement for the *Pt*CYS signal peptide in our co-expression system suggests that targeting the secretory pathway may be necessary to inhibit relevant proteases, and differences in trafficking efficiency between clones could modulate *Pt*CYS access to its targets.

Hence, this study highlights the potential of *Pt*CYS overexpression as a new strategy to improve the production of an episomal-encoded heterologous protein. The mode of action of *Pt*CYS, likely associated with the *in situ* inhibition of endogenous Cys protease, could also involve non-inhibitory effects, as illustrated for instance by the stabilizing effect of tomato *Sl*CYS8 on the overall structure of human α1ACT in *N. benthamiana* leaves [43]. However, our results also suggest that endogenous proteases would impede optimal YFP accumulation in some transconjugants, adding to other recently proposed silencing mechanisms [22] that could be targeted to improve metabolic engineering endeavors.

The function and activity of inhibitory cystatins in microalgae still need to be better understood. Further research is required to elucidate their role in the apoplast, plasma membrane, and the reorganization and polymerization of the cell wall, as well as in the modulation of the transition between *P. tricornutum* cell morphotypes. It also remains to be seen whether relative activities Cys protease and cystatin activities in this microalga are involved in processes such as cell-cell communication, population dynamics, or the formation of benthonic clusters.

Beyond molecular farming, diatoms are increasingly explored as sustainable sources of proteins and bioactive peptides for food and nutraceutical applications. Recent studies have highlighted the growing interest in alternative protein matrices and hydrolysates for functional foods and health-related applications [78–80]. In this context, strategies aimed at limiting intracellular proteolysis, such as cystatin overexpression, could prove relevant for improving the recovery of endogenous diatom proteins. Although the present study focused on heterologous protein accumulation, future work will be required to determine whether modulation of cysteine protease activity may impact total protein yield or peptide profiles in diatoms.

4. Conclusion

Over the years, the diatom *P. tricornutum* has been used as a sustainable heterologous protein production platform. However, several challenges remain, as some target proteins are difficult to express [8,75,81]. In other protein expression systems, such as bacteria or yeast, various strategies have been developed, including the engineering of strains better equipped to tolerate the expression of toxic proteins or more suitable for the cytoplasmic expression of disulfide bond–rich proteins [82], as well as optimization of culture media to reduce protease activity and oxidative stress. Here, we transfer knowledge and approaches previously developed in *N. benthamiana* [43] to a microalgal protein production platform. As the first study to characterize a diatom cystatin and explore its possible mode of action, this work provides a proof of concept for the stabilization of heterologous protein production in *P. tricornutum*, even though the model target used here, YFP, is not a direct substrate of cysteine proteases [36]. Our results show that overexpression of *Pt*CYS can induce pleiotropic effects in *P. tricornutum*, leading to increased or stabilized accumulation of valuable recombinant proteins in specific contexts. This study adds to growing knowledge on the functions of cystatins in photosynthetic organisms and their possible use to improve molecular engineering outcomes. Such advances are essential for fully deploying genetic engineering tools to express recombinant proteins with increased ease and control in new molecular farming hosts such as *P. tricornutum*.

CRediT authorship contribution statement

Elisa Fantino: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Natacha Mérindol:** Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Carolina Gabriela Gajón Robles:** Writing – review & editing, Formal analysis. **Fatima Awwad:** Writing – review & editing, Writing – original draft, Data curation, Conceptualization. **Karen Cristine Gonçalves dos Santos:** Writing – review & editing, Writing – original draft, Software, Data curation. **Kimy-Li Rhéaume:** Writing – review & editing, Formal analysis. **Marie-Claire Goulet:** Writing – review & editing, Conceptualization. **Dominique Michaud:** Writing – review & editing, Funding acquisition, Conceptualization. **Isabel Desgagné-Penix:** Writing – review & editing, Resources, 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 Quebec Research Fund Nature and Technology. Isabel Desgagne-Penix reports financial support was provided by Canada Research Chairs Program. Isabel Desgagne-Penix reports financial support was provided by Mitacs

Inc. 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 Centre SEVE – Concours Nouvelles Initiatives, volet Equipe Inter-institutionnelle to DM and IDP and by the Canada Research Chairs on plant specialized metabolism Award No CRC-2018-00137 to IDP. 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 CGGR from Mitacs - Globalink and to KCGS from Mitacs – Elevate program grant number IT28769 to IDP is also acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.algal.2026.104670.

Data availability

Data will be made available on request.

References

[1] A.A. Jamali, et al., Applications of diatoms as potential microalgae in nanobiotechnology, in: BioImpacts, 2012, pp. 83–89.
[2] S.S. Slattery, et al., Phosphate - regulated expression of the SARS - CoV - 2 receptor - binding domain in the diatom *Phaeodactylum tricornutum* for pandemic diagnostics, Sci. Rep. (2022) 1–15. Nature Publishing Group UK.
[3] S. D'Adamo, et al., Engineering the unicellular alga Phaeodactylum tricornutum for high-value plant triterpenoid production, Plant Biotechnol. J. (2019) 75–87.
[4] Fabris, M., et al., Extrachromosomal genetic engineering of the marine diatom Phaeodactylum tricornutum enables the heterologous production of monoterpenoids, in ACS Synth. Biol. 2020. p. 598–612.
[5] George, J., et al., Metabolic engineering strategies in diatoms reveal unique phenotypes and genetic configurations with implications for algal genetics and synthetic biology, in Front. Bioeng. Biotechnol. 2020. p. 1–19.
[6] F. Awwad, 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. 24 (23) (2023).
[7] Fantino, E.A., 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 Res., 2024.
[8] E. Fantino, et al., Extrachromosomal expression of functional Cannabis sativa cannabidiolic acid synthase in Phaedodactylum tricornutum, Algal Res. 85 (2025) 103889.
[9] F. Pruckner, et al., Remodeling of the terpenoid metabolism during prolonged phosphate depletion in the marine diatom Phaeodactylum tricornutum, J. Phycol. 61 (3) (2025) 512–528.
[10] N. Trevisan, et al., Metabolic engineering for improved heterologous pinene production in the chloroplast of Phaeodactylum tricornutum, Microb. Cell Factories 25 (1) (2026) 46.
[11] Sene, N.G.d.S., K.C.; Merindol, N.; Gélinas, S-E.; Custeau, A.; Awwad, F.; Fantino, E.; Meddeb-Mouelhi, F.; Germain, H. and Desgagné-Penix, I., Impact of heterologous expression of *Cannabis sativa* tetraketide synthase on *Phaeodactylum tricornutum* metabolic profile. Biotechnol. Biofuels Bioprod., 2025.
[12] Hao, X.H., F.; Pan, Y.; Yin, W.; Hu, H. , Heterologous phospholipid:diacylglycerol acyltransferase channels eicosapentaenoic acid to triacylglycerol in *Phaeodactylum tricornutum*. Algal Res., 2024.
[13] X. Liu, et al., Bioinformatic analysis and genetic engineering approaches for recombinant biopharmaceutical glycoproteins production in microalgae, Algal Res. 55 (2021) 102276.
[14] Zhou, C., et al., Structural and spectroscopic insights into fucoxanthin chlorophyll a/c-binding proteins of diatoms in diverse oligomeric states. Plant Commun., 2024. 5(11).
[15] Laukens, B., et al., Off-target glycans encountered along the synthetic biology route toward humanized N-glycans in *Pichia pastoris*. Biotechnol. Bioeng., 2020. 117(8): p. 2479–2488.
[16] B. Wolf, et al., Therapeutic antibody glycosylation impacts antigen recognition and immunogenicity, Immunology 166 (3) (2022) 380–407.
[17] F. Fernandes, et al., Flipase-mediated cassette exchange in Sf9 insect cells for stable gene expression, Biotechnol. Bioeng. 109 (11) (2012) 2836–2844.
[18] M.J. Gramer, Product quality considerations for mammalian cell culture process development and manufacturing, Adv. Biochem. Eng. Biotechnol. 139 (2014) 123–166.
[19] P.W. Barone, et al., Viral contamination in biologic manufacture and implications for emerging therapies, Nat. Biotechnol. 38 (5) (2020) 563–572.
[20] M.J. Barbosa, et al., Hypes, hopes, and the way forward for microalgal biotechnology, Trends Biotechnol. 41 (3) (2023) 452–471.
[21] A. Diamond, et al., Instability of extrachromosomal DNA transformed into the diatom Phaeodactylum tricornutum for heterologous vanillin production, Algal Res. 70 (2023) 102998.
[22] A.M. Diaz-Garza, et al., No two clones are alike: characterization of heterologous subpopulations in a transgenic cell line of the model diatom Phaeodactylum tricornutum, Microb. Cell Factories 23 (1) (2024) 286.
[23] A. Messaabi, et al., In vivo thrombin activity in the diatom Phaeodactylum tricornutum: biotechnological insights, Appl. Microbiol. Biotechnol. 108 (1) (2024) 481.
[24] W. Liu, et al., The papain-like cysteine protease HpXBCP3 from Haematococcus pluvialis involved in the regulation of growth, salt stress tolerance and chlorophyll synthesis in microalgae, Int. J. Mol. Sci. 22 (21) (2021).
[25] Y. Zou, P.V. Bozhkov, Chlamydomonas proteases: classification, phylogeny, and molecular mechanisms, J. Exp. Bot. 72 (22) (2021) 7680–7693.
[26] A.E. Ikui, et al., Control of pre-replicative complex during the division cycle in Chlamydomonas reinhardtii, PLoS Genet. 17 (4) (2021) e1009471.
[27] P. Sharma, D. Gayen, Plant protease as regulator and signaling molecule for enhancing environmental stress-tolerance, Plant Cell Rep. 40 (11) (2021) 2081–2095.
[28] C. Ovide, et al., Comparative in depth RNA sequencing of P. tricornutum's morphotypes reveals specific features of the oval morphotype, Sci. Rep. 8 (1) (2018) 14340.
[29] M. Novinec, B. Lenarcic, Papain-like peptidases: structure, function, and evolution, Biomol. Concepts 4 (3) (2013) 287–308.
[30] A. Schiermeyer, et al., Production of Desmodus rotundus salivary plasminogen activator alpha1 (DSPAalpha1) in tobacco is hampered by proteolysis, Biotechnol. Bioeng. 89 (7) (2005) 848–858.
[31] J.M. Watson, et al., RNA silencing platforms in plants, FEBS Lett. 579 (26) (2005) 5982–5987.
[32] C. Goulet, et al., A protease activity-depleted environment for heterologous proteins migrating towards the leaf cell apoplast, Plant Biotechnol. J. 10 (1) (2012) 83–94.
[33] D. Rivard, et al., An in-built proteinase inhibitor system for the protection of recombinant proteins recovered from transgenic plants, Plant Biotechnol. J. 4 (3) (2006) 359–368.
[34] C. Goulet, et al., A companion protease inhibitor for the protection of cytosol-targeted recombinant proteins in plants, Plant Biotechnol. J. 8 (2) (2010) 142–154.
[35] S. Komarnytsky, et al., Cosecretion of protease inhibitor stabilizes antibodies produced by plant roots, Plant Physiol. 141 (4) (2006) 1185–1193.
[36] K. Duwadi, et al., Identification, characterization and down-regulation of cysteine protease genes in tobacco for use in recombinant protein production, PLoS One 10 (7) (2015) e0130556.
[37] S. Afonso, L. Romagnano, B. Babiarz, The expression and function of cystatin C and cathepsin B and cathepsin L during mouse embryo implantation and placentation, Development 124 (17) (1997) 3415–3425.
[38] Henskens, Y.M., et al., Cystatins S and C in human whole saliva and in glandular salivas in periodontal health and disease. J. Dent. Res., 1994. 73(10): p. 1606–14.
[39] A.O. Grubb, Cystatin c—properties and use as diagnostic marker, Adv. Clin. Chem. 35 (2000) 63–99.
[40] S. Robert, et al., Protection of recombinant mammalian antibodies from development-dependent proteolysis in leaves of Nicotiana benthamiana, PLoS One 8 (7) (2013) e70203.
[41] P.V. Jutras, et al., An accessory protease inhibitor to increase the yield and quality of a tumour-targeting mAb in nicotiana benthamiana leaves, PLoS One (2016) 1–17.
[42] P. Pillay, et al., Use of transgenic oryzacystatin-I-expressing plants enhances recombinant protein production, Appl. Biochem. Biotechnol. 168 (6) (2012) 1608–1620.
[43] F. Sainsbury, et al., Tomato cystatin SlCYS8 as a stabilizing fusion partner for human serpin expression in plants, Plant Biotechnol. J. 11 (9) (2013) 1058–1068.
[44] Rastogi, A.M., U.; Dorrell, R.G.; Rocha Jimenez Vieira, F.; Maumus, F.; Kustka, A.; McCarthy, J.; Allen, A.E.; Kersey, P.; Bowler, C. & Tirichine, L. , Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. Sci. Rep., 2018.
[45] Camacho, C., et al., BLAST+: architecture and applications. BMC Bioinformatics, 2009. 10: p. 421.
[46] Girard, C., et al., A multicomponent, elicitor-inducible cystatin complex in tomato, *Solanum lycopersicum*. New Phytol., 2007. 173(4): p. 841–851.
[47] S. Kumar, et al., MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for adaptive and green computing, Mol. Biol. Evol. 41 (12) (2024).
[48] F. Teufel, et al., SignalP 6.0 predicts all five types of signal peptides using protein language models, Nat. Biotechnol. 40 (7) (2022) 1023–1025.
[49] M. Blum, et al., The InterPro protein families and domains database: 20 years on, Nucleic Acids Res. 49 (D1) (2021) D344–D354.
[50] L. Kall, A. Krogh, E.L. Sonnhammer, Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server, Nucleic Acids Res. 35 (Web Server issue) (2007) W429–W432.

[51] A.D. Yates, et al., Ensembl Genomes 2022: an expanding genome resource for non-vertebrates, Nucleic Acids Res. 50 (D1) (2022) D996–D1003.
[52] H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide Protein Data Bank, Nat. Struct. Biol. 10 (12) (2003) 980.
[53] T.K. Kassaw, A.J. Paton, G. Peers, Episome-based gene expression modulation platform in the model diatom phaeodactylum tricornutum, ACS Synth. Biol. (2022) 191–204. American Chemical Society.
[54] Karas, B.J., et al., Designer diatom episomes delivered by bacterial conjugation, in Nat. Commun.. 2015, Nature Publishing Group. p. 1–10.
[55] J. Tremblay, et al., Harnessing the functional diversity of plant cystatins to design inhibitor variants highly active against herbivorous arthropod digestive proteases, FEBS J. (2022) 1827–1841.
[56] M. Martinez, et al., Comparative phylogenetic analysis of cystatin gene families from arabidopsis, rice and barley, Mol. Gen. Genomics. 273 (5) (2005) 423–432.
[57] A. Augustine, et al., Multifactorial interaction and influence of culture conditions on yellow fluorescent protein production in Phaeodactylum tricornutum, Bioresour. Technol. 425 (2025) 132336.
[58] R. Beauchemin, et al., Successful reversal of transgene silencing in *Chlamydomonas reinhardtii*, Biotechnol. J. 19 (1) (2024) e2300232.
[59] A. Shamsi, B. Bano, Journey of cystatins from being mere thiol protease inhibitors to at heart of many pathological conditions, Int. J. Biol. Macromol. 102 (2017) 674–693.
[60] J. Wu, et al., Expression profile of cystatin gene family in alfalfa (Medicago sativa L.) related to biotic and abiotic stress response, BMC Genomics 26 (1) (2025) 987.
[61] M. Martinez, et al., Carboxy terminal extended phytocystatins are bifunctional inhibitors of papain and legumain cysteine proteinases, FEBS Lett. 581 (16) (2007) 2914–2918.
[62] M. Martinez, I. Diaz, The origin and evolution of plant cystatins and their target cysteine proteinases indicate a complex functional relationship, BMC Evol. Biol. 8 (2008) 198.
[63] M. Benchabane, et al., Plant cystatins, Biochimie 92 (11) (2010) 1657–1666.
[64] M. Tian, et al., A Phytophthora infestans cystatin-like protein targets a novel tomato papain-like apoplastic protease, Plant Physiol. 143 (1) (2007) 364–377.
[65] J. Abramson, et al., Accurate structure prediction of biomolecular interactions with AlphaFold 3, Nature 630 (8016) (2024) 493–500.
[66] W. Bode, et al., The 2.0 A X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases, EMBO J. 7 (8) (1988) 2593–2599.
[67] Erdene-ochir, E., et al., Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from *Phaeodactylum tricornutum*, in Sci. Rep. 2019, Springer US. p. 1–10.
[68] Hempel, F., et al., From hybridomas to a robust microalgal-based production platform: molecular design of a diatom secreting monoclonal antibodies directed against the Marburg virus nucleoprotein Microb. Cell Factories 2017, BioMed Central. p. 1–10.
[69] O. Kilian, P.G. Kroth, Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids, Plant J. (2005) 175–183.
[70] Apt, K.E., et al., In vivo characterization of diatom multipartite plastid targeting signals. J. Cell Sci., 2002. 115(Pt 21): p. 4061–9.
[71] L. Wu, M.A. Barry, Fusion protein vectors to increase protein production and evaluate the immunogenicity of genetic vaccines, Mol. Ther. 2 (3) (2000) 288–297.
[72] BioRender, BioRender, 2023.
[73] A.C. Fisher, M.P. DeLisa, Laboratory evolution of fast-folding green fluorescent protein using secretory pathway quality control, PLoS One 3 (6) (2008) e2351.
[74] S. Persson, et al., When a day makes a difference. Interpreting data from endoplasmic reticulum-targeted green fluorescent protein fusions in cells grown in suspension culture, Plant Physiol. 128 (2) (2002) 341–344.
[75] Diaz-Garza, A.M.L.-M., F.; Merindol, N.; Diamond, A. and Desgagné-Penix, I., Preprint. What is true for plants may not be true for Phaeodactylum tricornutum: the case of *Vanilla planifolia* vanillin synthase (VpVAN) targeted to four subcellular compartments of the diatom 2024, bioRxiv.
[76] J. Tremblay, M.C. Goulet, D. Michaud, Recombinant cystatins in plants, Biochimie 166 (2019) 184–193.
[77] F. Sainsbury, et al., Discrimination of differentially inhibited cysteine proteases by activity-based profiling using cystatin variants with tailored specificities, J. Proteome Res. 11 (12) (2012) 5983–5993.
[78] Zu, X.-y.H., Y-q.; Zhao, Y-j.; Xiong, G-q.; Liao, T.; Li H-I., Peptide extraction from silver carp (*Hypophthalmichthys molitrix*) scales via enzymatic hydrolysis and membrane filtration Ital. J. Food Sci., 2023.
[79] Singh, H.S., R.; Gupta, A.; Joshi, S.; Dar, B.N.; Singh, B.; Sharma, S., Characterization of jackfruit seed enriched pasta: product-functionality profile, secondary protein structures, bioactive composition and molecular morphology. Quality Assurance and Safety of Crops & Foods, 2023.
[80] Park, J.H.L., S.I.; Kwon, W.S; Cho, S.C.; Kim, I.H., Health benefits of co-supplementing mealworm protein hydrolysate and cranberry fruit extract Ital. J. Food Sci., 2023.
[81] Sene, N.L., B.; Gélinas, S-E.; Custeau, A.; Merindol, N.; Meddeb-Mouelhi, F.; Desgagné-Penix, I., Production of *Dictyostelium discoideum* hybrid type enzyme SteelyA in the diatom *Phaeodactylum tricornutum* Appl. Sci., 2025.
[82] Bertelsen, A.B., et al., Discotune: versatile auxiliary plasmids for the production of disulphide-containing proteins and peptides in the *E. coli* T7 system. Microb. Biotechnol., 2021. 14(6): p. 2566–2580.