Microalgal Enzymes with Biotechnological Application

Thesis

How to cite:


For guidance on citations see FAQs.

© 2022 Giorgio Maria Vingiani

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Version of Record

Link(s) to article on publisher’s website:

http://dx.doi.org/doi:10.21954/ou.ro.00015444

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.

oro.open.ac.uk
Microalgal enzymes with biotechnological application

The Open University Ph.D. Course - XX Cycle (2018 - 2021)

Stazione Zoologica Anton Dohrn

Ecosustainable Marine Biotechnology Department

Naples, Italy

Doctor of Philosophy

School of Life, Health and Chemical Sciences

Giorgio Maria Vingiani (M.Sc.)

Director of studies: Dr. Chiara Lauritano

Stazione Zoologica Anton Dohrn

Internal supervisors: Dr. Pasquale De Luca
Dr. Adrianna Ianora

External supervisor: Prof. Alan D.W. Dobson
University College Cork

May 2022
Thesis Abstract

The aim of this thesis was to address the current state of art of microalgal enzymes with biotechnological application, and to expand the current knowledge of enzymes with the potential to find proper application in industry or novel market niches, via their production and biochemical characterization. Available transcriptomic datasets were used to identify possible enzymes of interest. Specifically, after a review of the available scientific literature (Chapter 1) my experimental activities focused on different enzymes identified in the transcriptome of the marine diatom *Cylindrotheca closterium*.

Studied enzymes were two di-n-butyl phthalate hydrolases (CcDBPH1-2), putatively involved in phthalate degradation (Chapter 2), two platelet activating factor- acetylhydrolase (CcPAF-AH1-2), involved in lipid metabolism and inflammation modulation, which role in diatom is still unknown (Chapter 3), a L-asparaginase (CcASNase), that is an enzyme with a well-supported market niche in cancer treatment and acrylamide mitigation (Chapter 4).

The experimental pipeline of CcDBPH1-2, starting from transcript identification till protein functional activity studies, was completed and reached the stage of protein purification and activity assay. In addition, different bioinformatics analyses and *cc_dbph1-2* expression by RT-qPCR in *C. closterium* grown with different concentrations of DBP were also performed confirming evidences of the protein bioremediation activity.

The expression of CcPAF-AH1-2 was hampered by the strong toxic effect that recombinant proteins production induced in *E. coli* cells. However, the mutation of the catalytic serine 243 of CcPAF-AH2 allowed enzyme accumulation in 1L cultures.

Finally, the expression of CcASNase was successful in BL21 (DE3) and Rosetta 2 (DE3) *E. coli* strains. However, the low amount of produced protein will need further optimization of the growth conditions.

Overall, this PhD thesis represents a contribution to the growing number of studies aimed to unravel the application of microalgae as an eco-friendly and eco-sustainable source of novel enzymes for human applications.
“Let me have a look.”

Woland stretched out his hand palm uppermost.

“Unfortunately I cannot show it to you,” replied the master, ‘because I burned it in my stove.’

‘I’m sorry but I don’t believe you,’ said Woland.

“You can’t have done. Manuscripts don’t burn.”

— Mikhail Bulgakov, *The Master and Margarita*

“Jasper Gwyn mi ha insegnato che non siamo personaggi, siamo storie, disse Rebecca. Ci fermiamo all’idea di essere un personaggio impegnato in chissà quale avventura, anche semplicissima, ma quel che dovremmo capire è che noi siamo tutta la storia, non solo quel personaggio. Siamo il bosco dove cammina, il cattivo che lo frega, il casino che c’è attorno, tutta la gente che passa, il colore delle cose, i rumori.”

“Jasper Gwyn taught me that we are not the character, but the story, Rebecca said. We stick to the idea of being a character committed to god-knows-what adventure, even a plain one, while we should grasp that we are the entire story, not only that character. We are the woods where he gets lost or the villain that tricks him. We are the mess, the people passing, the colours around him. We are the noises.”

— Alessandro Baricco, *Mr. Gwyn*

“There is no unemployment in eco-systems.”

— Gunter Pauli
Acknowledgements

It is not easy to acknowledge all the people that, in [one] way or another, helped this four-years-long experience to be shaped the way it turned out in the end. I will try to list as many people as my oh-so-flawed memory allows me, asking for forgiveness if I forget someone. You *know* you meant something in this journey.

First of all, I would like to thank Dr. Chiara Lauritano, my Director of Studies. I am grateful (and feel sorry for her!) for having been her first “official” PhD Student. I hope to have given her back at least a fraction of all she gave me in terms of compassion, humanity and scientific competence. Thank you for having taught me about discipline, commitment and perseverance. You are one of the kindest human being I ever met, and I wish you the best.

To Dr. Pasquale De Luca, my constant guide and the most charismatic boss the CSAM lab could ever have: your expertise, kindness and almost infinite patience have been the lighthouse of my PhD. Surely without you I wouldn’t have been able to arrive at the end of this journey. You know very well how much you did for me, and how much I made you endure my protean behaviour. I feel sorry for how many times you had to rescue me from my worst shape, and I hope to have made you proud sometimes.

I thank Dr. Adrianna Ianora, my Internal Supervisor, for her constant interest in my project and for always being close to me (even when distant) and having warmly made me feel as one of her students.

I thank Prof. Alan D.W. Dobson, my External Supervisor, for his far away guidance and advice, that was always precious and gladly welcomed.

While not officially part of my PhD panel, Dr. Serena Leone was the most unexpected and vital addition to my mentors, teaching me everything I know (not that much, actually) about protein biochemistry and structural biology, and enriching my thesis with tons of cool stuff. I feel really lucky to have met such a great scientist and friend, and her companionship and wisdom keeps enlightening me every day.

Thank you, Elvira. You were like a second mother, and I will never forget your constant support (and food providing) during these years. Nothing could be the same at CSAM without you. I am sorry for every excel sequence I sent you too late, and thank you for having made them nonetheless.

Thank you, Raimondo, you are the perfect counterpart of our CSAM ‘mommy’: ironic, collected and simply based. Among all the stuff I have to thank you for, thank you for waiting for me when ordering the primers oh so many times.
I thank all the colleagues and friends who have followed this path with me and shared a laugh, a drink and a kind word. You were the building blocks of these years.

Thank you, Elio, Monia, Pina, Ivan and Angela. You were always eager to give me your ear to dispel any doubt about my scientific path. Every one of you was a big milestone towards the end of this path.

Thank you, Kevin, for all the moments and time we shared together, inside and outside the lab. Thank you, Filippo, for the insights of clarity you gifted me, the endless coffees and the talks of “men of culture”. Thank you, Federica, for having found together a way to move through the bad days. Thank you, Maria and Vittoria, for our ‘musical’ first days. Thank you, Viviana, Antonella, Maria, Anna, Francesco and Annalisa, for all the coffees, food, celebrations and laughs we shared together. Thank you, Carola, for having shared with me the heavy ‘crown’ of PhD representatives and many moments together. Thank you, Lorena, Hung and Louise for having moved together the first steps of this journey.

Thank you Gennaro, Susy and Anita for your mentorship, the endless coffees, the laughs... your example is the best I could ever follow for my career and my life. Last met, thank you Laura for your empathy and kindness. You were the right person at the right moment, and I hope to be able to reciprocate if you ever need to.

Lastly, I thank my mentor and friend from the University of Alcalá de Henares, Prof. Leonardo Casano: thank you for having sparked [in me] the passion for microalgae and molecular biology during my year in Spain, and for the affection you showed me in the years thereafter. Ours is a bond that transcends the first year we spent together, and that’s a beautiful thing.

I feel incredibly lucky to be surrounded by many friends, and I haven’t experienced the companionship of half of you half as well as I should like, and I like less than half of you half as well as you deserve. You really are too many to receive a kind phrase each, but please understand that it is for mere text space reasons, and not because I am lazy. Probably.

Thank you, Mary for being exactly like the sister I never had. Wherever your hopes and dreams bring you, you will find me always at your side.

Instead, thank you Marzio, for being as far as possible from my idea of a ‘big brother’. And that’s perfect nonetheless.

To my Kings, my Based, my best friends: Macera, the fiend, the unerring machine; you all know I cannot say more or I’d be arrested. Roberto, the “Did your therapist tell you to say this?”. Dario, the “Dario is SPEAKING, everybody shut up.”. Accardo the perfect ‘cucciolo’/‘iastemm/cringe’ combination. Marzio, again, the “Am I the only one to hear Noir music all of a sudden?".
Your companionship and friendship shaped me during these years. May I enjoy your love for a hundred years more.

A mention aside for my ‘Thanatolia’ players: Cozzolino, Manu, Jacopo, Simone, Macera and Marzio. We could very well have made the best RPG campaign anyone ever could, and our time together was one of the funniest of my life.

To my friends from Naples, from Castellammare, from my second family of Soliptica. To Alessia, Alfredo, Andrea, Angelo, Anna, Antonella, Bagongo, Bolognese, Chiara, Clelia, Fabiana, Filippo, Frank, Lo Schiavo F., Lo Schiavo M., Gianmarco, Isa, Laura, Lex, Lina, Luca, Luisa, Jolanda, Mario, Nunzio, Pacilio, Paola, Rossella, Stefano, Xander. Be it small or big, you all shaped a piece of my being. Thank you.

To my parents, that I have never felt so distant and at the same time so close like in the last three years. While I was taking the first wobbling steps in adulthood, I always found you by my side, ready to support, encourage and advise me. In Neapolitan we say “tené a’ magnatoria vascia”, having a low feeder, to be privileged and not having to fight hard to achieve something. That, on the other hand, also means having a kind environment that forgives your mistakes and is capable of putting you back on the right track. I know I have fruitfully re-invested at least one tenth of what you gave me. I can only hope among the many mistakes and false starts you attended to there is something to be proud of, and big enough to not be afraid for my future.

To you. You know how much you meant. I can only try to make you proud, knowing that will never be enough to reciprocate the time and energies you spent making me a better human being and helping me find the strength to finally love myself. But, nonetheless, I am confident it will create something somehow beautiful, in shapes and colours unfathomable. And that will “correct the world”, one gesture after the other.

To Chloe.

This journey and whatever will bloom from it is dedicated to you, and is mine as much as is of you all.
List of original papers included in this Thesis and author contribution

Chapter 1: Part of the review work presented (updated to 2019) in this chapter has been published previously:


Author contributions: G.M.V., P.D.L., A.I., A.D.W.D. and C.L. co-wrote the review.

Chapter 2: Results included in Paragraphs 2.3.1, 2.3.2, 2.3.3, 2.3.6 and related Methods have been published previously:


List of other publications during the candidature


Table of Contents

Chapter I
1.1 Background
  1.1.1 Microalgal enzymes for biotechnological applications 2
  1.1.2 Microalgal enzymes for high-value added lipids and biodiesel production 5
  1.1.3 Microalgal enzymes for healthcare applications 11
  1.1.4 Microalgal enzymes for bioremediation 15
1.2 Premises and objectives of the PhD project 18

Chapter 1 References: 22

Chapter II

Chapter 2 Abstract
2.1 Introduction
  2.1.1 The diatom Cylindrotheca closterium 35
  2.1.2 Phthalate esters: classification and posed dangers 37
  2.1.3 The PEs biodegradation and previous findings and applications of PEs-bioremediation enzymes 39
2.2 Methods
  2.2.1 C. closterium strain specifics, culturing and harvesting 43
  2.2.2 Bioinformatic mining of CcDBPH putative transcripts 44
  2.2.3 CcDBPHs modeling and docking simulations 44
  2.2.4 cc_dbph1-2 oligo design and reaction procedures for PCR and RT-qPCR 45
  2.2.5 cc_dbph12 heterologous expression, SDS-PAGE and purification 47
  2.2.6 Esterase activity assay 49
  2.2.7 DBP-degradation activity assay and measurement via HPLC 49
2.3 Results and discussion
  2.3.1 C. closterium (FE2) culture with DBP and growth data 50
  2.3.2 Sequences coding enzymes involved in DBP degradation response 51
  2.3.3 DBPHs catalytic site identification and 3D structure modelling 57
  2.3.4 PCR Isolation of di-n-butyl phthalate hydrolase (DBPH) coding sequences and purification 61
  2.3.5 Cloning, sequencing of DBPH in TOPO-TA vectors and intron retention resolution 64
  2.3.6 RT-qPCR of FE2 + DBP samples and data analysis 68
  2.3.7 Cloning and sequencing of dbph1-2 in pOPIN vectors 71
  2.3.8 Heterologous expression of cc_dbph1-cc_dbph2//pOPIN 74
  2.3.9 Enzyme purification of DBPH-pOPIN proteins 78
  2.3.10 Estrase and DBP-degrading assays 81
2.4 Conclusions 84

Chapter 2 References: 87
Chapter III

Chapter 3  Abstract

3.1  Introduction

3.1.1  Inflammation and the role of PAF in the inflammatory response
3.1.2  Microalgal anti-inflammatory activity
3.1.3  Previous evidences of platelet activating factor-acetylhydrolase (PAF-AH) enzymes

3.2  Methods

3.2.1  cc_paf-ah1-2 oligo design and reaction procedures for PCR
3.2.2  CcPAF-AH1-2 heterologous expression and SDS-PAGE
3.2.3  CcPAF-AH2 purification attempts
3.2.4  3.2.4 CcPAF-AH2 catalytic serine mutation

3.3  Results and discussion

3.3.1  Sequences coding PAF-AH enzymes
3.3.2  CcPAF-AHs sequence alignment and catalytic site identification
3.3.3  PCR Isolation of cc_paf-ah1 and cc_paf-ah2 coding sequences and purification
3.3.4  Cloning and sequencing of cc_paf-ah1-2 in pOPIN vectors
3.3.5  Heterologous expression of CcPAF-AH/pOPIN proteins
3.3.6  Cloning and sequencing of cc_paf-ah1-2 in pJ411 vectors
3.3.7  Heterologous expression of CcPAF-AH1-2/pJ411 proteins
3.3.8  Troubleshooting with CcPAF-AH2 expression and purification

3.4  Conclusions

Chapter 3 References:

Chapter IV

Chapter 4  Abstract

4.1  Introduction

4.1.1  Summary of L-asparaginase (ASNase) characteristics and its therapeutic applications
4.1.2  Previous applications of L-asparaginase (ASNase) enzymes in food treatment

4.2  Methods

4.2.1  CcAsnase oligo design and reaction procedures for PCR
4.2.2  CcASNase heterologous expression and SDS-PAGE

4.3  Results and discussion

4.3.1  Sequences coding CcASNase enzyme and catalytic site identification
4.3.2  PCR Isolation of cc_asnase coding sequence and cloning in TOPO-TA vector
4.3.3  Cloning and sequencing of cc_asnase in pJ411 vectors
4.3.4  Heterologous expression of CcASNase/pJ411 proteins

4.4  Conclusions

Chapter 4 References:
Chapter V

Chapter 5  Final Conclusions  171

Chapter 5 References:  175

Appendix M  176

M1. Microalgae inoculation, growth, scale-up and harvesting processes  177
M2. Evaluation of microalgae growth curves  179
M3. RNA extraction from microalgal samples  179
M4. cDNA production  180
M5. Primer designing for endpoint polymerase chain reaction (PCR) and reverse transcription-
quantitative PCR (RT-qPCR)  181
M6. PCR reaction and electrophoretic separation  183
M7. Purification of PCR products and sequencing  184
M8. RT-qPCR reactions and data analysis  184
M9. Cloning in TOPO® vectors  185
M10. Restriction enzyme reactions  186
M11. Preparation of vectors and PCR fragments for homologous recombination-based cloning  187
M12. Cloning in expression vectors  190
M13. Protein Heterologous Expression  191
M14. Protein Extraction and Sodium Dodecyl Sulphate– Polyacrylamide Gel Electrophoresis (SDS-PAGE)  192
M15. Western-Blot Analysis  193
M16. Bioinformatic analyses and databases  194
M17. Chemicals and products  195
List of Figures

Chapter I

Figure 1.1 “Microalgae Biotechnology” PubMed Search Results (modified from Vingiani, 2019) 4
Figure 1.2 Examples of fatty acids of biotechnological interest (from Vingiani et al., 2019) 6
Figure 1.3 Main studied and engineered enzymes for TAGs and PUFAs in microalgae for the production of high value-added lipids (modified from Vingiani, 2019) 9
Figure 1.4 Enzymes for healthcare applications (from Vingiani, 2019) 11
Figure 1.5 Enzymes for bioremediation (from Vingiani, 2019) 17
Figure 1.6 PhD final project structure 19

Chapter II

Figure 2.1 Photographs of C. closterium 35
Figure 2.2 Common structure of PEs and most common PEs (from Ren et al., 2018) 38
Figure 2.3 Representation of hydrolysis of phthalate esters (PEs) by phthalate esterase to phthalic acid via the formation of phthalate monoester intermediate (from Huang et al., 2013) 40
Figure 2.4 DNA sequence of the 18S sequence of strain FE2 of Cylindrotheca closterium 43
Figure 2.5 C. closterium growth curve in control and DBP+ conditions 50
Figure 2.6 DNA sequence of the transcript TR3123|c0_g1_i1 len=2609, where the two new PE esterases (provisionally named DBPH1 and DBPH2) were found 56
Figure 2.7 Protein sequence alignment of the newfound CcDBPH1 and CcDBPH2 (boxed in red) with other Class IV bacterial dialkyl esterases (from Vingiani, 2022) 58
Figure 2.8 Full length, refined homology models of CcDBPH1 (left) and CcDBPH2 (right) coloured according to their Verify3D score (from Vingiani et al., 2022) 59
Figure 2.9 Structural detail (A) and 2D map (B) of the interaction between DBP and CcDBPH1 (from Vingiani et al., 2022) 60
Figure 2.10 Agarose gel of cc_dbph1 CDS amplified with different primer couples (F1R2 and F1R1) and temperatures 61
Figure 2.11 Agarose gel of cc_dbph1 F1R2 CDS purified and quantified via indirect quantification with λ-2 ladder 62
Figure 2.12 Agarose gel of cc_dbph2 CDS amplified with different primer couples (F1R2, F2R2, F2R3, F1R3 and F2R1) at 56°C 63
Figure 2.13 Agarose gel of dbph2 F2R3, F1R3 and F2R1 CDS purified and quantified via indirect quantification with λ-2 ladder 64
Figure 2.14 CcDBPH1 CDS alignment from transcriptome and newly sequenced, as visualized via Jalview. In light blue are the areas of non-homology 65
Figure 2.15 DBPH2 CDS alignment from transcriptome and newly sequenced, as visualized via Jalview. In light blue are the areas of non-homology 65
Figure 2.16 Agarose gel at 2% of cc_dbph1-2 CDS internal fragments using respectively a newly produced cDNA treated with gDNAse (NEW) and a previous DNA preparation positive for the intron presence (OLD) 66
Figure 2.17 CcDBPH1 CDS and translated aminoacid sequence alignment from transcriptome and newly intron-less sequenced, as visualized via Jalview. In light blue are the areas of non-homology.

Figure 2.18 CcDBPH2 CDS and translated aminoacid sequence alignment from transcriptome and newly intron-less sequenced, as visualized via Jalview. In light blue are the areas of non-homology.

Figure 2.19 Reference gene assessment for C. closterium (from Vingiani et al., 2022).

Figure 2.20 Expression levels of cc_dbph1 and cc_dbph2 in C. closterium cells cultured with DBP exposure for 24 h and 48 h at 5 and 10 mg/L DBP compared to the control condition (from Vingiani et al., 2022).

Figure 2.21 Plasmid maps of cc_dbph1/pOPINE-F and cc_dbph2/pOPINE-F produced using SnapGene. A. cc_dbph1/pOPINF; B. cc_dbph1/POPINE; C. cc_dbph2/POPINF; D. cc_dbph2/POPINE.

Figure 2.22 SDS-PAGE of soluble and insoluble fractions of protein extracts of Lemo21 (DE3) cells transformed with cc_dbph1/pOPINE-F and cc_dbph2/pOPINE-F.

Figure 2.23 SDS-PAGE of soluble fractions of protein extracts of Rosetta 2 (DE3) and BL21 (DE3) cells transformed with cc_dbph1/pOPINE-F and cc_dbph2/pOPINE-F.

Figure 2.24 SDS-PAGE of insoluble fractions of protein extracts of Rosetta 2 (DE3) and BL21 (DE3) cells transformed with cc_dbph1/pOPINE-F and cc_dbph2/pOPINE-F.

Figure 2.25 ECL Chemoluminescent resolution of Western-Blot transfer of CcDBPH1 and CcDBPH2. A. Western-Blot of the CcDBPH1/pOPINE and CcDBPH2/pOPINE soluble protein extracts from the expression screening performed in Rosetta 2 (DE3) and BL21 (DE3) cells.

Figure 2.26 SDS-PAGE of the different steps of protein purification of CcDBPH1-2. A. CcDBPH1 purification from long term-induced Rosetta 2(DE3) pellets.

Figure 2.27 p-NPB degradation of CcDBPH1 and CcDBPH2.

Figure 2.28 Plasmid map of cc_dbph1-2/pOPINF and cc_dbph2/pOPINE-F produced using SnapGene.

Chapter III

Figure 3.1 Structure of platelet-activating factor (PAF).

Figure 3.2 DNA sequence of the transcript TR16369|c0_g1_i1 length: 1566 nt, where the longest newfound PAF-AH (prov. named CcPAF-AH1) was found.

Figure 3.3 DNA sequence of the transcript TR12571|c0_g1_i1 length: 1460 nt, where the shortest newfound PAF-AH (prov. named CcPAF-AH2) was found.

Figure 3.4 Protein sequence alignment of C. closterium PAF-AH1-2 and other PAF-AH II enzymes.

Figure 3.5 Agarose gel of cc_paf-ah1-2 CDS amplified with different primer pairs.

Figure 3.6 Agarose gel of cc_paf-ah1 CDS with pOPINE-F adaptors amplified and with quantification with λ-2 ladder.

Figure 3.7 Agarose gel of cc_paf-ah2 CDS with pOPINE-F adaptors amplified.

Figure 3.8 cc_paf-ah1 CDS alignment from transcriptome and newly sequenced, as visualized via Jalview.

Figure 3.9 cc_paf-ah2 CDS alignment from transcriptome and newly sequenced, as visualized via Jalview.

Figure 3.10 cc_paf-ah2 protein sequence alignment from transcriptome and newly sequenced, as visualized via Jalview. In light blue are the areas of non-homology.

Figure 3.11 Plasmid maps of cc_paf-ah1/pOPINE-F-S3C and cc_paf-ah2/pOPINE-F-S3C produced using SnapGene. A. cc_paf-ah1/POPINF; B. cc_paf-ah1/POPINE; C. cc_paf-ah1/POPINS3C.
Figure 3.12 SDS-PAGE of soluble fractions of protein extracts of Rosetta 2 (DE3) pLySs cells transformed with cc_paf-ah1/pOPINE-F and cc_paf-ah2/pOPINE-F

Figure 3.13 SDS-PAGE of insoluble fractions of protein extracts of Rosetta 2 (DE3) pLySs cells transformed with cc_paf-ah1/pOPINE-F and cc_paf-ah2/pOPINE-F

Figure 3.14 Plasmid maps of cc_paf-ah1-2/pJ411_NTER-CTER produced using SnapGene

Figure 3.15 SDS-PAGE of soluble and insoluble fractions of protein extracts

Figure 3.16 LB Agar plates of SHuffle cells transformed with A. cc_paf-ah2/pJ411_NTER and B. with empty pJ411_NTER


Figure 3.18 Alignment of the peptide sequences extracted from SDS-PAGE protein bands with CcPAF-AH2 aminoacid sequence on the Mascot search engine

Figure 3.19 Graphical summary of the mode of action of recombinant CcPAF-AH2 and S243A mutated CcPAF-AH2 in E. coli host

Chapter IV

Figure 4.1 DNA sequence of the transcript TR25145|c0_g1_i1 len=1778 nt, where the newfound ASNase (prov. named CcASNase) was found

Figure 4.2 Protein sequence alignment of the newfound CcASNase (boxed in red) with other other type II and type I l-asparaginases

Figure 4.3 Preliminary homology model via SWISS-MODEL of CcASNase shown as a crescent-shaped dimer

Figure 4.4 Agarose gel of cc_asnas CDS amplified with different primer couples (F1R1, F2R2, F1R2, F2R1)

Figure 4.5 cc_asnase CDS alignment from transcriptome and newly sequenced

Figure 4.6 Agarose gel of cc_asnase CDS amplified with different adaptors for homologous recombination-based cloning

Figure 4.7 Plasmid maps of cc_asnase/pJ411_NTER-CTER produced using SnapGene

Figure 4.8 SDS-PAGE of soluble fractions of protein extracts of Lemo 21, BL21 (DE3) and Rosetta 2 (DE3) cells transformed with cc_asnase/pJ411_NTER and cc_asnase/pJ411_CTER

Figure 4.9 SDS-PAGE of insoluble fractions of protein extracts of Lemo 21, BL21 (DE3) and Rosetta 2 (DE3) cells transformed with cc_asnase/pJ411_NTER and cc_asnase/pJ411_CTER

Figure 4.10 ECL Chemoluminescent resolution of Western-Blot transfer of CcASNase/pJ411_NTER-CTER soluble protein extracts from the expression screening done in BL21 (DE3) cells

Figure 4.11 Plasmid maps of cc_asnase/pOPINS3C-TRX-HALO7 produced using SnapGene. A. cc_asnase/pOPINS3C; B. cc_asnase/pOPINTRX; C. cc_asnase/pOPINHALO7

Chapter V
List of Tables

Chapter I
Table 1.1 Enzymes from microalgae for lipid and biodiesel production (modified from Vingiani, 2019) 9
Table 1.2 Enzymes from Microalgae for Healthcare Applications (from Vingiani, 2019). 14
Table 1.3 Enzymes from microalgae for bioremediation (from Vingiani, 2019). Cu: Copper; Cr: Chromium 18

Chapter II
Table 2.1 Primer sequences of dbph genes used for CDS isolation, Sanger sequencing and cloning in pOPIN vectors 45
Table 2.2 Primer sequences of genes of interest and reference genes used for RT-qPCR (from Vingiani, 2022) 46
Table 2.3 PCR reaction setup for cc_dbph1 and cc_dbph2 CDS isolation and for DBPH1 and DBPH2 colony screening for intron-less sequences 46
Table 2.4 PCR reaction setup for cc_dbph1 and cc_dbph2 CDS isolation with pOPIN adaptors. Name of the PCR application, primer working couple, cycling conditions and PCR reagents are listed. 47
Table 2.5 TBLASTN alignments of the protein sequences of PE esterases available in the literature with the transcriptome of C. closterium FE2 51
Table 2.6 Promising transcripts ORFs visualization via ExPASy Translate and protein classification in NCBI CDD 54
Table 2.7 Restriction enzyme reactions performed in order to verify the presence of the insert 64
Table 2.8 Reference genes assessment using the softwares BestKeeper, NormFinder and geNorm (from Vingiani et al., 2022) 68

Chapter III
Table 3.1 Primer sequences of cc_paf-ah1 and cc_paf-ah2 genes used for CDS isolation, Sanger sequencing and cloning in pOPIN and pJ411 vectors 104
Table 3.2 PCR reaction setup for of cc_paf-ah1 and cc_paf-ah2 CDS isolation with pOPIN and pJ411 adaptors 104
Table 3.3 PCR reaction setup for of cc_paf-ah1 and cc_paf-ah2 CDS isolation with pOPIN and pJ411 adaptors 105
Table 3.4 Primer sequences for cc_paf-ah2/pJ411_N-CTER mutation of S243A 107
Table 3.5 PCR reaction setup for cc_paf-ah2/pJ411_N-CTER mutation of S243A 107
Table 3.6 Reaction setup for Kinase Ligase Digestion (KLD) reaction 108

Chapter IV
Table 4.1 Primer sequences of asnase gene used for CDS isolation, Sanger sequencing and cloning in pJ411 and pOPIN vectors 147
Table 4.2 PCR reaction setup for cc_asnase CDS isolation 148
Table 4.3 PCR reaction setup for cc_asnase CDS isolation with pJ411 and pOPIN adaptors 148
Table 4.4 Restriction enzyme reactions done in order to verify the insert presence 153
Table 4.5 Recombinant microalgal genes expressed in E. coli strains 161

Chapter V

Table 5.1 Key achievements and theorized future perspectives for each Chapter of this thesis 173
List of Common Acronyms and Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>Aminoacids</td>
</tr>
<tr>
<td>ASNase</td>
<td>L-Asparaginase</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment search tool</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>CDS</td>
<td>Coding sequence</td>
</tr>
<tr>
<td>DBP</td>
<td>Dibutyl phthalate</td>
</tr>
<tr>
<td>DBPH</td>
<td>Dibutyl phthalate hydrolase</td>
</tr>
<tr>
<td>DHA</td>
<td>Docosahexaenoic acid</td>
</tr>
<tr>
<td>EC_{50}</td>
<td>(half maximal) Effective concentration</td>
</tr>
<tr>
<td>EDC</td>
<td>Endocrine disruptive chemical</td>
</tr>
<tr>
<td>EPA</td>
<td>Eicosapentaenoic acid</td>
</tr>
<tr>
<td>GOI</td>
<td>Gene of interest</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>nt</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>p-NPB</td>
<td>4-nitrophenyl butyrate</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PAF</td>
<td>Platelet activating factor</td>
</tr>
<tr>
<td>PAF-AH</td>
<td>PAF- acetylhydrolase</td>
</tr>
<tr>
<td>PE</td>
<td>Phthalate ester</td>
</tr>
<tr>
<td>pLySs</td>
<td>plasmid (containing) T7 Lysozyme gene</td>
</tr>
<tr>
<td>POI</td>
<td>Protein of interest</td>
</tr>
<tr>
<td>PUFA</td>
<td>Poly-unsaturated fatty acid</td>
</tr>
<tr>
<td>rpm</td>
<td>revolution per minute</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>Reverse transcription – quantitative polymerase chain reaction</td>
</tr>
<tr>
<td>TAG</td>
<td>Triacyl glycerol</td>
</tr>
</tbody>
</table>
Chapter 1

General Introduction

Microalgal Bioreactors from the BLUEBIO SZN Department (Ph. Mario Di Pinto)
Chapter 1

1.1 Background

1.1.1 Microalgal enzymes for biotechnological applications

Enzymes are organic macromolecules capable of speeding up chemical reactions both in living organisms and in isolated systems, catalysing a huge range of valuable processes. In fact, enzymes are commonly used in several sectors (such as food processing, detergent, pharmaceuticals, biofuel and paper production (Chapman, Ismail and Dinu, 2018; Raveendran et al., 2018; Basso and Serban, 2020) to improve, scale-up and optimize industrial production.

For example, hydrolases are a class of enzymes that catalyse the hydrolysis of a chemical bond and have been extensively used in biofuel production (laccases and cellulases (Cao and Tan, 2002; Srivastava, Mishra and Upadhyay, 2020)) syrup production (amylases (Simair et al., 2017)), food processing (protease, phytases and galactosidases (Fernandes, 2010; Abedi, Pourmohammadi and Sayadi, 2022)), and in pharmaceutical applications (Vellard, 2003).

More efficient and stable enzymes, having optimal activities within a wider range of pH and temperature conditions and possessing a longer half-life, can increase the industrial production of products of interest and render processes, that are not normally economically sustainable, more feasible (such as the aforementioned production of biofuels (Hoffman et al., 2021)). The demand for more efficient enzymes is growing annually, and many financial reports expect the global enzyme market revenue value to reach the $15 billion mark by 2028 (https://www.alliedmarketresearch.com/enzymes-market, https://www.researchandmarkets.com/reports/5322192/enzymes-market-type-source-reaction-type-and), of which c.a. $7-9 billion alone will be for industrial applications (https://www.grandviewresearch.com/industry-analysis/enzymes-industry).

Likewise, novel sources of enzymes are evaluated with great interest by market stakeholders and among these marine organisms are considered an ideal source for investment. The 2014 infographic of the Blue Growth EU initiatives already listed blue biotechnology (including industrial enzymes) as one of the five sectors with a high potential for development in the field of the “blue economy” (https://ec.europa.eu/assets/mare/infographics/#_North_Sea); different projects funded by the European Union Seventh Framework program (EU-FP7; years: 2007-2013) focused on the identification and study of novel enzymes (e.g. project ‘GIAVAP’ (Genetic Improvement of Algae for Value Added Products; http://cordis.europa.eu/project/rcn/97420_en.html and ‘SUNBIOPATH’ ‘Towards a better sunlight to biomass conversion efficiency in microalgae’; http://cordis.europa.eu/project/rcn/92954_en.html; (La Russa et al., 2012)).

Similarly, the EU Horizon 2020 (H2020; years: 2014-2020) projects ‘INMARE’ (Industrial Applications of

Among the different types of marine organisms, microalgae have been receiving increased attention for the discovery of novel enzymes (Brasil et al., 2017). Microalgae are photosynthetic unicellular eukaryotic organisms (even if often the term ‘microalga’ can include both unicellular eukaryotes and cyanobacteria) that can reach high levels of biomass production (up to 100 t/ha year by dry weight under optimal conditions). In nature they contribute up to 50% of global carbon assimilation (Prasad et al., 2021), and are at the basis of aquatic food webs and generally have short generation times. They also possess a variety of genes that are exploitable for biotechnological use (Gimpel et al., 2013) and can be used as producers of industrially important enzymes (Brasil et al., 2017). In addition, the number of known microalgal species represents only a fraction of what is believed to exist (Hadi et al., 2016) and new species are being discovered each year.

Microalgae have a well-proven metabolic plasticity due to their ability to adapt to different environments that can stimulate the production of several enzymes/molecules with possible applications in various biotechnology sectors (i.e., food, energy, health, environment and biomaterials); in fact, recent studies have shown that microalgae produce lipases, phosphatases, amylases, gelatinases, ureases and catalases (Brasil et al., 2017), all of which are enzymes that have a strong biotechnological market niche of their own (Robinson, 2015).

Moreover, due to recent developments in cultivation technology, large amounts of microalgal biomass can be obtained, thereby rendering these microorganisms a renewable if as yet poorly-explored resource for biotech industrial development. In fact, traditional open cultivation methods used in the 1950’s (Venkatesan, Manivasagan and Kim, 2015), have mainly been replaced by the cultivation of microalgae in photobioreactors, under controlled conditions with relatively small quantities of micro- and macro-nutrients, thereby avoiding external contamination (Daneshvar et al., 2021). Overall, their capability to be cultivated with relatively inexpensive resources render them a pivotal source of sustainable processes in a circular bioeconomy model. While, in comparison to bacteria and fungi, microalgae are characterized by longer cultivation times and the lack of standardized and replicable protocols for the evaluation and recovery of downstream products, they are already present in the biotechnological market sector, even if mainly as food and feed supplements (Rahman, 2020; Fernández et al., 2021). On the other side, at the present moment massive culturing of microalgae is mainly impaired by the high costs related to their growth in tightly controlled condition,
harvesting of biomass and bioactives and maintenance of culture purity. (Grobbelaar, 2011; Trivedi et al., 2015; Dębowski et al., 2020). Other hurdles can be: absence of standardized pipelines aimed to the recovery of secondary products in combination with the main product of interest (typically biomass); lack of regulation for the commercial use of engineered strains; lack of low-cost procedures to avoid environmental damages (eutrophication, pollution and biodiversity damage during waste management; (Dębowski et al., 2020).

Even with the actual constrains in the industrial scale-up of microalgae cultivation, the available literature in the PubMed database for “microalgal biotechnology” clearly shows interest in the biotechnological potential of these microorganisms (Figure 1.1). The 20-year interval “2001-2021” indicates that there has been a rapid increase in the number of publications regarding “microalgae” and “biotechnology” as of 2014, with a peak in the year 2020.

![Figure 1.1 “Microalgae Biotechnology” PubMed Search Results (modified from Vingiani, 2019). N. of publications is on the y-axis, while the year is on the x-axis.](image)

The literature regarding the biotechnological applications of microalgae is dominated by four main research sectors: (1) Direct use of microalgal cells, e.g. as bioremediators/biomonitors and as food supplements (Khan, Shin and Kim, 2018); (2) Extraction of bioactives for different applications (e.g. cosmeceutical, nutraceutical, pharmaceutical applications and for biofuel production (Lauritano, Romano, et al., 2016; Bhalamurugan, Valerie and Mark, 2018; Martínez Andrade et al., 2018; Tiwari et al., 2021); (3) Use of microalgae as platforms for heterologous expression or endogenous gene editing and overexpression (Doron, Segal and Shapira, 2016; Jeon et al., 2017; Manfellotto et al., 2020); (4) Use of microalgae as sources of enzymes for industrial applications (Brasil et al., 2017; Abd Ellatif, El-Sheekh and Senousy, 2020). The latter field appears to be less
studied compared to the others, due to the high costs of enzyme extraction and characterization, and scarcity of annotated microalgae genomes and transcriptomes.

Recent projects, such as those funded under the aforementioned EU-FP7 and H2020 framework programs have resulted in an increase in –omics data (i.e. genomics, transcriptomics, proteomics and metabolomics data) available for microalgae, making it possible to develop pipelines aimed at finding new enzymes (Lauritano and Ianora, 2018; Lauritano, Ferrante and Rogato, 2019) from both marine and freshwater species. Microalgal enzymes and their biotechnological applications have been reviewed by different authors (Mogharabi and Faramarzi, 2016; Brasil et al., 2017; Vingiani et al., 2019).

The vast majority of research studies on microalgal enzymes with biotechnological applications belongs to one of three main applicative fields: enzymes for high-value added lipids and biodiesel production, enzymes for healthcare applications and enzymes for bioremediation. The following three subparagraphs are aimed at describing the state-of-the-art for microalgal enzymes identified and characterized in these three fields.

1.1.2 Microalgal enzymes for high-value added lipids and biodiesel production

Microalgae are known to accumulate large amounts of lipids (Chisti, 2007), with triglycerides (TAGs) and polyunsaturated fatty acids (PUFA) being the most studied, particularly for biodiesel and nutraceutical production (M. Sanghvi and Martin Lo, 2012; Chisti, 2013; Bellou et al., 2014; Khan, Shin and Kim, 2018). TAGs, esters derived from glycerol and three fatty acid (FA) chains usually stored in cytosol-located lipid droplets (Moriyama et al., 2018), can be used to produce biodiesel following acid- or base-catalysed transesterification reactions (Fukuda, Kondo and Noda, 2001; Chisti, 2007). PUFAs have well proven beneficial health effects (Wells et al., 2017; Caporgno and Mathys, 2018), especially Ω-3 fatty acids such as docosahexanoic acid (DHA) and eicosapentaenoic acid (EPA) and Ω-6 fatty acid dihomo-γ-linolenic acid (DGLA (Novichkova et al., 2020)). Finally, polar lipids such as phosphatidylcholine (PC) and phosphatidylethanolamine (PE) have proven to be beneficial against inflammation (Lordan et al., 2018) and were recently also discovered from microalgal extracts of Chlorococcum sp. (Shiels et al., 2021) (Figure 1.2).
Figure 1.2 Examples of fatty acids of biotechnological interest (from Vingiani et al., 2019)). a) Through various reactions of elongation and formation of double C-C bonds, polyunsaturated fatty acids (PUFA) can be synthetized, such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) with nutraceutical or food applications; b) Accumulation in triglycerides (TAGs) and biodiesel formation via chemical transesterification.

The most frequently studied enzyme involved in lipid synthesis is acyl-CoA diacylglycerol acyltransferase (DGAT), involved in the final reaction of the TAG biosynthetic pathway (Xu et al., 2018). Three independent groups of enzymes referred to as acyl-CoA diacylglycerol acyltransferases type 1, 2, and 3 (DGATs 1-2-3) take part in the acyl-CoA-dependent formation of TAGs from its precursor sn-1,2-diacylglycerol (DAG) (Lung and Weselake, 2006), with the individual contribution of each DGAT isozyme to the fatty acid profile of TAG differing between species (Shockey et al., 2006; Xu et al., 2018). A gene encoding DGAT1 was first functionally analysed in the green alga Chlorella ellipsoidea by Guo et al. (Guo et al., 2017) via its transformation in A. thaliana and Brassica napus. Other studies reported that the overexpression of DGAT1, or its isoform DGAT1A, and DGAT2 leads to an increase in TAG production in the diatoms Phaeodactylum tricornutum (Niu et al., 2013) and Thalassiosira pseudonana (Manandhar-Shrestha and Hildebrand, 2015), and in the green algae Nannochloropsis oceanica (Wei et al., 2017), Ostreococcus tauri (Wagner et al., 2010) and Neochloris oleoabundans (Klaitong, F-aaroonsawat and Chungjatupornchai, 2017). For N. oleoabundans, DGAT2 overexpression significantly altered fatty acid composition, with a doubling in the levels of saturated fatty acid C16:0, whereas there was a three-fold decrease in the levels of C18:0. In contrast, the overexpression of different isoforms of DGAT2 in the model green algal species Chlamydomonas reinhardtii did not always correlate with the amount of TAGs produced (La Russa et al., 2012).
Li first identified and successfully cloned the DGAT2 gene into an expression vector and used it to transform the *N. oceanica* genome, causing an increase in TAG and neutral lipid content (Li *et al.*, 2016). Different isoforms of DGAT2 (NoDGAT2A, 2C, 2D) have successively been identified in the same species and differentially overexpressed or knocked-out in separate strains, with each engineered strain having a different fatty acid-production profile, some optimized for nutritional applications and others for biofuel purposes (Xin *et al.*, 2017). Recently Cui *et al.* (Cui *et al.*, 2018) characterized a dual-function wax ester synthase (WS)/DGAT enzyme in *P. tricornutum*, whose overexpression leads to an accumulation of both TAGs and wax esters. This was the first report of this enzyme in a microalga, and a patent involving the enzyme was subsequently filed (Patent Code: CN107299090A, 2017). Finally, Ma *et al.* (Ma *et al.*, 2021) identified and cloned five DGAT-encoding genes from the green microalga *Haematococcus pluvialis* in the yeast *Saccharomyces cerevisiae* H1246 and tested their potential for TAG biosynthesis (HpDGAT1, TT1, TT2, TT3, TT4) via complementation. HpDGAT1 was identified as the main TAG synthase (and only its activity was detectable even in *in vitro* assays) while, surprisingly, HpDGTT2 showed a novel lysophosphatidic acyltransferase (LPAAT) activity, that catalyzes the second acylation step of the *de novo* TAG biosynthetic pathway, converting lysophosphatidic acid to phosphatidic acid (PA (Zhukovsky *et al.*, 2019)). This activity was further investigated by transformation in *C. reinhardtii* str. CC-400 cw15 leading to retarded cellular growth, enlarged cell size and enhanced TAG accumulation, mimicking ROS-induced stress signal via PA accumulation (Ma *et al.*, 2021).

Similarly, the overexpression of LPAAT1 resulted in an increase in TAG production in *N. oleoabundans*, both alone (Chungjatupornchai, Areerat and Fa-Aroonsawat, 2019) and combined with glycerol-3-phosphate acyltransferase, DGAT (Muñoz *et al.*, 2019) and DGAT2 (Chungjatupornchai and Fa-aroonsawat, 2021). In this last study a prolonged stability (a 4 year-period) in the overexpression was also observed.

In addition to DGAT and LPAAT, other genes of fatty acid biosynthetic pathway have been targeted in an effort to increase high-value added lipid production: These include: glucose-6-phosphate dehydrogenase (G6PD), phosphogluconate dehydrogenase, glycerol-3-phosphate acyltransferase (GPAT1-GPAT2), acetyl-CoA synthetase 2 (ACS2), phospholipase A2 and other lipases, ACP desaturase and a novel fatty acid photodecarboxylase (FAP). Overexpression of these enzymes resulted in similar increases in TAGs (Zhu *et al.*, 2017; Fukuda *et al.*, 2018; Rengel *et al.*, 2018). For example, overexpression of G6PD (which produces 6-phospho-D-glucono-1,5-lactone and NADPH from D-glucose 6-phosphate and NADP+ (Kletzien, Harris and Foellmi, 1994)) in *P. tricornutum* resulted in a 2.7-fold increase in lipid content (Xue *et al.*, 2017). Similarly, G6PD and phosphogluconate dehydrogenase overexpression led to a 1.5-fold increase in lipids in the diatom *Fistulifera solaris* (Osada *et al.*, 2017).

In addition, two patents for desaturases have been filed. One covers a Δ-6-desaturase from *Nannochloropsis spp.* which converts linoleic acid to γ-linolenic acid (GLA) and α-linolenic acid (ALA) to stearidnoic acid (Patent
Code: CN101289659A, 2010), the second covers a Ω-6-desaturase from *Arctic Chlamydomonas sp. ArF0006*, which converts oleic acid to linoleic acid (Patent Code: KR101829048B1, 2018).

Other approaches to increase lipid production have been employed. Examples include the knock-out of a phospholipase A2 (PLA2) gene via CRISPR/Cas9 ribonucleoproteins in *C. reinhardtii* (Shin et al., 2019), the knock-out of lipases in the diatom *Fistulifera solaris* JPCC DA0580 (Nomaguchi et al., 2018), microRNA silencing of the stearoyl-ACP desaturase (that forms oleic acid via removal of a double-bond in a lipid chain (Los and Murata, 1998)) in *C. reinhardtii* (Mensink et al., 2003), and meganuclease and TALE nuclease genome modification in *P. tricornutum* (Daboussi et al., 2014). This last procedure involved modifying the expression of 7 genes, potentially affecting the lipid content (UDP-glucose pyrophosphorylase, glycerol-3-phosphate dehydrogenase and enoyl-ACP reductase), the acyl chain length (long chain acyl-CoA elongase and a putative palmitoyl-protein thioesterase) and the degree of fatty acid saturation (Ω-3 fatty acid desaturase and Δ-12-fatty acid desaturase). This resulted in the generation of an enhanced lipid producing strain with a 45-fold increase in triacylglycerol accumulation. Thiyagarajan et al. (Thiyagarajan, Arumugam and Kathiresan, 2020a) were able to induce EPA biosynthesis in *E. coli* via a double heterologous expression: an elongase from *Isochrysis sp.* and a novel Δ5-desaturase from *Pavlova sp.*, paving the way towards novel and economic EPA production pipelines.

**Figure 1.3** provides a graphical overview of the subcellular localization of metabolic pathways and engineered enzymes in the aforementioned publications.
Figure 1.3 Main studied and engineered enzymes for TAGs and PUFAs in microalgae for the production of high value-added lipids (modified from Vingiani, 2019). Enzymes are divided in subcellular compartments. A lipid droplet shows where TAGs are accumulated. Abbreviations: Δ-6/Δ-12-Desaturase: delta-6/delta-12 fatty acid desaturase; Ω-3/Ω-6 - desaturase: omega-2/omega-6 fatty acid desaturase; ACS2: acetyl-CoA synthetase; DGAT: acyl-CoA diacylglycerol acyltransferase; ENR: enoyl-acyl carrier protein reductase; G6PD: glucose-6-phosphate dehydrogenase; GPAT: glycerol-3-phosphate acyltransferase; LPAAT: lysophosphatidic acyltransferase; 2; PLA2: phospholipase A2; TAG: triglyceride; UGPase: UDP-glucose pyrophosphorylase

Sorigué et al. (Sorigué et al., 2017) reported for the first time in microalgae, the presence of a photoenzyme called fatty acid photodecarboxylase (FAP) in Chlorella variabilis str. NC64A. FAP that converts fatty acids to hydrocarbons and may be useful in light-driven production of hydrocarbons.

Finally, Sahoo et al. (Sahoo et al., 2020) have developed the dEMBF database (http://bbprof.immt.res.in/embf/) that provides comprehensive and annotated sequences of enzymes involved in lipid synthesis from available microalgal genomes (e.g. C. reinhardtii, P. tricornutum, Volvox carteri). This database aims to identify and classify genes commonly considered as targets for genetic engineering, in order to improve the lipid content in algal species. dEMBF database also aims to be a reference for multiple annotations for uncharacterized genes. To date, the database has collected entries from 32 genomes, while providing different browsing options (Search by: “Enzyme Classification”, “Organism”, and “Enzyme Class”) and different web-based tools (NCBI’s Blast software integrated, sequence comparison, Motif predicament via the MEME software). The advancement made in the design of microalgal strains optimized for lipid accumulation has also been reviewed by Shahid et al. (Shahid et al., 2020).

All of the enzymes discussed in this section are outlined in Table 1.1.

Table 1.1 Enzymes from microalgae for lipid and biodiesel production (modified from Vingiani, 2019). DGAT: acyl-CoA diacylglycerol acyltransferase; WS: wax ester synthase; G6PD: glucose-6-phosphate dehydrogenase; PGD: phosphogluconate dehydrogenase; GPAT: glycerol-3-phosphate acyltransferase; ACS2: acetyl-CoA synthetase; PLA2: phospholipase A2; Δ6-Desaturase: delta-6 fatty acid desaturase; FAP: fatty acid photodecarboxylase; Ω6-Desaturase: omega-6 fatty acid desaturase.

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>Microalgae</th>
<th>Results</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>stearoyl-ACP desaturase</td>
<td>Chlamydomonas reinhardtii</td>
<td>Enzyme gene silencing via microRNA</td>
<td>(Mensink et al., 2003)</td>
</tr>
<tr>
<td>acyl-CoA diacylglycerol acyltransferase 2</td>
<td>Ostreococcus tauri</td>
<td>Gene identification, cloning and heterologous expression in Saccharomyces cerevisiae mutant strains</td>
<td>(Wagner et al., 2010)</td>
</tr>
<tr>
<td>acyl-CoA diacylglycerol acyltransferase 2</td>
<td>Chlamydomonas reinhardtii</td>
<td>Absent TAGs overproduction via overexpression</td>
<td>(La Russa et al., 2012)</td>
</tr>
<tr>
<td>acyl-CoA diacylglycerol acyltransferase 2</td>
<td>Pheodactilium tricornutum</td>
<td>Neutral lipids production enhanced via overexpression with EPA-PUFAs enrichment</td>
<td>(Niu et al., 2013)</td>
</tr>
<tr>
<td>UDP-glucose pyrophosphorylase, glycerol-3-phosphate dehydrogenase, enoyl-ACP reductase, long chain acyl-CoA elongase, putative palmitoyl-protein</td>
<td>Pheodactilium tricornutum</td>
<td>Massive increase in lipid production via 7-genes simultaneous mutagenesis via mecanuclease and TALE</td>
<td>(Daboussi et al., 2014)</td>
</tr>
<tr>
<td>Enzyme Activity</td>
<td>Organism</td>
<td>Function and Conditions</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>----------</td>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td>Thioesterase, ω-3 fatty acid desaturase and Δ-12-fatty acid desaturase</td>
<td>Thalassiosira pseudonana</td>
<td>Increase in TAGs production via overexpression with focus on the intracellular enzyme localization (Manandhar-Shrestha and Hildebrand, 2015)</td>
<td></td>
</tr>
<tr>
<td>acyl-CoA diacylglycerol acyltransferase 2</td>
<td>Nannochloropsis oceanica</td>
<td>Neutral lipids production enhanced via overexpression (Li et al., 2016)</td>
<td></td>
</tr>
<tr>
<td>acyl-CoA diacylglycerol acyltransferase 1</td>
<td>Chlorella ellipsoidea</td>
<td>TAG accumulation role confirmation in different heterologous systems (Guo et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>acyl-CoA diacylglycerol acyltransferase 2</td>
<td>Neochloris oleoabundans</td>
<td>Higher FAS production via overexpression, with doubling of FAs (C16:0) and triple reducing of FAs (C18:0) (Klaitong, Fa-Aroonsawat and Chungjatupornchai, 2017)</td>
<td></td>
</tr>
<tr>
<td>glucose-6-phosphate dehydrogenase; phosphogluconate dehydrogenase</td>
<td>Fistulifera solaris</td>
<td>Absent TAGs overproduction via overexpression (Osada et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>Fatty acid photodecarboxylase</td>
<td>Chlorella variabilis</td>
<td>Enzyme identification and alkane synthase activity test (Sorigué et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>acyl-CoA diacylglycerol acyltransferase 1A</td>
<td>Nannochloropsis oceanica</td>
<td>Increase in TAGs production via overexpression both in nitrogen-replete and -deplete conditions (Wei et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>acyl-CoA diacylglycerol acyltransferase 2A, 2C, 2D</td>
<td>Nannochloropsis oceanica</td>
<td>Differential DGAT2 isoforms expression in different engineered strains (Xin et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>glucose-6-phosphate dehydrogenase</td>
<td>Pheodactilium tricornutum</td>
<td>Modest neutral lipids production enhancement via overexpression with a FAS composition switch from polyunsaturated FAS to monounsaturated FAS (Xue et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>Δ6-desaturase</td>
<td>Pheodactilium tricornutum</td>
<td>Neutral lipids production enhanced via overexpression and relative increase of EPA content (Zhu et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>wax ester synthase/acyl-CoA diacylglycerol acyltransferase</td>
<td>Pheodactilium tricornutum</td>
<td>Neutral lipids and wax esters production enhanced via overexpression (Cui et al., 2018)</td>
<td></td>
</tr>
<tr>
<td>glycerol-3-phosphate acyltransferase 1, 2</td>
<td>Cyanidioschyzon merolae</td>
<td>Massive increase in TAGs production via overexpression (Fukuda et al., 2018)</td>
<td></td>
</tr>
<tr>
<td>Lipase genes</td>
<td>Fistulifera solaris</td>
<td>Genomic Screening and 16 putative lipases genes found (Nomaguchi et al., 2018)</td>
<td></td>
</tr>
<tr>
<td>acetyl-CoA synthetase</td>
<td>Chlamydomonas reinhardtii</td>
<td>Neutral lipids production enhanced via overexpression (Rengel et al., 2018)</td>
<td></td>
</tr>
<tr>
<td>lysophosphatidic acyltransferase</td>
<td>Neochloris oleoabundans</td>
<td>Increase in TAGs production via overexpression (Chungjatupornchai, Areerat and Fa-Aroonsawat, 2019)</td>
<td></td>
</tr>
<tr>
<td>acyl-CoA diacylglycerol acyltransferase 1, glycerol-3-phosphate acyltransferase and lysophosphatidic acyltransferase</td>
<td>Neochloris oleoabundans</td>
<td>Increase in TAGs production via overexpression (Muñoz et al., 2019)</td>
<td></td>
</tr>
<tr>
<td>phospholipase A2</td>
<td>Chlamydomonas reinhardtii</td>
<td>Increase in TAGs production after gene knockout via CRISPR/Cas9 (Shin et al., 2019)</td>
<td></td>
</tr>
</tbody>
</table>
### Microalgal enzymes for healthcare applications

Enzymes for Healthcare Applications can include: 1) enzymes directly used as “drugs”, or 2) enzymes involved in the synthesis of bioactives (Figure 1.4).

Figure 1.4 Enzymes for healthcare applications (from Vingiani, 2019). Enzymes for healthcare applications can include: a) enzymes used directly as “drugs”, such as L-asparaginase and Carboxypeptidase D; b) enzymes involved in the biosynthetic pathway of active compounds, such as polyketides, carotenoids or oxylipins. In the synthesis of polyketides, the studied enzymes are Polyketide synthases.
Regarding the first group, the most studied enzyme is L-asparaginase. This enzyme has a hydrolytic activity against L-asparagine (Ali et al., 2016), and is used for the treatment of acute lymphoblastic leukaemia, acute myeloid leukaemia and non-Hodgkin’s lymphoma (Batool et al., 2016), by reducing the availability of this amino acid for white blood cancer cells, which are unable to synthesize it autonomously. L-asparaginase was first discovered and industrially produced from bacteria (e.g. *Escherichia coli*, *Erwinia aroideae*, *Bacillus cereus* (Roberts, Prager and Bachynsky, 1966; Peterson and Ciegler, 1969; Thenmozhi et al., 2011)). However, in order to overcome economic and safety limitations associated with the enzyme actually on the market, there have been increased efforts focused on the identification and characterization of the enzyme from microbial sources (Ahmad, Pandit and Maheshwari, 2012; Vidya et al., 2017). For example, Paul (Paul, 1982) first purified an L-asparaginase in *Chlamydomonas spp.*, with limited antitumor activity, and tested it in an anti-lymphoma assay in vivo. Ebrahiminezhad et al. screened 40 microalgal isolates via activity assays, stressing *Chlorella vulgaris* as a novel potential feedstock for L-asparaginase production (Ebrahiminezhad et al., 2014). Another group also subsequently identified and characterized an L-asparaginase in *Spirulina maxima*. Following purification of the enzyme from N$_2$-enriched cultures they tested its activity under different pH and temperature conditions and reported the enzyme to be optimally active at pH 8.5 and 37 ± 0.42 °C (Abd El Baky and El Baroty, 2016; Abd El Baky and El-Baroty, 2019).

Regarding enzymatic pathways involved in the synthesis of bioactive compounds, many studies have focused on polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS). PKS produce polyketides while NRPS produce nonribosomal peptides, with both classes of secondary metabolites being formed by sequential reactions operated by these “megasynthase” enzymes (Berry, 2011; Sasso et al., 2012). Polyketides and nonribosomal peptides have been reported to have antipredator, allelopathic, anticancer and antifungal activities (Kobayashi, 2008; Kellmann et al., 2010; Sasso et al., 2012; Kohli et al., 2016). PKS can be multi-domain enzymes (Type I PKS), large enzyme complexes (Type II) or homodimeric complexes (Type III) (Jenke-Kodama et al., 2005). Genes potentially encoding these first two types of PKS sequences have been identified in several microalgae (e.g. *Karenia brevis*, *Amphidinium carterae*, *Azadinium spinosum*, *Gambierdiscus spp.*, *Alexandrium tamutum* (Monroe and Van Dolah, 2008; Kohli et al., 2015; Lauritano, Romano, et al., 2016; Lauritano et al., 2017; De Luca and Lauritano, 2020; Vingiani et al., 2020). Similarly, NRPSs have a modular organization similar to type I PKSs and genes potentially encoding NRPS have been found in different microalgae (Sasso et al., 2012); moreover, metabolites that are likely to derive from hybrid NRPS/PKS gene clusters, have been reported from *Karenia brevis* (López-Legentil et al., 2010). However, to
our knowledge, there are no studies reporting the direct correlation of a PKS or NRPS gene from a microalga with the production of a bioactive compound.

Other microalgal enzymes which have been widely studied are those involved in the synthesis of compounds with nutraceutical and cosmeceutical applications and are involved in carotenoid synthesis (e.g. astaxanthin, β-carotene, lutein and canthaxanthin). Carotenoids are isoprenoid pigments that represent one of the main value-added products studied and extracted from microalgae for their health (antioxidant, anticancer, anti-inflammatory) and industrial (natural colorants) applications (Silva et al., 2020). Antioxidant carotenoid properties can potentially protect humans from a compromised immune response, premature aging, arthritis, cardiovascular diseases and/or certain cancers (Gong and Bassi, 2016). The most studied enzymes involved in their synthesis are: phytoene synthase, phytoene desaturase, β-carotene oxygenase, zeaxanthin epoxidase, violaxanthin de-epoxidase and β-carotene ketolase (Saini et al., 2020). In order to improve the production of carotenoids, different approaches have been employed. The initial method used was to induce random or site directed mutations in an attempt to improve the activity of enzymes involved in the carotenoid metabolic pathway. Increased production of carotenoids can also be achieved by changing the parameters involved in the culture condition or by employing genetic modifications (Saini et al., 2020). For example, phytoene synthase (PSY), which converts geranylgeranyl pyrophosphate to phytoene (Wang et al., 2014), was up-regulated in *Chlorella zofingiensis* under conditions of nitrogen starvation, which induced a 4-fold increase in astaxanthin production (Mao et al., 2018).

Regarding genetic modifications, Cordero transformed the green microalga *Chlorella zofingiensis* by overexpressing phytoene synthase, resulting in a 2.0- and 2.2-fold increase in violaxanthin and lutein production, respectively (Cordero et al., 2011). A phytoene desaturase (PDS), which transforms the colourless phytoene into the red-colored lycopene (Fraser et al., 1992), was mutated in *Haematococcus pluvialis* by Steinbrenner and Sandmann resulting in the up-regulation of the enzyme and an increase in astaxanthin production (Steinbrenner and Sandmann, 2006). Galarza expressed a nuclear PDS gene in the plastidial genome of *H. pluvialis*, resulting in an up to 67% higher astaxanthin accumulation compared to controls (Galarza et al., 2018). Alternative methods include the use of CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9) for precise and highly-efficient knock-out of key genes (Cong et al., 2013). For example, Baek has used CRISPR/Cas9 to “knock-out” the zeaxanthin epoxidase (ZEP) gene in *C. reinhardtii* (Baek et al., 2018). This enzyme is involved in the conversion of zeaxanthin to violaxanthin (Frommolt, Goss and Wilhelm, 2001), and in doing this they obtained a 45.2-fold increase in zeaxanthin production. The current state-of-art involved in metabolic engineering for carotenoid production in microalgae is further discussed in different reviews (Gimpel, Henríquez and Mayfield, 2015; Gong and Bassi, 2016; Saini et al., 2020).
Other studies have focused on enzymes involved in the synthesis of oxylipins, which are secondary metabolites that have previously been shown to have antipredator and anticancer activities (De Los Reyes et al., 2014; Lauritano, Romano, et al., 2016; Ávila-Román et al., 2018). While oxylipin chemistry and putative biosynthetic pathways have been extensively studied in both plants and microalgae (Pohnert, 2002; Matos and Pham-Thi, 2009; Cutignano et al., 2011), only recently, the related enzymes and genes have been identified and characterized in microalgae. Adelfi and co-workers have studied genes involved in the biosynthesis of oxylipins in *Pseudo-nitzschia multistriata* and performed transcriptome analysis on these genes in *Pseudo-nitzschia arenysensis* (Adelfi et al., 2019). They characterized for the first time in diatoms two patatin-like lypolitic acid hydrolases (LAH1) involved in the release of fatty acid precursors of oxylipins and tested their galactolipase activity in vitro. Transcriptomic analysis also revealed three of seven putative patatin genes (g9879, g2582, and g3354) in *N. oceanica* and demonstrated that they were up-regulated under nitrogen-starvation conditions (Li et al., 2014). Similarly, Lauritano and co-workers analysed the transcriptome of the green alga *Tetraselmis suecica* and reported that three PLAT (Polycystin-1, Lipoxygenase, Alpha-Toxin)/LH2 (Lipoxygenase homology) domain transcripts, were down-regulated under nitrogen-starvation conditions (Lauritano et al., 2019). The group also performed in silico domain assessment and structure predictions.

All of the enzymes discussed in this section are described in Table 1.2.

**Table 1.2** Enzymes from Microalgae for Healthcare Applications (from Vingiani, 2019).

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>Microalgae</th>
<th>Results</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-asparaginase</td>
<td><em>Chlamidomonas spp.</em></td>
<td>Enzyme purification and antilymphoma in vivo assay</td>
<td>(Paul, 1982)</td>
</tr>
<tr>
<td>phytotene desaturase</td>
<td><em>Haematococcus pluvialis</em></td>
<td>Increase in astaxanthin production via overexpression</td>
<td>(Steinbrenner and Sandmann, 2006)</td>
</tr>
<tr>
<td>phytotene synthase</td>
<td><em>Chlorella zofengiensis</em></td>
<td>Increase of violaxanthin and lutein of 2.0- and 2.2-fold via overexpression</td>
<td>(Cordero et al., 2011)</td>
</tr>
<tr>
<td>L-asparaginase</td>
<td><em>Chlorella vulgaris</em></td>
<td>Screening of 40 microalgal isolates searching for new L-asparaginase sources</td>
<td>(Ebrahiminezhad et al., 2014)</td>
</tr>
<tr>
<td>Enzymes</td>
<td>Microalgae</td>
<td>Notes</td>
<td>Patent Code (Year)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>L-asparaginase</td>
<td><em>Spirulina maxima</em></td>
<td>Enzyme finding and characterization</td>
<td>(Abd El Baky and El Baroty, 2016; Abd El-Baky and El-Baroty, 2019)</td>
</tr>
<tr>
<td>lypolitic acid hydrolase 1</td>
<td><em>Pseudo-nitzschia multistrata</em>, <em>Pseudo-nitzschia arenysensis</em></td>
<td>Enzyme identification, characterization and retrieving of homologous sequences in other diatoms</td>
<td>(Adelfi et al., 2019)</td>
</tr>
<tr>
<td>zeacantine epoxisase</td>
<td><em>Chlamydomonas reinhardtii</em></td>
<td>Increase of zeaxanthin of 45.2 fold after gene knock-out via CRISR/Cas9</td>
<td>(Baek et al., 2018)</td>
</tr>
<tr>
<td>phytoene desaturase</td>
<td><em>Haematococcus pluvialis</em></td>
<td>Astaxanthin accumulation up to 67% higher via plastidial overexpression</td>
<td>(Galarza et al., 2018)</td>
</tr>
<tr>
<td>polycystin-1, Lipoxigenase, Alpha-Toxin/lipoxigenase homology 2</td>
<td><em>Tetraselmis suecica</em></td>
<td>Three putative enzyme sequences identification and in silico domain assessment and structure prediction</td>
<td>(Lauritano et al., 2019)</td>
</tr>
<tr>
<td>Light-dependant CPD repair enzyme</td>
<td><em>Antarctic Chlamydomonas</em></td>
<td>Enzyme sequence was amplified and isolated from microalgae RNA, and the enzyme was expressed in E. Coli</td>
<td>CN103160488B (2016)</td>
</tr>
</tbody>
</table>

### 1.1.4 Microalgal enzymes for bioremediation

Bioremediation is the use of microorganisms and their enzymes for the degradation and/or transformation of toxic pollutants into less dangerous metabolites/moieties. The capability of microalgae to live in environments rich in nutrients (i.e. eutrophic environments) and to bio-sequester heavy metal ions (HMs) in specific organelles makes them ideal candidate organisms for bioremediation strategies focusing on microalgal metabolism, thus known as “phycoremediation” (de la Noüe, Laliberté and Proulx, 1992; Mathimani and Pugazhendhi, 2019). The most desirable goal in this field is to combine bioremediation with the possibility of extracting lipids and other high-value added compounds from the produced biomass (Kuo et al., 2015; Kim et al., 2016; Hemalatha et al., 2017; Rugnini et al., 2018) in order to reduce overall costs and to recycle materials. Phycoremediation mainly consists of the removal of HMs from contaminated environments via the contribution of metal transporter proteins (MTPs) and metal-binding peptides (metallothioneins and phytochelatins).

The majority of studies in this field have focused on the identification of novel metal-binding peptides in microalgae (reviewed in (Balzano et al., 2020)) while the few transgenic approaches aimed to investigate the mechanisms of HM-resistance more than proteins’ potential biotechnological application. Only Ibuot et al.
obtained an improved Cd uptake and resistance by overexpressing the CrMTP4 gene in *C. reinhardtii* (Ibuot *et al.*, 2017); even in this case the acquired *C. reinhardtii* overall metal tolerance was lower than that of three chlorophyte microalgal strains that had been previously adapted to wastewater growth.

A novel approach for bioremediation, termed enzymatic bioremediation, involves the direct use of purified or partially purified enzymes from microorganisms in order to detoxify a specific toxicant/pollutant (Sharma, Dangi and Shukla, 2018). This method has recently started to produce promising results by using bacterial enzymes (Thatoi *et al.*, 2014; Sivaperumal, Kamala and Rajaram, 2017; Sharma, Dangi and Shukla, 2018). Examples are enzymes for bioremediation of industrial waste and the recent use of chromate reductases, found in chromium resistant bacteria, known to detoxify the highly toxic chromium Cr(VI) to the less-toxic Cr(III) (Langrrd, 1990; Joutey *et al.*, 2015). In microalgae, a recent study focused on Cr(VI) reduction involving *C. vulgaris* (Yen *et al.*, 2017). This activity was suggested to be both via a biological route, through the putative enzyme chromium reductase, and via a non-biological route, using the scavenger molecule glutathione (GSH).

With respect to chromium removal several strains of microalgae have been reported to be capable of removing Cr(IV) from water bodies, including *Scenedesmus* and *Chlorella* species (Daneshvar *et al.*, 2019). In the aforementioned transcriptome study on the green algae *Tetraselmis suecica*, a transcript for a putative nitrilase was found (Lauritano *et al.*, 2019). Given that nitrilases are enzymes that catalyze the hydrolysis of nitriles to carboxylic acids and ammonia (Raczynska *et al.*, 2011) and that this enzyme has recently been used for cyanide bioremediation in wastewaters (Park, Trevor Sewell and Benedik, 2017), this nitrilase in *T. suecica* may be useful in the treatment of cyanide contaminated water bodies.

Other enzymes have been reported to be over-expressed in microalgae when they are exposed to contaminants, even if they are not directly involved in their degradation, but are produced as a stress response (e.g. to detoxify reactive oxygen species produced after exposure to contaminants). Examples of these enzymes include peroxidases (Px), superoxide dismutase (SOD), catalase (CAT) and glutathione reductase (GR). SOD, Px and CAT typically function in helping detoxifying the cell from oxygen reactive species (Lauritano *et al.*, 2015), while GR replenishes bioavailable glutathione, catalyzing the reduction of glutathione disulfide (GSSG) to the sulfhydryl form (GSH). These two typologies of enzymes are summarized in *Figure 1.5*. 
Figure 1.5 Enzymes for bioremediation (from Vingiani, 2019). Enzymes for bioremediation can be: a) Enzymes directly used for the degradation of toxicant compounds into less or non-toxic products (e.g., hexavalent Chromium is converted into the less toxic trivalent Chromium due to the contribution of Chromium Reductase); b) Enzymes involved in cellular stress response mechanisms, such as SOD, Px and CAT, that detoxify reactive oxygen species that are formed, and GR, that replenishes bioavailable glutathione, catalyzing the reduction of glutathione disulfide (GSSG) to the sulfhydryl form (GSH). Abbr. Cr, Chromium; peroxidases, Px; superoxide dismutase, SOD; catalase, CAT; glutathione reductase, GR.

For example, peroxidase activity has been reported in extracts of the green alga *Selenastrum capricornutum* (now named *Raphidocelis subcapitata*) (Suzuki *et al*., 2018), which was highly sensitive to low copper (Cu) concentrations (0.1 mM). Previous authors proposed that the enzyme could be employed as a sensitive bioindicator of copper contamination in fresh waters (Sauser, Liu and Wong, 1997). Levels of Px, SOD, CAT and GR have been reported to be up-regulated following Cu contamination in *P. tricornutum* and following lead (Pb) contamination in two lichenic microalgal strains from the *Trebouxia* genus (prov. names, TR1 and TR9) (Morelli and Scarano, 2004; Álvarez *et al*., 2012). In Morelli’s work, an increase in CAT activity of 200% evidenced its primary role in Cu detoxification. Alvarez, instead, showed how the enzymes resulted in being more active in TR1 than in TR9 in control conditions (with the exception of the catalase), while the prolonged exposure to Pb brought the enzymatic activities of the two microalgae to similar levels, reflecting the different physiological and anatomical adaptations of the two microalgae. The enzymes discussed in this section are reported in Table 1.3.
Table 1.3 Enzymes from microalgae for bioremediation (from Vingiani, 2019). Cu: Copper; Cr: Chromium

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>Microalgae</th>
<th>Results</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Putative ascorbate peroxidase</td>
<td><em>Selenastrum capricornutum</em></td>
<td>High sensitivity to Cu concentration activity</td>
<td>(Sauser, Liu and Wong, 1997)</td>
</tr>
<tr>
<td>Superoxide-dismutase, catalase, glutathione reductase</td>
<td><em>Phedactilum tricornutum</em></td>
<td>Higher detected enzymatic activity after Cu accumulation</td>
<td>(Morelli and Scarano, 2004)</td>
</tr>
<tr>
<td>Superoxide-dismutase, catalase, glutathione reductase, ascorbate peroxidase</td>
<td><em>Trebouxia 1 (TR1), Trebouxia 9 (TR9)</em></td>
<td>Costitutive higher detected enzymatic activity in TR1, while exposition to Pb brings TR1 and TR9 enzymatic activities to comparable levels</td>
<td>(Álvarez et al., 2012)</td>
</tr>
<tr>
<td>Metal transporter protein CrMTP4</td>
<td><em>Chlamidomonas reinhardtii</em></td>
<td>Overexpression and increased Cd accumulation and Cd-toxicity resistance</td>
<td>(Ibuot et al., 2017)</td>
</tr>
<tr>
<td>Putative Cr Reductase</td>
<td><em>Chlorella vulgaris</em></td>
<td>Enzymatic Cr conversion (Cr(VI) -&gt; Cr(III)) detected</td>
<td>(Yen et al., 2017)</td>
</tr>
<tr>
<td>Nitrilase</td>
<td><em>Tetraselmis suecica</em></td>
<td>Putative enzyme sequence identification</td>
<td>(Lauritano et al., 2019)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Enzymes</th>
<th>Microalgae</th>
<th>Notes</th>
<th>Patent Code (Year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NADPH-cytochrome P450 reductase</td>
<td>N.D.</td>
<td>Method of enzymatic extraction in order to use the enzyme for petroleum hydrocarbon degradation</td>
<td>CN106222147A (2016)</td>
</tr>
</tbody>
</table>

1.2 Premises and objectives of the PhD project

Among marine organisms, microalgae appear to be particularly promising for biotechnological applications. Recently, new insights have been gained into both the ecological role and biotechnological potential of numerous microalgae thanks to genome and transcriptome sequencing projects (Sasso et al., 2012; Lauritano and Ianora, 2018). Not many microalgal genomes have been sequenced to date, particularly due to their genome sizes that can also reach 112 Gbp. Even if microalgal genomic availability is rapidly changing (Chen et al., 2020; Nelson et al., 2021), transcriptomic approaches are still preferentially used in order to characterize microalgal functional ecology and recently also to explore the biotechnological potential of species (Keeling et al., 2014).

The EU FP7-funded project PharmaSea ([https://cordis.europa.eu/project/id/312184](https://cordis.europa.eu/project/id/312184)), that ended in 2017, aimed at identifying bioactive marine microalgae and bacteria for possible applications in three market sectors, namely personal care, human health and nutrition. The project identified various microalgae with
bioactivities that could prove beneficial in the treatment of human pathologies (e.g. anti-inflammatory, anticancer, antimicrobial and anti-neurodegenerative activities (Lauritano et al., 2016; Giordano et al., 2018)). The transcriptome of the most active species was sequenced: the dinoflagellate *Amphidinium carterae* with extracts that showed antifungal activity; the dinoflagellate *Alexandrium tamutum* with extracts that showed anticancer activity; the green algae *Tetraselmis suecica* with extracts that showed antioxidant activity; the diatom *Cylindrotheca closterium*, with extracts that showed anti-inflammatory activity (Lauritano et al., 2017, 2019; Elagoz, Ambrosino and Lauritano, 2020; Vingiani et al., 2020).

Based on these premises, my PhD project was designed with the following objectives: (1) analysis of the transcriptomes of previously annotated active microalgal species, focusing on genes that could have direct correlation with ecological and physiological evidences emerged from literature search; (2) identification of gene clusters involved in secondary metabolite production or other enzymes with direct biotechnological applications and selection of at least six enzymes for cloning and heterologous expression in bacterial host; (3) selection of the two most promising enzymes for further functional characterization. This project structure is summarized in Figure 1.6.

![Figure 1.6 PhD final project structure.](image)

After a preliminary bibliographic study, that focused on the four bioactive microalgae with available transcriptomes in my research group, the diatom *Cylindrotheca closterium* was selected as the “most
promising” microalga for enzyme mining and heterologous expression in bacterial hosts. In fact, in *C. closterium* different bioactivities are well supported by scientific literature, such as the anti-inflammatory activity of its extracts (Lauritano et al., 2016), and the capacity to degrade DBP in marine sediments more efficiently than other microalgae (Gao and Chi, 2015; Zhang, Zhao and Chi, 2020). These bioactivities were potentially linked to specific gene sequences already annotated via Blast2GO or found via TBLASTN multiple alignment (in the case of phthalates degradation, where an abundant literature of homologous bacterial enzymes was available (Bhattacharyya et al., 2021). During the PhD period the transcriptome of *A. tamutum* (Vingiani et al., 2020) was also annotated and analysed but without finding gene sequences worthy of further investigation.

It is worth mentioning that, while bacteria are common organisms for the expression of heterologous genes and protein production and characterization, the scientific literature of microalgal enzymes with biotechnological applications expressed in bacterial hosts is not abundant. Different successful heterologous expressions were achieved with enzymes involved in fatty acid production, already mentioned as one of the three main fields of study for microalgal enzymes. First, a β-ketoacyl-coenzyme A synthase from the diatom *Dunaliella salina*, that catalyses the first and rate-limiting step in fatty acid elongation, was cloned in *E. coli* HB101 and indicated how recombinant PUFA elongases could interact with the endogenous enzymes of *E. coli* elongation complex to perform the complete elongation of fatty acids (Azachi et al., 2002).

Moreover, the expression of a C20-elongase from the green alga *Pavlova viridis* (Niu et al., 2009), of a Δ4-desaturase from *P. viridis* (Xu, Niu and Kong, 2011), a Δ6-desaturase from *Isochrysis sp.* (Thiyagarajan et al., 2018) and of a Δ6-elongase and Δ5-desaturase from *Isochrysis galbana* and *Pavlova sp.* respectively, as already mentioned (Thiyagarajan, Arumugam and Kathiresan, 2020), was achieved in *E. coli* BL21 (DE3) strain. In these works, the amount of produced recombinant microalgal protein was variable, and in case of C20 elongase and Δ4-desaturase from *P. viridis* it also caused a moderate toxicity effect in bacterial colonies. A recent work, again from Thiyagarajan et al., also achieved the production of eicosapentaenoic acid and arachidonic acid in BL21 (DE3) via the heterologous expression of the Δ5-desaturase from *Isochrysis galbana* (Thiyagarajan et al., 2021). In these aforementioned cases, the recombinant protein, even when expressed only in small amounts, was enough to make functional characterization of the recombinant enzyme activity in *E. coli*. Previously, the heterologous expression of a squalene synthase from the microalga *Botryococcus braunii*, involved in the synthesis of the squalene, a C20 hydrocarbon derived from the isoprenoid pathway, was also achieved in BL21 (DE3) (Okada, Devarenne and Chappell, 2000). Moreover, between 1990 and 2000, many scientific works aimed to identify a *E. coli* strain optimized for carotenoid production, finding different bottlenecks and rate-limiting steps that halted the diffusion of this trend of metabolic engineering (Paniagua-Michel, Olmos-Soto and Ruiz, 2012). To my knowledge, this was never attempted with microalgal genes.
Overall, this PhD thesis is formed by six Chapters, with the following content:

- **Chapter 1** describes the state of art regarding the main biotechnological applications that involve microalgal enzymes, and the premises and the rationale behind this manuscript.

- **Chapter 2** describes the identification in the transcriptome of the diatom *C. closterium* of the enzymes involved in phthalates degradation (*cc_dbph1-2*), and the relative protein production and characterization.

- **Chapter 3** describes the cloning of genes belonging to the Platelet Activating Factor – Acetylhydrolase family (*cc_paf-ah1-2*) from *C. closterium* transcriptome, and different production attempts of the relative proteins.

- **Chapter 4** describes the cloning of an L-asparaginase gene, whose protein has application in both cancer treatment and acrylamide mitigation, from *C. closterium* transcriptome.

- **Chapter 5** summarizes the outcomes of the PhD thesis and the future steps for the study of the different novel enzymes identified.

- **Chapter M** describes the common experimental methods used in the previous chapters.

The aim of this PhD thesis is to address the current state of art regarding microalgal enzymes with biotechnological application, and to expand the current knowledge of enzymes of interest, via their production and biochemical characterization, starting from available transcriptomic datasets, nowadays still the paramount gateway for massive amount of microalgal genetic data.
Chapter 1 References:


Daboussi, F. et al. (2014) ‘Genome engineering empowers the diatom Phaeodactylum tricornutum for


Chapter II

First Identification of Phthalate Ester-Degrading Enzymes in the Marine Diatom Cylindrotheca closterium

Graphical abstract of the experimental pipeline described in Chapter 2
Chapter 2  Abstract

The transcriptome of the diatom *Cylindrotheca closterium* was previously sequenced under control and silica starvation (Si-) conditions in order to stimulate metabolic pathways of potential biotechnological interest. Considering that previous evidence had shown that *C. closterium* was capable of degrading phthalate esters, in particular dibutyl-n-phthalate (DBP), the transcriptome was examined for enzymes involved in the degradation of phthalate esters (PEs), belonging to the esterase family. PEs are used as plasticizers, solvents and fixatives and due to their tendency to leach from the host matrix and contaminate the surrounding environment, they are classified as endocrine disrupting chemicals (EDCs) and linked to different diseases.

Among various potential transcripts, a transcript containing two dibutyl phthalate hydrolases (provisionally called CcDBPH1 and CcDBPH2, while the corresponding genes were called *cc_dbph1* and *cc_dbph2*), was identified; that were the targets of subsequent approaches.

The CDS of these two genes were isolated via PCR and their eukaryotic origin was proven by the presence of introns in their non-mature RNA sequence, as observed via sequencing. The mature cDNA sequences of *dbph1* and *dbph2* were successfully cloned in TOPO-TA and pOPIN vectors.

Moreover, cultures of *C. closterium* grown with different concentrations of DBP were set-up and RNA samples were collected. RT-qPCR of *dbph* genes showed an increase in their expression levels (time and concentration-dependent) in cultures enriched with DBP, further corroborating the hypothesis of their involvement in DBP metabolism.

Furthermore, CcDBPH1 and CcDBPH2 were successfully expressed in bacterial hosts (as proven via SDS-Page and Western-Blot) and produced in 1L cultures for pilot-scale purification and characterization.

Using *ad hoc* designed purification protocols CcDBPH1 and CcDBPH2 were partially purified and tested against the generic esterase substrate 4-Nitrophenyl-butyrate (p-NPB), with both enzymes showing esterase activity, as expected. However, *in vitro* degradation assays, conducted by incubating the isolated enzymes with the expected substrate DBP, did not produce any conclusive result. Future approaches will include the complete purification of CcDBPH1-2, the co-production of CcDBPH1-2 in the same host organism and degradation assays with bacterial pellets.
2.1 Introduction

2.1.1 The diatom Cylindrotheca closterium

*Cylindrotheca closterium* is a marine cosmopolitan pennate diatom commonly found in the neritic zone, on seaweeds and polar sea-ice (Hargraves, 1998; Von Quillfeldt, Ambrose and Clough, 2003). It can be benthic, pelagic, or part of algal assemblages attached, or not, to various marine organisms (Cupp, 1943; Underwood and Smith, 1998).

The species is characterised by solitary and motile cells which have a lanceolate shape, a nucleus and two chloroplasts in the central region with long, thin, and flexible ends (Cupp, 1943). The size of each cell is 25 - 400 μm in length and 2.5 - 8 μm in width (Figure 2.1).

![Photographs of C. closterium](http://nordicmicroalgae.org/taxon/Cylindrotheca%20closterium).

Photographer/artist: Gert Hansen;  
B: *C. closterium* single cell from [https://www.algaebase.org/search/images/detail/?img_id=5417](https://www.algaebase.org/search/images/detail/?img_id=5417),  
Photographer/artist: Karl Bruun (skogenman@earthlink.net);  
C. closterium can grow quickly in laboratory cultures, even in poorly enriched media (Ohgai, Iwano and Hoshijima, 1986). The favourable conditions for blooming of this species are reached when there is a combination of moderate temperature (20°C) with high salinity (25–35 psu) and nutrient concentrations (total concentration of NH₄-N, NO₂-N and PO₄-P of 6 µg-at·L⁻¹; Affan et al., 2009), even if the evidence from many existing strains has shown that the species is able to adapt to a wide range of temperatures (growth optimum ranging between 9°C and 27°C; Stock et al., 2019). From a reproductive point of view C. closterium displays heterothallic sexual reproduction (Vanormelingen et al., 2013) with pheromones involved in both mate finding and sexual reproduction itself (Klapper et al., 2021).

From an ecotoxicological point of view, it is well characterised for being used as a microbenthic model in environmental toxicity assessments (Moreno-Garrido et al., 2006; Araújo, Diz, Tornero, et al., 2010; Ruocco et al., 2018, 2020) and in investigations on the detrimental effects of heavy metals on diatoms (Moreno-Garrido et al., 2003; Araújo, Diz, Lubián, et al., 2010; Mišić Radić et al., 2021).

C. closterium is suggested to also have antioxidant capacity (evaluated only by enzymatic assay, without using human cells) (Affan et al., 2009) and its extracts have shown anti-inflammatory activities, that was assessed by monitoring the inhibition of tumour necrosis factor α (TNFα), which is a potent mediator of the inflammatory response (Newton and Dixit, 2012), in human THP-1 cells (Lauritano et al., 2016). At the same time, C. closterium extracts did not show anticancer or antimicrobial properties (Lauritano et al., 2016). Recently, its anti-inflammatory activity was proposed to be related to the production of lysophosphatidylcholines and a breakdown product of chlorophyll, pheophorbide a, by chemically analysing the active fraction via a dereplication approach (Lauritano et al., 2020). Moreover, C. closterium was also found to have a high fucoxanthin, a carotenoid with well-proven beneficial health bioactivities (Woo et al., 2010), content and productivity (up to 1.8 mg/L/day; (Wang et al., 2018).

At the present time, the genome of C. closterium is not available but the transcriptome has been recently sequenced and annotated, using an Illumina HiSeq 2500 System, in both control and in silica-starvation (Si-) conditions (Elagoz, Ambrosino and Lauritano, 2020). Nutrient starvation conditions were used to trigger the activation of the broadest range of metabolic pathways of interest. However, stressful growth conditions were suggested to impair anti-inflammatory molecule/s production and the expression of related genes, as observed in the down-regulation of enzymes involved in the synthesis of anti-inflammatory compounds such as the Phosphatidylinositol 3-kinase catalytic subunit type 3 (PIK3C3), Phosphatidylinositol N-acetylglucosaminyltransferase subunit A (PIGA).
Phthalate esters (PEs) are defined as di-esters of 1,2/1,3/1,4 of benzene dicarboxylic acid. They are mainly used to increase plastic flexibility, but are also employed as solvents, fixatives in fragrances, and as additives in cosmetics and personal care products (Abdel daiem et al., 2012). They were introduced as plasticizers in 1920, vitally contributing to the spread of polyvinyl chloride (PVC) as the most common group of plastics worldwide (Tokiwa et al., 2009), which today constitutes 70% of the global plasticizer market (Ren et al., 2018). The most used PEs are diethylhexyl phthalate (DEHP), dibutyl phthalate (DBP), diethyl phthalate (DEP), di-isononyl phthalate (DiNP), and di-iso-decyl phthalate (DiDP), where they are mainly used in the plastic industries as plasticizers to produce PVC. PEs can be synthetized from three isomeric forms (ortho, meta and para) of the benzenedicarboxylic acid backbone, respectively called phthalate (or phthalic acid, PA), isophthalate and terephthalate, and their esters. The length, the ramification and the nature (-alkyl or -aryl groups) of the side chains is described by the nomenclature of the PE, and by default PEs are intended with the –ortho position of the dicarboxylic acids (Figure 2.2). Mono-alkyl/mono-aryl PEs (MAPs and MArPs) are generally included among the PEs, even if they are usually metabolites and reaction intermediates.
Figure 2.2 Common structure of PEs and most common PEs (from Ren et al., 2018). DMP: dimethyl phthalate; DEP: diethyl phthalate; DEHP: diethylhexyl phthalate; BBP: dibenzylbutyl phthalate; DCHP: dicyclohexyl phthalate.
PEs are not chemically bound to the host polymers, so they tend to slowly detach from the host matrix and migrate into the surrounding environment (Benjamin et al., 2015). For this reason, PEs are one of the most common contaminants humans come into contact with, from practically everywhere, and are exposed to PEs via dermal, inhalation, oral, and intravenous routes (Wang and Qian, 2021). PEs are classified as endocrine disrupting chemicals (EDCs) and their presence in the human body has been linked to health diseases and disorders in the reproductive, cardiovascular, respiratory and nervous systems (Gao and Wen, 2016; Verma et al., 2016; Stojanoska et al., 2017; Tsatsakis et al., 2019). This is particularly problematic in children, where interacting with vinyl toys, plastic toys and plastic food packages, usually incurs a higher risk of PEs exposure and linked detrimental effects. For example, different studies on child populations in China and France have shown how the exposure to different PEs was linked to increased risk of development of overweight and obesity (Xia et al., 2018), altered respiratory health among boys aged under 5 years (Vernet et al., 2017) and attention deficit hyperactivity disorder (ADHD) among children of 8–11 years (Kim et al., 2009).

Even if the health risks posed by PEs have only been studied mainly in the last two decades, different countries have already set bans and restrictions to decrease exposure to PEs. In 2008, the USA banned products containing DEHP, DBP, and BBP at levels >0.1% by weight (https://www.congress.gov/110/plaws/publ314/PLAW-110publ314.pdf); similarly in 2007, the EU completely banned DEHP, DBP, BBP, and, since 2018, di-isobutyl phthalate (DiBP) in all plastic materials for toys and childcare products (https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R2005&from=EN). Until recently such regulations did not protect consumers from products imported from China, the biggest manufacturer and consumer of PEs (Gao and Wen, 2016; Wang and Qian, 2021).

Among commonly banned PEs, DEHP accounts for 30% and 80% of the PEs produced in the European Union (EU) and China, respectively (Huang et al., 2013; Meng et al., 2014) and globally makes up 37.1% of the PE’s market (Ren et al., 2018), surpassing a global consumption of 3 million tons (Ren et al., 2018; Wang and Qian, 2021). DBP is also extensively used in nail products as a solvent for dyes (Janjua et al., 2007) and its EU production in 2007 accounted for at least 10.000 tons (Dibutyl phthalate - Substance Information - ECHA, 2009), which presumably has increased in recent years.

2.1.3 The PEs biodegradation and previous findings and applications of PEs-bioremediation enzymes

Abiotic degradation of PEs happens mainly via atmospheric hydrolysis and photolysis (Huang et al., 2013; Fernández-Amado et al., 2017), while the PEs half-life in aquatic or soil environments is usually high enough that the abiotic contribution to their degradation is considered negligible (Lertsirisopon et al., 2009; Net et al., 2015).
The biological pathway of degradation of PEs mainly involves the action of carboxylesterases (EC 3.1.1.1) belonging to the α/β fold hydrolase superfamily, that break the ester bond(s), releasing benzene dicarboxylic acid and its alcoholic moieties (Huang et al., 2013; Benjamin et al., 2015; Bhattacharyya et al., 2021; Figure 2.3).

![Figure 2.3 Representation of hydrolysis of phthalate esters (PEs) by phthalate esterase to phthalic acid via the formation of phthalate monoester intermediate (from Huang et al., 2013).](image)

The conversion of longer alkyl residues to shorter residues by unknown enzymes has also been reported, but very infrequently (Boll et al., 2020). Several microorganisms have been identified that are capable of degrading PEs by using them as the only carbon source (Yan and Pan, 2004; Ahuactzin-Pérez et al., 2014; Kumar, Sharma and Maitra, 2017; Zhao et al., 2018; Nahurira et al., 2019), even if the complete mineralization of PEs usually requires the cooperation of a consortia of different microorganisms (Wang, Fan and Gu, 2004; Wu et al., 2010; He et al., 2013; Pereyra-Camacho, Balderas-Hernández and De Leon-Rodriguez, 2021).

While the degradation of phthalic acid and its isoforms has been studied since the 1960s (Ribbons and Evans, 1960; RIBBONS and EVANS, 1962) the genetic and amino acid sequences of the esterases involved in the transformation of PEs into PAs were characterized much more recently and are mainly from bacterial sources. Moreover, it was observed in different bacterial genomes that genes involved in PE degradation were clustered separately from other genes involved in PA catabolism (Eaton, 2001; Han, 2008; Zhao et al., 2018); the reason for this was related to the secondary involvement of PE hydrolases in other pathways, due to their substrate promiscuity. Moreover, the enzymes involved in PE-degradation are generally only able to convert PEs in their equivalent phthalate monoester, or to convert phthalate monoesters into PAs.
Esterases capable of degrading phthalate di-esters to monoesters have been discovered and their sequences annotated in: *Arthrobacter keyseri* 12B (Eaton, 2001), *Acinetobacter* sp. strain M673 (Wu et al., 2013), *Sulfobacillus acidophilus* DSM10332 (Zhang et al., 2014), *Sphingobium* sp. SM42 (Whangsuk et al., 2015), *Sphingomonas glacialis* PAMC 26605 (Hong, Jang and Lee, 2016), *Gordonia* sp. 5F (Huang et al., 2019), *Bacillus velezensis* SYBC H47 (Huang et al., 2020) and *Microbacterium* sp. PAE-1 (Lu et al., 2020).

Instead, esterases capable of degrading phthalate monoesters to phthalic acid were found in: *Gordonia* sp. strain P8219 (Nishioka et al., 2006), *Rhodococcus jostii* RHA1 (Hara, Stewart and Mohn, 2010), *Rhodococcus* sp. EG-5 (Iwata et al., 2016), *Gordonia* sp. YC-RL2 (Nahurira et al., 2017), *Gordonia* sp. YC-JH1 (Fan et al., 2018) and *Microbacterium* sp. PAE-1 (Lu et al., 2020).

Moreover, a promising trend for the identification of novel PE-esterases in metagenomics libraries has become evident in recent years (Jiao et al., 2013; Wu et al., 2019; Qiu, Yang, et al., 2020; Qiu, Zhang, et al., 2020; Sarkar et al., 2020; Yan et al., 2021).

Finally, different enzymes capable of hydrolysing both of the carboxylic ester linkages, thus fully converting PEs to PAs, have recently been discovered. At present, the known “double action” esterases are: estG from *Sphingobium* sp. SM42, mainly capable of degrading DBP and also showing a weak activity against the MAP intermediate (Whangsuk et al., 2015); carboxylesterase CarEW, isolated from *Bacillus* sp. K91 capable of efficiently degrading both phthalate monoesters and phthalate diesters (Ding et al., 2015); class VIII esterase EstM2, recently isolated from a soil metagenome and modelled as being capable of interacting with a vast array of monoesters and diesters, with a slower efficiency towards MAPs (Sarkar et al., 2020); XtjR8 and EstJ6 esterases isolated from lotus pond sludge and soil metagenomes, respectively, but with both hydrolases showing a stronger activity against dialkyl PEs (Qiu, Yang, et al., 2020; Qiu, Zhang, et al., 2020).

Regarding eukaryotic organisms, there is less evidence of PE degradation. Different fungal species are well known to efficiently degrade PEs (Carstens et al., 2020), while PE metabolism in the human body is usually rapid (Hoppin et al., 2002) and is based on different hydroxylation and oxidation reactions (Wang and Qian, 2021).

However, few PE hydrolases have been characterized from such sources: the fungi *Candida cylindracea* and *Fusarium oxysporum* f. sp. *Pisi* have been reported to possess an esterase and a cutinase, respectively, with activity against dipropyl phthalate (Kim et al., 2005), while only recently the first fungal PE hydrolase with an annotated sequence was obtained from the thermophilic fungus *Malbranchea cinnamomea* (Duan et al., 2019). Among mammals, assays that tested the PE-degradation capability of porcine and bovine pancreatic cholesterol esterases have been performed (Saito et al., 2010), but evidence regarding specific PE-hydrolases and their sequences are still lacking.
Finally, microalgae appear to be the least studied class of PE-degrading microorganisms, and the scientific literature regarding their interaction with PEs is scarce: the green alga *Chlorella pyrenoidosa* has been investigated solely for its capability to grow using DMP as the only carbon source (Yan, Pan and Liang, 2002). DBP addition to the growth medium in the range of mg/L impaired microalgal growth in the freshwater algae *Scenedesmus obliquus* and *Chlorella pyrenoidosa* (Gu *et al.*, 2017) and the marine green alga *Dunaliella salina* (Wei *et al.*, 2021). On the contrary, concentrations of 5–20 μg/L of DBP increased the cell concentration of *Phaeocystis globosa*, *Prymnesium parvum*, *Prorocentrum donghaiense*, *Rhodomonas salina*, and *Chlorella sp.* but did not having significant affect *Skeletonema costatum* and *Phaeodactylum tricornutum* (Gu *et al.*, 2017). The vast majority of toxicity studies have focused on the interaction between microalgae and DBP because it is among the most abundant PE in soil and aquatic environments (Lee *et al.*, 2019), other than being one of the most common PEs as previously outlined.

Chi *et al* showed that the growth of the diatom *C. closterium* was affected by DBP concentrations of 3-5 mg/L, but the diatom also exhibited a strong adaptive ability to PEs: in fact, DBP concentration needed to obtain 50% inhibition (EC50) was higher at 96h DBP incubation time, compared to 48-72h, suggesting a timely reprogramming of its metabolism (Chi, Li and Gao, 2019).

However, *C. closterium* has also shown the ability to degrade DBP in marine sediments (Li *et al.*, 2015), and its efficiency was higher than in the green algae *Dunaliella salina* and the diatom *Chaetoceros muelleri* (Gao and Chi, 2015). Zhang *et al.* also showed how the presence of *C. closterium* resulted in a drastic improvement in DBP degradation in sediments and water in microcosm water-sediment experiments (compared to the sediment bacterial community contribution; (Zhang *et al.*, 2019)). Subsequent experiments suggested that this degradation activity could be influenced by the presence of environmental particles (via its bioavailability), while further corroborating *C. closterium’s* role as the most important biological actor in sediment DBP degradation (Zhang, Zhao and Chi, 2020). This potential use of *C. closterium* in PE degradation stimulated an interest in searching for DBP-degrading esterases in the *C. closterium* transcriptome as a possible source of enzymes for bioremediation applications.
2.2 Methods

2.2.1 *C. closterium* strain specifics, culturing and harvesting

The used strain of *Cylindrotheca closterium* was the FE2 strain, isolated by Dr. Francesco Esposito from an Adriatic Sea seawater sample in 1999 via optic microscopy. FE2 strain 18S sequence is reported in Figure 2.4.

**Cylindrotheca closterium** strain FE2 18S Sequence:

```
GCGGTAATCCAGCTCAATAGCGTATATTTAAGTTGTGCAGTTAAAAGCTCGTAGTTGTGGATTTGTGGTG
TGATGACGCCGCCGCCGCCACATGTGCGAGCTGCTGCCCTGCGCCTCGCCCATCTTTGGGTGGATCTCTGTG
GGCATTAAGTTGTGCGAGGGATGCCATCGTTTACTGTGAAATTATAGAGTGGTCTCAGAGAGACGCT
TAGCCGTTGAATATATTACGATTGAAATAATAGATAGGACCTTGATCTATTTTTGTGTTTGCGCACC
AAGGTAATATGAATAGGAACGTTGGGATCGATTTGCATATTGCAGAGGTGAAATTCTTGAGTTTT
TGAAAGACGAACTACTGGCAAAGGATTACAAATGCAAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGG
TCGAAAGATGATTTAATACACATCGTATGCCTCAAACATAACATTGCCAGAACAGGATGTTGTGAGATCTG
AACCTCCATCGACCTTGTGAGAAATCACAAGGTCCTTGGGTTCTCGGGGGGAGTATGATGTTCGCAAGGC
TGAAATTTAAAAAGATGACGGAAAGGGCAACCCAGGAGTTGAGCCTGCGCTTAAATTGACCTAACAC
GGGAAAACCTACAGTCCAGATCGCATAGGAGGATGACAGATGAGCTCTTTCTGTAGTGTG
```

Figure 2.4 DNA sequence of the 18S sequence of strain FE2 of *Cylindrotheca closterium*.

*Cylindrotheca closterium* FE2 was grown as discussed in General Methods, par. M1, using Guillard’s F/2 medium (GUILLARD and RYTHER, 1962). Routine culture stock maintenance was performed using 25 mL flasks and preparing a new inoculum approximatively every 5-6 days.

*C. closterium* FE2 was grown with the addition of three different concentrations of DBP, 5 mg/L, 10 mg/L and 20 mg/L. The lowest concentration used is similar to the one found in contaminated environments and which was used in previous experiments (2.06 mg/L (Zhang et al., 2019)), and higher to the EC₅₀ concentration established in previous experiments (2.5–3.5 mg/L (Chi, Li and Gao, 2019)).

Cultures were prepared using 2L carboys for each of the three treatments (5, 10 and 20 mg/L of DBP) and for the control growth without the addition of DBP, and each culture was performed in biological triplicates, with a total of 12 carboys. An approximate concentration of 5000 cells/mL was used as the inoculum.

DBP was added to 9 of the bottles after 24h of acclimation in the inoculum. Cell concentration was measured daily by sampling 1 mL from each carboy, as reported in General Methods, par. M1.

Culture aliquots (50 mL) were sampled every 24h for three days following DBP addition (at 24h, 48h and 72h) in technical triplicates for each bottle. In order to obtain a cellular concentration that was sufficient to obtain a good yield of RNA following extraction, culture aliquots at 24h of 100 mL were used.
2.2.2 Bioinformatic mining of CcDBPH putative transcripts

Transcripts potentially involved in PE degradation were identified as described in General Methods, par. M16. The probes used for the TBLASTN alignment search in C. closterium FE2 transcriptome were bacterial monoalkyl PE hydrolases and dialkyl PE hydrolases as listed in the review from Ren et al. (Ren et al., 2018), with the addition of: monoalkyl phthalate hydrolase (mphG1, Gene Bank Code: MH674097.1 (Fan et al., 2018)), dialkyl-ethylhexyl phthalate hydrolase (goEst15; Gene Bank Code: AYW76487.1 (Huang et al., 2019)), feruloyl esterase (BDS4; Gene Bank Code: AWR93191.1 (Wu et al., 2019)), carboxylesterase (baces06; Gene Bank Code: MK617184.1 (Huang et al., 2020)), mono/di-alkyl phthalate hydrolase (mpeH/dpeH; Gene Bank Codes: MK165156/ MK165157 (Lu et al., 2020)), phthalate hydrolase (estM2; Gene Bank Code: KP113669.1 (Sarkar et al., 2020)), EstJ6 esterase (Gene Bank Code: MK037455 (Qiu, Zhang, et al., 2020)), Xjtr8 esterase (Gene Bank Code: MK912039 (Qiu, Yang, et al., 2020)), alpha/beta hydrolase (estYZ5; Gene Bank Code: QWT77157.1 (Yan et al., 2021)). The lipase McLipB (Gene Bank Code: MH254885.1 (Duan et al., 2019)) from a fungal source was also used for the homology alignment. To the best of my knowledge, the probe list covers all the annotated PE hydrolase sequences that are currently available in the literature.

Given the average amino acid length of the probes used (300 aa) Bit-scores below 60 points (with an E-value below 0.0000001) were considered as not significant. The ORFs found in the transcript: TR3123|c0_g1_i1 were selected as putative PE-degrading genes and provisionally named cc_dbph1 and cc_dbph2 (with the related proteins being named CcDBPH1 and CcDBPH2, respectively).

2.2.3 CcDBPHs modeling and docking simulations

The CcDBPH1 and CcDBPH2 3D protein models of the newly-found putative PE hydrolases were built with I-TASSER (Roy, Kucukural and Zhang, 2010; Zheng et al., 2019) and refined with Gromacs v. 2019.6 (Abraham et al., 2015) according to the following outline: each model was immersed in a cubic box with periodic boundary conditions and solvated with water. The appropriate number of Cl− and Na+ ions were added to reach neutrality to simulate a salt concentration of 0.01 M. The Amber ff03 force field and TIP3P water models were used (Jorgensen et al., 1983; Duan et al., 2003). The models were submitted to initial energy minimization with 5000 steps of steepest descent, followed by 100 ps NVT and 100 ps NPT equilibration at 300 K with position restraints before performing 10 ns of unrestrained simulation. The conformational stability of the proteins was evaluated with the tools within the Gromacs suite. A clustering analysis was performed to extract the average representative conformation, which was further minimized before docking experiments. The quality of the final protein models was assessed with Molprobity (Williams et al., 2018), Whatcheck (Hooft et al., 1996) and Verify 3D (Lüthy, Bowie and Eisenberg, 1992). Docking experiments were
performed with AutoDock vina (Trott and Olson, 2009). Models and complexes were prepared and analysed with MGL tools (Sanner, Olson and Spehner, 1996) and UCSF Chimera (Pettersen et al., 2004).

2.2.4 cc_dbph1-2 oligo design and reaction procedures for PCR and RT-qPCR

The newly-identified putative cc_dbphs were selected as genes of interest (GOIs). Primers for regular PCR, for sequencing reactions and RT-qPCR were designed as reported in General Methodologies, par. M5.

For RT-qPCR reference genes (RGs) selection was carried out by testing a series of genes previously used as references for other microalgae (Adelfi et al., 2014; Orefice et al., 2015; Lauritano et al., 2019; Elagoz, Ambrosino and Lauritano, 2020). The selected RGs were: Actin (Act), Glyceraldehyde 3-Phosphate Dehydrogenase (GAPDH), Translation elongation factor-like protein (EFL), Calmodulin (CaM), α-tubulin (Tub-α) and β-tubulin (Tub-β).

The sequences of primers used for endpoint PCR are reported in Table 2.1, while sequences of the primers used for RT-qPCR are reported in Table 2.2.

Table 2.1 Primer sequences of dbph genes used for CDS isolation, Sanger sequencing and cloning in pOPIN vectors. For each gene of interest forward and reverse primers are listed, as well as the amplicon size. In case of pOPIN Cloning primers the adaptor tail is evidenced in bold.

<table>
<thead>
<tr>
<th>Gene of Interest - Application</th>
<th>Forward Primers</th>
<th>Reverse Primers</th>
<th>Length of Amplicons (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcDBPH1 - CDS Isolation</td>
<td>F1: GCCAGTCAGCCCATCAAAAAG</td>
<td>R1: ACCTGACTACGTTCAACTACAC R2: CCCGAATGCATGCTCCTATC</td>
<td>F1R1: 1109 F1R2: 1151</td>
</tr>
<tr>
<td>CcDBPH1 - Sequencing</td>
<td>S1: GATGTGTTCCAAGATAGTCC</td>
<td>S2: GCCACTGGGAGCTTTGGTTT S3: CCTGATGAACATCCCATAATGG</td>
<td>N.D.</td>
</tr>
<tr>
<td>CcDBPH2 - Sequencing</td>
<td>N.D.</td>
<td>R1: TAAATTTGCTTTCTTCAGTGGC R2: CCCTGGAGAGGCGAATTTATTC</td>
<td>N.D.</td>
</tr>
<tr>
<td>CcDBPH1 / pOPINE - Insert Cloning</td>
<td>AGGAGATATACCATGTGTTCAAGAGGAT GCTTCC</td>
<td>GTGATGCTGATGTTCCTTACCTCTCAAGATAAATCTCTTC</td>
<td>1092</td>
</tr>
<tr>
<td>CcDBPH2 / pOPINE - Insert Amplification</td>
<td>AGGAGATATACCATGTGTTCAAGAGGAT GCTTCC</td>
<td>GTGATGCTGATGTTCCTTACCTCTCAAGATAAATCTCTTC</td>
<td>1041</td>
</tr>
<tr>
<td>CcDBPH1 / pOPINF-B - Insert Amplification</td>
<td>AAGTTCTGTTTCAGGCGGCGTATGCTTCCAA</td>
<td>ATGGTCTAGAAGCTTATATCCAGTACTT AGACCC</td>
<td>1102</td>
</tr>
<tr>
<td>CcDBPH2 / pOPINF-B - Insert Amplification</td>
<td>AAGTTCTGTTTCAGGCGGCGTATGCTTCCAA</td>
<td>ATGGTCTAGAAGCTTATATCCAGTACTT AGACCC</td>
<td>1051</td>
</tr>
</tbody>
</table>
Table 2.2 Primer sequences of genes of interest and reference genes used for RT-qPCR (from Vingiani, 2022). For each gene the amplicon size, the primer efficiency (E%) and the coefficient of correlation (R²) are listed.

<table>
<thead>
<tr>
<th>Genes of Interest &amp; Reference Genes</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>Length of Amplicon, bp</th>
<th>E (%)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Act</td>
<td>CCCTACAGCTGGTATCATGG</td>
<td>CTCGGTAAAGGAGCAGGTT</td>
<td>208</td>
<td>100,0</td>
<td>1,00</td>
</tr>
<tr>
<td>CaM</td>
<td>GGTATGAGAAAGCAGGCTG</td>
<td>CAATCGAACCACCCTGTCA</td>
<td>200</td>
<td>99,5</td>
<td>0,99</td>
</tr>
<tr>
<td>EFL</td>
<td>CAATTGTCGATGCCCCAGGA</td>
<td>CAGCAGCAGCAATAACAGCA</td>
<td>197</td>
<td>93,0</td>
<td>0,99</td>
</tr>
<tr>
<td>GAPDH</td>
<td>TGAACGCGCAAGGCTTCTGTT</td>
<td>GATTCGGGCTTTCTGCAA</td>
<td>233</td>
<td>100,0</td>
<td>1,00</td>
</tr>
<tr>
<td>Tub-α</td>
<td>TGCGACACCTGTACAGGTC</td>
<td>CAAGAGGGCATGGTGGAAGA</td>
<td>211</td>
<td>100,0</td>
<td>0,99</td>
</tr>
<tr>
<td>Tub-β</td>
<td>GCTACACTTTCGGTCA</td>
<td>TGGTGGTGTGTGATGTTTCAAATTCTCCTGCCAGGTT</td>
<td>217</td>
<td>100,0</td>
<td>0,99</td>
</tr>
<tr>
<td>CcDBPH1 - qPCR</td>
<td>CGATCTGACCACACCCACCACAA</td>
<td>ACCTCGATTTGCCCCTTTCTT</td>
<td>194</td>
<td>124,0</td>
<td>0,98</td>
</tr>
<tr>
<td>CcDBPH2 - qPCR</td>
<td>ATGGCGGTTGGAAGGATCATG</td>
<td>TGGCATTGCCAGTCTTCTTTCT</td>
<td>208</td>
<td>95,1</td>
<td>0,99</td>
</tr>
</tbody>
</table>

Different PCR conditions and setups were tested for the amplification of each PCR fragment: the Expand™ High Fidelity^PLUS^ PCR System (Roche, Switzerland) was used for the isolation of cc_dbph1 and cc_dbph2 CDS; the PrimeSTAR® GXL DNA Polymerase (Takara, Japan) was used for the amplification of cc_dbph1 and cc_dbph2 CDS with the pOPIN adaptors. PCR reactions were set-up using the cycling conditions and PCR reagents listed in Table 2.3 and Table 2.4.

Table 2.3 PCR reaction setup for cc_dbph1 and cc_dbph2 CDS isolation and for DBPH1 and DBPH2 colony screening for intron-less sequences. Name of the PCR application, primer working couple, cycling conditions and PCR reagents are listed.

<table>
<thead>
<tr>
<th>Application</th>
<th>Working Couple</th>
<th>Cycling Conditions</th>
<th>PCR reagents</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcDBPH1 - CDS Isolation</td>
<td>F1R2</td>
<td>94°C for 2 min; (94°C for 30s; annealing at 60°C for 30s and elongation at 72°C for 1:20 min) for 35 cycles; terminal extension at 72°C for 7 min</td>
<td>2.5 U Expand™ High Fidelity PCR^PLUS^, 10 µL Expand HiFIPRPLUS Reaction Buffer 5x, 0.2 mM dNTPs, 0.4 µM primers, 1 µL template cDNA (20 ng cDNA), MilliQ H₂O up to 50 µL</td>
</tr>
<tr>
<td>CcDBPH2 - CDS Isolation</td>
<td>F2R3</td>
<td>94°C for 2 min; (94°C for 30s; annealing at 55°C for 30s and elongation at 72°C for 1 min) for 35 cycles; terminal extension at 72°C for 7 min</td>
<td>2.5 U Expand™ High Fidelity PCR^PLUS^, 10 µL Expand HiFIPRPLUS Reaction Buffer 5x, 0.2 mM dNTPs, 0.4 µM primers, 1 µL template cDNA (20 ng cDNA), MilliQ H₂O up to 50 µL</td>
</tr>
</tbody>
</table>
Table 2.4 PCR reaction setup for cc_dbph1 and cc_dbph2 CDS isolation with pOPIN adaptors. Name of the PCR application, primer working couple, cycling conditions and PCR reagents are listed.

<table>
<thead>
<tr>
<th>Application</th>
<th>Cycling Conditions</th>
<th>PCR reagents</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcDBPH1 without Intron Check–Colony PCR</td>
<td>95°C for 5 min; (95°C for 30s; annealing at 56°C for 30s and elongation at 72°C for 0:30 min) for 30 cycles</td>
<td>1U Xtra Taq Polymerase, 6 µL XtraTaq Pol White Buffer 5x, 0.2 mM dNTPs, 0.5 µM primers, MilliQH2 up to 50 µL</td>
</tr>
<tr>
<td>CcDBPH2 without Intron Check–Colony PCR</td>
<td>95°C for 5 min; (95°C for 30s; annealing at 56°C for 30s and elongation at 72°C for 0:30 min) for 30 cycles</td>
<td>1U Xtra Taq Polymerase, 6 µL XtraTaq Pol White Buffer 5x, 0.2 mM dNTPs, 0.5 µM primers, MilliQH2 up to 30 µL + picked colony</td>
</tr>
</tbody>
</table>

RT-qPCR reactions were set-up as reported in General Methodologies, par. M7.

2.2.5 cc_dbph12 heterologous expression, SDS-PAGE and purification

CcDBPH1 and CcDBPH2 heterologous expression screening was performed using pOPIN vectors, Lemo21 (DE3) Rosetta 2 (DE3) and BL21 (DE3) strains, overnight at 16°C and 1-hour at 37°C induction conditions, as described in the General Methodologies, par. M12. High-scale production of CcDBPH1 and CcDBPH2 was performed in 1L growth volume, as described in the General Methodologies, par. M12. Selected conditions for the scale-up growth and expression were: cc_dbph1/pOPINE vector in Rosetta 2 (DE3) strain, cc_dbph2/pOPINE vector in BL21 (DE3) strain. Total proteins were extracted from pellets and visualized on SDS-PAGE as described in the General Methodologies, par. M13.

cc_dbph1/pOPINE/Rosetta 2 (DE3) 1L liquid culture was separated in four pellets of approximately 400 OD600 cell number each and purification was attempted involving one pellet at a time. Each pellet was suspended in 15 mL Binding Buffer A composed of: 50 mM Phosphate pH 7.4, 500 mM NaCl, 20 mM Imidazole, 2 mM BME, 1% Triton-X-100. The cell suspension was lysed via 15min sonication with the Sonifier 250 sonicator (Branson, USA), with care being taken to keep the sample on ice. After lysis, the sample was centrifuged for 15min at 12,500 rpm at 4°C in an Avanti™ JA-30I centrifuge (Beckman Coulter, USA). The
supernatant was collected and incubated with agitation for 40 min at 4°C with 400 µL of Ni-NTA Agarose Resin (ABT, USA), previously equilibrated with buffer A, following the manufacturer’s instructions. The sample and resin were loaded on a gravity flow column and the flowthrough volume was collected. At this point the protein, possessing a polyhistidine motif on its C-terminal should have been strongly bound with the nickel charges cross-linked to the resin. The resin was washed with 10 bed volumes (4 mL) of Buffer A.

Finally, the sample was eluted by adding 6 bed volumes of Elution Buffer B (50 mM Phosphate pH 7.4, 500 mM NaCl, 500 mM Imidazole, 2 mM BME, 0.5% Triton-X-100; high-concentration imidazole is used as a competitive agent for the elution of His-tagged proteins (Bornhorst and Falke, 2000)), and each fraction was collected separately. UV absorbance at 280 nm was measured in order to identify the elution fractions that had detectable amounts of proteins (usually fractions 2-4). Protein-rich fractions were loaded on a Econocolumn (BioRad) containing Sephadex G-25 resin (previously equilibrated using manufacturer’s instructions) for buffer change to Purification Buffer C (20 mM Phosphate pH 7.4, 150 mM NaCl, 5 mM BME, 20% Glycerol). 10-20 500 µL fractions were collected and their UV absorbance at 280 nm was measured in order to identify protein-containing fractions. Aliquots of samples which were being processed in the aforementioned steps were taken and visualized via SDS-PAGE, as per General Methodologies, par. M11.

Protein-rich fractions in buffer C were quantified via the Bradford protein assay (Sigma Aldrich) following the manufacturer’s instructions and the 96 well plate assay protocol. Before quantification samples were properly diluted due to incompatibility with glycerol and reducing agents, as per assay’s instructions.

CcDBPH2 purification followed a similar protocol: cc_dbph2/pOPINE/BL21 (DE3) 1L liquid culture was separated in four pellets of approximately 500 OD_{600} cell number each. Each pellet was suspended in 15 mL of Binding Buffer A’ (50 mM Phosphate pH 8, 500 mM NaCl, 20 mM Imidazole, 2 mM dithiothreitol (DTT), 1% Triton-X-100. The cell suspension was lysed via 15’ sonication and centrifuged for 15’ at 12,500 rpm at 4°C. The supernatant was collected and incubated in agitation for 40’ at 4°C with 400 µL of Ni-NTA Agarose Resin (ABT, USA), previously loaded with buffer A’, following the manufacturer’s instructions. Sample and resin were loaded, washed and eluted with Elution Buffer B’ (50 mM Phosphate pH 8, 500 mM NaCl, 500 mM Imidazole, 2 mM DTT, 0.5% Triton-X-100) as previously described. The UV absorbance at 280 nm of the elution fractions was measured and protein-rich fractions were loaded on a Econocolumn (BioRad) containing Sephadex G-25 resin (previously equilibrated using the manufacturer’s instructions) for buffer change to Purification Buffer C’ (20 mM Phosphate pH 8, 150 mM NaCl, 2 mM DTT, 20% Glycerol). 10-20 500 µL fractions were collected and protein-rich fractions were processed as previously described.
2.2.6  **Esterase activity assay**

The esterase activity of the recombinant CcDBPH1 and CcDBPH2 was measured via the degradation of the generic esterase substrate 4-nitrophenyl butyrate (p-NPB); hydrolysis by this class of enzymes releases the yellow chromophore 4-nitrophenol. The increased amount of this product was spectrophotometrically analysed at 405-415 nm (Anderson *et al.*, 1994; Zhu *et al.*, 2015; Huang *et al.*, 2019). 3 µg of each partially purified protein sample was incubated with 1 mM p-NPB in 200 µL buffer (20 mM Phosphate pH 7.4, 150 mM NaCl with 0.5% Triton). A 100 mM p-NPB stock solution was prepared in H₂O and added to the reaction mix to a final concentration of 1 mM.

Samples were prepared into a Costar® 96 multi-well plate (Corning Incorporated, USA) and agitated at room temperature for 18h, with absorbance at 405-415 nm being measured with the Microplate Reader Infinite® M1000 PRO (Tecan, Switzerland) or the Multiskan PlateReader (Thermo Fisher Scientific, USA).

In both cases, the measurement times were: every minute for 30’, every 5’ for 60’ (total: 1:30h) and every 30’ for 12h (total kinetic measurement time: 13:30h).

2.2.7  **DBP-degradation activity assay and measurement via HPLC**

The DBP-degradation capability of the recombinant CcDBPH1 and CcDBPH2 was assessed by incubating partially purified protein samples (5 µg) in 500 µL buffer (20 mM Phosphate Buffer pH 7.4, 150 mM NaCl) with or without 0.5% Triton-X.100. A 1mM DBP stock solution was prepared in methanol and added to the reaction mix to a final concentration of 100 µM.

Additionally, a combination of the two proteins, using 5 µg from each protein sample, was also tested.

After substrate addition, the samples were agitated at room temperature for 18h. Subsequently the samples were stored at -80°C, after which two volumes of methanol (1mL) were added, to precipitate the proteins. The samples were centrifuged at 12,000 rpm, 4°C, for 15’ and analysed by HPLC.

Analyses were performed on an Agilent 1260 Infinity II HPLC (Agilent Technologies, USA) equipped with a Poroshell 120 EC-C18 column (Agilent Technologies). The elution was performed as follows: Buffer A was 0.1% Formic Acid in H₂O, buffer B was 0.1% Formic acid in methanol with a flow rate of 0.5 mL/min.
2.3 Results and discussion

2.3.1 C. closterium (FE2) culture with DBP and growth data

The growth curve of C. closterium FE2 culture triplicates in control and under DBP+ conditions are reported in Figure 2.5.

![C. closterium Growth Curve](image)

**Figure 2.5** C. closterium growth curve in control and DBP+ conditions.

From the growth curve it is clear that the microalgal cultures entered the exponential phase on day 3, which is similar to what has previously been reported (Affan et al., 2009). The growth rates of C. closterium FE2 DBP+ samples in the exponential phase decreased in comparison to the control, as expected. In fact, the DBP concentration interval used (5-20 mg/L) was higher than the C. closterium FE2 DBP EC50 established in previous experiments (2.5–3.5 mg/L at 72h (Chi, Li and Gao, 2019)). On the other hand, the increasing concentration of DBP did not impair growth on a concentration-dependent basis, at least in the time span under study.

Nonetheless, paragraph 2.3.7 will later report that 24h after DBP addition, increased expression of the putative dbph genes was observed in a time-dependent basis. However, as more extensively discussed in the
conclusions section, it was difficult to link these observations with the similarly observed decrease in growth with 5 mg/mL DBP, 10 mg/mL DBP and 20 mg/mL DBP.

All sampled aliquots were freeze-dried and stored at -80°C. Successively, their RNA was extracted and cDNA produced as reported in General Methods par. M3-M4. The obtained samples were used for both CDS isolation of genes involved in PAE-degradation and for RT-qPCR.

2.3.2 **Sequences coding enzymes involved in DBP degradation response**

The *C. closterium* FE2 transcriptome was screened for sequences of enzymes that could be involved in PEs degradation. As described above, PEs degradation is performed by different esterases capable of performing the hydrolysis of PE side chains. Therefore, a simple bioinformatic strategy was used to identify members of this enzyme superfamily, using the program and queries previously mentioned in the Methods section.

The protein sequences of bacterial PEs hydrolases previously identified in the literature were used as probes, using the TBLASTN algorithm (Gerts et al., 2006) included in the BioEdit program. The TBLASTN, among the different available BLAST alignments, was preferred because the esterase sequences that were used were almost all bacterial sequences. Thus simply using BLAST, risked omitting results because of the different codon usage of bacteria and microalgae. The TBLASTN results are shown in **Table 2.5**.

<table>
<thead>
<tr>
<th>Enzyme (Gene Name)</th>
<th>Substrate</th>
<th>Esterase Family</th>
<th>GenBank Protein ID</th>
<th>Aminoacid Probe Length</th>
<th>TBLASTN 1° Best Match Transcript</th>
<th>1° Best Match Bit-Score (E-score)</th>
<th>TBLASTN 2° Best Match Transcript</th>
<th>2° Best Match Bit-Score (E-score)</th>
<th>Source Bacterium</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>phthalate ester hydrolase (pehA)</td>
<td>DAP</td>
<td>XVII</td>
<td>AAK165 32.1</td>
<td>218</td>
<td>TR19157</td>
<td>c0_g1_i1 len=407</td>
<td>34 (0.097)</td>
<td>TR13068</td>
<td>c0_g1_i1 len=1552</td>
<td>32 (0.37)</td>
</tr>
<tr>
<td>mono-ethylhexyl phthalate hydrolase</td>
<td>MPE</td>
<td>MEHP hydrolase</td>
<td>BAE785 00.1</td>
<td>311</td>
<td>TR13492</td>
<td>c0_g1_i2 len=1532</td>
<td>45 (7e-005)</td>
<td>TR7884</td>
<td>c0_g2_i1 len=1151</td>
<td>45 (7e-005)</td>
</tr>
<tr>
<td>Enzyme Name</td>
<td>Type</td>
<td>Accession</td>
<td>Scaffold ID</td>
<td>E-value</td>
<td>Percentage</td>
<td>Species/Strain</td>
<td>Source</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
<td>-----------</td>
<td>-------------</td>
<td>---------</td>
<td>------------</td>
<td>----------------</td>
<td>-------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phthalate Ester Hydrolase (pateE)</td>
<td>MPE</td>
<td>XVII</td>
<td>ABG992 14.1</td>
<td>221</td>
<td>35 (0.058)</td>
<td>TR19157</td>
<td><em>Rhodococcus</em> Jostii RHA1</td>
<td>(Hara, Stewart and Mohn, 2010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diakyl PE Hydrolase</td>
<td>DPE</td>
<td>IV</td>
<td>AFK313 09.1</td>
<td>365</td>
<td>74 (2e-011)</td>
<td>TR3123</td>
<td>Acinetobacter sp. strain M673</td>
<td>(Wu et al., 2013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diakyl Phthalate Hydrolase (dphB)</td>
<td>DPE</td>
<td>IV</td>
<td>AGY559 60.1</td>
<td>302</td>
<td>102 (4e-002)</td>
<td>TR3123</td>
<td>Uncultured bacterium clone 2</td>
<td>(Jiao et al., 2013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triacylglycerol Lipase (<em>EstS1</em>)</td>
<td>DPE</td>
<td>IV</td>
<td>AEW03 609.1</td>
<td>304</td>
<td>87 (2e-017)</td>
<td>TR1865</td>
<td><em>Bacillus</em> sp. K91</td>
<td>(Ding et al., 2015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diakyl Phthalate Hydrolase (<em>estG</em>)</td>
<td>DPE &amp; MPE</td>
<td>IV</td>
<td>AJO678 04.1</td>
<td>315</td>
<td>104 (1e-022)</td>
<td>TR15089</td>
<td><em>Sphingobium</em> sp. SM42</td>
<td>(Whang suk et al., 2015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diakyl Phthalate Hydrolase (<em>estB</em>)</td>
<td>DPE</td>
<td>N.D.</td>
<td>AJO678 03.1</td>
<td>384</td>
<td>54 (3e-007)</td>
<td>TR13492</td>
<td><em>Gordonia</em> sp. YC-RL2</td>
<td>(Nahurira et al., 2017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monoethylhexylphthalate Hydrolase (mehpH)</td>
<td>MPE</td>
<td>MEHP hydrolase</td>
<td>BAU220 81.1</td>
<td>303</td>
<td>60 (2e-009)</td>
<td>TR11852</td>
<td><em>Rhodococcus</em> sp. EG-5</td>
<td>(Iwata et al., 2016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diakyl Phthalate Hydrolase (<em>estSP1</em>)</td>
<td>DPE</td>
<td>IV</td>
<td>WP_010186968.1</td>
<td>314</td>
<td>92 (5e-019)</td>
<td>TR11334</td>
<td><em>Sphingomonas</em> glacialis PAMC 26605</td>
<td>(Hong, Jang and Lee, 2016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monoethylhexylphthalate Hydrolase (mepH)</td>
<td>MPE</td>
<td>MEHP hydrolase</td>
<td>KU2122 13.1</td>
<td>311</td>
<td>50 (2e-006)</td>
<td>TR13492</td>
<td><em>Gordonia</em> sp. YC-RL2</td>
<td>(Nahurira et al., 2017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Type</td>
<td>Accession</td>
<td>Length</td>
<td>E-value</td>
<td>Similarity</td>
<td>Source</td>
<td>Year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
<td>-----------</td>
<td>--------</td>
<td>---------</td>
<td>------------</td>
<td>-------------------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>monoalkyl phthalate hydrolase (mphG1)</td>
<td>MPE</td>
<td>MEHP</td>
<td>311</td>
<td>48</td>
<td>(1e-005)</td>
<td>Gordonia sp. YC-JH1</td>
<td>2018</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dialkyl-ethylhexyl phthalate hydrolase (goE015)</td>
<td>DPE</td>
<td>VII</td>
<td>539</td>
<td>137</td>
<td>(3e-032)</td>
<td>Gordonia sp. 5F</td>
<td>2019</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>feruloyl esterase (B054)</td>
<td>DPE</td>
<td>IV</td>
<td>317</td>
<td>100</td>
<td>(2e-031)</td>
<td>Unculture d microorganism</td>
<td>2019</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>carboxyl esterase (baces04)</td>
<td>DPE</td>
<td>VI</td>
<td>289</td>
<td>35</td>
<td>(0.067)</td>
<td>Bacillus velezensis strain SYBC H47</td>
<td>2020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>diethyl phthalate hydrolase (deph1)</td>
<td>MPE</td>
<td>V</td>
<td>280</td>
<td>32</td>
<td>(0.54)</td>
<td>Methylobacterium populic yc-xj1</td>
<td>2020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>monoalkyl phthalate hydrolase (mpeH)</td>
<td>MPE</td>
<td>MEHP</td>
<td>304</td>
<td>33</td>
<td>(0.27)</td>
<td>Microbacterium sp. PAE-1</td>
<td>2020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dialkyl phthalate hydrolase (dpeH)</td>
<td>DPE</td>
<td>MEHP</td>
<td>240</td>
<td>37</td>
<td>(0.013)</td>
<td>Microbacterium sp. PAE-1</td>
<td>2020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>family VIII phthalate hydrolase (estM2)</td>
<td>MPE  &amp; DPE</td>
<td>VIII</td>
<td>392</td>
<td>34</td>
<td>(0.22)</td>
<td>Unculture d bacterium clone pETESTM 2</td>
<td>2020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>esterase gene (estJ6)</td>
<td>MPE  &amp; DPE</td>
<td>IV</td>
<td>298</td>
<td>96</td>
<td>(3e-020)</td>
<td>Unculture d organism</td>
<td>2020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hydrolase gene (xtjr8)</td>
<td>MPE  &amp; DPE</td>
<td>IV</td>
<td>314</td>
<td>91</td>
<td>(2e-018)</td>
<td>Unculture d organism</td>
<td>2020</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Starting from the original 24 sequence entries, 12 gave significant results, showing homology for a total of 9 different transcripts. In order to further investigate the transcriptome entries identified by this alignment, the sequences were visualized and translated via the ExPASy translate tool, to identify potential Open Reading Frames (ORFs) that could be associated with esterase enzymes. Complete ORFs, containing at least 100 residues, were aligned in the NCBI CDD database. Results of this second identification step are shown in Table 2.6.

Table 2.6 Promising transcripts ORFs visualization via ExPASy Translate and protein classification in NCBI CDD. Transcript codes, Gene names that are aligned with the transcript, position and length and NCBI CDD protein classification of the realistic ORF(s) identified in the transcripts are reported. In yellow are evidenced the most promising ORFs, in blue ORFs that were found with high levels of homology with bacterial esterases but with unusual aa number, in green are evidenced the ORFs selected for further experiments.

<table>
<thead>
<tr>
<th>Transcript Code</th>
<th>Aligned with Gene (Bit-score, E-score)</th>
<th>ORF1 Start Position (RF)</th>
<th>ORF1 Aminoacid Aminoacid Length</th>
<th>ORF1 Protein Classification in NCBI CDD</th>
<th>ORF2 Start Position</th>
<th>ORF2 Aminoacid Aminoacid Length</th>
<th>ORF2 Protein Classification in NCBI CDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR3123</td>
<td>c0_g1_i1</td>
<td>dialkyl PE hydrolase (74, 2e-013); dphB (102, 4e-022); EstS1 (87, 2e-017); estG (104, 4e-022); estSP1 (92, 5e-019); BDS4 (100, 2e-021); estj6 (96, 3e-20); xtjr8 (80, 2e-015)</td>
<td>38 (5'-3'+1)</td>
<td>355</td>
<td>alpha/beta hydrolase</td>
<td>481 (5'-3'+2)</td>
<td>338</td>
</tr>
<tr>
<td>TR19133</td>
<td>c0_g1_i1</td>
<td>EstS1 (63, 2e-010); estG (63, 3e-019); xtjr8 (91, 2e-008)</td>
<td>34 (5'-3'+3)</td>
<td>725</td>
<td>acetyl esterase/lipase (Interval: 472-697)</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
<tr>
<td>TR1865</td>
<td>c0_g1_i1</td>
<td>carEW (164, 1e-040); goEst15 (137, 3e-013); McLipB (100, 4e-021)</td>
<td>23 (5'-3'+2)</td>
<td>563</td>
<td>alpha/beta hydrolase (Interval: 34-452)</td>
<td>N.D.</td>
<td>N.D.</td>
</tr>
</tbody>
</table>
The phthalate diesterase probes were classified as belonging to family IV of the 19 Esterase Families classification (Bhattacharyya et al., 2021), with the exception of pehA (family XVII) and carEW and goEst15 (family VIII). On the other hand, phthalate monoesterases were distributed in a different phylogenetic clade, separated from all the other families (Bhattacharyya et al., 2021). All of the family IV PE esterase, with the exception of estYZ5, gave the transcript “TR3123|c0_g1_i1 len=2609” as 1° or 2° Best Match. This transcript was also the one that appeared more often as 1° or 2° Best Match in all the alignments (8 out of 12 sequences that gave significative entries).

The transcript TR3123|c0_g1_i1 len=2609 possessed two ORFs (RF 5'-3' +1 and 5'-3' +2 respectively) of 355 and 338 aa in length, respectively, and both were identified by NCBI CDD as “alpha/beta hydrolases”. The amino acid number of these two novel hydrolases was coherent with the average amino acid number of the PE hydrolase (300 aa).

Considering that *C. closterium* is capable of degrading PEs has to date mainly been studied with dialkyl PEs, the two ORFs contained in this transcript were evaluated to be the most feasible PE-degrading target genes obtained using this gene mining approach. The related proteins were considered adequate to perform heterologous expression and the *in vitro* PEs degradation assay. ORF1 and ORF2 were provisionally named *cc_dbph1* and *cc_dbph2*, and the related proteins were named CcDBPH1 and CcDBPH2, respectively. The sequences of CcDBPH1 and CcDBPH2 shared a 57% sequence identity and a 69% similarity to each other, respectively.
It is worth mentioning that the presence of two different ORFs in the same transcript is a rare but plausible event: in fact, while polycistronic transcripts are mainly a bacterial characteristic, the extensive usage of operon units by green algae such as *Chlamydomonas reinhardtii* and *Chromochloris zofingiensis* have recently been discovered and characterized ([Gallaher et al., 2021; Zou et al., 2021]), while previously it has been also found in fungi, plants and other animals such as *Drosophila* ([Pauli, Tonka and Ayme-Southgate, 1988; Blumenthal, 2004; Matsuda and Dreher, 2006; Gordon et al., 2015; Mouilleron, Delcourt and Roucou, 2016; Wang et al., 2019]). While similar studies have not been reported yet in diatoms it is likely that CcDBPH1 and CcDBPH2 have a similar organization. TR3123|c0_g1_i1 len=2609 sequence is reported in Figure 2.6.

**Figure 2.6** DNA sequence of the transcript TR3123|c0_g1_i1 len=2609, where the two new PE esterases (provisionally named DBPH1 and DBPH2) were found. *dbph1* CDS is evidenced in yellow; *dbph2* CDS is evidenced in blue.
The other significant transcripts listed in Table 2.6 matched only against diesterase probes (including some esterase with dual activity against DPE and MPE). The ORFs contained in these transcripts translated in amino acid sequences that were always confirmed as hydrolases/esterases from NCBI CDD, with the exception of “TR23311|c0_g1_i1 len=3116”.

On the other hand, these sequences often showed an amino acid number that was higher than the average amino acid number of bacterial PE hydrolases, with the highest homology with the queries of higher size (such as: CarEW 489 aa, goEst15 539 aa, McLipB 565 aa). This could potentially be linked to functional similarities. In fact, Huang et al. speculated that the enzyme size of goEst15 could be related to its higher affinity for large PE substrates (DEHP, DNOP, DCHP and BBP), requiring a greater flexibility (and size) to accommodate substrates in different binding modes (Huang et al., 2019). The scientific knowledge about PE hydrolases (especially from eukaryotic sources) is still too scarce to make broad generalizations, especially in the absence of any structural data. However, the ORFs coding for shorter amino acid sequences contained in transcripts “TR1865|c0_g1_i1 len=1870”, TR11852|c0_g1_i1 len=1778 and TR6682|c0_g1_i1 len=1588 nonetheless gave significant homology scores (especially the transcript “TR1865|c0_g1_i1 len=1870”, that had a Bit-score with carEW of 164 (1e\(^{-040}\)), the highest among all the alignments) which could thus be interesting targets for future experiments. Finally, entries associated with monoesterases, as already mentioned, did not give homogeneous results and lower Bit-scores and E-scores, never reaching the appropriate significance threshold.

2.3.3 **DBPHs catalytic site identification and 3D structure modelling**

The protein sequences of CcDBPH1 and CcDBPH2 were aligned with other previously described bacterial dialkyl PEs hydrolases of family IV (dialkyl PE hydrolase, dphB, estG, estP1) in order to infer the presence of the putative catalytic triad of our proteins of interest and investigate their conserved motifs (Figure 2.7). Based on the sequence alignment, it was possible to deduce the putative catalytic triads composed of S180, H323 and D293 in DBPH1 and by S170, H309 and D279 in DBPH2.

The eight conserved motifs which typically characterize the sequences of family IV bacterial PE esterases were also present (Ren et al., 2018). The sequences of CcDBPH1 and CcDBPH2 from C. clustorium displayed the HGGG, GXSAGG and HG motifs of bacterial PE esterases (ARPIGNY and JAEGGER, 1999; Long and Cravatt, 2011; Ren et al., 2018). On the other hand, they lacked the TH motif, while the YP motif was conserved in DBPH1 and replaced by NP in CcDBPH2. Similarly, the GV motif was conserved in DBPH2 but replaced by HV in CcDBPH1. Finally, for the YXLAPE motif, which is conserved in bacterial dialkyl PEs hydrolases, the final glutamic acid residue was replaced by an asparagine in both proteins. Another major divergence, shared by
both CcDBPH1 and CcDBPH2, was the absence of the TAXXD motif, which comprises the catalytic aspartic, which was replaced by a conserved VGDLD stretch in the two microalgal proteins.

<table>
<thead>
<tr>
<th>Protein sequence alignment of the newfound CcDBPH1 and CcDBPH2 (boxed in red) with other Class IV bacterial dialkyl esterases (from Vingiani, 2022). The protein sequences used in the alignment are: dialkyl PE hydrolase from Acinetobacter sp. strain M673 (GenBank: JQ478494.1; Wu, 2013); dph8 from an uncultured bacterium clone (GenBank: AGY59840.1 Jiao, 2013); estG from Sphingobium sp. SM42 (GenBank: KM386873.1; Whangsuk, 2015); EstP1 from Sphingomonas glacialis PAMC 26605 (GenBank: WP010186968.1; Hong, 2016). Conserved regions are boxed in black, while the putative aminoacids of the catalytic triad are marked with a red star.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Figure 2.7</strong> Protein sequence alignment of the newfound CcDBPH1 and CcDBPH2 (boxed in red) with other Class IV bacterial dialkyl esterases (from Vingiani, 2022). The protein sequences used in the alignment are: dialkyl PE hydrolase from Acinetobacter sp. strain M673 (GenBank: JQ478494.1; Wu, 2013); dph8 from an uncultured bacterium clone (GenBank: AGY59840.1 Jiao, 2013); estG from Sphingobium sp. SM42 (GenBank: KM386873.1; Whangsuk, 2015); EstP1 from Sphingomonas glacialis PAMC 26605 (GenBank: WP010186968.1; Hong, 2016). Conserved regions are boxed in black, while the putative aminoacids of the catalytic triad are marked with a red star.</td>
</tr>
</tbody>
</table>

In the two *C. closterium* DBPHs, the first variable residue (X1) of the GXSXG pentapeptide motif is a glutamic acid (E) which has a negative-charged side chain. Such negative-charged residues are present in all the aligned dialkyl PEs hydrolases which showed a major degradation activity against DPEs (Ren et al., 2018). Even GoEst15 and baces04, belonging to family VII and VI respectively, had a D residue in position X1 (Huang et al., 2019, 2020). The conserved presence of a negative amino acid in position X1 can be explained by the repulsive action of the –COOH residue against the negative charge of the carboxylate anion that is formed after the first ester bond cleavage of the dialkyl PE, helping to free the active site after hydrolysis while guaranteeing reaction specificity towards dialkyl PEs.

A similar effect was proposed about the presence of positively-charged amino acids in position X1 and/or X2 for monoalkyl PEs hydrolases (Maruyama et al., 2005; Nishioka et al., 2006; Fan et al., 2018). In this case positively charged residues in position X1 and X2 in monoalkyl PEs esterases neutralize the negative charge of the carboxylate anion allowing it to attach to the reaction site. In fact, the mutation of X1 residue of the GXSXG motif in the monoalkyl phthalate hydrolase “MphG1” isolated from *Gordonia sp. YC-JH1* caused a
severe loss of activity, confirming its role in the formation of a proper electrostatic interaction (Fan et al., 2018). About the exceptional capability of diesterase-monoesterases enzymes to degrade both DPE and MPE as primary substrates, no explanation is available to date, except that this is a property that is not linked to a specific esterase family. This dual substrate specificity was found in family VII (CarEW), VIII (EstM2) and IV (estj8, xtr8 and, even if with a very weak activity, estG). Nonetheless, even if there is not yet an explanation for the capability of *C. closterium* to degrade MAPs, it is highly possible that CcDBPH1 and CcDBPH2, due to their negative charges in position X1 are most likely to be involved in the first hydrolysis step.

Moreover, to confirm the possible DBP degrading function of CcDBPH1 and CcDBPH2, homology models of the two proteins were constructed (as explained in par. 2.2.3). The first difficulty that was encountered related to the fact that even if the sequences of several microbial PE hydrolases are known, to date no structure is available in the Protein Data Bank (PDB) for this class of esterase. To detect suitable templates for homology modelling, the PDB database was searched with HHpred (Hildebrand et al., 2009). Several templates were detected for both CcDBPH1 and CcDBPH2, all belonging to the subfamily IV of bacterial α/β hydrolases, as expected. All of them, however, shared quite low identity (< 25%) with the targets, with the highest sequence divergence at the N-terminus of the proteins. To build full length models of the two proteins, the iterative fragment assembly approach implemented by the automatic pipeline of I-TASSER was used (Roy, Kucukural and Zhang, 2010; Yang et al., 2014), which has been shown to function well when compared to other methods, particularly when modelling difficult targets (Croll et al., 2019). Using I-TASSER structures with acceptable confidence levels, considered the low homology (c-score 0.22 for CcDBPH1 and 0.38 for CcDBPH2, Figure 2.8), were obtained.

![Figure 2.8](image) Full length, refined homology models of CcDBPH1 (left) and CcDBPH2 (right) coloured according to their Verify3D score (from Vingiani et al., 2022). A higher score indicates a stronger reliability of the model, given its amino acid sequence.
Prior to docking analysis, both models were refined with a 10 ns run of unrestrained molecular dynamics. The trajectories underwent clustering analysis and average conformations were further minimized and evaluated with Molprobity (Williams et al., 2018), Whatcheck (Hooft et al., 1996) and Verify 3D (Lüthy, Bowie and Eisenberg, 1992).

The latter analysis in particular highlighted some incongruence between the sequence and the N-terminus of both modelled structures. Since this portion of the protein is not believed to participate in substrate binding during catalysis, this stretch (residues 1–20 in DBPH1 and 1–40 in DBPH2) was removed; before attempting to dock DBP in the catalytic pocket of the enzymes. In the case of CcDBPH1, the binding of DBP was found to produce a favourable affinity (−31.0 kJ/mol). Analysis of the binding site confirmed that DBP fitted correctly, in proximity to the residues of the catalytic triad, S180, D293 and H323 (Figure 2.9). A significant energetic contribution was predicted to be provided by the pi-interactions between the aromatic ring of DBP and the side chain of the tryptophan residue W240, which is therefore likely to assist the positioning of the indole ring of the substrate in the enzymatic pocket, in a similar fashion to what has recently been reported for the class VIII PE hydrolase EstM2 (Sarkar et al., 2020).

Figure 2.9 Structural detail (A) and 2D map (B) of the interaction between DBP and CcDBPH1 (from Vingiani et al., 2022). Residues involved in substrate binding with DBP are labelled.

In the case of CcDBPH2, despite attempts with different docking protocols, no binding poses with favourable energy were detected. This could indicate that the enzyme is not actually involved in the first step of DBP hydrolysis but could nonetheless be involved in a regulatory function, highlighting the possible role of DBPH1
as the primary actor in the degradation of DBP (this hypothesis will be further corroborated by other evidence in the following paragraphs). Given the polycistronic nature of *cc_dbph1* and *cc_dbph2* transcripts, it is difficult to imagine that the two proteins are not involved in the same degradation pathway.

2.3.4 **PCR isolation of di-n-butyl phthalate hydrolase (DBPH) coding sequences and purification**

The *cc_dbph1-2* coding sequences were successfully isolated via PCR using 20 ng of cDNA template, as reported in General Methodologies par. M6. The followed strategy was to first clone isolated sequences in the TOPO-TA vector, in order to have a higher PCR efficiency when using pOPIN primers (the rationale for this is explained in General Methodologies, par. M8 and M9).

*cc_dbph1* CDS isolation was achieved by trying different primer combinations (amplicon length, as well as primer sequences and reaction final setup are listed in par. 2.2.4) and annealing temperatures (AT). Both primer combinations F1R2 and F1R1 showed efficient amplification of expected band length with different degrees of amplification depending on the AT (with an absence of amplification for the couple F1R2 at 56°C AT). Agarose gel electrophoresis resolution is shown in Figure 2.10.

![Figure 2.10 Agarose gel of cc_dbph1 CDS amplified with different primer couples (F1R2 and F1R1) and temperatures.](image)

F1R2 primer couple with an AT of 60°C was arbitrarily selected as the best reaction setup and selected for further purification. However, because of the faint amplification of *cc_dbph1* three individual reactions with
the same setup were merged in the same tube in order to retrieve sufficient quantities of DNA after the purification step.

The \textit{cc\_dbph1} PCR product did show aspecific amplifications and was purified using the AMPure XP for PCR Purification kit (Beckman Coulter) as reported in General Methodologies, par. M6 (\textbf{Figure 2.11})

![Image of agarose gel with λ-2 ladder and purified cc\_dbph1 CDS](image)

\textbf{Figure 2.11} Agarose gel of \textit{cc\_dbph1} F1R2 CDS purified and quantified via indirect quantification with λ-2 ladder.

Based on the standardized quantity of the λ-2 fragments, the \textit{dbph1} CDS was equated for intensity of the band to the third upper band of the ladder, so that a quantity of 65 ng/µL was determined for the purified sample.

The \textit{cc\_dbph2} CDS isolation was performed following a similar approach since the \textit{cc\_dbph1}. \textit{cc\_dbph2} primer couple F1R1 was unable to produce any amplification, even after changing the annealing temperature or doubling the amount of used template or primers (data not shown). Other primer pairs (all tested at AT of 56°C, as reported in methods) gave different results. Couples F2R3, F1R3 and F2R1 showed efficient amplification of expected band length (near the 1200-1000 bp ladder bands). Agarose gel electrophoresis resolution of the products is shown in \textbf{Figure 2.12}. 

62
Figure 2.12 Agarose gel of \textit{cc_dbph2} CDS amplified with different primer couples (F1R2, F2R2, F2R3, F1R3 and F2R1) at 56°C.

F2R3, F1R3 and F2R1 primer couples were selected as the most efficient and produced fragments selected for further purification. The \textit{cc_dbph2} PCR products did show aspecific amplifications and were purified using GenElute Gel Extraction kit (Sigma-Aldrich) as reported in General Methodologies, par. M6 (Figure 2.13).
Via λ-2 indirect quantification the cc_dbph2 CDS was equated for intensity of the band to the third upper band of the ladder, so that quantities of 50 ng/µL for F2R3 and 25 ng/µL for F1R3 and F2R1 were determined for the purified samples. F2R3 fragment was used for further experiments since it was the most abundant.

2.3.5 Cloning, sequencing of DBPH in TOPO-TA vectors and intron retention resolution

The purified cc_dbph1 and cc_dbph2 CDS were cloned in TOPO-TA vectors as reported in General Methodologies, par. M8. Successful cloning was confirmed via enzymatic digestion as reported in General Methodologies, par. M9.

The enzymes used were selected based on the restriction map of the multiple cloning site of TOPO-TA vector (available on the manufacturer’s website at the link: https://tools.thermo Fisher.com/content/sfs/vectors/prcr2_1topo_map.pdf) and on the restriction map of the inserted fragment, predicted via the ApE program.

The restriction enzymes used and the expected size of product are listed in Table 2.7.

<table>
<thead>
<tr>
<th>Cloned Gene of Interest</th>
<th>Used Enzyme</th>
<th>N. of Insert Restriction Sites in (position)</th>
<th>Expected Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcDBPH1</td>
<td>EcoRI</td>
<td>1 (543-548)</td>
<td>Linearized vector of 3900 kb + 609 bp fragment + 542 bp fragment</td>
</tr>
<tr>
<td>CcDBPH2</td>
<td>HindIII</td>
<td>1 (508-513)</td>
<td>Linearized vector of 5080 (3973 + 1107) bp</td>
</tr>
</tbody>
</table>

Figure 2.13 Agarose gel of dbph2 F2R3, F1R3 and F2R1 CDS purified and quantified via indirect quantification with λ-2 ladder.
Once the presence of the sequence of interest was demonstrated via restriction enzyme digestion, the DNA MiniPrep preparations of \textit{cc\_dbph1} and \textit{cc\_dbph2} were sequenced, using the original primers already used for CDS isolation, generic primers external to the sequence of interest (e.g. M13 Forward, M13 Reverse, T7 Promoter, Sp6) and also “internal” primers (listed in par. 2.2.4), in order to ensure full coverage of the double-strand sequence.

Alignments between the sequenced \textit{CcDBPH1} and \textit{DBPH2} CDS and the original transcriptome sequence are shown in Figure 2.14 and 2.15.

The main differences between the two sequences are the presence in both \textit{CcDBPH1} and \textit{CcDBPH2} of intron sequences, respectively, of 69 bp starting from position 453 for \textit{DBPH1} and of 74 bp starting from position 423 for \textit{DBPH2}. Both introns were characterized by the typical 60-70 nt length and GT-AG spliceosome ends.

While this result was a useful confirmation of the eukaryotic source of the identified sequences, the presence of introns could be linked to different explanations: were the introns from genomic contamination of RNA source samples or were they functionally inserted in the pre-mRNA as part of a regulatory post-transcriptional mechanism? The first possibility was the least plausible, considering that the iScript™ cDNA...
Synthesis Kit (BioRad) used, implemented a primer blend mainly composed of oligo(dT). Moreover, the TRizol RNA extraction protocol should have precipitated the vast majority of the DNA in the interphase. In any case, the cDNA retro-transcription kit was changed, using the PrimeScript RT Reagent Kit with gDNA Eraser (Takara) to remove any possible genomic DNA contamination, and a new cDNA from the same RNA sample was prepared.

PCR was performed using primers localized in proximity to the 5’ and 3’ of the two introns, so that a shorter product was expected in case of a sequence without introns. For cc_dbph1 primers SEQ Forward and qPCR Reverse were used (products of 320 and 389 bp were expected, without or with the intron respectively), while for cc_dbph2 primers qPCR-F and qPCR-R were used (products of 200 and 274 bp were expected, without or with the intron respectively).

Moreover, a positive control reaction using the DNA preparations of the cc_dbph1 -2 included in the TOPO-TA (thus with introns) as template was performed, from which only the longer sequences should be produced. The PCR results are shown in Figure 2.16. All reactions were run with an annealing temperature of 56°C and 30” of elongation time and visualized using a 2% agarose gel.

Even if the gel electrophoresis failed to show the cc_dbph2 positive control it was evident that the two new cDNA still had a heterogeneous population of fragments with and without introns of the cc_dbph1-2 genes. Probably, even the purified CcDBPH1 and CcDBPH2 CDS contain a mixture of sequences with and without the intron that are impossible to separate. In the majority of cases, pre-mRNA has a shorter half-life in
comparison to mature mRNA and intron removal may have occurred before the addition of the poly-A tail. Therefore, the presence of such mature mRNA with intron retention could have a post-transcriptional purpose, and be involved in the post-transcriptional regulation of CcDBPH1-2 products.

While further approaches to study post-transcriptional regulation in *C. closterium* were evaluated, a colony PCR screening approach was set up in order to grow in liquid culture and purify only the colonies that showed PCR amplification of the shorter fragment, using the aforementioned primer couples and reaction settings. A colony containing an intron-less clone for each gene was grown overnight, purified via Miniprep and sequenced as already described. Alignments between the newly sequenced CcDBPH1 and CcDBPH2 CDS and the transcriptome sequence and alignments of their translated sequences are shown in Figure 2.17 and 2.18.

Figure 2.17 CcDBPH1 CDS and translated aminoacid sequence alignment from transcriptome and newly intron-less sequenced, as visualized via Jalview. In light blue are the areas of non-homology.

Figure 2.18 CcDBPH2 CDS and translated aminoacid sequence alignment from transcriptome and newly intron-less sequenced, as visualized via Jalview. In light blue are the areas of non-homology.
This time the sequences did not show introns, as expected; however, they presented multiple errors (base changes of the transcriptomic “original” sequence was considered real if present in both with-intron and intron-less de novo sequences of CcDBPH1/CcDBPH2). These errors could be caused by to the use of a regular Taq Polymerase in the PCR reaction (c.a. 1 per 3300 bp; https://www.neb.com/tools-and-resources/feature-articles/polymerase-fidelity-what-is-it-and-what-does-it-mean-for-your-pcr) or by errors in the transcriptomic source sequence.

In the case of CcDBPH1 these differences involved four ammnoacid changes: L98M, N140K, Y166D and Q246R; and in the case of DBPH2 these differences involved two amino acid changes: R31L and R69G.

These changes did not involve the catalytic triad of the enzymes, nor the putative residues involved in substrate binding.

In any case, in parallel to the CDS isolation and following protein production, RT-qPCR analyses of cc_dbph1 and cc_dbph2 in samples grown at different concentrations of DBP for different times were performed in order to further substantiate the previously theorized roles of CcDBPH1-2 in C. closterium and their 3D model; the observed errors in the transcriptomic sequence did not involve the primers designed for RT-qPCR.

2.3.6 **RT-qPCR of FE2 + DBP samples and data analysis**

RT-qPCR reactions were set-up as reported in General Methodologies, par. M5. Before the analysis of the expression levels of the DBPH GOIs, putative reference genes (RGs) were first screened, selecting them from those already tested as RGs for the diatom C. closterium and already used for other diatoms in previous studies. Table 2.8 and Figure 2.19 summarize the results obtained on the RGs assessment, by using the three softwares BestKeeper, NormFinder and geNorm.

<table>
<thead>
<tr>
<th>RANKING</th>
<th>BESTKEEPER</th>
<th>NORMFINDER</th>
<th>GENORM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Act</td>
<td>Act</td>
<td>Tub-β - GAPDH</td>
</tr>
<tr>
<td>2</td>
<td>Tub-β</td>
<td>Tub-α</td>
<td>Act</td>
</tr>
<tr>
<td>3</td>
<td>EFL</td>
<td>EFL</td>
<td>EFL</td>
</tr>
<tr>
<td>4</td>
<td>GAPDH</td>
<td>CaM</td>
<td>CaM</td>
</tr>
<tr>
<td>5</td>
<td>Tub-α</td>
<td>GAPDH</td>
<td>Tub-α</td>
</tr>
<tr>
<td>6</td>
<td>CaM</td>
<td>Tub-β</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.19 Reference gene assessment for C. closterium (from Vingiani et al., 2022). Results on the ranking of the most stable reference genes according to (a) BestKeeper (lowest standard deviation), (b) NormFinder (lowest expression stability value) and (c) geNorm (lowest average expression stability M) softwares. Tested genes were: Actin (Act), Glyceraldehyde 3-Phosphate Dehydrogenase (GAPDH), Translation elongation factor-like protein (EFL), Calmodulin (CaM), α-tubulin (Tub-α) and β-tubulin (Tub-β).

According to the results obtained by BestKeeper, the lowest standard deviation (SD) was obtained for Act, followed by Tub-β and EFL (Figure 2.19, a). Instead, according to NormFinder, the lowest stability values were found for Act, followed by Tub-α and EFL (Figure 2.19, b). Finally, according to the results of geNorm, the two most stable genes were Tub-β and GAPDH (Figure 2.19, c). Considering the best RGs assigned by each software, Act, Tub-β and GAPDH were selected as RGs for RT-qPCR analyses.

The best predicted RGs were in part different in comparison to previous results, where C. closterium was exposed to silica-starvation (in that case the best RGs were Tub-α, GAPDH, Act and EFL (Elagoz, Ambrosino and Lauritano, 2020) or RGs used for other diatoms, such as Skeletonema marinoi, exposed to silica-starvation, phosphate-starvation, CO₂-starvation (Lauritano et al., 2015) and aging (Orefice et al., 2015) or Pseudo-nitzschia multistriata and P. arenysensis exposed to nitrogen-starvation (Adelfi et al., 2014). This was expected, since RG stability varies depending on the studied species and strain, growth phases, nutrient culturing conditions and exposure to different environmental stressors.

Relative expression levels of the selected GOIs, cc_dbph1 and cc_dbph2, were analyzed in order to investigate the effect of DBP direct exposure on the diatom cultures (Figure 2.20).
Effectively, results showed a significant concentration-dependent increase in both \textit{cc\_dbph1} and \textit{cc\_dbph2} at both 24 h and 48 h DBP incubation (Figure 5; \(p < 0.05\) for all except \(p < 0.01\) for \textit{CcDBPH1} after 48 h DBP exposure at both 5 and 10 mg/L and \(p > 0.05\) for \textit{CcDBPH2} after 48 h exposure at 5 mg/L). Relative expression data after 72h and at 20 mg/L DBP gave results with very high standard deviation values and were omitted from the final graphs.

![Figure 2.20](image.png)

\textbf{Figure 2.20} Expression levels of \textit{cc\_dbph1} and \textit{cc\_dbph2} in \textit{C. closterium} cells cultured with DBP exposure for 24 h and 48 h at 5 and 10 mg/L DBP compared to the control condition (no DBP exposure, 0 mg/L DBP; expressed in the figure by the x-axys (from Vingiani \textit{et al.}, 2022). Data are represented as log2 x-fold expression ratio ± SD (\(n = 3\)). Asterisks indicate significance of the observed data based on standard deviation (SD) estimation via \textit{Taylor algorithm} as per (Pfaffl, Horgan and Dempfle, 2002); one asterisk means a P value ≤ 0.05, two asterisks mean a P value ≤ 0.01.

The data showed a significant response in the expression of the two putative \textit{cc\_dbph} genes at 5 and 10 mg/L.

Specifically, at the lower DBP concentrations the expression of \textit{cc\_dbph1} and \textit{cc\_dbph2} showed an increase at 24h (at 5- and 3- fold respectively) and a subsequent decrease at 48h, with \textit{cc\_dbph2} expression in particular that returned to the level shown in the control conditions. It is possible that at 48h the diatom had already degraded most of the PE in solution, coherently reprogramming \textit{cc\_dbph1-2} expression to basal levels; a somewhat similar degradation time was previously shown by Gao and Chi where \textit{C. closterium}
intracellular crude cellular extract, incubated with 0.4 mg/L DBP, decreased DBP concentration to less than 0, 2 mg / L in 24h, until it became no longer detectable at 72h (Gao and Chi, 2015).

Nonetheless, at the 10 mg/L DBP concentration the expression levels of CcDBPH1 and CcDBPH2 increased after 24h (with 6- and 5- fold increase, respectively) and 48h (with 8- and 6- fold increase, respectively), indicating a concentration-dependent expression intensity and duration. When comparing the overall levels of expression of the two cc_dbph genes at both concentrations, it appears that the expression level of cc_dbph1 is more clearly impacted by DBP concentration, also showing smaller standard deviations. To the best of my knowledge this is the only RT-qPCR experiment to date focused on genes involved in phthalate esters degradation in microalgae.

The only other RT-qPCR experiments related to the pathway of phthalate degradation genes were performed by Dr. Ruyang Han in his PhD Thesis. In this work the expression levels of three genes involved in the first step of phthalate’s isomers degradation pathway, that is phthalate dioxygenase gene opHA2, isophthalate dioxygenase gene ipHA2, and terephthalate dioxygenase gene tphA2, were measured in Comomonas testosteroni strain YZW-B grown on succinate, phthalate, isophthalate, terephthalate, and two or three phthalate isomer combinations. This experiment showed that these three genes were highly induced by the corresponding phthalate isomers (332 to 1117 fold), while showing more ambiguous gene induction in the case of growth with phthalate mixtures (Han, 2008).

In the case of microalgae, an apparently fainter expression induction could be associated with other forms of post-transcriptional regulation or be only coherent with the role of microalgae in PEs degradation but inability to further proceed in phthalate catabolism. In previous work, the charophyte freshwater green algae Closterium lunula was reported to be able to degrade di-n-methyl phthalate (DMP), with an overall efficiency that depended on the available concentration of inorganic carbon, and also a fainter capability to further process the degradation intermediate PA (Yan and Pan, 2004). Even so, there is a clear need for more data and evidence to explain the modulation of such genes towards the different PEs and exposure times, and the metabolic differences regarding PEs degradation between bacteria and microalgae.

2.3.7 Cloning and sequencing of dbph1-2 in pOPIN vectors

cc_dbph1 and cc_dbph2 CDS sequences with pOPIN adaptors were produced using 5 ng of DNA preparation of dbph1/2-TOPO-TA template; with CDS produced being treated with DpnI enzyme, then purified and cloned into pOPIN vectors as reported in General Methodologies, par. M10.

The growth of cc_dbph1 /pOPINA-B and cc_dbph2 /pOPINA-B showed slower growth rates in comparison to cc_dbph1 /pOPINE-F and cc_dbph2 /pOPINE-F, and also their DNA preparations had lower concentrations (30-40 ng/µL against the average 150-200 ng/µL concentration of cc_dbph1-2/pOPINE-F).
This lower concentration meant that $cc_{dbph1}$ /pOPINA-B and $cc_{dbph2}$ /pOPINA-B sequencing would require higher volumes and further eventual steps of re-transformation. Therefore, for expression screening it was decided to proceed with the sequencing of $cc_{dbph1}$ /pOPINE-F and $cc_{dbph2}$ /pOPINE-F (in brief, every GOI inserted in two different expression plasmids), keeping aside the use of $cc_{dbph1}$ /pOPINA-B and $cc_{dbph2}$ /pOPINA-B for a second moment, in case of need.

Results of DNA sequences of $cc_{dbph1}$ /pOPINE-F and $cc_{dbph2}$ /pOPINE-F did not show further base changes. $cc_{dbph1}$ /pOPINE-F and $cc_{dbph2}$ /pOPINE-F plasmid maps were produced using SnapGene and are shown in Figure 2.21.
**Figure 2.21** Plasmid maps of cc_dbph1/pOPINE-F and cc_dbph2/pOPINE-F produced using SnapGene. A. cc_dbph1/pOPINF; B. cc_dbph1/POPINE; C. cc_dbph2/POPINF; D. cc_dbph2/POPINE. The inserted gene DNA is showed in red and the related protein in dark purple (CcDBPH1) or in light purple (CcDBPH2). All plasmids show their size (bp), the lac operator (blue), the T7 promotor (arrowhead white), the replication origin (ori, yellow) and the Ampicillin resistance gene (AmpR, light green). pOPINE maps show the C-terminus His-tag (6xHis, light purple); pOPINF maps show the N-terminus His-tag (6xHis, light purple) and the cleavage site for protease 3C (dark blue).

### 2.3.8 Heterologous expression of cc_dbph1-cc_dbph2/pOPIN

The molecular mass of CcDBPH1 and CcDBPH2, based on their primary amino acid structure, was predicted using Protpi. CcDBPH1 was predicted to have a mass of **41.69 kDa** with the N-terminus fusion tag of pOPINF (MAHHHHHHSSGLEVLFQP...) and **40.63 kDa** with the C-terminus fusion tag of pOPINE (...KHHHHHHH), respectively. DBPH2 was predicted to have a mass of **39.35 kDa** with the N-terminus fusion tag of pOPINF and **38.29 kDa** with the C-terminus fusion tag of pOPINE, respectively.

Gel images are shown in **Figure 2.22-2.24**.
Figure 2.22 SDS-Page of soluble and insoluble fractions of protein extracts of Lemo21 (DE3) cells transformed with cc_dbph1/pOPINE-F and cc_dbph2/pOPINE-F. Each lane triplet is occupied from left to right by samples from 37°C Induction, 16°C Induction and Control, respectively. Used marker (M) was the Precision Plus Protein™ All Blue Standards (BioRad, USA).

Figure 2.23 SDS-Page of soluble fractions of protein extracts of Rosetta 2 (DE3) and BL21 (DE3) cells transformed with cc_dbph1/pOPINE-F and cc_dbph2/pOPINE-F. Each lane triplet is occupied from left to right by samples from 37°C Induction, 16°C Induction and Control, respectively. Used marker (M) was the Precision Plus Protein™ All Blue Standards (BioRad, USA). Lanes with asterisks show a plausible high-level production of the protein of interest.
Figure 2.24 SDS-PAGE of insoluble fractions of protein extracts of Rosetta 2 (DE3) and BL21 (DE3) cells transformed with cc_dbph1/pOPINE-F and cc_dbph2/pOPINE-F. Each lane triplet is occupied from left to right by samples from 37°C Induction, 16°C Induction and Control, respectively. Used marker (M) was the Precision Plus Protein™ All Blue Standards (BioRad, USA). Lanes with asterisks show a plausible high-level production of the protein of interest, without a relevant “leaky” expression from the control sample.

Loading the same volumes of protein extracts from different induction conditions was an inconvenient approach, because different samples had different cellular concentrations and therefore different protein amounts in the same loaded volume. For the future applications of this method an O.D. reading of each culture or a BSA assay of the protein extracts was used in order to normalize the amount of protein that was loaded (see Chapter 3 and 4). Despite this it was clearly visible that CcDBPH1-pOPINE had been successfully produced at visible levels on the SDS-PAGE by Rosetta 2 cells at 16°C long-term induction, while CcDBPH2-pOPINE had been successfully produced by BL21 cells at 37°C short-term induction.

The production of soluble heterologous proteins is considered to be of paramount importance in order to have biologically active proteins (Ghosh et al., 2004), in comparison to their precipitation in insoluble protein aggregates which is usually determined to result in the loss of their native structure (McCormick et al., 2014). Thus, even if desired protein production was also observed in many Rosetta and BL21 insoluble fractions the conditions that were selected as optimal for the 1L scale-up heterologous expression were: Rosetta 2 (DE3) at 16°C for 16h for CcDBPH1-pOPINE and BL21 (DE3) at 37°C for 2-3h for CcDBPH2-pOPINE.
Before proceeding with 1L growth and purification attempts, the observed heterologous protein production was confirmed via Western-Blot analysis, as described in General Methodologies, par.M14. 1-2 µL of the soluble fractions of CcDBPH1-pOPINE/Rosetta2/16°C and CcDBPH2-pOPINE/BL21/37°C were loaded adjacent to their control sample (not induced, NI). The protein amount that was loaded was in each case normalized using the BCA protein assay (Thermo Fisher Scientifics, USA) following the manufacturer’s instructions and 96 well plate assay protocol. The total protein concentration of the Rosetta samples was: 0.87 mg/mL for the 16°C sample, and 1.24 mg/mL for the NI sample. The total protein concentration for the BL21 samples was: 0.68 mg/mL for the 37°C sample, and 1.02 mg/mL for the NI sample. Approximately 1 µg of protein extract was loaded in each lane. Photo of the detected filter was taken with ChemiDoc™ and can be seen in Figure 2.25 (panel A).

![Figure 2.25](image)

**Figure 2.25** ECL Chemoluminescent resolution of Western-Blot transfer of CcDBPH1 and CcDBPH2. A. Western-Blot of the CcDBPH1/pOPINE and CcDBPH2/pOPINE soluble protein extracts from the expression screening performed in Rosetta 2 (DE3) and BL21 (DE3) cells, respectively; CcDBPH1/pOPINE protein extracts are from the 16°C and the not induced (NI) conditions, CcDBPH2/pOPINE protein extracts are from the 37°C and the not induced (NI) conditions. B. Western-Blot of the CcDBPH1/pOPINE and CcDBPH2/pOPINE soluble protein extracts from the 1L culture performed in Rosetta 2 (DE3) and BL21 (DE3) cells, respectively; CcDBPH1/pOPINE protein extracts are from the 37°C, 16°C and the NI conditions, CcDBPH2/pOPINE protein extracts are from the 37°C, 20°C and the NI conditions. White arrows show the lanes with the highest amount of protein after normalization with BCA protein assay. Used marker (M) was the Precision Plus Protein™ All Blue Standards (BioRad).

In both Rosetta 2 and BL21, a faint induction of the His-tagged proteins of the expected molecular weight in NI samples could be detected. However, in 16°C and 37°C CcDBPH1 and CcDBPH2 protein amounts were much more abundant (even more so considering that total protein concentration of NI samples was higher in comparison to 16°C/37°C samples).
For each of these selected conditions, 1L volume was prepared as described in General Methodologies, par. M14.

For both genes the 1L volume was split in two 500 mL Erlenmeyer flasks and GOI expression was induced following both short-term and long-term induction, in order to better study the plasmid expression pattern in bigger volumes. As in the previous experiments, pellets were harvested and 1 mL aliquots of different treatments were lysed and their soluble fractions normalized using the BCA protein assay and analysed on Western-Blot before starting with protein purification. Also in this case 1 µg of protein extract was loaded in each lane. Photo of the detected filter was taken with ChemiDoc™ and can be seen in Figure 2.25 (panel B).

In this scale-up culture experiment, CcDBPH1 and CcDBPH2 production patterns differed in comparison to what was observed in previous expression screening and Western-Blot. For both proteins the NI expression levels was higher, indicating that the protein production in high volumes was “leakier” in comparison with the expression screening. This could be due to the fact that the LB medium of the overnight culture was not enriched with 1% glucose, which would lower cAMP levels in the cell (Grossman et al., 1998; Briand et al., 2016), and consequently T7 polymerase production. Moreover, while in the case of CcDBPH1 short- and long-term induction times did not increase the relative amount of produced POI per µg of total protein. In the case of CcDBPH2 the short- term induction did not improve POI production in comparison to NI, while the long-term induction showed a higher level of protein, contrary to what was observed in the screening, where CcDBPH2 production at 16°C was practically undetectable on the polyacrylamide gel. These different results are not unusual: switching from a 3 mL volume-scale of expression screening to 500 mL/1L volumes implies experimental changes that could easily cause the observed differences. Even if 1% glucose could have been added to the next scale-up experiments in order to avoid protein expression in NI samples, it was decided to focus on the protein purification independently, using all the harvested pellets in which the POIs were detected to test different purification buffers and protocols.

2.3.9 Enzyme purification of DBPH-pOPIN proteins

Purification of CcDBPH1 and CcDBPH2 was performed as reported in paragraph 2.2.4. Considering that the described method was the result of different trials, some considerations on the selected final method are discussed here. First purification approaches based on chemical lysis with NPI-10 Tween buffer and freeze cycling (3-4 cycles of rapid freezing and thawing of the suspended pellet with dry ice and at 42°C, respectively) caused the protein of interest to rapidly precipitate in the insoluble fraction before binding with the Ni-NTA resin, resulting in a poor final protein yield.
Generally speaking, sonication-based pellet lysis is more efficient than chemical lysis, but the consequent heat generation can easily cause protein denaturation during the process with a consequent loss of proteins in the soluble fraction (Islam, Aryasomayajula and Selvaganapathy, 2017). The addition of a non-ionic detergent has been shown to be effective in preventing this outcome (Johnson, 2013), thus Triton-X-100 was selected for its minor tendency to form bubbles following agitation in comparison to Tween 20.

The buffer used for CcDBPH1 had BME added as a reducing agent before bringing the buffer to pH 7.4, because of CcDBPH1 having a predicted isoelectric point of pH 8.8. Similarly, the buffer used for CcDBPH2 had DTT added as a reducing agent before bringing the buffer to pH 8, because of CcDBPH2 having a predicted isoelectric point of pH 6.36. The general idea was to use a buffer with a pH of at least one point distant from the predicted isoelectric point pH of the protein and selecting a reducing agent (added in order to prevent oxidative stress and unwanted sulphide bonds) that was capable of remaining soluble at the designed pH.

Finally, after transferring elution aliquots to a Sephadex G-25 column (needed to remove detergent), glycerol was added to the final purification buffer, which has a stabilizing and cryoprotectant effect.

SDS-Page gels of CcDBPH1-2 purifications were visualized via ChemiDoc™ and images generated are shown in Figure 2. 26.
Following the rationale explained above, CcDBPH1 and CcDBPH2 were successfully extracted from the Rosetta/BL21 pellets, and the contribution of detergents and reducing agents resulted in the majority of the protein remaining in solution (the CcDBPH2 insoluble fractions show bands of the CcDBPH2 molecular weight, CcDBPH1 insoluble fractions did not seem to contain the protein of interest). On the other hand, the buffer change between Ni-NTA and G-25 buffer caused some differences in the final protein retention between CcDBPH1 and CcDBPH2.

CcDBPH1 was found in different G-25 elution fractions, and composed, in percentage, the most abundant protein in the protein mixture. CcDBPH2, on the other hand, co-purified with an abundant protein of 25 kDa from *E. coli*. 

---

**Figure 2.26** SDS-PAGE of the different steps of protein purification of CcDBPH1-2. A. CcDBPH1 purification from long term-induced Rosetta 2(DE3) pellets. IN: Insoluble fraction; LD: Loaded fraction; FW: Flowthrough fraction; P2-P3: 2°-3° Ni-NTA Elution fraction; G4-G9: 4°-9° Sephadex G-25 Elution fraction. B. CcDBPH2 short term-induced BL21 (DE3) pellets. IN1-IN2: Insoluble fraction (before and after resuspension of the insoluble pellet in water); LD: Loaded fraction; FW: Flowthrough fraction; P2-P4: 2°-4° Ni-NTA Elution fraction; G1-G6: 1°-6° Sephadex G-25 Elution fraction. Used marker (M) was the Precision Plus Protein™ All Blue Standards (BioRad, USA). The white arrows show lanes where fractions enriched in the protein of interest were loaded.
2.3.10 Estrase and DBP-degrading assays

Partially purified CcDBPH1 and CcDBPH2 were tested for their esterase activity against p-nitrophenyl butyrate (p-NPB), as reported in paragraph 2.2.5. Among the different commercially available 4-nitrophenol esters the p-nitrophenyl butyrate was selected for its structural similarity with DBP, that was hypothesized to be the PE that would be most efficiently degraded by *C. closterium* and its enzymes. CcDBPH1 and CcDBPH2 were tested several times, both to compare their respective esterase activity with this generic substrate and to check if conservation at -20°C affected their activity in comparison to samples stored at 4°C for a week. The degradation efficiencies of CcDBPH1 and DBPH2 are shown in Figure 2.27.
Figure 2.27 p-NPB degradation of CcDBPH1 and CcDBPH2. A. Comparison of p-NPB degradation between CcDBPH1 aliquots stored 7 days at 4°C and -20°C. B. Comparison of p-NPB degradation between CcDBPH2 aliquots stored 7 days at 4°C and -20°C. C. Comparison of p-NPB degradation between CcDBPH1 and CcDBPH2. Degradation of p-NPB due to its autohydrolysis was used as control in all of the three experiments. Absorbance is expressed in mAU (milli-absorbance unit).
The graphs clearly show how CcDBPH1 esterase activity was far more efficient than that of CcDBPH2 and, at the same time, that glycerol was successful in helping to cryopreserve protein activity while it was stored at -20°C. Given that neither of the two protein purification solutions is monodispersed, it is difficult to say if the difference in activity between CcDBPH1 and CcDBPH2 could be related to a different degradation capability of the two esterases or to their different relative concentration in the partially purified protein samples (CcDBPH1 was far more represented in its protein mixture than CcDBPH2).

Even if protein modelling has evidenced a favourable affinity energy with DBP for CcDBPH1 (−31.0 kJ/mol) and not for CcDBPH2, an absolute correlation of CcDBPH1-2 with the observed esterase activity cannot be inferred. In fact, it is difficult to say if there could be other esterases present in the purified protein mixture that may be contributing to p-NBP degradation.

Moreover, the incubation of the CcDBPH 1-2 solutions with DBP, as described in par. 2.2.6, did not show any substrate degradation, in either the CcDBPH1-2 + DBP incubations, or in the combined CcDBPH1 + CcDBPH2 + DBP incubation. Also in this case it is impossible to correlate this lack of esterase activity to a unique cause. Possible hypotheses are: 1) purified CcDBPH1-2 retain their esterase activity but a partial denaturation (or the absence of unknown post-translation modifications) hampers the interaction with DBP; 2) CcDBPH1-2 work together and are cooperative involved in PE-degradation (as already suggested in par. 2.3.3) and individual recombinant enzymes cannot properly interact with DBP, or functionally interact with each other when added to the same solution; 3) co-purified proteins interfere with the interaction between CcDBPH1-2 and DBP; 4) the selected reaction conditions are not compatible with the reaction. This fourth scenario is the least plausible reason, because the buffer used and incubation times are similar to those used in other publications for other PE esterases (Jiao et al., 2013; Huang et al., 2019; Wu et al., 2019; Qiu, Yang, et al., 2020).

Regardless of the hypotheses outlined, future biochemical characterizations and activity assays of CcDBPH1-2, as well as their comparison with previously characterized bacterial and fungal PE hydrolases, will need a monodisperse protein solution in order to provide robust data.
PE-degrading esterases were selected as the first target proteins in the thesis. The aim was to investigate the enzymatic role of *C. closterium* DBP degradation that had previously been reported in the literature (Gao and Chi, 2015; Zhang *et al.*, 2019), and, more generally, to elucidate the presence of PE-degrading enzymes in microalgae, that had not previously been discovered or characterized. The direct application of bacterial PE-degrading enzymes in PEs bioremediation has already been discussed (Ren *et al.*, 2018) and the direct application of similar microalgal enzymes in the bioremediation field of microalgal biotechnology appeared promising.

Using the protein sequences of all the previously identified PE esterases as queries, different esterase sequences were identified in the *C. closterium* transcriptome. Among these the most promising sequences were considered the ones that matched with the highest number of queries and with the highest Bit-scores. The two highest matching sequences were found in the same transcript as an uncommon polycistronic unit. The relative proteins were provisionally called CcDBPH1-2 and were elected as POI for the experimental pipeline.

Experiments were subsequently undertaken following three different approaches: 1) the main approach included CcDBPH1-2 CDS isolation, their cloning in the TOPO-TA and pOPIN expression vectors, their heterologous expression and esterase activity assessment; 2) the second approach included further bioinformatic analyses and the modelling of the CcDBPH1-2 3D structure and their docking with DBP; 3) the third approach involved the growth of *C. closterium* with the addition of different DBP concentration to the medium, and the acquisition of samples for RNA extraction and RT-qPCR experiments aimed at examining potential differential expression of cc_dbph1-2 in the presence of DBP. The second approach corroborated the initial bioinformatic findings (in particular for DBPH1, whose docking simulation yielded favourable affinity for DBP), while the RT-qPCR analyses highlighted an increment in the expression levels of both genes in correlation both to DBP concentration and exposure time. This experimental part was included in a recently published work for Science of the Total Environment (Vingiani *et al.*, 2022).

Regarding the first approach, the CDS isolation of cc_dbph1-2 and their cloning in TOPO-TA and pOPIN vectors was successfully achieved, and via expressions screening, ideal conditions for the 1L growth and recombinant proteins production were identified. Both CcDBPH1-2 were partially purified by designing ad hoc purification protocols and their esterase activity was assessed using the generic esterase substrate p-NPB. However, even if CcDBPH1 esterase activity appeared more efficient than CcDBPH2, a more complete purification of the enzymes will be required to generate definitive data on their activity. Moreover, partially purified proteins incubated with DBP did not show any degradation activity.
For this enzyme family and its biotechnological potential different paths could be followed for future approaches: 1) the purification design of CcDBPH1-2 should be optimized (potentially using automated instruments such as the ÄKTA chromatography systems (Cytiva) to obtain monodispersed protein solutions. Once pure proteins are obtained their biochemical properties can be characterized (e.g. optimum temperature of the enzyme reaction, thermostability, pH optimum activity, effect of metal ions or organic solvents), and precise kinetic measurements of their potential PE-degrading activity can be measured.

2) Alternatively, measurements of PE-degrading activity could be attempted directly using lysates of bacterial cellular pellets that expressed CcDBPH1-2 or both proteins at the same time. In this way the eventual partial denaturation caused by the purification can be circumvented, and, in the case of the combined production of CcDBPH1 and CcDBPH2, the theorized cooperation of the two proteins could also be achieved. The co-production of CcDBPH1 and CcDBPH2 can be ideally obtained by the cloning of both GOI in the same vector. In fact, all pOPIN vectors use the same pUC replication origin (shared by the vast majority of commercially available plasmids), making the co-transformation of two plasmids in the same bacterial cell an unpredictable and uncommon event. On the other hand, cc_dbph1/pOPINF has an HindIII restriction site at the 3’ of the GOI, and can be linearized without affecting the GOI sequence. At the same time, an ad hoc designed cc_dbph2 CDS sequence could be produced via PCR and cloned in cc_dbph1/pOPINF via homologous recombination. This CDS should have, other than the sequence adaptors, a 3’ C-terminus His-tag (needed for affinity purification) and a 5’ Shine-Dalgarno sequence (that could favour ribosome binding). The plasmid map of this cc_dbph1-2/pOPINF construct was produced using SnapGene and is shown in Figure 2.28.
Figure 2.28 Plasmid map of cc_dbph1-2/pOPINF and cc_dbph2/pOPINE-F produced using SnapGene. Inserted gene DNA is shown in red and the related proteins in dark purple (CcDBPH1) and light purple (CcDBPH2). The plasmid shows its size (bp), the lac operator (blue), the T7 promoter (arrowhead white), the replication origin (ori, yellow), the Ampicillin resistance gene (AmpR, light green), the C-terminus His-tag (6xHis, arrowhead light purple), the N-terminus His-tag (6xHis, light purple), the cleavage site for the protease 3C (dark blue) and the Shine-Dalgarno sequence (grey). The position of the primers used for the PCR amplification of cc_dbph2 CDS is also shown.

Overall, this chapter represents the first attempt to identify and characterize PE-degrading enzymes in marine microalgae, and specifically in the marine diatom Cylindrotheca closterium. Bioinformatic and RT-qPCR analyses suggest the potential of C. closterium and its enzyme system in bioremediation strategies targeting PEs. Finally, the recombinant protein production and purification, while not yet completed, still represents the first step in the discovery of microalgal PE-degrading enzymes, biochemical properties and activity kinetics, and will surely benefit future approaches in this research field.
Chapter 2 References:


Chapter III

First Identification of Platelet Activating Factor – Acetylhydrolases in the Marine Diatom Cylindrotheca closterium

Graphical abstract of the experimental pipeline described in Chapter 3.
Chapter 3  Abstract

Many microalgal extracts have shown anti-inflammatory activity. However, the inconsistency between the methods used and the difficulty in correlating such an activity to a specific metabolite or protein has somewhat limited the advancement of such discoveries for potential biotechnological applications. *Cylindrotheca closterium* extracts have previously been shown to have anti-inflammatory activity, possibly linked to lysophosphatidylcholine and pheophorbide, while the extracts under nutrient starvation condition lost this activity. Searching the *C. closterium* transcriptome for enzymes that could be related to this activity, two transcripts annotated as “platelet activating factor–acylhydrolase” (PAF-AH) were evaluated as the most interesting targets. In fact, PAF-AH is mainly involved in the metabolism of the phospholipid PAF, a potent inflammatory mediator, but in different organisms its phospholipase activity has also been shown to be involved in morphogenesis, cancer development and lipid metabolism processes. Given the multifaceted role of PAF-AHs and the lack of data on their role in microalgal lipid metabolism and overall physiology, the possibility of investigating these transcripts in *C. closterium* appeared promising.

The CDS of these two genes (prov. called *cc_paf-ah1* and *cc_paf-ah2*, with the related proteins CcPAF-AH1 and CcPAF-AH2) were isolated via PCR and different strategies were developed to clone them in pOPIN expression vectors. However, CcPAF-AH1 and CcPAF-AH2 lacked visible expression in different bacterial hosts and under different induction conditions. A second expression screening system was set-up at the Biomedical Sciences Research Complex (BSRC) at the University of St. Andrews (Scotland, UK) as part of a 1-month EU-ASSEMBLE+ Transnational Access (TNA) 8th Call project, in collaboration with Dr. Clarissa Melo Czekster and Dr. Rafael Guimaraes da Silva. The aim of the TNA was to express CcPAF-AH1-2 using pJ411 expression vectors, in order to obtain high-scale production of these protein and to undertake related crystallographic data.

At the BSRC, an abundant expression of CcPAF-AH2 was obtained in the insoluble fraction, but at the same time, difficulties were encountered in replicating expression in large-scale 1L cultures; with growth under these conditions resulting in the production of toxic proteins (C43 cells) and the correct formation of disulfide bonds (SHuffle T7 cells) was completely impaired. Consequently, the protein was evaluated to be highly toxic to the bacterial cells and, in order to obtain structural data regarding this novel class of phospholipases, a mutation of the catalytic Serine 243 was initiated. This way, a conspicuous amount of CcPAF-AH2 was observed in the insoluble fraction of 1L cultures of SHuffle T7 and Rosetta (DE3) cells. The persistent presence of the protein in the insoluble fraction suggests a tight association of CcPAF-AH2 with the cellular membrane. Future purification approaches will take into account this lipophilic characteristic of CcPAF-AH2 as well as its high toxicity, an uncommon trait in comparison with other PAF-AHs identified from other organisms.
3.1 Introduction

3.1.1 Inflammation and the role of PAF in the inflammatory response

Inflammation is a biological protective response in body tissues to harmful stimuli, associated with innate immunity, involving immune white cells and molecular mediators. At a tissue level, inflammation is characterized by redness, swelling, heat, pain, and loss of tissue function. These characteristics are so evident and common that they were already listed by the Latin scholar Celsus as rubor, calor, tumor, dolor and functio laesa, respectively (Ayoub, 2010). The final role of the inflammatory response is to counteract the initial cause of insult and clear necrotic cells and damaged tissues; the specifics of this response depend on the precise nature and location of the initial detrimental stimulus. Inflammation is always characterized by the following common steps: 1) cell surface receptors stimulus recognition, 2) activation of the inflammatory pathways, 3) release of the inflammatory markers and 4) recruitment of the inflammatory cells (Chen et al., 2018).

Inflammatory markers are small secreted protein cytokines (e.g. Interleukins (IL-1β, IL-6), tumour necrosis factor-α (TNF-α)) and other molecules (proteins, enzymes, phospholipids); that recruit and activate effector cells such as neutrophils, monocytes, lymphocytes and mast cells (Van Linthout, Miteva and Tschöpe, 2014) to the site of inflammation, cause the initiation of the inflammatory response, resulting in the release of a vast array of second-step inflammatory mediators, and, finally, the consequent resolution of the inflammation. Inflammation is a process which is normally tightly regulated both from a spatial and temporal point of view. The impairment of this regulation brings a status of chronic inflammation (Headland and Norling, 2015), whose effects are linked to many diseases, such as: cardiovascular diseases (CVD), atherosclerosis, diabetes mellitus, chronic kidney disease, rheumatoid arthritis, different forms of cancer, as well as other autoimmune and neurodegenerative morbidities (Chen et al., 2018; Furman et al., 2019). Overall, chronic inflammatory diseases are the most significant cause of death in the world (Fleit, 2014).

Among the different inflammatory markers, the platelet activating factor (PAF) is subject to very complex regulation. In fact, PAF is involved in a vast array of processes that encompass not only physiological and chronic inflammation, but also include apoptosis, wound healing, reproduction, angiogenesis, and related chronic diseases such as cancer, renal diseases, cerebrovascular and central nervous system disorders, allergies, asthma and sepsis (Lordan et al., 2019). The structure of PAF includes a glycerol backbone like all phospholipids, with an alkyl group at the sn-1 position, an acetyl group at the sn-2 position and phosphocholine at the sn-3 position (sn-1,2,3 is the nomenclature given to glycerol carbons (Ashraf and Nookala, 2020)). The PAF structure is shown in Figure 3.1.
Figure 3.1 Structure of platelet-activating factor (PAF); A. PAF structural model drawn with ChemDraw; B. PAF ball and stick model data from PubChem (circled in red is the glycerol backbone). Carbon atoms are shown in grey, oxygen atoms in red, phosphorus atoms in orange, nitrogen atoms in blue.

PAF itself is synthesised constitutively or under certain stimuli by a variety of cells such as endothelial cells, platelets, mast cells, macrophages, monocytes and granulocytes (Lordan et al., 2019). The *de novo* synthetic pathway of PAF, that involves a series of reactions via transacylases, phosphohydrolases and choline-phosphotransferase (PAF-CPT), starts from alkyl-lyso-glycerophosphate involved in the maintenance of constitutive low basal levels of PAF; alternatively the “remodelling” synthetic pathway starts with a molecule of phosphatidylcholine, where a sn-2 arachidonic fatty acid acyl residue is substituted with an acetyl residue, and is responsible for the increase in PAF production in acute and chronic inflammation (Ashraf and Nookala, 2020). This “remodelling” synthesis is mainly performed by endothelial cells and neutrophils that possess high levels of phosphatidylcholine. In physiological conditions, the PAF pro-inflammatory effect is blocked by PAF receptor (PAFR) desensitization and by the effect of a highly specific family of intracellular and extracellular phospholipases A₂, called PAF acetylhydrolases (PAF-AH; (Yost, Weyrich and Zimmerman,
that hydrolyses PAF at the sn-2 position, generating lyso-PAF and acetic acid, and biologically inactivating PAF (Dong, Li and Wu, 2021). PAF-AH was historically considered as a negative regulator of PAF pro-inflammatory effects (Tjoelker et al., 1995). However, its release of short or oxidised fatty acids linked to its hydrolytic activity was also found to be linked to pro-inflammatory effects (Addis et al., 1995); this resulted in an intense debate regarding PAF-AH as a pro- or anti-inflammatory enzyme, which is still ongoing (Chen, 2004; Marathe et al., 2014; Stafforini and Zimmerman, 2014; Lordan et al., 2019).

3.1.2 Microalgal anti-inflammatory activity

As previously described in Chapter 1, microalgae have been extensively studied for their production of valuable bioactives (e.g. β-carotene, lutein, astaxanthin, long-chain PUFAs, oxylipins, sulphated polysaccharides) with beneficial effects for human health and, in particular; for the anti-inflammatory activity of various microalgal extracts/compounds (Lauritano et al., 2020; Saide, Lauritano and Ianora, 2020; Saide et al., 2021). Among these bioactives, the most studied compound is astaxanthin(3,3′-dihydroxy-β-carotene-4,4′-dione), which is a bright red secondary carotenoid, and is considered the most powerful anti-oxidant in nature (54 times higher than β-carotene, 65 times higher than vitamin C (Gong and Bassi, 2016)), due to the presence of a keto- and hydroxyl group on each end of its structure; making it extremely efficient in quenching free radicals (Higuera-Ciapara, Félix-Valenzuela and Goycoolea, 2006). In particular, the freshwater microalga Haematococcus pluvialis is known to accumulate high amounts of astaxanthin under different environmental stress conditions and the commercial growth of H. pluvialis and asthaxantin production is a well-defined biotechnological field, with a global market revenue which has been estimated to be worth USD 1 billion by 2025 (Oslan et al., 2021).

Astaxanthin has been tested in vitro for its capability to inhibit the production of pro-inflammatory cytokines in different cell lines (Zhang et al., 2014; Choo et al., 2020); moreover, in rat models, a supplementation of 100 mg/kg astaxanthin was as strong as that of a 10-mg/kg dose of prednisolone, an anti-inflammatory drug, in anti-inflammatory effects (Ohgami et al., 2003). Similarly, anti-oxidative properties have also been observed for other carotenoids (Esteban et al., 2015), such as lutein (Jahns and Holzwarth, 2012) and sulphated polysaccharides (Sun, Wang and Zhou, 2012). There is a clear correlation between oxidative stress and various inflammatory skin diseases such as vitiligo (Qiu, Song and Setaluri, 2014), with reactive oxygen species (ROS) that act directly as inflammatory markers and mediators (Choo et al., 2020).

However, while many experiments have been performed by adding chemicals or manipulating metabolic fluxes to induce oxidative stress in microalgae and increase the amounts of antioxidant bioactives produced (Yu, Chen and Zhang, 2015; Montero-Lobato et al., 2018), the commercial feasibility of producing such anti-
inflammatory compounds is still far from being optimized in the vast majority of microalgae that have to date been studied.

There are numerous reports in the literature on cellular extracts that are capable of inducing the reversion of the inflammatory phenotype (often induced via the addition of lipopolysaccharides (Li et al., 2005)) in many different cell lines. For example, Lauritano et al. identified extracts from *Cylindrotheca closterium* FE2/1, *Odontella mobiliensis* FE326/1, and *Pseudo-nitzschia pseudodelicatissima* FE1098_1/1 as being capable of blocking the release of the inflammation marker TNF-α in acute monocytic leukaemia THP-1 cells (Lauritano et al., 2016). Surprisingly perhaps, this activity was lost when testing extracts from microalgae that had been grown under nutrient starvation conditions, implying that the induction of stressful conditions and of ROS accumulation in microalgae does not always lead to an increased accumulation of anti-inflammatory compounds (Montero-Lobato et al., 2018). More recently ethanol-extracts of *Chloromonas reticulata* were tested in RAW 264.7 macrophages and displayed anti-inflammatory activity (Suh et al., 2019).

However reports in the literature in this field present several limitations: 1) the anti-inflammatory potential of microalgal extracts is implied via the analysis of a few inflammatory markers, and this is far from representative of the complex inflammatory signal network; 2) the studies applied to animal models are much less abundant (Ávila-Román et al., 2014; Gutiérrez-Pliego et al., 2018) and there is a complete absence of clinical trials; 3) it is usually very difficult to correlate the activity of cellular extracts to a specific compound, even if transcriptomic (Elagoz, Ambrosino and Lauritano, 2020) and dereplication (Lauritano et al., 2020; Shiels et al., 2021) approaches have both generated interesting results. Specifically, the transcriptomic approach can elucidate the identity of key transcripts involved in the production of anti-inflammatory compounds, and their differential expression under different culture conditions; and on the other hand, dereplication approaches can be used to investigate the biochemical nature of anti-inflammatory compounds. The contribution of microalgae to solving problems related to inflammation have mainly been correlated to their lipid fractions (or lipid isolates, such as the ω-3 PUFAs EPA and DHA (Conde et al., 2021)); however, other potential anti-inflammatory compounds have been identified, such as pheophorbide a from *Cylindrotheca closterium*, as previously mentioned in Chapter 2.

Finally, it is worth mentioning how dietary microalgal polar lipids, especially those of marine sources, such as phosphatidylcholine (PC), and digalactosyldiacylglycerol (DGDG) from different marine *Chlorococcum sp.* strains have shown the potential to compete with PAF for the active site of its receptor PAFR (Shiels et al., 2021), thus blocking its cascade of pro-inflammatory signals, which can result in different morbidities as previously mentioned. In fact, microalgal-based supplements have recently seen an increase in scientific interest also due to the potential of the dietary assumption of PAF-inhibitors to prevent and treat inflammation-based cancer and other morbidities (Lordan et al., 2019).
3.1.3 Previous evidences of platelet activating factor-acetylhydrolase (PAF-AH) enzymes

While PAF-AH is mainly known for its involvement in the metabolism of PAF, and consequently in its ambiguous pro/anti-inflammatory role; it is an underestimation to consider that its role is limited to the inflammatory pathway. The three isoforms of PAF-AHs that have been characterized in humans are conventionally named PAF-AH I, PAF-AH II, and plasma PAF-AH (Lp-PLA2 or pPAF-AH), with PAF-AH I and PAF-AH II being intracellular while pPAF-AH is the extracellular isoform; associated with low-density or high-density lipoproteins (LDL and HDL, respectively (Karasawa and Inoue, 2015)). Schaloske and Dennis classified these phospholipase A₂ isoforms as groups VIII, IIB, and IIA, respectively (Schaloske and Dennis, 2006).

Although PAF is recognized as the primary substrate for these enzymes, the structural requirements for substrate recognition are different among the different isoforms. pPAF-AH and PAF-AH II hydrolyze short-chain diacylglycerols, triacylglycerols and acetylated alkanols (Min et al., 2001). The minimum structural requirement for a substrate was reported to be a glyceride derivative that includes an sn-2 ester and a reasonably hydrophobic chain in the position occupied by the sn-1 chain (Kono and Arai, 2019). On the other hand, PAF-AH I has shown poor activity towards PAF derivatives that have sn-2 residues that differ from the acetyl-group (Hattori et al., 1995). While PAF-AH I is a heterotetrameric enzyme with low homology across different organisms, PAF-AH II and pPAF-AH are monomeric enzymes of c.a. 40 kDa that share 43% identity (Hattori et al., 1995) and with an amino acid sequence that is highly conserved across different species. All PAF-AH II and pPAF-AH identified to date include a lipase/esterase motif(G-X-S-X-G), a S/D/H catalytic triad and independence from Ca²⁺ cofactors. Moreover, all the PAF-AH II identified to date also share a consensus motif for N-myristoylation (NH₂-Met-Gly-X-X-Ser-) and consequently are localized in the membrane. PAF-AH II has been reported to have three catalytic activities: lysophospholipid transacetylase (TAₗ), sphingosine transacetylase (TAₛ), and acetylhydrolase (AH) activity, with the TA activity that transfers the acetyl group from PAF to lysophospholipids and sphingosine, producing different PAF derivatives and expanding the biological functions of PAF. Both TA and AH activities require the same S/D/H catalytic centre, while the TA activity is also related to the G2 and C120 residues in humans (Bae et al., 2000).

Other than in PAF degradation, PAF-AH II may function as a protector against oxidative stress-induced cell injury and dysfunction, via its release of sn-2 acyl residue in phospholipids, undergoing peroxidation due to the presence of free radicals in the cellular environment. In fact, involvement in the inflammation phenotype of truncated oxidized phospholipids has been widely studied (Davies and Guo, 2014; Karki and Birukov, 2020), and the removal of these oxidation products is essential for protection of the cell from oxidative damage. Even the unicellular organism *Saccharomyces cerevisiae*, that does not use PAF as a signal molecule involved in inflammation, uses the acetylhydrolase activity of PAF-AH II to avoid the deleterious effects of oxidative
damage (Foulks et al., 2008). Finally, the presence of pro-oxidative (such as ROS) or anti-oxidative compounds have been correlated with the membrane and cytoplasmic translocation of PAF-AH II, respectively (Thévenin et al., 2011).

On the other hand, integral oxidized phospholipids, such as DHA, EPA and AA, have been found to be strong anti-inflammatory mediators (Feige et al., 2010), acting as precursors for a series of bioactive lipids which are involved in the inhibition of mast cells activation, neutrophil migration and other hypersensitivity mechanisms, via their binding with the PPARγ nuclear receptor (Dong, Li and Wu, 2021). The enzymatic role of PLA2, and particularly of PAF-AH II, is to maintain sufficient levels of these lipid metabolites and mediators.

With its ability to hydrolyse oxidized truncated or integral acyl chains of membrane phospholipids and its consequent anti-oxidative and anti-inflammatory roles, PAF-AH II may have broader patho/physiological functions that still need to be fully elucidated. Even if it is not possible to speculate on the function of PAF-AH II in microalgae, it is clear that it does have a pivotal role in the metabolism of short-length phospholipids or those having oxidatively modified sn-2 chains.

Given the aforementioned multifaceted role of PAF-AH IIs and the lack of data regarding their role in microalgal lipid metabolism and in their overall physiology, the possibility to investigate two different transcripts of *C. closterium* annotated as “platelet activating factor – acetylhydrolase” appeared promising in order to answer both biotechnological and ecological question regarding this enzyme. To my knowledge, PAF-AHs have never been characterized in microalgae and this chapter provides the first report on this subject. The discovery and annotation of sequences from microalgal sources may allow *paf-ah* genes to become a target for genetic modification, that is nowadays one of the most used methods to develop microalgal strains optimized for the production of value-added lipids (Vingiani et al., 2019).

### 3.2 Methods

#### 3.2.1 *cc_paf-ah1-2 oligo design and reaction procedures for PCR*

The Blast2GO software annotated both the transcripts “TR16369|c0_g1_i1 length: 1566 nt” and “TR12571|c0_g1_i1 length: 1460 nt” as “platelet activating factor – acetylhydrolase” The CDS contained in these transcripts were provisionally called *cc_paf-ah1* and *cc_paf-ah2*, respectively, and selected as genes of interest (GOIs). *cc_paf-ah1* CDS has a length of 1362 bp, *cc_paf-ah2* CDS has a length of 1146 bp.

Primers for endpoint PCR and for sequencing reactions were designed as reported in General Methodologies, par. M5 and M9. Sequences of used primers for PCR are reported in Table 3.1.
Table 3.1 Primer sequences of cc_paf-ah1 and cc_paf-ah2 genes used for CDS isolation, Sanger sequencing and cloning in pOPIN and pJ411 vectors. For each gene of interest forward and reverse primers are listed, as well as amplicon size. In the case of pOPIN/pJ411 cloning primers the adaptor tail is evidenced in **bold**.

<table>
<thead>
<tr>
<th>Gene of Interest - Application</th>
<th>Forward Primers</th>
<th>Reverse Primers</th>
<th>Length of Amplicons (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcPAF -AH1 - Sequencing</td>
<td>TCGATGCGACACCATTACC</td>
<td>TTCCTGCTCCTTGTTTTC</td>
<td>N.D.</td>
</tr>
<tr>
<td>CcPAF -AH2 - Sequencing</td>
<td>GCATCGGTTCCAGGAAATGA</td>
<td>TGGAGGCTTTGCTGGAAA</td>
<td>N.D.</td>
</tr>
<tr>
<td>CcPAF -AH1 - pOPINE Cloning</td>
<td>AGGAGATACCATGATTGATGATATCTGGCATTTC</td>
<td>ATGGTCTAGAAAGCTTTATTGGGCTTTTGCCATTTCC</td>
<td>1397</td>
</tr>
<tr>
<td>CcPAF -AH2 - pOPINE Cloning</td>
<td>AGGAGATACCATGTTGCAGGTAAACATATGTTGGTGATGATATCTGGCATTTC</td>
<td>ATGGTCTAGAAAGCTTTATTGGGCTTTTGCCATTTCC</td>
<td>1395</td>
</tr>
<tr>
<td>CcPAF -AH1 - pJ411_NTER Cloning</td>
<td>ATTTGTATTTCCAGGGGATGTTGGTGATGATATCTGGCATTTC</td>
<td>ATGGGGAACGCTTTATTGGGCTTTTGCCATTTCC</td>
<td>1179</td>
</tr>
<tr>
<td>CcPAF -AH2 - pJ411_NTER Cloning</td>
<td>ATTTGTATTTCCAGGGGATGTTGGTGATGATATCTGGCATTTC</td>
<td>ATGGGGAACGCTTTATTGGGCTTTTGCCATTTCC</td>
<td>1176</td>
</tr>
<tr>
<td>CcPAF -AH1 - pJ411_CTER Cloning</td>
<td>CTTTTAAGGGAATTTACATATGGTTGTGATATCTGGCATTTC</td>
<td>GTGGTGTTGGTGCCTCTGGGCTTTTGCCATTTCC</td>
<td>1180</td>
</tr>
<tr>
<td>CcPAF -AH2 - pJ411_CTER Cloning</td>
<td>CTTTTAAGGGAATTTACATATGGTTGTGATATCTGGCATTTC</td>
<td>GTGGTGTTGGTGCCTCTGGGCTTTTGCCATTTCC</td>
<td>1176</td>
</tr>
</tbody>
</table>

Different PCR conditions were tested for the amplification of each PCR fragment. For the amplification of cc_paf-ah1 and cc_paf-ah2 CDS the Expand™ High FidelityPLUS PCR was used; for the amplification of cc_paf-ah1 and cc_paf-ah2 CDS with the pOPIN/pJ411 adaptors the PrimeSTAR® GXL DNA Polymerase was used.

PCR reactions were set-up using the cycling conditions and PCR reagents listed in Table 3.2 and 3.3.

Table 3.2 PCR reaction setup for of cc_paf-ah1 and cc_paf-ah2 CDS isolation with pOPIN and pJ411 adaptors. Name of the PCR application, primer working pair, cycling conditions and PCR reagents are listed.

<table>
<thead>
<tr>
<th>Application</th>
<th>Working Pair</th>
<th>Cycling Conditions</th>
<th>PCR reagents</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcPAF-AH1 - CDS Isolation</td>
<td>FR1</td>
<td>95°C for 5 s; (95°C for 30 s; annealing at 56°C for 30 s and elongation at 72°C for 1:30 s) for 30 cycles</td>
<td>1U Xtra Taq Polymerase, 6 µL XtraTaq Pol White Buffer 5x, 0.2 mM dNTPs, 0.5 µM primers, 50 ng cDNA, MillQH₂O up to 30 µL</td>
</tr>
</tbody>
</table>
### Table 3.3 PCR reaction setup for of cc_paf-ah1 and cc_paf-ah2 CDS isolation with pOPIN and pl411 adaptors. Name of the PCR application, primer working pair, cycling conditions and PCR reagents are listed.

<table>
<thead>
<tr>
<th>Application</th>
<th>Cycling Conditions</th>
<th>PCR reagents</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcPAF-AH1 / pOPINE-F - Insert Amplification</td>
<td>98°C for 2 ; (98°C for 10s; annealing at 60°C*,** for 15s and elongation at 68°C for 1:30 ¤) for 34 cycles; terminal extension at 68°C for 7 ¦</td>
<td>1.25 U PrimeSTAR® GXL DNA Polymerase, 10 µL PrimeSTAR® GXL Buffer 5x, 0.2 mM dNTPs, 0.4 µM primers, 1 µL template cDNA (20 ng), MillQ H2O up to 50 µL</td>
</tr>
<tr>
<td>CcPAF-AH2 / pOPINE-F - Insert Amplification</td>
<td>98°C for 2 ; (98°C for 10s; annealing at 60°C*,** for 15s and elongation at 68°C for 1:30 ¤) for 34 cycles; terminal extension at 68°C for 7 ¦</td>
<td>1.25 U PrimeSTAR® GXL DNA Polymerase, 10 µL PrimeSTAR® GXL Buffer 5x, 0.2 mM dNTPs, 0.4 µM primers, 1 µL template Purified PCR dil. 1/50, MillQ H2O up to 50 µL</td>
</tr>
<tr>
<td>CcPAF-AH1 / pJ411_NTER-CTER - Insert Amplification</td>
<td>98°C for 2 ; (98°C for 10s; annealing at 54°C* for 15s and elongation at 68°C for 1:30 ¤) for 34 cycles; terminal extension at 68°C for 7 ¦</td>
<td>1.25 U PrimeSTAR® GXL DNA Polymerase, 10 µL PrimeSTAR® GXL Buffer 5x, 0.2 mM dNTPs, 0.4 µM primers, 1 µL template DNA (2 ng plasmid cc_paf-ah1/pOPINE + Miniprep product), MillQ H2O up to 50 µL</td>
</tr>
<tr>
<td>CcPAF-AH2 / pJ411_NTER-CTER - Insert Amplification</td>
<td>98°C for 2 ; (98°C for 10s; annealing at 57°C* for 15s and elongation at 68°C for 1:30 ¤) for 34 cycles; terminal extension at 68°C for 7 ¦</td>
<td>1.25 U PrimeSTAR® GXL DNA Polymerase, 10 µL PrimeSTAR® GXL Buffer 5x, 0.2 mM dNTPs, 0.4 µM primers, 1 µL template DNA (2 ng plasmid cc_paf-ah2/pOPINF Miniprep product), MillQ H2O up to 50 µL</td>
</tr>
</tbody>
</table>

*annealing temperature was selected based on the formula \( Tm (°C) = \left(\frac{[\text{the number of A and T}] \times 2}{} + [\text{the number of G and C}] \times 4\right) - 5 \), considering for the calculation only the part of the primers that annealed with the template.

**annealing temperature started at 45°C with a ramp rate of 1°C per each cycle up to the annealing temperature calculated as per * . Afterwards the annealing temperature followed with a ramp rate of 1°C per cycle up to 60°C, which was maintained as annealing temperature for the rest of the reaction.

### 3.2.2 CcPAF-AH1-2 heterologous expression and SDS-PAGE

Heterologous expression of CcPAF-AH1-2 was attempted a number of times, both at the Stazione Zoologica and at the Biomedical Sciences Research Complex (BSRC) at the University of St. Andrews.

Heterologous expression screening at the Stazione Zoologica was performed using the pOPIN vectors (pOPINE and pOPINF) and Rosetta 2 (DE3) pLySs strain as the heterologous expression host, using overnight growth at 20°C and 1-hour at 37°C induction conditions and 1 mM and 0.5 mM IPTG concentrations, as in the General Methodologies, par. M12. Total proteins were extracted from pellets and visualized on SDS-PAGE as in General Methodologies par. M13. The absence of the protein of interest (POI) was verified via Western-Blot, as in General Methodologies par. M14.

Heterologous expression screening at the BSRC was performed using the pl411 vectors (pl411_NTER, pl411_CTER) and BL21 (DE3), Rosetta (DE3), OverExpress™ C43 (DE3) (Lucigen, USA), Shuffle® T7 (NEB) strains as heterologous expression hosts, using an IPTG concentration of 0.5 mM. The induction temperatures (and expression times) that were tested include: 37°C (1:30h), 37°C (16h), 25°C (16h) and 16°C (16h). Promising combinations of plasmid-bacterial strain-induction temperatures were scaled up to 1L
growth volumes (with the addition of 1% glucose to the pre-inoculum and the growth volume), as in General Methodologies, par. M10.

In this case total proteins were extracted by adding 100 µL BugBuster™ Protein Extraction Reagent (Novagen) to every 1 mL cell pellet, with the addition of 20 U of DNase I (NEB) per mL of lysis buffer. After 20 min of agitation at 37°C (necessary for DNase I activation), samples were centrifuged at 10,000 rpm for 15 min in a Heraeus Fresco21 Centrifuge (ThermoFisher, USA). After centrifugation, the supernatant was transferred to a new tube without disturbing the ‘insoluble’ pellet, which was further suspended with the additional 100 µL of BugBuster™ + DNase I buffer. Samples were prepared as in General Methods par. M11 and loaded on NuPAGE™ polyacrylamide gels, run using XCell SureLock Mini-Cell and PowerEase Touch 120W (Invitrogen) and visualized via a ChemiDoc XRS+ Imaging System (BioRad).

Individual protein bands were subsequently cut out and sent to the BSRC Mass Spectrometry and Proteomics Core Facility for sequence analysis. Mass spectrometry data were analysed via the Mascot search engine, that identifies protein identity from primary sequence databases (SwissProt; EMBL ESTs; cRAP; selected UniProt proteomes) or from an internal database where the primary sequences of proteins of interest are uploaded (https://newbsrcmascot.st-andrews.ac.uk/mascot/search_intro.html; Perkins et al., 1999).

3.2.3 *CcPAF-AH2 purification attempts*

The different *CcPAF-AH2* 1L liquid cultures were separated into two pellets of approximatively 1500-2000 OD<sub>600</sub> cell numbers each and purification was attempted for each pellet at a time. The pellet was suspended in 25 mL of Purification Buffer A composed of: 50 mM HEPES, 500 mM NaCl, 10 mM Imidazole pH 8.0 and supplemented with 1 mg/mL of Lysozyme (Sigma-Aldrich, USA) and 1 U/mL of DNase I and half tablet? of cOmplete™ Mini Protease Inhibitor (Sigma Aldrich). The pellet was then lysed by passing it through a CF1 cell disruptor (I&L Biosystems, Germany) at 30 kpsi, keeping the sample cooled at 4°C for the entire lysis process. After homogenization the sample was centrifuged for 15 min at 12,500 rpm and 4°C on an Avanti™ JA-30I centrifuge (Beckman Coulter, USA). After centrifugation, the supernatant was transferred to a new falcon tube with a syringe and a 0.45 µm filter and loaded on a His-Trap FF 5 mL column (Cytiva, USA) using an ÄKTA Start chromatography system (Cytiva). The sample loading was performed at a flux speed of 0.5 mL/min. After that a washing step with 10 CV (50 mL) of Buffer A followed, and elution was performed with 20 CV of a gradient from 0% to 100% of Purification Buffer B (50 mM HEPES, 500 mM NaCl, 500 mM Imidazole pH 8.0), collecting 18, 6 mL fractions. Aliquots of samples from each of the aforementioned purification steps were taken and visualized via SDS-PAGE, as described in the previous paragraph.

Alternatively, in order to dissolve inclusion bodies and attempt to solubilize insoluble proteins, the following protocol was followed: after cell disruption and centrifugation the supernatant was discarded and the pellet was agitated with 20 mL of Buffer A' composed of 50 mM HEPES, 500 mM NaCl and 8M Urea pH 8.0 for 2h.
After resuspension the buffer supernatant was loaded on a His-Trap FF 5 mL column using the ÄKTA Start, at a flux speed of 0.5 mL/min. After that a wash with 100 CV (500 mL) with a gradient from 100% Buffer A% to 100% Buffer B’ (50 mM HEPES, 500 mM NaCl pH 8.0) followed, with the aim of slowly removing the urea and renaturing the unfolded protein while it was on the column. Finally, the elution was performed with 20 CV of Purification Buffer B (50 mM HEPES, 500 mM NaCl, 500 mM Imidazole pH 8.0), collected and analysed as previously described.

3.2.4 3.2.4 CcPAF-AH2 catalytic serine mutation

In order to overcome the suspected toxicity of CcPAF -AH2 in the E. coli cells, a mutation of its catalytic serine was performed in both pJ411_NTER and pJ411_CTER vectors. The catalytic serine was identified as the 243th amino acid residue (S243) in the CcPAF -AH2 aa sequence, by aligning the CcPAF -AH2 sequence with homologous intracellular PAF-AH II from other species using the TCoffee alignment software; because of the highly conserved catalytic amino acids in these enzymes. The mutation that was targeted involved replacing a Ser with an Ala (S243A), which was achieved by replacing the TCG codon with GCG. The mutation primers were designed using the NEBaseChanger online tool (https://nebasechanger.neb.com/) and are reported in Table 3.4.

Table 3.4 Primer sequences for cc_paf-ah2/pJ411_N-CTER mutation of S243A. Forward and reverse primers are listed, as well as the amplicon size. Mutated codon is evidenced in bold.

<table>
<thead>
<tr>
<th>Gene of Interest - Application</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>Length of Amplicon (bp)</th>
</tr>
</thead>
</table>

The PCR reaction used to insert the desired mutation was set-up using the cycling conditions and PCR reagents listed in Table 3.5.

Table 3.5 PCR reaction setup for cc_paf-ah2/pJ411_N-CTER mutation of S243A. Name of the PCR application, cycling conditions and PCR reagents are listed.

<table>
<thead>
<tr>
<th>Application</th>
<th>Cycling Conditions</th>
<th>PCR reagents</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcPAF -AH2-S243A/pJ411_NTER-CTER - Insert Amplification</td>
<td>98°C for 1’; (98°C for 30s; annealing at 60°C* for 30s and elongation at 72°C for 5’) for 25 cycles; terminal extension at 72°C for 3’</td>
<td>1U Q5® High-Fidelity DNA Polymerase (New England Biolabs), 5 µL Q5® Reaction Buffer 5x, 0.2 mM dNTPs, 0.5 µM primers, 1 µL template DNA (10 ng plasmid ccppaf-ah2/pJ411_NTER Miniprep product), MilliQH2O up to 25 µL</td>
</tr>
</tbody>
</table>
1 µL of the PCR product was used for the Kinase Ligase Digestion (KLD) reaction. This reaction combines ligase and kinase activities with the circular plasmid template digestion. The reaction setup is summarized in Table 3.6.

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration per reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR product</td>
<td>1 µL</td>
</tr>
<tr>
<td>T4 Ligase buffer 10X</td>
<td>1 µL</td>
</tr>
<tr>
<td>10 U DpnI (NEB)</td>
<td>0.5 µL</td>
</tr>
<tr>
<td>7.5 U Polynucleotide kinase (NEB)</td>
<td>0.75 µL</td>
</tr>
<tr>
<td>7.5 U T4 ligase (NEB)</td>
<td>0.75 µL</td>
</tr>
<tr>
<td>Nuclease-free water</td>
<td>6 µL</td>
</tr>
<tr>
<td>Total volume</td>
<td>10 µL</td>
</tr>
</tbody>
</table>

The reaction was incubated at room temperature for 1:30h and the entire volume was used for a single transformation reaction. Transformation, colony transfer to liquid growth medium and MiniPrep for plasmid DNA recovery were performed as in General Methodologies par. M8. DNA sequences were sent to the MRC PPU Unit of the University of Dundee for sequencing and mutation validation.

3.3 Results and discussion

3.3.1 Sequences coding PAF-AH enzymes

The C. closterium FE2 transcriptome was screened for enzymatic sequences that could be involved in the anti-inflammatory response. As explained in paragraph 3.1.4, among the anti-inflammatory enzymes, PAF-AH plays an important role in the metabolism of the pro-inflammatory phospholipid PAF; moreover, its phospholipase activity has been reported to have a broad-spectrum (Min et al., 2001) and is involved in different cellular and physiological functions that are independent from the inflammatory pathway (Karasawa and Inoue, 2015). For this reason, PAF-AHs from C. closterium were considered as interesting candidates for the expression and study of enzymes with biotechnological potential. Two transcripts, “TR16369|c0_g1_i1 length: 1566 nt” and “TR12571|c0_g1_i1 length: 1460 nt” were found and annotated as “platelet activating factor – acetylhydrolase”. The CDS contained in these transcripts were provisionally called cc_paf-ah1 and cc_paf-ah2 (with cc_paf-ah1 as the longest gene, 1362 bp, and cc_paf-ah2 as the shortest
gene, 1146 bp). TR16369|c0_g1_i1 length: 1566 nt and “TR12571|c0_g1_i1 length: 1460 nt sequences are reported in Figure 3.2 and Figure 3.3.

**Figure 3.2** DNA sequence of the transcript TR16369|c0_g1_i1 length: 1566 nt, where the longest newfound PAF-AH (prov. named CcPAF-AH1) was found. cc_paf-ah1 CDS is evidenced in **yellow**.
Figure 3.3 DNA sequence of the transcript TR12571|c0_g1_i1 length: 1460 nt, where the shortest newfound PAF-AH (prov. named CcPAF-AH2) was found. cc_paf-ah2 CDS is evidenced in blue.

Both transcripts possessed a single ORF of 452 and 382 aa, respectively. The amino acid number of these two novel PAF-AHs was coherent with the average amino acid number of previously identified PAF-AH cytoplasmic isoform II (390 aa).

3.3.2 CcPAF-AHs sequence alignment and catalytic site identification

The protein sequences of CcPAF-AH1-2 were aligned with other PAF-AH annotated in the UniProt database in order to deduce the putative catalytic triad of our proteins of interest and investigate their conserved motifs. Unsurprisingly CcPAF-AH1-2 showed very low homology with the subunits of the heterotetrameric intracellular type I PAF-AH, while having consistent homology with other PAF-AH type II. Based on the sequence alignment and on the identity of the catalytic triad of the human PAF-AH II (Thévenin et al., 2011), it was possible to deduce the putative catalytic triads composed of S244, D280 and H349 in CcPAF-AH1 and by S243, D267 and H337 in CcPAF-AH2, respectively (Figure 3.4).

In comparison to other esterase classes, the PAF-AH II has less conserved motifs, the function of which has been elucidated by site-directed mutagenesis (Dong, Li and Wu, 2021). One of the most evident differences in the two C. closterium PAF-AHs is the absence of the domain (MGXXXS) at the N-terminus for the post-translation myristoylation, that is shared by proteins with a subcellular distribution in both the cytoplasm
and membranes (Hattori et al., 1996). Interestingly, they share the lack of a MGXXXS motif with the other two microalgal PAF-AHs. A mutation in the G2 resulted in the incorrect trafficking of PAF-AH II between the membrane and cytosol (as proven in a transfection experiment in mammalian cells with a mutated human PAF-AH II protein (Thévenin et al., 2011), denoting its important role in membrane translocation. Nonetheless, *C. closterium* PAF-AH1-2 have a hydrophobic residue in position 2 (L and V, respectively) that could have the same function. Similarly, another conserved motif identified in another PAF-AH II is a cluster of three hydrophobic amino acids involved in membrane binding (Thévenin et al., 2011), identified in humans as L327, I328 and F331; this triad of amino acids is F362, L363 and M367 in CcPAF-AH1 and F350, L351 and L355 in CcPAF-AH2, respectively. Finally, the residue C121 is associated in humans with the TA activity of PAF-AH II (Bae, 2000) and is highly conserved across species; this residue is likewise conserved in CcPAF-AH1 (C139) but not in CcPAF-AH2. A crystal structure of PAF-AH II is current not available (the protein was quite difficult to crystallize, probably due to its transient association with the membrane (Birch et al., 2018)), and the only available homology model is based on the crystal structure of pPAF-AH. The expertise at the BSRC in the crystallization of membrane proteins was a key element in the achievement of this task and in the more precise elucidation of PAF-AH II catalytic and membrane-binding mechanisms.
Figure 3.4 Protein sequence alignment of C. closterium PAF-AH1-2 and other PAF-AH II enzymes, the sequences of which were collected from the UniProt database. The protein sequences used in the alignment are from the following organisms: the two green microalgae Auxenochlorella prototechoides and Osteococcus tauri, Homo sapiens (HUMAN), Bos taurus (BOVIN), Mus musculus (MOUSE), Caenorhabditis elegans (CAEEL) and Schizosaccharomyces pombae (SCHPO). Conserved regions are boxed in black, while the putative aminoacids of the catalytic triad are marked with a red star.

3.3.3 PCR Isolation of cc_paf-ah1 and cc_paf-ah2 coding sequences and purification

cc_paf-ah1-2 coding sequences were successfully isolated via PCR by using 20 ng of cDNA template, as described in General Methods par. M6. cc_paf-ah1 CDS isolation was performed by employing different
primer combinations and AT. However, even if the PCR reaction was performed at different temperatures, the primer pair FR1 always resulted in a specificity of c.a. 1000 bp in combination with the expected band length of c.a. 1500 bp (Figure 3.5, lane A), while the primer pair FR2 never gave a fragment of the expected length, even when primer concentration or template amount were changed (Figure 3.5, lane B). In the case of cc_paf-ah2 the primer pair FR1 gave a clear band of the expected length of c.a. 1300 bp, with a faint nonspecific product of a slightly longer length (Figure 3.5, lane C), while the primer pair FR2 gave a faint band of the expected length, apparently without any nonspecific product (Figure 3.5, lane D).

Figure 3.5 Agarose gel of cc_paf-ah1-2 CDS amplified with different primer pairs. A. cc_paf-ah1 CDS amplified with FR1 primers. B. cc_paf-ah1 CDS amplified with FR2 primers. C. cc_paf-ah2 CDS amplified with FR1 primers. D. cc_paf-ah2 CDS amplified with FR2 primers. Every PCR reaction was paired with the relative blank, without the addition of DNA template.

Regarding cc_paf-ah1 PCR product purification, the c.a. 1500 bp fragment obtained by the primer pair FR1 was purified using GenElute Gel Extraction kit (Sigma-Aldrich, USA) as reported in General Methodologies, par. M6. The sample was quantified with NanoDrop™, showing a concentration of 5 ng/µL; however, when this purified PCR product was used as template for the production of cc_paf-ah1 CDS sequences with pOPIN adaptors, the reaction failed to give any consistent amplification. To obtain the desired PCR fragment, another approach was tested: a PCR reaction, using primers for cc_paf-ah1 amplification with pOPIN adaptors was directly used on C. clustertium cDNA as template, using a “ramping up” protocol (described in paragraph 3.2.1) in order to overcome the expected poor representation of the transcript of interest. A faint amplification for both cc_paf-ah1/pOPINE and cc_paf-ah1/pOPINF-B sequences was observed with this approach. This PCR product was diluted 1/50 and used as a template for a second round of PCR reactions. This time, as expected, the fragment amplification was quite abundant and the fragment was consequently purified and used for the following cloning reactions (Figure 3.6).
On the other hand, for cc_paf-ah2 PCR product purification, the faint 1300 bp fragment obtained by the primer pair FR2 was directly diluted 1/50 (without any purification step) and used as template for the production of cc_paf-ah2 CDS sequences with pOPIN adaptors. Also in this case a “ramping up” protocol was chosen in order to maximise the reaction efficiency.

In this case a strong amplification for cc_paf-ah2/pOPINE was observed, but the high amount of nonspecific products required a purification with GenElute Gel Extraction kit; a fainter but clear amplification for cc_paf-ah2/pOPINF-B was instead observed, and in this case it was sufficient to combine three separate reaction tubes before proceeding with the AMPure purification method (Figure 3.7).
Using the different approaches as previously outlined, both cc_paf-ahs CDS sequences were isolated and amplified with appropriate pOPIN adaptors and were ready for the In-Fusion cloning reaction without the need for TOPO-TA subcloning.

3.3.4 **Cloning and sequencing of cc_paf-ah1-2 in pOPIN vectors**

*cc_paf-ah1* and *cc_paf-ah2* CDS sequences with pOPIN adaptors were produced and purified as previously described, and cloned into the pOPIN vectors as reported in General Methodologies, par. M11. The plasmids pOPINE, pOPINF and pOPINS3C (that uses the same adaptors and restriction enzymes as pOPINF, as described in General Methods par. M10 and by Louise E. Bird (Bird, 2011)) were used.

Results from the DNA sequencing of *cc_paf-ah1/pOPINE-F-3C* did not show base changes compared to the original transcriptome sequence (Figure 3.8), while *cc_paf-ah2/pOPINE-F-3C* showed different base changes in comparison to the original sequence (Figure 3.9).
Figure 3. cc-paf-ah1 CDS alignment from transcriptome and newly sequenced, as visualized via Jalview. In light blue are the areas of non-homology.
It is unlikely that all of these mutations were generated during the first endpoint PCR reaction, also because they are identical in all the sequenced plasmids; additional multiple errors during the second PCR reaction is likewise an unlikely event.

Nonetheless, the majority of these mutations are synonymous and missense mutations that do not involve any of the amino acids of the catalytic triad or of other conserved regions, and in terms of translated sequence only two amino acids changed between the original transcriptome sequence and the final sequence. It is also worth noticing that the synonymous codons in positions 334 and 448 (Leu), that are GGA and CTA respectively in cc_paf-ah2/pOPINE-S3C and GGG and CTG respectively in cc_paf-ah2/pOPINF. Considering that pOPINS3C uses the same insert as pOPINF, the most plausible explanation for this codon heterogeneity is a polymorphism with both versions amplified in the first PCR for cc_paf-ah2 CDS isolation.
Alignment between the translated sequences of the newly sequenced cc_paf-ah2 and the “original” transcriptome sequence is shown in Figure 3.10.

Finally, cc_paf-ah1/pOPINE-F-S3C and cc_paf-ah2/pOPINE-F-S3C plasmid maps were produced using SnapGene and are shown in Figure 3.11.
Figure 3.11 Plasmid maps of cc_paf-ah1/pPOPINE-F-S3C and cc_paf-ah2/pPOPINE-F-S3C produced using SnapGene. A. cc_paf-ah1/POPINF; B. cc_paf-ah1/POPINE; C. cc_paf-ah1/POPINS3C; D. cc_paf-ah1/POPINF; E. cc_paf-ah2/POPINE; F. cc_paf-ah2/POPINS3C.
The inserted gene DNA is shown in red and the related protein in light purple (CcPAF-AH1) or in dark purple (CcPAF-AH2). All plasmids show their size (bp), the lac operator (blue), the T7 promotor (arrowhead white), the replication origin (ori, yellow) and the Ampicillin resistance gene (AmpR, light green). pPOPINE maps show the C-terminus His-tag (6xHis, light purple); pOPINF maps show the N-terminus His-tag (6xHis, light purple) and the cleavage site for the protease 3C (dark blue; pOPINS3C maps show the N-terminus His-tag (6xHis, light purple), the Small Ubiquitin-like Modifier (SUMO TAG, grey) and the cleavage site for the protease 3C (dark blue).

3.3.5 Heterologous expression of CcPAF-AH/pOPIN proteins

cc_paf-ah1 and cc_paf-ah2 heterologous expression screening was undertaken as reported in paragraph 3.2.2. For this first expression screening only pPOPINE and pOPINF were used; this choice employed the following rationale: while the addition of small protein tags at the N-terminus of the POI can be relevant in terms of amount and solubility of the expressed protein, they are usually not relevant in terms of expression.

It was then decided to prove the occurrence of expression, even at low levels, firstly, using two different IPTG concentrations (1 mM and 500 µM) and one bacterial host. Rosetta 2 (DE3) pLySs was selected because cc_paf-ah1-2 preliminary sequence analyses using the JCat online tool (Grote, 2005) showed a prevalence of eukaryotic codons. Consequently, Rosetta 2 strains, that are engineered to compensate for seven rare eukaryotic codons (Tegel, 2010), were evaluated as being potentially useful in overcoming these difficulties. Finally, the pLySs strain was employed in order to avoid “leaky” production of a phospholipase in bacterial cells that could cause detrimental and unpredictable effects on their growth.

Using Protpi the molecular mass of CcPAF-AH1 and CcPAF-AH2 was predicted, with CcPAF-AH1 being predicted to have a mass of 52,85 kDa with the fusion tag of pOPINF and 51,79 kDa with the fusion tag of pPOPINE, respectively. CcPAF-AH2 was predicted to have a mass of 44,77 kDa with the fusion tag of pOPINF and 43,74 kDa with the fusion tag of pPOPINE, respectively. The gene encoding for a bacterial-optimized Green Fluorescent Protein (GFP; Noah, 2005) cloned in pOPINF was included in the screening as positive control. Total proteins were extracted from pellets and visualized on SDS-PAGE as reported in General Methods, par. M13.

The gel images are shown in Figures 3.12 and 3.13.
Figure 3.12 SDS-PAGE of soluble fractions of protein extracts of Rosetta 2 (DE3) pLysS cells transformed with cc_paf-ah1/pOPINE-F and cc_paf-ah2/pOPINE-F. Each of the five lanes is occupied from left to right by samples from 37°C Induction, 20°C Induction using 1 mM and 500 µM of IPTG inductor and Control, respectively. Used marker (M) was the Precision Plus Protein™ All Blue Standards (BioRad, USA).
The protein of interest was not produced in any of the conditions tested (which were experimentally proven correct due to the observed expression of the GFP gene). A Western-Blot analysis of all the different expression conditions was considered necessary in order to confirm faint levels of protein production or their complete lack of expression, as in General Methodologies, par. M14. 5 µg of the soluble fractions of CcPAF-AH1-pOPINE-F/37-20°C and of CcPAF-AH2-pOPINE-F/37-20°C were tested for the presence of His-tagged proteins. However, following the use of different antibodies and incubation times with the primary antibody no signal of interest was observed in the Western-Blots.

3.3.6 **Cloning and sequencing of cc_paf-ah1-2 in pJ411 vectors**

In order to set-up the samples needed for the experiments planned at the BSRC, cc_paf-ah1-2 were cloned in the pExpress411-Kan vectors provided by Dr. Czekster and da Silva, characterized by fusion-His6 tags at the N-terminus (pJ411_NTER) or at the C-terminus (pJ411_CTER). cc_paf-ah1 and cc_paf-ah2 CDS sequences with pJ411 adaptors were produced and cloned in pJ411 vectors as reported in General Methodologies, par. M10-11. cc_paf-ah1/pOPINE and cc_paf-ah2/pOPINF were used as templates for the synthesis of the PCR fragments cc_paf-ah1/pJ411_NTER-CTER and cc_paf-ah2/pJ411_NTER-CTER, respectively.
DNA sequencing of cc_paf-ah2/pJ411_NTER-CTER did not show further base changes, while cc_paf-ah1/pJ411_NTER-CTER showed the same synonymous mutations previously observed in cc_paf-ah2/pOPINF. Plasmid maps were produced using SnapGene and are shown in Figure 3.14.
3.3.7 Heterologous expression of CcPAF-AH1-2/pJ411 proteins

The first set of CcPAF-AH1-2 heterologous expression screening regimes at the BSRC followed the specific approaches previously reported in paragraph 3.2.2: specifically, all plasmids were transformed in BL21 (DE3) cells and tested at induction temperatures of 37°C (1:30h), 37°C (16h), 25°C (16h) and 16°C (16h). Total proteins were extracted from the pellets and visualized on SDS-PAGE using the protocol described in paragraph 3.2.2. Gel images are shown in Figure 3.15.

![Figure 3.14 Plasmid maps of cc_paf-ah1-2/pJ411_NTER-CTER produced using SnapGene. A. cc_paf-ah1/pJ411_NTER; B. cc_paf-ah1/pJ411_CTER; C. cc_paf-ah2/pJ411_NTER; D. cc_paf-ah2/pJ411_CTER. The inserted gene DNA is shown in red and the related protein in light purple (CcPAF-AH1) or in dark purple (CcPAF-AH1). All plasmids show their size (bp), the lac operator (blue), the T7 promotor (arrowhead white), the replication origin (ori, yellow), the Shine-Dalgarno sequence (grey) and the Kanamicin resistance gene (KanR, light green). pJ411_NTER maps show the N-terminus His-tag (6xHis, light purple) and the cleavage site for the TEV protease (grey); pJ411_CTER maps show the C-terminus His-tag (6xHis, light purple).](image)
Figure 3.15 SDS-PAGE of soluble and insoluble fractions of protein extracts. A. soluble fractions of BL21 (DE3) transformed with cc_paf-ah1/pJ411_NTER and cc_paf-ah1/pJ411_CTER. B. soluble fractions of BL21 (DE3) transformed with cc_paf-ah2/pJ411_NTER and cc_paf-ah2/pJ411_CTER. C. insoluble fractions of BL21 (DE3) transformed with cc_paf-ah1/pJ411_NTER and cc_paf-ah1/pJ411_CTER. D. insoluble fractions of BL21 (DE3) transformed with cc_paf-ah2/pJ411_NTER and cc_paf-ah2/pJ411_CTER. The white arrows indicate lanes that show an overexpression of the POI. Used marker (M) was the PageRuler™ Plus Prestained Protein Ladder (Thermo-Fisher).
This second expression screening gave poor results in terms of POI synthesis, with the exception of CcPAF-AH2, which was produced in all the different induction conditions in BL21 (DE3) strains, in association with both the pJ411_NTER and _CTER vectors. The different bands were cut out and sent to the BSRC Mass Spectrometry and Proteomics Core Facility, confirming their identity as CcPAF-AH2. Scale-up 1L cultures of BL21 (DE3) transformed with cc_paf-ah2/pJ411_NTER-CTER were prepared and the pellets produced were subsequently used in different purification regimes.

3.3.8 Troubleshooting with CcPAF-AH2 expression and purification

Different purification regimes were performed as described in paragraph 3.2.3 but without the successful recovery of CcPAF-AH2 from the elution fractions. The main protocol employed consisted of the mechanical action of the CF1 cell disruptor, in combination with the lysis buffer to disrupt inclusion bodies and resolubilize the POIs (Singh, 2015). Lysis buffers that differed from the original composition (+ DTT 2mM; + Triton 0.1%; + DTT 2mM and Triton 0.1%; 300 mM NaCl instead of 500 mM NaCl) were also tested, but with no beneficial effects.

Moreover, the very clearly visible CcPAF-AH2 overexpression band that was observed in the expression screening described in paragraph 3.3.7 was not visible in the insoluble nor in the pre-His-Trap purification fractions.

In addition, no signal of interest was observed in the Western-Blots, which had previously been seen in the expression screening with pOPIN vectors.

Subsequently, the expression of cc_paf-ah2/pJ411_NTER-CTER in 1L cultures of additional E. coli strains was undertaken. The E. coli strains that were used were the OverExpress™ C43 (DE3) and the SHuffle T7 strains. C43 (DE3) cells possess genetic mutations that have been phenotypically selected to confer tolerance to toxic proteins, and can reduce the problem of plasmid instability (Dumon-Seignovert, Cariot and Vuillard, 2004). SHuffle cells for their part possess a genetic mutation resulting in an induced impairment in cytoplasmic reductive pathways, thus allowing for the formation of disulphide bonds the recombinant proteins in the cytoplasm. The aim of this new growth experiment was to obtain a consistent expression of CcPAF-AH2 in 1L cultures, and possibly, due to the contribution of the SHuffle cells, to facilitate this expression in the soluble fraction. In fact, the formation of incorrect disulphide bonds in the POI or between different proteins (an occurrence more likely to happen in a reducing cytoplasmic environment) is a known cause of protein precipitation and the formation of IBs (Chen and Leong, 2009). Surprisingly, neither C43 nor the SHuffle cells transformed with cc_paf-ah2/pJ411_NTER-CTER plasmids were able to stay vital scaling-up from the 15 mL pre-inoculum to the 1L growth volume. Even with the addition of 1% Glucose to the pre-inoculum and the 1L
growth volume both cells stopped growthing immediately after the addition of IPTG. Finally, plate-grown colonies of C43 or SHuffle cells transformed with cc_paf-ah2/pJ411_NTER-CTER appeared smaller and fewer in number when compared to the ones transformed with empty pJ411_NTER-CTER vectors (Figure 3.16).

![Figure 3.16 LB Agar plates of SHuffle cells transformed with A. cc_paf-ah2/pJ411_NTER and B. with empty pJ411_NTER. The different colony size and number is clearly visible between the two plates.](image)

These results suggested a high level of toxicity in the E. coli cells induced by CcPAF-AH2, the expression of which appeared to hamper the growth of C43 strains, even though this strain has been optimized for the expression of toxic proteins. It is likely that the oxidative environment of the SHuffle cells effectively benefits the correct folding of CcPAF-AH2 and consequently, its toxicity. It is likely that the plasmid instability of cc_paf-ah2/pJ411_NTER-CTER allowed the BL21 (DE3) 1L cultures to keep growing during the induction time. More specifically, the cells that received IPTG induction and expressed CcPAF-AH2 are likely to have died. Probably a similar outcome happened also in the case of CcPAF-AH1 and of the transformation of Rosetta 2 (DE3) pLySs cells with cc_paf-ah1-2/pOPINE-F.

It is difficult to explain why BL21 (DE3) was able to express CcPAF-AH2 only in the expression screening. A possible reason is that in BL21 (DE3), due to its “leaky” expression, CcPAF-AH2 started to accumulate in IBs
before the addition of IPTG and, when the expression levels increased, the newly produced protein naturally interacted with the already misfolded proteins. However, considering that this theoretical process did not happen in the case of 1L cultures, the observed PAF-AH2 accumulation in the expression screening was considered a stochastic event, the replicability of which is likely occasional and unreliable for subsequent processing steps.

Nonetheless, considering that the aim of the TNA project was to obtain pre-refined data for the CcPAF-AH1-2 crystal structure, a strategy was developed for the accumulation of CcPAF-AH2 and its purification to obtain structural data, including circumventing the toxicity phenotype via the mutation of the catalytic S243 in A (an aminoacid with a rigid and inert lateral chain). A mutation at S243 was performed as described in paragraph 3.2.4. cc_paf-ah2_243A/pJ411_NTER-CTER vectors were used to transform C43 (DE3), BL21 (DE3), Rosetta (DE3) and SHuffle cells. The strain selection was made in order to encompass all the different possibilities. BL21 was selected due to the positive results obtained in the previous expression screening. C43 was selected for its tolerance to toxic proteins, caused by mutations of the RNA Polymerase that slow the gene expression and, consequently, the effect of eventual toxicity effects in the host ((Wagner et al., 2008). Rosetta was selected for the eukaryotic codon usage of CcPAF-AH2 and SHuffle for their oxidative cytoplasm environment that can help in the formation of the right disulphide bonds. Each transformed strain was directly inoculated into a 1L growth volume and induced at 37°C and 16°C, in both cases for 16h. For each condition total proteins of 1 mL aliquots were extracted from pellets and visualized on SDS-PAGE. Gel images are shown in Figure 3.17.
In this third expression screening regime, CcPAF-AH2_S243A identity was confirmed by mass spectrometry data, via the analysis of protein bands of plausible molecular weight at the BSRC Mass Spectrometry and proteomics Facility via the Mascot search engine (Figure 3.18).
CcPAF-AH2 still appeared in the insoluble fraction, and the only conditions that resulted in an appreciable expression level were CcPAF-AH2_S243A/pJ411_NTER in Rosetta (DE3) cells, at both 37°C and 16°C induction conditions and CcPAF-AH2_S243A/pJ411_NTER in SHuffle cells, at 37°C induction condition. This result, other than confirming the toxicity of CcPAF-AH2 for SHuffle and C43 cells (that however this time normally reached the exponential phase), also confirmed the eukaryotic codon usage of cc_paf-ah2. In fact, the overexpression observed in the Rosetta samples probably benefitted from the Rosetta optimization for eukaryotic codon usage. However, in this case as well, the overexpressed protein did not appear in the soluble fraction, not even in the case of the SHuffle cells. This was possibly related to the membrane translocation of CcPAF-AH2 that, as previously discussed in par. 3.3.2, is probably independent from the presence of the MGXXXS motif, and especially from the presence of Glycine in position 2, in contrast to the key role the residue is known to play in human (Thévenin et al., 2011). Overall, the following conclusions can be drawn: in the expression screening described in par. 3.3.7 it is not possible to say if the presence of CcPAF-AH2 in the insoluble fractions of BL21 (DE3) is related to its inclusion in IBs or to its likely tight association with the cell membrane. In this third screening regime, given the inhibition of the catalytic activity of the POI and the presence of CcPAF-AH2 even in the insoluble fractions of SHuffle cells, its association with the cell membrane is more
plausible; in this case the overexpression of the POI was at last achieved in 1L LB cultures, and new purification regimes were thus attempted.

Specifically, a purification protocol based on the dissolution of the IBs using 8M Urea (as described in par. 3.2.3) was attempted, but even in this case no appreciable purification of CcPAF-AH2_S243A/pJ411_NTER was achieved. While this failed attempt further corroborated the hypothesis of the membrane-association of CcPAF-AH2, clear inferences can only be drawn after ad hoc purification protocols designed for membrane proteins have been attempted. Membrane protein purification involves the use of non-ionic detergents that keep the protein soluble while minimizing the interference with the elected purification technique (Loll, 2014; Pandey, 2018), such as the n-dodecyl-β-D-maltopyranoside (DDM), and involves the use of ultracentrifuges that are capable of reaching 100.000g in order to separate the membranes from other cellular components. This laborious purification process is one of the reasons why membrane proteins are currently underrepresented in protein databases such as PDB (Pandey et al., 2016).

3.4 Conclusions

CcPAF-AH1-2 were selected as proteins of interest in order to elucidate the role of the PAF-AH II protein family in microalgae, and specifically their role in the diatom C. closterium which has to date not been investigated. In fact, the role of CcPAF-AHs could be related both to the anti-inflammatory activity of extracts and to lipid metabolism, potentially contributing to the healthcare field of microalgal biotechnology.

Given this aim the planned experimental pipeline included CcPAF-AH1-2 CDS isolation, their cloning in expression vector and subsequent heterologous expression. These preliminary steps were finalized for the production of high levels of POIs, in order to obtain crystallographic data at the BSRC at the University of St. Andrews as part of the 1-month EU-ASSEMBLE+ Transnational Access (TNA) 8th Call project, in collaboration with Dr. Clarissa Melo Czekster and Dr. Rafael Guimaraes da Silva. It is worth mentioning that the resolution of a crystallographic structure of an enzyme belonging to the PAF-AH II family has never been achieved, and this novelty made the endeavour for both SZN and BSRC research groups both challenging and but provided the opportunity to potentially generate interesting new data to the research field dedicated to PAF-AHs.

The CDS isolation of the two genes cc_paf-ah1 and cc_paf-ah2 and their cloning in pOPIN and pJ411 vectors was successfully achieved. However, in terms of protein expression the use of E. coli as a heterologous host proved to be unfeasible. In fact, while BL21 (DE3)-derived strains have previously been successfully used to produce recombinant proteins from the same PLA2 superfamily (Valdez-Cruz et al., 2017; Gimenes et al., 2020), it was not possible to achieve expression of CcPAF-AH1-2 on a regular and predictable basis. Moreover, the growth of bacterial strains optimized for the production of toxic proteins (E. coli C43) or for the formation
of correct disulphide bonds (*E. coli* SHuffle T7) was completely impaired by induction of the expression of the protein of interest.

Given the time limits for the TNA, the best strategy to obtain crystallographic data was deemed to be to mutate the catalytic serine of CcPAF-AH2, in order to remove its enzymatic activity and thereby overcome its suspected high toxicity in *E. coli*, that is likely to involve the uncontrolled acetylhydrolase activity of the enzyme which targets the membrane phospholipids in the bacterial cell. In this case, even if CcPAF-AH2 production was achieved in 1L cultures, the persistent presence of POI in the insoluble fraction also suggests a tight association of CcPAF-AH2 with the cellular membrane. **Figure 3.19** summarizes the theorized mode of action of CcPAF-AH2, based on the available literature on this enzyme family and the aforementioned preliminary observations.

In the left panel, an active “normal” CcPAF-AH2 migrates to the cell membrane, cleaving key phospholipids and altering membrane stability. This likely leads to cell death. In fact, different studies have been performed to elucidate the substrate specificity of PAF-AH II enzymes. Regarding its non-regulated acetylhydrolase activity the minimal substrate requirement is the portion of a glyceride derivative that includes a sn-2 ester and a hydrophobic chain in position sn-1 (Dong, Li and Wu, 2021).

In the right panel, a mutated S243A CcPAF-AH2 normally migrates to the cell membrane, but the protein remains attached to the membrane, and still appears in the insoluble fraction on SDS-PAGE. From the literature available for human PAF-AH II, the catalytic serine is not involved in membrane translocation, and the residues involved in this function in *C. clouterium* are as yet unknown. Given that other PAF-AH IIs have previously displayed both cytoplasmic and membrane subcellular distribution (Rice, 1998), a PAF-AH II activity that is purely cytoplasmic cannot be completely ruled out.
The biotechnological potential of CcPAF-AH1-2 can be further elucidated by following two main routes: 1) In order to obtain pure CcPAF-AH2 and related crystallographic data, protocols optimized for membrane protein purification should be designed, involving the use of non-ionic surfactants and high-speed centrifugation. 2) In order to investigate the anti-inflammatory and lipase activity of CcPAF-AH2 an alternative host organism for recombinant protein production should be considered (such as the yeast *Pichia pastoris*, or the diatom *Phaeodactylum tricornutum*). Once a monodisperse protein solution is obtained, the activity assays could easily be performed using the different PAF hydrolase assays that are commercially available, as well as other potential phospholipid substrates. Also in this case, as proposed for CcDBPH1-2, a feasible alternative could be to measure PAF hydrolytic activity directly from cellular pellets.

Overall, this chapter represents the first attempt to produce and purify recombinant PAF-AHs enzymes belonging to the PAF-AH II family from the marine diatom *Cylindrotheca closterium*. While the protein purification itself has proven unsuccessful, the data reported here represent a first step in helping to elucidate the role of PAF-AH in the metabolism of marine diatoms and the potential of this family of enzymes (even if they have not as yet been characterized, are also known to be present in other microalgae) for microalgal healthcare biotechnology applications.
Chapter 3 References:


Chapter IV

First Identification of L-asparaginase in the Marine Diatom Cylindrotheca closterium

Graphical abstract of the experimental pipeline of Chapter 4.
Chapter 4  Abstract

L-Asparaginase (ASNase) is an interesting enzyme from a biotechnological point of view, with a well-defined commercial niche: ASNase is a l-asparaginase hydrolase that has been widely used in the treatment of acute lymphoblastic leukemia and other related blood cell cancers for more than 30 years. More recently, the additional glutaminase activity of many ASNases has increased the therapeutic application of this enzyme, that is considered interesting for the treatment of different solid tumors. However, ASNase purified from terrestrial microorganisms often causes cytotoxicity and detrimental effects.

Moreover, ASNase can also be used for the reduction of acrylamide, a carcinogenic compound that is generated from l-asparagine in different high temperature-cooked foods, as a product of the Maillard reaction. Consequently, novel ASNase can have a potential application not only as a therapeutic drug but also in the context of an overall improvement in food quality.

Many marine bacteria and microalgae have been suggested as promising sources of ASNase. A transcript annotated as “l-asparaginase i” has been found in the C. closterium transcriptome. The CDS of this gene (prov. called cc_asnase, with the related protein CcASNase) was isolated via PCR and cloned in a TOPO-TA vector and into pJ411 expression vectors. While in the expression screening CcASNase lacked an intense overexpression appreciable following SDS-PAGE, western-blot analysis subsequently confirmed its expression under different conditions. In order to increase the amount of expressed protein, the cc_asnase gene was cloned in the pOPINS3C-TRX-HALO7 vectors, characterized by N-terminus protein tags of high molecular weight, in order to try to increase protein production. Future approaches may involve the high-scale production of CcASNase, its purification and its activity characterization, via different assays, both in vitro against isolated substrates (L-Asparaginase, glutamine) and against different cell lines (leukaemia and carcinoma cancer cells).
4.1 Introduction

4.1.1 Summary of L-asparaginase (ASNase) characteristics and its therapeutic applications

L-asparaginase (ASNase or L-ASP) has a long history as a therapeutic enzyme with antineoplastic properties; it was the first anti-cancer enzyme to be studied by the global scientific community almost 70 years ago (Batool, 2016). Asparaginase (L-asparagine amidohydrolase, E.C. 3.5.1.1; ASNase) is defined as an enzyme that catalyses the hydrolysis of L-asparagine to aspartic acid and ammonia. ASNase was already briefly presented in Chapter 1 as a therapeutic enzyme for the treatment of ALL and other blood cancers (Batool et al., 2016), and homologous enzymes have already been identified in a few microalgae. This enzyme was initially linked to the anti-lymphoma activity of guinea pig serum (Broome, 1963), but microorganisms were selected as a preferential source for enzyme extraction and downstream production, more specifically bacteria (Sanches et al., 2003; Pieters et al., 2011; Farahat, Amr and Galal, 2020), fungi (Costa et al., 2016; Al Yousef, 2022) and, to a lesser extent, cyanobacteria (ELKOMY and FARAG, 2018; Abd El-Baky and El-Baroty, 2019) and microalgae (Paul, 1982; Ebrahiminezhad et al., 2014; Lauritano et al., 2016). Of the present five L-ASNase preparations used in the clinic, two are from the bacterium Erwinia chrysanthemi (Erwinase® and Rylaze® (Pieters et al., 2011)), and three from E. coli (Spectrila®, Oncaspar®, Asparlas® (Van Trimpont et al., 2022)).

The vast majority of microbial ASNases studied to date come from terrestrial organisms. Marine sources are also receiving increasing interest in order to identify enzymes with favourable features, such as the absence of immunological and non-immunological side effects, lack of resistance induction, low molecular weight and increased substrate specificity (Izadpanah et al., 2018).

E. coli possesses two classes of ASNases (type I and II, respectively): type I (AnsA) is constitutively expressed with cytosolic localization, has broad substrate specificity for both asparagine and glutamine, with low affinity (millimolar Km) for asparagine. Type II (AnsB) expression is instead induced under anaerobic conditions, the enzyme is localized in the periplasm and shows considerably higher affinity (micromolar Km) for asparagine, even if it retains some glutaminase activity (Schwartz, Reeves and Broome, 1966). Type I and II are overall very similar in their secondary and tertiary structures, and both have shown crystallographic tetrameric structures using X-ray analysis (Yun et al., 2007); with the tetramer consisting of four identical subunits with the catalytic site localized at the dimer interface (Sanches et al., 2003; Yun et al., 2007). Finally, the residues of the active site and the homodimer interface are conserved between type I and II, and throughout evolution (Yun et al., 2007; Batool et al., 2016). Given its higher Km, type II ASNase (and its homologue in other microorganisms) has been evaluated for therapeutic applications (Batool et al., 2016; Baral et al., 2021).
The AnsA homologue has also been widely investigated with results indicating that the l-asparaginase domain of a 60-kDa (573-residue) human lysophospholipase, named hASNase1, showed a $S_{0.5}$ 10 times higher than the bacterial homologue and exhibited strong positive allosteric modulation of l-asparagine concentrations (Karamitros and Konrad, 2014). ScASNase1 of the yeast *Saccharomyces cerevisiae* showed a high affinity for antineoplastic substrates similar to II ASNases, but also cytoplasmic localization and allosteric modulation such as type I (Costa *et al.*, 2016).

Even therapeutic ASNases showed some degree of glutaminase activity, hydrolysing L-glutamine into glutamate and ammonia (OLLENSCHLÄGER *et al.*, 1988). Glutaminase activity is, however, involved in some non-immune related toxicities that include pancreatitis, hepatotoxicity, coagulopathy, neurotoxicity and hyperglycemia (Costa *et al.*, 2016; Van Trimpont *et al.*, 2022). Moreover, while novel therapeutic ASNases are being researched with low or absent glutaminase activity (Nguyen *et al.*, 2018; Van Trimpont *et al.*, 2022), different studies have shown how a low level ASNase-induced cytotoxicity can be found in blood cancer cells (Anishkin *et al.*, 2015; Parmentier *et al.*, 2015). Glutaminase activity is also vital against leukemic cell lines that have a functional L-asparagine synthase (ASNS) enzyme (Chan *et al.*, 2014).

Other solid cancers have shown low ASNS levels in comparison to healthy cells, and have been proposed as likely targets for the therapeutic activity of ASNase (Scherf *et al.*, 2000; Dufour *et al.*, 2012). Studies on this matter have seen various degrees of success: in metastatic breast carcinoma cell lines, asparagine depletion caused a loss of metastatic potential (Knott, 2018). In the case of ovarian carcinoma, instead, no clinical studies were successful, even if different tested tissues proved to be ASNS-deficient.

Glutaminase activity of ASNase was beneficial in combination with other drugs in specific subsets of cancers: in combination with rapamycin, with ASNase arresting the growth of KRAS-mutant colorectal cancer (Toda *et al.*, 2016) and, in combination with glutamine synthase inhibitors, had an anti-cancer effect on hepatocellular carcinoma and triple-negative breast cancer (Fung and Chan, 2017). Finally, the efficacy of ASNase against brain tumours, such as glioblastoma, has been proven in cell lines (Panosyan *et al.*, 2014), while the full effects of asparagine depletion from cerebrospinal fluid are not yet fully understood (Van Trimpont *et al.*, 2022).

Overall, although there is a long history of therapeutic applications of ASNase, there is still a need for new developments and discoveries. Even if type II ASNases have always been considered as paramount for anticancer activity, bacterial classification cannot fully describe all the different homologous enzymes available in higher organisms that have been characterized in recent years. In particular, among marine organisms, microalgae are an almost-completely untapped source, and to my knowledge there is to date no scientific literature regarding ASNase purification and characterization from microalgal sources yet.
4.1.2 Previous applications of L-asparaginase (ASNase) enzymes in food treatment

Acrylamide (AA) is a small organic molecule with a molecular formula of CH$_2$=CH–CO–NH$_2$. AA is formed with asparagine reacting with sugars (especially glucose and fructose) due to the frying, baking, or grilling of starchy foods at over 120 °C through a non-enzymatic process called the Maillard reaction (Tareke et al., 2002; Zytk et al., 2003; https://www.fda.gov/food/chemical-contaminants-food/acrylamide-questions-and-answers#:~:text=Acrylamide%20is%20a%20chemical%20that,food%20packaging%20or%20the%20environment.). Acrylamide has been classified as a probable human carcinogen (Group 2A) by the International Agency for Research on Cancer, having been shown to have neurotoxic, genotoxic, carcinogenic and reproductive system toxicity effects in animal models (Zamani et al., 2017). The EU regulation 2017/2158 of 20 November 2017 was formulated to reduce the amount of acrylamide in food (https://eur-lex.europa.eu/eli/reg/2017/2158/oj), and there is great interest among food business operators to find novel and economic ways to comply with the new rules and mitigate AA production. Acrylamide mitigation strategies include raw materials selection, product composition alteration, optimization of processing conditions and food pre- and post-treatment procedures (Xu, Oruna-Concha and Elmore, 2016). However, the most feasible solution is believed to be enzymatic pre-treatment of the raw materials. ASNase can hydrolyse asparagine, reducing substrate availability for the Maillard reaction, and be deactivated naturally during the cooking process via heating or pH alteration. Two commercial ASNase-based food additives, Acrylaway® and PreventAse® (from fungi Aspergillus oryzae and Aspergillus niger, respectively) are both sold as products able to reduce acrylamide levels in processed food by up to 95% with no effect on taste or appearance of the final food. Both have been proven to have acrylamide mitigation efficacy by many authors (Dourado et al., 2020; Corrêa et al., 2021; Rottmann et al., 2021) and are considered safe as food additives (Xu, Oruna-Concha and Elmore, 2016; Jia et al., 2021). On a pilot lab scale, microbial ASNase-based mitigation of acrylamide has been tested and proven in various foods, such as: potato slices and strips for chips and in french fries’ production (Pedreschi et al., 2011; Sun et al., 2016; Jiao et al., 2020), together with flour and dough for bakery products (Anese, Quarta and Frias, 2011; Hong et al., 2014; Meghavarnam and Janakiraman, 2018) and coffee (Khalil, Rodriguez-Couto and El-Ghany, 2021).

Finally, the thermal stability of ASNase and the implementation of reutilization strategies are considered crucial elements for acrylamide mitigation on an industrial scale. Both thermostability and reusability were improved by immobilizing the enzyme with different materials, such as magnetic nanoparticles modified with aminopropyl triethoxysilane (APTES), with a three-fold increase in its thermostability and 90% of activity retention after the fifth reaction cycle (Atam, 2018) or agarose spheres, with 93.21% of activity retention after the sixth reaction cycle. Other highly thermostable ASNases from Thermococcus kodakaraensis (Chohan and Rashid, 2013), Bacillus amyloliquefaciens MKSE (Yim and Kim, 2019), Anoxybacillus flavithermus have been suggested as promising in acrylamide mitigation approaches (Maqsood et al., 2020) At present,
only one industrial scale pilot study has included the pre-treatment of an industrial batch of 15 tons of potato stripes, before their pan-frying with different concentrations of commercial ASNase preparations (PreventASe L®). However, in order to reach 59% acrylamide reduction, an extra production cost of up to 50 cent/kg was evaluated (Rottmann et al., 2021).

The investigation on a transcript of C. closterium annotated as “l-asparaginase i” appeared promising considering the absence of information on the molecular and biochemical characterization of microalgal ASNases and the two aforementioned rich commercial sectors that could make use of novel ASNases.

4.2 Methods

4.2.1 CcAsnase oligo design and reaction procedures for PCR

The Blast2GO software annotated the transcript “TR25145|c0_g1_i1 len=1778” as “l-asparaginase i”; the CDS contained in the transcript was provisionally called cc_asnase (and the related protein CcASNase), and selected as GOI. The cc_asnase CDS has a length of 1545 bp.

Primers for endpoint PCR and for sequencing reactions were designed as reported in General Methodologies, par. M5 and M9. Sequences of used primers for PCR are reported in Table 4.1.

<table>
<thead>
<tr>
<th>Gene of Interest - Application</th>
<th>Forward Primers</th>
<th>Reverse Primers</th>
<th>Length of Amplicons (bp)</th>
</tr>
</thead>
</table>
| CcASNase - CDS Isolation      | F1: GCCCCAGACGAAGATTTGTG  
F2: GCCGAATTTGATTCCTGG       | R1: TCTATGGAACGATGGGGCAT  
R2: GAACGATGGGGCATCGATC      | F1R1: 1684  
F1R2: 1678  
F2R1: 1634  
F2R2: 1628 |
| CcASNase - Sequencing         | 1: CGTGGTCACTCATCTTGCAGAC  
2: ACTGGGAGATGGTGGCTCAA  
3: GACGATCTGGTCAATGGGTT       | 1: CCTAACAAACGGATCGCGTAGGCT  
2: GTGTTGATCTCTGGTACGTCG  
3: GATCTCCGTAACCTCTCGC  
4: TAGCATCGTCTCGATGATCC       | N.D. |
| CcASNase – pj411_NTER Cloning | AAAATTTGTATTTCCAGGGGATGATCC  
GCTATAATTACATTAG          | ATGCTAGGGGGGAAGCTTCCAAAATGATCGACTGCGGATTTC       | 1585 |
| CcASNase – pj411_CTER Cloning | ACTTTTAAGGAGGTAACACATGATCC  
GCTATAATTACATTAG          | TGGTGTGTTGTTGTTCGCTCAATGATCGACTGCGATTTTC       | 1585 |
| CcASNase – pOPIN Family Cloning | AAGTTCTGTTTTCAGGGGCGATGATCC  
GC TAT AATTACATTGAA         | ATGGCTAGAAGACGGTTTAATGATCGACTGCGGATTTTC       | 1583 |

Different PCR conditions and set-ups were tested for the amplification of cc_asnase PCR fragments: specifically, the Expand™ High FidelityPLUS PCR System (Roche) was used for the isolation of cc_asnase CDS, while the cc_asnase CDS with the pj411/pOPIN adaptors was isolated from TOPO-TA template using the
PrimeSTAR® GXL DNA Polymerase. PCR reactions were set-up using the cycling conditions and PCR reagents listed in Table 4.2 and Table 4.3.

Table 4.2 PCR reaction setup for cc_asnase CDS isolation. Name of the PCR application, primer working couple, cycling conditions and PCR reagents are listed.

<table>
<thead>
<tr>
<th>Application</th>
<th>Working Couple</th>
<th>Cycling Conditions</th>
<th>PCR reagents</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcASNase - CDS Isolation</td>
<td>F2R2</td>
<td>95°C for 5 '; (95°C for 30s; annealing at 55°C for 30s and elongation at 72°C for 1:30 ') for 30 cycles</td>
<td>1U Xtra Taq Polymerase, 6 µL XtraTaq Pol White Buffer 5x, 0.2 mM dNTPs, .05 µM primers, 50 ng cDNA, MilliQH20 up to 30 µL</td>
</tr>
</tbody>
</table>

Table 4.3 PCR reaction setup for cc_asnase CDS isolation with pJ411 and pOPIN adaptors. Name of the PCR application, primer working couple, cycling conditions and PCR reagents are listed.

<table>
<thead>
<tr>
<th>Application</th>
<th>Cycling Conditions</th>
<th>PCR reagents</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcASNase/pJ411_NTER – CTER – Insert Amplification</td>
<td>98°C for 2 '; (98°C for 10s; annealing at 58°C* for 15s and elongation at 68°C for 1:40 ') for 34 cycles; terminal extension at 68°C for 7 '</td>
<td>1.25 U PrimeSTAR® GXL DNA Polymerase, 10 µL PrimeSTAR® GXL Buffer 5x, 0.2 mM dNTPs, 0.4 µM primers, 1 µL template DNA ( 2 ng plasmid CcASNase/TOPO-TA Miniprep product), MilliQH20 up to 50 µL</td>
</tr>
<tr>
<td>CcASNase /pOPINF – Insert Amplification</td>
<td>98°C for 2 '; (98°C for 10s; annealing at 56°C* for 15s and elongation at 68°C for 1:40 ') for 34 cycles; terminal extension at 68°C for 7 '</td>
<td>1.25 U PrimeSTAR® GXL DNA Polymerase, 10 µL PrimeSTAR® GXL Buffer 5x, 0.2 mM dNTPs, 0.4 µM primers, 1 µL template DNA ( 2 ng plasmid CcASNase/TOPO-TA Miniprep product), MilliQH20 up to 50 µL</td>
</tr>
</tbody>
</table>

*annealing temperature was selected based on the formula \[ Tm (^\circ C) = [(\text{the number of A and T}) \times 2] + [(\text{the number of G and C}) \times 4] – 5 \], considering for the calculation only the part of the primers that annealed with the template.

4.2.2 CcASNase heterologous expression and SDS-PAGE

Heterologous expression of CcASNase was performed using pJ411 vectors (pJ411_NTER and pJ411_CTER) using Lemo21, Rosetta 2 (DE3) and BL21 (DE3) as expression hosts, incubated overnight at 20°C and 1-hour at 37°C induction conditions and 1 mM IPTG concentration, as in General Methodologies par. M12. Total proteins were extracted from pellets and visualized on SDS-PAGE, as in General Methodologies par. M13. The presence of the protein of interest (POI) was verified via Western-Blot, as in General Methodologies par. M14.
4.3 Results and discussion

4.3.1 Sequences coding CcASNase enzyme and catalytic site identification

The importance of mining marine microbial sources for novel and safer ASNase enzymes (Izadpanah et al., 2018) has previously been addressed in paragraph 4.1.1, together with the potential application of this enzyme in the food industry in paragraph 4.1.2. For this reason, and given the robust amount of available literature regarding its activity assessment and protein purification, an ASNase from *C. closterium* was considered a viable candidate for heterologous expression and further activity assays. Consequently, the *C. closterium* FE2 transcriptome was screened for enzymatic sequences annotated as “l-asparaginase” by the Blast2GO annotation software. Only one transcript, “TR25145|c0_g1_i1 len=1778”, was found annotated as “l-asparaginase I”. The CDS contained in this transcript was provisionally named *cc_asnase* and selected as gene of interest (GOI). The transcripts possessed a single ORF of 515 aa. “TR25145|c0_g1_i1 len=1778” and is reported in Figure 4.1.

**Figure 4.1** DNA sequence of the transcript TR25145|c0_g1_i1 len=1778 nt, where the newfound ASNase (prov. named CcASNase) was found. *cc_asnase* CDS is highlighted in green.
Transcript “TR25145|c0_g1_i1 len=1778” possessed a single long ORF of 515 aa and via BlastP alignment its identity as a “l-asparaginase I” of microalgal origin was confirmed due to its high homology with other annotated ASNases from microalgae. The amino acid (aa) number of this novel ASNase was higher than the average aa number of previously identified ASNases (c.a. 350 aa), but similar to other ASNases I that have been annotated in diatom genomes (such as *Fistulifera solaris*, 467 aa (Tanaka *et al.*, 2015), Chaetoceros *tenuissimus*, 482 aa (Hongo *et al.*, 2021) and *Nitzschia inconspicua*, 502 aa (Oliver *et al.*, 2021)).

Two more transcripts, “TR4418|c0_g1_i1 len=673” and “TR4418|c0_g1_i2 len= 678” were annotated as “probable l-asparaginase periplasmic-like”. As previously mentioned, the periplasmic ASNase is the one historically used for ALL treatments and consequently these two transcripts were evaluated for the presence of CDS of congruous size. However, both transcripts contained short amino acid sequences (180 aa in both cases of “TR4418|c0_g1_i1 len=673” and “TR4418|c0_g1_i2 len= 678”) and BlastP aligned them with bacterial asparaginases. This probably indicates that these sequences belong to bacterial contaminants that evaded transcriptome filtering.

The protein sequence of CcASNase were aligned with other previously described type I and type II ASNases and with the other ASNase found in diatoms (using on NCBI Protein database the queries “l-asparaginase” and “60 kDa lysophospholipase; *Figure 4.2*). Based on the sequence alignment and on the know active site for AsnA (which is composed by T14, T91, D92 and K163; (Yun *et al.*, 2007)), it was possible to deduce the putative catalytic site of CcASNase composed of T157, T238, D239 and K310. All the other ASNase conserved domains are equally conserved in diatoms as well, with the exception of the F349 that is usually shared among ASNase II sequences, while a L residue is more common in ASNase of Class I. *C.closterium* shares this feature with *C. tenuissimus* (having a F316 residue).
Figure 4.2 Protein sequence alignment of the newfound CcASNase (boxed in red) with other type II and type I l-asparaginases. The protein sequences used in the alignment are: L-asparaginase from bacterium Erwinia chrisantemi (strain 3937) (UniProt: E0SD18_DICD3); L-asparaginase 2 from yeast Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (UniProt: ASP21_YEAST); L-asparaginase 1 and 2 from E. coli (strain K12) (UniProt: ASPG1_ECOLI and ASPG2_ECOLI, respectively); Chaetoceros tenuissimus (GenBank: GFH60632.1; Hongo et al., 2021); Fistulifera solaris (GenBank: GAX12934.1; Tanaka et al., 2015); Nitzschia incospicua (GenBank: KAG7373112.1; Oliver et al., 2021). Conserved regions are boxed in black, while the putative amino acids of the catalytic site are marked with a red star.

Finally, the preliminary modelling of CcASNase performed via SWISS-MODEL identified the enzyme native conformation as a homodimer as shown in Figure 4.3, with the crescent shape already observed in the two
known Class I ASNases crystal structures (Yao et al., 2005; Yun et al., 2007). However, the residues involved in dimerization and tetramerization were not conserved in CcASNase nor in any aligned diatom.

Figure 4.3 Preliminary homology model via SWISS-MODEL of CcASNase shown as a crescent-shaped dimer. Chain 1 is shown in yellow, chain 2 in light green.

4.3.2 PCR Isolation of cc_asnase coding sequence and cloning in TOPO-TA vector

The cc_asnase coding sequence was successfully isolated via PCR by using 20 ng of cDNA template, as in General Methods par. M6. cc_asnase CDS isolation was performed by employing different primer combinations and with an annealing temperature of 55°C (Figure 4.4). The combination F1R1 never gave a visible amplification product, while the primer couples F1R2 and F2R1 gave different products of shorter length than expected (a strong amplification of 1300 bp and another fainter amplification of 1000 bp). The selected annealing temperature (55°C) could have caused some of these nonspecific products. The PCR product obtained using the primer couple F2R2 was strong and without aspecific bands, and was selected for purification and subcloning in TOPO-TA vector. Before proceeding with the AMPure purification three separate reaction tubes using primers F2R2 were set-up and combined.
Figure 4.4 Agarose gel of cc_asnas CDS amplified with different primer couples (F1R1, F2R2, F1R2, F2R1).

Purified cc_asnase was successfully cloned into the pCR©2.1-TOPO™ vector (InvitroGen) as reported in General Methodologies, par. M8. Successful cloning was proven via enzymatic digestion as reported in General Methodologies, par. M9 following the same rationale already explained in paragraph 2.3.6. The restriction enzymes employed and the expected size of the products are listed in Table 4.4.

Table 4.4 Restriction enzyme reactions done in order to verify the insert presence.

<table>
<thead>
<tr>
<th>Cloned Gene of Interest</th>
<th>Used Enzyme</th>
<th>N. of Insert Restriction Sites in (position)</th>
<th>Expected Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcASNase</td>
<td>HindIII</td>
<td>1 (1440-1445)</td>
<td>Linearized vector of 4092 bp + 1439 bp fragment</td>
</tr>
</tbody>
</table>

Results of DNA sequencing of cc_asnase/ pCR©2.1-TOPO™ did not show base changes when compared to the original transcriptome sequence (Figure 4.5).
Figure 4.5 cc_asnase CDS alignment from transcriptome and newly sequenced, as visualized via Jalview.

4.3.3 Cloning and sequencing of cc_asnase in pJ411 vectors

cc_asnase CDS sequences with pJ411/pOPIN adaptors were produced using 5 ng of DNA preparation of cc_asnase/pCR©2.1-TOPO™ template. The PCR reactions produced the expected sequences, as shown in Figure 4.6.

Figure 4.6 Agarose gel of cc_asnase CDS amplified with different adaptors for homologous recombination-based cloning. A. cc_asnase CDS amplified with pJ411_NTER-CTER adaptors; B. cc_asnase CDS amplified with pOPINF adaptors.
The produced CDS were treated with DpnI enzyme, and then purified. The \textit{cc\_asnase} + pJ411\_NTER-CTER adaptors PCR product was purified using AMPure purification, while the \textit{cc\_asnase} + pOPINF adaptor PCR product was purified with the GenElute Gel Extraction kit, given the presence of aspecific products of c.a. 700 bp and possibly of 2000 bp. The produced \textit{cc\_asnase} + pJ411\_NTER-CTER were cloned in pJ411 vectors as reported in General Methodologies, par. M10, while the \textit{cc\_asnase} + pOPINF adaptor product was retained for future use. The sequences generated were cloned in pJ411 vectors as reported in General Methodologies, par. M11. Results of the DNA sequencing of \textit{cc\_asnase}/pJ411\_NTER did show a single base change compared to the previous sequence, corresponding to base n. 957, that changed from cytosine to thymine (C to T). Nonetheless it was decided to use the plasmid for heterologous expression, since the mutation was synonymous (both TTC and TTT codons encode for the amino acid Phenylalanine). Finally, \textit{cc\_asnase}/pJ411\_NTER-CTER plasmid maps were generated using SnapGene and are shown in Figure 4.7.
Figure 4.7 Plasmid maps of cc_asnase/pJ411_NTER-CTER produced using SnapGene. A. cc_asnase/pJ411_NTER; B. cc_asnase/pJ411_CTER. The inserted gene DNA is showed in red and the related protein in dark purple (CcASNase). All plasmids show their size (bp), the lac operator (blue), the T7 promotor (arrowhead white), the replication origin (ori, yellow), the Shine-Dalgarno sequence (grey) and the Kanamicin resistance gene (KanR, light green). pJ411_NTER maps show the N-terminus His-tag (6xHis, light purple) and the cleavage site for the TEV protease (grey); pJ411_CTER maps show the C-terminus His-tag (6xHis, light purple).

4.3.4 Heterologous expression of CcASNase/pJ411 proteins

CcASNase heterologous expression screening followed the specifics reported in paragraph 4.2.2. Using Protpi the molecular mass of CcASNase was predicted. CcPAF-AH1 was predicted to have a mass of 59,24 kDa with the fusion tag of pJ411_NTER and 57,58 kDa with the fusion tag of pJ411_CTER, respectively. Gel images are shown in Figure 4.8 and 4.9.
Figure 4.8 SDS-PAGE of soluble fractions of protein extracts of Lemo 21, BL21 (DE3) and Rosetta 2 (DE3) cells transformed with cc_asnase/pJ411_NTER and cc_asnase/pJ411_CTER. Each lane triplet is occupied from left to right by samples from 37°C Induction, 16°C Induction and Control, respectively. Used marker (M) was the Precision Plus ProteinTM All Blue Standards (BioRad, USA).
Even if the samples relative to \textit{cc\_asnase/pJ411\_CTER} expressed in BL21 (DE3) presented a faint signal (probably due to partial cell lysis), in no other case was the protein of interest expressed at appreciable levels in any of the tested conditions. In order to confirm the presence of faint levels of expression or the complete lack of expression, a Western-Blot analysis was performed, as in General Methodologies, par.M14.

1-4 µL of soluble fractions of \textit{Cc\_ASNASE-pJ411\_NTER-CTER/BL21 (DE3)} samples were loaded for every expression temperature (plus the non-induced control, NI).

BL21 (DE3) samples were selected for the first Western-Blot analysis in order to compensate for the partial lysis of the \textit{pJ411\_CTER} samples. The amount of protein loaded was normalized using the BCA protein assay (Thermo Fisher Scientifics) following the manufacturer’s instructions and the 96 well plate assay protocol.

The total protein concentration of the \textit{Cc\_ASNASE-pJ411\_NTER/BL21 (DE3)} samples was: .1.13 mg/mL for 37°C sample, 1.42 mg/mL for 20°C sample and 0.8 mg/mL for the NI sample. The total protein concentration of \textit{Cc\_ASNASE-pJ411\_CTER/ BL21 (DE3)} samples was: .0.32 mg/mL for 37°C sample, 0.28 mg/mL for 20°C sample, and 0.4 mg/mL for the sample. A photograph of the detected filter taken with ChemiDocTM is shown in Figure 4.10.
In both the BL21 (DE3) sample set, a faint induction of the His-tagged proteins of the expected molecular weight can be detected. In the case of CcASNase/pJ411_NTER, expression of the POI could be detected in all three samples, surprisingly with a stronger expression in the NI sample (given its lower total protein concentration), probably due to leaky expression. Otherwise, in the CcASNase/pJ411_CTER samples expression could be detected only in the 37°C sample. However, it is difficult to determine if this could due on a real lack of expression or on the quality of tested samples.

4.4 Conclusions

CcASNase was selected as a protein of interest in order to explore new sources for ASNase, that could benefit both the therapeutic and food treatment market sectors. While novel marine ASNase sources are not unique of (Izadpanah et al., 2018), to my knowledge there is a complete lack of biochemical and molecular data regarding isolated ASNase genes and proteins from microalgae and this chapter represents a first step in this direction.

The planned experimental pipeline included CcASNase CDS isolation and its cloning in both TOPO-TA and pJ411 vectors. However, even if the CDS isolation of cc_asnase and its cloning was successfully achieved, the E. coli- based heterologous expression system proved unsuccessful, at least with the vectors and expression conditions employed. However, more than a complete lack of expression, as in CcPAF-AH1-2, in this case an expression detectable only using Western-Blot analysis was achieved. Consequently, future approaches could make use of different expression vectors or growth conditions in order to increase the yield of the protein of interest up to mg/L of growth culture, to generate sufficient quantities for subsequent purification and extensive activity testing against isolated substrates and on different cell lines. A monodispersed protein
solution is required to perform robust biochemical and kinetic data that could then be subsequently with the vast literature available for bacterial and fungal ASNases.

While *E. coli* BL21 (DE3) is used on a routine basis for heterologous expression, purification and biochemical characterization of ASNases identified in other microorganisms (Wang *et al.*, 2021), in our case it appeared to be suboptimal, even if a faint expression was detected via Western-Blot.

The preferential usage of the translational machinery of the cell of some codons over the others, called codon usage bias (CUB), has not been extensively studied in diatoms, but the CUB of *C. closterium* has been studied in a recent work of Krasovec and Filatov (Krasovec and Filatov, 2019). *C. closterium* CUB was evaluated overall as very modest, a 55.57 score in the theoretical range from 61 (no bias) to 20 (maximal codon bias). Moreover, third position GC content of preferred codons was of 0.57, to say that C/G appeared the 57% of the times in the preferred codon third position; this means that there is only a weak GC-rich prevalence in *C. closterium* preferred codons.

In *E. coli* the most “difficult” codons (that can slow down and hamper translation and impact correct protein folding (Hanson and Coller, 2018)) are the AGA, AGG, AUA, CUA, GGA, CCC, and CGG rare codons whose related tRNA genes are provided in Rosetta 2 (DE3) strains (Tegel *et al.*, 2010). The supply of these codons was relevant in the expression screening of S243A CcPAF-AH2, that was more efficiently produced in Rosetta (DE3) strain.

In case of CcASNase, instead, the use of Rosetta 2 (DE3) did not provide a visible benefit, even if the sequence possessed a Codon Adaptation Index (CAI) higher than CcPAF-AH2, as estimated using JCat (Sharp and Li, 1987; Puigbò, Bravo and Garcia-Vallve, 2008). Even using the same host strain and growth and induction conditions, the supply of *E. coli* rare codons of course cannot guarantee alone high expression levels of the POI. In fact, in the previous literature the recombinant protein that was produced more abundantly (Δ5Des-Iso (Thiyagarajan *et al.*, 2021)) showed a modest CAI, while the gene with the highest CAI (elkj (Niu *et al.*, 2009)) showed very low levels of production.

In Table 4.5 the different microalgal genes expressed in *E. coli* (and mentioned in Chapter I) and their CAI calculated via JCat are shown.
Table 4.5 Recombinant microalgal genes expressed in E. coli strains. The source microalga, the gene name, their CAI calculated via JCat, the eventual induced toxicity in bacterial host and the references are reported.

<table>
<thead>
<tr>
<th>Microalga</th>
<th>Gene</th>
<th>Codon Adaptation Index</th>
<th>Toxicity effect</th>
<th>Reference (GenBank Accession Number)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Botryococcus braunii</td>
<td>Squalene synthase</td>
<td>N.D.</td>
<td>No</td>
<td>(Okada, Devarenne and Chappell, 2000) (AH009227.2, partial CDS)</td>
</tr>
<tr>
<td>Dunaliella salina</td>
<td>β-ketoacyl-coenzyme A synthases (Kcs)</td>
<td>0.27</td>
<td>No</td>
<td>(Azachi et al., 2002) (AF333040.1)</td>
</tr>
<tr>
<td>Pavlova viridis</td>
<td>C20 elongase (elkj)</td>
<td>0.339</td>
<td>Yes</td>
<td>(Niu et al., 2009) (EF486525.1)</td>
</tr>
<tr>
<td>Pavlova viridis</td>
<td>Δ4 fatty acid desaturase (Δ4Des4)</td>
<td>0.280</td>
<td>Yes (without fusion with Mistic)</td>
<td>(Xu, Niu and Kong, 2011) (GUS94191.1)</td>
</tr>
<tr>
<td>Isochrysis sp. SKMKUIso1</td>
<td>Δ6 fatty acid desaturase (Δ6Des-Iso)</td>
<td>0.229</td>
<td>No</td>
<td>(Thiyagarajan et al., 2018) (KR005946.1)</td>
</tr>
<tr>
<td>Pavlova sp.</td>
<td>Δ5-desaturase (Δ5Des-Pav)</td>
<td>N.D.</td>
<td>No</td>
<td>(Thiyagarajan, Arumugam and Kathiresan, 2020) (MK579398.1)</td>
</tr>
<tr>
<td>Isochrysis galbana</td>
<td>Δ5-Desaturase (Δ5Des-Iso)</td>
<td>0.268</td>
<td>No</td>
<td>(Thiyagarajan et al., 2021) (KR062001.1)</td>
</tr>
<tr>
<td>Cylindrotheca closterium</td>
<td>Dibutyl phthalate hydrolase 1 (cc_dbph1)</td>
<td>0.193</td>
<td>No</td>
<td>This thesis</td>
</tr>
<tr>
<td>Cylindrotheca closterium</td>
<td>Dibutyl phthalate hydrolase 2 (cc_dbph2)</td>
<td>0.235</td>
<td>No</td>
<td>This thesis</td>
</tr>
<tr>
<td>Cylindrotheca closterium</td>
<td>Platelet activating factor – acetylhydrolase 2 (cc_pafah2)</td>
<td>0.177</td>
<td>Yes (needed mutation of the catalytic S243)</td>
<td>This thesis</td>
</tr>
<tr>
<td>Cylindrotheca closterium</td>
<td>L-Asparaginase (cc_asnase)</td>
<td>0.195</td>
<td>N.D.</td>
<td>This thesis</td>
</tr>
</tbody>
</table>

An increase in protein expression levels could be obtained using plasmids of the pOPIN series that have a larger N-terminus fusion protein tag, such as pOPINS3C, pOPINTRX and pOPINHALO7. Higher sized fusion tags can, in fact, increase protein solubility (Bird, 2011; Rosano and Ceccarelli, 2014) and, in the case of eukaryotic proteins with “difficult” codon usage characteristics for bacteria, also help with the translation start and completion (Costa et al., 2014). The plasmid map of cc_asnase/pOPINS3C-TRX-HALO7 was generated using SnapGene and is shown in Figure 4.11.
If proper expression conditions for CcASNase are achieved, and a monodispersed protein solution is obtained, future approaches can also involve: 1) enzymatic *in vitro* assays in order to evaluate protein activity against the expected substrates (L-asparagine, glutamine), its pH and temperature range of activity and kinetic parameters such as Km, even if without an allosteric regulation (these data could already elucidate if CcASNase can have potential therapeutic or acrylamide-mitigation applications); 2) CcASNase activity and cytotoxicity testing in cell lines already used in the literature such as human T-lymphoblast acute lymphoblastic leukemia cell lines MOLT-4, the promyelocytic leukemia cells HL-60 and acute monocytic leukemia THP1 cells (Husain *et al.*, 2016; Belviso *et al.*, 2017; Saeed *et al.*, 2020), while in the case of solid tumors the adenocarcinoma human alveolar basal epithelial cells A549 (Saeed *et al.*, 2020; Dumina *et al.*, 2021) and the epidermoid carcinoma cells Hep-2 (El-Naggar *et al.*, 2016; El-Gendy *et al.*, 2021) have already been used.

Finally, a pipeline as the one presented in this Chapter could also involve the characterization of other
microalgal ASNases found in annotated transcriptomes and genomes, such as the L-asparaginase from the dinoflagellate *Amphidinium carterae* (Lauritano *et al.*, 2017).

Overall, this chapter represents the first attempt to identify and characterize from a molecular and biochemical point of view an ASNase from a marine microalga. Bioinformatic identification identified CcASNase as a type I asparaginase that is recently receiving more scientific attention after decades of being overshadowed by the higher-affinity type II enzymes. Recombinant protein production still has to be achieved to generate satisfactory protein yield, but the increasing number of papers on ASNase activity characterization is likely to make the future evaluation of the biotechnological potential of this novel enzyme more feasible.
Chapter 4 References:


Rottmann, E. et al. (2021) ‘Enzymatic acrylamide mitigation in French fries – An industrial-scale case study’, 167


Chapter V

Final Conclusions

PhD final results flowchart
Chapter 5  Final Conclusions

This PhD thesis is a small contribution to the novel field of microalgal biotechnological enzyme discovery, a field where, as discussed in Chapter I, annotated enzyme sequences are still scarce and mainly linked to the fields of high value-added lipids and carotenoids production (Brasil et al., 2017; Galarza et al., 2018; Vingiani et al., 2019; Avila-Roman et al., 2021; Montoya et al., 2021; Thiyagarajan et al., 2021; Yazdani et al., 2021; Hassan et al., 2022; Santin et al., 2022). In fact, while microalgae produce several bioactive compounds with pharmaceutical and nutraceutical applications (Martínez Andrade et al., 2018; Saide et al., 2021; Hassan et al., 2022), in many cases, the enzymatic pathway responsible for their synthesis is still unknown, hampering the engineering processes that could be developed to optimize compound production and applications. Even if recent discoveries have expanded the molecular toolkits available for microalgal strains, these still lag far behind those available for bacteria and yeast (Gimpel, Henríquez and Mayfield, 2015; Sproles et al., 2021).

Other lab-scale tested applications of microalgal enzymes include the use of laccase and lignin peroxidase enzymes in dye decolourisation (Abd Ellatif, El-Sheekh and Senousy, 2020), β-galactosidase in lactose removal in the food industry (Suwal et al., 2019; Zanette et al., 2019), phytase for phosphorus bioavailability in food products (Spier et al., 2020), lipase for the food and detergent industries (Brasil et al., 2017) and protease for therapeutic applications (Páblo Eugênio da Costa e et al., 2017). While many authors have been able to characterize the molecular weight, optimal pH and temperature of specific microalgal enzymes (Spier et al., 2020) via their direct purification from microalgal cultures, further characterization of enzymes’ biochemical properties is pivotal for their industrial application.

While my thesis mainly focused on the transcriptome of Cylindrotheca closterium, used experimental pipeline can be applied to other transcriptomes of microalgae of biotechnological interest virtually without any changes. Few -omic datasets are available for Cylindrotheca closterium other than its transcriptome sequencing and annotation (Elagoz, Ambrosino and Lauritano, 2020). Very recently, whole genome sequencing of Cylindrotheca closterium has been undertaken, which sequence deposited on GenBank (Deposited on 2022-01-18; https://www.ncbi.nlm.nih.gov/sra/SRX13820307). This additional resource, which may allow genetic engineering manipulations as well, will speed-up the possible biotechnological applications of C. closterium in the future.

My PhD thesis described four attempts of heterologous expression of microalgal genes in E. coli strains (CcDBPH1, CcDBPH2, CcPAF-AH2, CcASNase), with a variable degree of success and protein production. Even if different microalgal enzymes have already been produced in BL21 (DE3), this thesis is the first example of where a purification protocol (for CcDBPH1-2) was successfully attempted. Moreover, among the different vectors available for heterologous expression in bacterial host, the pOPIN and pJ411 vector families have not previously been employed to express microalgal genes.
CcDBPH1-2 clearly received most of the experimental based efforts in this thesis, and the focus given to their CDS isolation, PCR products purification and cloning in TOPO-TA and pOPIN vectors can be justified as these were the “first” enzymes approached during my PhD. As it was evident from Chapters 3 and 4, other enzymes’ characterization was not described to the same level of detail, to avoid redundant explanations. The following summarizes the different results presented in each chapter:

Chapter 1 elucidated the current state of art in the literature of biotechnological enzymes from microalgae, elucidating the most studied fields (value-added lipids- and carotenoids- involved enzymes), and the least studied fields (bioremediation-involved enzymes and enzymes directly usable in industry, such as ASNase). This chapter was integrated by an overview of the expected goals of my PhD. Future approaches could involve the writing of reviews more focused on the individual fields of application of microalgal enzymes, with critical evaluation of the “best” enzymes on which to invest further studies based on the abundance of literature and on the success ratio of the different approaches. Recently, a review of Santin et al. (Santin et al., 2022) similarly focused on microalgal enzymes involved in PUFAs production.

Chapter 2 focused on correlating the phthalate-degrading activity observed in C. closterium with the putative involved enzymes. To achieve this goal, RT-qPCR analysis to study expression levels of CcDBPH1-2 in different culture conditions of C. closterium, the 3D modelling of CcDBPH1-2 and CcDBPH1 docking simulation with the putative DBP substrate was undertaken and successfully implemented. The enzymes’ CDS isolation, production and purification of the experimental pipeline proceeded successfully up to their partial purification and activity testing. The primary follow-up of the presented pipeline is the absence of activity assays of DBP degradation from purified proteins or bacterial lysates, which would definitely confirm CcDBPH1-2 involvement in PE degradation. A more stringent purification of CcDBPH1-2, up to a monodispersed solution, on the other side is mandatory to obtain precise biochemical data of the newfound enzymes (optimal pH and temperature of activity, interaction with detergents/metal ions) and compare them with previously characterized bacterial PE-esterases.

Chapter 3 focused on the platelet activating factor acetylhydrolase enzymes (namely CcPAF-AH1 and 2). While their CDS isolation was hindered by aspecific amplifications, the main problem encountered was the inability to achieve protein production due to the hypothesized toxicity of both proteins in bacterial hosts. Thanks to participation in the transnational access within ASSEMBLE+ and the collaboration with researchers at the BSRC of the University of St. Andrews, this toxicity was circumvented via mutation of the catalytic serine. However, the protein was still found in the insoluble fraction and it was impossible to make it soluble via protein denaturation with urea. While the purification of the mutated CcPAF-AH2 could be achieved via a protocol for membrane proteins, involving high-speed centrifuge and non-ionic surfactants, it is also true that structure resolution of membrane proteins, while easier than in the past, is still a non-trivial challenge to overcome. Alternatively, the host organism could be changed, instead of focusing on the purification from
the membrane fraction: \textit{P. tricornutum}, as a diatom, could be useful to better assess the activity and the subcellular localization of PAF-AH1-2; the fungus \textit{P. pastoris} with its secretory pathway could completely circumvent the toxic phenotype caused by the enzymes; finally, insect cells could also be used, exploiting the features included in pOPIN vectors for the heterologous expression in insect and mammalian cell lines, other than in bacteria.

\textbf{Chapter 4} finally included some preliminary experiments on a type I L-asparaginase (CcASNase), that can have applications both in cancer treatment and in acrylamide remediation. While heterologous expression of this gene was successful, currently only low levels of production have been achieved, with expression only being detectable via Western-Blot. In order to do so different approaches can be used, such as expression plasmids with N-terminus big tags, that can positively impact on the production rate of the protein of interest; other strategies involve the use of \textit{ad hoc} modified growth medium in order to reach higher cell density and, consequently, higher amounts of CcASNase. Independently from the strategy set, when optimal expression conditions are achieved, the activity assays for ASNase will benefit from the previous literature on bacterial and eukaryotic ASNases.

\textbf{Table 5.1} summarizes the main achievements and future perspectives of each previous thesis \textbf{Chapter} while \textbf{Figure 5.1} shows an updated version of \textbf{Figure 1.6}, with the updated milestone structure.

<table>
<thead>
<tr>
<th>Thesis Chapter</th>
<th>Key Achievements</th>
<th>Future Perspectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 1</td>
<td>- A review of the actual state of art of scientific literature regarding microalgal enzymes with biotechnological applications, grouped in three main application fields (lipid expression, healthcare, bioremediation) updated to 2021 was produced</td>
<td>- New reviews, more focused on one particular application field, while more and more scientific interest focuses in this field, can be produced</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>- CcDBPH1-2 sequences were identified via TBlastN, using all the available PE hydrolases as queries - CcDBPH1-2 CDS was isolated and cloned in TOPO-TA and pOPIN vectors - CcDBPH1-2 RT-qPCR gene expression analysis showed an overexpression DBP concentration- and time- dependant - CcDBPH1-2 3D model was produced - CcDBPH1 docking simulation showed a favourable affinity energy between DBP and the protein active site - CcDBPH1-2 were produced in \textit{E. coli} and partially purified from 1L cultures - CcDBPH1-2 esterase activity was tested and confirmed</td>
<td>- Activity assay against DBP can be checked directly using bacterial lysates that have expressed CcDBPH1, CcDBPH2 or both proteins - A monodispersed solution of CcDBPH1 and CcDBPH2 can be obtained, in order to have robust data of their biochemical properties (optimal pH, temperature, interaction with detergents/metal ions, kinetics)</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>- CcPAF-AH1-2 CDS was isolated and cloned in pOPIN and pL411 vectors - CcPAF-AH2 S243A was mutated in order to circumvent its toxic phenotype in \textit{E. coli} - CcPAF-AH2 S243A was produced in 1L cultures</td>
<td>- CcPAF-AH2 S243A purification can be achieved with a protocol for membrane proteins (high-speed centrifuge, non-ionic surfactants) - Alternative heterologous expression in \textit{P. tricornutum} or \textit{P. pastoris} can be designed to circumvent the toxic phenotype without losing enzyme activity</td>
</tr>
</tbody>
</table>
Chapter 4

- CcASNase CDS was isolated and cloned in TOPO-TA and pJ411 vectors
- CcASNase was produced in E. coli in low amounts (detectable only via Western-Blot)

- High-rate expression and protein expression can be achieved using N-terminus tags or other growth and induction conditions
- Protein purification and biochemical analysis can be achieved and benefit from previous literature
- Activity assay can be done on target aminoacids (asparagine, glutamine) or cancer cell lines

2x Active Recombinant Enzymes:
- DBPH1 and DBPH2 (C. closterium)

5x Enzymes of Biotechnological Interest:
- DBPH1, DBPH2, PAF-AH1, PAF-AH2, L-ASP (from C. closterium)

Microalgae may be considered an eco-friendly and eco-sustainable source of novel enzymes that can potentially meet the growing industrial demand and market. While microalgal enzyme biotechnological exploitation seems still far, new studies (such as those focused on microalgae engineering) are focused in this direction and will increase the microalgal portfolio of products for industrial applications. Many microalgal biosynthetic pathways of biotechnological interest will benefit from approaches similar to the one followed in my work of thesis; this could elucidate their role and potential, benefitting from the increasing amount of –omics data that is adding up.

That, if scientific community and market stakeholders will be able to reinforce the academia-business communication, in order to avoid the common problem of “enzyme-hoarding”. Once discovered and characterized, enzymes for biotechnological application need to leave the lab and be framed in value propositions or patents that can enforce their application on higher scale projects and, overall, on answering to industrial needs.

Figure 5.1 PhD final milestone structure.
Chapter 5 References:


Appendix M

General Methodologies

AKTA start purification system and related buffers and consumables at the Biomedical Sciences Research Complex (BSRC) at the University of St. Andrews (St. Andrews, UK)
**M1. Microalgae inoculation, growth, scale-up and harvesting processes**

Microalgal inoculation, growth and scale-up processes were conducted under aseptic conditions using a laminar flow hood (CLEAN FLUX O/130; Folabo instruments, Italy). All microalgal species were taken from the Stazione Zoologica Anton Dohrn Culture Collection. For routine culture maintenance, stocks were maintained in 30 mL cell-culture flasks: once per week up to 5 drops (depending on the initial cell concentration) of a previous microalgal stock were added to a 30 mL aseptic cultivation flask together with 25 mL of Guillard’s F/2 medium, under a laminar flow hood. Guillard medium was prepared following the original recipe (Guillard, 1975). In the case of flagellates, Guillard’s F/2 medium was prepared without silica addition. In the case of experiments with bigger culture volumes stocks of 50 mL, 250 mL or 2L were prepared.

To start an experiment, the mean value of cells present in the inoculum was measured using an Axioscope 2 MOT (ZEISS, Germany) microscope and 1 mL of the culture sample was loaded on a 1 mL counting chamber (Hausser Scientific, USA). The samples for counting were fixed with 1 mL of Lugol solution (Merck, USA).

To harvest the microalgal biomass, aliquots of 50 mL were centrifuged in a Beckman Coulter Allegra® 6R centrifuge for 30 min at 4200 rpm. The supernatant was then discarded, leaving only 1 mL of volume to further resuspend microalgal biomass. Samples were transferred to a 1.5 mL Eppendorf and further centrifuged for 15’ at 12000 rpm and 4°C. The supernatant was again discarded. For each sample, 500 µl of TRIzol (Life technologies, USA) was added for RNA extraction experiments and the samples were incubated at 60°C for 3’ in agitation using the “Basic cooling and heating thermoshaker” (Carl Roth, Germany). TRIzol reagent is a monophasic solution of phenol, guanidine isothiocyanate, used for RNA extractions. After incubation, samples were frozen in liquid nitrogen and stored at -80°C.

The following material was used for inoculation, growth and scale-up processes:

- Pipettor ErgoOne® FAST (Starlab, Italy)
- Aseptic pipets (5 mL, 10 mL, 25 mL and 50 mL)
- Aseptic cultivation flasks w/ air filter (30 mL, 250 mL, 500 mL and 1L)
- Autoclaved carboys w/ aeration system (2L and 10L)
- Microscope ZEISS Axioscope 2 MOT (ZEISS)
- Microscope ZEISS Axiovert 200 (ZEISS)
- 1 mL counting chamber
- Lugol’s Solution

Microalgae were cultivated in climate chambers with controlled temperature and light conditions (i.e. at 20°C, on a 12h:12h light: dark cycle at 110 μmol·m-2·s-1 photons). In addition, filtered air (via 0.22 μm Millipore filters) was supplied.
Filtered sea water and Milli-Q high-purity water are the base solvents of both mediums. Fresh sea water was collected weekly from the gulf of Naples and filtered according to the protocol shown below: Milli-Q water was obtained from the system Millipore Synergy 185 (Merck) and was added to sea water in order to adjust its salinity to a 1:20 ratio (e.g. 50 mL Milli-Q for each 950 mL of seawater; **Figure M1**).

The following material was used for the preparation of filtered sea-water:

- Millipore filtration apparatus (90mm filter holder, 275 psi max inlet)
- Millipore 0.22 µm filter (White GSWP, 90mm)
- Millipore glass fibre pre-filter
- KNF Neuberger N811 KN.18 vacuum pump
- Trap (glass Erlenmeyer with vacuum adapter)
- 10L glass carboy with vacuum adapter (Schott DURAN, Germany)
- Kartell 10L sea-water plastic recipient

**Figure M1.** Sea water filtration assemblage (left); Millipore filtration apparatus (right). The water is collected in the Schott DURAN 10L glass carboy with vacuum adapter.

The sterilization of the prepared media was achieved using the autoclave model FVG3 (Fedegari, Italy).

In order to stress the microalgae and also stimulate the activation of metabolic pathways that could be silent in control conditions, nutrient starvation experiments (e.g. minus phosphate, minus silica) were also performed.
M2. Evaluation of microalgae growth curves

For the evaluation of a growth curve it is necessary to measure microalgal concentrations in a given volume over a certain period of time. Culture samples were collected every 24 hours starting from the day of inoculation. The shape of the curve and the total lifetime will depend on the microalgal species.

Cell counts were performed daily (using the same method described before) because growth rates differ for each microalgal species. The Axiovert 200 microscope (ZEISS) was also used to observe the viability of the cells.

Microalgae generally show a sigmoid-like growth pattern which includes: 1) an initial lag phase with a slow growth, 2) an exponential phase where algal growth rate is higher than death rate, followed by a 3) linear phase where the growth rate starts declining, 4) a stationary phase when the number of cells replicated is equal to the number of cells that die, and finally 5) a senescent phase when most cells are dying (Figure M2, (Price, Kelsey; Farag, 2013)).

![Figure M2. Phases of microalgal growth (Price, Kelsey; Farag, 2013)](image)

M3. RNA extraction from microalgal samples

To avoid contamination and degradation of the extracted nucleic acid, the entire protocol was performed under an ASEM Fume Cabinet T100 (ASEM, Italy) hood, using ice for storing samples between different steps. The following steps were performed:

1. The samples were defrosted and briefly agitated, half a spatula of glass beads (300 μm) was added to each sample and the tubes were shaken for 10’ at 60°C and 1400 rpm.
2. The supernatants were transferred into clean tubes and 0.1 ml of chloroform for 0.500 ml of TRIzol (proportion 1:5) was added. The tubes were then hand shaken and incubated at 25°C for 5’.

3. Tubes were centrifuged at 12000 rpm at 4°C for 15’. After this centrifugation 3 layers are expected to appear: an organic layer at the bottom (chloroform), an interphase of precipitated DNA and an aqueous phase on top containing the RNA. It was necessary to remove the top layer carefully and transfer this aqueous phase to a new tube.

4. An equal volume of chloroform to the collected fraction was added to the samples; then the tubes were hand shaken, centrifuged at 12000 rpm at 4°C for 10’, and the supernatant was transferred to a new tube. This wash step was repeated at least twice.

5. After 3 washes the same amount of isopropyl alcohol was added to the collected fraction, together with 1 μL of glycogen (Invitrogen), vortexed by hand and incubated at -20°C for 1h. The RNA will precipitate in the isopropyl alcohol/water mixture (to ensure the precipitation of RNA it is optimal to add frozen cold isopropyl alcohol). After 1h the mixture was centrifuged at 12000 rpm at 4°C for 15’. The RNA is visible after this step as a white pellet.

6. The supernatant was removed and 1 ml of 75% ethanol was added to each sample, vortexed and centrifuged at 12000 rpm at 4°C for 15’ in order to remove the isopropyl alcohol. This step was repeated twice, removing the supernatant each time.

7. Finally, the RNA pellet was air-dried by leaving the Eppendorf open for at least 20’. RNA was dissolved by adding 20-30 μl of DEPC H₂O (water with the addition of Di-ethyl Pyrocarbonate, an RNase inhibitor) and incubated at 60°C for 10’, agitated for 15” and centrifuged at 12000 rpm for 10”.

The RNA was stored at -20°C for immediate use, or stored at -80°C for longer periods.

RNA concentration and quality were checked using the NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific). RNA quality parameters were based on the 260/280 and 260/230 nm ratios of absorbance. According to the T009-TECHNICAL BULLETIN of NanoDrop™ 1000 Spectrophotometer and the NanoDrop™ 8000 Spectrophotometer, the 260/230 value for “pure” nucleic acid is usually higher than the respective 260/280 values. Expected 280/260 ratio of 2.0 is generally accepted as “pure” for RNA, while 260/230 values are commonly in the range of 2.0-2.2. As a rule of thumb, if both ratios are appreciably lower than expected, it may indicate the presence of contaminants, such as: proteins which absorb at 280 nm and phenols, which absorb both at 280 and 230 nm (TRIzol is a phenolic solution, and can contaminate RNA extraction until the end).

**M4. cDNA production**
Reverse transcription, conversion of RNA into the corresponding cDNA, was performed by using the iScript™ cDNA Synthesis Kit (BioRad, USA) or the PrimeScript RT Reagent Kit with gDNA Eraser (Takara, Japan) following the manufacturer's instruction. For both kits the starting amount of 1 µg of RNA was used.

Reactions were run on the T100™ Thermal Cycler or the C1000 Touch Thermal Cycler (BioRad).

**M5. Primer designing for endpoint polymerase chain reaction (PCR) and reverse transcription-quantitative PCR (RT-qPCR)**

Primers for RT-qPCR and PCR experiments regarding genes of interest (GOI) and for DNA sequencing of produced PCR products or plasmids were designed using the software Primer3 v. 0.4.0 (http://bioinfo.ut.ee/primer3-0.4.0/primer3/). The standard and adjusted parameters used for primer design for RT-qPCR are fully shown in Table M2.

**Table M2 Default and adjusted parameters used for RT-qPCR primer design**

<table>
<thead>
<tr>
<th>Setting primer design conditions for RT-qPCR</th>
<th>Objective Function Penalty Weights for Primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Primer Picking Conditions - Adjusted</td>
<td>Lt</td>
</tr>
<tr>
<td>Min</td>
<td>Opt</td>
</tr>
<tr>
<td>Primer Size</td>
<td>18</td>
</tr>
<tr>
<td>Primer Tm</td>
<td>59</td>
</tr>
<tr>
<td>Product Tm</td>
<td>-</td>
</tr>
<tr>
<td>Primer GC%</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Input</td>
</tr>
<tr>
<td>Product Size Ranges</td>
<td>150 - 250</td>
</tr>
<tr>
<td>General standards</td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>Input</td>
</tr>
<tr>
<td>Pick left primer, or use left primer below</td>
<td>+</td>
</tr>
<tr>
<td>Pick right primer, or use right primer below (5' to 3' on opposite strand)</td>
<td>+</td>
</tr>
<tr>
<td>Don't pick hybridization probe (internal oligo), or use oligo below:</td>
<td>-</td>
</tr>
<tr>
<td>Number To Return</td>
<td>5.00</td>
</tr>
<tr>
<td>Max Repeat Mispriming</td>
<td>12.00</td>
</tr>
<tr>
<td>Max Template Mispriming</td>
<td>12.00</td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Max 3' Stability</td>
<td>9.00</td>
</tr>
<tr>
<td>Pair Max Repeat Mispriming</td>
<td>24.00</td>
</tr>
<tr>
<td>Pair Max Template Mispriming</td>
<td>24.00</td>
</tr>
<tr>
<td>General Primer Picking Conditions - Standard</td>
<td></td>
</tr>
<tr>
<td>Max Tm Difference:</td>
<td>100</td>
</tr>
<tr>
<td>Max Self Complementarity:</td>
<td>8</td>
</tr>
<tr>
<td>Max 3' Self Complementarity:</td>
<td>3</td>
</tr>
<tr>
<td>Max #N's:</td>
<td>0</td>
</tr>
<tr>
<td>Max Poly-X:</td>
<td>5</td>
</tr>
<tr>
<td>Inside Target Penalty:</td>
<td>-</td>
</tr>
<tr>
<td>Outside Target Penalty:</td>
<td>0</td>
</tr>
<tr>
<td>First Base Index:</td>
<td>1</td>
</tr>
<tr>
<td>CG Clamp:</td>
<td>0</td>
</tr>
<tr>
<td>Concentration of monovalent cations:</td>
<td>50</td>
</tr>
<tr>
<td>Salt correction formula:</td>
<td>Schildkraut and Lifson 1965</td>
</tr>
<tr>
<td>Concentration of divalent cations:</td>
<td>0</td>
</tr>
<tr>
<td>Concentration of dNTPs</td>
<td>0</td>
</tr>
<tr>
<td>Annealing Oligo Concentration:</td>
<td>50</td>
</tr>
<tr>
<td>Liberal Base</td>
<td>+</td>
</tr>
<tr>
<td>Do not treat ambiguity codes in libraries as consensus</td>
<td>+</td>
</tr>
</tbody>
</table>

The most relevant parameters for RT-qPCR primer designing were: primer size of 18-22 bp, GC content of 50%, primer tails with preferably a G or C (for more stable attachment and primer stability) and amplicon size of 160-250 bp in size, in order to promote more reliable comparison of assays and ensure similar PCR efficiency. For determination of possible primer dimers, hairpin, bulge and internal loops, Gene Runner version 6.5.51 (http://www.generunner.net/) was used.

Primer designing parameters for endpoint PCR were less stringent, preferring primer size of 18-22 bp, GC content around 50%, a melting temperature (Tm) of 52-60°C, primer tails with a G or C at 3’ and with no more than 3 G/C in the last 5 nucleotides (for more stable attachment), G at the 5’ end (in order to increase the probability to add a terminal adenosine overhang, see next paragraph). For CDS isolation the primers were selected at 50-100 bp distance from the start and the end of the GOI, in order to design multiple primer pairs and try different forward/reverse combinations.
Finally, primer designing parameters for DNA sequencing were the same as for traditional PCR, but selecting the primer sequence from inside the CDS, considering that every sequencing reaction has an average efficiency of 800-1000 sequenced bases. The primers were received lyophilized from the manufacturer (Eurofins, Belgium) and stored at -20°C once resuspended at 100 μmol stock concentration and diluted to 20 μmol work concentration.

**M6. PCR reaction and electrophoretic separation**

PCR reactions were performed for different objectives (e.g. in order to check cDNA quality, to check primers efficiency or to amplify the CDS of a GOI).

Amplification reactions were performed using different TAQ Polymerases with their specific buffers following the manufacturer’s instructions; TAQ enzymes used were: TAQ DNA Polymerase (Roche, Switzerland), Expand™ High Fidelity PCR System (Roche, Switzerland), XtraTaq Pol White (GeneSpin, Italy), PrimeSTAR® GXL DNA Polymerase (Takara). The different TAQ Polymerases differed for many features, such as primer concentration needed, reaction total volume, other than length and temperature of the different reaction steps. The DNA template quantity varied between 1-5 ng (in the case of plasmid or purified PCR product template) and 20-50 ng (in the case of cDNA mix template). In order to use purified PCR products as template a good practice was to dilute 1/50 of the PCR product and use 1 μL of such dilution, in order to avoid reaction errors that are more prone to happen using PCR products as a template. Likewise, when using plasmids as templates, PCR samples were treated with DpnI enzyme (New England Biolabs, USA; abbr. NEB) following manufacturer’s instructions before purification, in order to remove residual closed vector template.

One of the most relevant features among different TAQ Polymerases is the addition of a single 3′-adenine to both DNA strands. This overhang nucleotide can be exploited for TOPO-TA cloning, a technique that uses linearized vectors with a-T residue protruding on each 3′ to hybridize with the 3′-overhang adenine left on the PCR fragment and, in the presence of DNA Topoisomerase I (https://www.thermofisher.com/it/en/home/life-science/cloning/topo/topo-resources/the-technology-behind-topo-cloning.html), become ligated together. TAQ DNA Polymerase, Expand™ High Fidelity PCR System and XtraTaq Pol White, for example, produce PCR products with –A protruding.

To avoid picking small volumes from the component tubes and guarantee the highest reproducibility, a mix with H2O + Buffer + dNTPs + Primers + Enzymes for (N + 0,5-1) samples was prepared (where N is the number of samples to amplify via PCR). Finally, in the case of TAQ enzyme with proofreading activity, it is unfeasible to add primers to the mix (because they are single-strand DNA and easily degraded).
Endpoint PCR reactions were performed using the BioRad C1000 Touch™ Thermal Cycler (BioRad). The obtained PCR products were resolved by horizontal electrophoresis in 0.8-2% agarose gels (normally using 1%). These gels were stained with MIDORI\textsuperscript{Green} dye (NIPPON Genetics, Japan) added to liquid agarose (4 µL per 100 mL of agarose gel); DNA dye allows the visualization of DNA bands using an UV-transilluminator. The transilluminators used were: Gel Logic 100 Imaging System UV (Kodak, USA) and the via ChemiDoc\textsuperscript{TM} MP Imaging System (BioRad). The 100 bp Plus, 1kb and λ-2 DNA ladders (Thermo Fisher Scientific) were used as standards. Usually samples of 6 µL in total were loaded in each gel lane: samples were prepared with 5 µL of the PCR reaction and 1µL of 6X Gel Loading Dye; in the case of quantification via λ-2 ladder, samples (or standards) were prepared with 1 µL of the purified PCR product (or DNA standard), 1 µL of 6X Gel Loading Dye and 4 µL of H\textsubscript{2}O.

\textbf{M7. Purification of PCR products and sequencing}

PCR products were purified using the GenElute Gel Extraction Kit (Sigma Aldrich, USA) or the Agencourt AMPure XP PCR purification system (Beckman Coulter, USA), depending if the PCR reaction gave aspecific products or not.

GenElute Gel Extraction Kit purification was performed following the manufacturer’s instructions.

AMPure XP was used following manufacturer’s instructions with the following change:

\begin{itemize}
  \item the volume of AMPure XP used for a single reaction was calculated with the following equation:
  \begin{equation}
  \text{(Volume of AMPure XP per reaction)} = 1.2 \times (\text{PCR reaction volume})
  \end{equation}
\end{itemize}

The purified PCR product was quantified via indirect quantification using a λ-2 ladder or the NanoDrop\textsuperscript{TM} 1000 Spectrophotometer, following the same quality parameters described in par. M3.

Nucleic acid sequencing was performed in order to verify the absence of possible mutations inserted by polymerase errors by the Centro Sequenziamento & Analisi Molecolari Core Facility (CSAM) (http://www.sbmweb.it/) of the Stazione Zoologica Anton Dohrn using the Applied Biosystems 3730 DNA Analyzer (Thermo Fisher Scientific).

PCR DNA samples were prepared in the reaction mix composed of: 15 fmol/µL of DNA template and 4.5 pmol/µL of a single primer, in a total volume of 10 µL. A complete sequence of the PCR fragments usually needed two or more reactions from each side of the double-strand, depending on the GOI length and the efficiency of used primers.

\textbf{M8. RT-qPCR reactions and data analysis}

The fluorescent dye SYBR GREEN was used to evaluate the expression levels of selected genes by RT-qPCR. RT-qPCR was performed in MicroAmp Optical 384-Well reaction plate (Applied Biosystems, USA) with Optical Adhesive Covers (Applied Biosystem) in the ViiA7 Real Time PCR System (Applied Biosystems).
The total volume for each well was 10 µl, with 5µl of 2X Fast Start SYBR Green Master Mix (Roche), 1µl of cDNA template and 0.33 pmol/µl for each primer (usually 4 µL of a 0.7 pmol/µl primer mix).

The RT-qPCR reaction was performed using the following protocol: 95°C for 10’, (95°C for 15” and 60°C for 1’) x 40, 72°C for 5’.

In RT-qPCR, the optimal cDNA template quantity for each species was assessed using serial dilutions, ranging from 1:1 to 1:1000 (approximately from 50 ng to 50 pg of template in each reaction). These dilutions were also used to determine the reaction efficiency for all primer pairs. These efficiencies were calculated generating for each primer pair standard curve with at least five dilution points by using the Cycle Threshold (Ct) value versus the logarithm of each dilution factor and using the equation E=10-1/slope. All RT-qPCR reactions were carried out in technical triplicates to capture intra-assay experimental variability. Each assay included three no-template negative controls (NTC) for each primer pair.

Best reference genes (RGs) for RT-qPCR experiments were identified using three different algorithms, i.e. BestKeeper (Pfaffl et al., 2004), geNorm (Vandesompele et al., 2002) and NormFinder (Andersen, Jensen and Ørntoft, 2004). To study expression levels for each gene of interest relative to the most stable RGs, we used the REST tool (Relative Expression Software Tool (Pfaffl, Horgan and Dempfle, 2002)). Finally, statistical analysis was performed using GraphPad Prime Statistic Software, V4.00 (GraphPad Software; http://www.graphpad.com/quickcalcs/).

**M9. Cloning in TOPO® vectors**

Purified PCR cDNA fragments were inserted in TOPO® vectors using the TOPO TA® or the Zero BluntTM PCR cloning kits (InvitroGen, USA) following the manufacturer instructions, except as follows:

- Insert to vector molarity ratio was set to 10:1 (e.g. If the plasmid size is c.a. 4kb and the insert size is c.a. 1kb, approximate quantity ratio should be 1 ng of vector for 2,5 ng of insert (molarity ratio x bp size ratio = 10/1 x 1/4 = 2,5). Plasmid was used at a concentration of 10 ng/µL and no less than 5 ng of TOPO® plasmid was used for each reaction.
- Once the reaction was set the tube was mixed gently and incubated for 30’ at room temperature.
- Chemically competent or electrocompetent TOP10 *E. coli* cells were used for transformation. In each 200 µL cell vial 3 µL of the TOPO® cloning reaction were added and gently mixed without pipetting. LB broth was used instead of S.O.C. medium. From the 1 mL total transformation reaction volume 20, 200 and 500 µL were spread onto a pre-warmed plate with the addition of selective antibiotic (Ampicillin, Kanamycin) and incubated overnight at 37°C.
Transformation efficiency (the number of clonal colonies that appears on a pre-warmed selective plate incubated overnight at 37°C per μg of template DNA) of used bacterial cells was also calculated as follows:

\[ \text{N. Observed colonies} \times \text{dil. factor/ DNA in the reaction (µg)} \]

(e.g. Using 0,05 µg of Template DNA (50 ng of a starting 100 ng amount, 3 µL of the original 6 µL cloning reaction, were used for transformation) 53 colonies were observed on a plate where 20 µL of the total transformation volume of 100 (Dil. factor = 5) were spread.

Efficiency is: \[53 \times 5/0,033 = 8030 \text{ colony forming units (CFU)} = 8 \times 10^3\].

Colonies on the plate were screened via colony PCR for the expected insert, using PCR reaction settings as described in paragraph M6 and cells from the selected colony (picked with a sterile pipette tip) as template. Positive colonies were transferred to 15 mL Falcon tubes with 3-5 mL of LB broth with selective antibiotic (Ampicillin) and incubated overnight at 37°C in constant agitation.

In order to recover plasmid DNA from bacterial samples the GenElute™ Plasmid Miniprep Kit (Sigma Aldrich) was used following the manufacturer’s instructions. The purified plasmid DNA was quantified and sequenced as described in paragraph M6. In the case of plasmid sequencing, plasmid DNA samples were prepared in the reaction mix composed of: 25 fmol/µL of DNA template and 4.5 pmol/µL of a single primer in a total volume of 10 µL.

### M10. Restriction enzyme reactions

To verify the presence of expected inserts in the purified DNA, or to linearize circular plasmid DNA for further manipulation (as described in par. M11) an enzymatic reaction with restriction endonucleases was performed. The enzymes and buffers used were bought from NEB manufacturer. Reaction setup is summarized in **Table M3**.

**Table M3** Reaction setup for restriction enzyme reactions

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration per reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffer Reaction Mix (10X)</td>
<td>1X</td>
</tr>
<tr>
<td>BSA (Bovine Serum Albumin)</td>
<td>100 µg/ml (if needed)</td>
</tr>
<tr>
<td>Enzyme(^1)</td>
<td>1 U per µg of DNA</td>
</tr>
<tr>
<td>Nuclease-free water</td>
<td>Variable</td>
</tr>
<tr>
<td>MiniPrep template</td>
<td>approx. 100-2000 ng(^2)</td>
</tr>
<tr>
<td>Total volume</td>
<td>Variable</td>
</tr>
</tbody>
</table>

\(^1\) The enzyme volume should not exceed 10% total reaction volume (otherwise the glycerol buffer will impair the reaction and favour “star” activity)
2. If the reaction objective is to check the digestion capacity of the enzyme, then 50-100 ng of template is the right amount to avoid comigration of different bands and noise signals; in case of fragment excision approximately 20 ng are the minimum needed amount to clearly visualize the shorter fragment (Formula: [DNA Amount in the reaction/ (Vector + Fragment total DNA size / Fragment size)]. If the formula result is lower than 20 ng, it is likely that it will not be clearly visualized on agarose gel and a higher template amount should be used).

**M11. Preparation of vectors and PCR fragments for homologous recombination-based cloning**

Once the sequence of interest was isolated (via PCR or cloning in TOPO® vectors) the next step was the designing of forward and reverse primers with adapters, 15-20 nucleotides added to each primer sequence needed to perform homologous recombination (and consequently insertion of the sequence of interest) with the 5’ and 3’ ends of the linearized expression vectors. The designed primers then combined these 15-20 nucleotide extensions and 18-20 nucleotides of the 5’ or 3’ ends of the CDS of the gene of interest. The stringency of primer designing and the usage of TAQ Polymerases that do not add –A protruding (such as the PrimeSTAR® GXL DNA Polymerase) is the reason why previous sequence isolation is a key step for a successful PCR reaction with extension primers.

It is worth mentioning that this PCR reaction with adapter primers can be performed directly on cDNA samples, but the efficiency of this process is extremely variable and generally unfeasible. Once primers were designed and successfully amplified the sequence of interest with adapters (using as template cDNA, purified PCR product or TOPO® vector with the inserted gene, alternatively), the PCR product was quantified and purified as reported in paragraph M6.

The expression vectors used were: pOPIN vectors, designed in the Oxford Protein Production Facility coordinated by Prof. Raymond J Owens (currently renamed OPPF-UK, www.oppf.ox.ac.uk ), described in Berrow et al. (Berrow et al., 2007; Berrow, Alderton and Owens, 2009) and Louise E. Bird (Bird, 2011) and ordered from AddGene, and the PJExpress411-Kan (ATUM, USA) vectors were kindly provided by Prof. Rafael Guimaraes da Silva’s lab (St. Andrews University, St. Andrews, UK). These vectors were used in combination with the commercially available homologous recombination enzyme kits, such as the In-Fusion™ (Takara) or the Gibson Assembly® (NEB).

Based on the abovementioned theory, PCR fragments preparation protocol, originally written by Dr. Luise Bird and available online on the OPPF-UK website (actually updated to the 14th July 2017 version; https://www.oppf.rc-harwell.ac.uk/OPPF/protocols/sop/OPPF-UK%20SOP%20Cloning%20and%20Small%20Scale%20Expression%2020170714.pdf), was followed. Details of the pOPIN plasmids used in this Thesis and of their complementary primer extensions are shown in Table M4.
Table M4. Details of the pOPIN vectors used in the present Thesis. Fusion Tags column lists the aminoacid sequence (1-letter code) fused to the POI. “…” at the beginning of the fusion tag sequence indicating that it is a C-terminus fusion tag; “…” at the end of the fusion tag sequence indicates that it is a N-terminus fusion tag; ↓ represents the 3C protease cleavage site (LFQ/GP). SUMO represents the Small Ubiquitin-like Modifier (Guerrero, 2015) fusion protein. Thioredoxin represents the thioredoxin (Levallie, 2000). HaloTag represents the HaloTag7 (Ohana, 2009) fusion protein. Bold ATG represents a start codon; ‡ symbol represents the beginning (or the end) of the GOI CDS; bold TTA represents a reverse complemen stop codon.

<table>
<thead>
<tr>
<th>Vector</th>
<th>Fusion Tag</th>
<th>Parent Vector/ Antibiotic resistance</th>
<th>Restriction Enzymes</th>
<th>Forward Primer Extension</th>
<th>Reverse Primer Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>pOPINA</td>
<td>...KHHHHHH</td>
<td>pET28a/Kanamycin</td>
<td>Ncol/Dral</td>
<td>AGGAGATATACCA TG‡</td>
<td>GTGGTGTTGGGTG‡</td>
</tr>
<tr>
<td>pOPINB</td>
<td>MGSSHHHHHHSSGLEVLFQ↓GP ...</td>
<td>pET28a/Kanamycin</td>
<td>KpnI/HindIII</td>
<td>AAGTTCTGTTCAG GGCCCG‡</td>
<td>ATGGTCTAGAAAG CTTTA‡</td>
</tr>
<tr>
<td>pOPINE</td>
<td>...KHHHHHH</td>
<td>pTriEx2/Ampicillin</td>
<td>Ncol/Pmel</td>
<td>AGGAGATATACCA TG‡</td>
<td>GTGATGGTGATGT TT‡</td>
</tr>
<tr>
<td>pOPINF</td>
<td>MAHHHHHHSSGLEVLFQ↓GP...</td>
<td>pTriEx2/Ampicillin</td>
<td>KpnI/HindIII</td>
<td>AAGTTCTGTTCAG GGCCCG‡</td>
<td>ATGGTCTAGAAAG CTTTA‡</td>
</tr>
<tr>
<td>pOPINS3C</td>
<td>MAHHHHHH SUMO- SSGLEVLFQ↓GP...</td>
<td>pTriEx2/Ampicillin</td>
<td>KpnI/HindIII</td>
<td>AAGTTCTGTTCAG GGCCCG‡</td>
<td>ATGGTCTAGAAAG CTTTA‡</td>
</tr>
<tr>
<td>pOPINRX</td>
<td>MAHHHHHH thioredoxin- SSGLEVLFQ↓GP...</td>
<td>pTriEx2/Ampicillin</td>
<td>KpnI/HindIII</td>
<td>AAGTTCTGTTCAG GGCCCG‡</td>
<td>ATGGTCTAGAAAG CTTTA‡</td>
</tr>
<tr>
<td>pOPINHALO7</td>
<td>MAHHHHHH HaloTag- SSGLEVLFQ↓GP...</td>
<td>pTriEx2/Ampicillin</td>
<td>KpnI/HindIII</td>
<td>AAGTTCTGTTCAG GGCCCG‡</td>
<td>ATGGTCTAGAAAG CTTTA‡</td>
</tr>
</tbody>
</table>

For plasmid preparation 5-10 ng of closed circular pOPIN vector were used for each transformation reaction in TOP10 E. coli cells, following the same steps listed in paragraph M9, using pre-warmed plates and liquid LB cultures with selective antibiotic (Ampicillin or Kanamycin, depending on the vector). Once pOPIN plasmid DNA was recovered from bacterial samples via GenElute™ Plasmid Miniprep Kit approximately 2 µg of each vector were digested with a pair of specific restriction enzymes, selected from the plasmid restriction map available on Addgene, in order to produce vector ends complementary with the PCR fragment extensions. Restriction enzymes were used following the manufacturer’s instructions (NEB) as detailed in par. M10 and are listed in the Table M5:

Table M5 List of used vectors, restriction enzymes used for their linearization (buffers and cofactors used) and size of the excised fragment.

<table>
<thead>
<tr>
<th>Vector</th>
<th>Restriction Enzymes</th>
<th>Buffer</th>
<th>Addition of BSA</th>
<th>Size of excised fragment (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pOPINA</td>
<td>Ncol/Dral</td>
<td>CutSmart Buffer</td>
<td>NO</td>
<td>392</td>
</tr>
<tr>
<td>pOPINB</td>
<td>KpnI/HindIII</td>
<td>NEBuffer 2</td>
<td>YES</td>
<td>333</td>
</tr>
<tr>
<td>pOPINE</td>
<td>Ncol/Pmel</td>
<td>CutSmart Buffer</td>
<td>NO</td>
<td>387</td>
</tr>
<tr>
<td>pOPINF*</td>
<td>KpnI/HindIII</td>
<td>NEBuffer 2</td>
<td>YES</td>
<td>333</td>
</tr>
</tbody>
</table>

*pOPINS3C, pOPINRX and pOPINHALO are derived from pOPINF and need the same restriction enzymes
After the reaction a small aliquot of the samples (20-50 ng) was observed on agarose gel and compared with the same amount of undigested plasmid. The presence on the gel of a c.a. 300 bp fragment corresponding to the excised lacZα fragment of β-galactosidase was considered as proof of a successful double digestion (Figure M3).

![Figure M3](image.png)

*Figure M3* Agarose gel visualization of pOPIN preparative double digestion; DD: Double Digested. CTRL: Control. The used ladder is the 1 kb DNA Ladder (Sigma-Aldrich, USA). The white arrow indicates the short fragments excised from the double digestion.

After this test the remaining reaction was run on agarose gel and the band corresponding to the linearized vector was purified via the GenElute Gel Extraction Kit, to avoid contaminations from undigested circular vectors and vector re-closing with the lacZα fragment. Linearized vectors were quantified on agarose gel as previously described (usually 1/10 of the starting amount was recovered) and stored at -20°C for further applications.

For the usage of PJExpress411-Kan (pJ411) the procedure was a synthesis via PCR of the plasmid backbone, using the NEBuilder Assembly Tool (version 2.5.4; [https://nebuilder.neb.com/#!/](https://nebuilder.neb.com/#!/)) for the designing of primers for the plasmid backbone and the PCR fragment with extensions bases. Details of the pJ411 plasmids used in this Thesis and of their backbone primers are shown in Table M6.

**Table M6.** Details of the pJ411 vectors used in the present Thesis. Fusion Tags column lists the aminoacid sequence (1-letter code) fused to the POI. "..." at the beginning of the sequence indicates a C-terminus fusion tag; "..." at the end of the sequence indicates
an N-terminus fusion tag; **TEV** represents the TEV protease cleavage site (ENLYFQ/G). Bold **TGA** represents a stop codon, added to the POI (in case of the pJ411-CTer plasmid right after the His$_6$ tag).

<table>
<thead>
<tr>
<th>Vector</th>
<th>Fusion Tag</th>
<th>Forward Backbone Primer</th>
<th>Reverse Backbone Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>pJ411-NTer</td>
<td>MHHHHHHDYDIPTT-TEV-...</td>
<td>TGAAGCTTCCCCCTAGC</td>
<td>CCCCTGGAATACAAAATTTTCA G</td>
</tr>
<tr>
<td>pJ411-CTer</td>
<td>...EHHHHH</td>
<td>GAGCACCACCCACCCACCAC</td>
<td>ATGTATTACCTCCTTTAAAAGTTA AACAAAAATTATTTCTAGAGGG GAATTG</td>
</tr>
</tbody>
</table>

For plasmid preparation 10-20 ng of closed circular pJ411 vector were used as templates for PCR reactions, PrimeSTAR® GXL DNA Polymerase was used.

PCR products were treated with DpnI and purified with GenElute™ Plasmid Miniprep Kit as previously described. Plasmid backbones were finally quantified using NanoDrop™ 1000 Spectrophotometer and stored at -20°C for further applications.

**M12. Cloning in expression vectors**

Once purified, linearized pOPIN or PCR-synthetized pJ411 vectors and PCR products with extensions were obtained, the cloning reaction was performed using the Gibson Assembly® kit or the In-Fusion™ HD Cloning kit, following manufacturer instructions, except for the following:

- 100 ng of linearized vector were used in each cloning reaction (and a 2:1 insert to vector molarity ratio was used)
- 1/2 or 1/3 volume of cloning mixture was added to each 50 µL aliquot of Stellar Competent Cells (Takara, Japan)

As described in paragraph M8, colonies were screened via colony PCR for the expected insert (using a primer of the insert and a primer of the vector, in order to discriminate among incorrect inserts and eventual contaminant plasmids). 2-3 positive colonies were transferred to 15 mL Falcon tubes with 3-5 mL of LB broth with selective antibiotic (Ampicillin, Kanamycin) and incubated overnight at 37°C in constant agitation. Kanamycin-based plasmids were incubated longer (18-20h) to reach a good plasmid concentration. After overnight incubation, culture plasmid DNA was recovered, quantified and sequenced as reported in paragraph M8.
M13. Protein Heterologous Expression

Once vectors with the GOI were obtained and sequences proved the absence of any mutations in the DNA sequence and a correct reading frame for the POI, a heterologous expression screening was performed, using a modified version of the aforementioned OPPF-UK protocol.

The following steps were performed:

1- *E. coli* expression strains were each transformed with 5 ng of recombinant vector. The strains used were *E. coli* Lemo21 (DE3) (NEB), Rosetta 2 (DE3) (Novagen, USA) and BL21 (DE3) (Novagen) strains. Eventually Rosetta 2 and BL21 (DE3) strains were also used with their pLysS variant (which includes the pLysS plasmid that contains the T7 lysozyme gene (LysS)) and a resistance gene for the Chloramphenicol). The transformation protocols used were the same as those of *E. coli* TOP10 cells, described in paragraph M8. Additionally, in order for the cells to retain Lemo™ or pLysS endogenous plasmids, Chloramphenicol (final concentration of 35 µg/mL) was added to plate media, together with the antibiotic needed to retain the expression plasmid.

2- After overnight incubation, individual colonies were picked and transferred to 10 mL Falcon bottom-round tubes with 3 mL of LB broth with selective antibiotics (Ampicillin, Kanamycin, Chloramphenicol) as in paragraph M8.

3- Approximately 200 µL (150-250 µL, depending on the growth rate of the strain) of the overnight liquid cultures were added to new tubes with 3 mL of fresh LB medium supplemented with the appropriate antibiotic. Each 200 µL aliquot was added to six different new tubes, two for each different induction conditions plus two for control (i.e., the experimental setup was done in duplicate).

4- New tubes were grown at 37°C in agitation for 3-5 hours (usually reaching an average OD_{600} of 0.6-0.8).

5- Induction of the POI expression was performed by addition of IPTG to a final concentration of 0.5 mM or 1mM per tube. Tubes were then grown for 16h in agitation at 20°C (long-term induction) or alternatively for 1:30h at 37°C (short-term induction), 2 mL of each culture volume were split into two 1,5 mL Eppendorf and cells were harvested via centrifugation at 10.000 rpm for 10"in a Heraeus Fresco21 Centrifuge (Thermo Fisher Scientific).

6- After the induction time had passed, the OD_{600} of each tube was measured and culture volumes were processed as in the previous point.

7- After centrifugation, the supernatant was carefully drained and samples were stored at -80°C until required.

In the case of larger growth volumes (400 mL, 1L) the same procedure was followed, except for these modifications:
- A single colony was inoculated in glass bottles or small 50 mL glass Erlenmeyer flasks, filled with 15 mL LB broth with antibiotic and grown overnight; then the entire pre-inoculum was added to the volume of growth medium the day after. 1% glucose can eventually be added to both overnight pre-inoculum and growth volumes in order to avoid eventual “leaky” protein expression.

- Growth volume was strictly monitored and IPTG was added after reaching OD\textsubscript{600} of 0.8–1.0.

- IPTG-added Erlenmeyer flasks were grown for 16h in agitation at 20°C (long-term induction) or alternatively for 2-3h at 37°C (short-term induction); this longer growth time at 37°C was finalised to obtain a higher yield of protein of interest (that, however, could be more prone to precipitation).

- Prior to IPTG-addition or post-induction time 1mL sample was taken from each flask for OD\textsubscript{600} reading, pelleted and later used for analysis.

- After induction time the growth volume was transferred in 1L polycarbonate centrifuge bottles and cells were harvested via centrifugation at 8.000 rpm for 15’at 4°C in an Avanti J-26S centrifuge (Beckman Coulter, USA). At the end cellular pellets were manually transferred to 50 mL falcon tubes.

**M14. Protein Extraction and Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis (SDS-PAGE)**

To visualize the protein patterns from the different 1mL samples with SDS-PAGE, proteins were extracted and denatured, following a slightly modified version of the aforementioned OPPF-UK protocol:

1. Previously stored cell pellets were completely defrosted and suspended in 100 µL of NPI-10- Tween buffer, supplemented with 1 mg/mL of Lysozyme (Sigma Aldrich) and 1.5U/mL of Benzonase (Merck) and left in agitation for 30’.
   NPI-10-Tween buffer was composed as follows: 50 mM NaH\textsubscript{2}PO\textsubscript{4}, 300 mM NaCl, 10 mM imidazole, 1% v/v Tween 20, adjusted to pH to 8.0 and filtered before use.

2. After waiting 30’for the action of the enzymes, samples were centrifuged at 10.000 rpm for 15’in a Heraeus Fresco21 Centrifuge.

3. The supernatant was then transferred to a new tube without disturbing the ‘insoluble’ pellet, which was further suspended with an additional 100 µL of NPI-10- Tween buffer.

4. Samples were prepared in 24 µL aliquots containing 18 µL of protein extract (soluble or insoluble fraction) and 6 µL of 4x dye mix. Dye mix was prepared using a 10:1 ratio of Laemmli dye (BioRad) to β-Mercaptoethanol (BME); BME was used to reduce protein disulphide bonds, further breaking native protein conformation and avoiding the formation of aggregates. After dye addition the samples were boiled at 95°C for 2’before loading.

5. 5-15 µL of samples were loaded in a 4–15% Criterion™ TGX Stain-Free™ Protein Gel (BioRad). Different samples had a different cellular amount (and, accordingly, protein amount) and were
usually normalized with the OD_{600} cell concentration or via Bradford (Sigma Aldrich or Thermo Scientific) or BCA (Thermo Scientific) protein assays, following the manufacturer’s instructions and loading different volumes accordingly.

As a rule of thumb, 5 µg of protein (c.a. 5 µL of a 100 µL suspension of a 1,5 mL 0.7 OD_{600} sample) is a fair amount for Western-Blot analysis, while 10-12 µg of protein (c.a. 10-15 µL of a 100 µL suspension of a 1,5 mL 0.7 OD_{600} sample) is a fair amount for Coomassie staining. SDS-PAGE run was performed using a Criterion™ Cell (BioRad) and a Power Pac 3000 Universal Power Supply (BioRad, USA). Running buffer was TGS 1X, from a TGS 10X stock (BioRad). Usually the first and the last lane of the protein gel were occupied by Precision Plus Protein™ WesternC™ or Precision Plus Protein™ All Blue Standards (BioRad). SDS-PAGE run lasted approximately 1:30h at 120 mV, soon after the purple Laemmli dye disappeared from the lanes.

6- After the run, polyacrylamide gels were removed from their plastic scaffold and stained with QC Colloidal Comassie Stain (BioRad) following the manufacturer instructions.

7- Finally, polyacrylamide gels were visualized via ChemiDoc™ MP Imaging System.

Protein gel visualization eventually showed differences among protein patterns of the control samples (not induced) and samples where the expression of the protein of interest was induced. These samples were further tested via Western-Blot analysis before proceeding with protein purification.

**M15. Western-Blot Analysis**

Western-Blot analysis was done after a regular SDS-PAGE, which was performed as described in the previous paragraph except for the following:

- In order to avoid excess of signal from the chemo luminescence reaction of the secondary antibody a lower amount of samples (c.a. 5 µg of soluble fraction of the protein extract or 0.5 µg (1:10 dilution) of insoluble fraction of the protein extract) was loaded in each lane.

- After the run, polyacrylamide gels were removed from their scaffold and directly visualized via ChemiDoc™ MP Imaging System, using the Bio-Rad proprietary trihalo contained in Stain-Free™ Gels, in order to check the presence and position of the POI.

Western-Blot transfer was performed using Trans-Blot® Turbo™ Transfer System (BioRad) and Trans-Blot Turbo 0.2 µm PVDF Transfer Packs (BioRad) using the TURBO protocol, following the manufacturer instructions. After transfer (that was verified observing on ChemiDoc™ the polyacrylamide gel post-transfer) the following steps were followed:
1. Blocking step of 1h was performed, soaking the filter with TBS-T (Tris Buffered Saline solution + Tween; TBS 1X + 0,1% v/v Tween-20) + 5% milk powder (PanReac AppliChem, USA) at 4°C in agitation. TBS 10X buffer stock recipe was: 200 mM Tris/Trizma; 1.5 M NaCl; pH 7.5.
2. Incubation step was performed overnight at 4°C in agitation, using primary anti-His antibody (Bethyl Laboratories, USA) diluted 1: 15,000 in TBS buffer + 5% milk powder.
3. After overnight incubation the filter was briefly left at room temperature (22-25°C) before proceeding with wash steps. Buffer with added primary antibody was recovered and stored at -20°C for further uses.
4. 2-3 wash steps were performed, soaking the filter with TBS-T buffer at room temperature in agitation for 30’.
5. Secondary incubation step of 1h was performed at room temperature in agitation, using secondary antibody Anti-Rabbit IgG (whole molecule)–Peroxidase produced in goat (Sigma Aldrich, USA) diluted 1:5000-1: 10,000 in TBS buffer + 5% milk powder.
6. 2-3 additional wash steps were performed, soaking the filter with TBS-T buffer at room temperature in agitation for 10’.
7. Finally, the filter was dried with paper and moistened with 1-2 mL of Clarity MaxTM Western ECL Substrate (BioRad) for chemiluminescent protein resolution.
8. After 3-5’ of reaction (in the dark) the filter was observed using ChemiDoc™ MP Imaging System.

M16. Bioinformatic analyses and databases

Genes identified as promising for further study and potential biotechnological applications were found in analysed transcriptomes using both the automatic annotation via Blast2GO and using the TBLASTN algorithm, included in the BioEdit sequence alignment editor (http://www.mbio.ncsu.edu/BioEdit/page2.html) using default settings and homologous genes as queries and different transcriptomes as subject databases.

Main used databases were: the diatom C closterium transcriptome strain FE2, first published by Elagoz et al. (Elagoz, Ambrosino and Lauritano, 2020) and available in the NCBI BioProject database, under the series entry PRJNA577416; the green alga Tetraselmis suecica strain FE205, first published by Lauritano et al. (Lauritano et al., 2019) and available in NCBI GenBank (GEO database) under series entry GSE109461; the dinoflagellates Alexandrium tamutum strain FE107, first published by Vingiani et al. (Vingiani et al., 2020) and available in the NCBI Sequence Read Archive (SRA) database under the series entry PRJNA632001 and Amphidinum carterae strain FE102, first published by Lauritano et al. (Lauritano et al., 2017) and available in NCBI GenBank (GEO database) under the series entry GSE94355.

ExPASy Translate tool (https://web.expasy.org/translate/) was used to identify ORFs in observed transcriptomes’ contigs and translate them. Nucleotide or aminoacid sequence alignment were performed
via Clustal-Omega (Madeira et al., 2019) while ExPASy BoxShading was used for alignment visualization (https://embnet.vital-it.ch/software/BOX_form.html) or via Jalview (Procter et al., 2021).

Preliminary protein modelling and docking were performed with the following programs: SWISS-MODEL (Waterhouse et al., 2018), Phyre2 (Kelley et al., 2015) and Yang Lab’s COACH-D (Yang, Roy and Zhang, 2013). Sequenced DNA traces were visualized and assembled via the programs EditSeq and Seqman (currently discontinued and merged in the DNAStar Lasergene suite https://www.dnastar.com/software/lasergene/).

Codon usage for GOIs was analysed using the JCat online tool (Grote et al., 2005).

ApE v3.0.0 (https://jorgensen.biology.utah.edu/wayned/ape/) program was used to identify ORFs and restriction sites on assembled sequences. SnapGene program (https://www.snapgene.com/) was used to visualize plasmid maps and simulate Cloning and PCR reactions.

**M17. Chemicals and products**

For this PhD thesis, chemicals and products not before mentioned were of known commercial companies (GE Healthcare, Invitrogen, Merck, Panreac, Promega, Pronadisa, Qiagen, Serva, Sigma Aldrich and Thermo Scientific), being chemically pure and of the highest quality possible.
Appendix M References:


