Deciphering the "OPR code" to further assess the physiological role of OPR proteins

Domitille Jarrige

THÈSE DE DOCTORAT DE SORBONNE UNIVERSITÉ

Préparée au Laboratoire de **Biologie du Chloroplaste et Perception de la Lumière chez les Microalgues**. UMR7141 CNRS/Sorbonne Université
INSTITUT DE BIOLOGIE PHYSICO-CHIMIQUE

École doctorale Complexité du vivant (ED515)

Présentée par DOMITILLE JARRIGE
Pour l'obtention du grade de docteur en Biologie

DÉCHIFFRER LE « CODE OPR »
POUR UNE MEILLEURE COMPRÉHENSION DU RÔLE PHYSIOLOGIQUE DES PROTÉINES OPR.

DECIPHERING THE "OPR CODE"
TO FURTHER ASSESS THE PHYSIOLOGICAL ROLE OF OPR PROTEINS.

Dirigée par YVES CHOQUET

Soutenue le 12 décembre 2019
Devant le Jury :

DR CLAIRE LURIN	Rapportrice
DR HAKIM MIREAU	Rapporteur
PR FELIX WILLMUND	Examinateur
PR CHRISTIAN SCHMITZ-LINNEWEBER	Examinateur
DR YVES CHOQUET	Directeur de thèse
PR FRÉDÉRIC DEVAUX	Examinateur

REMERCIEMENTS

Arrivée à l'IBPC en 2016 à la recherche d'un stage pour mon master, j'aurais difficilement pu tomber sur un meilleur laboratoire ! Cette thèse a été pour moi l'occasion d'une importante évolution et il est compliqué pour moi de trouver les mots pour remercier comme il se doit toutes les personnes qui ont jalonné mon parcours.

Je souhaite bien évidement remercier Yves avec qui pendant presque quatre ans (Eh oui le temps passe vite !) j'ai travaillé, réfléchi et évolué. Merci beaucoup pour le temps et l'énergie que tu m'as consacrés et les longues discussions scientifiques que nous avons eues. Merci beaucoup pour ton aide en cette dernière année de thèse et la relecture de ce manuscrit.

Un grand merci aussi à Felix Willmund, Christian Schmitz-Linneweber et Frédéric Devaux d'avoir accepté de faire partie de mon jury et particulièrement à Claire Lurin et Hakim Mireau d'avoir accepté en être rapporteurs.

Je souhaite aussi naturellement remercier les membres de mon comité de thèse Laure Teysset et Wojciech Majeran pour m'avoir donné de leur temps et pour m'avoir suivie et conseillée avec gentillesse lors de cette thèse.

Un grand merci également à Angela et Francis-André et à tous les membres du laboratoire, passés et présents, pour l'ambiance de travail (ou de détente !) excellente au cours de ces années, la qualité scientifique, l'esprit d'ouverture et de bienveillance, mais également les énigmes de la pause-café, les indénombrables galettes des rois et les parties de ping-pong endiablées !

Il serait bien sûr bien trop long de mentionner chacun pour tout ce qu'il m'a apporté, mais je souhaite remercier Katia, Sandrine, Catherine, Stefania, Richard et Oliver pour les nombreux conseils éclairés : pratiques, scientifiques ou même personnels, et les bons « tuyaux » expérimentaux.

Merci également mille fois à Sophie pour sa bonne humeur communicative et pour la préparation des innombrables milieux et aussi à Danielle pour la gestion attentive des commandes et missions et sa gentillesse.

Un grand merci aussi pour les riches discussions (notamment lors des fameuses réunion-marathon) avec Francis-André, Olivier, Ingrid, Stephan, Pierre, Marc et tous les autres, qui m'ont permis de mieux comprendre les rouages du chloroplaste et un peu mieux la photosynthèse.

Merci à Frédéric avec qui j'ai eu le plaisir de travailler en collaboration, et qui a été un formidable comparse de congrès ! Merci à Julien qui m'a aidé avec patience pour l'étude du système fluorescent. Merci à Julia qui a travaillé avec le sourire plusieurs semaines à l'établissement de ce même système.

Un immense merci également à Alexandra, Suzanne, Clotilde, Frédéric, Benjamin, Marcio, Carole, Matthieu, Alix, Léna, Alessandro, Céline, Jean-Pierre, Marianne et bien d'autres d'avoir prêté une oreille conciliante à mes coups de grogne ou de frustration et de m'avoir remotivée !

Merci aussi à mon ancien voisin de bureau Wojciech pour les vidéos de cuisine des bois qui m'ont mis l'eau à la bouche, à Alexandre le roi de la contrepèterie de nous avoir tant fait rire… Au ficus pour m'avoir abritée de son ombre…

Merci de tout cœur à toute ma famille qui m'a soutenue et encouragée pendant les coups durs. Votre présence a été essentielle…Bref…

MERCI à tous !

Enfin merci à la petite Chlamy d'avoir subi sans broncher toutes ces expériences pendant ces années, et sans qui rien de tout ce travail n'aurait été possible !

SUMMARY

REMERCIEMENTS 1

ILLUSTRATIONS TABLE 6

NOMENCLATURE 14

ABBREVIATIONS 14

INTRODUCTION 15

1. ENDOSYMBIOSIS AND THE CHLOROPLAST EMERGENCE 20
Dawn of the endosymbiotic theory 20
The ancestral cyanobacterium 21
Current views on endosymbiosis 22
A brief overview of the diverse photosynthetic eukaryotes: 23
Secondary photosynthetic eukaryotes 25

2. CHLAMYDOMONAS REINHARDTII 26
A flexible model organism 26
Structural properties 28
Photosynthesis in the thylakoids 29

3. THE CHLOROPLAST GENOME: EXPRESSION AND REGULATIONS 30
The chloroplast genome structure reflects its history 30
Expression of the chloroplast genome 33
Nucleus and chloroplast, interactions and regulations 36

4. THE OPR: A SUBFAMILY OF OTAF 45

5. THE ATP SYNTHASE 50

6. MAIN OBJECTIVES OF THIS THESIS: 52

I. FUNCTIONAL STUDIES OF THE MDB1/*ATPB* EXPRESSION SYSTEM 55

INTRODUCTION 56

RESULTS 60
atpB mRNA stability and maturation 60
atpB 5'UTR and translation 64
Identification of a new mutant of MDB1 67

DISCUSSION 70
MDB1 a true maturating factor controlling *atpB expression* 70
Could atpB maturation be influenced by its 3'UTR? 71
atpB mRNA endures rapid degradation 71
Inactivation of *MDB1* by TOC1 insertion 71

II. FUNCTIONAL STUDIES OF THE MTHI1 FACTOR 73

INTRODUCTION 74

RESULTS 78
Swapping *atpH* and *atpI* MTHI1 target 78
Is MTHI1 the only limiting factor for *atpI* and *atpH* expression? 80
MTH2, a putative factor implicated in *atpH* stabilisation 81
Other factors implicated in *atpH* accumulation? 83
MTHI1 is implicated in *atpI* mRNA stability 84
MTHI1 is implicated in *atpH* translation 86

DISCUSSION 89
MTHI1, a keystone to both atpI and atpH expression 89
MTHI1 might work with other factors 90

Toward a general model of OPR proteins action? 91

III. INITIAL STUDIES OF MDB1 AND MTHI1 INTERACTIONS WITH THEIR TARGET MRNA *IN VIVO* 93
INTRODUCTION 94
The "OPR code" 94
HOW TO "CRACK" THE CODE 97
RESULTS 98
Mutagenesis of *atpB* MDB1 target sequence 98
MUTAGENESIS OF THE MTHI TARGET SEQUENCE *AT THE 5' END OF THE ATPH MRNA.* 101
DISCUSSION 105
A PARADIGM SHIFT ON OPR M FACTOR SPECIFICITY 105
THE SEARCH FOR MTHI1 PARTNERS CONTINUES? 105
INSIGHTS ON MDB1 FROM ANOTHER STUDY 105

IV. BUILDING A CHIMERIC SYSTEM TO VALIDATE THE "OPR CODE" 107
INTRODUCTION 108
Implications of switching to a chimeric reporter 108
Design of an additional reporter with the Spinach2 aptamer 109
RESULTS 111
Comparing the two reporter systems 111
Studying MDB1 binding with the chimeric system 116
Combining competition and sequence variation 118
Validation of the OPR code 119
DISCUSSION 129
Chimeric systems and competition with *atpB* mRNA 129
atpB 3'UTR might mediates a diminution of transcript accumulation 131
Chimeric reporter constructs mimic an "*in vitro*" system 133
Validation of the OPR recognition code remains elusive 133

DISCUSSION AND CONCLUSION 137
MDB1 and MTHI1 are key factors for the plastid ATP synthase expression. 138
The "OPR code" and *in vivo* revelations: 139
OPR proteins, advocates of teamwork? 141
OPR: an opportunity for evolutionary prototypes 142

MATERIALS AND METHODS 143
MATERIALS 144
Strains 144
Plasmids 144
PCR primers 144
METHODS 144
culture conditions 144
Crosses 144
Phototrophy test 144
Fluorescence live-imaging 145
Cloning 145
Chloroplast transformation 146
RNA extraction and RNA-blot 146
Protein extraction and Immuno-blot 146
Genomic analysis 147

REFERENCES 148

ANNEXES **169**
1) PLASMIDS **170**
2) PRIMERS **174**
3) ALIGNMENT OF MTH2 ..... **176**
4) OVERLAPPING HPR AND OPR PROTEINS IN C. REINHARDTII **177**
5) SYNTHETIC MDB1 FRAGMENTS **178**
ARTICLE 1 ..... **179**
ARTICLE 2: **181**
ARTICLE 3: **183**
ARTICLE 4: **185**

RÉSUMÉ **187**

ABSTRACT **192**

ILLUSTRATIONS TABLE

Figure 1: Cartoon of membrane cross-sections showing proton relocation and electron transport and ATP generation. A. Type II reaction centre B. Type I reaction centre C.ATP synthase..... 18
Figure 2: A. Diagram suggesting a rough estimate of global inorganic carbon fixation, or primary production, throughout Earth history. B. Relative atmospheric partial pressure of O_2 relative to present atmospheric level, data taken from Lyons et al, 2014, *Nature*..... 19
Figure 3: Simplified cladogram of cyanobacterial diversity from (de Vries and Archibald, 2017). Two possible origins for plastids are indicated..... 21
Figure 4: Primary chloroplast endosymbiosis. A possible history of chloroplast emergence is suggested..... 22
Figure 5: Distribution of photosynthetic organisms in the tree of eukaryotes. Full lines are primary endosymbiotic organisms, dashed lined secondary or tertiary ones. Grey organisms are not photosynthetic. Modified from (Ponce-Toledo *et al.*, 2019)..... 23
Figure 6: Chlamydomonas reinhardtii, with its three genomes. All of them can be transformed. 26
Figure 7: Genome segregation in sexual reproduction of *C. reinhardtii.* 27
Figure 8: Transmission electron micrograph of a *C. reinhardtii* (y-1 mutant) cell. Originally published (Ohad *et al.*, 1967) made available by I. Ohad (2012) CIL:37252, Chlamydomonas reinhardtii. CIL. Dataset. https://doi.org/doi:10.7295/W9CIL37252..... 28
Figure 9: Schematic cross-section of the photosynthetic chain in the thylakoid membrane of *C. reinhardtii.* 29
Figure 10: Cartoon of the photosynthetic complexes in the thylakoid membrane, genome origins of the subunits across Viridiplantae are indicated in colours. 32
Figure 11: Gene shuffling: Left: Genes organised in operons (bacteria, plant chloroplasts to some degree) right: the polycistronic expression of genes (*Chlamydomonas*). Colours schematise the function of the gene products..... 32
Figure 12: Bacterial model of gene expression versus the *C. reinhardtii* plastid one. In bacteria transcription and translation occur at the same time for each protein of the operon. In chloroplasts the transcripts are processed from the polycistron are longer lived and are translated independently..... 35
Figure 13: Every photosynthetic complex in *C. reinhardtii* displays some form of CES mechanism. Black arrows indicate the hierarchy of subunit assembly, CES subunits are at the receiving end of dominant subunits..... 38
Figure 14: The many roles of OTAF proteins in organelles RNA metabolism..... 40
Figure 15: Crystal structure of a designer PPR bound to its cognate RNA, from (Shen *et al.*, 2016), the PPR wraps around the RNA. Red areas are positively charged, the blue ones negatively charged..... 42
Figure 16: Crystal structure of PPR10 bound to one of its *psaJ* target mRNA, from (Yin *et al.*, 2013). A. The bases of the *psaJ* mRNA are inserted between bulky residues of the internal helices of the PPR repeats. B. Interactions of the 5th residue (in cyan) of PPR10 repeats with the bases of *psaJ*. The dotted red lines represent hydrogen bonds..... 42
Figure 17: Consensus amino acid sequence of canonical P PPR motif, and its variants: L (long) and S (short). Cyan arrows indicate the key residues for establishing the recognition specificity of the repeat. The grey bars indicate

where the two α-helices of the canonical PPR motif lie. Modified from (Barkan and Small, 2014) 43
Figure 18: The consensus sequence of the OPR repeats found in photosynthetic organisms. The taller the residue, the most abundant it is. The positions of the two putative α-helices are indicated under the consensus. The red arrow indicates the 6th amino acid, which is expected to be crucial for the specific interaction with nucleotides of the target mRNA. 45
Figure 19: LOGOs of OPR repeats from different functional factors in *Chlamydomonas reinhardtii* show a divergence in sequence. The LOGOs were obtained with MEME suite (Bailey *et al.*, 2009), note that MEME finds motifs of 37 amino acids, with an offset of several residues. The M factor LOGO was adjusted to align with the T and splicing factors LOGOs. 46
Figure 20: LOGOs of OPR repeats in Chlorophyceae and diatoms differs. The LOGOs were obtained with MEME suite (Bailey *et al.*, 2009), note that MEME finds motifs of 37 amino acids, with an offset of several residues. 47
Figure 21: Cryo-electron microscopy structure of ASA2 from (Murphy *et al.*, 2019). The 6th residue of each OPR motif is indicated in red. Note the helical shape of the protein. 48
Figure 22: There could be an inverse correlation between the number of NCL proteins and the transition to multicellularity 49
Figure 23: Structures of C-rings in the rotor of varuous ATP synthases, from (Walker, 2013) 50
Figure 24: C. reinhardtii ATP synthase 51
Figure 25: The β subunits of the chloroplast ATP synthase. 56
Figure 26: The *atpB* gene. MDB1 footprint is indicated by the red arrow. 56
Figure 27: The *L35a* (*mdb1-2*) and *thm24* (*mdb1-1*) mutants 57
Figure 28: A. *MDB1* gene on chromosome 14. B. Right: a prediction of MDB1 structure by the program I-Tasser, Left: A cartoon of MDB1 OPR domains 58
Figure 29: cRT-PCR procedure, mock and RPP samples were treated and analysed in parallel. 61
Figure 30: 2 of several attempts of cRT-PCR amplification of *atpB* with the very distant B-FW and B-RV primers 61
Figure 31: Sequence data of the RPP and Mock samples of a PCR amplification by B-RV and cRT-atpB FW1, the corresponding sequence of the expected precursor 5' and complete 3' junction of a circularised *atpB* mRNA is indicated below. 62
Figure 32: Amplification of cRT-PCR reactions with progressively more distant primers 63
Figure 33: Predicted secondary structure forming between *atpB* 5'UTR and *rbcL* 3'UTR, obtained with the RNABows program (Markham and Zuker, 2008) 64
Figure 34: Immunoblot, anti-cyt. *f*, OEE2 and β (CF1) primary antibodies were used. OEE2 serve as loading control. 65
Figure 35: Immunoblots, two technical repeats, anti-AadA, OEE2 and β (CF1) primary antibodies were used. OEE2 serve as loading control. * indicate cross reactions of the anti-AadA antibody, → a possible specific AadA signal. 66
Figure 36: A. Growth phenotype of *K4.20+*. B. Kinetic of PSII fluorescence saturation in K4.20+ versus WT.T222+. Obtained with a SpeedZen camera. C. Immunoblot of K4.20 and other mutants of ATP synthase hybridised with anti AtpB, AtpH and PsaD antibodies. D. RNA blots of *K4.20*, filters were hybridised with *atpB* and *atpH* dig-dUTP labelled probes. 67

Figure 37: A. Insertion site of TOC1 in *MDB1* in the *K4.20* mutant, Illumina paired end whole genome sequencing visualised with IGV. Reads pairing with mates on other chromosomes are indicated in pastel colours. Analysis of the mates' sequences with BLAST revealed TOC1 sequences B. Reconstructed map of *MDB1* in the K4.20 mutant. 68
Figure 38: PCR amplification of the junction between TOC1 and exon3 of *MDB1* in *K4.20*$^{+}$ x WT.S24^{-} non-photosynthetic progeny. Mating type PCR amplification was done in the same reaction as a DNA quality control. 69
Figure 39: *MDB1* expression along the day-night cycle, transcriptomic data redrawn from (Zones *et al.*, 2015). 70
Figure 40: ATP synthase operons in cyanobacteria and chloroplasts 74
Figure 41: A. The *atpH* gene, MTHI footprint is indicated by the turquoise arrow. B. Transcription of *atpH*, *atpF* is transcribed from the *atpH* promotor as it lacks a dedicated one, as suggested by small RNA coverage (Cavaiuolo *et al.*, 2017). 75
Figure 42: A. The *atpI* gene. B. Transcription of *atpI*, scissors indicate sites of endonucleolytic cleavage. 75
Figure 43: A. the MTHI1 gene. B. Left: Predicted structure of MTHI by I-Tasser, right: Cartoon of the OPR domain of MTHI1 76
Figure 44: MTHI1 and its two putative target sequences, one is at the very 5' extremity of the *atpH* monocistronic mRNA, the other within the *atpI* 5'UTR, about 60 nt upstream of the initiation codon. Note the difference of the 5th nucleotide between the two chloroplast genes 77
Figure 45: *C. reinhardtii* chloroplast ATP synthase, AtpI/AtpH stoichiometry is indicated 78
Figure 46: Schematic presentation of the two mutated chimeric constructs. The *psaA* promotor (white rectangle) was added in front of *atpI 5'UTR* 78
Figure 47: RNA blot of the chimeric constructs, filters were hybridised with ^{33}P *petA* and *atpB* (loading control) radioactive probes. 79
Figure 48: RNA blot of 5'*atpH* transcripts. Filters were hybridised with ^{33}P labelled *petA* and *atpH* probes. 79
Figure 49: Immunoblots of the chimeric transformants 80
Figure 50: RNA blot of *dlf* transcripts. Filters were hybridised with ^{33}P labelled *petA, atpI* and *atpA (loading control)* probes. 81
Figure 51: *MTH2* (Cre10.g461700) gene model, its 27 exons are indicated by pink arrows, 26 introns are present in this gene. 81
Figure 52: A. Growth phenotypes of WT.T222+ and *mth2-2* B. RNA blot of the *mth2-2* and *mthi1-2* mutants, filter was hybridised with *atpB* and *atpH* dig-dUTP labelled probes. Quantified ratio of *atpH* on *atpB* transcripts normalised on WT levels is indicated below. The quantification was performed with *ImageLab.* 82
Figure 53: Two fragments of RNA blots, presented in detail in chapter IV. Filters were hybridised with *atpB*, *gfp* and *atpH* dig-dUTP labelled probes. *atpH* was intended as a loading control but was surprisingly more accumulated in a *mdb1-2* background 83
Figure 54: RNA blot from Dominique Drapier, unpublished 83
Figure 55: Genes deleted in *mdb1-2*, genes in green are predicted to be targeted to the chloroplast by WoLF Psort. 84
Figure 56: Two hypotheses to explain the smaller levels of a chloroplast transcript when its T factor is absent. A. Ternary complex of M and T factors on the mRNA B. Translating ribosomes protects the mRNA from its degradation by endonucleases. 84

Figure 57: Mutation of the ATG start codon of *atpI,* replaced by the TAG stop codon. 85
Figure 58: A. RNA blot of the $atpI_{Ct}$ and $atpI_{St}$ mutants, filters were hybridised with dig-dUTP labelled *atpI* and *petB* probes. Three independent transformants were analysed. B. corresponding growth tests (two droplets were made for each transformant). 85
Figure 59: Insertion of a polyG track in *atpH* 5'UTR, directly before the MTHI1 target..... 86
Figure 60: Expression of *atpH* in the strains prior to *aadA* cassette excision. A. RNA blot, filter was hybridised with ^{33}P *atpH* and *atpA* (loading control) radioactive probes. B. Immunoblots of whole cell extracts, anti AtpH and OEE2 (loading control) primary antibodies were used. 86
Figure 61: *atpH* expression after *aadA* cassette excision A. cartoon of the *aAdI* construct B. Growth test C. RNA blot, filters were hybridised with dig-dUTP labelled *atpI*, *petB*, *atpH* and *psbD* DNA probes. D. Immunoblot, anti-tubulin and AtpH primary antibodies were used. Corresponding regions of the filter in Ponceau red stain are under the chemiluminescent signals..... 87
Figure 62: The lowest energy structure calculated at 25°C by RNA Folding Form (M-Fold: (Zuker, 2003)) of *atpH* 5'UTR and the first 25 nt of its CDS. The footprint of MDH1 is grey-shaded, while the *atpH* initiation codon is pink-shaded 89
Figure 63: A. Mode of action of maize PPR10, redrawn from (Prikryl *et al.*, 2011). B. Putative mode of action of MTHI1..... 89
Figure 64: Models of OTAFs interaction on mRNA. A. Ternary complex of M and T factors on mRNA B. Higher order complexes could be common in the organelles..... 91
Figure 65: The consensus sequence of the OPR repeats found in photosynthetic organisms. The taller the residue, the most abundant it is. Position of the two putative α-helices is indicated under the consensus. The red arrow indicates the 6th amino acid, which is expected to be crucial for the specific interaction with nucleotides of the target mRNA. 94
Figure 66: Our initial strategy to study the interaction of the OPR M factor MDB1 with its cognate *atpB* binding sequence *in vivo*. 97
Figure 67: RNA blot of the first *atpB* MDB1 target sequence variants, filter were hybridised with ^{33}P-labelled *atpB* and *atpA* radioactive probes. On top are depicted the mutated target sequences with the respective mutation black shaded. 98
Figure 68: A. Growth phenotypes of the mutants, a table aside shows the placement of the strains. Droplets of liquid culture of the strains were put on TAP and minimum media and grown for 12 days under 55 $\mu E.m^{-2}.s^{-1}$ illumination. B. RNA blot of *ΔatpB* strains transformed with the mutated *atpB* MDB1 target sequence. Corresponding mutations are depicted on top of each blot lane. Mutated nucleotides are depicted in black squares. An orange dot denotes the introduction of a steric clash in the mutant target. *atpB* and *petA* (loading control) mRNA quantifications were performed with the image lab software and normalised on dBCt levels. The ratio of *atpB* on *petA* transcripts is depicted under the blot. 100
Figure 69: A. Growth phenotypes of the mutants, a table on the left shows the placement of the strains. Droplets of liquid culture of the strains were put on TAP and minimum media under 55µE illumination for 8 days. B. RNA blot of *ΔatpH* strains transformed with the mutated *atpH* MDB1 target sequence.

Corresponding mutations are depicted on top of each blot lane. Mutated nucleotides are depicted in black squares. An orange dot denotes the introduction of a steric clash in the mutant target. *atpH* and *psbD* (loading control) mRNA quantifications were performed with the image lab software and normalised on *ΔdHCt 2* levels. The ratio of *atpH* on *psbD* transcripts is depicted under the blot. 102
Figure 70: A. Growth phenotypes of the mutants, a table on the left shows the placement of the strains. Droplets of liquid culture of the strains were put on TAP and minimum media under 55 $\mu E.m^{-2}.s^{-1}$ illumination. B. RNA blot of *ΔatpH* and *mth2-* strains transformed with the mutated *atpH* MTHI1 target sequences. Corresponding mutations are depicted on top of each blot lane. Mutated nucleotides are depicted in black squares. An orange dot denotes the introduction of a steric clash in the mutant target. *atpH* and *atpB* (loading control) mRNA quantifications were performed with the image lab software and normalised on *ΔdHCt 2* levels. The ratio of *atpH* on *atpB* transcripts is depicted under the blot. 104
Figure 71: Northern blot adapted from (Anthonisen *et al.*, 2001), *uidA* mRNA (GUS) compared to the control (C), on top are depicted the mutations inserted by Anthonisen and colleagues in the *atpB* MDB1 target sequence. Note the absence of loading control. 106
Figure 72: A. Comparison between some of our mutants and Anthonisen and colleagues' ones. B. Hypothesis to explain the differences observed: unknown factor(s) recognising a part of *atpB* transcript deleted in the *uidA* chimera can stabilise MDB1 on its target sequence. 106
Figure 73: The *gfp* chimeric construct at scale. 108
Figure 74: A. DFHBI B. Structure of Spinach aptamer and its interaction with DFHBI. From (Trachman *et al.*, 2017) C. Mechanism of DFHBI activation by Spinach, stuck in a planar conformation it cannot dissipate excitation energy by rotation and emits more fluorescence. 109
Figure 75: Mutations introduced by Strack and colleagues In Spinach2 improve thermostability and folding, Stem1 and Stem loop 3, in blue, were modified. From (Strack *et al.*, 2013). 110
Figure 76: The three chimeric reporters, drawn at scale. 110
Figure 77: RNA blot of the Spinach2 tagged chimeric transcripts, filter was hybridised with *atpB*, *gfp* and *atpH* dig-dUTP labelled DNA probes. 111
Figure 78: RNA blot of the "basic" *gfp* chimeras, either in WT or *ΔatpB* background. The filter was hybridised with *atpB*, *gfp* and *atpH* dig-dUTP labelled DNA probes. The ratios of *gfp* and *atpB* on *atpH* transcripts are depicted under the blot. 112
Figure 79: Two main hypotheses to explain how the mutated transcripts could be less destabilised in the presence of the whole atpB sequence. A. Other regions of the *atpB* transcript such its 3'end could be physically implicated in the stabilisation process. B. Auxiliary factors might limit *atpB* transcript destabilisation by maintaining MDB1 on its target sequence. 113
Figure 80: RNA blot of Spinach2 tagged chimeric constructs, with either *atpB* or *rbcL* 3'UTR, *gfp*, *atpB* and *atpH* (loading control) transcripts levels were probed. mRNA quantifications were performed with the image lab software and normalised on either {*ΔatpB dBWTgfp* Spix3-3'rbcL} 1 for *gfp* or WT levels for *atpB*. Reference levels are underlined. The ratios of *gfp* and *atpB* on *atpH* transcripts are depicted under the blot. 113

Figure 81: Chlorophyll auto-fluorescence and green fluorescence of cells treated or not with DFHBI, no difference in green fluorescence could be observed between the two treatments. The bright spots of green fluorescence correspond to the carotenoid-rich eyespots of the cells. 115
Figure 82: RNA blot of *ΔatpB* transformants chimeric constructs bearing *atpB* MDB1 target variants. Filter was hybridised with gfp and atpH dig-dUTP labelled probes. Transcript quantifications were done with ImageLab, and normalised on *ΔatpB*::dB_{WT}gfp 3 levels. Ratio of *gfp/atpH* transcripts is depicted under the corresponding lanes, the mutations are on top. Two technical repeats were made and give the same results. 116
Figure 83: Beginning of the amino acid sequence of the tested OPR motifs in MDB1, the probably crucial sixth residue of the repeat is highlighted in red. 117
Figure 84: Two RNA blots showing the difference of dB_M*gfp-3'rbcL* transcript in WT or *ΔatpB* genetic background. Filters were hybridised with *atpB*, *gfp* and *atpH* dig-dUTP labelled probes. Four technical repeats displayed the same patterns. 118
Figure 85: Modification inserted in MDB1, 6th and 7th or 11th and 12th OPR motif, they were chosen following the draft OPR code, for the CC1 and CC2 variant, one of the "unreadable" residue combination was picked at random. 119
Figure 86: The *MDB1mut* chassis construct, with the *aphVIII* paromomycin resistance gene for selection of transformants. 119
Figure 87: Photosynthetic activity of *MDB1mut* transformants, PSII fluorescence of cells on TAP media, φPSII, which reflect the proportion of open centres, was measured at the end of the continuous illumination phase. 120
Figure 88: Growth phenotypes of *mdb1-1::MDB1mut* transformants. Cells were grown for seven days under 55µE/m².s on minimum or TAP media. 121
Figure 89: Characterisation of three photoautotrophic independent transformants for the MDB1-HA, MDB1.CC_2 and MDB1.AA_2 variants by immunoblots. Primary antibodies against HA, β CF1 (AtpB), tubulin and cyt. *f* were used. Underlined transformants were selected for the subsequent crosses. 121
Figure 90: Characterisation of three photoautotrophic independent transformants for the MDB1-CC_1, MDB1.UU_1 and MDB1.GG_1 variants by immunoblots. Primary antibodies against HA, β *CF1* (AtpB), tubulin and cyt. *f* were used. Underlined transformants were selected for the subsequent crosses. 122
Figure 91: A. PCR amplification of MDB1 exon 5, B. a *Bsr*I restriction site is induced by the deletion of one A in *mdb1-1*. C. Digestion of subsequent amplicons with *Bsr*I. the *mdb1-1* allele A insertion creates a *Bsr*I site that does not exist otherwise in MDB1 5th exon. 124
Figure 92: Preliminary RNA blots showing the first descendant of each of the *MDB1mut* x dB_M*gfp* cross. A. Filter was hybridised with ^{33}P labelled *gfp* and *psaB* (loading control) probes. B. Filter was hybridised with ^{33}P labelled *atpB* and *aadA* probes. P are the MDB1mut parents. 125
Figure 93: Preliminary RNA blots of the MDB1.CC_2 MDB1.AA_2 and MDB1.GG_2 descendants. Filters were hybridised with ^{33}P labelled *gfp* and *psaB* (loading control) probes. P are the dB_M*gfp* parents, 1; 2 and 3 the descendants of a same cross. 128
Figure 94: RNA blot of a chimeric construct bearing the entire *atpB* 5' UTR, from Yves Choquet. The highest band hybridised with the 5'atpB probe is the endogenous *atpB* transcript, the lower one the chimeric one. Lincomycin prevents chloroplast translation. 129

Figure 95: A tentative model based on a specific 5'$_{sh}$*atpB*/3'*rbcL* UTR destabilising interaction, does not seems likely as both atpB and rbcL 3' UTR gfp chimeric transcript are destabilised. 130
Figure 96: A more probable model explaining the different competitiveness of the BKR and *dB$_{WT}$gfp.Spix3-3'rbcL* chimera. The short *atpB* 5'UTR lacks specific sequences that improve *atpB* MDB1-mediated stabilisation. 131
Figure 97: RNA blot of a chimeric constructs bearing the entire *atpB* 5' UTR, from Yves Choquet. The highest band hybridised with the 5'atpB probe is the endogenous *atpB* transcript, the lower one the chimeric one. Lincomycin prevents chloroplast translation. 131
Figure 98: Putative model explaining the variation in accumulation of *atpB* mRNA in presence of the *dB$_{WT}$gfp Spix3-3'atpB or dB$_{WT}$gfp Spix3-3'rbcL* transcripts. A specific endonuclease of *atpB* 3'UTR would be distributed between the endogenous 3'*atpB* and the chimeric one. 132
Figure 99: Comparison between some of our initial mutants (Chapter III) versus our mutated chimeric transcripts. The cartoon on top depicts a putative model of MDB1 interaction with the studied transcripts. Below are the relative accumulation levels of *atpB* MDB1 target variants compared to a control target..... 133
Figure 100 MDB1 and MTHI1 are part of the cross-talks between the chloroplast and the nucleus..... 138
Figure 101: Observations of the importance of parts of *atpB* target sequence for MDB1 binding..... 139
Figure 102: What we might find by immunoprecipitation of our tagged MDB1 and MTHI1..... 141
Figure 103: Modèle de l'interaction tripartite entre facteurs OTAF M et T sur leur ARNm cible, pourrait s'appliquer à MTHI1 et MDB1. 187
Figure 104: Gel d'ARN. Des mutations (surlignées en noir) réalisées dans la cible de MDB1 sur le 5' d'*atpB* ne conduisent qu'à une faible diminution de l'accumulation et donc de la stabilité du transcrit *atpB*. MDB1 parvient à s'y lier dans tous les cas, sauf pour la mutation dB12, très étendue. *petB* sert de contrôle de charge. 189
Figure 105: Gel d'ARN présentant l'accumulation du transcrit d'*atpH*. Les mutations (surlignées en noir) introduite dans la cible de MTHI1 sur le 5' d'*atpH,* n'empêchent pas la fixation de la protéine. *psbD* sert de contrôle de charge. 189
Figure 106: Gel d'ARN présentant l'accumulation du transcrit chimérique *gfp*. Les mutations (surlignées en noir) de la cible MDB1 entrainent cette fois une baisse plus importante du niveau de transcrit. 191

Table 1: The draft OPR recognition code, established by Yves Choquet. 48
Table 2: Maturation of chimeric *5'atpB* driven transcripts, data gathered by Blandine Rimbault and Marina Cavaiuolo. The nomenclature is as follow: (*B=atpB, A=atpA, F=petA, R=rbcL* and *K=aadA*) the first letter is the 5'UTR, the second the CDS, the third the 3'UTR..... 64
Table 3: The draft OPR code..... 95
Table 4: Prediction of the recognised nucleotide for each OPR motif of MTHI1 and MDB1 following the draft OPR code. In red is the 6th residue, deemed crucial, in orange the residues modulating the recognition. 96
Table 5: The draft OPR code: what is predicted and what we know so far. 117
Table 6: Expected descendants from a [mt+ {*ΔatpB::dB$_M$gfp* SpecR}] X [mt-*MDB1mut*-ParoR] cross. Grey shaded strains are sensitive to either spectinomycin or paromomycin and should not survive our double selection screen. 123
Table 7: The OPR code: where we stand. 134
Table 8: « Code OPR » préliminaire : le nucléotide reconnu est indiqué en bas; le résidu en position 6 de la répétition OPR, capital pour la spécificité de reconnaissance, en grands caractères. X dénote un acide aminé quelconque, – des exclusions. 188

NOMENCLATURE

In *Chlamydomonas reinhardtii*

- ***NUCLEAR GENES***
- **CYTOSOLIC PROTEINS**
- ***chloroplast genes***
- **Chloroplast proteins**

are written this way.

ABBREVIATIONS

ADP: adenosine diphosphate
AMP: antimicrobial peptide
ATP: adenosine triphosphate
CDS: coding sequence
CES: Control by Epistasis' of Synthesis
DFHBI: 3,5-difluoro-4-hydroxybenzylidene imidazolinone
DNA: deoxyribonucleic acid
cpDNA: chloroplast DNA
EGT: Endosymbiotic gene transfer
GFP: green fluorescent protein
LGT: Lateral gene transfer
LUCA: last universal common ancestor
NEP: Nucleus encoded polymerase
OPR: Octotricopeptide repeat
OTAF: Organellar Trans-Acting Factor
PBS: Phosphate-buffered saline (buffer)
PEP: Plastid encoded polymerase
PPR: Pentatricopeptide repeat
PS: photosystem
RC: reaction centre
RNA: ribonucleic acid
mRNA: messenger RNA
tRNA: transfer RNA
rRNA: ribosomal RNA
RuBisCO: ribulose biphosphate carboxylase/oxygenase
TIC: translocon at the inner envelope of chloroplast
TOC: translocon at the outer envelope of chloroplast
TP: transit peptide
UTR: untranslated region
WT: wild type

INTRODUCTION

Modified from Dartmouth Electron Microscope Facility, Dartmouth College

The ocean floors 4 billion years (Ga) ago, by an oceanic ridge… Seawater seeping deep through the fractured oceanic crust encounter the molten upper mantle rocks. This mineral enriched water is then ejected as a high temperature hydrothermal plume, a "black-smoker". Those vents have a transient lifetime (a few years or decades) and a hellish temperature unsuitable for life emergence.

However, seawater percolating through cracks in the oceanic crust further away from the ridge or in hotspot volcanoes or islands arcs, can react exothermically with peridotites and produces serpentine and dihydrogen (H2). This creates, as in the modern hydrothermal field of *Lost City* in the Atlantic Ocean, warm (around 80°C), alkaline (pH 9–10) and CO_2 depleted hydrothermal vents (Kelley *et al.*, 2001), called "white-smoker" (Corliss, 1990). Those systems are not only more hospitable but also longer lived, as less prone to eruptions. The *Lost City* field is more than 120 000 years old (Ludwig *et al.*, 2011). Crucially, abiotic formation of organic compound has been observed in the *Lost City* system (Proskurowski *et al.*, 2008). Similar hydrocarbon discharge might have occurred in the ancient effluents.

Mixing with the ambient cold and CO_2 rich seawater the white smokers form microporous calcium carbonate precipitates that present pH, thermic and redox gradients, from the vent aperture toward the sea. Those small compartments located at a source of organic compounds and H_2 have been suggested as one of the many possible starting points for life. They could help concentrate those molecules and provide a niche for primitive endolithic chemosynthetic organisms (Baross, 1985; Nisbet, 1995; Lane *et al.*, 2010; Sleep *et al.*, 2011; Sousa *et al.*, 2013; Martin *et al.*, 2018).

Thanks to the H_2 present in the plume, an abundant electron donor source, autotroph organisms could have performed CO_2 fixation to produce organic matter by the Wood-Ljungdahl pathway. This pathway is unique because it is exergonic and does not need ATP. (Fuchs, 2011; Sousa *et al.*, 2013; Sousa *et al.*, 2018). This theory is supported by phylogenetic data suggesting that the last universal common ancestor (LUCA) was a thermophilic autotroph relying on the Wood-Ljungdahl pathway (Weiss *et al.*, 2016).

To shift toward photosynthesis those ancient organisms obviously needed light. The famous “prebiotic soup” theory, which suggests the apparition of life in surface waters, under the action of UV irradiation or lighting, does provide a very luminous niche, maybe a too luminous one. The high sunlight exposure would pose a serious threat to early life. While high-light induced formation of chlorophyll triplets, does not produce much damage in absence of oxygen, as when photosynthesis began, the high UV radiations could severely damage the DNA and proteins of the organisms, rendering the evolution of early photosynthesis difficult without some kind of protective measures, or a protected niche.

Hydrothermal plumes can attain several hundred degrees Celsius and emit infrared radiation, but an intriguing very low visible light has been reported from black smokers (Van Dover *et al.*, 1996). It has been suggested long ago (Nisbet, 1995) that such dim light might have favoured infrared phototaxis, allowing the organisms to locate and move toward the smokers and indirectly induced a shift toward photosynthesis, without exposing early life to dangerous UV radiation that could photo-oxidise chlorophyll (Martin *et al.*, 2018). The isolation of an obligate phototrophic green sulphur bacterium from a deep ocean black smoker(Beatty *et al.*, 2005) gives credence to this hypothesis. Hydrothermal areas combining white and black smokers could have been compelling niches to evolve light sensing and transition to photosynthesis.

Another, possibility could lie in shallow hydrothermal vents on continental shelves (Arndt and Nisbet, 2012). But the probable very low levels of dissolved organic matter and plankton in the ancient water compared to today coastal areas, would not let only visible light but also UV radiation penetrate deep in the water column (Tedetti and Sempéré, 2006) exposing the cell to potential damage.

3.3 Ga ago much closer to the ocean surface, in the euphotic zone. Sun light penetrating the warm waters shines on a photosynthetic microbial mat laid on a shallow littoral (Tice, 2004; Westall *et al.*, 2006; Westall *et al.*, 2011). By harnessing this energy, the interwoven microbes became able to perform anoxygenic photosynthesis.

Current anoxygenic bacteria are scattered in highly diverse groups (Fischer *et al.*, 2016). Lateral gene transfer (LGT) probably played an essential role in granting photo-autotrophy to multiple phyla. For example, a whole photosystem I gene cluster has been found in marine bacteriophages (Sharon *et al.*, 2009). Photosynthetic reaction centres are complicated membrane embedded protein complexes. When their special pair of chlorophyll absorbs a photon, it generates a charge separation and an electron is transmitted to successive acceptors. RCI reduce ferredoxin as a terminal acceptor, while RCII reduce quinones. Those electrons then go to a complex III, which shuttles protons and generate a proton gradient across a membrane. This gradient is then used to produce ATP (Hohmann-Marriott and Blankenship, 2011) with a protein stemming from the last universal ancestor: the ATP synthase.

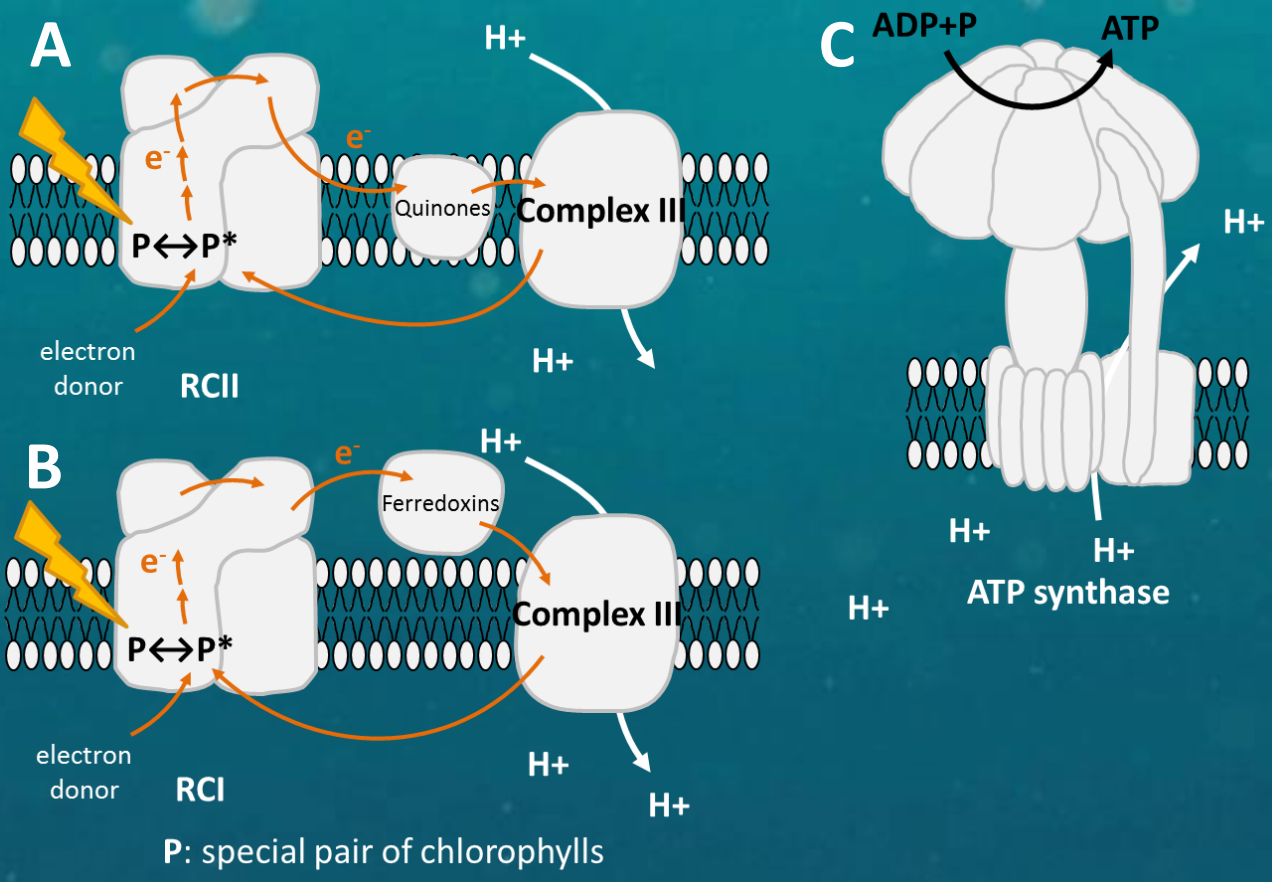


Figure 1: Cartoon of membrane cross-sections showing proton relocation and electron transport and ATP generation. A. Type II reaction centre **B.** Type I reaction centre **C.**ATP synthase

About 2.6 Ga ago an emerging class of Cyanobacteria the Oxyphotobacteria have assembled RCII and RCI in series. Among several competing hypotheses, it has been suggested that by LGT the ancestral cyanobacteria, which would have not been photosynthetic, acquired the two photosystems stemming from different lineages in a fusion process (Fischer *et al.*, 2016; Shih *et al.*, 2017; Soo and Fischer, 2017).

The Oxyphotobacteria are able to use light energy to split water, stripping electrons from it and to produce organic matter from CO_2. This innovation made them highly successful. However, this mechanism releases oxygen as a by-product. The Earth atmosphere, which was previously practically devoid of oxygen, soon saw its oxygen level shoot up (Figure 2). This event known as the *GOE* (*Great oxidising Event*), occurred 2.4 Ga ago. The oceans oxidised far slower (Lyons *et al.*, 2014; Sahoo *et al.*, 2012). This abundant oxygen induced a strong selective pressure on exposed organisms (its strong oxidative power and its reactive species block enzymes and damage proteins and DNA). But it also provided a new energy source as an electron acceptor. This paved the way for the advent of respiration.

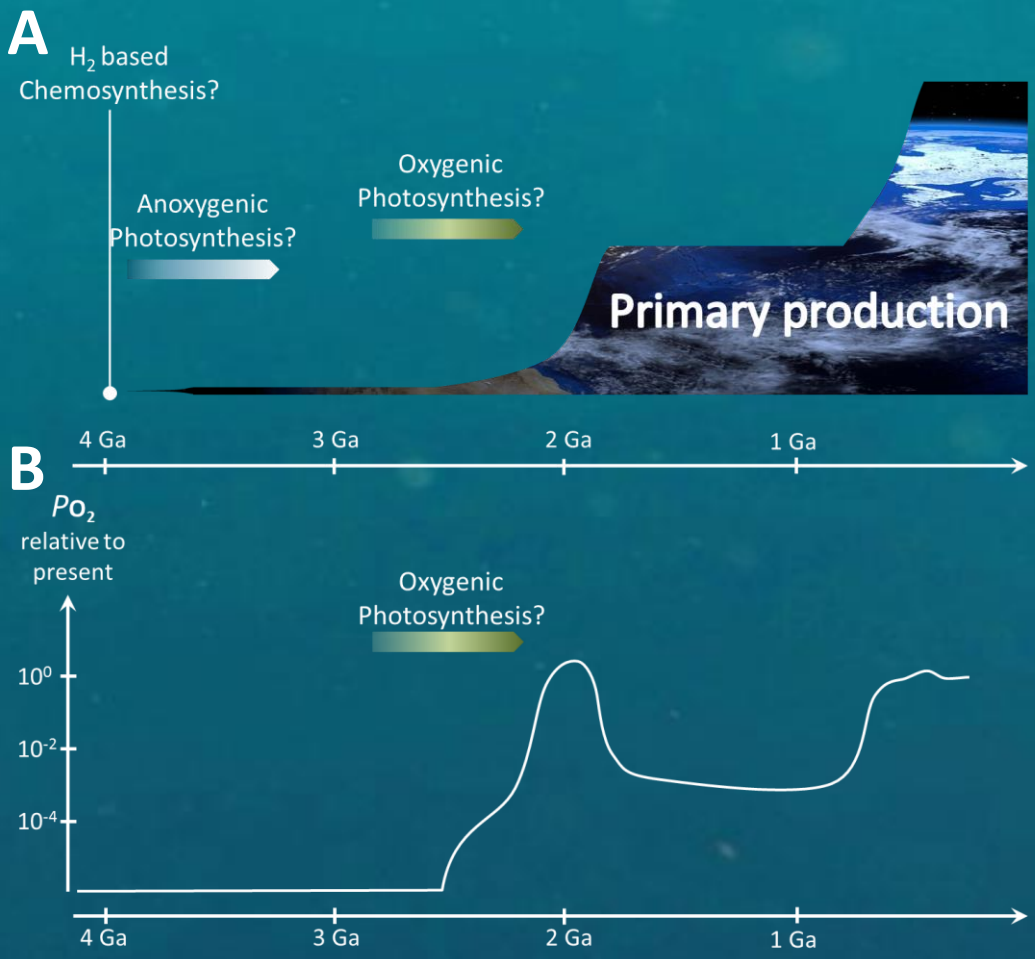


Figure 2: A. Diagram suggesting a rough estimate of global inorganic carbon fixation, or primary production, throughout Earth history**. B.** Relative atmospheric partial pressure of O_2 relative to present atmospheric level, data taken from Lyons et al, 2014, *Nature*.

1. ENDOSYMBIOSIS AND THE CHLOROPLAST EMERGENCE

DAWN OF THE ENDOSYMBIOTIC THEORY

Constantin Mereschkowsky, a Russian botanist, noticed the physiological similarities between what were called at the time cyanophyceae (cyanobacteria) and chromatophores (chloroplasts). He noted that both had round shapes, simple structures, green pigments, no true nucleus and that they were both able to assimilate CO_2 in the light. In 1905, drawing from previous studies describing chloroplast proliferation by division, their hereditary transmission in plant cells and their ability to survive transiently outside of the cytoplasm, he established the endosymbiotic theory. He suggested that chloroplasts derive from ancestral cyanobacteria following an endosymbiotic process [(Mereschkowski, 1905) English translation by (Martin and Kowallik, 1999)]. However, his hypothesis encountered fierce detractors and was progressively forgotten.

About sixty years later multiple microscopic and biochemical studies revealed the existence of DNA in the chloroplast of various organisms (Ris and Plaut, 1962; Gibor and Izawa, 1963) and the endosymbiotic theory was remembered (Ris and Plaut, 1962). It was brought back to the forefront by Lynn Margulis (Sagan, 1967) and has been since then largely accepted.

The advance of biochemistry built up the body of evidence for a cyanobacterial origin of the chloroplast. For instance, the composition of the plastid membranes is very different from that of all the other cells compartments. Unlike them, the chloroplast envelopes are composed mostly of galactolipids (Wintermans, 1960), produced in part from the stroma, as the cyanobacterial thylakoid (Holzl and Dormann, 2007). The chloroplast gene expression machinery was shown to derive from a bacterial one, with a 70s ribosome, whose rRNAs are encoded in the plastid genome (Miller and McMahon, 1974), and a PEP (Plastid encoded Polymerase) RNA polymerase (Sijben-Muller *et al.*, 1986). This is illustrated by the chloroplast sensitivity to bacterial transcription and translation inhibitors (rifampicin, spectinomycin and chloramphenicol).

Early phylogenetic trees illustrated how the evolutionary histories of mitochondria and plastid were distinct from that of their hosts (Schwartz and Dayhoff, 1978; Gray *et al.*, 1984; Giovannoni *et al.*, 1988). The organisation and composition of the ATP synthases from cyanobacteria and chloroplast were found to be related (Cozens and Walker, 1987) for instance. Then, the rise of genomics further cemented the endosymbiotic theory. Among many examples: hundreds of genes originating from cyanobacteria were found in *Arabidopsis thaliana* genome (Martin *et al.*, 2002).

THE ANCESTRAL CYANOBACTERIUM

To this day, there is still no definite consensus on which cyanobacterial clade is the most closely related to modern plastids. Some phylogenies propose a very early branching of the chloroplast ancestor in the oxygenic cyanobacteria, others on the contrary a late branching with nitrogen-fixing cyanobacteria (Deusch *et al.*, 2008; Dagan *et al.*, 2013; Ochoa de Alda *et al.*, 2014). This uncertainty might be explained in part by an uncomplete view of Cyanobacteria diversity and the extent of cyanobacterial evolution since the plastid was formed (de Vries and Archibald, 2017). Moreover, extensive intraphylum and interphylum LGT have been reported in cyanobacteria (Zhaxybayeva *et al.*, 2006), further complicating phylogenomic studies.

The discovery of the Gloeomargarita genus (Couradeau et al., 2012), introduced another potentially closely related group to primary plastids. The *Gloeomargarita* described so far are confined to fresh waters, with a preference for hot springs, and form biofilms. This kind of context, where microbes form inter-species communities, to benefit from oxygen production for example, could be an alluring melting pot to initiate endosymbiosis. Phylogenetical data based on plastid and nuclear genes support this hypothesis by placing *Gloeomargarita* as the sister group of the primary chloroplast ancestor (Ponce-Toledo *et al.*, 2017; Sanchez-Baracaldo *et al.*, 2017).

Both nitrogen-fixing cyanobacteria and Gloeomargarita are found in freshwater. This could perhaps mean that the plastid endosymbiosis occurred in a freshwater system. Molecular clocks (Parfrey *et al.*, 2011) and geochemical studies (Sahoo *et al.*, 2012) suggest that this primary endosymbiosis occurred before the slow oxygenation of the oceans. Possibly, the gain of a symbiont producing O_2 in the host, for the respiratory mitochondria, would allow them to venture into the anoxic oceans (Dagan *et al.*, 2013). And so, Rhodophyta and Chloroplastida progressively colonised the oceans, while glaucophytes kept to freshwater niches.

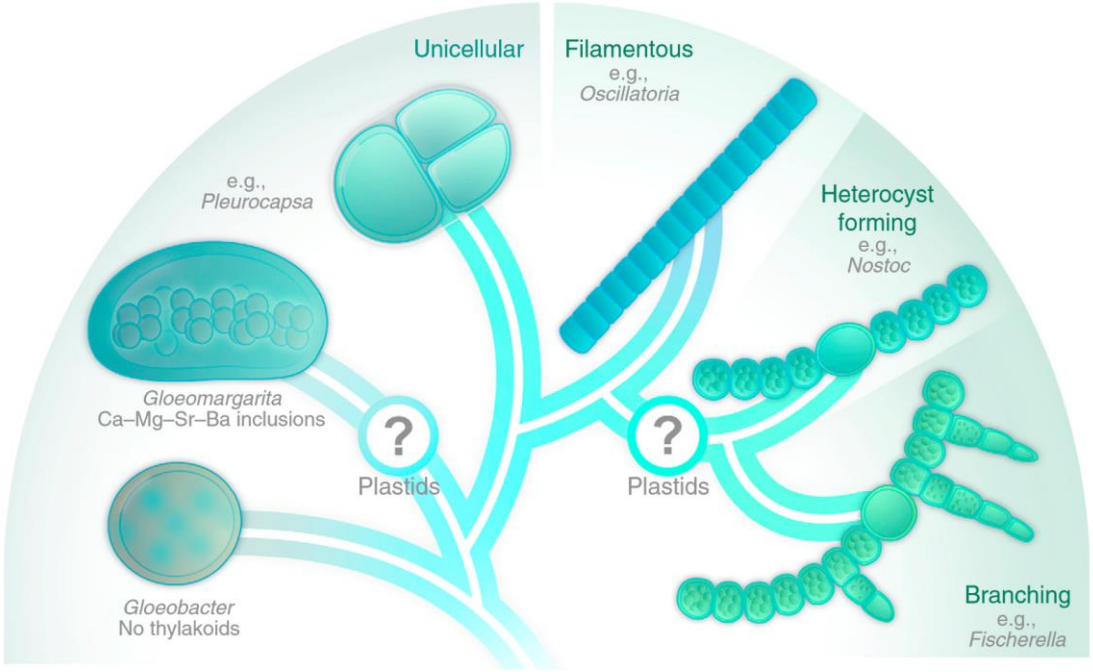


Figure 3: Simplified cladogram of cyanobacterial diversity from (de Vries and Archibald, 2017). Two possible origins for plastids are indicated.

CURRENT VIEWS ON ENDOSYMBIOSIS

Host nuclear genomes are a patchwork of genes from numerous origins. LGT and endosymbiotic gene transfer (EGT) in photosynthetic organisms seem to have occurred massively in a relatively short period and since then slowed down considerably. While the implication of viruses as vectors for LGT and EGT could be crucial in algal and bacterial communities (Derelle *et al.*, 2008), links between phagocytosis and gene transfer have also been suggested (Ford Doolittle, 1998; Keeling and Palmer, 2008). Digestion of preys is a great opportunity for organisms to integrate exogenous DNA of various sources. One could imagine that eukaryotes feeding on cyanobacteria could have over time started to accumulate significant amounts of cyanobacterial genes. Then, a prey could have become able to escape digestion in some way, and the symbiotic relationship would have started to evolve (Larkum *et al.*, 2007). As the host became fully autotroph, thanks to its new organelle, phagocytosis would cease and so, the opportunity for gene transfer would recede.

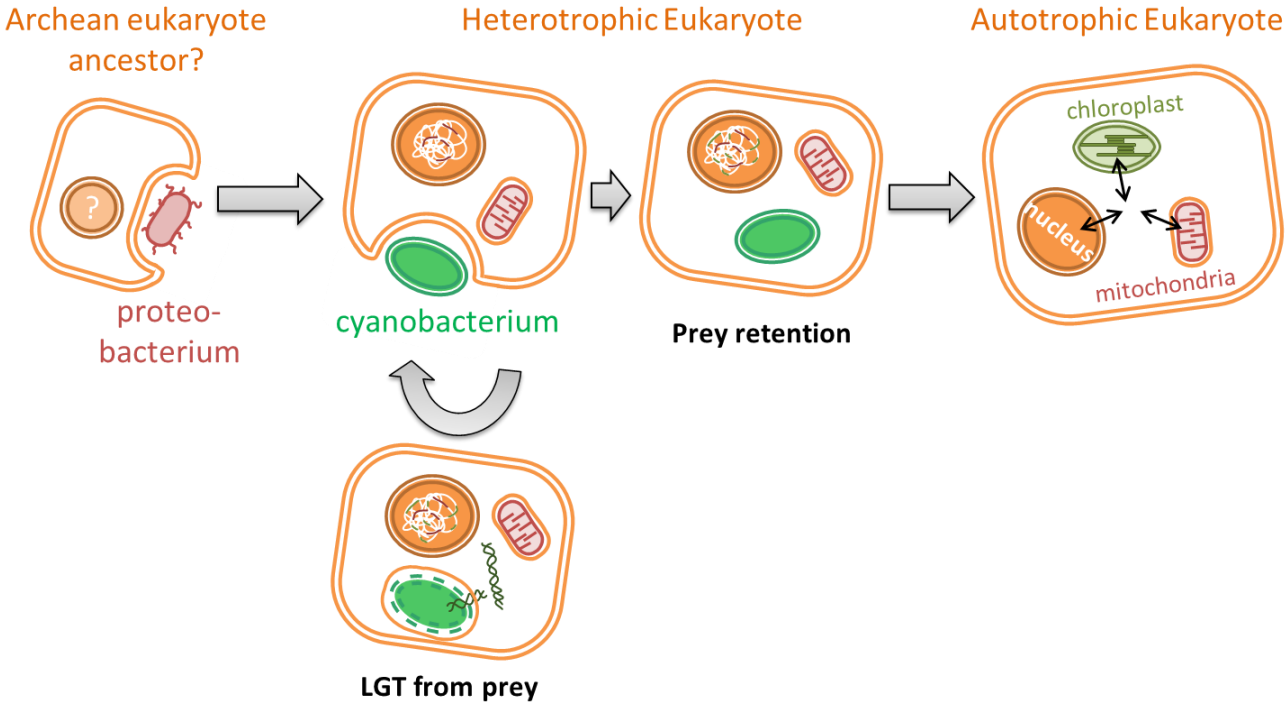


Figure 4: Primary chloroplast endosymbiosis. A possible history of chloroplast emergence is suggested.

Crucial steps in establishing endosymbiosis are first the ability of the prey to survive in the host, then the possibility for them to beneficially interact, the synchronisation of the endosymbiont proliferation and the host divisions to ensure its inheritance in both daughter cells, and lastly the ability to integrate the two organisms by importing proteins back into the organelle.

One theory to explain this rare event, based at first on the observation of chlamydial LGT in plants (Brinkman *et al.*, 2002), is that a chlamydial pathogen could have entered the eukaryote cell at the same time as a cyanobacterium and shielded it from digestion into an inclusion vesicle. This hypothesis dubbed *ménage à trois* also suggests that essential proteins of chlamydial origins were implicated in the metabolic connexion between the chloroplast and its host, thru glycogen metabolism (Ball *et al.*, 2013; Cenci *et al.*, 2017).

A complementary theory posits that the captured cyanobacteria developed a mechanism to protect themselves from antimicrobial peptides (AMP) produced by the host by importing them then degrading them. This mechanism closely resembles the import mechanism of nucleus-encoded proteins into the organelles (Wollman, 2016). It was proved recently that a subclass of helical amphipathic AMP can target proteins into the chloroplast or the mitochondria and conversely, that targeting peptides have retained some antibacterial activity (Garrido et al, submitted). Both theories are currently studied in our laboratory.

A BRIEF OVERVIEW OF THE DIVERSE PHOTOSYNTHETIC EUKARYOTES:

Photosynthetic organisms are spread across the tree of Eukarya (Figure 5). Nearly all of them stem directly or indirectly from a single primary endosymbiotic event (*Paulinella* is a notable exception that we will talk about later). Many of them are unicellular algae. The increasing number of algal genomes published (reviewed in (Blaby-Haas and Merchant, 2019)) provides a great opportunity to elucidate the origin of the various eukaryotic algae.

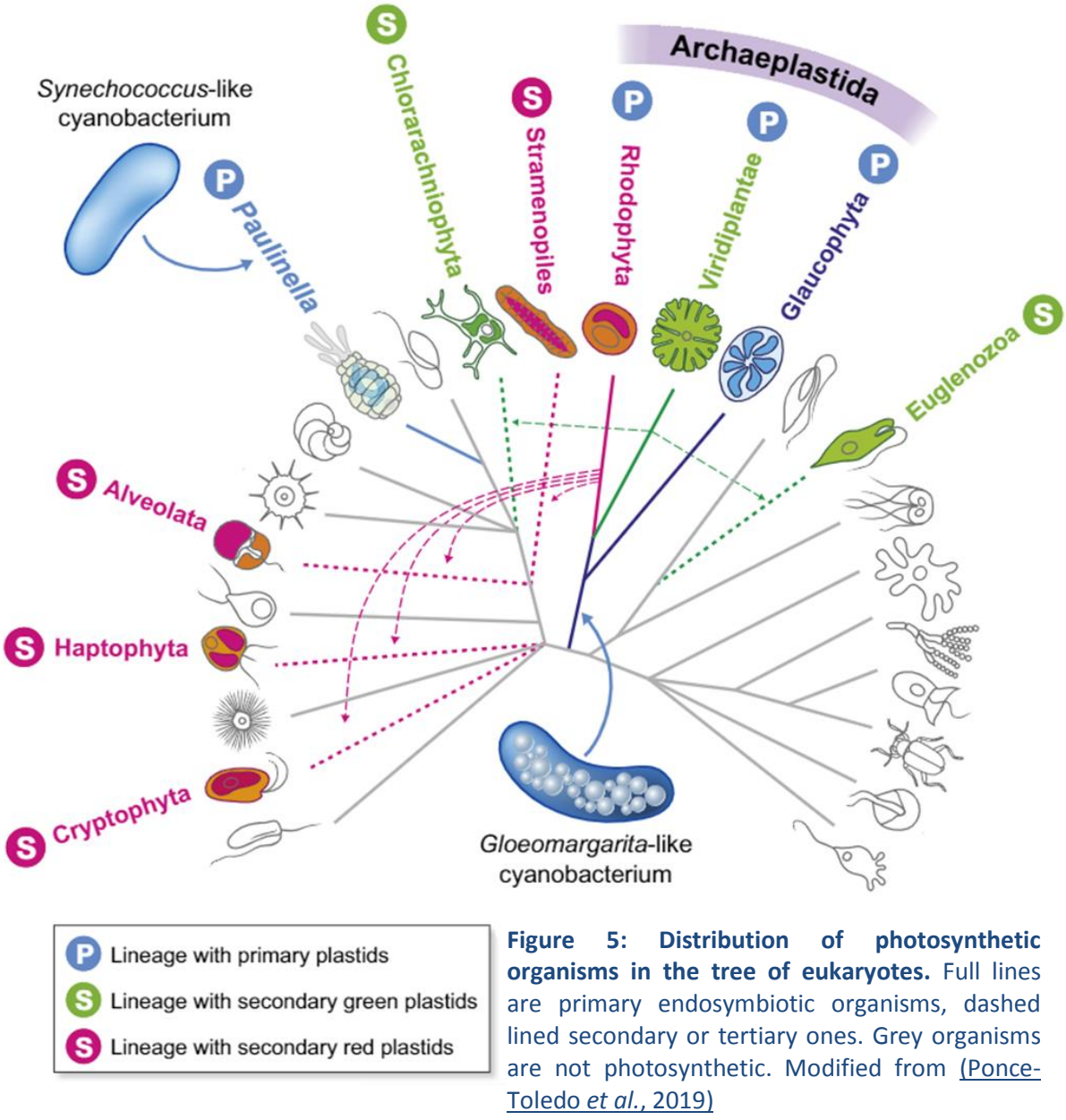


Figure 5: Distribution of photosynthetic organisms in the tree of eukaryotes. Full lines are primary endosymbiotic organisms, dashed lined secondary or tertiary ones. Grey organisms are not photosynthetic. Modified from (Ponce-Toledo *et al.*, 2019)

The Archaeplastida

Archaeplastida derive from a single primary endosymbiotic event, when an ancestral cyanobacterium was absorbed by a primitive eukaryote approximately 1.5 billion years ago (Parfrey *et al.*, 2011). Their plastid genomes contain inverted repeats around the rRNA operon (Palmer, 1985; Keeling, 2010).The common origin of their chloroplasts, which have two membranes of cyanobacterial origin, has been reinforced by phylogenetical data (Rodriguez-Ezpeleta *et al.*, 2005). The monophylly of the Archaeplastida is also supported by mitochondrial DNA phylogenies (Burger *et al.*, 1999). Studies have shown for instance the common origins of the subunits of the chloroplast import machinery TIC-TOC (**t**ranslocon at the **i**nner envelope of **c**hloroplast, **t**ranslocon at the **o**uter envelope of the **c**hloroplast) (McFadden and van Dooren, 2004; Price *et al.*, 2012).

Glaucophyta:

Glaucophytes are unicellular algae living in fresh water. Their blue tinged plastids have been historically called cyanelles. Their resemblance with cyanobacteria led old studies to classify the cyanelles as symbiotic cyanophyceae (Hall and Claus, 1963). This small group is often viewed as a witness of the chloroplast ancestry as the cyanelle does not only superficially look like a cyanobacterium. In fact, it bears a peptidoglycan layer between its membranes (Pfanzagl *et al.*, 1996). It also uses phycobilisomes as light-harvesting antenna complexes, much like its cyanobacterial ancestor. For a recent review on glaucophytes see (Jackson *et al.*, 2015).

Rodophyta

Rhodophytes are unicellular and multicellular algae. They are found either in fresh or seawater. They possess phycobilisomes associated with their PSII and a Light Harvesting Complex (LHC) associated with their PSI (Durnford *et al.*, 1999). They express chlorophyll c. Their genomes underwent considerable gene losses (Qiu *et al.*, 2015) that deprived them of cilia for example. Cyanidiophytes gained many new functions from LGT with bacteria or archaea, allowing them to adapt to extreme environments. In contrast their chloroplast kept a slightly higher number of genes (Green, 2011).

Viridiplantae

Viridiplantae comprise unicellular and multicellular organisms, aquatic and terrestrial ones, green algae and land plants. Their photosynthetic apparatus is different from the cyanobacterial one, they lost the phycobilisomes and rely solely on LHCs (LHCI and LHCII). They produce chlorophyll b, their chloroplast stocks polysaccharides as starch. Their thylakoids are appressed. Multicellular forms have evolved separately several times in the green algae. One ancestral alga gave rise to the terrestrial plants and their elaborate architecture (De Clerck *et al.*, 2012).

SECONDARY PHOTOSYNTHETIC EUKARYOTES

Many eukaryotic groups have acquired photosynthesis independently by engulfing primary endosymbiotic organisms in a secondary endosymbiosis. Euglena absorbed a green alga. Cryptophytes, dinoflagellates, diatoms, haptophytes have absorbed red algae (Keeling, 2010; Ponce-Toledo *et al.*, 2019). Some might even have arisen after tertiary endosymbiosis, as the secondary red plastids appear to be monophyletic (Munoz-Gomez *et al.*, 2017) but their hosts are not. Their origins are for some still quite hotly debated. The complex systems of 3 or 4 membranes around their chloroplast have been a strong argument of their acquisition by phagocytosis. The existence of relict of the primary eukaryotic host genome, the nucleomorph, found in the complex chloroplast of cryptomonads or chlorarachniophytes, is also proof of their origin (Gilson *et al.*, 1997). In my thesis I will not focus on secondary endosymbionts.

INDEPENDENT PRIMARY ENDOSYMBIOSIS

Nearly all known plastids descend from the unique event of cyanobacterial capture in the primary endosymbiosis 1.5 Ga ago, whether from direct inheritance or thru secondary absorption. This paucity suggests that successful and durable endosymbiosis must be difficult to establish. One famous exception is the thecamoeba *Paulinella chromatophora* and its sister lineage *Paulinella micropora*. They recently (~100 Million years ago) engulfed a α-cyanobacterium and potentially reflect early endosymbiotic stages. *P. chromatophora* was discovered at the turn of the 19th century by Robert Lauterborn, and its blue chromatophore led scientists of the time to think that this organism was parasitized by a cyanobacteria [(Mereschkowski, 1905) English translation by (Martin and Kowallik, 1999)]. Their photosynthetic organelle is called chromatophore, is surrounded by a peptidoglycan wall, and has similarly to the chloroplast undergone genome reduction of about two thirds, but kept almost all subunits of its photosynthetic machinery (Nowack *et al.*, 2008). *P. chromatophora* is phototrophic and does not use phagocytosis anymore, unlike its close relatives, such as *Paulinella ovalis*. This independent endosymbiotic event is an interesting model to compare to plastid evolution.

2. *Chlamydomonas reinhardtii*

A flexible model organism

Chlamydomonas reinhardtii is a unicellular green alga of the Chlorophyceae class. The Chlamydomonas genus is spread worldwide and in a wide variety of ecosystems; from tropical to artic zones, from sea-water to fresh water, damp soils or polluted sewages (Salome and Merchant, 2019)… *C. reinhardtii* and some of its "sociable" multicellular relatives such as *Gonium pectorale* and *Volvox carteri* were described in 1838 by Christian Ehrenberg.

C. reinhardtii has the advantages of microbial models: it can be cultivated either in liquid or solid media, grows fairly fast (about 8 hours between vegetative divisions) and rare events can be easily studied, as very large populations can be screened. *C. reinhardtii* can grow heterotrophically if supplied with acetate. The cell contains a single large chloroplast that can be readily transformed by biolistic means, DNA integrates by homologous recombination. However, the plastome is highly polyploid, present in about 80 copies per cell. So, after transformation, mutations need to undergo homoplasmisation: each copy of the chloroplast genome must bear the mutated allele. This is achieved by successive selective rounds of sub-cloning on selective medium over several generations. This process can take about 2 months, depending on the growth rate of the strain. The nuclear genome can also be modified but transforming DNA inserts randomly. Recently, CRISPR tools have been adapted to *C. reinhardtii* (Shin et al., 2016). However, our laboratory has not mastered its use yet. Thankfully, the nuclear genome is haploid, so random extinction of genes can be produced relatively easily. Lastly, it is also possible to transform the small mitochondrial genome of *C. reinhardtii* with a biolistic approach (Remacle *et al.*, 2006).

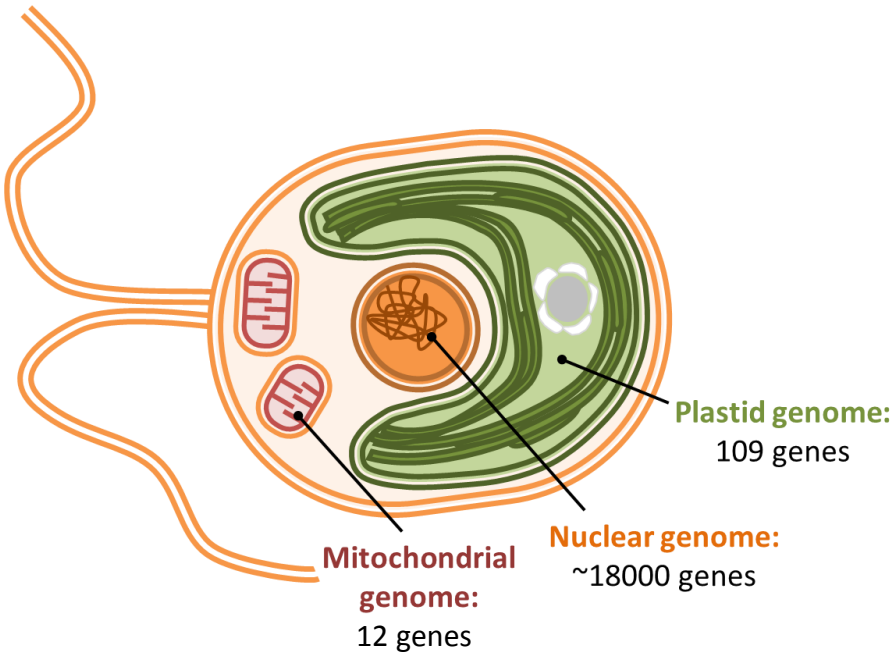


Figure 6: Chlamydomonas reinhardtii, with its three genomes.
All of them can be transformed.

Under nitrogen starvation *C. reinhardtii* produces gametes that can fecund the other mating type (either + or -). This allows classical genetic studies (Harris, 2001). Genetic inheritance is Mendelian for the nuclear genome, and uniparental for the chloroplast and mitochondrial genomes (Figure 7). The mating type + (mt^+) parent transmit its chloroplast genome to the progeny, while the mt^- progenitor transmits its mitochondrial genome. A small percentage of zygotes can fail to perform meiosis and generate stable diploid vegetative cells. They can be used to determine if mutations are dominant or recessive.

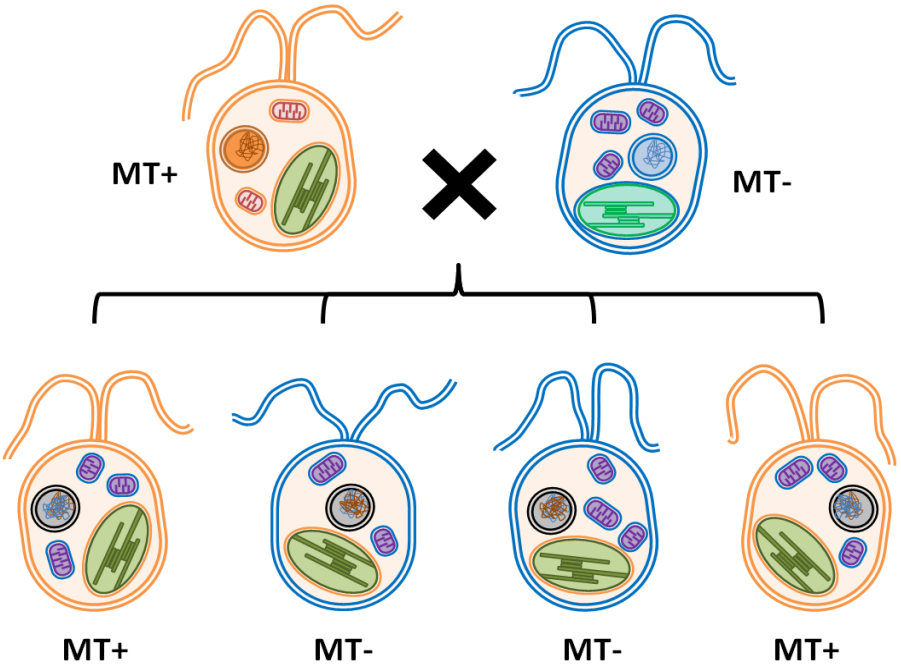


Figure 7: Genome segregation in sexual reproduction of *C. reinhardtii*.

So, in *Chlamydomonas reinhardtii*, all genetic compartments can be transformed, this gives great possibilities to use elaborate reverse genetic techniques. Classical genetic approaches can be used as well, thanks to its ability to reproduce sexually and the possibility to recover stable diploid vegetative cells. Its microbial nature allows screening millions of cells at once to hunt rare genetic events. Spontaneous mutations in the nucleus can produce phenotypes more readily, since the genome is haploid…

All those properties make *Chlamydomonas reinhardtii* an invaluable eukaryote model to study both photosynthesis and respiration, but also the relationship between the organelles and the nucleus, among many processes. Cilia biogenesis and function, for example, are studied in *C. reinhardtii*.

STRUCTURAL PROPERTIES

Chlamydomonas reinhardtii cells measure about 10 µm in diameter, are surrounded by a cell wall made of glycoproteins and have two cilia of equal length at their anterior pole, allowing them to swim toward or away from light (this phenomenon is called phototaxis). Those cilia are also necessary for the gamete agglutination in the reproductive cycle; cells with no cilia are sterile.

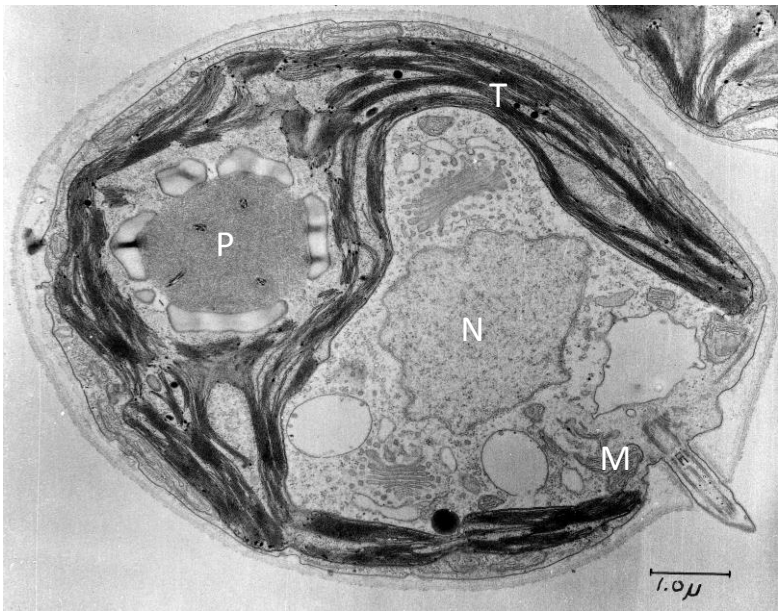


Figure 8: Transmission electron micrograph of a *C. reinhardtii* (y-1 mutant) cell. Originally published (Ohad *et al.*, 1967) made available by I. Ohad (2012) CIL:37252, Chlamydomonas reinhardtii. CIL. Dataset. https://doi.org/doi:10.7295/W9CIL37252
M: mitochondrion
N: nucleus
P: pyrenoid
T: Thylakoids

The single chloroplast of *C. reinhardtii* is "cup" shaped and is nested at the posterior side of the cell with its three lobes pointing toward the anterior side. It contains a structure rich in pigments called eyespot, that senses light and allows phototaxis. Long membrane structures, the thylakoids, form well defined appressed and non-appressed domains. Thylakoids are the sites where the photochemical part of photosynthesis takes place. A large pyrenoid is present at the centre of the base of the chloroplast. This liquid-like structure is composed of a multitude of RuBisCOs aggregated and bound by EPYC1 (Mackinder *et al.*, 2016) and forming a fluid matrix. The pyrenoid is surrounded by a starch sheath, potentially implicated in containing the matrix in a single pyrenoid (Itakura *et al.*, 2019). The high CO_2 concentration of the pyrenoid enhances the enzymatic efficiency of the RuBisCO to fix more inorganic carbon. The pyrenoid is crossed by thin thylakoids (Engel *et al.*, 2015). As the lumen is enriched in HCO_3^- and as a carbonic anhydrase (CAH3) is present in those intersecting thylakoids, dehydration of HCO_3^- might supply the necessary CO_2 to the pyrenoid.

PHOTOSYNTHESIS IN THE THYLAKOIDS

The major photosynthetic complexes of the electron transfer chain (PSII, cyt. b_6f and PSI) are embedded in the membrane along the ATP synthase. LHCs, mobile pigments complexes, collect the light energy. Then this excitation energy is transmitted to the special pairs of chlorophylls of PSII (chlorophylls P680) and PSI (chlorophylls P700). The excited chlorophylls release an electron, which goes through multiple acceptors out of the photosynthetic complexes. The PSII transmits its electrons to the membrane soluble plastoquinones (PQ). PQs transmit the electrons to the cyt. b_6f which donates it to the plastocyanin (PC). The PC shuttles the electrons to the special chlorophylls of the PSI, replacing the electrons released by the photon-driven charge separation. The PSI transfer the electrons to ferredoxin (Fd), then lastly to ferredoxin-$NADP^+$ reductase (FNR) that reduce an $NADP^+$ to produce NADPH that can be used in carbon fixation. To replenish the lost PSII electrons, the water-oxidising complex split two water molecules in O_2 and $4H^+$ and $4e^-$. The cyt. b_6f complex also perform the Q cycle, which transport more protons into the lumen. All those reactions generate a proton gradient across the thylakoid membrane. This proton motive force is exploited by the ATP synthase to produce ATP. The ATP synthase will be further described from p50.

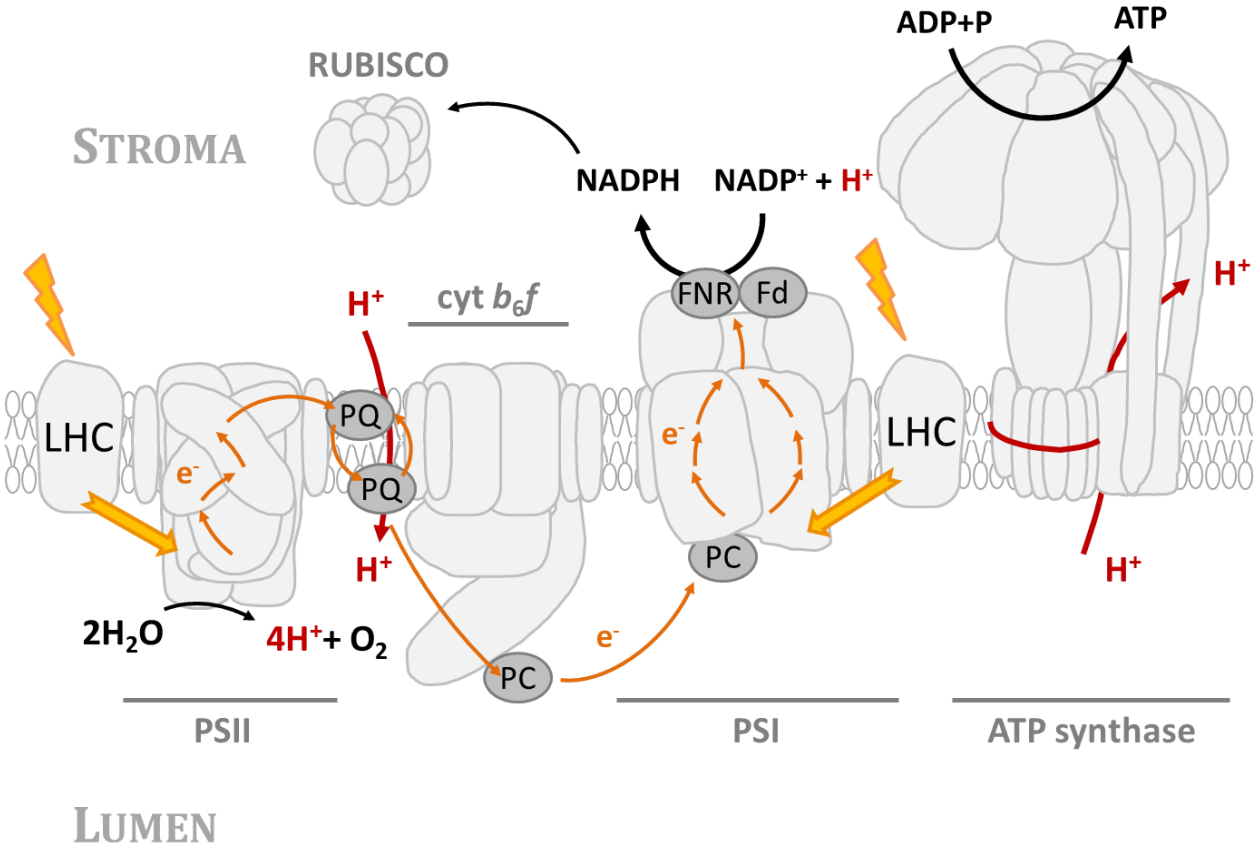


Figure 9: Schematic cross-section of the photosynthetic chain in the thylakoid membrane of *C. reinhardtii*.

3. The chloroplast genome: expression and regulations

The chloroplast genome structure reflects its history

Chloroplast genomes underwent drastic size reduction

Compared to their related free-living cyanobacteria, chloroplasts have much smaller genomes. *Gloeomargarita lithophora* has a genome of 3 Mb, while primary chloroplasts tend to have genomes of around 0.10 to 1.15 Mb. The number of genes encoded also dropped down in plastids from potentially 3000 in the ancestral cyanobacteria to about 100. In captive life, many genes of the cyanobacterial ancestor certainly were obsolete, for instance those implicated in motility, and could be lost without consequence for the cyanobacterium survival. Other genes became redundant, therefore dispensable, when the host and plastid integrated, such as those encoding DNA polymerase. Last, many genes were transferred to the nucleus of the host cell (Martin *et al.*, 1998; Dyall *et al.*, 2004; Timmis *et al.*, 2004). Interestingly, genome reduction occurs independently of lineages, it has also happened in the chromatophore of *P. chromatophora* and in mitochondria.

Those transfers have slowed down throughout evolution but can still occur in plant cells. For example; numerous chloroplast sequences transferred to the spinach nucleus have been found already decades ago (Timmis and Scott, 1983). The relative recent nature of those transfers is indicated by the fact that the sequences duplicated from the chloroplast to the nucleus are still quite similar. Another example is a study on the rice genome (*Oriza sativa*). The author observed that very large fragments of cpDNA (chloroplast DNA) integrate frequently in the nuclear genome, where they rapidly get shuffled around. Those cpDNA sequences predominantly integrate directly in the nucleus and not via retro-transcription of RNA, as genes are not transferred more frequently that intergenic DNA (Matsuo *et al.*, 2005). Interestingly, DNA transfers from chloroplast to mitochondria were also observed. Studies of mitochondrial EGT illustrated that the mechanism of integration in the nuclear genome relied on double-strand break repair events (Hazkani-Covo *et al.*, 2010).

Recent gene transfers in some organisms are rarer. In *C. reinhardtii* (Lister et al., 2003) the transfer rate is extremely low compared to tobacco in artificial gene transfer experiments. *C. reinhardtii* lone, unique chloroplast might be the cause. Indeed, organisms with only one or two plastid apparently barely endure EGT nowadays. Perhaps it is because a ruptured chloroplast is a relatively common opportunity for gene transfers to happen; in cells with many chloroplasts this would not induce much trouble for the cell. However, when the cell has very few (or even one) chloroplasts to its disposal, dispensing of one of them could have dramatic consequences (Bock, 2017).

But why are those gene relocations favoured by evolution? A first explanation is that the integration of new DNA in the host genome creates opportunities for evolutionary innovations; thru chimerisation, or simply by providing the cell with gene duplicates that can evolve and diverge to create new functions. Hence, hosts where those events took place would gain an evolutionary edge over others.

Another discussed hypothesis is based on population genetics. Genomes of the isolated endosymbiotic *Buchnera* bacterial lineages, with small populations, no recombination and no sexual reproduction, diverge faster by accumulating mildly deleterious mutations, in a phenomenon called Muller's ratchet (Moran, 1996).

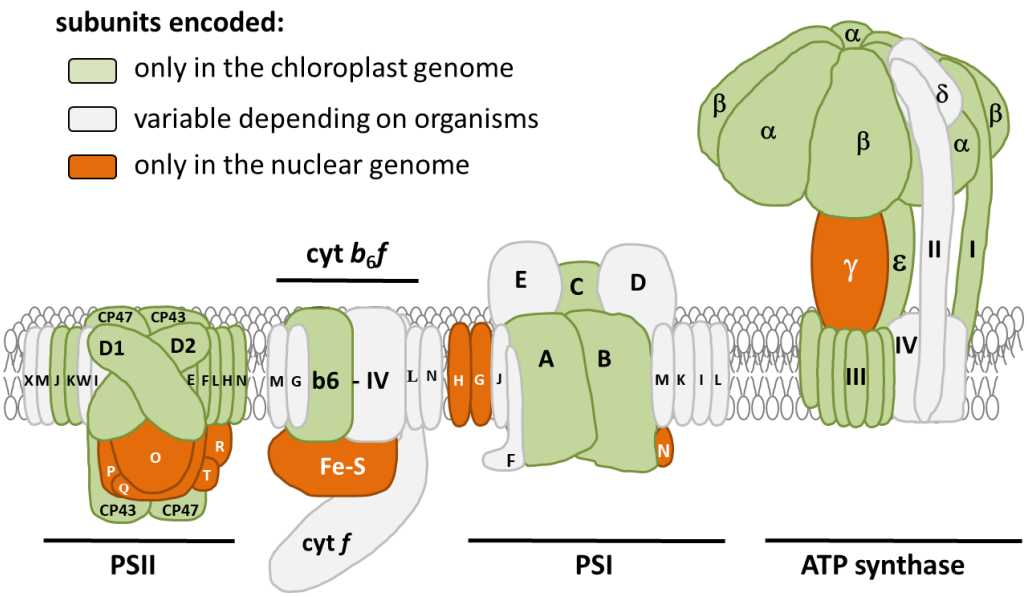
Interestingly, those bacteria also undergo a diminution of their GC contents, as is observed in modern chloroplast genomes that are AT rich. If slight reduction of fitness in sheltered organisms might not impact much their short-term persistence in the host, they could become problematic for the host-endosymbiont consortium. The relocation of organelle genes in the "sexually active" nuclear genome could protect them from this degeneration (Martin and Herrmann, 1998). However, contrary to metazoan mitochondria, which fit this view, chloroplasts tend to have fairly low substitution rates (Smith, 2015), maybe thanks to efficient repair mechanisms or to their high polyploidy. So attributing gene transfer mainly as a strategy to alleviate Muller's ratchet would be farfetched.

Yet another potential reason could be to relinquish the regulatory functions to the nucleus, to better control organelle expression and harmonise the cell functions. Altogether, reasons for gene relocation are probably numerous and complex. However, as we will see later, there seems to be common pressures that discourage transfer of certain genes.

Notwithstanding the varying frequency of those recent EGT, all chloroplast genomes (bar the peculiar recent chromatophore of *P. chromatophora*) transferred the genes encoding many essential subunits of photosynthetic complexes to the nucleus. Many of those gene handovers to the nucleus proceeded independently. Subunits are encoded in different compartments across species (Figure 10). But there is striking convergence of which subunit remained in the plastid. Two salient theories, not mutually exclusive, were proposed to explain this hypothesis.

One is called the CoRR (Co-location for Redox Regulation) hypothesis. It posits that the subunits retained in both chloroplast and mitochondria are key components in the electron transport chain and that their expression is regulated following redox fluctuations (Allen and Raven, 1996; Allen, 2017). An argument in favour of this hypothesis is that hydrogenosomes and mitosomes, former mitochondria that do not perform respiratory electron transport, have completely lost their genomes, presumably because they did not have to regulate themselves other redox sensitive functions. However, some key redox proteins such as the Rieske iron-sulphur protein have been transferred to the nucleus in many organisms, and so this sole theory is unlikely to fully explain the pattern of gene relocation from organelles to nucleus.

The other theory is based on the fact that proteins presumed difficult to import tend to stay in the chloroplast. Highly hydrophobic proteins with more than 3 transmembrane helices seem mostly confined to their organelle genome, mitochondrial or chloroplastic (Popot and de Vitry, 1990). Additionally, some proteins might require concerted translation and membrane integration to be functional. It was observed long ago in *Chlamydomonas* that part of the chloroplast ribosomes associate with the unstacked thylakoid membranes (Chua *et al.*, 1973; Chua *et al.*, 1976) and that this membrane bound fraction increases with light exposure, suggesting that those ribosomes are implicated in the synthesis of photosynthetic transmembrane proteins. In maize around half of the thylakoid membrane proteins integrate in the membrane co-translationally, as soon as the first transmembrane segment protrudes from the ribosome (Zoschke and Barkan, 2015). The short trans-membrane proteins, whose transmembrane domain does not exit the ribosome before translation is finished, are targeted post-translationally.



based on ~100 chloroplast genome sequences of plants and green algae

Figure 10: Cartoon of the photosynthetic complexes in the thylakoid membrane, genome origins of the subunits across Viridiplantae are indicated in colours.

Genome organisation and composition

Plastid genomes are quite diversified in their structure; they tend to be circular, can sometimes oligomerise and can take up a linear form. They often have large inverted repeats comprising the rRNA operon inherited from the cyanobacterial ancestor (Palmer, 1985; Keeling, 2010). Their size varies from 0.15 to 1.35 Mb, but this is not necessarily correlated with chloroplast gene number, as it can be imputed to variations of intergenic region length. Plastome size variation can also stem from the length of the inverted repeats (IR); in longer ones more genes will be found and so the genome size will increase with two copies of the IR genes.

In *C. reinhardtii*, as in other Chlorophyceae, short dispersed repeats are particularly prominent (Maul *et al.*, 2002). A putative transposon, Wendy, acquired by *C. reinhardtii* (Fan *et al.*, 1995) might have induced gene shuffling of the genome by homologous recombination. In *C. reinhardtii* synteny has been mostly lost compared to plant chloroplasts. This caused the ancestral operons to be completely modified: genes are still transcribed in polycistronic units but irrespective of their functions.

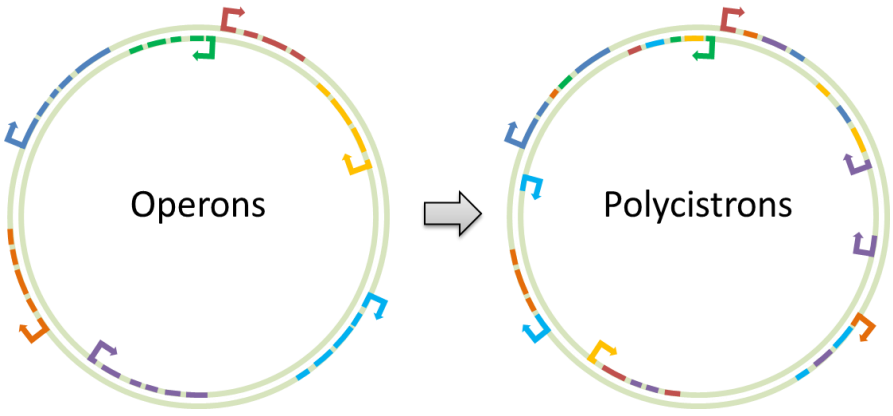


Figure 11: Gene shuffling: Left: Genes organised in operons (bacteria, plant chloroplasts to some degree) **right:** the polycistronic expression of genes (*Chlamydomonas*). Colours schematise the function of the gene products.

Rhodophyta and Glaucophyta have generally retained more genes in their plastome than Viridiplanta (Green, 2011; Jackson *et al.*, 2015) but still have compact genomes of small sizes.

The core set of plastid genes conserved across the Viridiplantae encodes: the core subunits of a bacterial type RNA polymerase, a full set of tRNAs and rRNAs, some of the ribosomal proteins of both of the small and large subunit, and subunits of the photosynthetic apparatus: the large subunit of RuBisCO (*rbcL*), many of the subunits of the thylakoid membrane electron transfer chain: PSI (*psaA, psaB, psaC, psaI*), PSII (*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbN, psbT*), cytochrome *b6f* (*petA, petB, petD, petG*) and half of the subunits of the ATP synthase (*atpA, atpB, atpE, atpF, atpH*). The following part will be focused mainly on *Chlamydomas reinhardtii.*

EXPRESSION OF THE CHLOROPLAST GENOME

EXPRESSION MACHINERY

Chloroplast genomes retain characteristics of their cyanobacterial ancestor. The gene expression machinery is one of the key functions that the plastid keeps, at least in part, encoded in its genome. Their ribosome is of bacterial type, with 30s and 50s subunits. Ribosomal rRNAs are preserved in the plastid genome and are quite conserved across lineages. But the chloroplast ribosome includes a mix of chloroplast and nucleus encoded ribosomal proteins. Land plant chloroplasts contain two types of RNA polymerases: one of bacterial type; the plastid encoded polymerase (PEP), and one or two of phage type; the nucleus encoded polymerases (NEP)(Allison *et al.*, 1996). In contrast, *C. reinhardtii* chloroplast only has a plastid encoded RNA polymerase of bacterial type, with a single nucleus encoded sigma factor (Surzycki and Shellenbarger, 1976). Plastids also have a full set of tRNAs.

While most of the gene expression machinery of the chloroplast is at least partially inherited from the cyanobacterial ancestor, the DNA replication system is not. Plastids do not encode a DNA polymerase (except some rare case of acquisition by LGT like in one cryptophyte (Khan *et al.*, 2007)), and depend on the nucleus genome for their replication. Plant organellar polymerases (POP), encoded in the nucleus and found both in the mitochondria and plastid of most bikonts, are not related to bacterial polymerases (Moriyama *et al.*, 2008). This could suggest that neither the DNA polymerase from the mitochondria or the chloroplast were successfully transferred to the nucleus, and that they were completely lost. This implicates that a DNA polymerase was recruited to replicate both the mitochondrial and chloroplast genomes. Surprisingly, no POP orthologue was found in *Chlamydomonas*, which instead targets a v-type polymerase to its organelles. Altogether, the loss of the probably redundant DNA polymerases of the organelle, suggests a great dependency of their genomes to the host.

TRANSCRIPTION, RNA EDITING, SPLICING AND DECAY

Some chloroplast transcripts have a dedicated promotor and are directly expressed in a monocistronic form. However, many mature transcripts in the plastome are produced from polycistronic transcription units. In *C. reinhardtii* particularly, the genes of a polycistron do not generally contribute to the same function nor are co-regulated; gene regulation is accordingly expected to occur mostly at a post-transcriptional level (Rochaix, 1996; Choquet and Wollman, 2002).

Chloroplast mRNAs, both in algae and plants, much like their prokaryotic counterparts do not have a protective 5' cap. The stroma contains exonucleases that degrade transcripts from 5'→3' and 3'→5' directions. To accumulate the transcripts are protected in their 3' end by a stem-loop (Drager *et al.*, 1996) and in the 5'end by specific factors that we will later study in detail (Drager *et al.*, 1998; Pfalz *et al.*, 2009). Those protective mechanisms prevent the progression of exonucleases on the mRNA; thus, they also define the boundaries of the mature mRNA.

In land plants, organellar mRNAs undergo cytidine to uridine editing, and sometimes U to C editing, often to restore a conserved amino acid or prevent a STOP codon (Smith *et al.*, 1997; Green, 2011). Defective unedited transcripts usually produce impaired or non-functional products. Editing is organelle specific; as mitochondrial sequences could not be edited in chloroplasts and vice versa. In fact editing is site specific and the editing machinery depends on specific nucleus-encoded factors, the first one was identified in 2005 (Kotera *et al.*, 2005). No RNA editing occurs in *C. reinhardtii* chloroplast.

Cis-splicing of plastid genes is quite rare in *C. reinhardtii* plastome; only two introns containing gene are present: *psbA* and *rrnL,* while 20 exist in *A. thaliana* plastome. A few cases of trans-splicing happen in chloroplasts, one example exists in *C. reinhardtii*: the *psaA* gene is transcribed from three exons at three distant loci. Those three transcripts are then trans-spliced together in two independent reactions that require several nuclear products (Choquet *et al.*, 1988).

Like their prokaryote relatives and unlike their cytosol counterparts, chloroplast mRNA are destabilised by the addition of poly(A) tails in their 3' end, those tails are partially heteropolymeric, meaning that they are composed of all four nucleotides but enriched in Adenosine (Schuster and Stern, 2009). Poly(A) tails are found both in the CDS and 3'UTR of chloroplast genes in *C. reinhardtii*, indicating that they are implicated in active degradation of mRNA (Bell *et al.*, 2016).

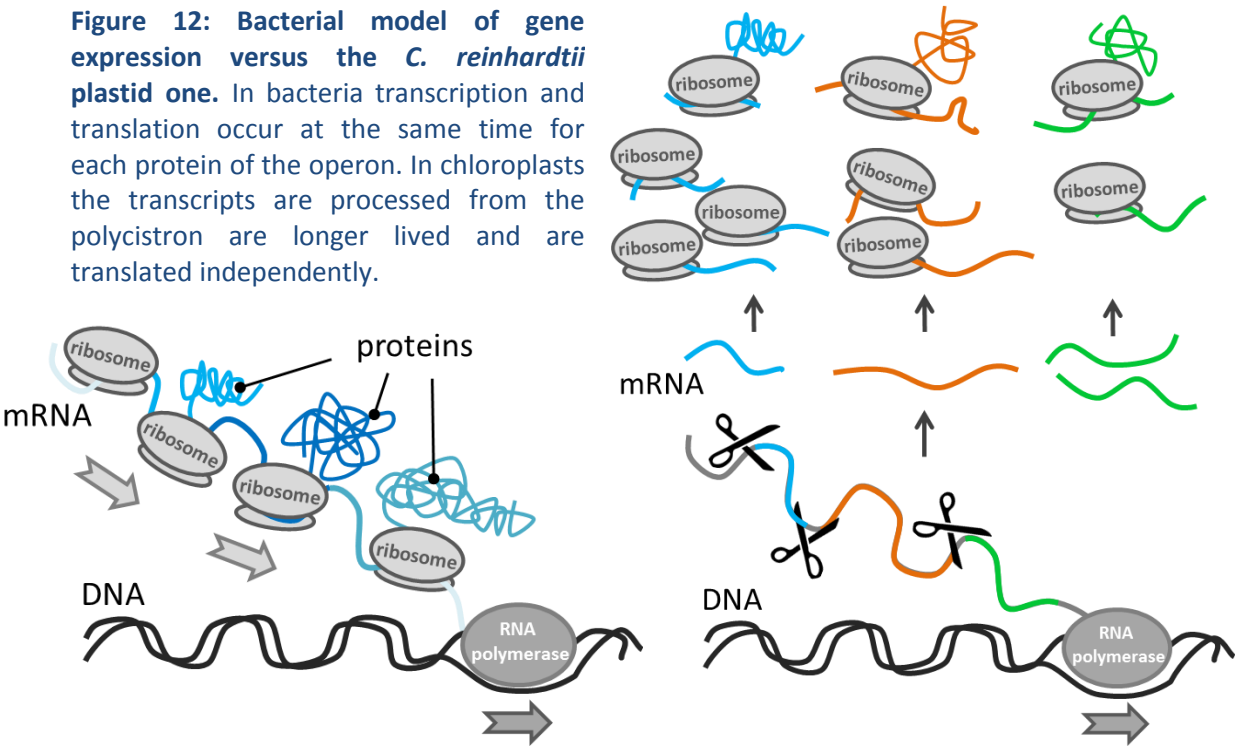
TRANSLATION

Plastid mRNA are long-lived compared to bacterial ones; they persist in the order of hours compared to the minutes of prokaryotic transcript lifetime. Experiments artificially decreasing the number of copies of the plastid genome in *C. reinhardtii* proved that gene dosage impacted mRNA transcription but not significantly translation (Hosler *et al.*, 1989; Eberhard *et al.*, 2002). Moreover, direct inhibition of chloroplast transcription with rifampicin in cells grown in mixotrophic conditions (with both light and acetate) did not affect much the chloroplast transcripts levels, while cells grown in phototrophic conditions displayed for many genes much lower levels of mRNAs, suggesting that they were more degraded. But even in those low mRNA conditions, translation remained stable and strong. This showed that there is no direct correlation between the level of transcript and the rate of translation, algae were able to bypass abnormal low transcript accumulation to translate normal levels of plastid protein (Eberhard *et al.*, 2002). This could also imply that translation actively depletes the mRNA stocks, and that it is unsustainable with limited mRNA levels. This decorrelation between transcript abundance and translation seems to be less pronounced in steady-state and was shown to be stronger in plants plastids than in *C. reinhardtii* (Trosch *et al.*, 2018). However, it appears that in many cases of mutations leading to diminished accumulation of a given transcript, translational regulations allow *C. reinhardtii* to cope with defect in gene expression.

Unlike in prokaryotes, transcription and translation do not massively co-occur in plastids, particularly in *C. reinhardtii*. Polycistronic transcripts in *C. reinhardtii* are accumulated at very low levels and are quickly processed in monocistronic mRNAs (Cavaiuolo *et al.*, 2017). Unlike in plants, translation of polycistronic transcripts appears rare. Considering the extent of operon shuffling in this alga this makes functional sense; genes found on a same polycistronic transcript might need to be expressed at very different rates. By being separately matured they can be translated and accumulated independently.

And so, mRNAs in the plastid are transcribed in excess and are "stored" without being translated. Control of gene expression in the chloroplast accordingly happen mostly post-transcriptionally (Rochaix, 1996; Choquet and Wollman, 2002; Germain *et al.*, 2013). A simplified comparative model of gene expression in bacteria and *C. reinhardtii* is drawn in Figure 12.

Figure 12: Bacterial model of gene expression versus the *C. reinhardtii* plastid one. In bacteria transcription and translation occur at the same time for each protein of the operon. In chloroplasts the transcripts are processed from the polycistron are longer lived and are translated independently.



Bacterial gene expression **Plastid gene expression in *C. reinhardtii***

An interesting property of translation in *C. reinhardtii* is that some plastid encoded genes are translated in localised areas of the chloroplast, presumably for an easier assembly or targetting, (Uniacke and Zerges, 2009; Sun and Zerges, 2015). For example, the large subunit of RuBisCO (encoded by the plastid *rbcL*) is translated next to the pyrenoid, and it appear that this targetting depend on the *rbcL* mRNA sequence and not on the polypeptide itself. As we saw before, in chloroplats, some of photosynthetic transmembrane proteins translation is initiated in the stroma, then on the thylakoids membranes where they integrate co-translationnally as soon as their first trans-membrane domain is exposed out of the ribosome (Chua *et al.*, 1973; Chua *et al.*, 1976; Zoschke and Barkan, 2015)

Protein degradation

Misfolded and damaged proteins in the chloroplast stroma can aggregate and become problematic for the cell. The major actor of the degradation of those proteins is the Clp (caseinolytic protease) protease complex (reviewed in (Rodriguez-Concepcion *et al.*, 2019)). FtsH is a major thylakoid membrane protease complex (reviewed in (Kato and Sakamoto, 2018)), implicated in the quality control of the membrane embedded photosynthetic complexes, notably PSII or cyt. b_6f, and its activity is strengthened under high light by a redox activated mechanism (Wang *et al.*, 2017). For a comprehensive review on chloroplast proteases see (Nishimura *et al.*, 2017).

Nucleus-encoded subunits of the photosynthetic complexes that over accumulate in the plastid are degraded, for example the small subunit of the RuBisCO is rapidly degraded in the chloroplast in absence of the large subunit (Schmidt and Mishkind, 1983).

But chloroplast localised proteins are not only degraded when they are damaged or unassembled. For example: under sulphur (Malnoe *et al.*, 2014; De Mia *et al.*, 2019) or nitrogen starvation (Wei *et al.*, 2014) photosynthetic complexes are degraded by FtsH and Clp to recover nutrients.

Nucleus and chloroplast, interactions and regulations

Import of nucleus-encoded proteins

Since so many genes chloroplast genes have been transferred to the nucleus, import mechanisms had to be established to allow the essential gene products back into their original compartment. Most proteins targeted to the plastid need to bear a chloroplast targeting peptide (cTP) at their N terminus. Those cTP are not conserved at the sequence level but share chemical properties that cause them to form an amphipathic helix. The hydrophobic part of the helix comes into contact with the chloroplast outer envelope, then the helix is pulled in the stroma by the TOC-TIC complex, the protein following along. The targeting peptide is then cleaved by the SPP (stromal processing peptidase) and degraded (Jarvis and Soll, 2001). The Hsp90 and Hsp70 cytosolic chaperones may be involved in the delivery of unfolded cytosolic pre-proteins to the TOC system (Paila *et al.*, 2015). The TOC-TIC import complex is composed partly of proteins of cyanobacterial origin and other that arose from the host cell (Price *et al.*, 2012). *In vitro* experiments suggest that the cTP interact preferentially with the chloroplast outer envelope thanks to its unique composition, as the only cytosol-exposed cell membrane containing galactolipids (Pinnaduwage and Bruce, 1996). Moreover, chloroplasts isolated from an *Arabidopsis* mutant deficient in production of digalactosyldiacylglycerol were defective in protein import (Chen and Li, 1998).

Proteins imported to the thylakoid lumen, have a second TP that similarly docks on the thylakoid membrane, then get imported in the lumen (Smeekens *et al.*, 1986).

RETROGRADE SIGNALLING

The nucleus/chloroplast interaction is not a one-way street. Retrograde signalling is a process of intracellular communication from the organelles to the nucleus. It is the counterpart of anterograde signalling, from the nucleus to the organelles. Retrograde signalling was suspected ever since it was discovered that nuclear encoded protein synthesis was modified following impairment of chloroplast translation (Bradbeer *et al.*, 1979), and has been studied mostly in plants. The chloroplast can be submitted to different types of stresses: Reactive oxygen species (ROS) production under high light conditions, high temperatures, pathogen infection, drought or starvation… Those stresses induce damages in the DNA, membranes and proteins of the chloroplast. And necessitate the production of emergency products from the nucleus: ROS scavenger or chaperones for instance (Rea *et al.*, 2018; Rochaix and Ramundo, 2018). Chloroplast sensors can detect those adverse conditions and transmit a signal either to the nucleus, for a transcriptional response, or to the cytosol for a post-transcriptional modulation. I will mention here two of the several known retrograde signals.

- **$^1O_2$**

  Under high light, excited chlorophylls cannot transmit their electron to the saturated electron transfer chain, already reduced. Carotenoids in the LHC can partially quench this excitation energy, and chlorophylls can dissipate the excess energy by emitting heat or fluorescence. However, they are susceptible of reacting with oxygen, particularly the P680 of the PSII, near the water oxidising system. In this case, singlet oxygen, a highly reactive species, is produced. This $^1O_2$ damage the photosystem and can cause lipid peroxidation (Dogra *et al.*, 2018). β-carotene and the nucleus encoded EXECUTER1 protein in *Arabidopsis,* or PSBP2 in *Chlamydomonas* (Brzezowski et al., 2012) are two separate singlet oxygen sensors in the chloroplast. Oxidised β-carotene derivatives are volatile and induce acclimation to high light stress in *Arabidopsis* (Ramel *et al.*, 2012) and could also be implicated in *Chlamydomonas* (Ledford *et al.*, 2007).

- **Linear tetrapyrroles: bilins**

  Phototropins, Flavin-based blue light receptors, modulate the expression of light harvesting proteins through the bilin pathway in *C. reinhardtii* (Im *et al.*, 2006). Studies of a mutant of the heme oxygenase (*hmox1*) in *C. reinhardtii*, which cannot produce biliverdin (a bilin) showed that those molecules were critical for photo acclimation. Addition of ectopic biliverdins partially rescued the *hmox1* phenotype (Duanmu *et al.*, 2013). Transcriptomic analyses revealed that numerous nuclear genes were expressed differently in response to the bilin signalling pathway and were implicated in the dark-light transition (Wittkopp *et al.*, 2017; Duanmu *et al.*, 2013). Interestingly, the synthesis of tetrapyrroles and photoreceptors is not regulated by this pathway. This could allow the cells to always be ready to detect light changes.

CES OF ORGANELLE COMPLEXES

Another process implicated in protein complexes assembly in organelle is the CES (Control by Epistasy of Synthesis) process (reviewed in (Choquet and Wollman, 2009)). CES subunits undergo a modulation of their translation according to their assembly state. This causes a sequential order in subunits synthesis that parallels their sequential assembly into a complex. Dominant subunits are necessary for the translation of the CES subunits. All photosynthetic complexes in *C. reinhardtii* chloroplast display at least one CES subunit (Figure 13).

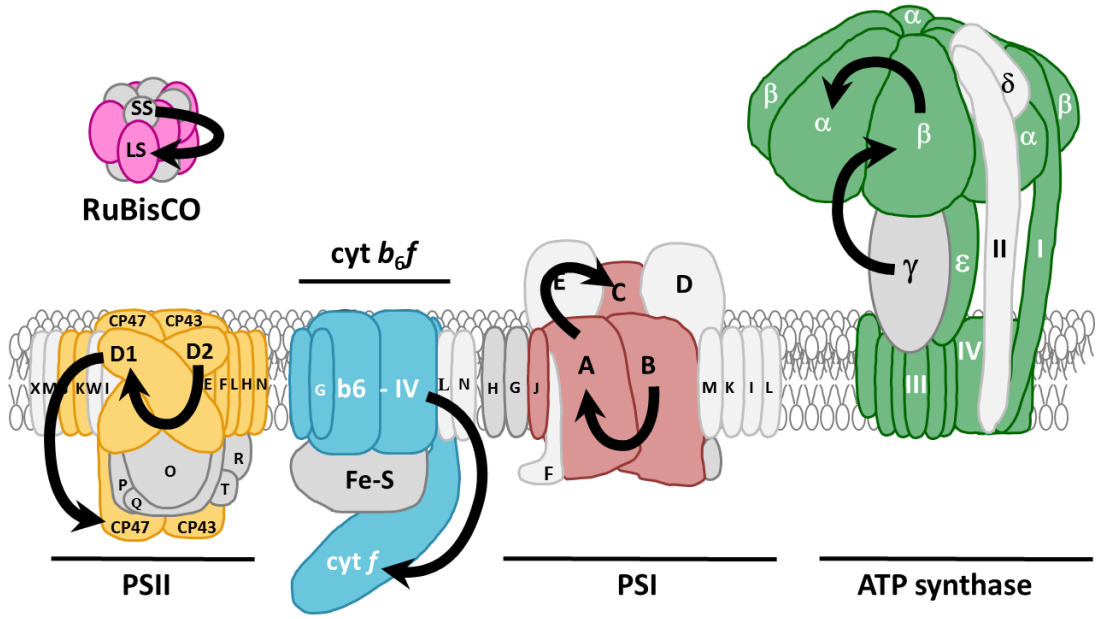


Figure 13: Every photosynthetic complex in *C. reinhardtii* displays some form of CES mechanism. Black arrows indicate the hierarchy of subunit assembly, CES subunits are at the receiving end of dominant subunits.

The most common form of CES described so far is a negative autoregulation: when over accumulating, an unassembled subunit prevents its own translation. When all the subunits assemble properly that inhibition is lifted, and translation can resume. The inhibition is exerted on the 5'UTR of the mRNA and can be studied thru chimeric genes. This mechanism has been observed for Cyt. *f* translation, unassembled Cyt. *f* embedded in the membrane induces the degradation of MCA1, the stabilisation factor of its mRNA, which is also implicated in its translation activation, thus in turn, less *petA* mRNA is accumulated and less of it is translated (Choquet *et al.*, 1998; Choquet *et al.*, 2001; Boulouis *et al.*, 2011). Other auto-negative controls have been observed for PsaA and PsaC (Wostrikoff *et al.*, 2004), D1 and apoCP47 (Minai *et al.*, 2006), β (Drapier *et al.*, 2007), the LS subunit of RuBisCO in tobacco (Wostrikoff and Stern, 2007) and in *C. reinhardtii* (Khrebtukova and Spreitzer, 1996).

CES can also sometimes rely on activating interactions, a rare example is the synthesis of the α subunit of the ATP synthase, which is stimulated in trans by the β subunit (Drapier *et al.*, 2007).

CES was not only observed in chloroplasts of tobacco or *C. reinhardtii* but also in yeast mitochondria (eg:(Calder and McEwen, 1991; Zambrano *et al.*, 2007; Bietenhader *et al.*, 2012)).

ORGANELLAR TRANS ACTING FACTORS

Historically, nuclear mutations, outside of photosynthetic subunits, affecting the photosynthetic complexes have been isolated and characterised (Kuchka *et al.*, 1988; Lemaire and Wollman, 1989; Drapier *et al.*, 1992; Monod *et al.*, 1992). They specifically impair the expression of one or a few chloroplast-encoded subunits, either at the mRNA, or translation stages. A wider picture soon emerged: transcripts in the chloroplast stroma but also the mitochondria need nuclear gene products (Barkan and Goldschmidt-Clermont, 2000) to be stabilised, matured and translated: the organellar trans-acting factors (OTAF). Those crucial factors can recognise and act on specific mRNA and as such are gene specific factors. I will call those specifically recognised mRNA: target mRNA in this manuscript. The OTAFs can be classified into big functional groups:

- M factors

They are necessary for the Maturation and stabilisation of their target mRNA. They specifically bind on the 5'UTR of an mRNA and protect it from 5'→3' exonucleases. Without its dedicated M factor a transcript cannot accumulate. The addition of an artificial structure called polyG track, which we will also use in our studies, in the 5'UTR allows the constitutive stabilisation of downstream sequences, even in the absence of the cognate M factor (Drager *et al.*, 1998). This illustrates the protective function of M factors against 5'→3' exonucleases. In polycistronic transcript of land plants the binding of a single M factor on the intergenic space can define the mature boundaries of both the 3'UTR of the upstream transcript, and the 5'UTR of the downstream transcript (Pfalz *et al.*, 2009). However, this has not been observed in *C. reinhardtii*.

- T factors

They are necessary for the translation of a specific mRNA. Their mode of action is poorly understood. Some of them act by opening secondary structures to uncover a sequestrated translation initiation signal, like TAB1 for *psaB* (Stampacchia *et al.*, 1997) or RBP40 for *psbD (Schwarz et al., 2007)* both in *C. reinhardtii*, or PPR10 for *atpH* in maize (Prikryl *et al.*, 2011). Some of them might interact with ribosomes or mRNA as they are translated and have been found to be associated with polysomes (as we will see in ***ARTICLE 3***). T factors could be implicated in ribosome recruitment, as interactions between helical repeat proteins and ribosome have been previously observed. An atypical generic PPR translation factor rPPR1 is part of the ribosome machinery in *Arabidopsis* mitochondria, along with 9 other rPPR (ribosomal PPR) (Uyttewaal *et al.*, 2008 ; Waltz *et al.*, 2019). The PPR proteins KRIPP1 and KRIPP8, also associated with the mitochondrial ribosome, have selective translation activation properties in *Trypanosoma brucei* (Aphasizheva *et al.*, 2016).

- Editing factors

Editing factors define the specific sites where edition of organellar RNA will occur. They recognise specific target sequences a few nucleotides upstream of the RNA editing site. They are prevalent in plants, but obviously not in green algae where editing does not occur. They are mainly represented by PPR proteins of a special subtype: PLS that I will present further below (Kotera *et al.*, 2005; Kobayashi *et al.*, 2012; Barkan and Small, 2014). This PPR-PLS group has expended in parallel to the expansion of editing in plants (Schmitz-Linneweber and Small, 2008).

- Splicing factors

These factors are required for the splicing of organellar mRNA. PPR5 is involved in the tRNA *trnG-UCC* splicing in maize chloroplasts (Williams-Carrier *et al.*, 2008) RAA1, RAA3, RAA8, RAT2, are 4 OPR proteins implicated in the trans-splicing of the *psaA* mRNA in *Chlamydomonas reinhardtii* (Merendino *et al.*, 2006; Marx *et al.*, 2015).

- Endonucleolytic factors

Putative endonucleolytic factors have been found both among PPR and OPR. They bear in their C terminal region domains related to endonucleolytic ones described in other organisms: a RNA Binding Abundant in Apicomplexa (RAP) domain (Lee and Hong, 2004; Boehm *et al.*, 2017) or small MutS-related (SMR) domain (Zhou *et al.*, 2017). But other factors devoid of such domains could also recruit endonucleases like RPF5 (Hauler *et al.*, 2013) RFL9 (Arnal *et al.*, 2014) and RFL2 (Fujii *et al.*, 2016), PPR proteins in *Arabidopsis* mitochondria.

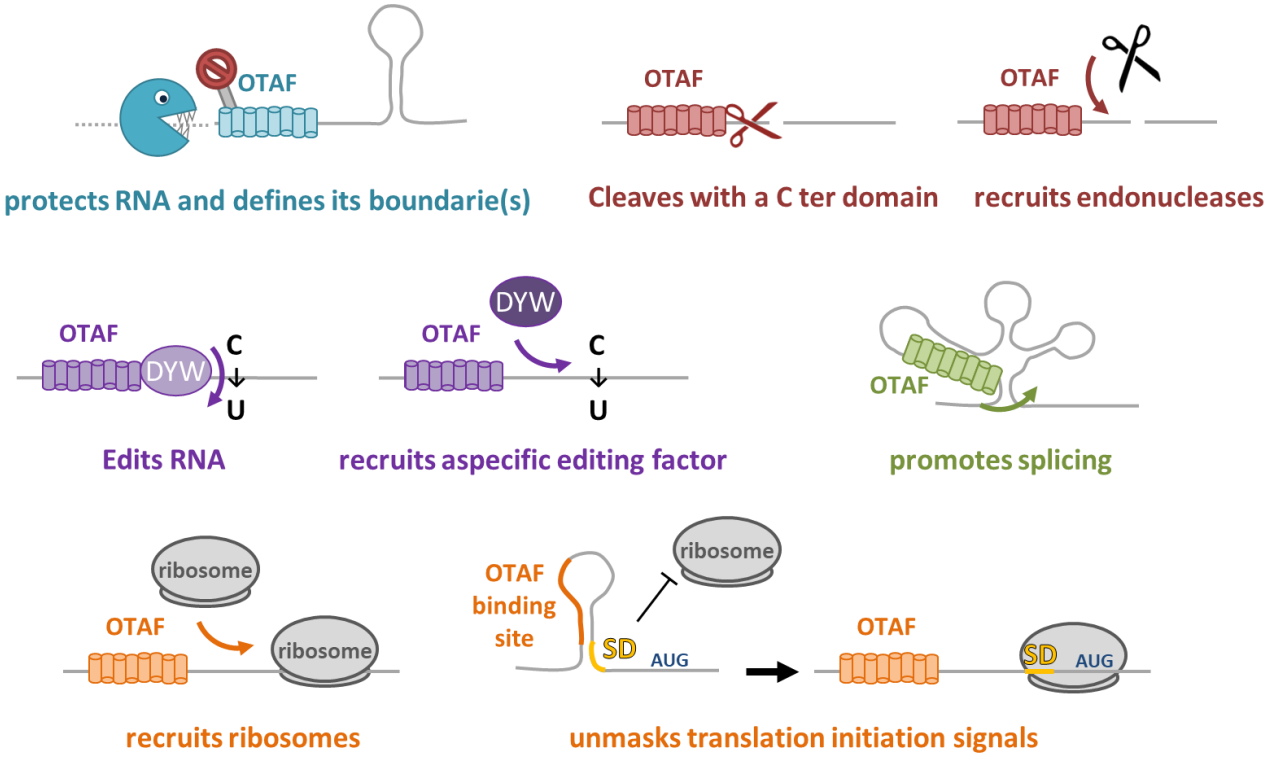


Figure 14: The many roles of OTAF proteins in organelles RNA metabolism.

OTAF PROTEINS BELONG TO MANY INDEPENDENT PROTEIN FAMILIES

OTAF belong to many different protein families of various origins, with different preferred families in different lineages. Many of those families belong to the superfamily of α-solenoid proteins (Kobe and Kajava, 2000), a type of super helical protein formed by tandem repeats of α helices hairpins stacked together and held by van der Waals interactions. Their extended groove can accommodate long biomolecules, like proteins, DNA or RNA strands. Those atypical proteins, contrary to globular proteins, are quite flexible, and can fold and unfold rapidly without dissipating much energy, some of them have even been dubbed biological springs (Kim *et al.*, 2010). These α-solenoid proteins, formed of repeated motifs, can evolve rapidly, because of unequal crossing over, duplication or losses of the similar repeat sequences in the genome. However, repeats of different families are usually not found in a same protein, because of their different stacking properties (Kajava, 1998). And so, residues critical to the overall super helical structure are preserved, while the other can vary because of the large pool of duplicated repeats. This gave rise to the convergent evolution of independent protein families with similar super-structures and physiological roles, which cannot intermingle but evolve dynamically on their own.

I will give a brief overview of some of the OTAF protein families in the next few pages. This super-group embodies the convergence of regulatory mechanisms that arose from endosymbiosis, and the crucial nature of gene expression control from the nucleus to the organelles.

TPR and HAT

TPR are α-solenoid proteins with a defining repeat of 34 amino acids, the tetratricopeptide repeat, folding into two antiparallel α-helices. TPR have mainly been described for their role in protein/protein interactions and are present in all the tree of life (Blatch and Lassle, 1999). But a derivative of the TPR, the HAT (half a tetratricopeptide repeat) had been proposed to bind RNA (Preker and Keller, 1998) and rightly, some of them do. For instance, in the chloroplast of *C. reinhardtii*: MAC1, a factor stabilising *psaC* mRNA (Douchi *et al.*, 2016), NAC2, which stabilises *psbD* mRNA (Boudreau *et al.*, 2000), or MBB1 which stabilises the *psbB* and *psbH* transcripts (Vaistij *et al.*, 2000; Loizeau *et al.*, 2014), and has an orthologue in *Arabidopsis* HCF107 that acts on the same genes (Sane *et al.*, 2005) and is also implicated in translation by uncovering the translation initiation signal of *psbH* (Hammani *et al.*, 2012).

mTERF

The mTERF family, whose first example was a transcription termination factor observed in human mitochondria (Daga *et al.*, 1993), is characterised by a motif of 30 amino acids. Those α-solenoid proteins are found in animals and plants (Roberti *et al.*, 2009). mTERF6 is required for the maturation of *Arabidopsis trnI.2* in the plastid (Romani *et al.*, 2015), mTERF4 for the splicing of type II introns in maize plastids (Hammani and Barkan, 2014).

PPR

The PPR (pentatricopeptide) protein family is a subgroup of OTAF, which was defined twice independently twenty years ago (Aubourg *et al.*, 2000; Small and Peeters, 2000) The PPR family is spread across the tree of eukaryotes but has expanded massively in plants (more than 450 in *Arabidopsis* versus 15 in yeast or 6 in human) probably because the plastome needs many factors to cater to its expression, while other organisms, like metazoans, with only a mitochondria with a small genome, have less organellar genes to regulate. But PPR are also rare in green algae (14 in *Chlamydomonas* for instance; (Tourasse *et al.*, 2013)), where other OTAF families are more abundant. And so, PPR have been extensively studied in land plants (Lurin *et al.*, 2004; Schmitz-Linneweber and Small, 2008; Barkan and Small, 2014). PPR are mostly targeted to the organelles, either plastid or mitochondria, or sometimes both (Colcombet *et al.*, 2013; Lurin *et al.*, 2004).

The PPR proteins bear degenerate 35 amino acids motifs (Figure 17), which fold into two antiparallel α-helices. Those tandem repeats stack into a solenoid super helix, with the first helix of each repeat inside the groove. This motif has been predicted to interact with the mRNA thanks to positive residues forming a continuous surface inside the groove (Small and Peeters, 2000).

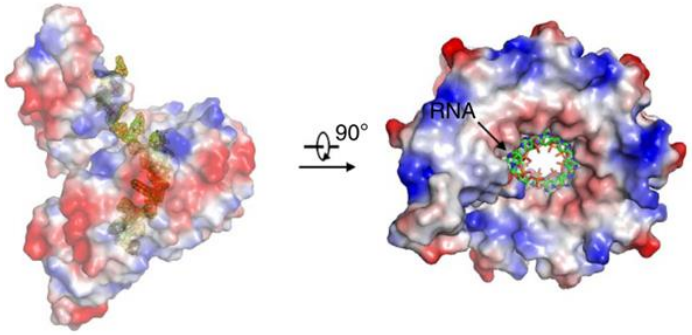


Figure 15: Crystal structure of a designer PPR bound to its cognate RNA, from (Shen *et al.*, 2016), the PPR wraps around the RNA. Red areas are positively charged, the blue ones negatively charged.

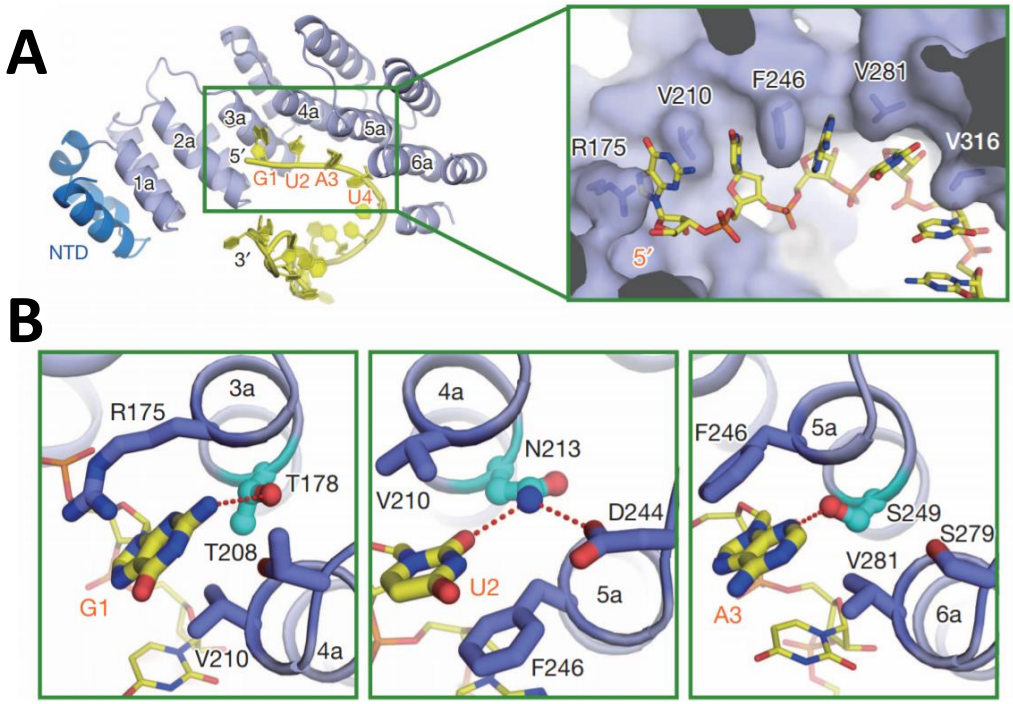


Figure 16: Crystal structure of PPR10 bound to one of its *psaJ* target mRNA, from (Yin *et al.*, 2013). A. The bases of the *psaJ* mRNA are inserted between bulky residues of the internal helices of the PPR repeats. **B.** Interactions of the 5th residue (in cyan) of PPR10 repeats with the bases of *psaJ*. The dotted red lines represent hydrogen bonds.

Further understanding of the mRNA/PPR motif interactions was achieved by molecular, computational and structural studies in the past decade (Prikryl *et al.*, 2011; Yin *et al.*, 2013; Ban *et al.*, 2013; Gully *et al.*, 2015) and a PPR recognition code was established in several studies (Barkan *et al.*, 2012; Kobayashi *et al.*, 2012; Yagi *et al.*, 2013; Kobayashi *et al.*, 2019) and demonstrated *in vitro* on the maize PPR10 protein by recoding it to recognise modified RNAs (Barkan *et al.*, 2012). This PPR code links the nature of residues at specific positions in the PPR repeat with the affinity for a specific nucleotide. Notably residues in position 5 and 35 (6 and 1' following the authors nomenclature) were proved to be essential to recognise specific bases. Crystal structures of this same PPR10 with one of its cognate mRNA *psaJ*, revealed that the bases of the mRNA are "locked" in place by bulky residues at the 2nd position of the PPR motif, and that the 5th residue forms hydrogen bonds with the base (Yin *et al.*, 2013). Electrostatic forces, with succession of positive residues at the 13th position on the PPR attracting the negative RNA, and a succession of negatively charged residues repulsing it, seem crucial to establish the PPR/mRNA interaction (McDermott *et al.*, 2018).

Two main classes of PPR proteins have been defined according to their type of PPR repeats (Figure 17):

- The P-type with proteins containing successive canonical PPR motifs. They are implicated in a variety of physiological role; maturation and stabilisation of mRNA, translation or splicing… around 240 have been found in *A. thaliana*.
- The PLS type with proteins containing a distinctive trio of PPR derived motifs: one canonical **P** repeat, then a **L**ong PPR motif (about 35-36 amino acids), then a **S**hort one (average of 31 amino acids). This class of PPR is implicated in RNA editing. About 200 are present in *A. thaliana*.

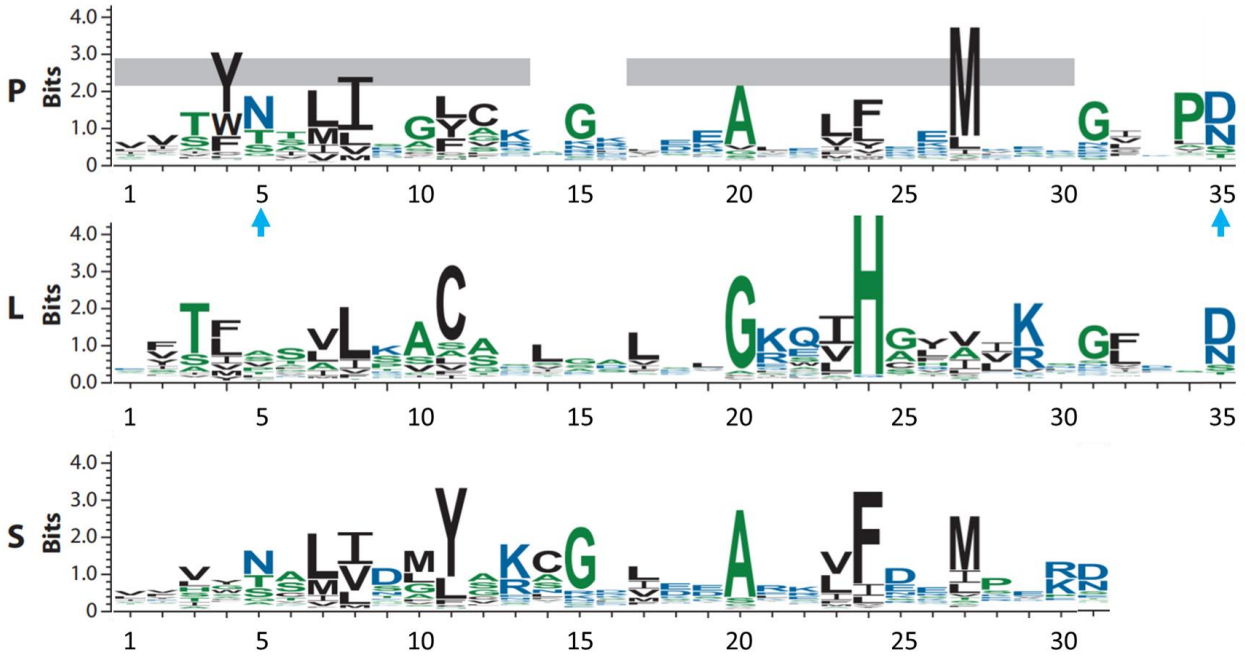


Figure 17: Consensus amino acid sequence of canonical P PPR motif, and its variants: L (long) and S (short). Cyan arrows indicate the key residues for establishing the recognition specificity of the repeat. The grey bars indicate where the two α-helices of the canonical PPR motif lie. Modified from (Barkan and Small, 2014)

Some PPR proteins also have more conserved additional domains in their C terminal region (Aubourg *et al.*, 2000; Lurin *et al.*, 2004; Rivals *et al.*, 2006) . The PLS class PPR have most of the time an E domain, plus often an E^+ one, in turn regularly followed by a DYW domain. The DYW domain (named after its 3 last residues) has a motif homologous to cytidine deaminases, which is necessary for RNA editing (Boussardon *et al.*, 2014). PPR editing factors devoid of DYW domains interact with nonspecific edition factors (Guillaumot *et al.*, 2017). For example, CRR4 interacts with DYW1 (a protein devoid of PPR repeats) to edit *ndhD* in Arabidopsis chloroplast (Boussardon *et al.*, 2012). E (extended) motifs might rather be implicated in nucleotide recognition (Ruwe *et al.*, 2019). A few of P-class PPR protein bear in their C terminus a small MutS-related (SMR) domain, which might have an endonucleolytic activity (Zhou *et al.*, 2017).

PPR proteins are implicated in all steps of organellar RNA expression, to cite only a few examples; transcription (Ikeda and Gray, 1999) stabilisation and maturation (Meierhoff *et al.*, 2003; Loiselay *et al.*, 2008; Beick *et al.*, 2008; Pfalz *et al.*, 2009; Johnson *et al.*, 2010), editing (Kotera *et al.*, 2005; Okuda and Shikanai, 2012), splicing (Williams-Carrier *et al.*, 2008; Wang *et al.*, 2018), translation initiation (Prikryl *et al.*, 2011) and cleavage (Zhou *et al.*, 2017).

80% of PPR genes in Arabidopsis do not have introns (Lurin *et al.*, 2004), and so it is probable that they were transposed and spread around the genome by a retro-transposition mechanism. While most PPR proteins are quite conserved across terrestrial plants species, where they accomplish crucial roles on the often conserved target RNAs, clusters of PPR proteins that are highly variable between species have been described. Those clusters of paralogous PPR genes, stemming from duplications, undergo diversifying selection across species and thus gain new RNA targets (Fujii *et al.*, 2011). Those restorers of fertility like PPR (RFL-PPR), in part counteract mitochondrial RNAs causing cytoplasmic male sterility in plants but can also play subtle roles in other mitochondrial mRNAs maturation (Dahan and Mireau, 2013). Altogether, this fast-evolving subgroup of PPR forms a reservoir of mitochondrial factors that can duplicate and acquire random new targets to respond to new CMS developed by the mitochondria, in a form of “arm-race”.

4. THE OPR: A SUBFAMILY OF OTAF

As we just saw, PPR are quite scarce in chlorophytes (only 14 PPR in *C. reinhardtii*). While algae do not edit their organellar transcripts, dispensing for the need of specific editing factors, their chloroplast genomes are of the same length and complexity that those of terrestrial plants. About a hundred genes must be controlled in the chloroplast but only a few in the mitochondria. To do so Chlorophyceae rely on another family of OTAF factors: the octotricopeptide repeat (OPR) proteins.

The octotricopeptide repeat is a degenerated motif of around 38 amino acids that folds in a pair of antiparallel α-helices in a structure reminiscent of PPR (Figure 18), as confirmed by the recent structure of the OPR protein ASA2 (see below). OPR repeats, when stacked together, form an α solenoid structure. However, PPR and OPR seem to have arisen from different origins separately by convergent evolution, as no similarity can be detected between the two motifs. In contrast to PPR that are highly abundant in land plants, OPR are mostly found in Chlorophyceae, and are very scarce in plants: only one lone OPR is found in *Arabidopsis, Physcomitrella,* rice or maize (Kleinknecht *et al.*, 2014). This suggest that both OPR and PPR were present in the Viridiplantae ancestor and that one of the two groups subsequently exploded independently in streptophytes and chlorophytes to fill the same roles of organelles control.

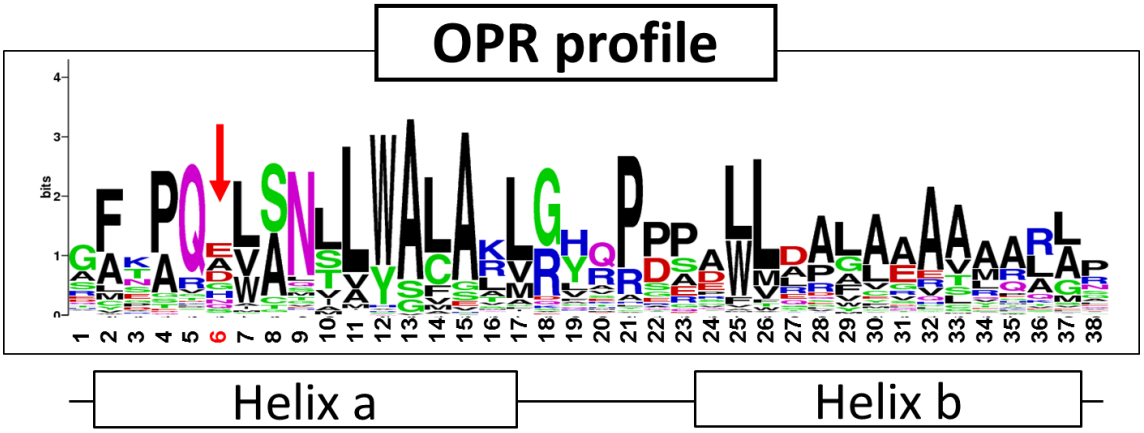


Figure 18: The consensus sequence of the OPR repeats found in photosynthetic organisms. The taller the residue, the most abundant it is. The positions of the two putative α-helices are indicated under the consensus. The red arrow indicates the 6th amino acid, which is expected to be crucial for the specific interaction with nucleotides of the target mRNA.

The PPPEW conserved motif described in (Eberhard *et al.*, 2011; Rahire *et al.*, 2012), based on subsets of respectively 42 and 43 OPR proteins of *Chlamydomonas* is less prominent in our LOGO stemming from a larger OPR proteins sample (originating both from *C. reinhardtii* and other organisms). Our definition of the OPR repeat instead of only 43 proteins, recover 127 in *C. reinhardtii.* In this LOGO drawn from a larger set of OPR repeats, another motif seems more conspicuous than PPPEW, the LWALA at the end of the first α helix.

The OPR families defined in (Eberhard *et al.*, 2011; Rahire *et al.*, 2012) were based on and contain mostly T and splicing factors. The difference with our newer LOGO could originate from differences between the T factors, that seem to interact briefly with RNA, and the numerous M factors included in our sample, which bind stably on RNA. It is conceivable that this difference in function results from variations in the first helix, which is believed to interact with the nucleotide (Figure 19).

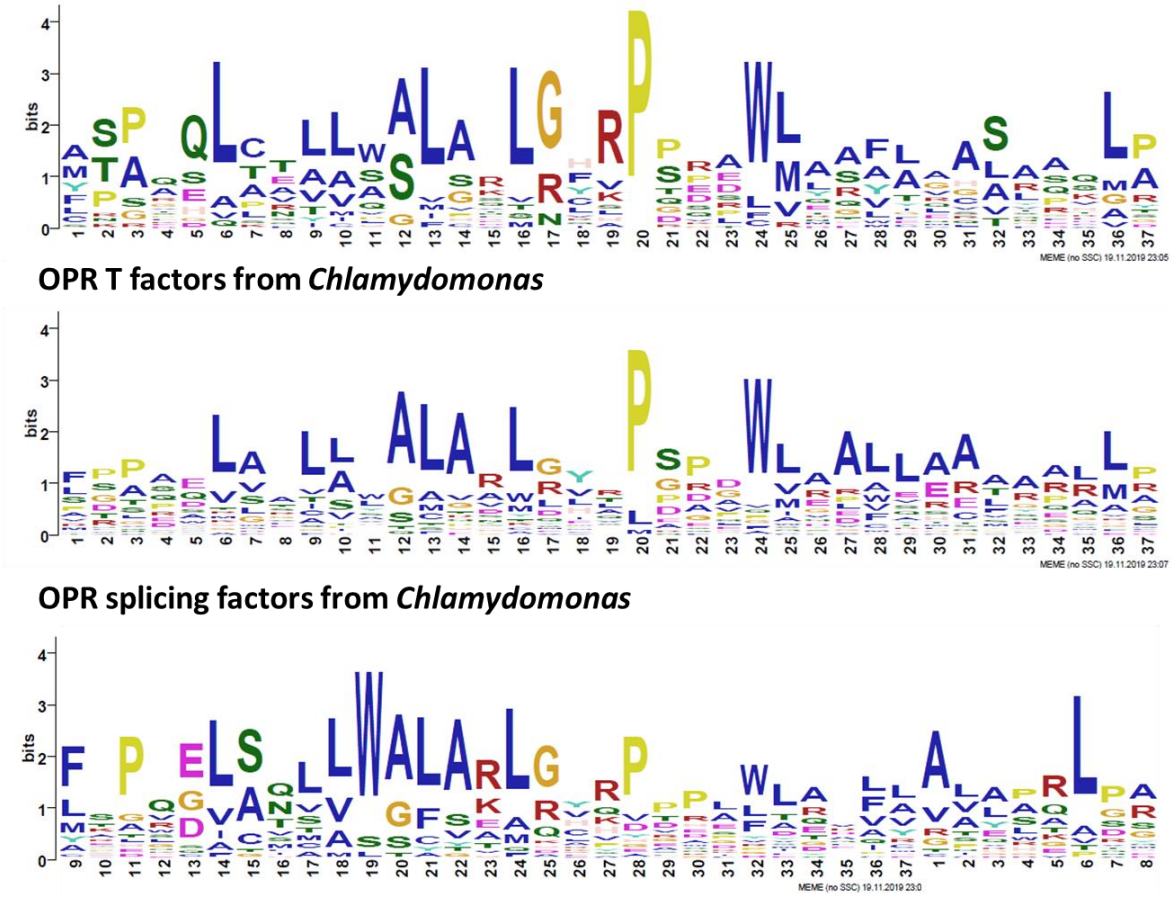


Figure 19: LOGOs of OPR repeats from different functional factors in *Chlamydomonas reinhardtii* show a divergence in sequence. The LOGOs were obtained with MEME suite (Bailey *et al.*, 2009), note that MEME finds motifs of 37 amino acids, with an offset of several residues. The M factor LOGO was adjusted to align with the T and splicing factors LOGOs.

As PPR proteins (Lipinski *et al.*, 2011), OPR proteins evolve rapidly in different clades. The LOGO in Figure 18 was mostly drawn from photosynthetic organisms but may differ in others. The two LOGOs in Figure 20 illustrate this divergence between clades, even if both are photosynthetic: in diatoms for instance the LWALA motif becomes AWAFA, and the tryptophan in the 29th position is replaced by another aromatic amino acid: phenylalanine. Thus, one can consider the OPR proteins, the HeptatricoPeptide Repeat (HPR) proteins, recently identified in *Plasmodium* (Hillebrand *et al.*, 2018), and the human FASTK proteins (Boehm *et al.*, 2017), as distant members of a same polymorphic family. Indeed, using our OPR motif to scan the HPR proteins found in *Chlamydomonas* we recover nearly all of them (33 out of 36) as OPR proteins (see ***Annex 4***). In addition, this extended OPR/HPR/FASTK family is characterised by the possible addition at the C-terminus of a RNA Binding Abundant in Apicomplexa (RAP) domain, possibly implicated in RNA cleavage.

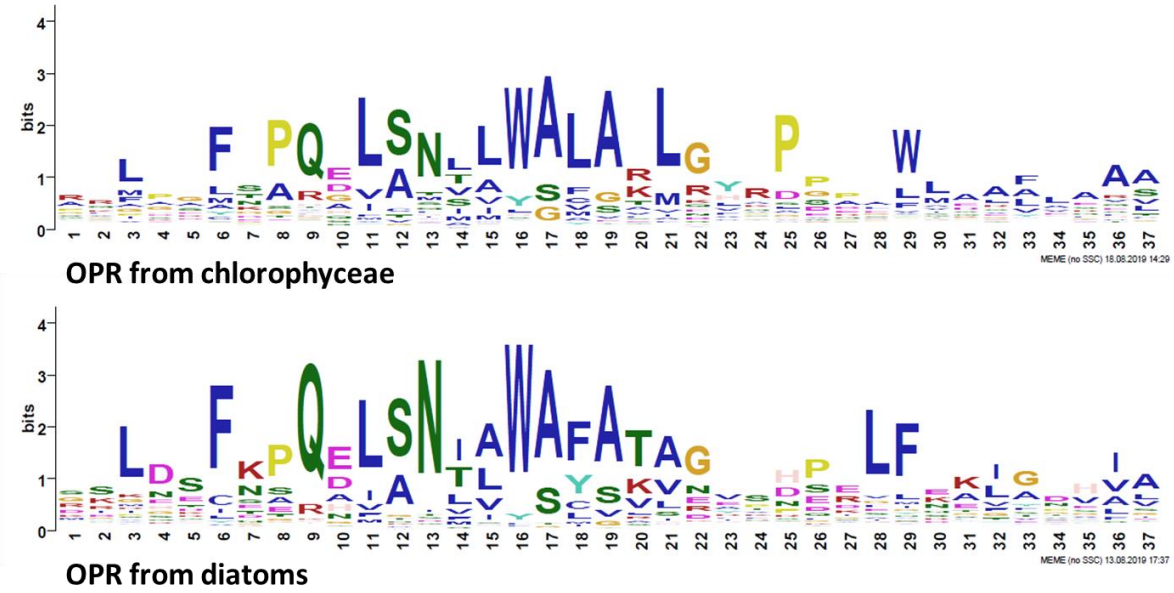


Figure 20: LOGOs of OPR repeats in Chlorophyceae and diatoms differs. The LOGOs were obtained with MEME suite (Bailey *et al.*, 2009), note that MEME finds motifs of 37 amino acids, with an offset of several residues.

OPR proteins have mostly been studied in *C. reinhardtii* where they thrived, with more than 120 OPR factors. Much like other OTAF, they are implicated in every possible post-transcriptional step of organelle gene expression, such as RNA stabilisation and maturation (Drager *et al.*, 1998; Murakami *et al.*, 2005; Wang *et al.*, 2015; Viola *et al.*, 2019; Cavaiuolo *et al.*, 2017) (***ARTICLE 1*** and ***ARTICLE 3***), splicing (Rivier *et al.*, 2001; Merendino *et al.*, 2006; Balczun *et al.*, 2006; Jacobs *et al.*, 2013; Marx *et al.*, 2015), translation initiation (Zerges and Rochaix, 1994; Stampacchia *et al.*, 1997; Auchincloss *et al.*, 2002; Eberhard *et al.*, 2011; Rahire *et al.*, 2012; Lefebvre-Legendre *et al.*, 2015; Cline *et al.*, 2017), and possibly RNA degradation as well (Drapier, 2002; Boulouis *et al.*, 2015).

Recently a single particle cryo-electron microscopy structure at a resolution of 2.7 to 2.8 Å of the ASA2 (**A**TP **S**ynthase **A**ssociated) OPR protein, containing 8 OPR repeats and a degenerated RAP domain, confirmed that OPR repeats fold into helical hairpins that stack on each other to form a "half a donut" structure (Murphy *et al.*, 2019) (Figure 21).

The first helix of the OPR repeat lays in the groove of the super-helix. This first helix is more conserved than the second one, hinting that it is more important for the function of the motif (Figure 18). However, specific positions in this conserved helix are highly variable. Those variable residues, notably the sixth one, protrude at the inner face of the groove and are predicted to be crucial for the specific recognition of a given nucleotide.

However, ASA2 is a peculiar OPR that does not bind RNA whose position is occupied by the N-terminal arm of the ASA7 protein. It is associated with *Polytomella* sp. mitochondrial ATP synthase, as part of the atypical stator-stalk found in the mitochondria of Chlorophyceae (Vazquez-Acevedo *et al.*, 2006). Interestingly a PPR protein is associated with the CF1 of *Trypanosoma brucei* mitochondrial ATP synthase (Montgomery *et al.*, 2018). Whether this parallel just stems from random conversion

of available organelles proteins to assume other functions or whether α-solenoid proteins bring useful structural properties to the ATP synthase complex remain to be elucidated.

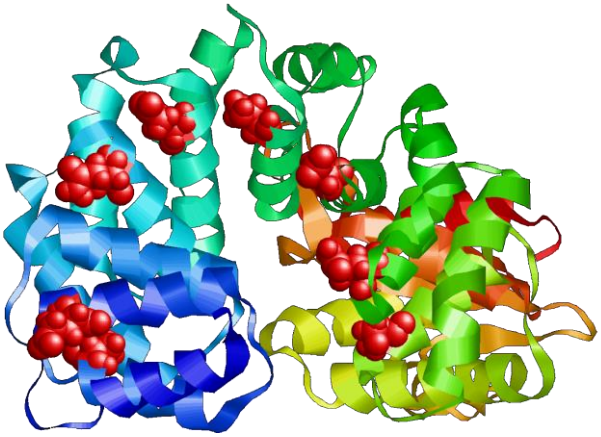


Figure 21: Cryo-electron microscopy structure of ASA2 from (Murphy *et al.*, 2019). The 6th residue of each OPR motif is indicated in red. Note the helical shape of the protein.

A preliminary recognition code (Table 1) based on confirmed OPR proteins/mRNA pairs has been drafted by Yves Choquet but remains putative and incomplete (as we will see from **Chapter III**). OPR proteins have been less studied than PPR proteins so far, and so, only a few of them have been characterised. Consequently, few validated OPR/mRNA pairs are available, making a recognition code more difficult to establish. Conversely, the absence of a confirmed molecular recognition code precludes the prediction of potential targets for many cryptic OPRs and hamper their study.

Position	Residue										
3	X	X	X	X	R, K	X	X	X	X	X	X
4	X	P	X	X	X - P	X	X	X	X	X	X
5	X	Q	X - R, K	R, K	X	R	X - R	R, Q	X	R	R
6	E	G	D	D	D	Q	Q	A	H	S	N
Recognised nucleotide	U	A	G	U	U	U	?	A	?	A	?

Table 1: The draft OPR recognition code, established by Yves Choquet.

Similarly to PPR proteins, most OPR proteins have conserved orthologues across chlorophyceae (or at least within the same order) and act on conserved sequences within their target mRNA (see alignments in ***Articles 1 and 3***).

An intriguing subfamily of OPR-RAP proteins dubbed the NCL (NCC-like) was found in C. reinhardtii. Two dominant mutants, *ncc1* and *ncc2* (Drapier, 2002; Boulouis *et al.*, 2015) with point mutations in the sixth OPR repeats of NCC1 and NCC2, respectively induced the degradation of the *atpA* and *petA* transcripts. This degradation signal is borne by two specific small RNA sequences in *atpA* or *petA* CDS, respectively. The mutations in those two OPR proteins are expected to modify their specificity and thus provide them these two new targets.

Indeed, the point mutation in NCC1 changes D_{6th} to A_{6th} in its sixth OPR motif and is expected to modify the nucleotide recognised by this OPR motif. The point mutation in NCC2 changes S_{8th} to R_{8th} in its sixth OPR motif and somehow appears to change the recognition of the OPR protein.

NCC1 and NCC2 belong to a family of 38 paralogues, the NCL proteins for NCC-Like, stemming from local gene duplication. Most of them lie together in a cluster on chromosome 15. Similar NCL clusters were found in *Chlamydomonas debaryana* (17), *Chlamydomonas asymetrica* (5), *Chlamydomonas spheroides* (20), *Tetrabaena socialis* (34) and *Gonium pectorale* (9) but not in some other closely related species such as *Volvox carteri*, *Yamagishiella unicocca, Eudorina species* or *Chlamydomonas applanata*. This suggests either a recent formation of those clusters or a secondary loss in some organisms. Interestingly, it appears that the more the organism displays a "social" organisation, with progressively more complex colonies, the more the NCL cluster tends to shrink and finally disappears as in *Volvox carteri, Eudorina species,* or *Yamagishella unicocca.* This might be a coincidence, but the OPR-RAP could perhaps become dispensable upon transition to multicellularity (Figure 22).

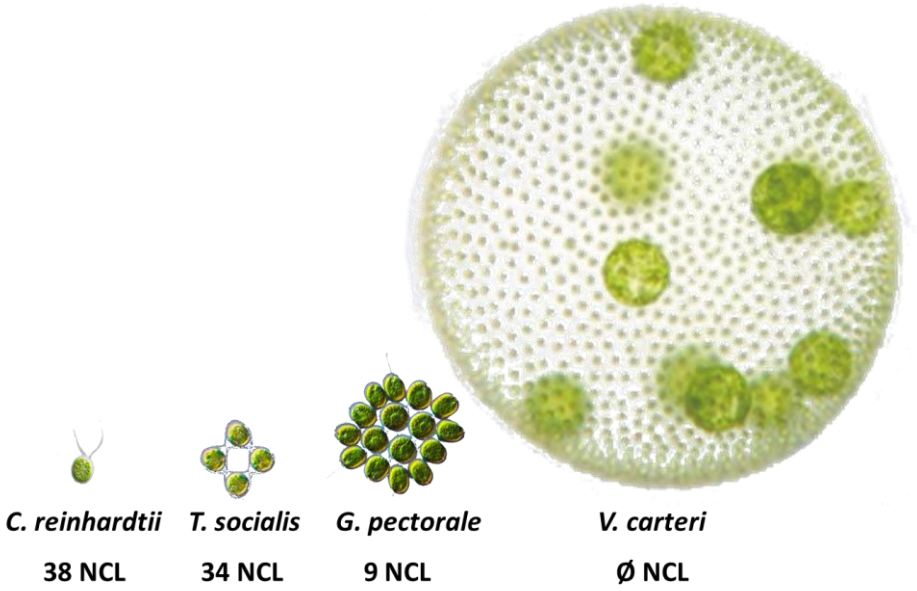


Figure 22: There could be an inverse correlation between the number of NCL proteins and the transition to multicellularity.

The function of the NCL family remains cryptic. It is likely, like the RFL-PPR, a reservoir under diversifying selection pressure for evolving OPR factors with new targets. Their association with a RAP domain suggests that they might directly cleave mRNA. Perhaps could they be involved in pathogen resistance by destroying intruding RNAs?

Many aspects of the OPR protein family function remain understudied for now. This family holds certainly many more secrets. In this thesis, I contributed to the growing knowledge of OPR proteins, by studying the physiological roles of some of them and by grappling with their specific RNA binding activity to try to understand how it is established.

5. The ATP Synthase

The ATP synthase is a very complex protein machinery that is ubiquitous across the tree of life. It is thought to have been present in LUCA (Weiss *et al.*, 2016). How such a complicated molecular motor, with its many subunits, came to be so early in evolution remains mysterious. But leaving aside its origins, the ATP synthase is an invaluable complex for life. Embedded in a membrane this motor acts as a selective channel to dissipate H^+ or Na^+ gradients while generating ATP, the major chemical energy source for most biological processes. The proton or sodium motive force creates a rotation of the ATP synthase "rotor" (Junge *et al.*, 1997). The central stalk, part of the rotor, then rotates into the catalytic head of the CF1 (Sabbert *et al.*, 1996), at the exterior of the membrane, inducing conformational changes in the α and β subunits that catalyse the production of ATP from ADP and inorganic PO_4^{2-} (Elston *et al.*, 1998). A peripheral stalk stabilises the stator by connecting AtpI to the catalytic head of CF1, thereby preventing fruitless rotation of the CF1. The ATP synthase can also work in reverse and consume ATP to equilibrate cations gradients.

In eukaryotes, the ATP synthase is present in mitochondria as in the plastids of photosynthetic organisms. In mitochondria, this molecular motor forms quite stiff dimers that induce folds in the mitochondrial membrane (Giraud *et al.*, 2002; Dudkina *et al.*, 2005). Interestingly, the stator-stalks of mitochondrial ATP synthase from land plant and green algae are very different (Vazquez-Acevedo *et al.*, 2006; Murphy *et al.*, 2019), the chlorophycean one includes notably an OPR! It seems that available proteins imported to the mitochondria could see their role tweaked to integrate protein complexes, and that this phenomenon is quite adaptive; since the mitochondria is an ancient endosymbiont it is fascinating to witness such differences in an absolutely key protein complex.

In the chloroplast, ATP synthases are monomeric and flexible (Hahn *et al.*, 2018; Kuhlbrandt, 2019). In spinach kept in the dark, the proton gradient is null across the thylakoid membrane and the ATP synthases are blocked by a redox control switch in their central stalk. Thus, they do not consume ATP to pump proton back into the lumen (Junesch and Gräber, 1987; Hahn *et al.*, 2018). This is an efficient mechanism to prevent wasteful hydrolysis of ATP at night.

Another intriguing property of the ATP synthase is how the number of subunits c (AtpH in *C. reinhardtii*) varies a lot across different phyla. The bovine mitochondrial ATP synthase has a ring of 8 c (c_8). The mitochondrial ATP synthase from *Saccharomyces cerevisiae* has a c_{10} ring. The ATP synthases from *Ilyobacter tartaricus* have a c_{11} ring, from *Caldalkalibacillus thermarum* a c_{13} ring, from spinach chloroplast a c_{14} ring and from the cyanobacterium *Spirulina platensis* a c_{15} ring. The biggest found so far is a c_{17} ring in

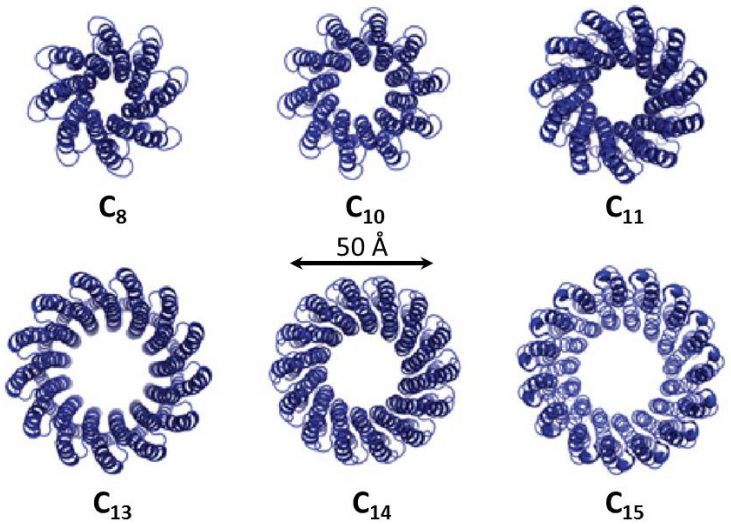


Figure 23: Structures of C-rings in the rotor of varuous ATP synthases, from (Walker, 2013)

an ATP synthase of *Burkholderia pseudomallei* (Schulz et al., 2017), a human pathogen. As a c-ring needs as many H^+ (or Na^+) as it contains subunits to make a full rotation, and since a 360° rotation is needed for CF1 to produce 3 ATP, the cation cost to produce one ATP is the number of c-subunits divided by 3 (Walker, 2013). The more c subunits, the less ATP is produced per cation. In other terms, bigger barrels should be less efficient but are strangely widespread in bacteria and chloroplasts.
The number of subunits in the c-ring could instead be linked to other functions than primarily producing ATP, for optimal growth or to adapt to the physiological environment (Kuhlbrandt, 2019). The c_{17} ring of *B. pseudomallei* for instance, could allow the ATP synthase to function as a high throughput proton pump by consuming ATP, to help this pathogen cope inside the acidic phagosome of macrophages (Schulz et al., 2017). Na^+ ATP synthases could also be used to endure salt stress.

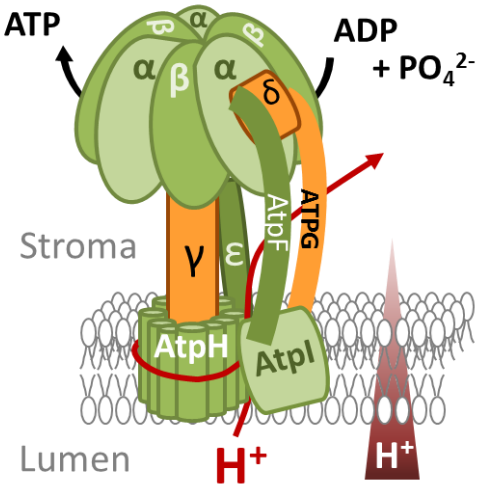


Figure 24: C. reinhardtii ATP synthase

In *C. reinhardtii* the chloroplast ATP synthase is composed of: CF1: 3α: 3β: 1γ: 1δ: 1ε and CFo: 14 AtpH: 1 AtpI: 1 AtpF: 1 ATPG. As for all photosynthetic complexes, the subunits are encoded in part in the nucleus genome, while others have been retained in the chloroplast one (Figure 24).

The CFo contains 14 AtpH subunits and 1 AtpI, their interface constitutes the two hemi-channels of the proton channel, one open on the lumenal side and one on the stromal side. After entering from the lumen, the proton is accepted by a glutamic acid that does a near complete rotation along the AtpH barrel to reach the exit channel. The high pH of the stroma deprotonates the glutamic acid, the proton is released.

The catalytic "head" of the ATP synthase is formed by 3 α and 3 β subunits, they are similar in their structure and both have a central nucleotide-binding domain. The β subunits are pushed by the asymmetrical γ stalk as it rotates and undergo conformational changes that change their catalytic properties. They go thru three states: an open one where the newly produced ATP is released, a loose state allowing the binding of ADP and Pi, and a tight one where ATP is formed from ADP and Pi. The α subunits do not exhibit an open state, and are not thought to directly catalyse ATP formation (Walker, 2013).

Aside from those mechanical properties, it is interesting to note that those subunits: AtpH, AtpI, α and β, are influenced by the redox potential and are still encoded in the chloroplast, following CoRR hypothesis. But in contrast, γ, which exhibit a redox-switch property to stop the ATP synthase at night is now encoded in the nucleus, seemingly in contradiction with CoRR theory. Another sticking point is that while AtpH and AtpI are also both hydrophobic transmembrane proteins, following the difficulty of import hypothesis in *C. reinhardtii*, in some species *atpI* has been transferred to the nucleus… Altogether, explaining why some subunits are now expressed in the nucleus while other stayed anchored in the chloroplast genome is quite a difficult task.

6. MAIN OBJECTIVES OF THIS THESIS:

While I also contributed to the study of mutants not affected in OPR factors (see ***ARTICLE 2***), the bulk of my work has been the study of OPR proteins in *C. reinhardtii*, to better grasp their biological roles and their RNA-binding activities. To do so, I focused on two OPR factors implicated in the expression of the chloroplast ATP synthase: MDB1 and MTHI1.

As we saw previously, the OPR protein family has not been studied as extensively as the PPR protein family and many aspects of their functions remain shrouded in mystery. The range of their physiological roles has yet to be completely explored. This thesis aimed to improve our knowledge and broaden our understanding of the **physiological functions of those OPR factors**.

- **In the first chapter, I will develop the functional study of the M factor: MDB1 and its chloroplast target mRNA *atpB*.**
- **In the second chapter, I will describe the study of MTHI1, a peculiar OPR protein targeting two mRNAs, *atpH* and *atpI*, with, for both genes, a dual role in mRNA stabilisation and translation initiation.**

Those parts focusing on the physiological roles of M factors revealed unsuspected properties of the OPR protein functions in the chloroplast of *Chlamydomonas reinhardtii*, which led us to start reconsidering our previous theories on those proteins.

Considering the OPR potential as modular RNA binding proteins, the OPR repeat is a motif of great interest to design or modify proteins. It could be used for the production of designer OPR proteins to bind desired targets, as has begun with PPR (Filipovska and Rackham, 2013; Coquille *et al.*, 2014; Yagi *et al.*, 2014; Shen *et al.*, 2016; Spahr *et al.*, 2018; Yu *et al.*, 2019; Rojas *et al.*, 2019). But to capitalise on this flexible motif, it is crucial to first uncover its primary property: its capacity to recognise specific nucleotides. This ability might follow a recognition code and a hypothetic one was drafted prior to this thesis.

This preliminary "OPR code" remains putative and uncomplete; no target prediction is possible for T factors; whose mode of action remains quite nebulous. Indeed, T factors seem to interact only transiently with their target mRNA and their footprints are not easily recoverable. Moreover, they tend to show a bias for OPR motifs that are not widespread in the better-characterised M factors. This creates a situation where the preferred OPR repeats of T factors stay "unreadable". Similarly, for the NCL family whose function might be subtle we can barely predict a target for a few of them.

Altogether, the study of the molecular recognition code of OPR proteins would be greatly beneficial to improve our understanding of OPR factors, and to allow the study of the many uncharacterised OPR proteins of *Chlamydomonas*, or other organisms. In addition, it would pave the way for designing OPR proteins. The main objective that I pursued in the three years of this thesis was to **decipher this potential "OPR code" *in vivo***.

- **In the third chapter, I will present our initial work on the OPR/RNA interaction *in vivo*, and how it turned on its head our view of M factors.**
- **In the fourth chapter, I will explain and detail the system I developed *in vivo* to "crack" the OPR recognition code.**

My thesis underpins how the functional landscape of OPR protein *in vivo* is far more complex than anticipated. And was a watershed point in our understanding of OTAF in *C. reinhardtii* chloroplast.

I. FUNCTIONAL STUDIES OF THE MDB1/*ATPB* EXPRESSION SYSTEM

Modified from Dartmouth Electron Microscope Facility, Dartmouth College

atpB

atpB is the chloroplast gene encoding the β subunit of the plastid ATP synthase (Figure 25). In *C. reinhardtii*, unlike in plants, *atpB* is not co-expressed with *atpE* in a polycistron, *atpE* is more than 100 kb away, sitting at the opposite side of the plastid genome. *atpB* has its own promotor and is expressed as a monocistron. The two genes are controlled independently. Like most plastid genes *atpB* does not have any intron, simply a CDS of 1446 nt. A secondary structure in *atpB* 3'UTR is implicated in transcript stability and 3' end maturation but not in transcription efficiency (Stern *et al.*, 1991). 3' end maturation of *atpB* transcript also appears to stimulate its translation (Rott *et al.*, 1998a), a topic that we will further explore in this chapter.

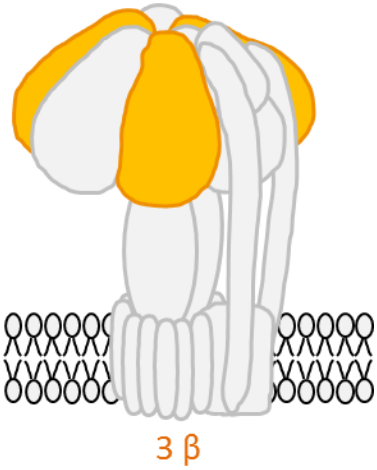


Figure 25: The β subunits of the chloroplast ATP synthase.

atpB 3'end maturation proceeds by two successive steps; first an endonucleolytic cleavage 10 nt downstream of the secondary structure, then the trimming of the remaining 10 nucleotides in a 3'→5' degradation. (Stern and Kindle, 1993), meanwhile the residual transcript fragment cleaved downstream of the endonucleolytic cleavage site is rapidly degraded by 5'→3'exonucleases (Hicks *et al.*, 2002). This 3' UTR secondary structure defines the 3' boundary of *atpB* mature mRNA, but as we saw previously (Introduction p33) the 5' end of *C. reinhardtii* plastid mRNAs is defined by specific factors, that bind and protect the mRNA from 5'→3' exonucleases. In ***ARTICLE 1*** (attached at the end of this manuscript) we studied this specific factor and the maturation process of the *atpB* 5' end.

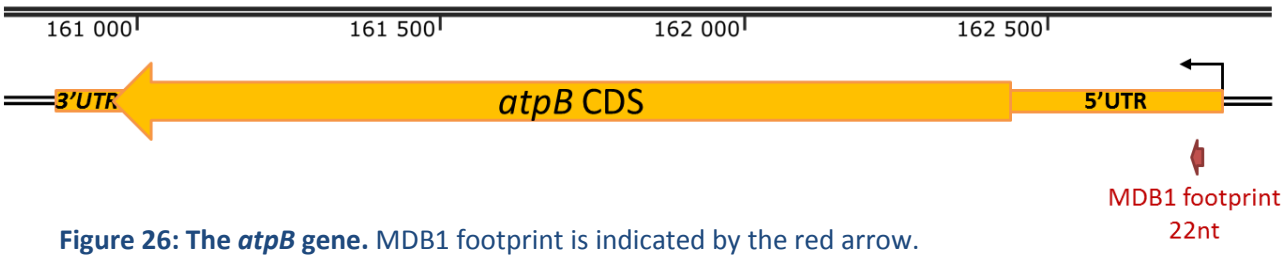


Figure 26: The *atpB* gene. MDB1 footprint is indicated by the red arrow.

MDB1

A nuclear mutant: *thm24,* isolated by Schmidt in 1977 was later identified as a chloroplastic ATP synthase mutant (Piccioni *et al.*, 1981). It was found to translate neither of the chloroplastic ATP synthase subunits α (AtpA) or β (AtpB) (Lemaire and Wollman, 1989). Further studies revealed that in this mutant no *atpB* mRNA accumulated (Drapier *et al.*, 1992) and that the defective translation of subunit α was due to a CES control by subunit β of subunit α synthesis (Drapier *et al.*, 2007): in absence of the *atpB* mRNA, subunit β is not synthesised and cannot activate the translation of the *atpA* mRNA.

Another mutant, obtained by Laura Houille in an insertional mutagenesis campaign (Houille-Vernes *et al.*, 2011), called *L35a*, was found to have the same phenotype, as we will see in ***ARTICLE 1***.

Both *thm24* and *L35a* are affected in the same gene identified by whole genome sequencing, which encodes an OPR protein, and was named *MDB1* (**M**aturation-stabilisation of complex **D** (ATP synthase) *atp**B*** transcript) (Figure 27). *thm24* displays a single deletion of one A in MDB1 exon 5 that transforms a *Bst*XI restriction site into a *Bsr*I site and leads to the formation of a premature STOP codon and to translation abortion. This mutant will be dubbed *mdb1-1* in the following manuscript. *L35a*, bears a large deletion of 30 kb encompassing *MDB1* and six other genes, *L35a* will be named *mdb1-2*.

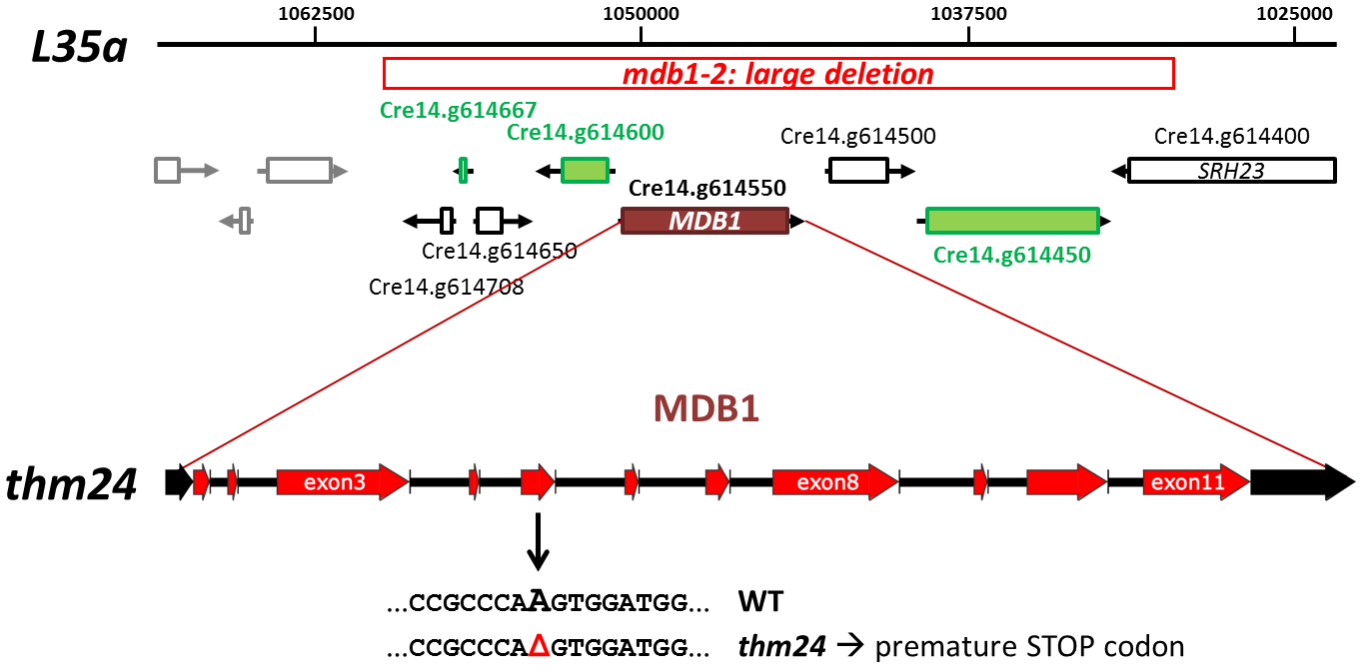


Figure 27: The *L35a* (*mdb1-2*) and *thm24* (*mdb1-1*) mutants

MDB1 (Cre14.g614550) is a nuclear gene of 7127 nt on chromosome 14 (Figure 28.A). It contains 10 introns and encodes a protein of 1137 amino acids, which is predicted to be addressed to the chloroplast by the Wolf PSort (Horton *et al.*, 2007), Predotar (Small *et al.*, 2004) and ChloroP (Emanuelsson *et al.*, 1999) programs. This protein has disordered regions in its N terminus and C terminus and 13 OPR motifs in its central region. The OPR motifs are separated in two groups of seven and six continuous repeats by a "hinge" of three α-helices (Figure 28.B). We suspect that this "hinge" could relieve tensions exerted when binding on the cognate *atpB* mRNA. Indeed, interaction of a long continuous PPR protein with mRNA induces a contraction of the PPR protein helical structure (Shen *et al.*, 2016) that probably creates a strain on the mRNA. Addition of "breathing" structures between extended helical repeat stacks might help alleviate those tensions, or inversely some nucleotides in the centre of the target could also be "ignored" by the repeats, to loosen the interaction in the middle, like might happen with PPR10.

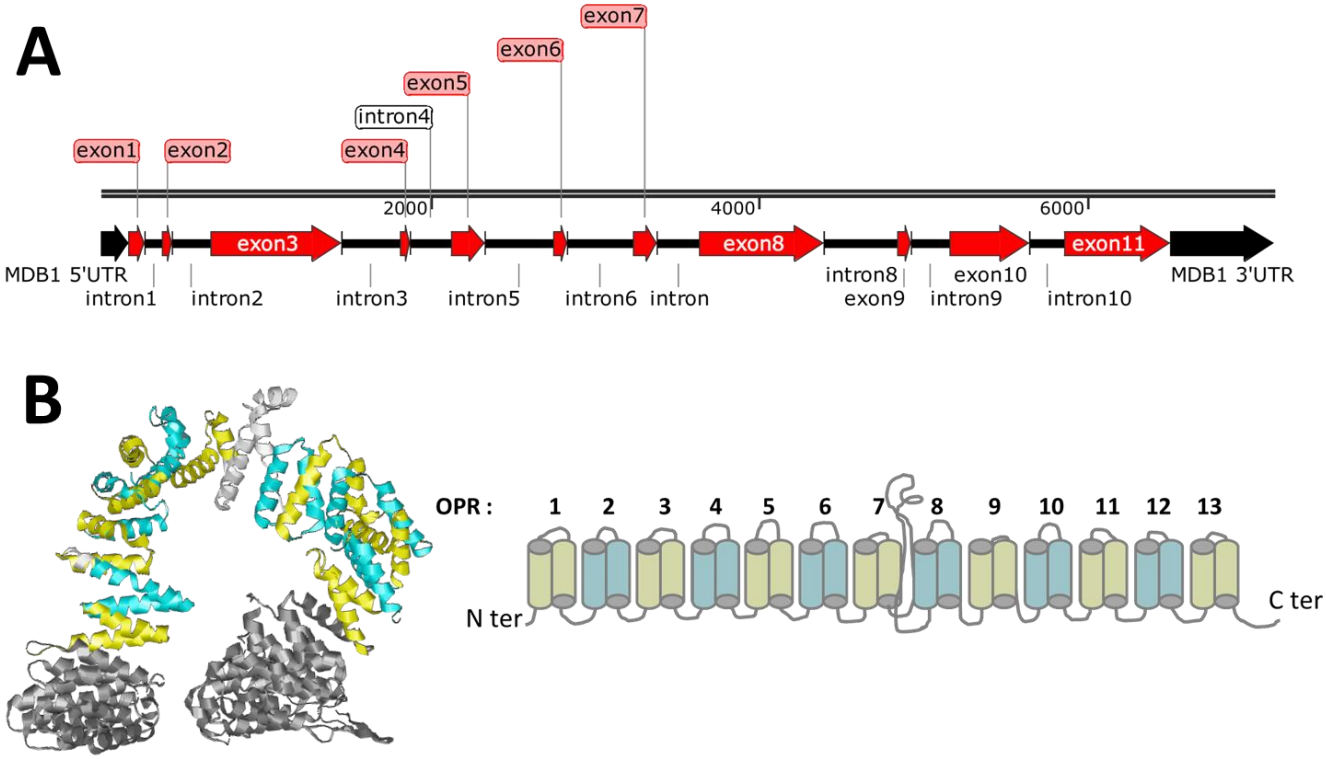


Figure 28: A. *MDB1* gene on chromosome 14. B. Right: a prediction of MDB1 structure by the program I-Tasser, Left: A cartoon of MDB1 OPR domains.

Crosses between *mdb1-1* and strains carrying chimeric transcripts under the control of the *atpB* 5'UTR, revealed that this 5'UTR was sufficient to induce an MDB1-dependant accumulation of chimeric mRNAs. Moreover, immunoprecipitation proved that tagged MDB1 interacts specifically, indirectly or directly, with the *atpB* mRNA 5' UTR (figure 2 from ***ARTICLE 1***). The addition of a polyG track a sequence of 18 successive G, that forms a strong and voluminous secondary structures in RNA, effectively blocking the path of exonucleases, (Drager *et al.*, 1998) in *atpB* 5'UTR, rescued the transcript accumulation in absence of MDB1, in the *mdb1-1* mutant (figure 4 from ***ARTICLE 1***).

Chloroplast small RNA sequence data has been used previously to look for the accumulation of footprints, short nucleotide sequences protected from degradation by a stably bound protein, to identify the putative sites of RNA binding proteins (Ruwe and Schmitz-Linneweber, 2012; Loizeau *et al.*, 2014; Cavaiuolo *et al.*, 2017). Sequencing of natural small RNAs in WT or *mdb1-2* performed by Marina Cavaiuolo revealed the accumulation of a 22 nt MDB1 dependant footprint in the *atpB* mRNA 5'UTR, one nucleotide downstream of the matured 5'extremity (Cavaiuolo *et al.*, 2017) (figure 3 from ***ARTICLE 1***). This footprint comprises a sequence essential to *atpB* 5'UTR-driven mRNA stability (Blowers *et al.*, 1990; Anthonisen *et al.*, 2001). Considering that MDB1 has 13 OPR repeats, we assume its target sequence to be 13 nt long, it encompasses a sequence defined by Anthonisen and colleagues as essential for the stable accumulation of the transcript (Anthonisen *et al.*, 2001). I have studied the binding sequence of MDB1 in great details as we will see later (Chapters III and IV).

My main contributions to the studies presented in ***ARTICLE 1*** (attached at the end of this manuscript) have focused on the maturation and degradation dynamics of *atpB* mRNA and the impact of 5' end maturation on protein expression.

RESULTS

atpB mRNA stability and maturation

Addition of a poly G track in *atpB* 5'UTR recovers another shorter form of *atpB* transcript, even when MDB1 is present. This shorter form is more abundant than the *atpB* mRNA in the wild type. This suggests that much of *atpB* mRNA is still exposed to 5'→3' exonucleases when its M factor is normally expressed. Thus, *atpB* mRNA is transcribed in excess amounts and MDB1 is a limiting factor for *atpB* mRNA accumulation (see figure 4 from ***Article 1***). Unprotected *atpB* transcripts should be rapidly degraded.

In addition, the *atpB* transcript accumulates in two forms in the chloroplast: a precursor one, from the transcription start-site, tri-phosphorylated, and an abundant mono-phosphorylated one, trimmed by 27 nucleotides (Woessner *et al.*, 1986; Blowers *et al.*, 1990; Anthonisen *et al.*, 2001; Cavaiuolo *et al.*, 2017). The extremity of this second matured form corresponds to the footprint of MDB1 on *atpB* 5'UTR. A primer extension experiment on *atpB* 5'UTR performed by Blandine Rimbault on a WT strain, revealed the low abundance of the precursor form relative to the mature one. In the *mdb1-1* mutant, the mature *atpB* mRNA form was not present anymore (figure 5 from ***Article 1***). This indicates that MDB1 is necessary for both *atpB* mRNA stabilisation and 5'end maturation. By blocking the path of 5'→3' exonucleases it protects the downstream mature *atpB* transcript.

The maturation of the *atpB* 3' end had been studied previously (Stern *et al.*, 1991). It was defined at the downstream boundary of a protective stem-loop secondary structure. By using the circular Reverse-transcription PCR (cRT-PCR) technique we wanted to confirm this 3'end and to study the phosphorylation state of *atpB* transcript 5'end experimentally as primer extension do not give any indication on this parameter. As summarised in Figure 29, the cRT-PCR consists in:

- Ligating extracted RNAs in circular molecules. This can only occur on mono-phosphorylated RNAs. By treating half of the samples with RPP (RNA 5' Polyphosphatase) prior to circularisation, tri-phosphorylated RNAs, stemming directly from transcription start sites, are converted in mono-phosphorylated forms that can be self-ligated.

- Circularised mRNAs of interest are then retrotranscribed with a specific primer

- This cDNA is then used as a matrix to amplify the junction between the 3' and 5' end by PCR amplification.

- Lastly, the amplicons are purified and sequenced, if the junctions are heterogenous the sequence signal become ambiguous from the divergence point.

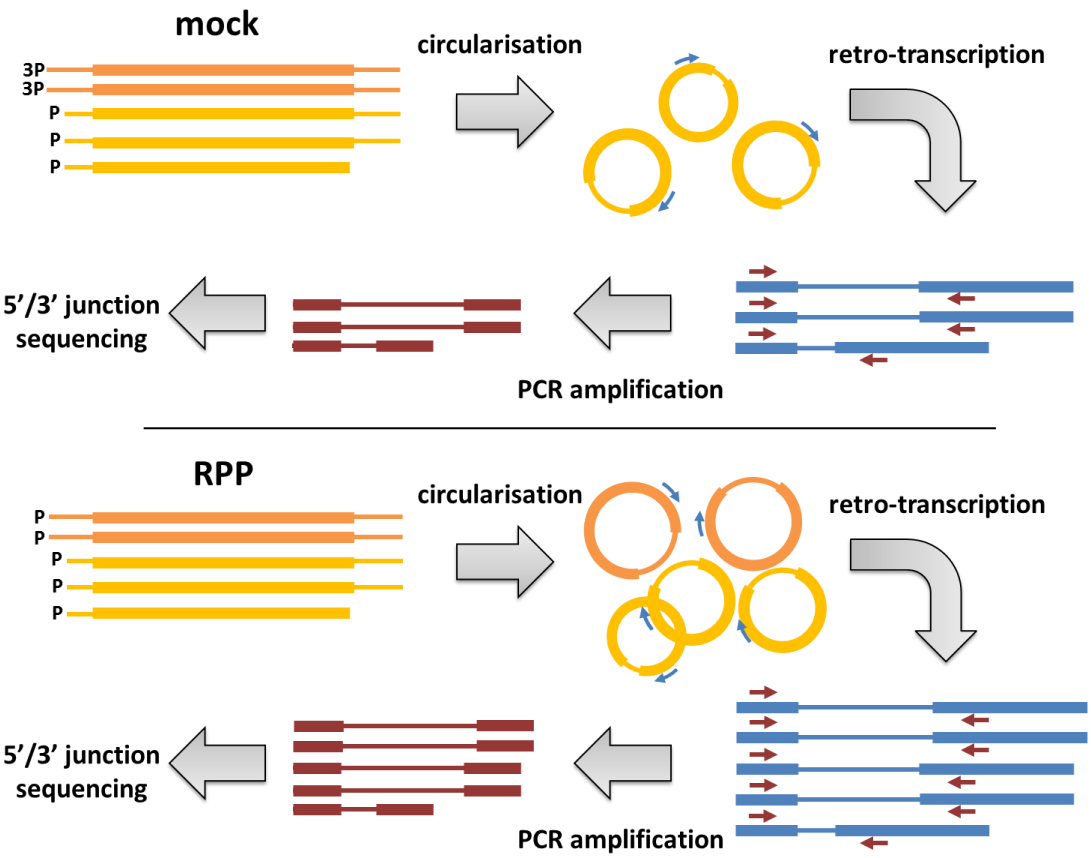
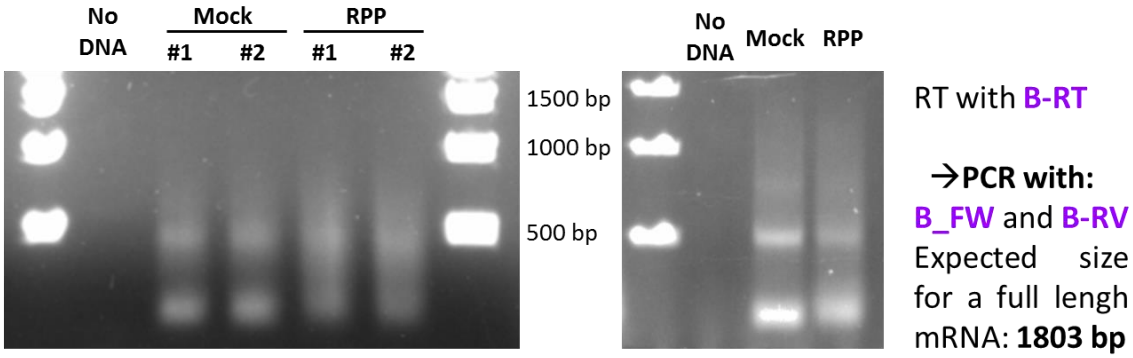


Figure 29: cRT-PCR procedure, mock and RPP samples were treated and analysed in parallel.

I set out to work with this protocol to study the occurrence of 5' and 3' ends and their phosphorylation state in a WT *atpB* mRNA.



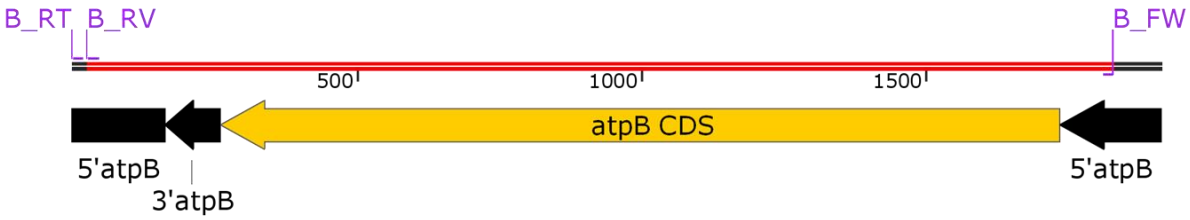


Figure 30: 2 of several attempts of cRT-PCR amplification of *atpB* with the very distant B-FW and B-RV primers.

I had difficulties to amplify a long amplicon covering the junction, and this in several independent attempts, with different polymerases, different temperatures and several independent circularisations and retro-transcriptions (Figure 30).

By designing and using primers closer to *atpB* extremities I could recover a fragment of the expected size (Figure 32), with as expected: a mature 5' end in the Mock sample, and surprisingly, a clear precursor 5' end in the RPP sample. This suggest that far more precursor mRNAs than mature ones are recovered by the cRT-PCR experiment than in the primer extension performed by Blandine Rimbault (figure 5 from ***ARTICLE 1***), where the processed 5'UTR form was far more abundant than the precursor one in the WT. But if primer extension experiments are quantitative, they do not reveal the 3' processing state of the *atpB* transcript. Here, looking at the sequencing data of the Mock amplicon, it appears that while the 5' extremity is clear, the 3' extremity has a very low-quality sequence, with seemingly many different signals mixed together. It is possible, considering the difficulties to amplify a long complete fragment of circularised *atpB* mRNA, that the *atpB* transcript endures 3' end degradation. This degradation would produce junctions with widely different 3' extremities, as is suggested by our sequencing data.

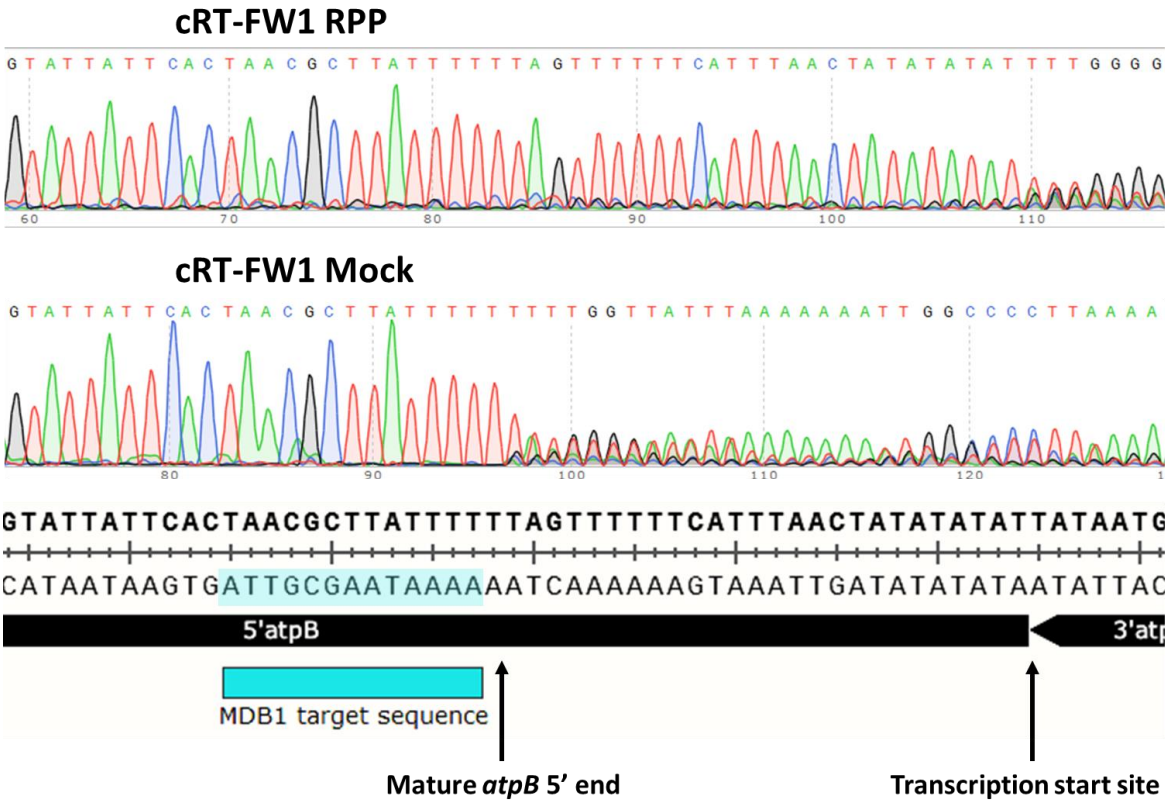


Figure 31: Sequence data of the RPP and Mock samples of a PCR amplification by B-RV and cRT-atpB FW1, the corresponding sequence of the expected precursor 5' and complete 3' junction of a circularised *atpB* mRNA is indicated below.

To get a clearer picture of the *atpB* state in my experiments, I tried to design primers progressively farther away from the junction. One of those experiments is depicted in Figure 32. The more the primers were distant, the less amplicons of the predicted size I recovered. In some reactions I recovered a faint "ladder" of amplicons of various sizes, while with very distant primers couples I only observed diffuse smears. While this could be caused by experimental issues, RNAs are sensitive to degradation and PCRs can be capricious, the repetition of strange PCR patterns over several independently treated samples with various primers and polymerases was suspicious.

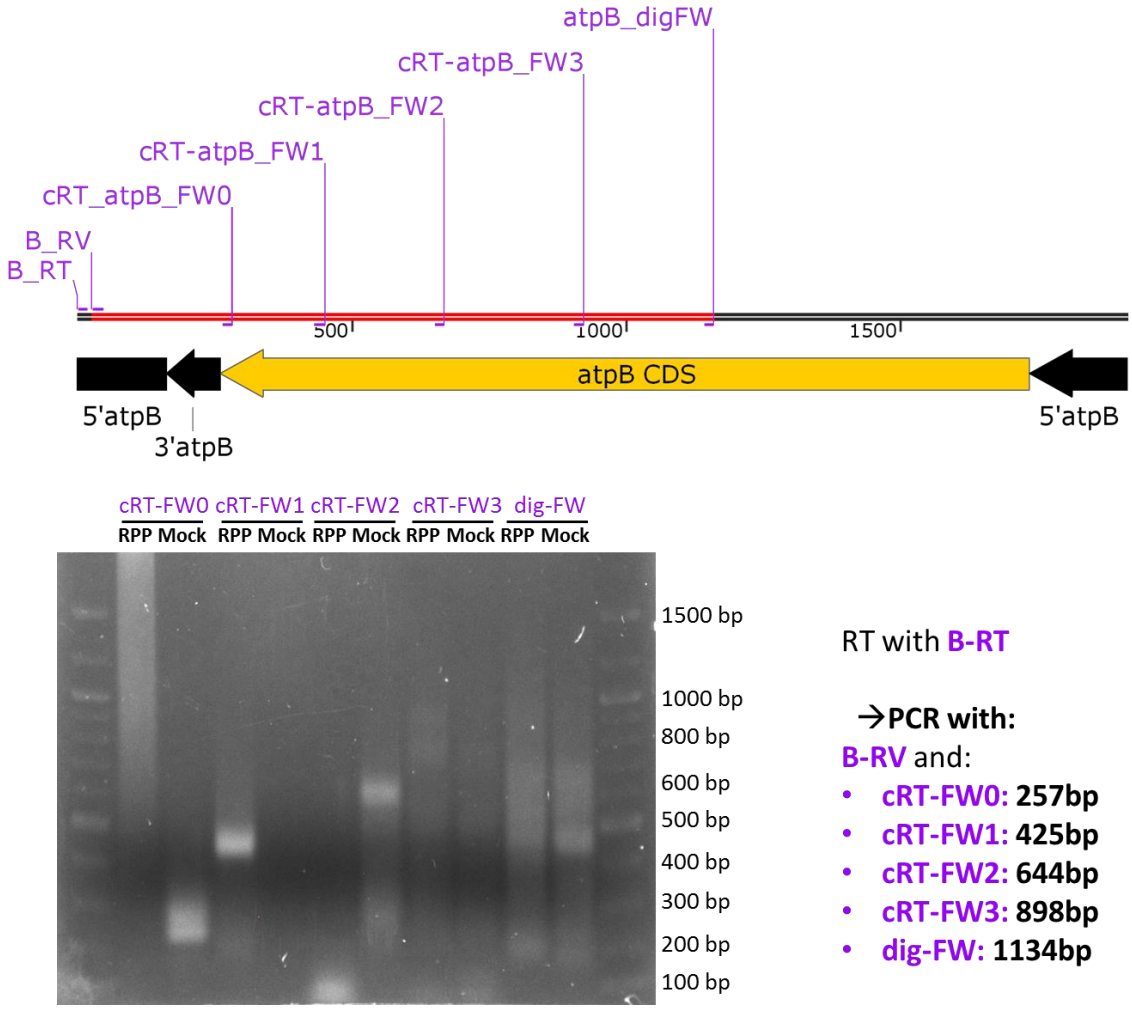


Figure 32: Amplification of cRT-PCR reactions with progressively more distant primers.

Moreover, in separate cRT-PCR experiments, Marina Cavaiuolo, after cloning cRT-PCR products, recovered amplicons in a range of variable sizes. Notably, after sequencing, the 3' extremities of the mature but not of the precursor *atpB* mRNA appear to be of a highly different lengths in our mRNA samples. Some were also poly-adenylated, which suggested that they are *atpB* mRNA committed to degradation (Schuster *et al.*, 1999; Schuster and Stern, 2009). This suggests that *atpB* mRNA is degraded from 3' → 5', even in WT samples. Such a degradation of *atpB* transcripts from this direction was unexpected, previous *in vitro* and *in vivo* studies had suggested that *atpB* mRNAs remained stable after 3'end trimming in contrary to the downstream RNAs (Stern and Kindle, 1993; Hicks *et al.*, 2002). However, *atpB* mRNAs have previously been shown to have a longer life span when translation is inhibited (Kato *et al.*, 2006), suggesting that translation of the *atpB* mRNAs might destabilise them.

Following this question, considering that precursor *atpB* mRNAs did not appear to endure degradation from their 3'end, we can wonder whether the precursor transcripts are translatable. If not, this might explain the absence of 3' degradation if translation was linked to this instability. But it is conceivable that 5' and 3' end might interact together and influence the processing of the other end. Alternatively, 5' processing, if very fast, might always occurs before the 3' end degradation might start.

Altogether, our data suggests that the mature *atpB* mRNA might endure a rapid 3'→5' degradation in the wild type, when it is translated. To assess this last point, we could look whether degradation also occurs when the *atpB* transcript is rendered untranslatable by a stop codon, or perhaps with the untranslatable *dB$_{WT}$gfp.Spix3-3'atpB* chimera that I developed in Chapter IV.

atpB 5'UTR and translation

To learn more on the role of MDB1 in *atpB* expression, we wondered whether *atpB* 5' end maturation is also necessary for its translation. In this case, MDB1 could be in some respect be both a M and T factor for *atpB*. PolyG *atpB* transcripts were not translated in previous experiments, but this happened irrespective of MDB1 absence. Presumably, the polyG track next to the translation initiation signals hampered translation directly. And so, no clues on MDB1 influence on translation could be gathered.

Impact of atpB 5'UTR maturation on translation

To assess which part of *atpB* are needed for transcript processing and translation, chimeric transcripts were built. Various combinations of CDS and 3'UTR were associated with the *atpB* 5'UTR. Some chimeric transcripts could still be matured at their 5' end, but it appeared that transcripts bearing the *rbcL* 3'UTR together with the *atpB* 5'UTR, irrespective of the CDS between them, could not be matured, whether MDB1 was present or not (Table 2). In contrast, a chimera built with the *atpA* 5'UTR and the *rbcL* 3'UTR could be matured normally (figure 7 from ***Article 1***).

chimera	*BKF*	*BBR*	*BFR*	*BRR*	*BKR*	*AKR*
5' maturation?	yes	no	no	no	no	yes

Table 2: Maturation of chimeric *5'atpB* driven transcripts, data gathered by Blandine Rimbault and Marina Cavaiuolo. The nomenclature is as follow: (*B=atpB, A=atpA, F=petA, R=rbcL* and *K=aadA*) the first letter is the 5'UTR, the second the CDS, the third the 3'UTR.

This specific impairment of the 5' processing could be caused by a secondary structure formed between the *rbcL* 3'UTR and the *atpB* 5'UTR according to RNABows, a secondary structure prediction program (Markham and Zuker, 2008). This putative structure would not prevent MDB1 binding its target on atpB 5'UTR, nor would it constitutively stabilise the transcript, as the BKR chimera does not accumulate in *mdb1-1* cells (figure 2 from ***Article 1***).

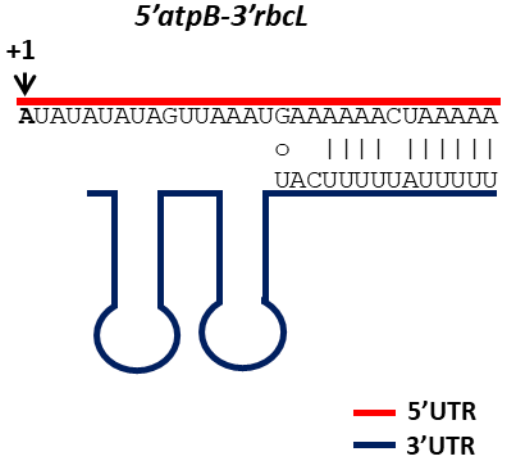


Figure 33: Predicted secondary structure forming between *atpB* 5'UTR and *rbcL* 3'UTR, obtained with the RNABows program (Markham and Zuker, 2008).

Considering that those precursor chimeric transcripts were still stabilised by MDB1 (figure 6 from ***ARTICLE 1***) but not matured, the next question was: are those transcripts translatable? The fact that the *aadA* chimeric mRNAs conferred spectinomycin resistance even without being matured indicated that some translation should still occur. *aadA* encodes an aminoglyside 3' adenyl transferase, and confers spectinomycin resistance to *C. reinhardtii cells* (Goldschmidt-Clermont, 1991). To ascertain that the precursor transcripts could be translated, I performed immunoblots after extracting whole cell proteins.

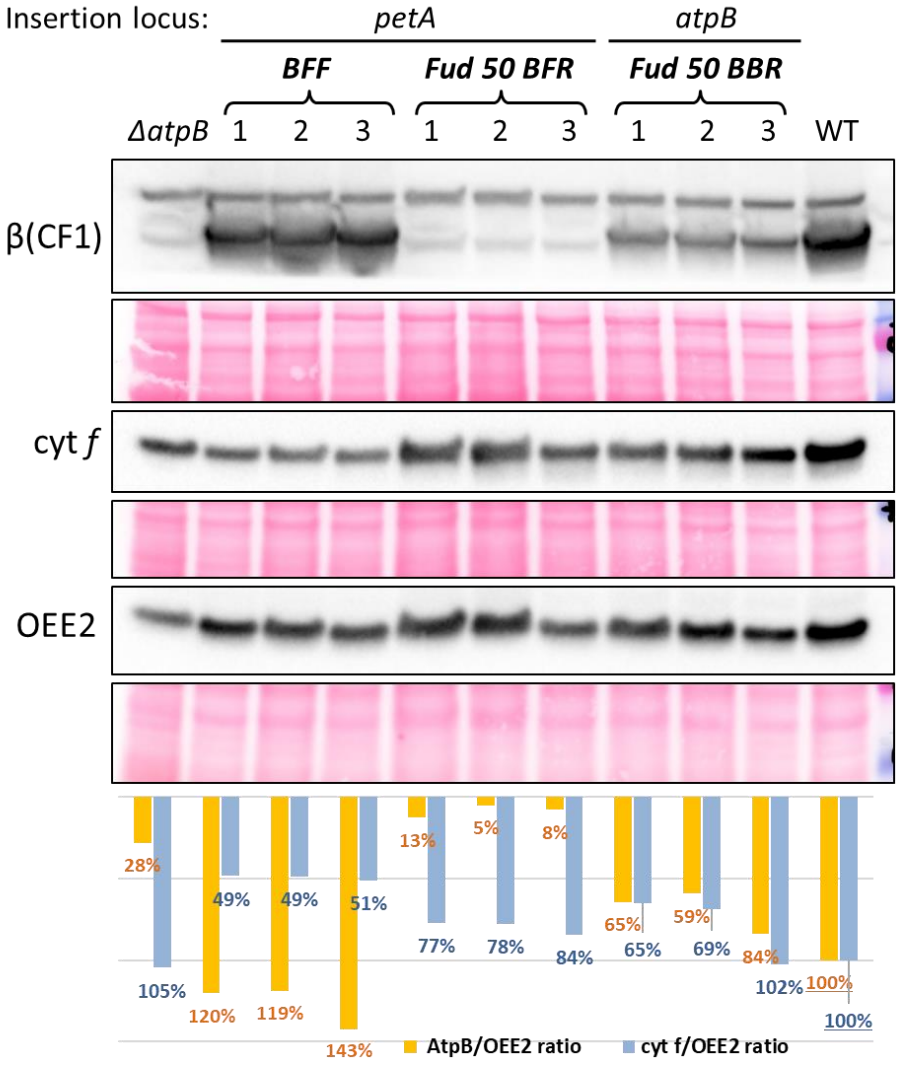


Figure 34: Immunoblot, anti-cyt. *f*, OEE2 and β (CF1) primary antibodies were used. OEE2 serve as loading control. Corresponding regions of the filter in Ponceau red stain are under the chemiluminescent signals. The band above the β (CF1) signal, appearing even in the *ΔatpB* lane, is the cross reacting mitochondrial ATP synthase β subunit. The faint band just below is an unrelated cross-reacting contaminant. Fud50 is a chloroplast mutant with a partially deleted atpB. Protein levels were quantified with the Image Lab program, and are normalised to the accumulation of the OEE2 protein and expressed as ratio of the wild-type

Note that the accumulation of the ATP synthase β subunit in these quantifications is not null as predicted for the Fud50 BFR transformants, because of the presence of a contaminating polypeptide, also present in the *atpB* deletion strain.

The mature *BFF* chimeric transcript could be translated as evidenced by Cyt. *f* accumulation. Both the BFR and BBR transcripts, that cannot be matured could also be translated, Cyt. *f* and the ATP synthase β subunit accumulated, indicating that 5' end maturation is not a prerequisite for *atpB* 5'UTR-driven translation. However, the levels of β subunit in the {*Fud 50*} strain (a chloroplast mutant with a partial deletion of *atpB*) transformed with *BBR* were about 30% lower than in the WT. The accumulation level of Cyt. *f* in the {*Fud50 BFR*} was about 20% lower than in the WT.

This suggests that either the maturation states of *atpB* 5'UTR slightly affect translation, or maybe that the *rbcL* 3'UTR is inherently a weaker 3'UTR to induce translation than the *atpB* 3'UTR, but not much weaker than the *petA* 3'UTR. The observation that *rbcL* 3'UTR reduces the translation of *atpB* was also made previously (Rott *et al.*, 1998b).

I also performed immunoblots of *aadA* chimeric transformants, but our anti-AadA antibodies gave so strong cross-reactions (red asterisk) that clear conclusions could not be drawn so far on *aadA* accumulation. A very faint band (red arrow), absent in the WT could be AadA, as it can be observed only in the *AKR* and *ΔatpB* strains, this could be coherent as this deletion of *atpB* was created by replacing it with an 5'*psaA*-driven *aadA* recycling cassette. Obviously, this immunoblot should be retried with a better anti-AadA antibody. However, all strains are spectinomycin resistant, thus AadA should be at least slightly translated from the chimeric transcripts.

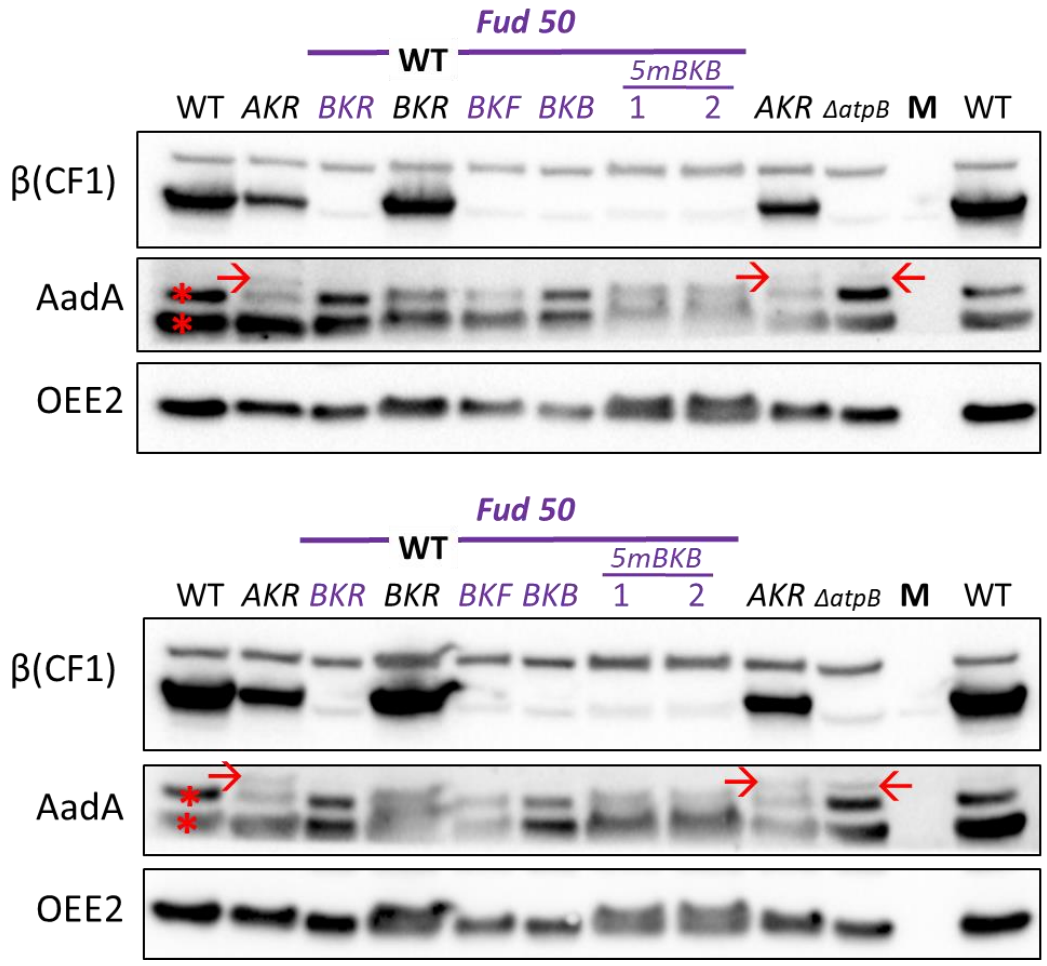


Figure 35: Immunoblots, two technical repeats, anti-AadA, OEE2 and β (CF1) primary antibodies were used. OEE2 serve as loading control. * indicate cross reactions of the anti-AadA antibody, → a possible specific AadA signal.

IDENTIFICATION OF A NEW MUTANT OF MDB1

While working on establishing an efficient CRISPR edition protocol of *C. reinhardtii* nuclear genome, Catherine de Vitry, Marcio Rodrigies-Azevado, and Frédéric Chaux-Jukic obtained new ATP synthase mutants in a screen for photosensitive mutants. I helped them characterise those by performing RNA blots and took part in the genomic analysis. We notably identified a mutant of *ATPG*, interrupted by a TOC1 retrotransposon. This gene was to our knowledge never mutated previously in *C. reinhardtii*. This side project is described in ***ARTICLE 2***, whose draft is attached in the annexes of this manuscript.

Using the methods described in this article I also studied with Frédéric Chaux-Jukic a mutant obtained independently by Katia Wostrikoff while she was also trying to set up the CRISPR protocol. She found it with a negative screen on minimum media; it could not grow photo-autotrophically.

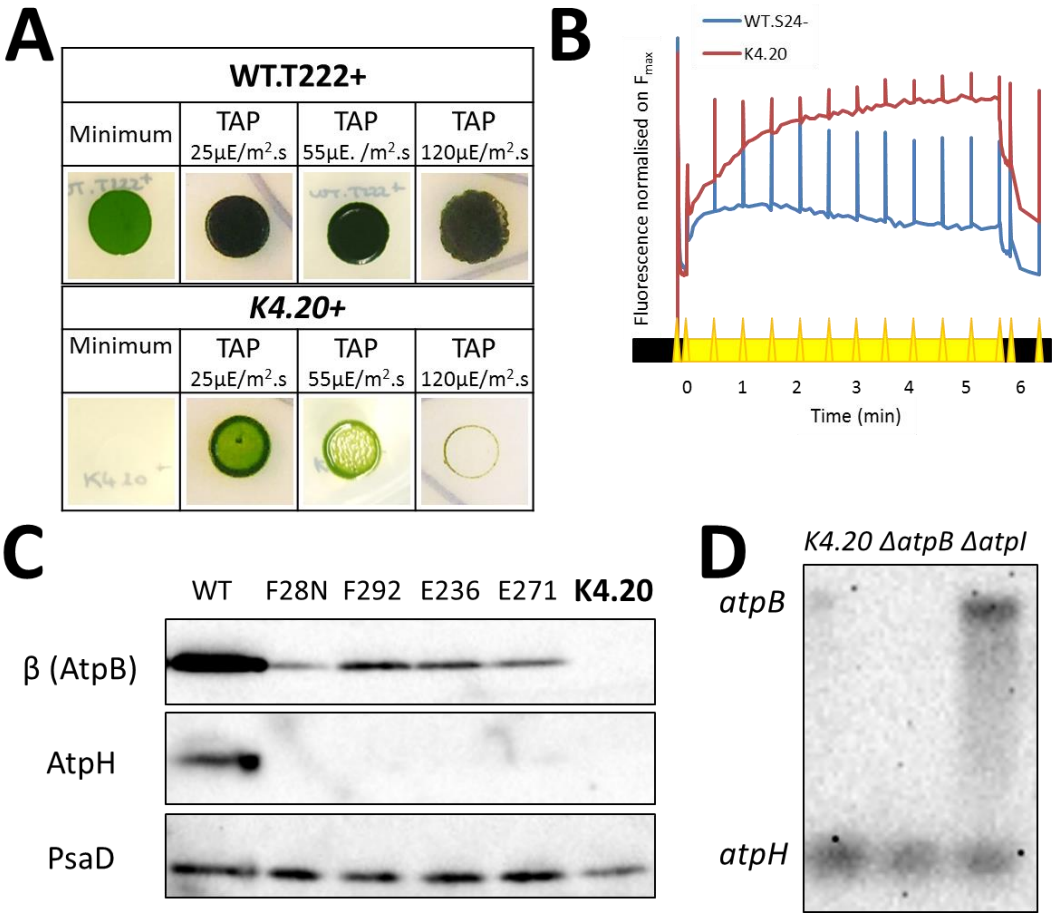


Figure 36: A. Growth phenotype of *K4.20+*. **B.** Kinetic of PSII fluorescence saturation in K4.20+ versus WT.T222+. Obtained with a SpeedZen camera. **C.** Immunoblot of K4.20 and other mutants of ATP synthase hybridised with anti AtpB, AtpH and PsaD antibodies. **D.** RNA blots of *K4.20*, filters were hybridised with *atpB* and *atpH* dig-dUTP labelled probes.

It also displayed a photosensitive phenotype under strong illumination (120 $\mu E.m^{-2}.s^{-1}$) (Figure 36.A) and its photosensitivity makes it grow slower than a WT under moderate illumination (25 $\mu E.m^{-2}.s^{-1}$). Fluorescence kinetics of PSII revealed that the PSII yield decreased with time under constant illumination, which is a typical phenotype of ATP synthase defective mutants. PSII becomes progressively completely reduced, the photosynthetic process gradually interrupted by the excessive proton gradient that cannot get consumed by the ATP synthase as it would in a WT.

Immunoblots revealed that this *K4.20* mutant did not accumulate the chloroplast ATP synthase β subunit (Figure 36.C). This blot also illustrates the concerted accumulation of the chloroplast ATP synthase in *C. reinhardtii*: when the CF1 is absent the CFo does not accumulate either.

After extracting RNAs from *K4.20* we blotted them and assessed the accumulation of several chloroplast transcripts. While both *atpA*, *atpE* and *atpH* transcripts accumulated, *atpB* mRNA was undetectable, as in *mdb1* mutants (Figure 36.D).

Whole genome Illumina sequencing of *K4.20* was performed by the *Eurofins* company. While screening candidate genes implicated in the chloroplast ATP synthase biogenesis, I found, in the *MDB1* first intron of the *K4.20* strain but not in the WT, a suspicious accumulation of reads with mates pairing on other chromosomes. This suggested a sequence mapping on several chromosomes, i.e. present in numerous copies in the genome (Figure 37.A). By aligning those promiscuous sequences, I found that they corresponded to the **T**ransposon **o**f ***C****hlamydomonas* 1 (TOC1) retrotransposon, abundant in the nuclear genome of *C. reinhardtii* (Day *et al.*, 1988). The stress induced by the transformation treatment probably induced the spontaneous jump of the retrotransposon in *MDB1*. While insertions in introns do not necessarily impair the expression of genes, this recently inserted, whole, functional TOC1 did prevent *MDB1* expression. A reconstructed probable map of *MDB1* of *K4.20*, interrupted by TOC1, is included in Figure 37.B.

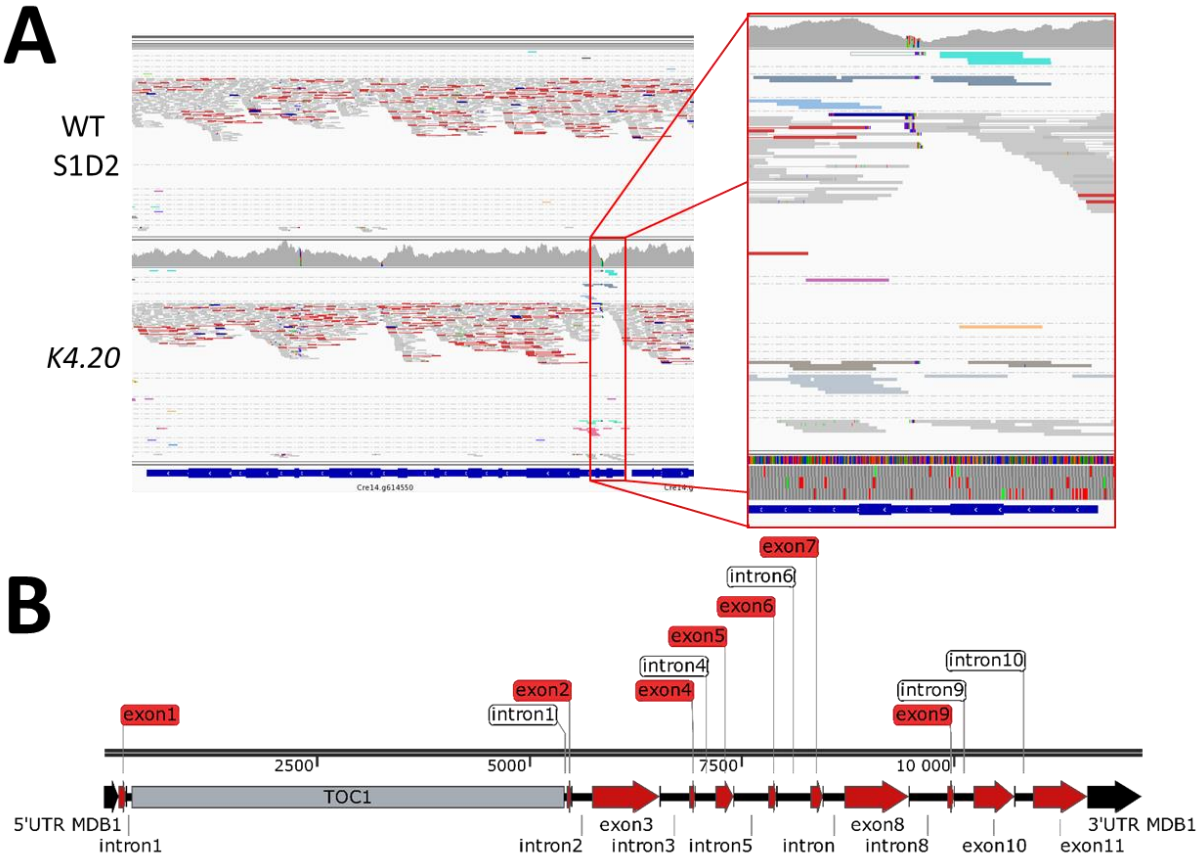


Figure 37: A. Insertion site of TOC1 in *MDB1* in the *K4.20* mutant, Illumina paired end whole genome sequencing visualised with IGV. Reads pairing with mates on other chromosomes are indicated in pastel colours. Analysis of the mates' sequences with BLAST revealed TOC1 sequences **B. Reconstructed map of *MDB1* in the K4.20 mutant.**

To further confirm the presence of this insertion in *K4.20* and its link to the ATP synthase deficiency I crossed *K4.20*$^{+}$ with WT.S24^{-}. Descendants were selected on photosynthetic deficiency and their DNA extracted.

PCR amplification with a probe specific to *MDB1* exon 3 and one specific to TOC1 yielded a specific amplicon in *K4.20* and its non-photoautotrophic descendants, not in the wild type (Figure 38). Probes specific to MT+ and MT- were added in the same reaction, both as a control of the DNA quality and to make sure that both mating types were represented in the progeny, as we should expect from a successful cross.

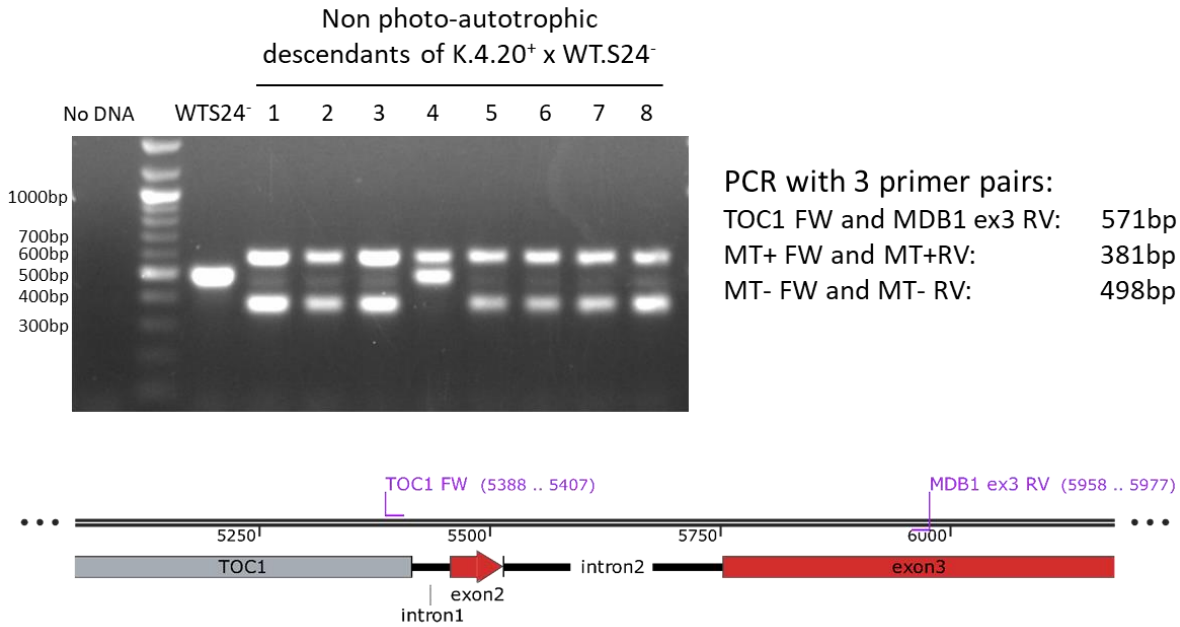


Figure 38: PCR amplification of the junction between TOC1 and exon3 of *MDB1* in *K4.20*$^{+}$ x WT.S24^{-} non-photosynthetic progeny. Mating type PCR amplification was done in the same reaction as a DNA quality control.

Therefore, the mutant phenotype and genotype are congruent. I characterised a new mutant of *MDB1*, further cementing its importance in *atpB* mRNA stabilisation. This *K4.20* mutant could be now called *mdb1-3*.

This spontaneous mutant with no antibiotic resistance linked to its MDB1 deficiency could be useful in other studies but did display phenotypic reversion when millions of cells were plated on minimum media (data not shown). I recently also found out that some of the descendants of the cross became photoautotroph. This is probably because the retrotransposon can sometimes get inactivated. This instability renders the strain perilous to work with.

DISCUSSION

MDB1 A TRUE MATURATING FACTOR CONTROLLING *ATPB EXPRESSION*

MDB1 is required for the stabilisation of *atpB* transcript. A specific footprint of MDB1 (figure 3 from ***ARTICLE 1***), and the pull down of *atpB* mRNA in immunoprecipitation experiments (figure 2 from ***ARTICLE 1***) strongly suggests that this OPR protein directly binds on *atpB* 5'UTR. Addition of a voluminous and stable polyG track in *atpB* 5'UTR produces the same protective effect. However, the 5' end maturation of this artificially protected *atpB* is different (figure 4 from ***ARTICLE 1***). This suggests that the bound MDB1 on the mRNA protects it from 5'→3' exonucleases and is used to define the 5' boundary of *atpB* mRNA. Moreover, in mutant of MDB1 the trace amounts of *atpB* transcript that can be observed correspond only to the precursor transcript, the form that is produced directly from transcription. The polyG addition allowed the recovery of higher levels of *atpB* mRNAs than in the WT, recovering *atpB* transcript otherwise doomed to degradation even when MDB1 is present. This indicates that this M factor is present in limiting amounts in the chloroplast. A similar situation has been observed previously for *petA* mRNA stabilisation by MCA1, another M factor (Loiselay *et al.*, 2008).

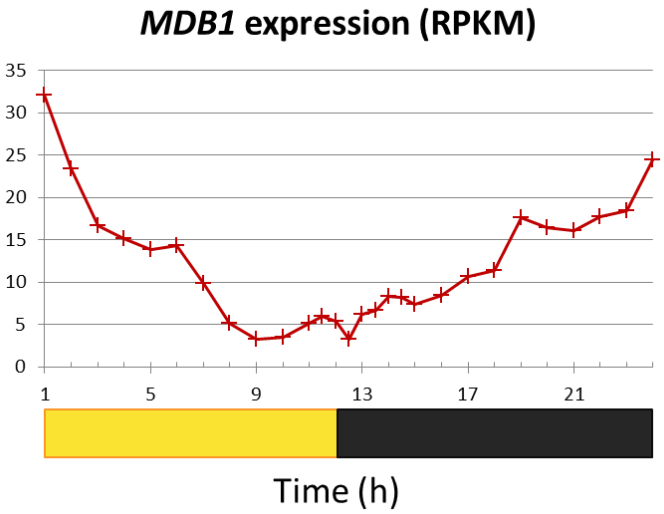


Figure 39: *MDB1* expression along the day-night cycle, transcriptomic data redrawn from (Zones *et al.*, 2015).

MDB1 is more expressed at the end of the night until the onset of light in circadian experiments (Zones *et al.*, 2015) (Figure 39) probably to prepare the photosynthetic apparatus in prevision of the day. Moreover, MDB1 appear to be activated under light by the bilin signalling pathway, in the *hmox1* mutant, unable to use the bilin signals, the expression of MDB1 was not modified by light, addition of biliverdin partially rescued a stronger expression under light. (Duanmu *et al.*, 2013; Wittkopp *et al.*, 2017). This could imply that the nucleus senses the state of the chloroplast thanks to the bilin retrograde signals and adjust the production of this M factor to tune the accumulation of ATP synthase β CF1 according to the chloroplast needs.

COULD ATPB MATURATION BE INFLUENCED BY ITS 3'UTR?

Surprisingly, *atpB* mRNA maturation might also involve secondary structures and 5'/3' end interactions as chimeric transcripts bearing *atpB* 5'UTR and *rbcL* 3' UTR were not properly processed and accumulated only as precursor mRNAs. However, this was not noted for other transcripts bearing *petA* 3'UTR in conjunction with *atpB* 5'UTR, nor with chimeras based on *atpA* 5'UTR and *rbcL* 3'UTR. This impairment of maturation could be caused by a specific incompatibility of *atpB* 5'UTR and *rbcL* 3' UTR. If this interaction modulates MDB1 activity it should not affect it dramatically, as the un-matured *BKR* transcript could still accumulate in a MDB1-dependant fashion (figure 2 from ***ARTICLE 1***). The maturation status of *atpB* transcript 5'UTR does not seem to significantly affect translation, as un-matured reporter constructs based on *aadA* are spectinomycin resistant. Moreover, immunoblots proved that un-matured transcripts BFR and BBR still allowed translation and accumulation of around 70-80% of Cyt. *f* and the ATP synthase β subunit.

ATPB MRNA ENDURES RAPID DEGRADATION

The rapid maturation of *atpB* 5'UTR and the apparent large portion of *atpB* transcripts stabilised by the polyG track, indicate that *atpB* mRNA is transcribed in large excess and that most the transcripts are fated to be rapidly degraded by 5'→3' exonucleases. But not only 5'→3' degradation of the transcript seems to occur. Indeed, our cRT-PCR experiments suggest that mature *atpB* mRNAs are also progressively degraded from their 3' end. Considering that it has previously been observed that the *atpB* transcripts appear longer lived when chloroplast translation is interrupted (Kato *et al.*, 2006), we ponder whether this effect is connected to translation, to test this hypothesis further cRT-PCR experiment with untranslatable transcripts, with an initiation codon replaced by a STOP one for instance, would be insightful. Alternatively, cRT-PCR experiments could be repeated for comparison on strains expressing chimeras treated with lincomycin. For now, it is impossible to conclude with certainty on this matter.

INACTIVATION OF *MDB1* BY TOC1 INSERTION

We identified and characterised a new mutant of MDB1 interrupted by a TOC1 retrotransposon in its first intron. The fact that this insertion prevents *MDB1* expression is not trivial. It was previously observed that an insertion of a full transposon in *PSB1* (encoding OEE1) prevents the expression of the *PSB1* mRNA and that a spontaneous partial deletion of the transposon partially restores the accumulation of the *PSB1* transcript and OEE1 synthesis. This suggests that the partial transposon allows the transcription and splicing of *PSB1* (Mayfield *et al.*, 1987) The situation in *mdb1-3* could be similar, considering its propensity for reversion. Indeed, the consensus splicing sequences of the intron 1 are preserved by the retrotransposon insertion, and so, the mRNA could in theory be properly spliced. But maybe the *MDB1* transcript is not properly transcribed because of the conflicting transcription termination signals of TOC1 that is inserted in the same orientation as *MDB1*. We could answer this question by performing a RT-PCR of the of the *MDB1* mRNA between exon 1 and 2.

However, even if transcription was not prevented by the transposon, introns are involved at many steps of eukaryotic gene expression (Le Hir *et al.*, 2003; Hernandez-Garcia and Finer, 2014). They can affect transcription; notably promotor-proximal introns may enhance transcription by re-initiation of RNA polymerase. Introns can also affect nuclear export; consensus splice sites serve as signals to prevent the export of unspliced transcripts out of the nucleus. But splicing can also actively stimulate mRNA export by recruiting export factors. Introns can also modulate RNA degradation or translation. It has been reported that the introns of *Chlamydomonas reinhardtii* nuclear genes can be crucial for gene expression. Expression of a transformed endogenous *ALS* (acetolactase synthase) was proved to be improved when all introns were preserved (Kovar *et al.*, 2002). The introns of *RBCS2* have a positive effect on post-transcriptional expression of foreign genes: *ble* (Bleomycin resistance protein) (Lumbreras *et al.*, 1998), of aph7" (aminoglycoside phosphotransferase) (Berthold *et al.*, 2002) or of the *Renilla*-luciferase gene (Eichler-Stahlberg *et al.*, 2009) among others. Furthermore, the first intron of RBCS2 contains a transcriptional enhancer (Lumbreras *et al.*, 1998; Baier *et al.*, 2018). It is thus also conceivable that a critical enhancing sequence might be interrupted by the insertion of TOC1 in *MDB1* first intron.

II. FUNCTIONAL STUDIES OF THE MTHI1 FACTOR

Modified from Dartmouth Electron Microscope Facility, Dartmouth College

INTRODUCTION

EXPRESSION SYSTEMS OF ATP SYNTHASE CFO IN *C. REINHARDTII*

The chloroplast ATP synthase is composed of two domains: the soluble domain CF1 catalyses the formation of ATP from ADP and Pi and comprises five subunits, α, β, γ, δ, ε in a 3:3:1:1:1 stoichiometry. The membrane-embedded domain, CFo, is a selective proton channel made of four subunits, AtpF, ATPG, AtpH and AtpI (formerly called Subunits, I, II, III, IV) in a 14:1:1:1 stoichiometry. The plastid genome of *C. reinhardtii* still encodes six of the nine ATP synthase subunits: α (*atpA*), β (*atpB*) and ε (*atpE*) of the CF1 domain and AtpF, AtpH and AtpI of the CFo domain. *ATPC (γ)*, *ATPD (δ)* and *ATPG* (subunit II) have been transferred to the nucleus as in the rest of Viridiplantae. However, unlike most Archaeplastida which have largely kept the two original cyanobacterial operons, in Chlorophyceae, ATP synthase genes have been shuffled around the chloroplast genome (Figure 40).

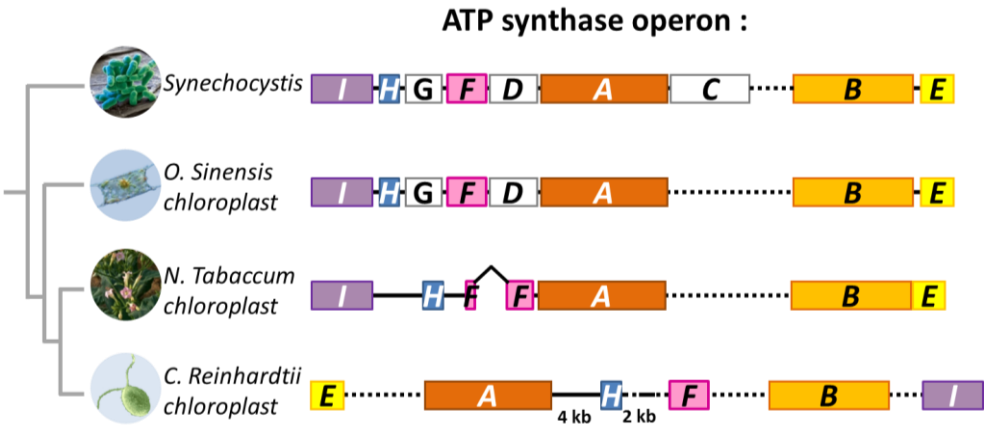


Figure 40: ATP synthase operons in cyanobacteria and chloroplasts

There is nearly no trace left of ATP synthase operons in *C. reinhardtii*. Only *atpA* and *atpH* on one hand and *atpH* and *atpF* on the other hand can be transcribed together in polycistrons but they are still quite far apart on the genome and are even intercepted with unrelated genes in the case of *atpA* and *atpH* (Figure 41.B). *atpI* is about 40 kb apart from *atpH* and is transcribed in a polycistronic unit with genes of unrelated functions.

To be correctly assembled the CFo subunits need to accumulate at the right stoichiometry. Since the subunits are not expressed in operon like in bacterial systems, with common promotors, and are not even synthesised in the same compartment, novel regulatory mechanisms are necessary in the chloroplast.

This control, as we saw previously (Introduction), can occur either dynamically:

- By degradation of surplus subunits, as is often the case for nucleus-encoded subunits when they accumulate unassembled in the chloroplast.
- By CES, an assembly-dependant regulation of translation, via a reduction of the translation of surplus subunits, or stimulation of the translation of depleted subunits.

Or in a constitutive manner: a control mechanism may link the expression of the different subunits thru a common regulator in limiting amount. If the regulator level fluctuates, the expression of the subunits would fluctuate together, therefore keeping a lock on the systems. If the nucleus is to control the chloroplast function, we would expect this factor to be nucleus encoded.

The question of how CFo assembly is regulated had not been studied previously. I contributed to the description of this regulation by working on *atpH*, *atpI* and on an OTAF: MTHI1. Most of this work has been included in ***ARTICLE 3*** (attached at the end of this manuscript).

atpH

atpH is a small gene encoding subunit AtpH of the plastid ATP synthase CFo. It can be transcribed from *atpA* promotor but is expressed mainly from its own promotor (Figure 41.B). *atpH* CDS is only 249 nt long and its 5'UTR is also very short, 41 nt. Small RNA sequencing revealed a very abundant 22 nt footprint specific of MTHI1 mapping to the triphosphorylated 5'end of the *atpH* mRNA (Figure 41.A). *atpH* is the third most abundant mRNA in the chloroplast of *C. reinhardtii* (Cavaiuolo *et al.*, 2017).

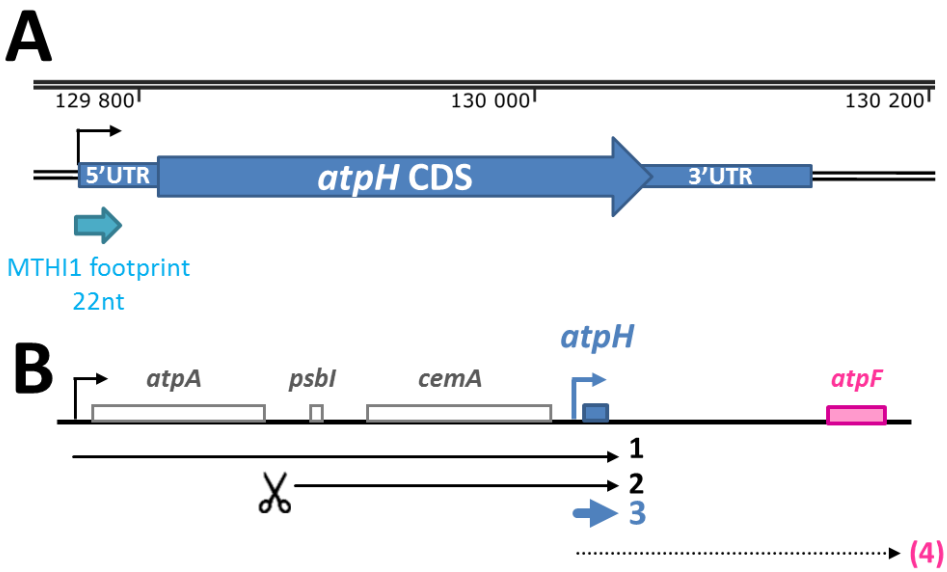


Figure 41: A. The *atpH* gene, MTHI footprint is indicated by the turquoise arrow. **B. Transcription of *atpH*,** *atpF* is transcribed from the *atpH* promotor as it lacks a dedicated one, as suggested by small RNA coverage (Cavaiuolo *et al.*, 2017).

atpI

atpI encodes the subunit AtpI of ATP synthase CFo. Its CDS is 717 nt long and its extensive 5'UTR is 493 nt long, its 3'UTR has not been precisely described but is probably around 100 nt long. (Figure 42.A). *atpI* does not have a dedicated promotor. It is exclusively transcribed as a polycistronic transcript starting at the promotor of *psbD* that undergoes several endonucleolytic cleavages (Figure 42.B).

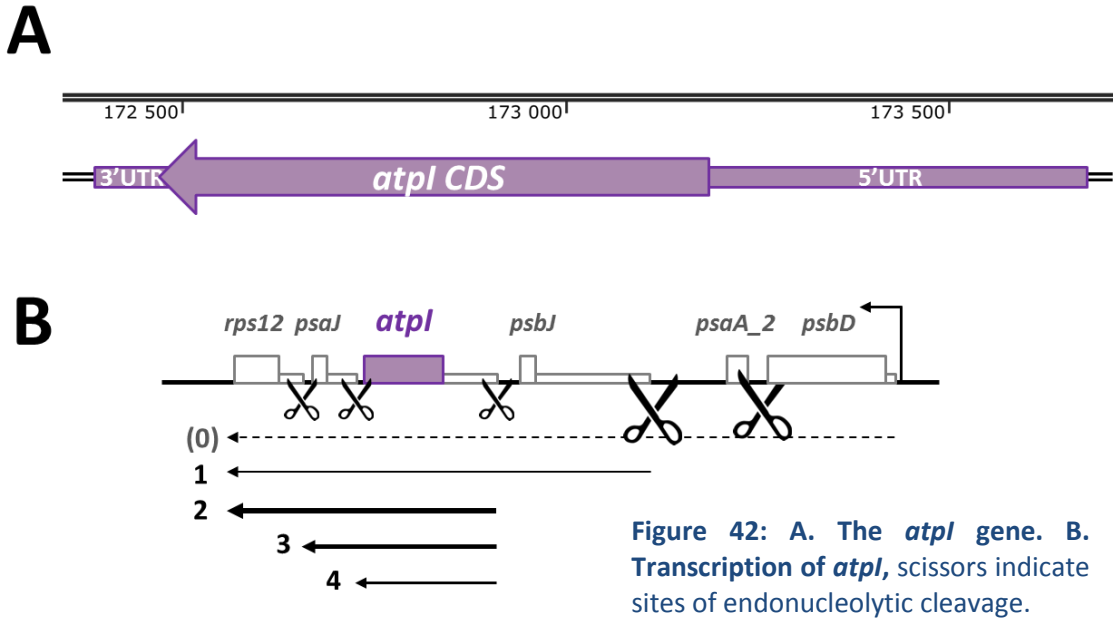


Figure 42: A. The *atpI* gene. B. Transcription of *atpI*, scissors indicate sites of endonucleolytic cleavage.

MTHI1

The *ac46* mutant, for acetate requiring mutant, isolated in 1960 (Levine, 1960), is a nuclear mutant which does not express AtpH nor AtpI. It does not accumulate the *atpH* monocistronic mRNA (Majeran *et al.*, 2001), thus doesn't translate AtpH and barely translate AtpI, but still translates AtpF (Lemaire and Wollman, 1989). The nuclear factor affected in *ac46* mutant has been named MTHI1 (**M**aturation and **T**ranslation of *atp**H*** and *atp**I***). For the rest of this study *ac46* will be dubbed *mthi1-1*. Another *MTHI1* mutant was later characterised: CAL014.01.38 (*mthi1-2)*, which displays the same phenotype as *mthi1-1,* no *atpH* monocistronic mRNA and a strongly reduced translation of *AtpI*.

MTHI1 is erroneously mapped on chromosome 17 in v5.5 of *C. reinhardtii* genome. It is 6019 nt long and contains 10 introns. It encodes a protein of 828 amino acids, predicted to be targeted to the chloroplast by the Predalgo software (Tardif *et al.*, 2012). It bears 9 OPR contiguous motifs at its N-terminal region and a bulky disordered domain at its C-terminal. Orthologues of MTHI1 in Chlorophyceae and Ulvales show conservation of the OPR domain but not of the C-terminal one (see figure S6 from ***ARTICLE 3***).

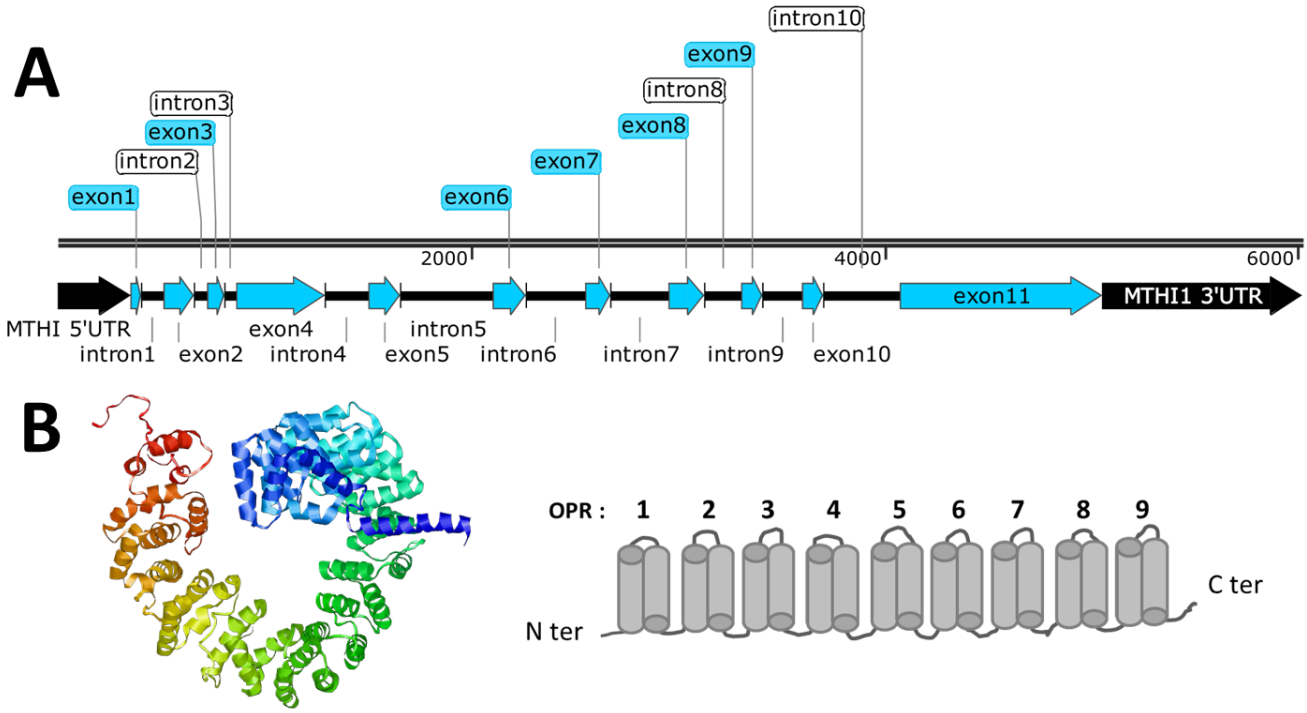


Figure 43: A. the MTHI1 gene. B. Left: Predicted structure of MTHI by I-Tasser, right: Cartoon of the OPR domain of MTHI1

The double phenotype displayed by MTHI1 mutants has long been intriguing; it could either imply a CES relationship between AtpH and AtpI: when AtpH is absent AtpI translation would be halted. Or it could imply a dual action of MTHI1 on both *atpI* and *atpH* mRNA. It was later proved in the article that AtpI is translated normally even in absence of AtpH while AtpH is translated at wild-type levels in the absence of AtpI (see figure 2 from ***ARTICLE 3***). This disproves a CES relationship between AtpH and AtpI and reinforces our view that the MTHI factor has a double function.

MTHI1 is an OPR protein with 9 OPR repeats, that most probably binds tightly on *atpH* 5'UTR (as revealed by a footprint, see Figure 41.A and figure 10 of ***ARTICLE 3***), but only weakly on *atpI* mRNA (since not footprint have been recovered so far). The specific footprint of MTHI1 on *atpH* transcript is recovered only in RNA 5' Polyphosphatase (RPP) treated samples. RPP converts tri-phosphorylated RNAs, which cannot be sequenced otherwise, in mono-phosphorylated RNAs. This proves that MTHI1 interacts directly at the 5'end of *atpH* monocistronic transcript which is only expressed from its dedicated promotor. A 9 nucleotides sequence at the very beginning of *atpH* mRNA, GGUUGUUAU, well conserved in other Chlorophyceae, Ulvales and Pedinophyceae (see figure S8 from ***ARTICLE 3***), likely corresponds to the target of this 9 OPR motif factor.

A sequence nearly identical to the binding site of MTHI1 in the *atpH* mRNA, GGUU**A**UUAU, was found in the 5'UTR of the *atpI* mRNA. Interestingly, in an otherwise evolutionary poorly conserved *atpI* 5'UTR, this sequence is in a conserved region. These putative target sequences are shown in Figure 44.

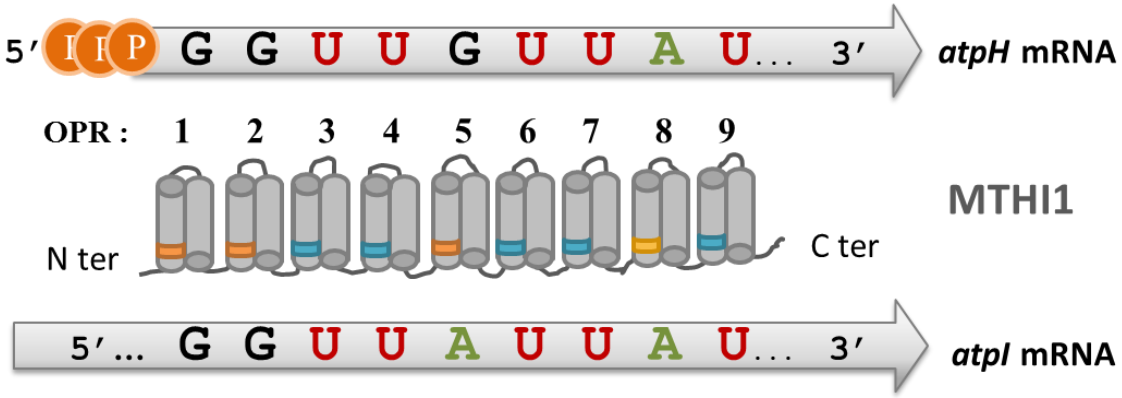


Figure 44: MTHI1 and its two putative target sequences, one is at the very 5' extremity of the *atpH* monocistronic mRNA, the other within the *atpI* 5'UTR, about 60 nt upstream of the initiation codon. Note the difference of the 5th nucleotide between the two chloroplast genes.

Mutation of the putative target sequence in a chimera driven by the *atpI* 5'UTR results in a similar phenotype as *mthi1-1*, lower transcript accumulation and halted translation, confirming that this sequence is necessary for *atpI* expression (see figure 13 of ***ARTICLE 3***).

These observations spurred us to investigate how exactly MTHI1 influences *atpI and atpH* expression.

RESULTS

SWAPPING *ATPH* AND *ATPI* MTHI1 TARGET

One of the most striking features of MTHI1 is that it acts specifically on two targets. This is very rare for an OTAF in *C. reinhardtii*, as they usually target a single gene. MTHI1 is thus doubly indispensable to the ATP synthase CFo synthesis. Moreover, *atpH* and *atpI* are not expressed at similar levels. Indeed, the *atpH* mRNA is 10 times more abundant than the *atpI* mRNA (Cavaiuolo *et al.*, 2017). The stoichiometry of the two subunits is also vastly unbalanced: 1 AtpI for 14 AtpH (Figure 45). Therefore, if both share a same key factor, and if that factor is limiting, how is the proper stoichiometry established? Small RNA sequencing (Cavaiuolo *et al.*, 2017) only revealed a MTHI1 footprint on the *atpH* mRNA 5' end, but none along the *atpI* mRNA, suggesting that the binding of this MTHI1 to the *atpI* transcript could be transient or weak. This could be caused by the difference of one nucleotide in the middle of MTHI1 target sequences: GGUU**G**UUAU for *atpH,* GGUU**A**UUAU for *atpI*. In fact, if the affinity of MTHI1 for *atpI* sequence is weaker, we could expect less and weaker interactions with this mRNA, explaining why this OPR is an M factor of *atpH*, tightly binding to its mRNA, while it would only act transiently on the *atpI* transcript, to initiate translation. To test whether the different functions of MTHI1 result from this variation in the target sequence, we swapped them.

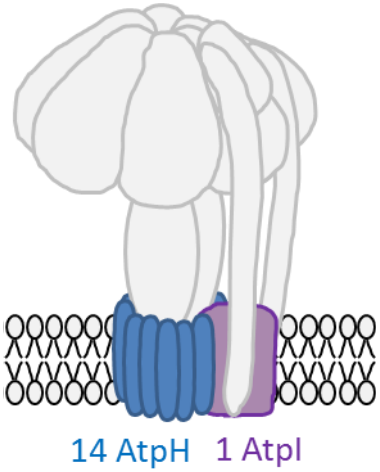


Figure 45: *C. reinhardtii* chloroplast ATP synthase, AtpI/AtpH stoichiometry is indicated

Chimeric constructs (Figure 46) bearing either *atpH* 5'UTR or *atpI* 5'UTR, mutated or not in their MTHI1 target, followed by *petA* coding sequence, were generated by Shin-Ichiro Ozawa and Yves Choquet and transformed in place of *petA*. *petA* is a good reporter gene in the chloroplast: it is a stable protein whose accumulation faithfully reflects the rate of synthesis and it is easy to monitor both at the mRNA and protein level. Moreover, the use of such chimeras allowed us to monitor the impact of the 5' UTR on expression, without the possible influence from the CDS or 3'UTR.

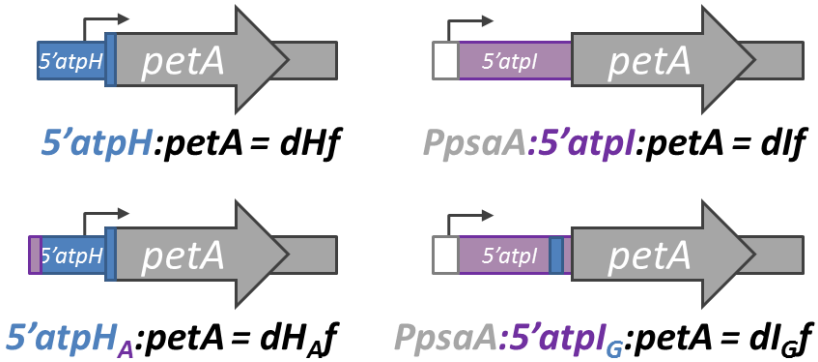


Figure 46: Schematic presentation of the two mutated chimeric constructs. The *psaA* promotor (white rectangle) was added in front of *atpI 5'UTR*

All the transformants were photoautotroph, indicating that *petA* could be expressed at least partially from all constructs. I extracted the RNA from several independent transformants for each of the four constructs and assessed the accumulation levels of the *petA* transcript by RNA blots (Figure 47). The samples were blotted twice independently. The *dIf* and *dI$_G$f* chimeras displayed the same transcript accumulation in the two technical repeats, as did the *dHf* and *dH$_A$f* chimeras.

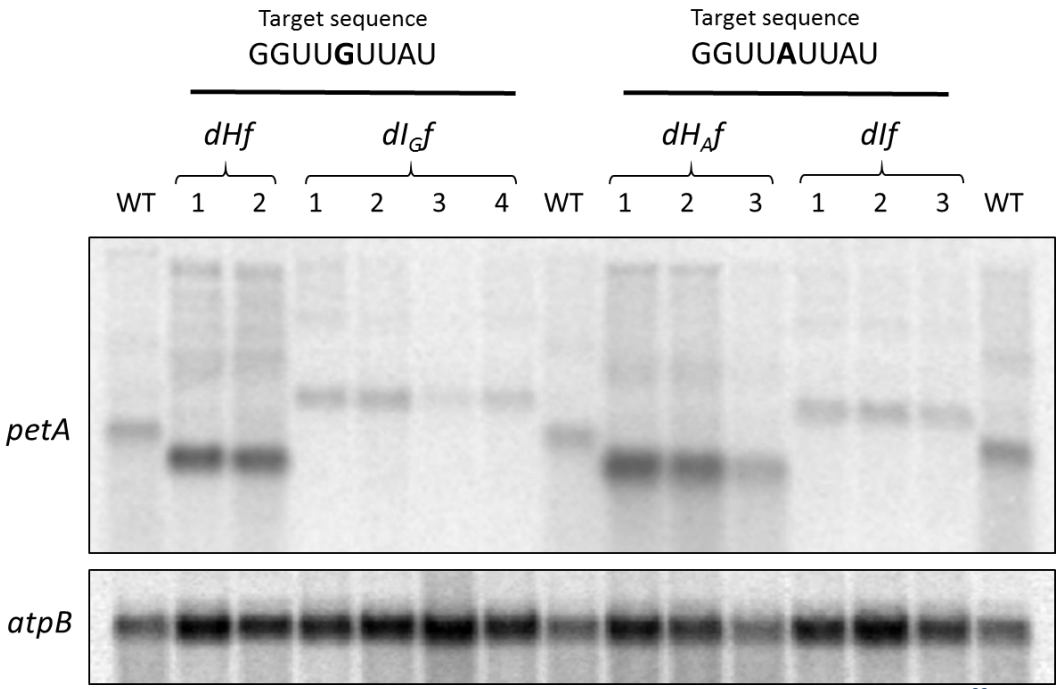


Figure 47: RNA blot of the chimeric constructs, filters were hybridised with ^{33}P *petA* and *atpB* (loading control) radioactive probes.

As seen on Figure 47, the chimeric mRNAs with the 5'*atpH* were shorter than the endogenous *petA* mRNA, while those with the *atpI* 5'UTR were longer. Stronger expression levels were obtained from *atpH* 5'UTR. The difference in the target sequences apparently did not have significant effect on transcript accumulation, when comparing *dIf* and dI_Gf, or *dHf and* dH_Af chimeras. And so, transcript accumulation seems to rely on other properties of the 5'UTR rather than on small variations in the target.

I also compared a mutant strain, $5'_A atpH$, where the MTHI1 target in the endogenous *atpH* gene has been modified to GGUU**A**UUAU with an *atpH* control strain. In this case an influence of the CDS or 3'UTR might emerge. As shown on Figure 48, *atpH* mRNA accumulation was not affected by the modification of its MTHI target. This result again suggested that the one nucleotide difference in the target sequence was not essential for MTHI *atpH* stabilisation mechanism.

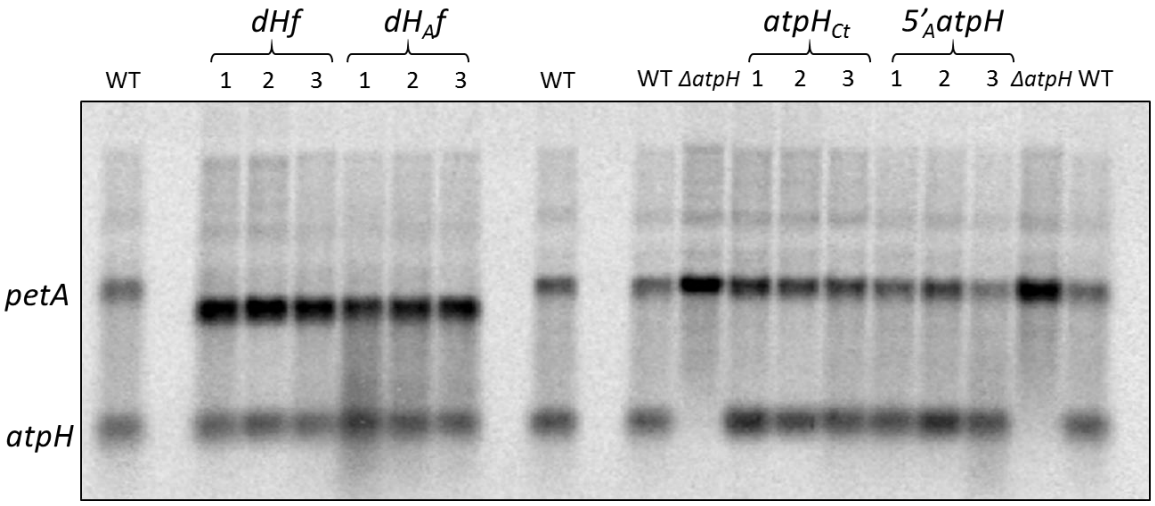


Figure 48: RNA blot of 5'*atpH* transcripts. Filters were hybridised with ^{33}P labelled *petA* and *atpH* probes.

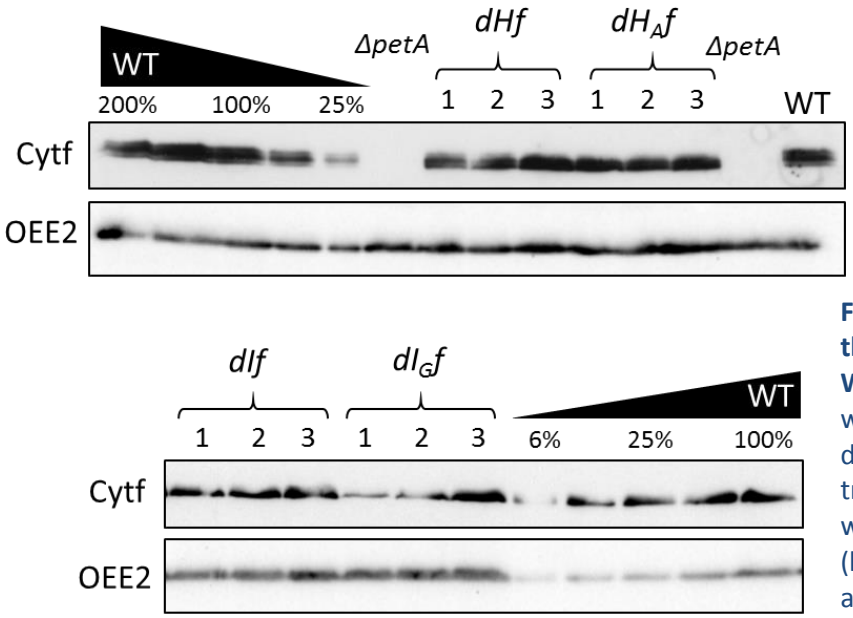


Figure 49: Immunoblots of the chimeric transformants. **W**hole cell protein extracts were separated under denaturising conditions, transferred, and incubated with Cyt. *f* and OEE2 (loading control) primary antibodies.

We also wondered whether the variation in the target sequence influenced protein accumulation, MTHI1 being also a translation activator. Whole cell protein extracts from the various transformants were blotted (Figure 49). The target sequence does not modulate strongly the accumulation of the Cyt. *f* reporter protein. Altogether, it appears that the MTHI1 target sequences alone are not sufficient to induce an *atpI* or *atpH*-like expression pattern of *petA* chimeras. The remainder of the 5'UTR is visibly essential in establishing mRNA stability and its subsequent translation.

Is MTHI1 the only limiting factor for *atpI* and *atpH* expression?

The use of chimeric genetic constructs revealed interesting aspects of *atpI* and *atpH* expression systems. If MTHI1 is present in a limiting amount, and because the expression of both genes relies on MTHI1, we expected that *atpH* or *atpI* deletion strains would allow more MTHI1 to interact with the other transcript, which would be therefore more expressed. And this proved true (see figures 4 and 5 from ***Article 3***).

But a surprising effect was observed with the *atpI* constructs. The *atpH* mRNA is 10 times more abundant than the *atpI* transcripts. Furthermore, MTHI1 binds stably to the 5'end of the *atpH* transcript (as revealed by its footprint) but probably only transiently to the *atpI* mRNA. As such, we would expect far more MTHI1 being sequestrated on the *atpH* mRNA and that lifting this constraint would free many MTHI1 factors to act on *atpI*. Conversely, the deletion of *atpI* should only moderately lighten MTHI1 functional burden. Unexpectedly, more *dIf* transcripts were accumulated in a *ΔatpI* strain than in a *ΔatpH* strain. This implies that other limiting factor(s) could act specifically on the *atpI* 5'UTR and become available for the expression of the chimera when the endogenous *atpI* is absent. This could well be some still unknown *atpI* M factor.

I contributed to those observations by comparing the expression of *dIf* and *dI$_G$f* chimeras in presence of the endogenous *atpI* 5'UTR, or in its absence using instead the *aAdI* construct that places *atpI* expression under the control of *psaA* promoter and 5'UTR. As shown in Figure 50, irrespective of the putative MTHI1 target sequence, the chimeric *dIf* transcripts were far more accumulated when no other *atpI* 5'UTR could compete.

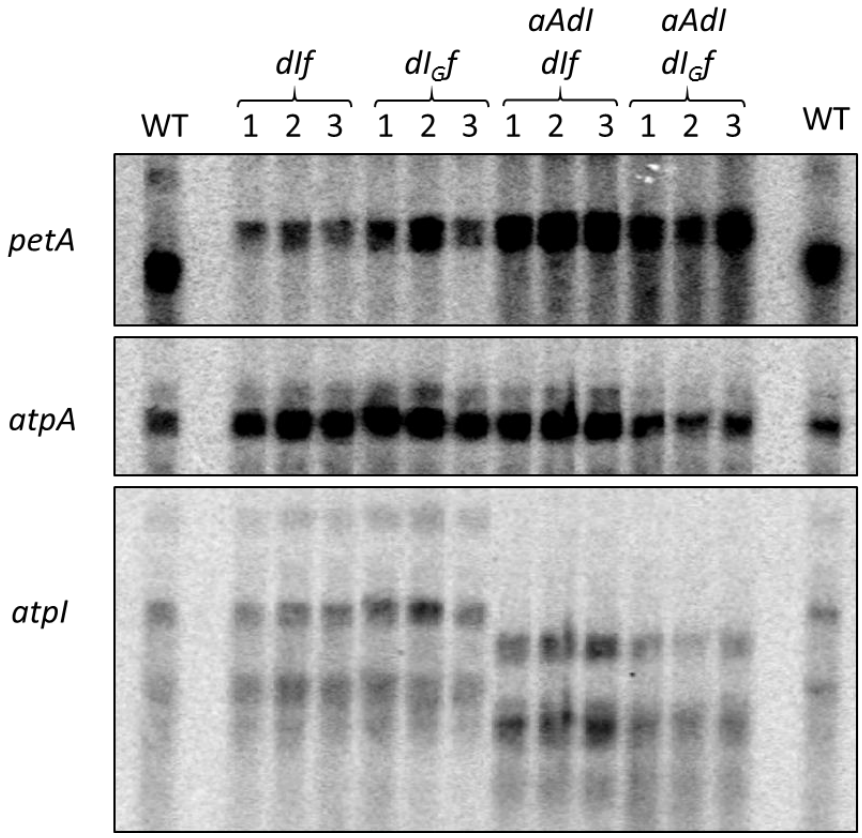


Figure 50: RNA blot of *dlf* transcripts. Filters were hybridised with ^{33}P labelled *petA, atpI* and *atpA (loading control)* probes.

This experiment and those discussed in ***ARTICLE 3*** (attached at the end of this manuscript) proved that some unknown factors act on the *atpI* 5'UTR, possibly a dedicated *atpI* M factor.

Moreover, a new factor, which has not been discussed in the article, seems to be implicated in *atpH* expression.

MTH2, A PUTATIVE FACTOR IMPLICATED IN *ATPH* STABILISATION

L63a, a nuclear mutant generated by insertional mutagenesis by Laura Houille, previously characterised by Dominique Drapier and Shin-Ichiro Ozawa, displayed a lower *atpH* mRNA accumulation, about one tenth compared to a WT strain. But this mutant remained photoautotroph, indicating that the mutated factor was dispensable for AtpH translation. The mutant was crossed, and its *atpH* phenotype segregated with the insertion, proving a genetic link. The strain was sequenced, and the nuclear gene Cre10.g461700 was found to be interrupted by the insertion. Thus, the putative protein product affected in this mutant was named MTH2 and the mutant *mth2-1*.

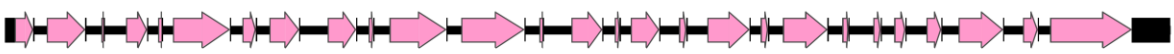


Figure 51: *MTH2* (Cre10.g461700) gene model, its 27 exons are indicated by pink arrows, 26 introns are present in this gene.

MTH2 (Cre10.g461700) is a massive gene encoded in chromosome 10; it contains 27 exons and encodes MTH2, a big protein of 3219 amino acids, with no discernible domains. It is not a helical repeat protein and we do not know if it is able to bind to RNA. It is predicted to be targeted to the chloroplast by WoLF Psort (Horton *et al.*, 2007) and ChloroP (Emanuelsson *et al.*, 1999). MTH2 has orthologues in other Chlamydomonaceae (see alignment in ***ANNEX***).

To ensure that this putative factor was really implicated in *atpH* expression, I characterised another mutant of MTH2. We obtained it from the clip library of insertional mutant (Li *et al.*, 2019). This *mth2-2* was also photoautotroph but displayed a slower growth on minimum medium than a WT strain. After RNA extraction and blotting, I confirmed that the phenotype of *mth2-2* was congruent with that of *mth2-1*: 12% of *atpH* transcript accumulated, compared to a WT strain (Figure 52). The *mthi1-2* strain, as expected, did not accumulate any *atpH* monocistronic mRNA. Unlike MTHI1, required for *atpH* mRNA stabilisation, MTH2 seems to be an auxiliary M factor of *atpH, whose* action on *atpI* expression remains to be studied

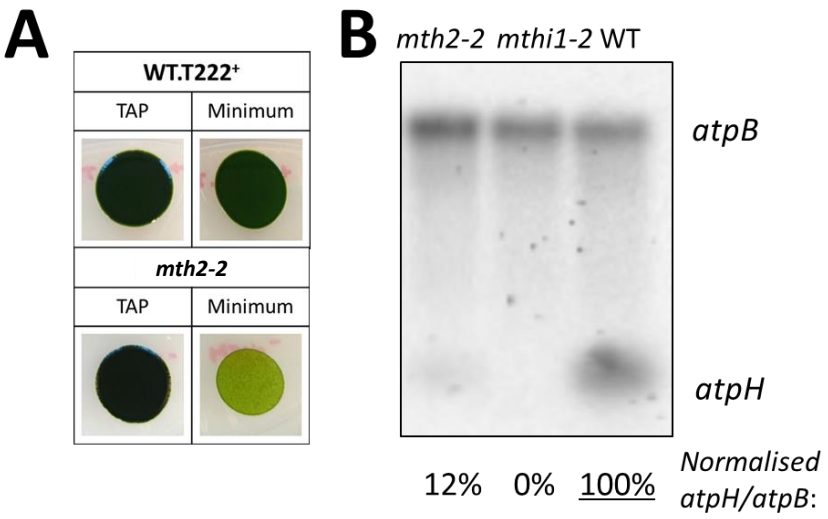


Figure 52: A. Growth phenotypes of WT.T222+ and *mth2-2* B. RNA blot of the *mth2-2* and *mthi1-2* mutants, filter was hybridised with *atpB* and *atpH* dig-dUTP labelled probes. Quantified ratio of *atpH* on *atpB* transcripts normalised on WT levels is indicated below. The quantification was performed with *ImageLab.*

Could this factor act on MTHI1 and help it stabilise the *atpH* transcript? As AtpH is translated in the absence of MTH2, MTH2 seems dispensable for the translation activation mechanisms. Maybe it could then be also dispensable for *atpI* translation. AtpI must be at least slightly translated in the *mth2-1* and *mth2-2* strains since the cells are photoautotroph. But we still do not know whether MTH2 is implicated in *atpI* expression or not.

On the other hand, MTH2 may be a true specific factor for *atpH* expression, which can interact specifically with the *atpH* mRNA or perhaps with another yet unknown factor that can recognise specifically the *atpH* mRNA. Because we do not currently know if MTH2 can bind RNA or proteins, it is for now impossible to distinguish between those hypotheses.

OTHER FACTORS IMPLICATED IN *ATPH* ACCUMULATION?

While working on the main project of my thesis (see Chapter IV), I inadvertently discovered some strange property in one strain. My aim was to look at a *gfp* chimeric transcript accumulation versus *atpB* (see Chapter IV), and I probed *atpH* mRNA as a loading control, its small size making it a convenient control when looking at several transcripts on a same blot. Surprisingly, we noticed than in a *mdb1-2* background *atpH* levels appeared higher, but quantification was impossible, due to the absence of another independent loading control. However, in all *mdb1-2* transformants probed over several blots, the accumulation of the *atpH* transcript appeared stronger (Figure 53).

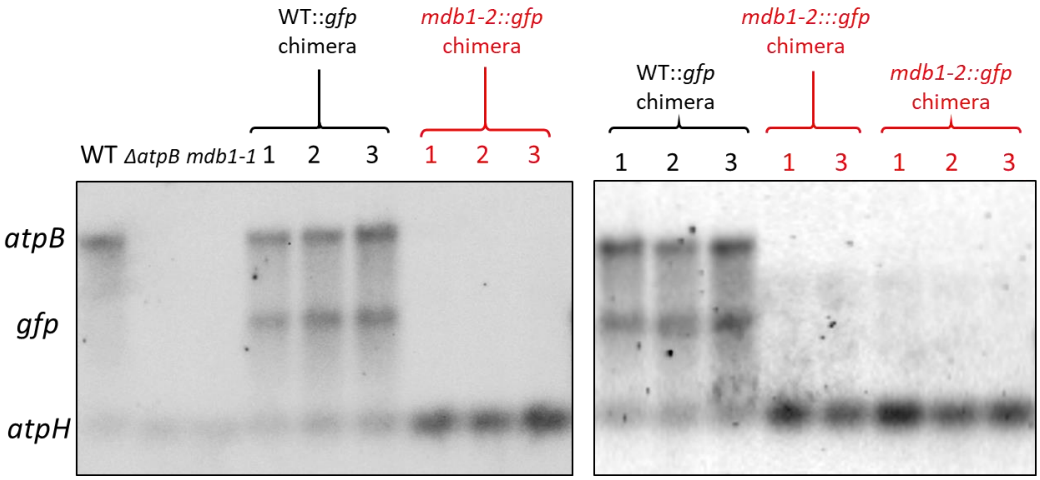


Figure 53: Two fragments of RNA blots, presented in detail in chapter IV. Filters were hybridised with *atpB*, *gfp* and *atpH* dig-dUTP labelled probes. *atpH* was intended as a loading control but was surprisingly more accumulated in a *mdb1-2* background

Unfortunately, I could not test yet if this is also the case in the simple *mdb1-2* strain. Data from Dominique Drapier (Figure 54) seems to suggest so. However, no interconnection between *atpH* and *atpB* transcript levels has ever been observed so far. The difference would be difficult to attributes to the chimeric insertion, which is in a neutral locus of the chloroplast genome, near *petA*, and has no *atpI* or *atpH* related sequence. Moreover, this strong accumulation of *atpH* transcript did not seem to occur in *mdb1-1::gfp chimera* transformants in several technical repeats.

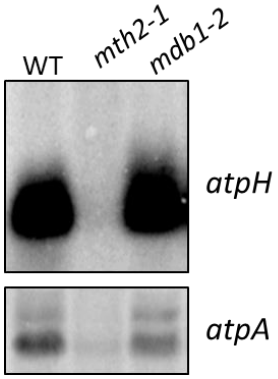


Figure 54: RNA blot from Dominique Drapier, unpublished.

I thus wonder whether the difference in the *mdb1-2* background might be caused by its large deletion in chromosome 14. Indeed, this strain lacks six genes around *MDB1* (Cre14.g614708, Cre14.g614667, Cre14.g614650, Cre14.g614600, Cre14.g614500, Cre14.g614450 and Cre14.g614400) (Figure 55). 5 of them have an unknown function, and of those, 3 are predicted to be targeted to the chloroplast by WolF Psort (Horton *et al.*, 2007): Cre14.g614450, Cre14.g614600 and Cre14.g614667.

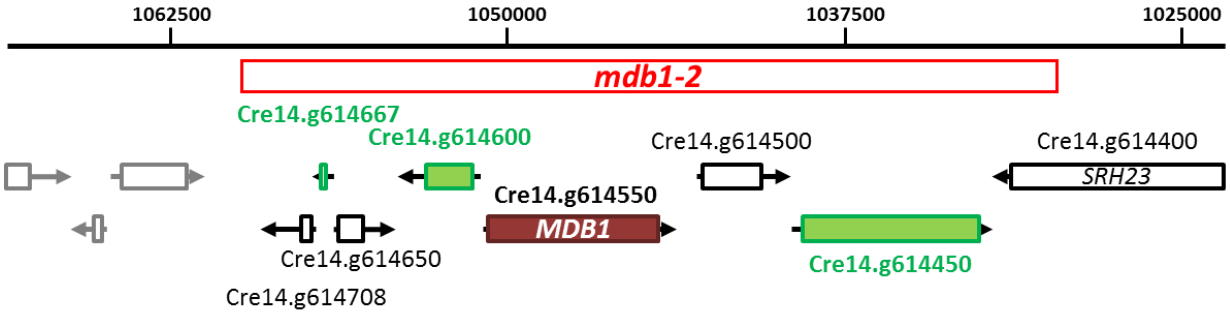


Figure 55: Genes deleted in *mdb1-2*, genes in green are predicted to be targeted to the chloroplast by WoLF Psort.

Those genes could potentially be linked to *atpH*, perhaps by acting in *atpH* transcript degradation. Or they might instead be linked to *atpI* expression. In this case, their absence might indirectly free more MTHI1 for *atpH* mRNA stabilisation. To answer those questions a first step would be to look in detail at *atpI* and *atpH* transcripts accumulation patterns in this *mdb1-2* strain. Then if a difference with the *mdb1-1* strain is confirmed, mutants of the putative genes could be looked at, to assess whether they are implicated in *atpH* or *atpI* expression.

MTHI1 IS IMPLICATED IN *ATPI* MRNA STABILITY

We knew that MTHI1 was necessary for *atpI* translation, and that *atpI* mRNA was also less accumulated in its absence. From there two main hypotheses were drawn (Figure 56). Either:

- The *atpI* mRNA is protected from degradation by the ribosomes when translated, as occurs in bacteria.
- Or, much like MCA1 and TCA1 (Loiselay *et al.*, 2008; Boulouis *et al.*, 2011), the M and T factors would co-stabilise the mRNA in a ternary complex.

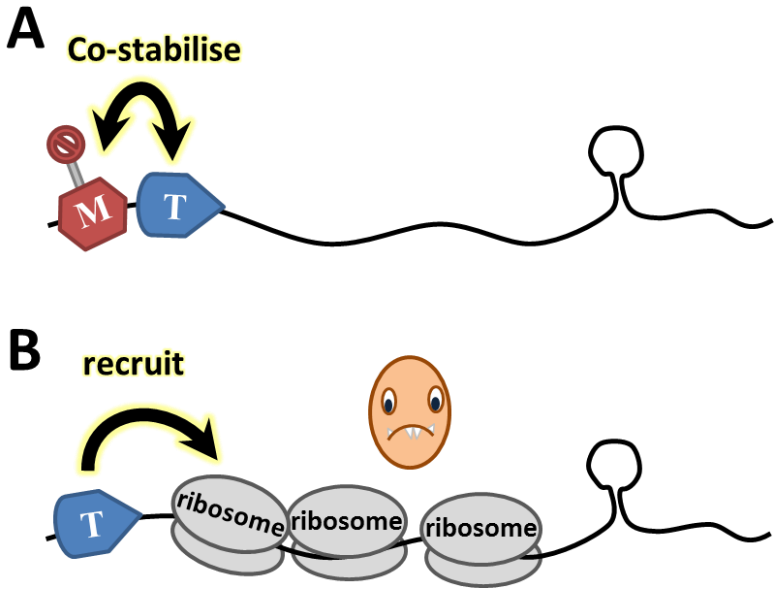


Figure 56: Two hypotheses to explain the smaller levels of a chloroplast transcript when its T factor is absent. A. Ternary complex of M and T factors on the mRNA **B.** Translating ribosomes protects the mRNA from its degradation by endonucleases.

To determine whether the destabilisation of the *atpI* mRNA is due to the lack of translation or directly to the absence of MTHI1, we rendered *atpI* untranslatable even in the presence of this later by creating an *atpI* construct where the start codon was replaced by a stop codon (Figure 57). A drop in *atpI* mRNA would infer that translation is directly implicated in the transcript stability.

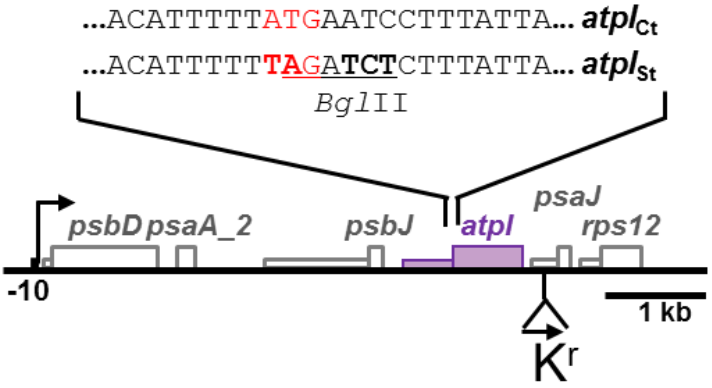


Figure 57: Mutation of the ATG start codon of *atpI*, replaced by the TAG stop codon.

This construct was transformed into the *ΔatpI* strain. As expected, the *atpI*$_{Ct}$ control transformants (carrying the selection marker without the *atpI* mutation) recovered phototrophy whereas the *atpI*$_{St}$ transformants did not, suggesting they cannot translate the *atpI* mRNA (Figure 58, B). No significant difference in *atpI* transcript accumulation was observed between the control and *atpI*$_{St}$ constructs (Figure 58.A).

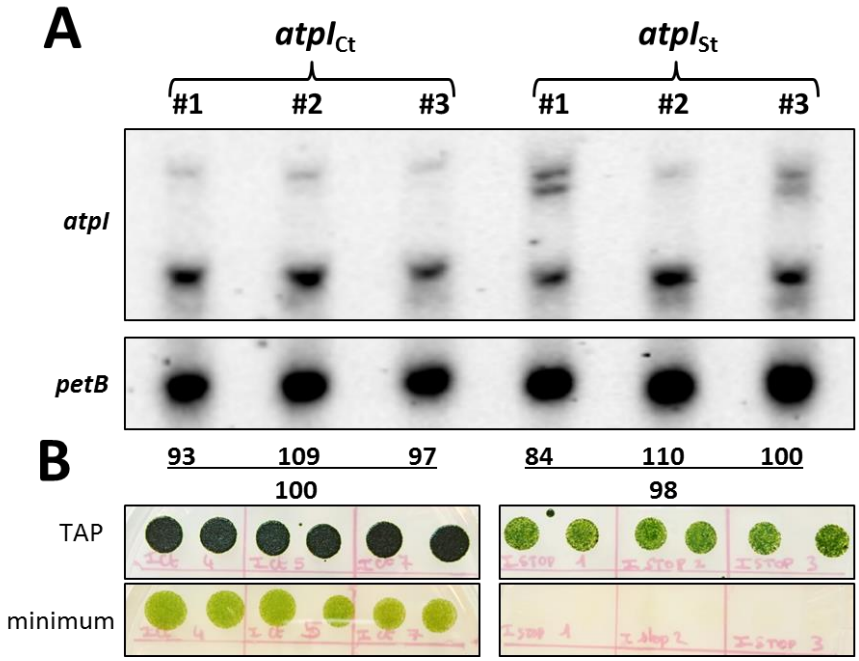


Figure 58: A. RNA blot of the *atpI*$_{ct}$ and *atpI*$_{st}$ mutants, filters were hybridised with dig-dUTP labelled *atpI* and *petB* probes. Three independent transformants were analysed. **B. corresponding growth tests** (two droplets were made for each transformant).

Thus, lack of translation of the *atpI* transcript was not responsible for its destabilisation. We thus infer that MTHI1 is directly implicated in the stabilisation of this mRNA, perhaps by contributing to a complex with some presently unknown *atpI* M factor.

MTHI1 IS IMPLICATED IN *ATPH* TRANSLATION

The more we studied MTHI1 the more it seemed implicated in several functions. Considering its dual partial stabilising effect on *atpI* mRNA and its necessity for translation, a potential implication of MTHI1 in *atpH* translation needed to be investigated. To assess whether MTHI1 is a T factor for *atpH*, a first hurdle needed to be passed: the absence of the *atpH* transcript in absence of its M factor. To this end, a polyG track, was used. PolyG tracks are voluminous structures that block the path of exonucleases on RNA (Drager *et al.*, 1998). By putting one in *atpH* 5'UTR the transcript could become constitutively stable, even in the absence of MTHI1.

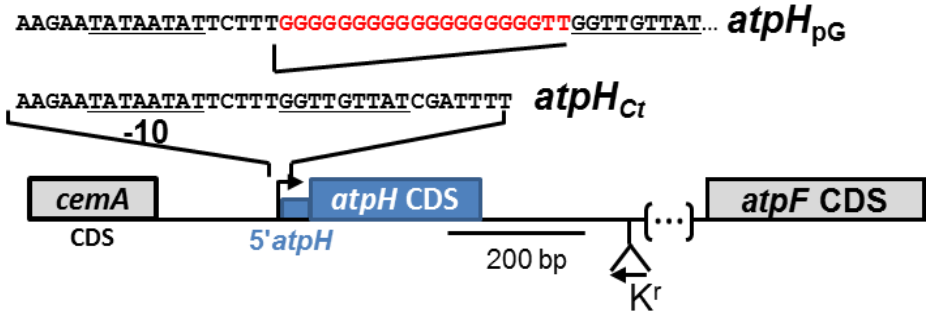


Figure 59: Insertion of a polyG track in *atpH* 5'UTR, directly before the MTHI1 target.

This construct (Figure 59), followed by an excisable 5'*psaA-aadA* cassette (Fischer *et al.*, 1996; Wostrikoff *et al.*, 2004) was transformed in the WT.T222^{+}, *mthi1-1* and *mthi1-2* strains. However, the *mthi1-1* mutants proved to be prone to phenotypic reversion, and I worked only with the *mthi1-2* mutant. Initial analysis of the transformants by RNA blot showed that *atpH* transcript accumulation was restored in the *mthi1-2* strain (Figure 60.A). However, the accumulation of AtpH in immunoblots was very weak, even in the WT {atpH$_{Ct}$} transformants (Figure 60.B). We suspected that this could be due to the disruption, caused by the *aadA* cassette, of the expression of the downstream located *atpF* gene, co-transcribed with *atpH*. Indeed, CFo subunits accumulate in a concerted fashion. If AtpI or AtpF are absent, unassembled AtpH cannot accumulate.

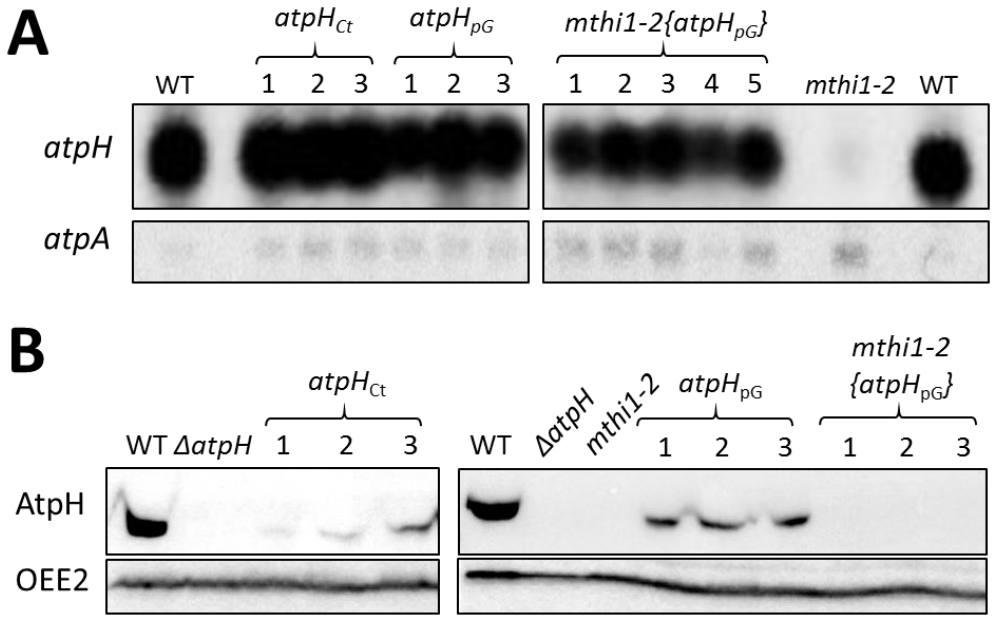


Figure 60: Expression of *atpH* in the strains prior to *aadA* cassette excision. A. RNA blot, filter was hybridised with ^{33}P *atpH* and *atpA* (loading control) radioactive probes. B. Immunoblots of whole cell extracts, anti AtpH and OEE2 (loading control) primary antibodies were used.

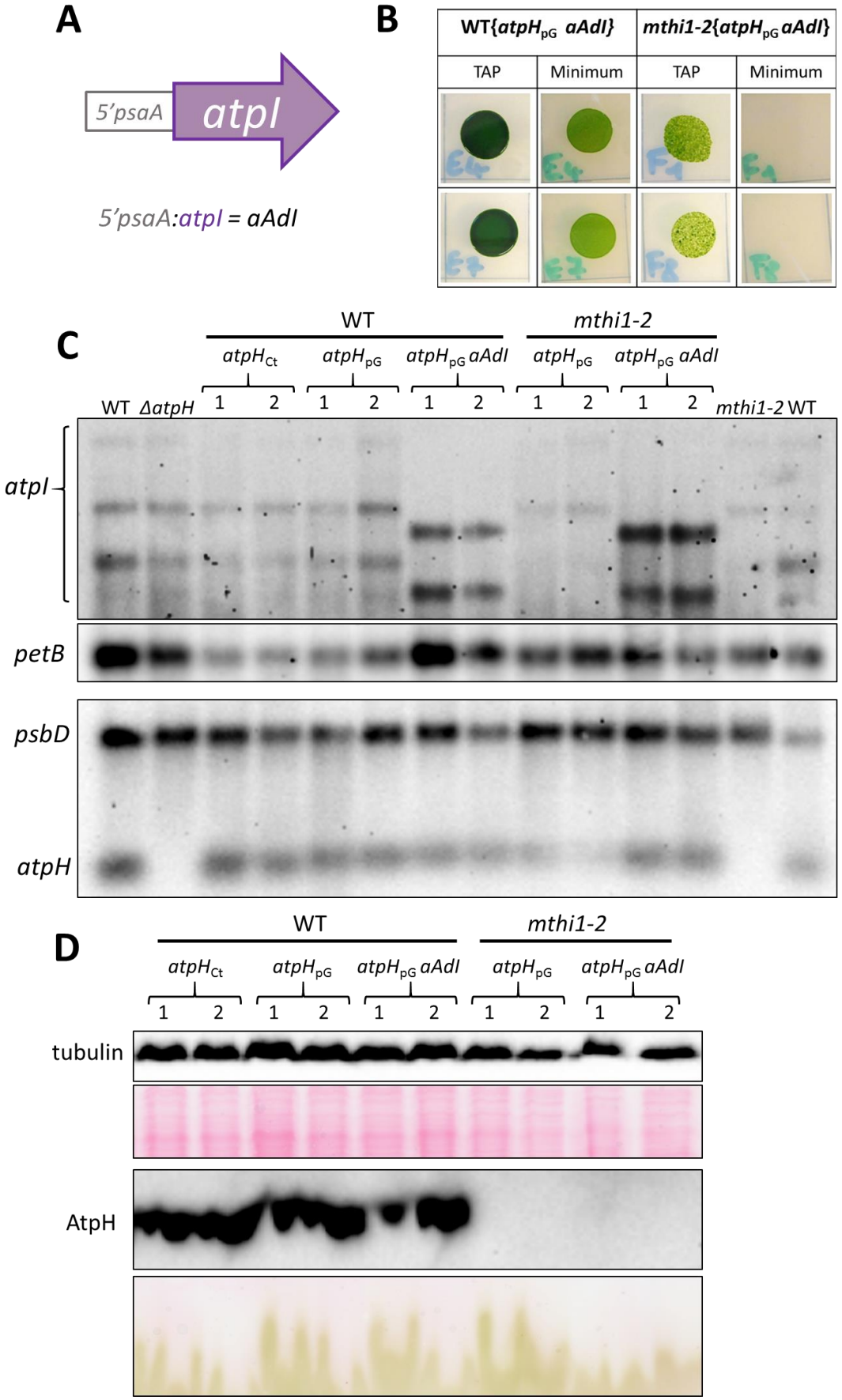


Figure 61: *atpH* expression after *aadA* cassette excision A. cartoon of the *aAdI* construct **B.** Growth test **C.** RNA blot, filters were hybridised with dig-dUTP labelled *atpI*, *petB*, *atpH* and *psbD* DNA probes. **D.** Immunoblot, anti-tubulin and AtpH primary antibodies were used. Corresponding regions of the filter in Ponceau red stain are under the chemiluminescent signals.

We attempted to check *atpF* mRNA accumulation in RNA blots, but it proved very difficult to observe, owing to its low accumulation and to its smeary distribution. And so, we decided to excise the selection cassette.

After excision of the cassette, I extracted once more the mRNA and proteins from the mutant strains and monitored *atpH* expression. As can be seen in Figure 61, AtpH accumulated in both the control WT {$atpH_{Ct}$} and WT {$atpH_{pG}$} mutants, whereas it still was under detection level in *mthi1-2* {$atpH_{pG}$}.

However, another problem to overcome to study *atpH* in absence of MTHI1 lies in *atpI* expression. As previously explained, AtpI is not synthesised in absence of MTHI1. And so, we cannot infer from the previous results whether AtpH is not translated or if it is degraded in absence of its partner subunit AtpI. ^{14}C pulse labelling experiments of the strains were done but were inconclusive due to the tendency of AtpH to co-migrate with chlorophylls, muddling the signals.

We then decided to produce AtpI independently from MTHI1; using the *aAdI* construct (Figure 61.A). Both WT {$atpH_{pG}$} and *mthi1-2* {$atpH_{pG}$} were transformed with *aAdI* at the *atpI* endogenous locus. *aAdI* permitted AtpI synthesis in the control transformants and in *mthi1-2* (Figure 61.B). RNA blot analysis confirmed the expression of *aAdI* in both strains, detected as shorter forms of *atpI* mRNA (Figure 61.C). In the *mthi1-2* background, $atpH_{pG}$ transcripts still accumulated but AtpH was not detected in immunoblots, confirming the absence of phototrophy of the strain in growth tests (Figure 61.D).

Thus, AtpH is not synthesised in the absence of MTHI1, even when the *atpH* transcript does accumulate while the AtpI and AtpF subunits are synthesised. Therefore, MTHI1 appears essential for *atpH* translation, and is both an M and a T factor for *atpH*.

Most of the aforementioned work has been included in a wider study of the MTHI1 factor: ***ARTICLE 3***, an article under review, which is attached at the end of this manuscript.

DISCUSSION

MTHI1, A KEYSTONE TO BOTH ATPI AND ATPH EXPRESSION

Altogether, we confirmed that MTHI1 is truly a bifunctional factor, essential for both *atpH* and *atpI* expression.

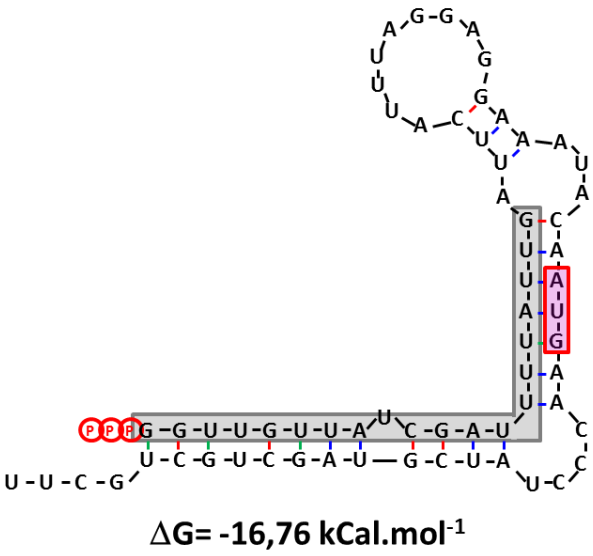


Figure 62: The lowest energy structure calculated at 25°C by RNA Folding Form (M-Fold: (Zuker, 2003)) of *atpH* 5'UTR and the first 25 nt of its CDS. The footprint of MDH1 is grey-shaded, while the *atpH* initiation codon is pink-shaded

It is necessary both to *atpH* mRNA stabilisation and to its translation. Structure prediction program M-Fold (Zuker, 2003) proposes that a secondary structure sequestrating the initiation codon of *atpH* might exist (Figure 62). MTHI1 might act by opening this secondary structure when binding to its target sequence. Most interestingly, a similar mechanism has been described in maize (Prikryl *et al.*, 2011) for PPR10, a PPR factor, also stabilising and activating the translation of *atpH* mRNA (Figure 63). This convergent evolution of proteins from evolutionary distinct family could indicate that this mechanism is especially suited to *atpH* expression. As AtpH is needed in great quantities for the ATP synthase assembly, perhaps this mechanism could allow the chloroplast to produce rapidly a large amount of AtpH when the need arise, with the induction of a single nuclear protein.

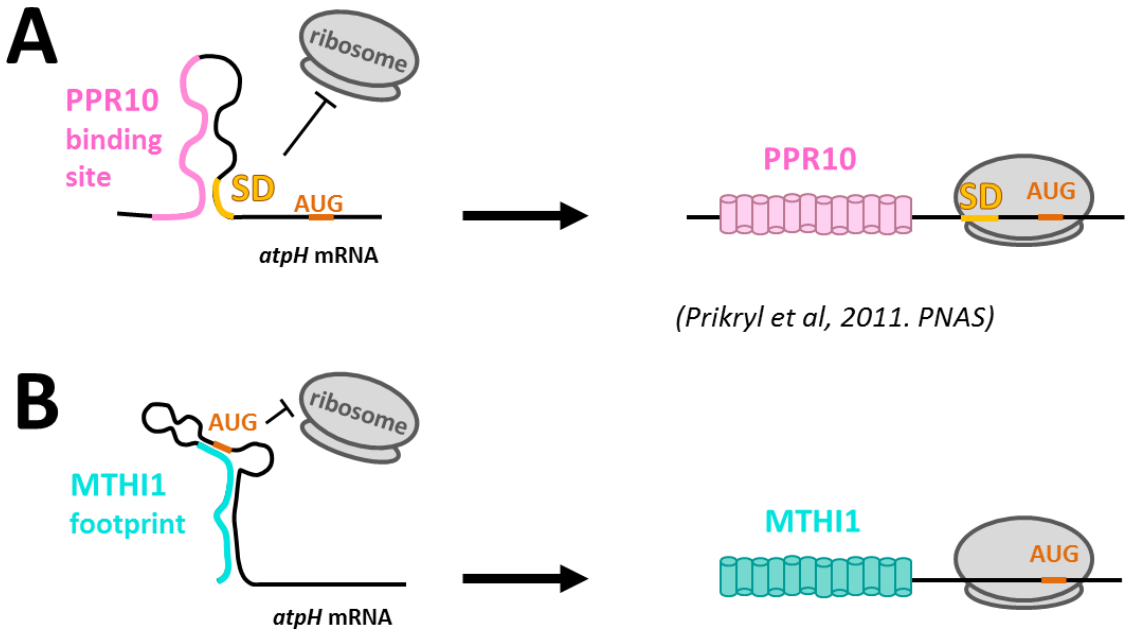


Figure 63: A. Mode of action of maize PPR10, redrawn from (Prikryl *et al.*, 2011).
B. Putative mode of action of MTHI1.

MTHI1 is also important for *atpI* mRNA translation and directly implicated in its stability. This is effectively a way for the cell to activate the expression of both subunits together. MTHI1 is strongly expressed at the beginning of the day in circadian gene expression studies (see figure 14 from ***ARTICLE 3***), certainly to prepare the biogenesis of the photosynthetic apparatus (Zones *et al.*, 2015). Accordingly, MTHI1 is more expressed at the onset of light that in the dark, a pattern that is attenuated in a *hmox1* null mutant, deficient in bilin biosynthesis. That mutant barely grow under a diurnal 12h dark/12h light cycle (Duanmu *et al.*, 2013; Wittkopp *et al.*, 2017). MTHI1 could be among key nuclear genes that are upregulated by a retrograde signalling mediated by bilins. MTHI1 appears to be a true regulator, implicated in chloroplast/nucleus cross talks.

MTHI1 MIGHT WORK WITH OTHER FACTORS

The different functions of MTHI1 do not seem established thru the differences of the target sequences of MTHI1. No variations in expression of the transcripts, either in stabilisation or translation, could be detected after swapping the targets. Moreover, in the chloroplast transcriptome the GTTGTTAT sequence is found 3 times (*atpH*, *rpoC2* and *rpoC1*) GGTTATTAT is found 2 times (*atpI* and *rps3*), and of all those sequences, only the *atpH* one is recovered in a MTHI1 footprint. No other sRNA than *atpH* was recovered either in MTHI1-RIP experiments.

It appears that the 5'UTRs of the mRNAs bear additional signals that control their expression and induce the strong affinity of MTHI1 for *atpH* and its effects on *atpI*. Competition between chimeric reporter constructs and endogenous genes revealed that MTHI1 was not the only limiting factor for *atpI* expression. While tweaking the MTHI1 target sequence did not affect the stability of the chimeric mRNA, the absence of a concurrent *atpI* 5'UTR increased the accumulation of 5'*atpI-petA* chimeras, much more than the absence of *atpH*. This suggests that at least one other specific factor acts on *atpI* mRNA 5'UTR, presumably a specific M factor, as a foot-print is found at the beginning of the mature *atpI* transcript (figure 10 from ***ARTICLE 3***). But this factor probably relies on MTHI1 quite a bit to stabilise *atpI* mRNA, since MTHI1 is truly implicated in the stabilisation process.

As for *atpH*, we characterised other mutants in a putative MTH2 protein, which display a drop in *atpH* mRNA accumulation to about 10% of the WT level but are still photoautotrophic. Such a factor could be an auxiliary M factor of MTHI1. Alone, MTH2 is not able to rescue any accumulation of *atpH* mRNA, but its presence could help stabilise MTHI1 on *atpH* mRNA. In fact, we do not know if MTH2 can bind RNA. It could well be a specific factor recognising a sequence in *atpH* mRNA or it might instead interact with yet another factor(s), specific to *atpH* mRNA.

The case for multifactor complexes has been strengthened biochemically. MTHI1 was found in high molecular weight fractions in size exclusion chromatography experiments (figure 9 from ***ARTICLE 3***). The action of RNases on the samples dissociated the complexes and produced monomeric MTHI1, certainly tethered to its mRNA footprints. The MTHI1 complexes of high molecular weight could form in presence of either *atpH* or *atpI* transcripts but shifted to even larger fractions in absence of both, presumably because MTHI1 aggregates. Those complexes formed upon *atpI* or *atpH* mRNA could involve MTHI1 and a suite of partner factors. Possible candidates were uncovered in the *mdb1-2* strain and could perhaps be implicated in *atpI* or *atpH* expression. Further studies are required to determine if they are linked or not to MTHI1 functions.

TOWARD A GENERAL MODEL OF OPR PROTEINS ACTION?

The present results point toward the implication of several actors to induce the expression of *atpH* and *atpI*. MTHI1 seems to have a relatively weak affinity for its targets but the implication of other specific factors could anchor MTHI1 on *atpH* mRNA and induce its actions on *atpI* mRNA. Such models have been described previously for other OTAF; MCA1 and TCA1 that interact together in a ternary complex on *petA* mRNA. When MCA1 is absent no stabilisation of the transcript is possible, but even if the transcript is artificially stabilised, its translation is drastically reduced. And when TCA1 is absent, *petA* mRNA is partially destabilised and only accumulates to about 15% of its normal level (Wostrikoff *et al.*, 2001; Raynaud *et al.*, 2007; Loiselay *et al.*, 2008; Boulouis *et al.*, 2011). When its partner is absent the other factor must rely only on its own specificity for the target sequence and thus induce a weaker action on the mRNA. Other described ternary complexes include MDA1 with TDA1 on *atpA* mRNA (Viola *et al.*, 2019) and MBD1 (NAC2) with Rp40 on the *petD* 5'UTR (Schwarz *et al.*, 2007).

The case of MTH2 is even more subtle than those ternary M/T/mRNA complexes; it is auxiliary to *atpH* expression and its action, not well characterised as of now, could be to act on MTHI1. It would be very interesting to study this MTH2 factor in more details and find out whether it interact with MTHI1 or even if it might, after all, bind RNA, or maybe act in a completely different way. As this protein does not contain evident domains, the search for this kind of factor in *C. reinhardtii* genome would be very difficult. Furthermore, such auxiliary factors are expected to induce subtle phenotypes when mutated, rendering their discovery even more complicated. It is impossible to say for now if those auxiliary factors are rare or common partners of OTAFs.

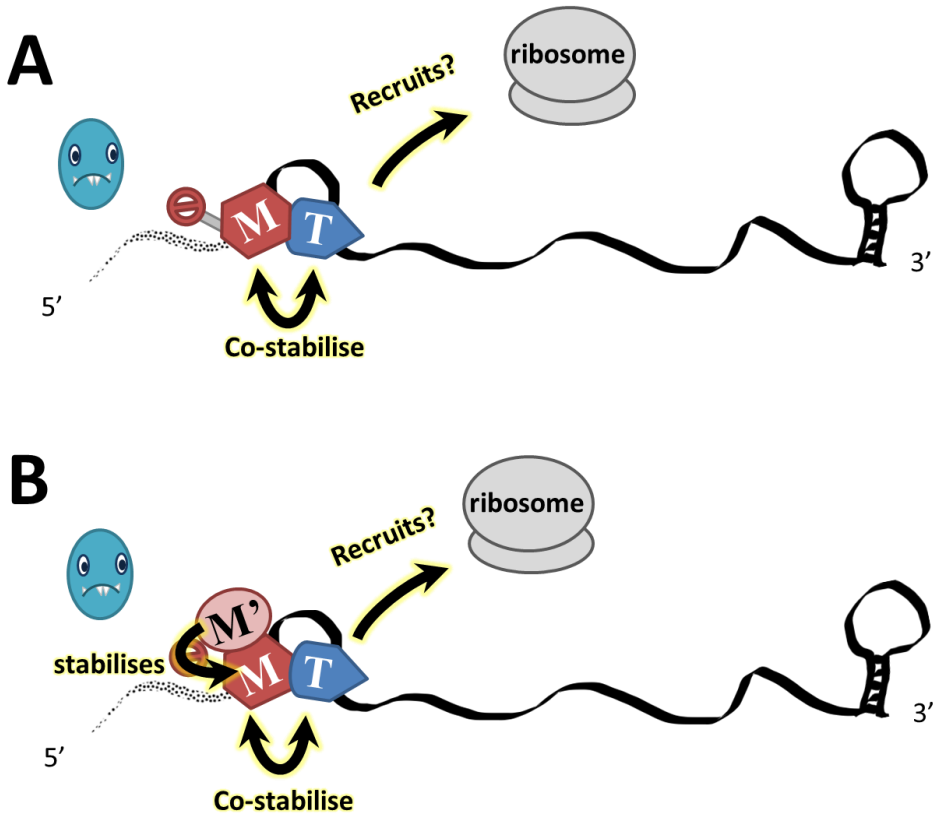


Figure 64: Models of OTAFs interaction on mRNA. A. Ternary complex of M and T factors on mRNA **B.** Higher order complexes could be common in the organelles.

Ternary or higher order multi-factor systems, relying on different proteins with modest affinity for a target mRNA could, when assembled together, display a strong specificity for the transcript and a higher resilience. Small mutations in one factor could be compensated by the others. But nonetheless it still appears that some factors are vital for expression and their absence cannot be compensated by auxiliary factors. Another important step would be to study the non-OPR domains of OPR proteins to understand how they might interact with other proteins, or how they activate translation.

We will see in the next chapters how those complicated complexes might greatly influence the specificity of OPR factors. And how studying mRNA/OPR interactions *in vivo* brought valuable new data on M factors and led us to radically rethink their roles in a wider scheme.

III. Initial studies of MDB1 and MTHI1 interactions with their target mRNA *in vivo*

Modified from Dartmouth Electron Microscope Facility, Dartmouth College

INTRODUCTION

THE "OPR CODE"

To better understand how the OPR factors establish their specificity in *Chlamydomonas reinhardtii* we set out to study the "OPR code". Starting from the amino acid sequence of an OPR repeat, the OPR code would predict which nucleotide should be recognised, much like the PPR code established seven years ago (Barkan *et al.*, 2012; Kobayashi *et al.*, 2012; Yagi *et al.*, 2013; Kobayashi *et al.*, 2019).

Improving target sequence prediction would allow the study of cryptic OPR proteins, which might have subtle roles, difficult to spot in phenotypic characterisation, notably auxiliary factors or factors acting on non-essential chloroplast products. It would ease particularly the study of T factors; not only are they difficult to recover in footprint experiments, they appear to be enriched in certain repeats that are currently "unreadable". It could also make it possible to assign some OPR factors to essential genes that are lethal when not expressed. This could also create opportunities to design and build artificial OPR protein as has been discussed and started with PPR proteins (Filipovska and Rackham, 2013; Coquille *et al.*, 2014; Yagi *et al.*, 2014; Spahr *et al.*, 2018).

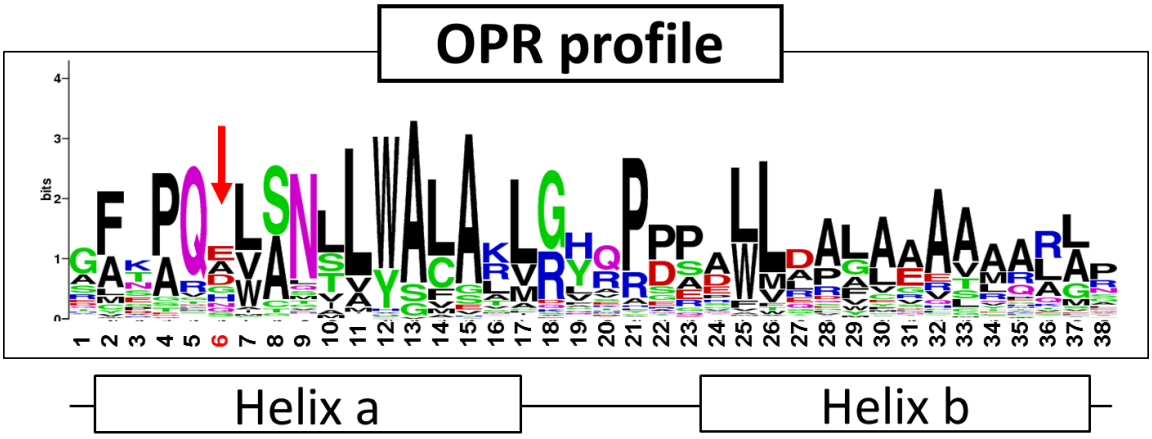


Figure 65: The consensus sequence of the OPR repeats found in photosynthetic organisms. The taller the residue, the most abundant it is. Position of the two putative α-helices is indicated under the consensus. The red arrow indicates the 6th amino acid, which is expected to be crucial for the specific interaction with nucleotides of the target mRNA.

A preliminary version, a "draft" OPR code, based on known OPR/mRNA pairs had been established by Yves Choquet before I joined our laboratory. When looking at the variability of residues in the repeat and at the predicted structure, it appeared that the sixth residue, which should be exposed in the groove of the super helix, could be critical to establish specificity. The occurrence of amino acids at this sixth position was then linked to corresponding nucleotides in available characterised binding sites. This residue is most often polar which could allow the formation of hydrogen bounds with the Watson-Crick face of the bases, as has been observed with PPR motifs. Alternatively, if the 6th residue is apolar, it tends to be a small one: glycine or alanine, which would leave enough space to fit even pyrimidine (A or G) without inducing steric clashes. The nature of the residues in the 3rd, 4th and 5th position could also be implicated in the recognition mechanism. Prolines in position 4, for instance, create a turn, tweaking the structure quite differently.

But this draft code remains theorical and uncomplete. This is due in part to the relative "secrecy" of OPR factors. Only few laboratories have studied OPR proteins so far, making the sample size of validated mRNA/OPR pairs quite small. As can be noticed in Table 3 the observation of known OPR/mRNA could not resolve which combination should recognise C. Moreover, while some combinations show quite a robust correlation with precise nucleotides, some combinations have not been attributed a preferred nucleotide, as the characterised OPR/nucleotide sample had several equally represented bases. Those "unreadable" repeats are abundant in T factors, which make their study even more difficult. Overall, we needed to gather molecular data to substantiate this draft OPR code.

Position	Residue										
3	X	X	X	X	R, K	X	X	X	X	X	X
4	X	P	X	X	X - P	X	X	X	X	X	X
5	X	Q	X - R, K	R, K	X	R	X - R	R, Q	X	R	R
6	E	G	D	D	D	Q	Q	A	H	S	N
Recognised nucleotide	U	A	G	U	U	U	?	A	?	A	?

Table 3: The draft OPR code

To study the affinity of an OPR repeat for a nucleotide, a strong interaction should be favoured. More subtle interactions could prove difficult to untangle at first. Following this line of though, M factors, which are known to bind strongly on their target mRNA are a safer choice that T factors for which footprints have never been recovered so far, and which might interact only transiently or weakly with their target transcript. And so, to study this recognition code I worked with two M factors that have been presented at length in the first and second chapters: MDB1 and MTHI1.

To exemplify how the draft OPR code works, here I used it on MDB1 and MTHI1 OPR motifs, to try to predict their binding sequence (Table 4). When the draft OPR code does not have enough examples of a specific combination to draw a robust correlation, the combination was excluded. In consequence this version of the draft code is restrictive, notably for the recognition of adenine.

		Residues: 12345**6**789…	Prediction:	Actual target sequences:
MTHI1	OPR1	NYDSWDTTL	G	G
	OPR2	IFKPVDCAN	G	G
	OPR3	DYAPGEVCQ	U	U
	OPR4	GYGTQELGM	U	U
	OPR5	HMAPGDIAI	G	G/A
	OPR6	KCRSSELCS	U	U
	OPR7	GLSHHEVAT	U	U
	OPR8	SFSPQGLAM	A	A
	OPR9	AFKPLELSQ	U	U
MDB1	OPR1	QPPPSSLQA	a	A
	OPR2	VYHSCAALS	a	A
	OPR3	ELHPRAVVV	A	A
	OPR4	ALQPRGLAS	a	A
	OPR5	RFAPREVST	U	U
	OPR6	SFCGRSLSN	A	A
	OPR7	AFSPQGLTQ	A	A
	OPR8	PYNGLDLST	G	G
	OPR9	RLEANQLCN	?	C
	OPR10	YFKPVDLSQ	G	G
	OPR11	MFSPVEVAN	U	U
	OPR12	SFKPQELCS	U	U
	OPR13	SMSGWCLAT	?	A

Table 4: Prediction of the recognised nucleotide for each OPR motif of MTHI1 and MDB1 following the draft OPR code. In red is the 6th residue, deemed crucial, in orange the residues modulating the recognition.
Predicted nucleotides in lowercase correspond to lesser known repeats which have a well characterised 6th residue, but with a rare 5th residue.

As *C. reinhardtii* is a great model where both nuclear and plastid genomes are transformable, this study was performed *in vivo*. Less was known at the time on the collaborative nature of OTAF factors in *C. reinhardtii* and the impact they might have on mRNA stabilisation. The strong specificity of the M factors for their target and their essential nature implied a stringent relationship, based on precise molecular properties. But while working *in vivo* entails unexpected difficulties to decipher the interaction of the M factor with its target, *in abstentia* of the other interacting factors, it brings a lot of information on how these factors work physiologically. Moreover, OPR proteins tend to aggregate when overexpressed and their purification remains laborious so far, which further spurred us to tackle this question directly *in vivo*.

HOW TO "CRACK" THE CODE

The directive idea of our strategy is similar to that developed in (Barkan *et al.*, 2012). First: modify the target mRNA and look at its interaction with the M factor, second: modify the M factor following the putative code and look again to see if the interaction is restored (Figure 66).

This project is thus based first on chloroplast transformation to introduce a mutated RNA target. As we work with subunits of the chloroplast ATP synthase, the interruption of the M factor/transcript interaction should render the cells non-photo autotrophic. Then nuclear transformation would be used to introduce the modified M factor. If it can restore the accumulation of the mRNA and rescue phototrophy of the chloroplast mutant strains, the recognition combination would be validated. Our *in vivo* proxy to assess the strength of the interaction will be the accumulation level of the target transcript, which depends on its stabilisation and so on the binding affinity of the OPR factor for the RNA sequence. Looking at the effective stabilisation of mutant transcripts by mutated OPR repeats we could draw conclusions for their preference: which nucleotides are stably recognised by the modified repeat? Is this specificity the one expected from the putative code?

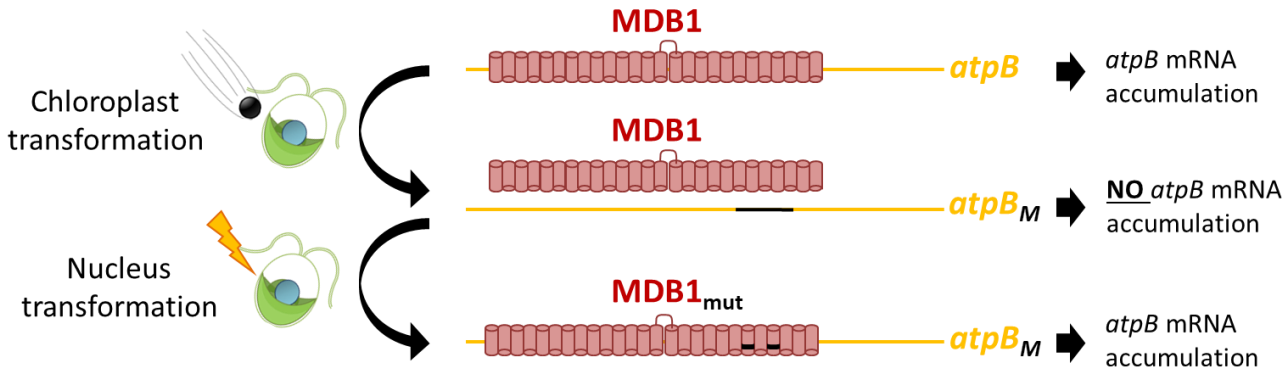


Figure 66: Our initial strategy to study the interaction of the OPR M factor MDB1 with its cognate *atpB* binding sequence *in vivo*.

In this chapter we will focus on the first step of this approach, where we discovered previously unsuspected properties of OPR protein/mRNA interactions.

MUTAGENESIS OF *ATPB* MDB1 TARGET SEQUENCE

THE FIRST ATTEMPT: MODERATE MUTATIONS

At the start of the project, I begun by transforming four mutated targets of MDB1 at *atpB* endogenous locus, together with a spectinomycin resistance cassette to select the transformants. Since the *ΔatpB* strain that was available at the time was already spectinomycin resistant, the constructs were transformed by biolistic in the WT.T222 strain. After homoplasmisation by successive sub-cloning, the presence of the mutations and the absence of the WT version of the target were checked by PCR and sequencing (Eurofins). Three independent transformants were selected for each version of the target.

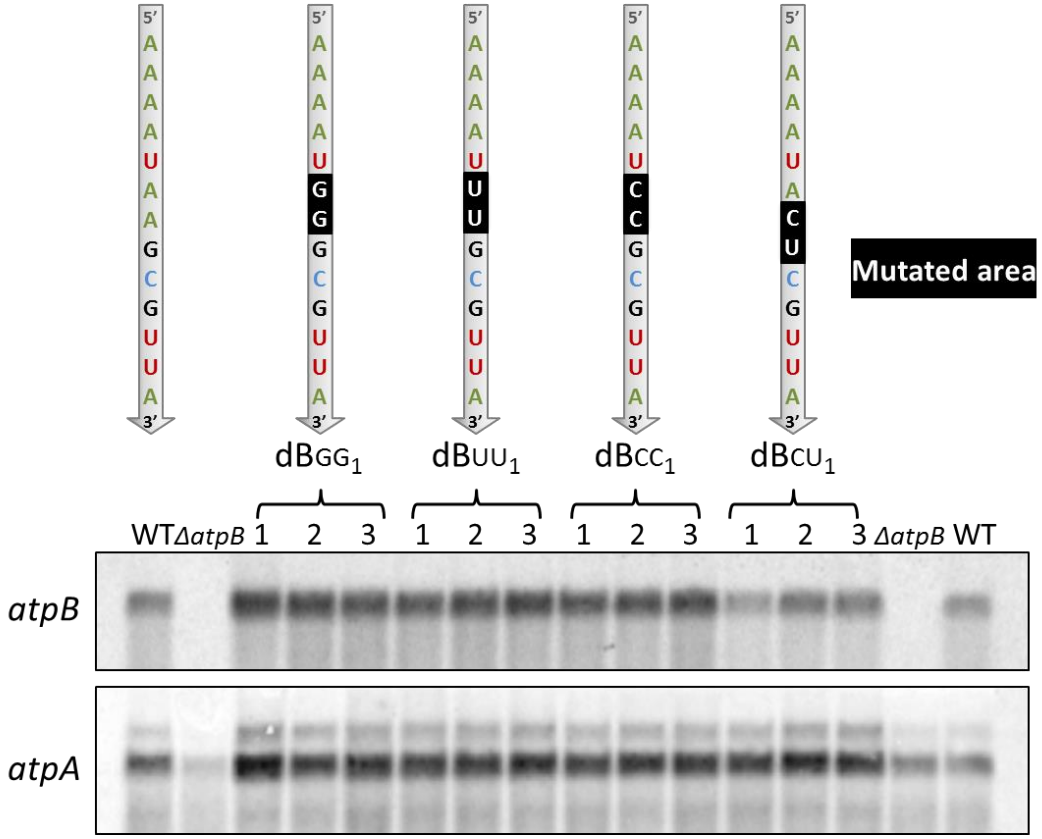


Figure 67: RNA blot of the first *atpB* MDB1 target sequence variants, filter were hybridised with ^{33}P-labelled *atpB* and *atpA* radioactive probes. On top are depicted the mutated target sequences with the respective mutation black shaded.

The *atpB* transcript accumulation levels were then checked by RNA blots (Figure 67). To our surprise the mutations had almost no effect on *atpB* mRNA accumulation, except for the *dBCT* one, which straddle the target sequence corresponding to the "hinge" of MDB1, i.e. an intervening sequence between two blocks of adjacent OPR repeats. This junction could be implicated in the relaxation of tensions in the MDB1 target sequence; indeed, it is possible that stretches of more than nine successive OPR repeat could inflict a strong tension on a substrate RNA. Perhaps this more flexible part of MDB1 is important for the efficient binding of the *atpB* transcript? In any case, it still does not fully prevent the binding of MDB1 to the *atpB* mRNA.

Considering how MDB1 is so crucial to *atpB* mRNA stabilisation and maturation and how it is specific to this transcript, we were quite baffled by the apparent resilience of the interaction. Could this predicted target sequence be accessory? To test this hypothesis, we decided to retry this strategy with another series of mutations.

EXPLORING THE WHOLE TARGET

Mutations spanning the whole putative target were designed. Some were still quite small, with only two nucleotides modified, other were much larger with four consecutive nucleotides altered. When a pyrimidine (U or C) was replaced with a purine (A or G) a stronger impact on the binding of MDB1 was expected, as puric bases are bulkier than pyrimidic ones, and could induce steric clashes. Those mutations were generated by two steps PCR mutagenesis and integrated into the pK^r*atpB* plasmid. As the spectinomycin cassette had been excised from the *ΔatpB* strain in the meantime, the constructions were transformed into this strain, to accelerate the sub-cloning phase.

After several rounds of sub-cloning, the *atpB* regions of the strains were sequenced, and three independent transformants by target variant picked. Growth test of the strains suggested that only *dB12* was severely affected (Figure 68.A).
Again, the *atpB* transcript levels were monitored by RNA blots, no great differences were observed between transformants of the same constructs. A blot with one transformant for each target variant is presented in Figure 68.B. The mRNA levels were normalised on the *dBct* control strain.

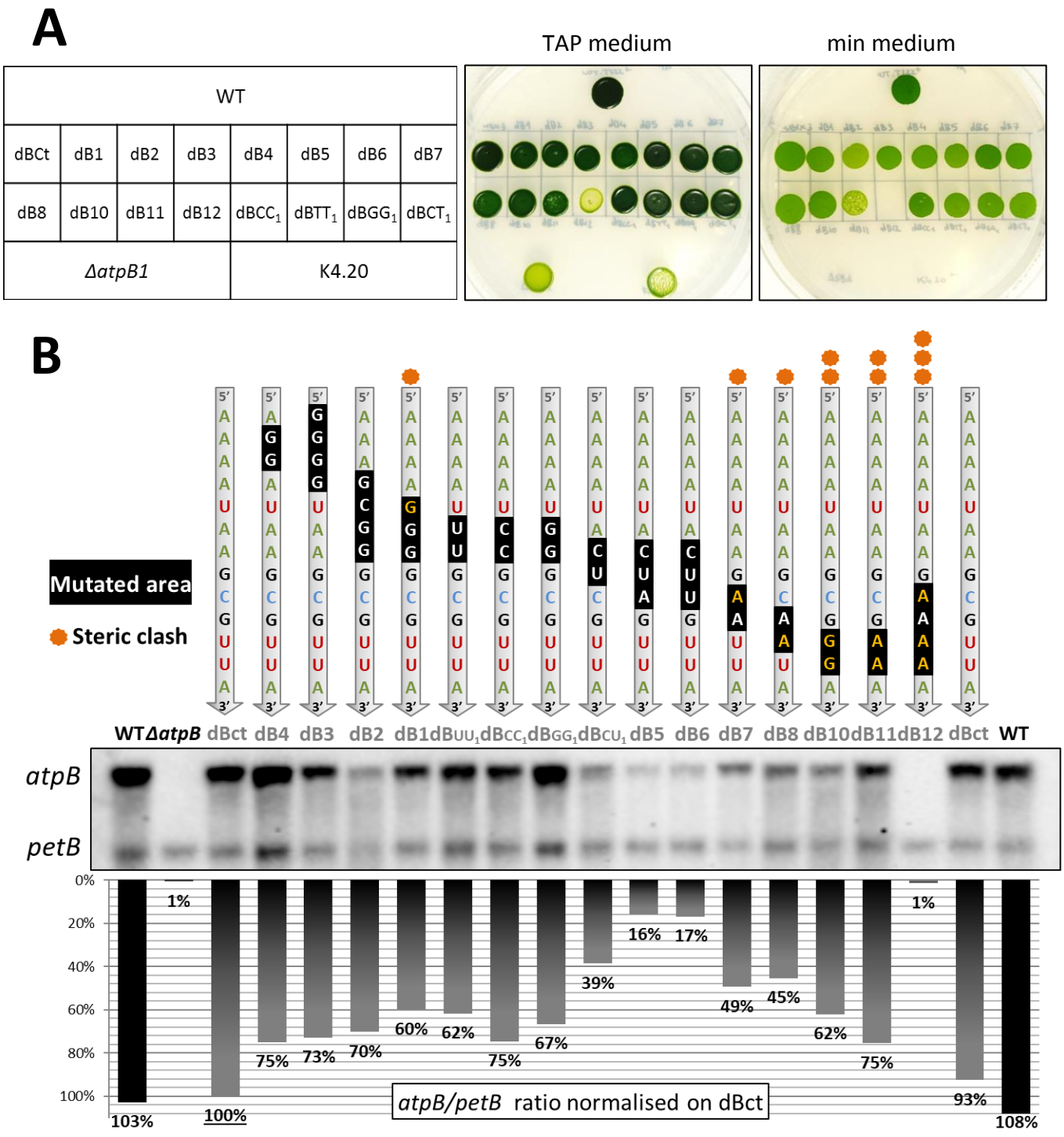


Figure 68: A. Growth phenotypes of the mutants, a table aside shows the placement of the strains. Droplets of liquid culture of the strains were put on TAP and minimum media and grown for 12 days under 55 µE.m^{-2}.s^{-1} illumination. **B. RNA blot of *ΔatpB* strains transformed with the mutated *atpB* MDB1 target sequence.** Corresponding mutations are depicted on top of each blot lane. Mutated nucleotides are depicted in black squares. An orange dot denotes the introduction of a steric clash in the mutant target. *atpB* and *petA* (loading control) mRNA quantifications were performed with the image lab software and normalised on dBct levels. The ratio of *atpB* on *petA* transcripts is depicted under the blot.

The first, blaring conclusion to draw from those experiments was that *atpB* mRNA stability was barely altered by most mutations. The target sequence was still bound by MDB1 and the transcript protected, even when four nucleotides were modified in the 5' part of the sequence, as in strains *dB2* and *dB3*. Overall, mutations in the 3' half of the target induced somewhat stronger effects, but this observation also seems to correlate with the amount of steric clashes induced, so it is difficult to conclude from this experiment whether it is the specific position or the steric clashes that induce this difference. We again noticed a stronger effect of mutations in the centre of the target where the MDB1 hinge lies, despite the absence of induced steric clashes. This part seems truly important for the interaction. Maybe the two groups of OPR repeats in MDB1 "clamp" down independently on each side of the target, and the junction is crucial to "lock" MDB1 in a stable interaction?

Nonetheless, the only mutant lacking *atpB* transcript was *dB12*, with four modified nucleotides and three steric clashes. This mutant proves that MDB1 is not completely insensitive to changes in the target sequence. However, this mutant would be quite impractical to study the code. An MDB1 with four mutated OPR repeats would have to be introduced and it could be difficult to identify the independent participation of each of the four repeats.

The surprising resilience of the MDB1/*atpB* mRNA interaction prompted us to assess whether it was an interesting but unfortunate (for our goal of studying the OPR code) exception or perhaps a more general property of OPR M factors.

MUTAGENESIS OF THE MTHI TARGET SEQUENCE *AT THE 5' END OF THE ATPH MRNA.*

INITIAL MUTAGENESIS

MTHI1, as an M factor with a well-defined binding sequence in *atpH* mRNA (see Chapter II), was thus picked to observe the effect of nucleotide modifications in its target sequence *in vivo*. We already saw that swapping the central nucleotides of the MTHI1 targets within the *atpH* and *atpI* 5'UTRs (see Chapter II from p78) had seemingly no effect on mRNA stability, pointing toward a tolerance for point mutations. Nonetheless, this could also be due to a putative ability of the 5th OPR repeat to recognise both A and G, even if the draft code point to a strong preference for G. Or perhaps the middle of this short binding sequence is less critical for the establishment of the interaction than the extremities.

Three target variants of *atpH* were designed. They were generated by two step PCR mutagenesis and inserted in the p^{Kr}atpH plasmid, then transformed in *ΔatpH* cells at the *atpH* locus. Transformants were selected for spectinomycin resistance and sub-cloned on selective media until they reached homoplasmy. Three independent transformants were selected for each mutation. At this step, growth tests revealed that, of the three mutants, only one was non-phototrophic, *ΔdH3* (Figure 69.A). Then, as usual, total RNAs were extracted from the transformants, blotted, and the accumulation of the *atpH* mRNA estimated (Figure 69.B).

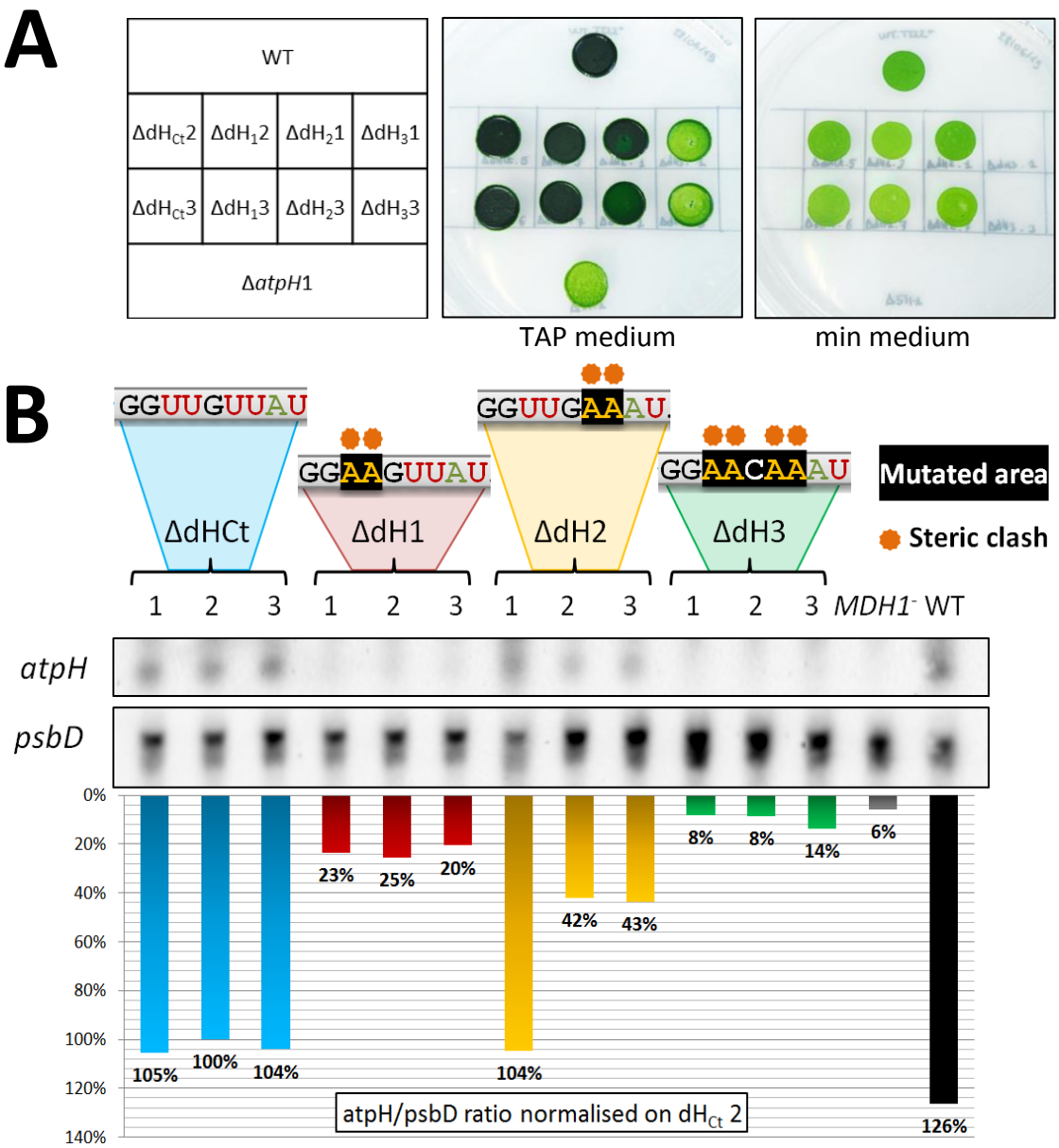


Figure 69: A. Growth phenotypes of the mutants, a table on the left shows the placement of the strains. Droplets of liquid culture of the strains were put on TAP and minimum media under 55µE illumination for 8 days. **B. RNA blot of *ΔatpH* strains transformed with the mutated *atpH* MDB1 target sequence.** Corresponding mutations are depicted on top of each blot lane. Mutated nucleotides are depicted in black squares. An orange dot denotes the introduction of a steric clash in the mutant target. *atpH* and *psbD* (loading control) mRNA quantifications were performed with the image lab software and normalised on *ΔdHCt 2* levels. The ratio of *atpH* on *psbD* transcripts is depicted under the blot.

The levels of *atpH* mRNA in the *ΔdH1* and *ΔdH2* mutants were quite sizeable and sufficient to allow some expression of AtpH and thus photo-autotrophy. In contrary *atpH* transcript levels were very low in *ΔdH3*. Quantifications here are probably biased by the smear of the strong *psbD* signal: indeed, in all three technical repeats the *ΔdH3* levels were of the same order than *mthi1-2* strain that should be null. The *atpH* mRNA probably does not accumulate in *ΔdH3* transformants, in coherence with the fact that they are non-photo autotrophic.

This suggests that MTHI1, much like MDB1, binds quite resiliently to its target RNA. It is resistant to changes in its target sequence in the *atpH* 5'UTR, even when steric clashes are induced, only a large mutation covering more than of half the sequence could prevent its binding. It seems that, like MDB1, MTHI1 does not exert a stringent recognition of its target sequence. This apparent low specificity of OPR M factors is paradoxical considering that their only observed specific footprints are found at the 5' end of their cognate chloroplast mRNA and that mutations in M factors usually affect the expression of one single chloroplast gene (Trosch *et al.*, 2018), but is consistent with the weak affinity of MTHI1 for its target as discussed in Chapter II.

TARGET SEQUENCE MODIFICATIONS IN ABSENCE OF MTH2

As suggested for MTHI1 in the second chapter, other factors could play a crucial role in establishing the specificity of M factors. For MTHI1 a putative auxiliary factor had been identified: MTH2. And so, we decided to assess if MTHI1 would be more easily dissociated from *atpH* transcript in absence of MTH2.

The plasmids containing the *dHct* and *dH2 atpH* variants were transformed in *mth2-2* cells. After sub-cloning, homoplasmy was tested by PCR amplification. Growth tests (Figure 70.A) indicated that the *mth2-2* {*dHct*} transformants grew as the *mth2-2* recipient strain: more slowly on minimum media than the WT but they were still photo-autotrophic. In contrast, the *mth2-2* {*dH2*} transformants could barely grow with photosynthesis alone, but still survived on minimum medium. Moreover, they did not display a photosensitive phenotype like the *ΔatpH* strain on TAP media, suggesting that AtpH was still produced at trace levels, enough to allow some photosynthesis and some dissipation of the proton gradient.

RNA blot of the transformants (Figure 70.B) revealed that the effect of MTH2 deficiency and of the *dH2* mutation were cumulative. Trace amounts of *atpH* transcript must accumulate in the *mth2-2* {*dH2*} transformants since they are phototrophic but were under detection level in this blot.

MTH2 does not seem to be the only factor anchoring MTHI1 on *atpH* 5'UTR: even in its absence and with a mutated target sequence, MTHI1 still manages to bind the *atpH* transcript.

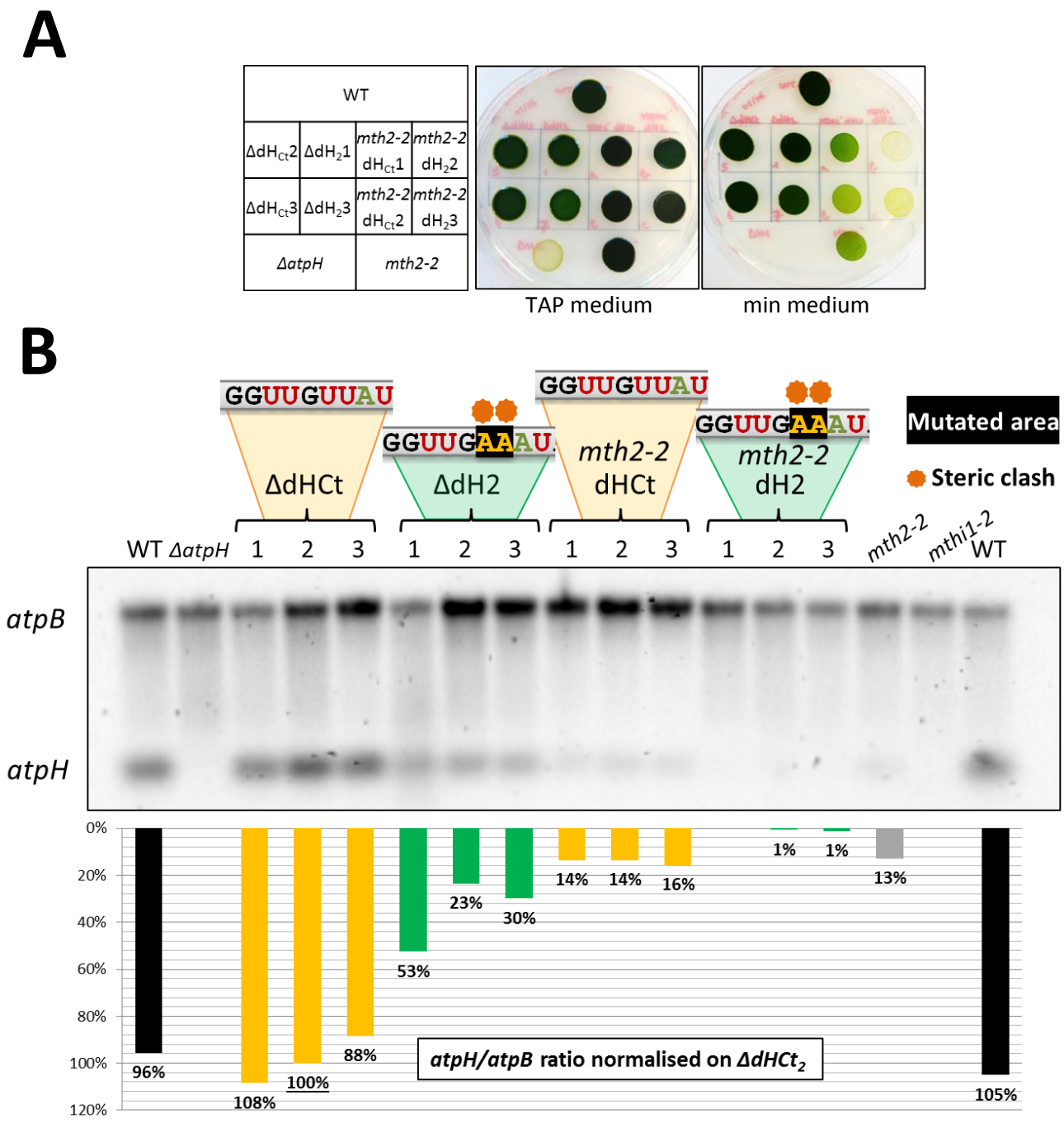


Figure 70: A. Growth phenotypes of the mutants, a table on the left shows the placement of the strains. Droplets of liquid culture of the strains were put on TAP and minimum media under 55 µE.m^{-2}.s^{-1} illumination. **B. RNA blot of *ΔatpH* and *mth2*$^{-}$ strains transformed with the mutated *atpH* MTHI1 target sequences.** Corresponding mutations are depicted on top of each blot lane. Mutated nucleotides are depicted in black squares. An orange dot denotes the introduction of a steric clash in the mutant target. *atpH* and *atpB* (loading control) mRNA quantifications were performed with the image lab software and normalised on *ΔdHCt 2* levels. The ratio of *atpH* on *atpB* transcripts is depicted under the blot.

DISCUSSION

A PARADIGM SHIFT ON OPR M FACTOR SPECIFICITY

Both of the OPR M factors that we tested displayed a "lenient" recognition for their chloroplast mRNA target: mutations in the binding sites sequences, corresponding to the specific footprints of these M factors, did not affect strongly the accumulation of either the *atpB* or *atpH* transcripts. Those resilient interactions seem contradictory with the reported specificity of OPR factors for their target in *C. reinhardtii* and came as a surprise for us at the time. But when substantial modifications in the targets were made, a near complete loss of the mRNA could be achieved, confirming that OPR factors do recognise their target. And when M factors cannot bind at all their target mRNA, no mechanism can alleviate the loss of this interaction.

As we saw in Chapter II, some OTAF factors are suspected to act together in complexes on their specific mRNAs. We suspect that those unknown factors could be critical for the compensation of defects in the target sequence of the M factors. The collaborative interaction between specific factors with moderate affinity could create a stronger affinity for the mRNA.

Another putative mechanism could rely on other parts of the mRNA itself. Indeed, as we saw in Chapter I, putting *rbcL* 3'UTR in a chimeric construct based on *atpB* 5'UTR can affect the 5'end maturation of the transcript. Perhaps secondary structures formed between the 5' and 3'UTR of a chloroplast mRNA could also impact the interaction of the M factor with its binding site.

THE SEARCH FOR MTHI1 PARTNERS CONTINUES?

Combining the *dH2* mutation in the absence of MTH2 did not completely prevent the accumulation of the *atpH* mRNA. The effects on transcript accumulation of both mutations appeared rather cumulative, suggesting either than the MTH2-mediated stabilisation is independent from the MTHI1-mediated one or that MTH2 forms only a part of the "rescue" mechanism tethering MTHI1 on the *atpH* mRNA. And so, yet another partner factor(s) likely contributes to the interaction of MTHI with the *atpH* transcript. We do not have enough data to counter or confirm these theories for now. Analysing the super complexes formed with *atpH* mRNA and MTHI1 (described in ***ARTICLE 3***) is required to answer this question.

INSIGHTS ON MDB1 FROM ANOTHER STUDY

Our results for *atpB* and MDB1 appeared even more surprising when compared to those of an older study (Anthonisen *et al.*, 2001). The authors made point mutations in the target sequence of MDB1 in the *atpB* 5'UTR and witnessed clear effects on transcript accumulation (Figure 71). While their results did not include loading controls, potentially lessening their observations, the differences between our results appeared striking enough to warrant further consideration. Smaller mutations led to a strong impairment of the MDB1/target sequence interaction, as depicted in Figure 72.A.

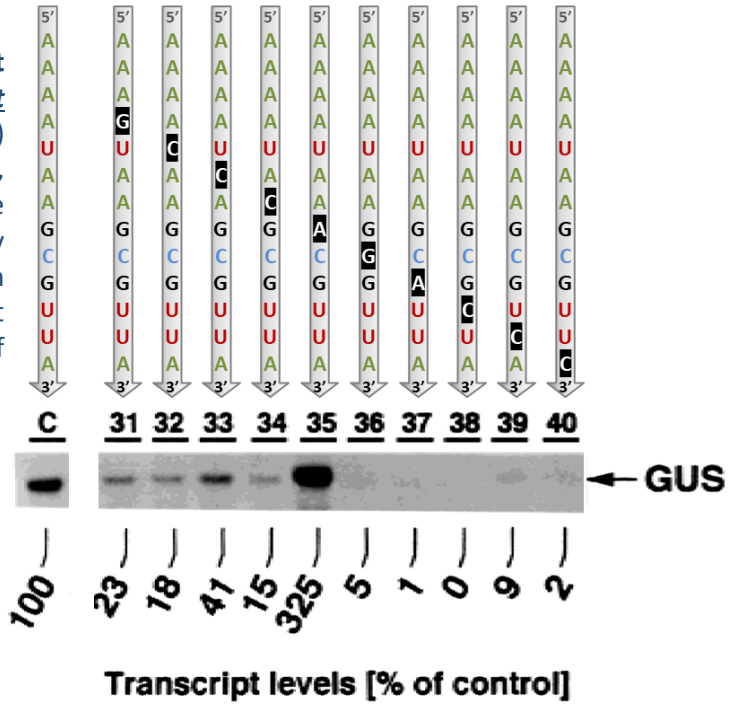


Figure 71: Northern blot adapted from (Anthonisen *et al.*, 2001), *uidA* mRNA (GUS) compared to the control (C), on top are depicted the mutations inserted by Anthonisen and colleagues in the *atpB* MDB1 target sequence. Note the absence of loading control.

They did also work in *C. reinhardtii*; but used a reporter construct based on *uidA*, an exogenous gene, and on the *psaB* 3'UTR. Only a small part of the mature *atpB* 5'UTR (31 nt) was put in front of *uidA* CDS. We suspect that the parts of *atpB* that were deleted are implicated in tethering MDB1 on its target (Figure 72.B). Their absence effectively "isolated" MDB1, allowing its study without interference of other factors. And so, we decided to design a similar construct, as we will see in the next chapter.

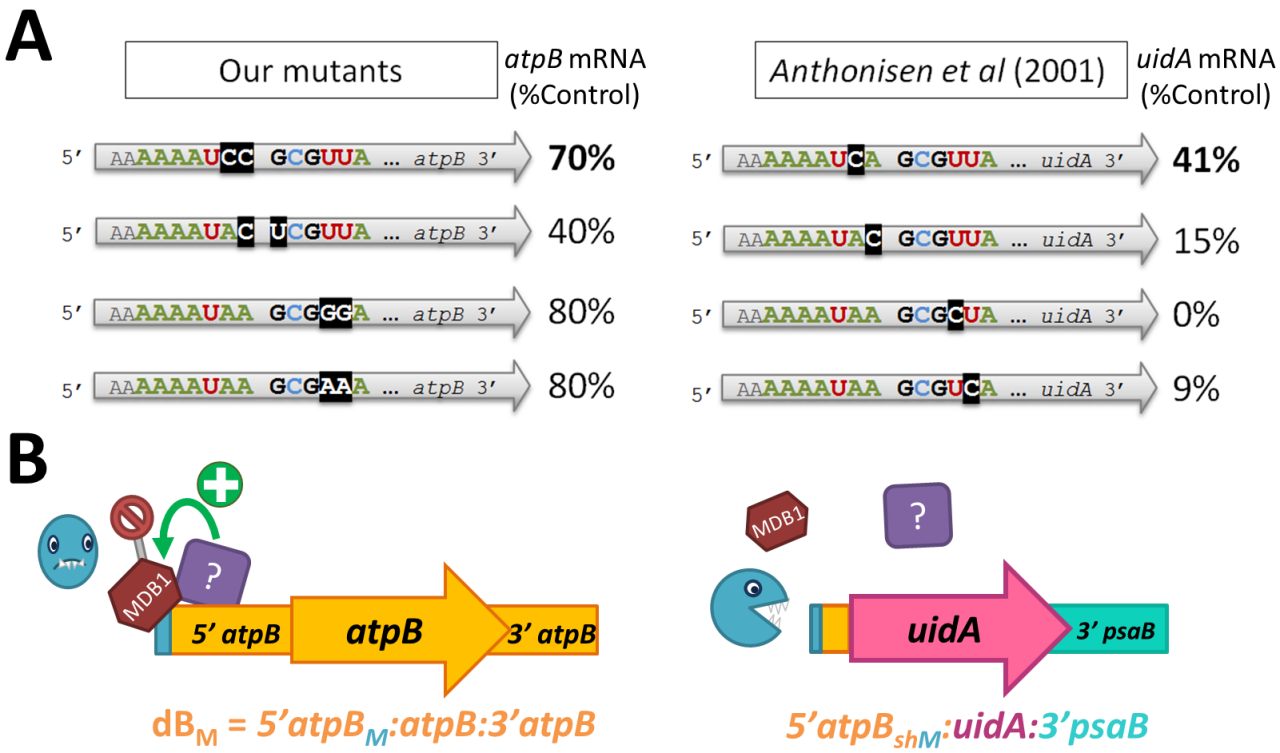


Figure 72: A. Comparison between some of our mutants and Anthonisen and colleagues' ones. **B.** Hypothesis to explain the differences observed: unknown factor(s) recognising a part of *atpB* transcript deleted in the *uidA* chimera can stabilise MDB1 on its target sequence.

IV. Building a chimeric system to validate the "OPR code"

Modified from Dartmouth Electron Microscope Facility, Dartmouth College

IMPLICATIONS OF SWITCHING TO A CHIMERIC REPORTER

Following our previous results (Chapter III) and the observations of Anthonisen and colleagues (Anthonisen *et al.*, 2001), we continued our study of the OPR code on the MDB1/*atpB* pair, using an exogenous sequence instead of the *atpB* CDS and only a short fraction of the *atpB* 5'UTR, which does not contain the translation initiation signals. This thus deprived us from a convenient screen based on the restoration of phototrophy or on expression of a reporter gene. Selecting and analysing transformants would require monitoring the accumulation of the chimeric transcript in each transformant without prior selection. This would render the reporter construct unsuitable to design future experiments with larger scale randomised RNA target variants for instance.

We used a part of the exogenous CDS encoding *Azotobacter vinelandii* GFP (green fluorescent protein) as a reporter transcript to lower as much as we could the possibilities of interactions with OTAFs. We also decided to keep our chimera as similar as possible to the one from (Anthonisen *et al.*, 2001) to get similar results. Indeed, the whole *atpB* 5'UTR is strongly suspected to interact with other factors that would tether MDB1 on its target (Chapter III). We did not add translation initiation signals from other chloroplast genes as they could be the target of some OTAFs, defeating our purpose of looking at the sole interaction between MDB1 and its *atpB* target.

The chimera was built as follow: a 144 bp fragment of the *atpB 5'* region, containing the *atpB* promoter (Klein *et al.*, 1992) up to the first 34 nt of the *atpB* mature 5'UTR, thus including the MDB1 target and the whole MDB1 footprint, was placed in front of the *gfp* sequence. The construct was terminated by a short version of the *rbcL* 3'UTR, as stem loops are critical for mRNA 3'end stability (Figure 73). Funnily enough (or rather not), at that time, we did not know the impact of the *rbcL* 3'UTR on the maturation of the *atpB* 5'UTR (Chapter I) …

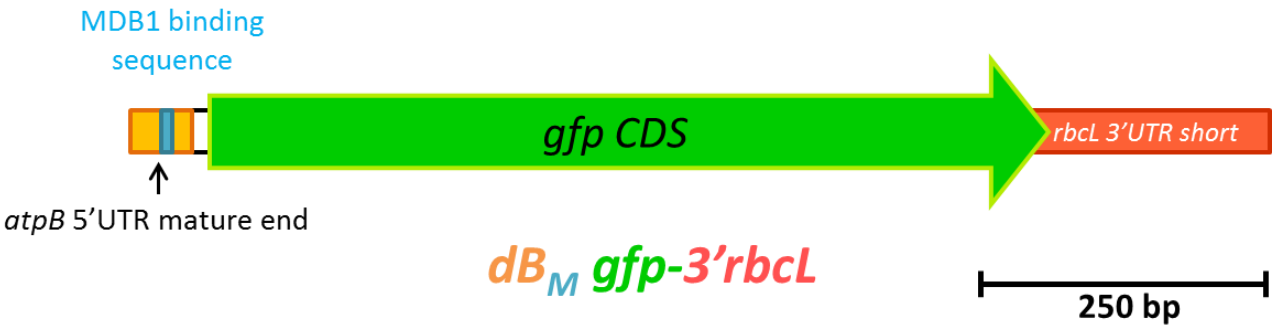


Figure 73: The *gfp* chimeric construct at scale.

DESIGN OF AN ADDITIONAL REPORTER WITH THE SPINACH2 APTAMER

Since translation was not possible in our chimeric system, I looked at possible RNA reporter systems, enabling us to screen transformants for stable transcript accumulation without extracting RNA from every one of them. I found a good potential reporter: Spinach (Paige *et al.*, 2011). Spinach is an RNA aptamer, a sequence that folds into a secondary structure, able to interact with specific substrates. Its structure is formed of planar layers, among which two G quadruplexes. The space between two of the layers allows a flat molecule to enter the structure (Figure 74.B). Spinach is a light-up aptamer, as it can bind the fluorophore 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) and strongly increases its fluorescence emission (Figure 74.A). DFHBI is not toxic for the cells and membrane permeant, making its use possible *in vivo*. In its free state, DFHBI dissipates the excitation energy by rotation of its cycles (Figure 74.C). When DFHBI interact with Spinach it is locked in a stiff state and cannot dissipate energy by conformational changes. Therefore, when excited by light of the appropriate wavelength (469 nm) it emits fluorescence at 501 nm.(Paige *et al.*, 2011; Bouhedda *et al.*, 2017).

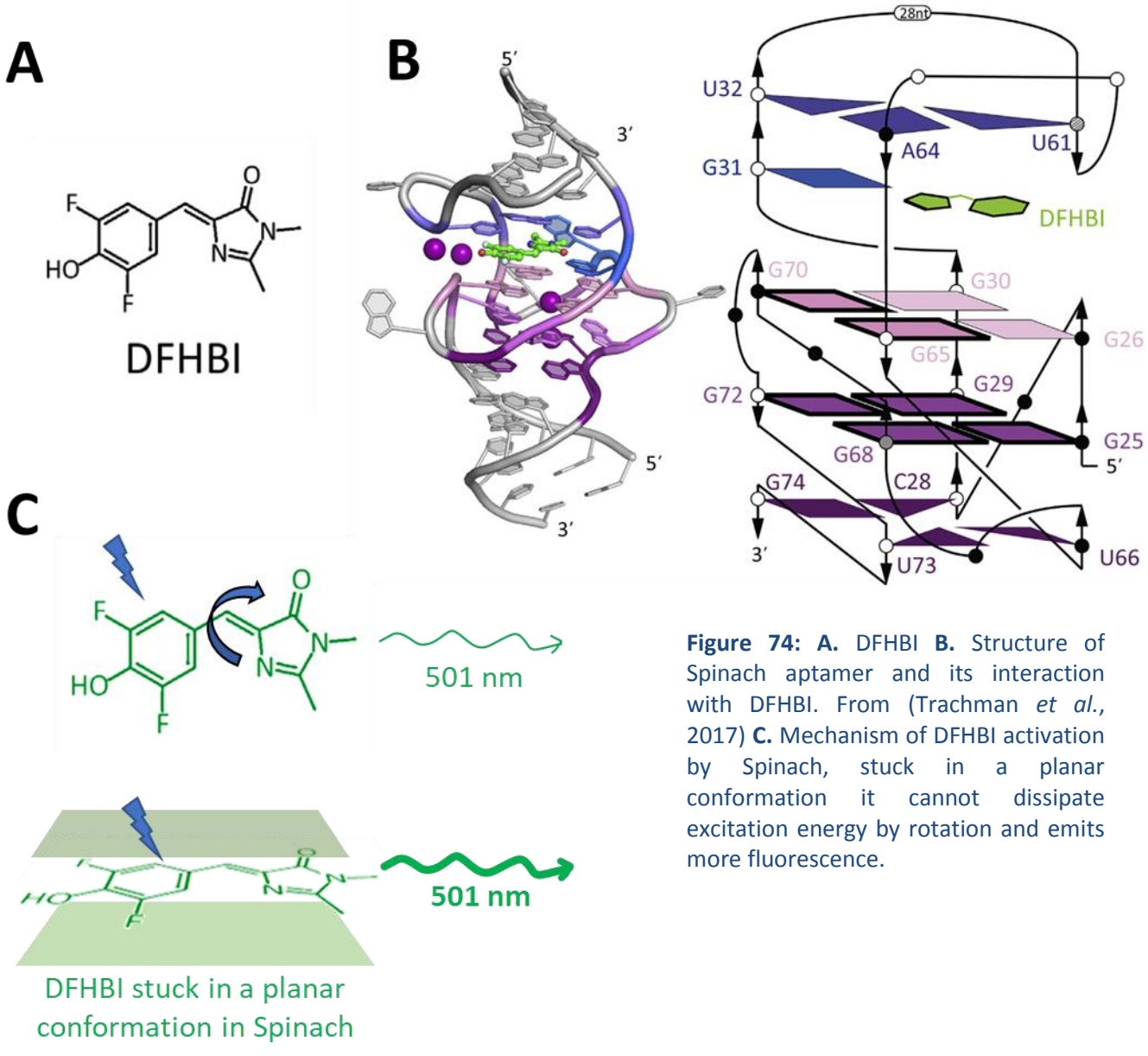


Figure 74: A. DFHBI **B.** Structure of Spinach aptamer and its interaction with DFHBI. From (Trachman *et al.*, 2017) **C.** Mechanism of DFHBI activation by Spinach, stuck in a planar conformation it cannot dissipate excitation energy by rotation and emits more fluorescence.

This reporter system had been used previously in *C. reinhardtii* (Guzman-Zapata *et al.*, 2017) to image living cells and assess the accumulation of a transcript, making this system seemingly suited to our needs.

I thus developed in parallel of the streamlined chimera, a Spinach-based reporter. As Spinach/DFHB1 fluorescence is rather faint in part because of poor folding properties in cells (Han *et al.*, 2013; Ilgu *et al.*, 2016), we used Spinach2 a 95 nt long optimised version of Spinach, that is more thermostable and folds better *in vivo* (Figure 75) (Strack *et al.*, 2013).

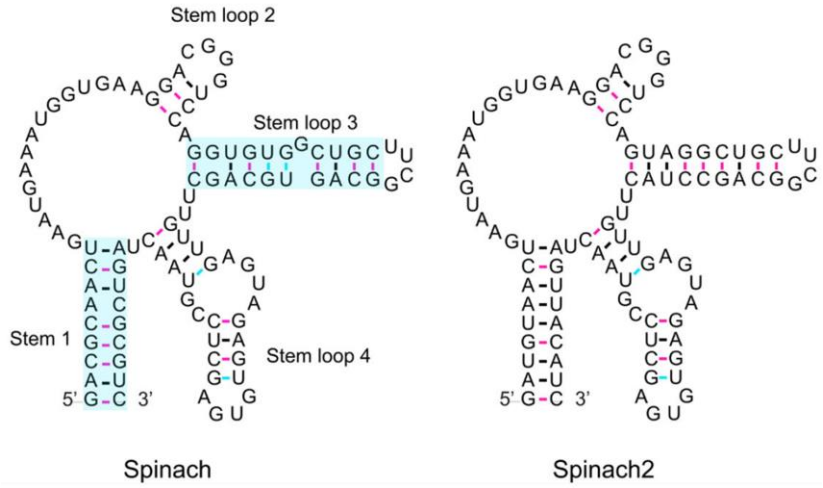


Figure 75: Mutations introduced by Strack and colleagues In Spinach2 improve thermostability and folding, Stem1 and Stem loop 3, in blue, were modified. From (Strack *et al.*, 2013).

We also tried to add several consecutives Spinach2 in our reporter, to enhance the fluorescence emission of a transcript. The Spinach2 were put between the *gfp* sequence and the 3'UTR (Figure 76). As the Spinach2 sequences were inserted at specific restriction sites, the *gfp* CDS had to be trimmed at various lengths in its 3' end. This should be inconsequential since the reporter is not intended to be translatable.

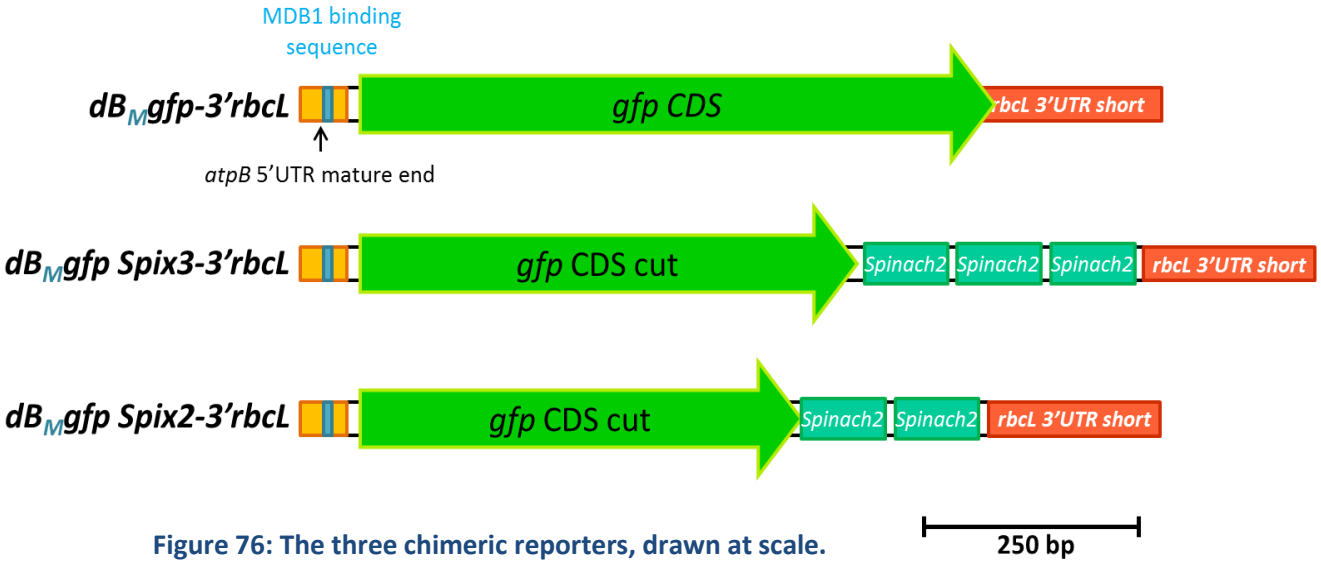


Figure 76: The three chimeric reporters, drawn at scale.

RESULTS

COMPARING THE TWO REPORTER SYSTEMS

Part of the work described in this chapter was done with Julia lo Turco an undergraduate student in a training internship.

DESIGNING THE REPORTERS

The construction of the simple *dB$_{WT}$gfp-3'rbcL* chimera was achieved as described in Materials and Methods p145. The construction of the Spinach2 reporters, derived from the *dB$_{WT}$gfp-3'rbcL* chimera, is described p146 and illustrated in Figure 76. All chimeras are embedded in plasmids containing the excisable 5'*psaA-aadA* cassette for selection of the transformants (Wostrikoff *et al.*, 2004).

We transformed the *dB$_{WT}$gfp-3'rbcL* construct in the WT T222+ or *ΔatpB* strains, while the *dB$_{WT}$gfp Spix2-3'rbcL* and *dB$_{WT}$gfp Spix3-3'rbcL* constructs were introduced in the WT.T222+, *mdb1-2* and *ΔatpB* backgrounds. After sub-cloning, homoplasmy was tested by PCR amplification. Total RNAs were extracted from the transformants and blotted.

REPORTER ACCUMULATION, COMPETITIONS WITH *ATPB* MRNA

Spinach2 tagged transcripts accumulate in a MDB1 dependant manner

We worried whether the Spinach2 aptamer, a dense and stable structure, reminiscent of a polyG track, could stabilise the chimeric transcript, irrespective of the MDB1 binding. The reporter being designed to assess the MDB1-mediated stabilisation, such an intrinsic effect of Spinach2 on mRNA stability could be highly misleading.

The chimeric transcript accumulated in WT or *ΔatpB* strains (Figure 77). However, transformants lacking MDB1 did not accumulate it at all, indicating that Spinach2 in the 3'end of the transcript does not constitutively stabilise it. Moreover, the presence of two or three copies of Spinach2 did not increase the accumulation of the *gfp* chimeric mRNA in the WT, when compared to the version devoid of the aptamer. Those results confirmed that the missing 3' part of the *atpB* 5'UTR is not necessary for MDB1-mediated stabilisation, as could be expected from the study of (Anthonisen *et al.*, 2001).

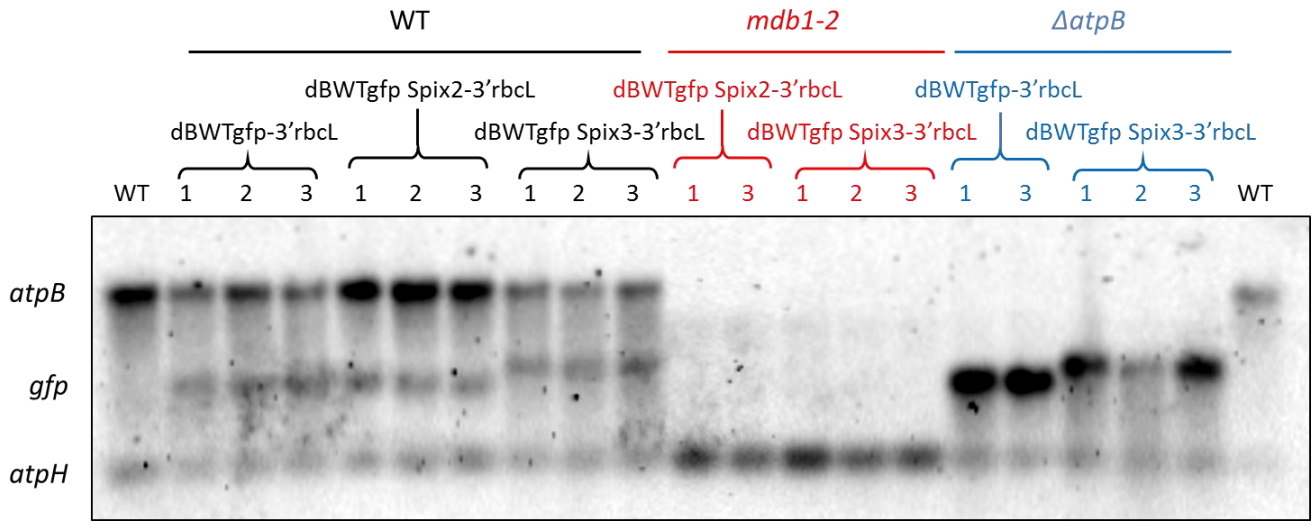


Figure 77: RNA blot of the Spinach2 tagged chimeric transcripts, filter was hybridised with *atpB*, *gfp* and *atpH* dig-dUTP labelled DNA probes.

Note that *dB$_{WT}$gfp Spix2-3'rbcL* and *dB$_{WT}$gfp-3'rbcL* migrate similarly in the gel, while *dB$_{WT}$gfp Spix3-3'rbcL* is retarded. In the *dB$_{WT}$gfp Spix2-3'rbcL* construct, part of the *gfp* CDS was trimmed (Figure 76). This explains the fast migration of the transcripts, similar to that of the *dB$_{WT}$gfp-3'rbcL* chimeric mRNA (Figure 77).

Chimeric transcripts compete with *atpB* mRNA

This RNA blot also revealed that the chimeric Spinach2 x3 construct is far more accumulated in the *ΔatpB* background than in a WT background (Figure 77), probably because more MDB1 is available in absence of its endogenous target transcript. Similarly, the accumulation of the simple chimeric *gfp* mRNA was increased about ten-fold in the absence of the endogenous *atpB* transcript (Figure 78). In contrast, *atpB* mRNA accumulation was only slightly affected by the presence of the chimera and remained at ~75 % of its level in the wild-type strain.

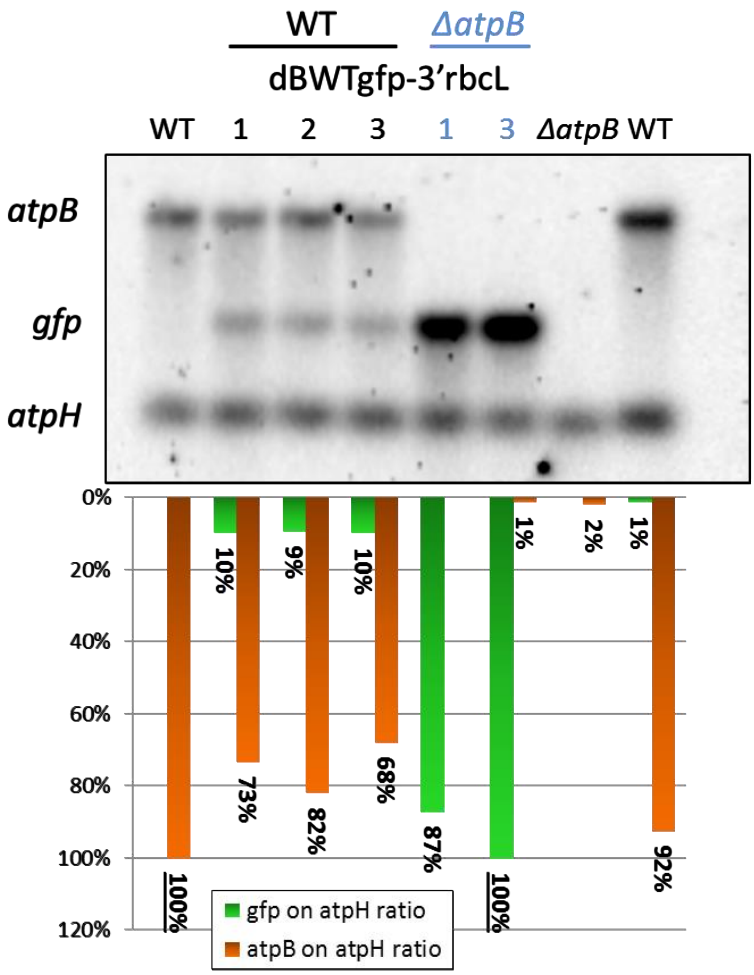


Figure 78: RNA blot of the "basic" *gfp* chimeras, either in WT or *ΔatpB* background. The filter was hybridised with *atpB*, *gfp* and *atpH* dig-dUTP labelled DNA probes. The ratios of *gfp* and *atpB* on *atpH* transcripts are depicted under the blot. mRNA quantifications were performed with the image lab software and the ratios normalised on WT for *atpB* and *ΔatpB*::*dB$_{WT}$gfp* 3 for gfp.

Those results indicate that *atpB* is preferentially stabilised in a competition with the chimera. *atpB* could either be transcribed at higher levels than the chimeric transcripts, even if the promoter fragment retained in our construct is as efficient as the full length *atpB* promoter (Klein *et al.*, 1992) and would sequester more of the MDB1 stock, or, based on the results in ***ARTICLE 1***, the *rbcL* 3'UTR could prevent the maturation of 5'*atpB*-driven transcripts and indirectly destabilise the chimeric transcript. However, the most likely explanation is that the missing sequences of the *atpB* 5'UTR, participate in its stabilisation. These could include the binding sites of other specific factors, anchoring MDB1 on its mRNA (Figure 79).

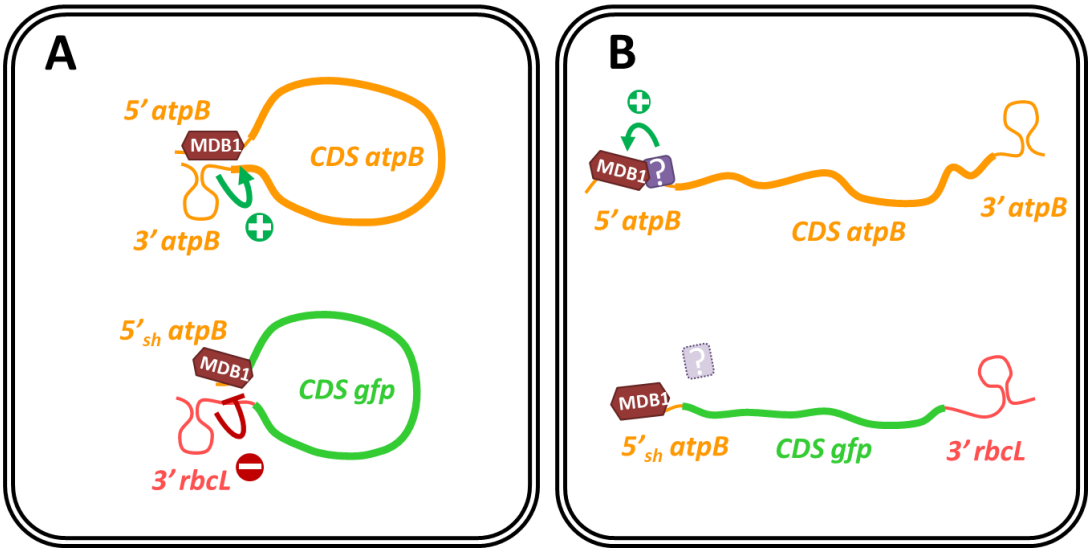


Figure 79: Two main hypotheses to explain how the mutated transcripts could be less destabilised in the presence of the whole atpB sequence. A. Other regions of the *atpB* transcript such its 3'end could be physically implicated in the stabilisation process. **B.** Auxiliary factors might limit *atpB* transcript destabilisation by maintaining MDB1 on its target sequence.

Effects of atpB 3'UTR on chimeric transcript levels

To test those hypotheses, the chimeric *dB$_{WT}$gfp.Spix3-3'rbcL* construct was modified to replace the *rbcL* 3'UTR by the *atpB* 3'UTR. Again, the construction was transformed into WT.T222+, *mdb1-2* and *ΔatpB* strains. After homoplasmisation, total RNAs were extracted and blotted (Figure 80).

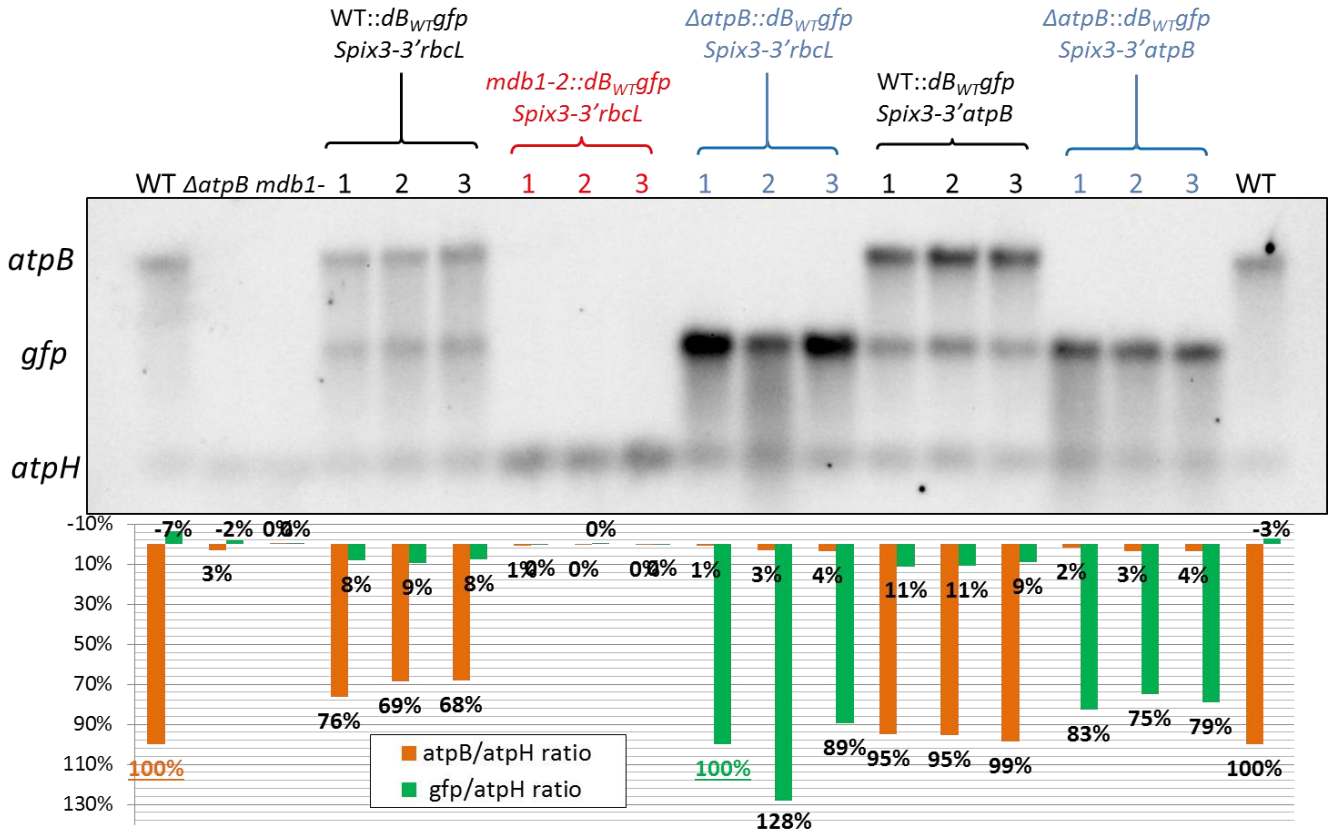


Figure 80: RNA blot of Spinach2 tagged chimeric constructs, with either *atpB* or *rbcL* 3'UTR, *gfp*, *atpB* and *atpH* (loading control) transcripts levels were probed. mRNA quantifications were performed with the image lab software and normalised on either {*ΔatpB dBWTgfp* Spix3-3'rbcL} 1 for *gfp* or WT levels for *atpB*. Reference levels are underlined. The ratios of *gfp* and *atpB* on *atpH* transcripts are depicted under the blot.

Firstly, this blot confirmed that the chimeric transcript accumulates in a MDB1 dependant manner, and is in competition with the endogenous *atpB* transcript for its stabilisation: there is about ten times more *gfp* signal in the *ΔatpB* background that in the WT. Surprisingly, the levels of the *dBWTgfp.Spix3-3'atpB* transcript were slightly lower than those of the *dBWTgfp.Spix3-3'rbcL* transcript in the *ΔatpB* background! *atpB* 3'UTR seems to have a slight negative impact on transcript accumulation. It is quite possible that the *rbcL* 3'UTR inherently protects more efficiently the transcript from exonucleases that the *atpB* 3'UTR, because of a stronger stem loop structure. However, we did not observe a stronger accumulation of a chimeric *BKR* (*atpB 5'UTR:aadA:rbcL 3'UTR*) transcript compared to a *BKB* (*atpB 5'UTR:aadA:atpB 3'UTR*) one (see Figure 97 p131). Admittedly, our dB_{WT}*gfp.Spix3* chimeras are different because they only bear a small part of *atpB* 5'UTR. So a difference in secondary structure formation between the 5' and 3'ends is possible, as further explored in the discussion (p130). But a direct weak negative effect of *atpB* 3'UTR on transcript stability is also conceivable. This effect could appear when the 5'UTR is not complete, maybe because auxiliary stabilising factors would not mitigate it.

Moreover, the endogenous *atpB* transcript is more accumulated when in competition with the dB_{WT}*gfp.Spix3-3'atpB* transcript than with the dB_{WT}*gfp.Spix3-3'rbcL* one. The chimeric transcript levels appear a bit stronger with *atpB* 3'UTR but only in the presence of the *atpB* transcript. This slightly higher accumulation of dB_{WT}*gfp.Spix3-3'atpB* in the presence of *atpB* is not detrimental to *atpB* endogenous transcript accumulation and probably does not indicate an improved recruitment rate of MDB1.

From this blot, I suggest that a low abundance specific endonuclease might target the *atpB* 3'UTR, whose presence in the chimera exposes it more to degradation than *rbcL* 3'UTR. When both the *atpB* and the dB_{WT}*gfp Spix3-3'atpB* transcripts are present together they both draw part of the specific endonucleases, thus sharing this burden, and diminishing the degradation rate of the *atpB* endogenous transcript.

I suggest that *atpB* 3'UTR is not really significantly implicated in MDB1 stabilisation process but mostly in the degradation of the *atpB* transcript. Chimeras with the *atpB* 5'UTR and *petA* 3'UTR are matured at their 5' end while those with the *atpB* 5'UTR *and rbcL* 3'UTR are not (as seen in ***ARTICLE 1***), the effect on maturation could be specifically caused by *rbcL* 3'UTR. The natural process of *atpB* 5' end maturation might not implicate the 3'UTR normally.

A LOOK AT THE SPINACH2 REPORTER FLUORESCENCE

Since the Spinach2-tagged transcripts accumulated in our cells, I tried to observe their fluorescence *in vivo*. Following the (Guzman-Zapata *et al.*, 2017) protocol, *C. reinhardtii* cells of various strains were incubated in 200 µM DFHB1 (in PBS) for 10 min at 37°C and washed three times. Afterward, the cells were deposited in poly-lysine coated chambered microscopy slides and observed with a Zeiss Axio Observer Z1 microscope with various green filters. No difference in green fluorescence could be observed between strains accumulating or not the Spinach2 tagged transcripts. Figure 81 shows the fluorescence of the *ΔatpB::dB$_{WT}$gfp-Spix3-3'rbcL* strain, which accumulates the most Spinach2 tagged mRNAs, yet does not display any detectable differential green fluorescence when incubated with DFHBI.

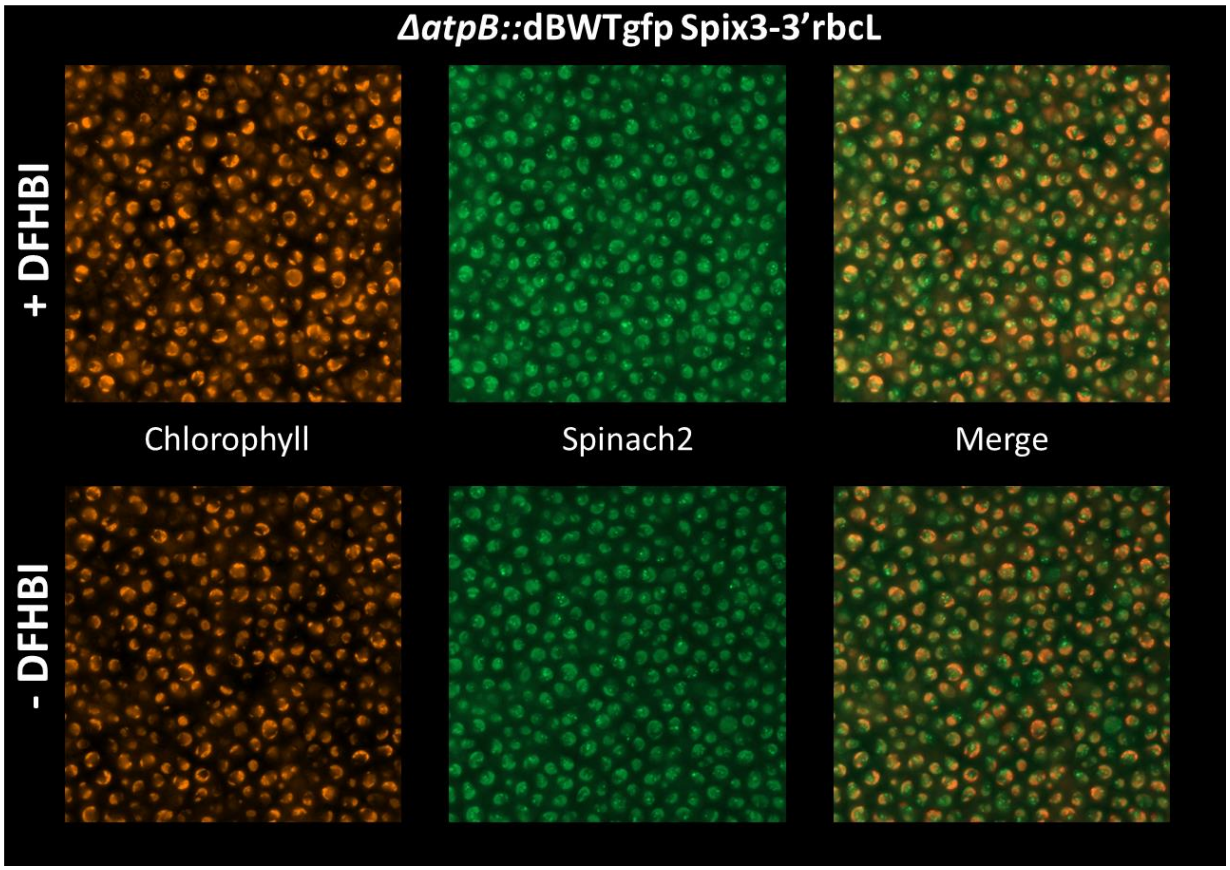


Figure 81: Chlorophyll auto-fluorescence and green fluorescence of cells treated or not with DFHBI, no difference in green fluorescence could be observed between the two treatments. The bright spots of green fluorescence correspond to the carotenoid-rich eyespots of the cells.

Subsequent tests performed with Julien Sellés using several biophysical systems of our laboratory could not detect any differential signal either. We suspect that the weak DFHBI fluorescence is masked by the broad and strong carotenoid fluorescence in *C. reinhardtii*. Unfortunately, time was running short and I did not have enough time to optimise this fluorescence reporter system.

Since the Spinach2 reporters do not provide any added value for screening, we decided to use the simpler *dBgfp-3'rbcL* reporter to study the OPR code. From this point onward, all chimeras are based on *dBgfp-3'rbcL* unless otherwise specified.

Studying MDB1 binding with the chimeric system

Considering the previous results from Anthonisen and colleagues, we opted for small mutations of two nucleotides, predicting that, in the absence of the rest of the *atpB* 5'UTR, they should disturb MDB1 binding on its target sequence. Following the apparent importance of the 3' half of the target (see Chapter III Figure 68), we decided to test and compare mutations in both sides of the target sequence. MDB1 target variants were generated by PCR mutagenesis or recovered from the pK^r*atpB* WT, CC_1, UU_1 or GG_1 plasmids and integrated in the dBgfp-*3'rbcL* construct. After transformation in the WT.T222+ or *ΔatpB* strains, and several rounds of sub-cloning, homoplasmy was tested by PCR and three independent transformants were selected for each sequence variant. After RNA extraction, the samples were blotted.

Binding affinity of MDB1 for the target variants

In *ΔatpB* transformed cells the control transcript accumulated strongly; in comparison the levels of the target variants constructs were far weaker (Figure 82). Overall those results confirm our first observations, the 5' half of the target sequence is less important than the 3' one for the MDB1/*atpB* mRNA interaction, moreover, this effect is also observed for the $dBCC_2$ variant that does not have steric clashes.

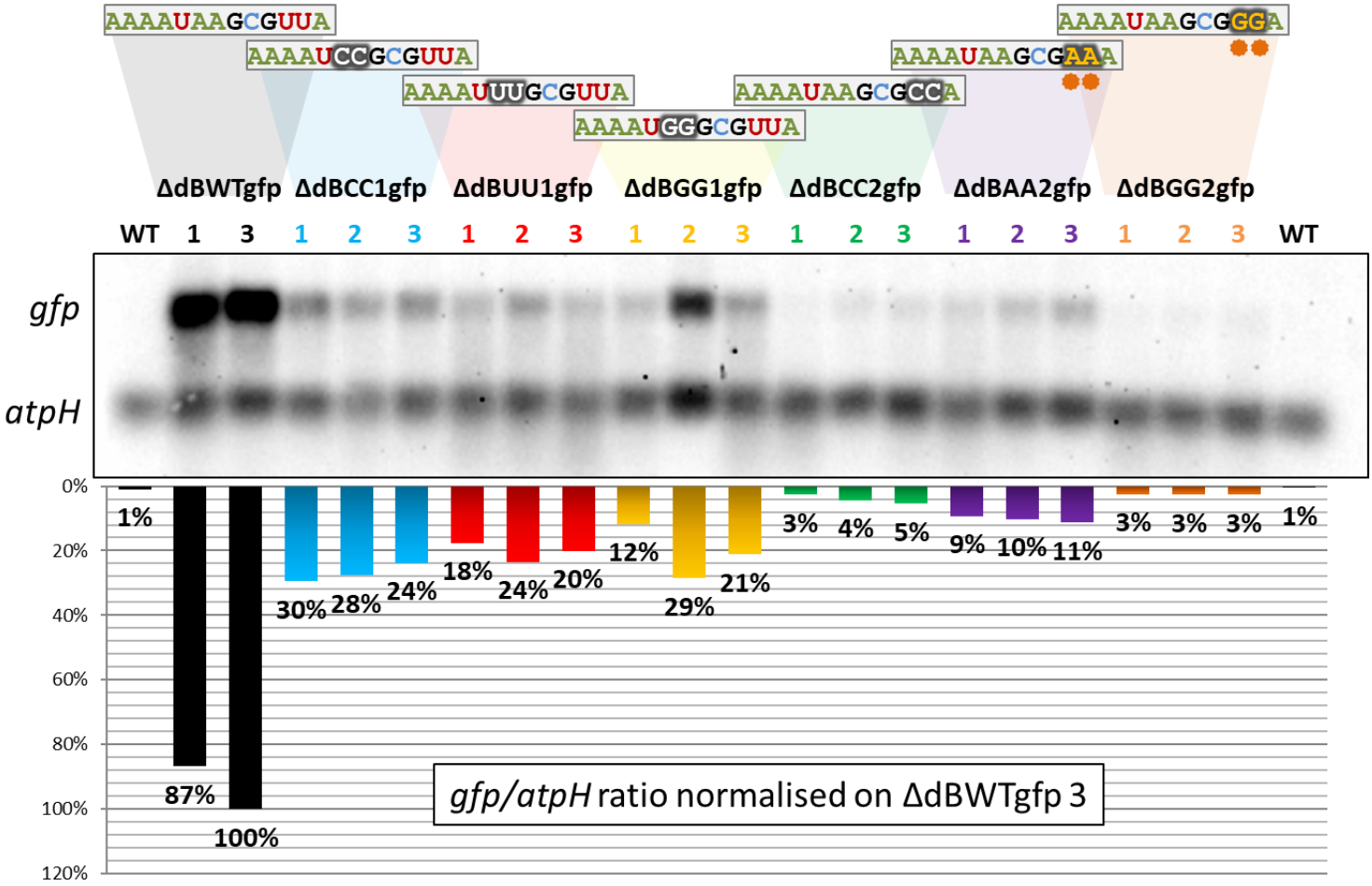


Figure 82: RNA blot of *ΔatpB* transformants chimeric constructs bearing *atpB* MDB1 target variants. Filter was hybridised with gfp and atpH dig-dUTP labelled probes. Transcript quantifications were done with ImageLab, and normalised on *ΔatpB*::dB_{WT}gfp 3 levels. Ratio of *gfp/atpH* transcripts is depicted under the corresponding lanes, the mutations are on top. Two technical repeats were made and give the same results.

The peptide sequence of corresponding OPR repeats (number 6, 7 for the mutations in the 5'part of the target and 11 and 12 for the mutations in the 3'part of the target) of MDB1 are depicted in Figure 83.

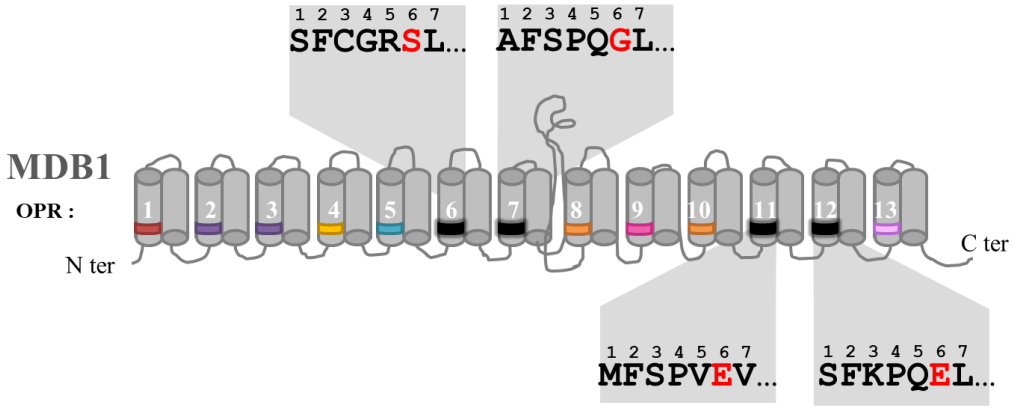


Figure 83: Beginning of the amino acid sequence of the tested OPR motifs in MDB1, the probably crucial sixth residue of the repeat is highlighted in red.

As the repeat 6 and 7 of MDB1 have different residues in position 6, drawing conclusion on their individual affinity for nucleotides is not possible; however, they broadly follow the putative code. They do bind far better AA than UU or GG and seems to tolerate CC slightly more. For the repeats 11 and 12, stronger conclusions can be made, as they both have a glutamate at their sixth position. From this RNA blot we can say that repeats 11 and 12 that have a E in 6th position both recognises preferentially: U>>A>C≥G. Interestingly, it appears in this case that pyrimidines are strongly differentiated by those two repeats. This is quite different from what was observed for the PPR code (Barkan *et al.*, 2012), in which some residues combinations binding U or C were "promiscuous" for the other pyrimidine. It remains to be seen if this is a general rule for the other OPR repeats or if the other recognises pyrimidines more loosely. Table 5 summarises our observations, so far, they go along the predictions of the putative OPR recognition code. But to validate the importance of the 5th and 6th residues, we need to modify just them in repeats to see the impact on nucleotide recognition.

Position	Residue										
3	X	X	X	X	R, K	X	X	X	X	X	X
4	X	P	X	X	X - P	X	X	X	X	X	X
5	X	Q	X - R, K	R, K	X	R	X - R	R, Q	X	R	R
6	E	G	D	D	D	Q	Q	A	H	S	N
Recognised nucleotide	U	A	G	U	U	U	?	A	?	A	?

Not tested yet
Coherent

Table 5: The draft OPR code: what is predicted and what we know so far.

COMBINING COMPETITION AND SEQUENCE VARIATION

By comparing the levels of the *gfp* chimeric transcript in a WT versus a *ΔatpB* context, we witnessed a drop in the accumulation of the target variant transcript under detection levels in presence of the concurrent endogenous *atpB* (Figure 84); only the chimeric *gfp* transcript with the WT target sequence could be detected. The combination of the target mutations and the competition for MDB1 with *atpB* mRNA completely prevents the accumulation of the reporter transcript, even if recognition by MDB1 remains possible as seen in Figure 82.

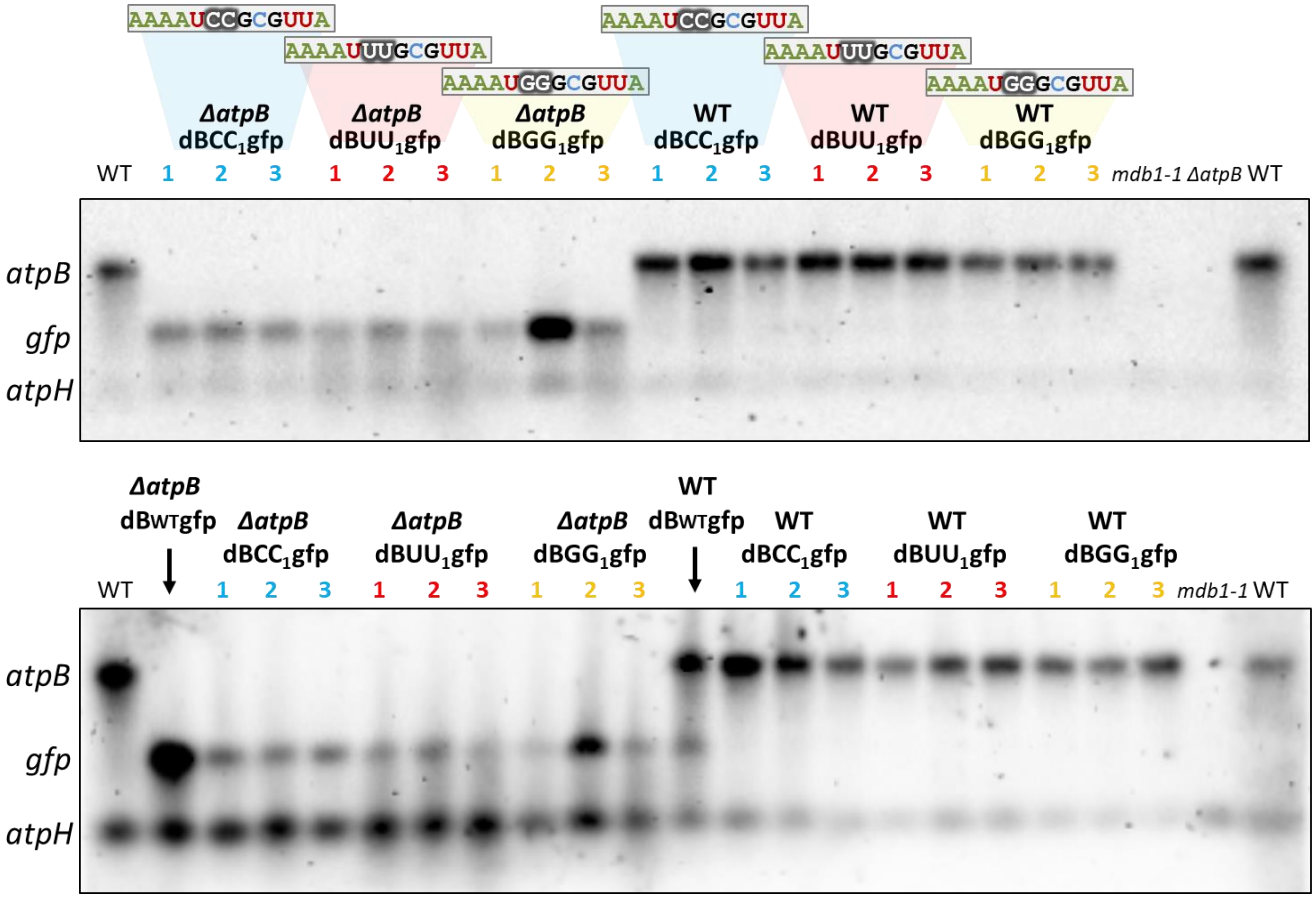


Figure 84: Two RNA blots showing the difference of *dB$_M$gfp-3'rbcL* transcript in WT or *ΔatpB* genetic background. Filters were hybridised with *atpB*, *gfp* and *atpH* dig-dUTP labelled probes. Four technical repeats displayed the same patterns.

The WT situation is more complicated than the *ΔatpB* one because we encounter the competition effects. Because of this competition, we could not see a difference of accumulation between different variants, nor assess their relative importance for the binding affinity. To study the code, a system with MDB1 entirely dedicated to the stabilisation of the reporter is more powerful: it is closer to an "*in vitro*" system, with MDB1 isolated from its partner factors and able to interact with the sole chimeric transcript. Thus, we decided to validate the code in the *ΔatpB* strain.

I will now describe the introduction of mutated MDB1 proteins in *C. reinhardtii*, and our strategy to confront the various mutated OPR with each target variant.

VALIDATION OF THE OPR CODE

MDB1 MUTAGENESIS:

Mutations in MDB1 sequence were introduced to modify the putative key residues at the fifth and sixth positions in the OPR repeat 6 and 7 or 11 and 12 according to the draft OPR recognition code (Figure 85). Since, no specificity conferring residues have been linked so far to C recognition we used residues similar or close to those found in the 9th OPR repeat of MDB1 that is predicted to interact with C, but are more generally found in "unreadable" repeats.

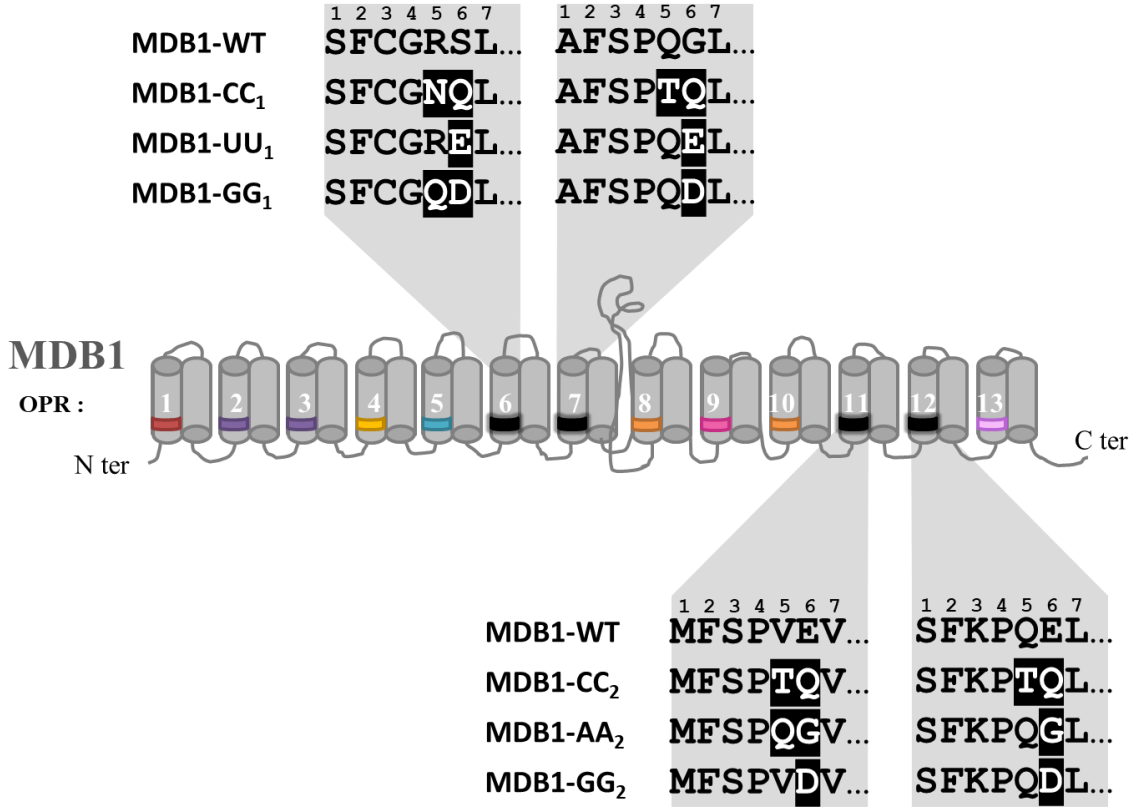


Figure 85: Modification inserted in MDB1, 6th and 7th or 11th and 12th OPR motif, they were chosen following the draft OPR code, for the CC1 and CC2 variant, one of the "unreadable" residue combination was picked at random.

The mutations were introduced in vectors containing a partially spliced *MDB1* sequence, tagged with either an HA tag, or with both HA and Strep tags (Figure 86). The vectors also bear a paromomycin resistance gene for transformants selection.

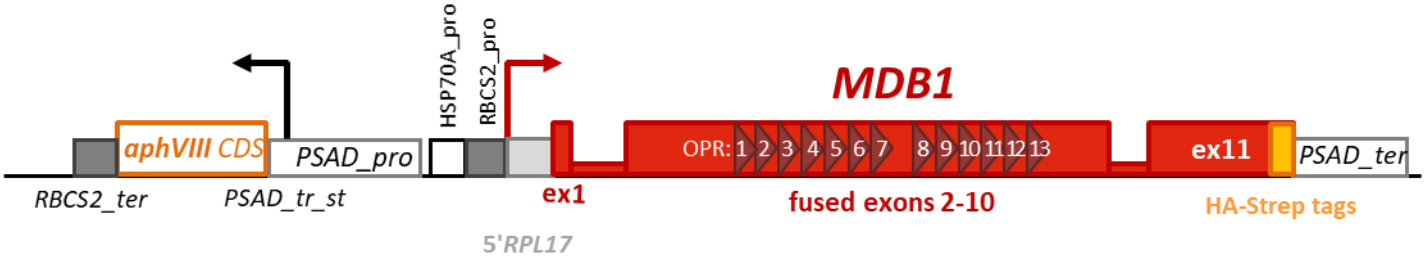


Figure 86: The *MDB1mut* chassis construct, with the *aphVIII* paromomycin resistance gene for selection of transformants.

NUCLEAR TRANSFORMATION

As the aim was to study the interaction of the modified MDB1 as isolated as possible on their target sequence, care was needed to make sure that the endogenous MDB1 would not interfere. The modified *MDB1* genes were transformed into *mdb1-1* (mt-) cells by electroporation. The transformants were selected on paromomycin resistance, then for restoration of photo-autotrophy. Indeed, the *mdb1-1* strain is incapable of photosynthesis, as it cannot accumulate at all *atpB* mRNA. Moreover, as we saw in Chapter III, MDB1 binding on a modified target *atpB* is quite resilient. And so, we expected that conversely, modified MDB1 might also be able to bind to some degree the endogenous WT *atpB* target sequence. Transformant plates were screened with a SpeedZen camera, to find partially rescued ATP synthase phenotypes.

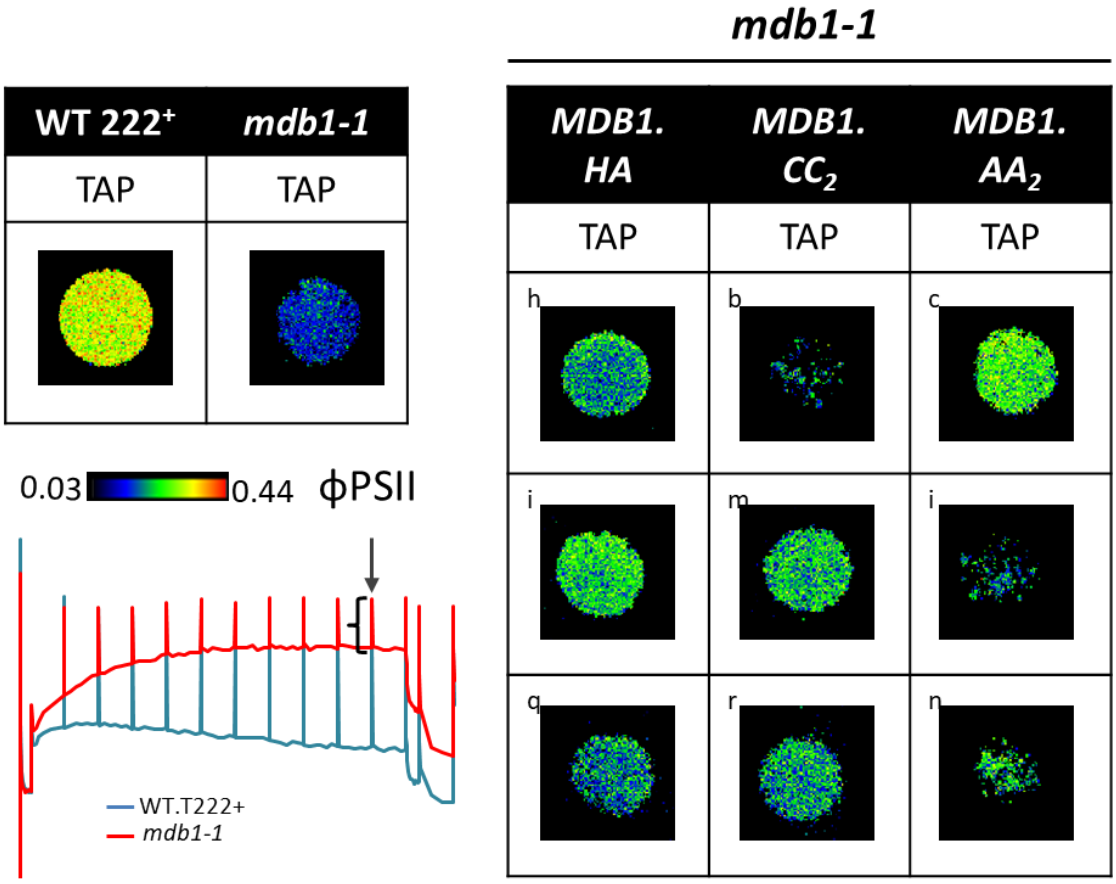


Figure 87: Photosynthetic activity of *MDB1mut* transformants, PSII fluorescence of cells on TAP media, ϕPSII, which reflect the proportion of open centres, was measured at the end of the continuous illumination phase.

Several independent candidates for each mutated *MDB1mut* variant were plated on minimum medium, to assess their ability to grow from photosynthesis alone. The three photo-autotrophic transformants with the best restoration of WT fluorescence kinetics were selected for each *MDB1mut* variant, for further characterisation.

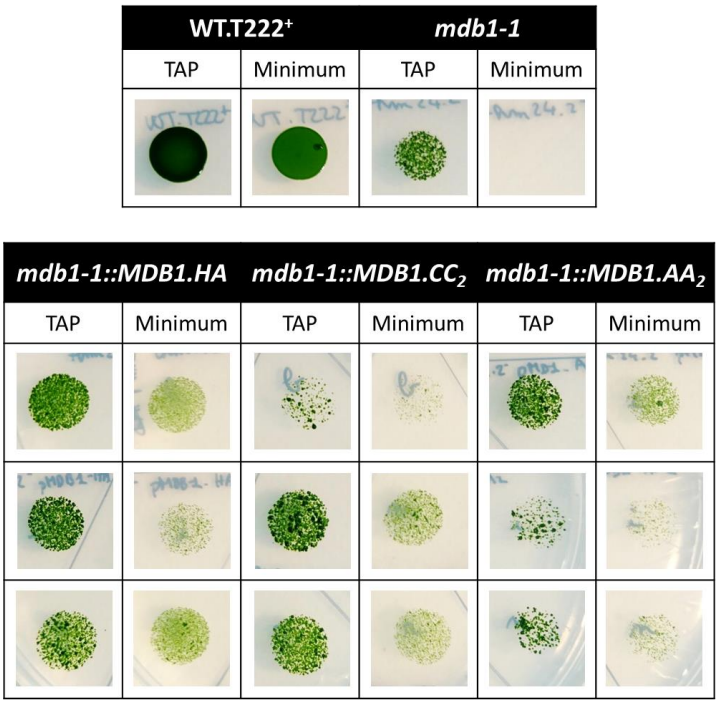


Figure 88: Growth phenotypes of *mdb1-1::MDB1mut* transformants. Cells were grown for seven days under 55µE/m^2.s on minimum or TAP media.

DNA integration in the nuclear genome is random, so we finally had to monitor MDB1mut expression. Proteins were extracted from three photoautotrophic transformants for each mutation. The levels of β CF1, and HA-tagged MDB1mut protein was assessed by immunoblot, an example is shown in Figure 89. The transformants expressing the highest level of the tagged MDB1mut were selected for subsequent work.

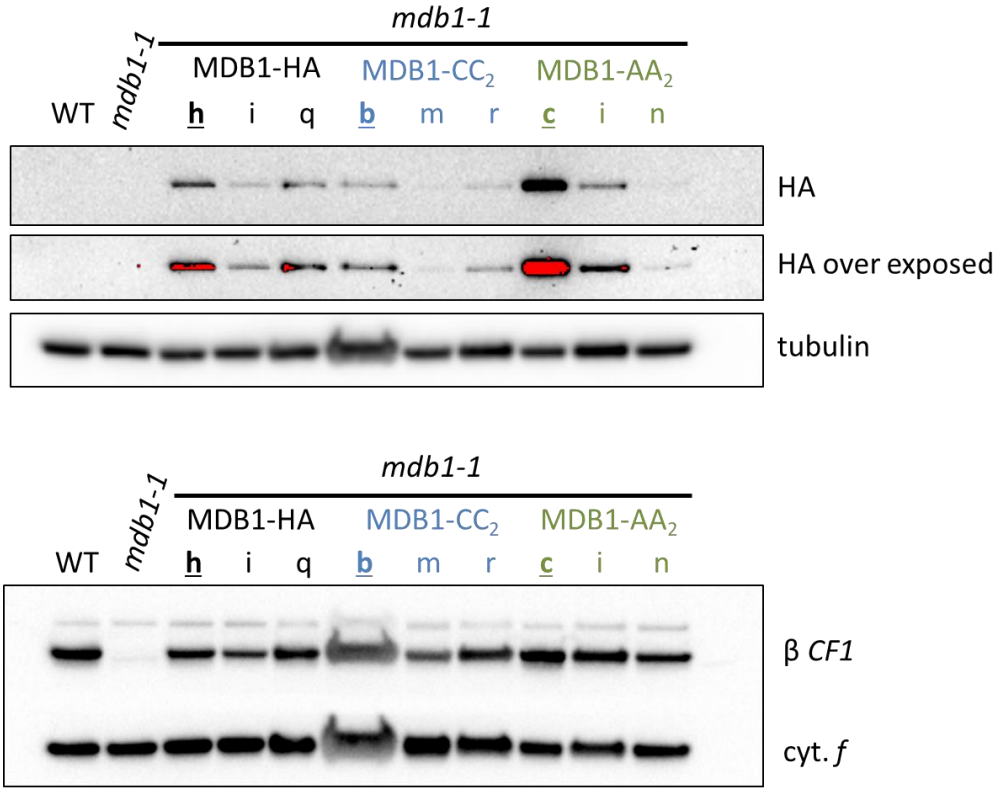


Figure 89: Characterisation of three photoautotrophic independent transformants for the MDB1-HA, MDB1.CC$_2$ and MDB1.AA$_2$ variants by immunoblots. Primary antibodies against HA, β CF1 (AtpB), tubulin and cyt. *f* were used. Underlined transformants were selected for the subsequent crosses.

During this process I noted an interesting property of the MDB1.CC$_1$ transformants: recovering photo-autotrophic cells was far more difficult. Most of the transformants still had a deficient ATP synthase phenotype, even if they were resistant to paromomycin. After selection on minimum media I could recover a few transformants able to perform photosynthesis. And after looking at their MDB1.CC$_1$ expression levels on immunoblots (Figure 90), all three of them were accumulating far more mutated MDB1 than the transformants of the other constructs. While this might be simply due to chance, I suspect that it might indicate that the introduced mutations: Q in position 6, in repeat 6 and 7 lower considerably the binding affinity of MDB1 for WT *atpB* target sequence.

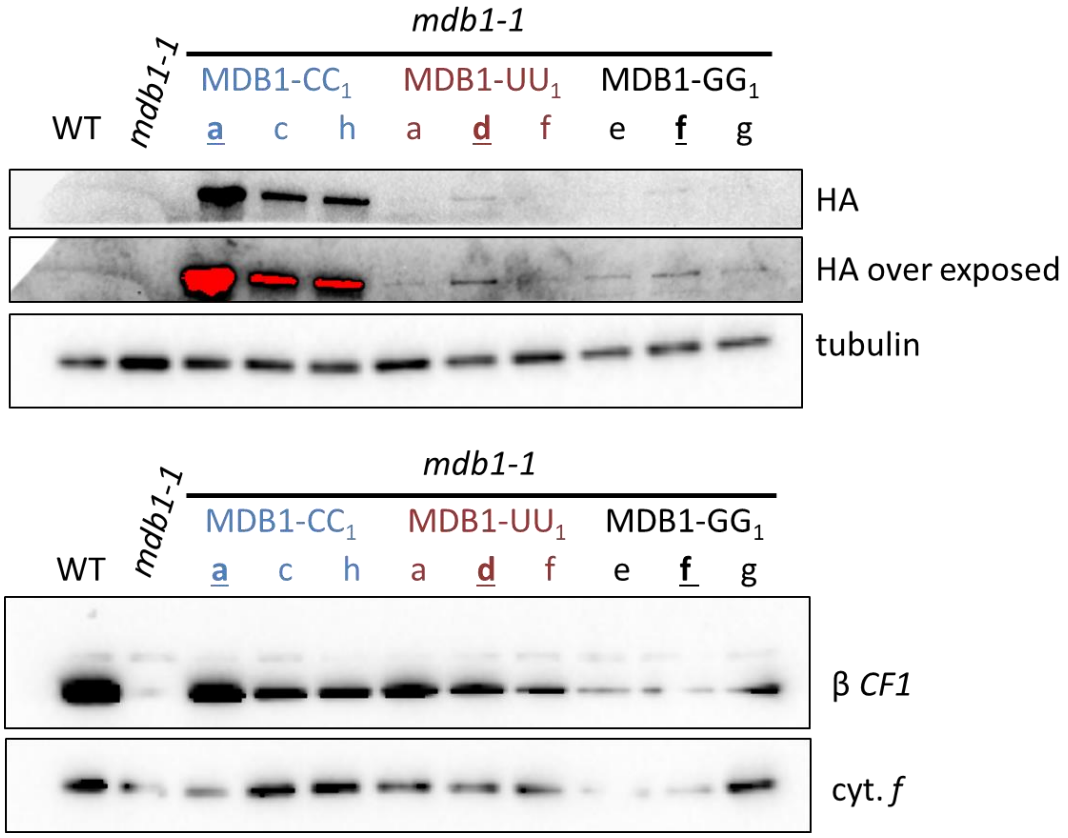


Figure 90: Characterisation of three photoautotrophic independent transformants for the MDB1-CC$_1$, MDB1.UU$_1$ and MDB1.GG$_1$ variants by immunoblots. Primary antibodies against HA, β *CF1* (AtpB), tubulin and cyt. *f* were used. Underlined transformants were selected for the subsequent crosses.

CONFRONTING THE OPR VARIANTS TO EACH TARGET MRNA BY CROSSES

To test the OPR recognition code, the different MDB1 mutants, with their modified OPR repeats, must be confronted with each mutated chloroplast target sequence. One of the difficulties with nuclear mutations, as we do not yet master the CRISPR system, is that insertions are random. Therefore, the expression levels of insertions vary widely in independent transformants, as can be seen in (Figure 89 and Figure 90). This would prevent us from comparing the accumulation of the reporter transcript in independent {*ΔatpB::dB$_M$gfp*} transformed strains. Moreover, if we had decided to first recover nuclear mutants then transform their chloroplast, it would have been very long to first delete *atpB*, excise the cassette, and then insert our chimeras in a neutral locus. I thus performed crosses, as the expression level of a transgene is similar in the progeny of a cross to that observed in the transgenic parent (Raynaud *et al.*, 2007; Boulouis *et al.*, 2011).

I crossed the *mdb1-1::MDB1mut* transformants previously selected with the {*ΔatpB::dB$_M$gfp*} strains. Since the MDB1/target sequence interactions seemed more affected in target variants modified in the 3' half of the target sequence: CC_2, AA_2 and GG_2, we decided to work first with the series of mutant affected in the second part of the target: *mdb1-1::MDB1-HA-strep* (the control MDB1), *mdb1-1::MDB1.CC$_2$*, *mdb1-1::MDB1.AA$_2$*, *mdb1-1::MDB1.GG$_2$* on one hand and {*ΔatpB::dB$_{WT}$gfp*}, {*ΔatpB::dB$_{CC2}$gfp*}, {*ΔatpB::dB$_{AA2}$gfp*} and {*ΔatpB::dB$_{GG2}$gfp*} on the other hand. Following the rule of genetic segregation in *C. reinhardtii*, where the mt+ parent transmits uniparentally its chloroplast genome to the whole progeny, while nuclear genes follow Mendelian inheritance, and assuming independence of the *MDB1mut* insertion with the *MDB1* locus, the descendant strains should be as described in Table 6.

	mt-, *mdb1-1::MDB1mut*-Paro^R {WT}	
mt+, *MDB1$_{WT}$* {*ΔatpB::dB$_M$gfp* Spec^R}	*MDB1$_{WT}$* {*ΔatpB::dB$_M$gfp* Spec^R}	*mdb1-1* {*ΔatpB::dB$_M$gfp* Spec^R}
	mdb1-1, *MDB1mut*-Paro^R {*ΔatpB::dB$_M$gfp* Spec^R}	*MDB1$_{WT}$*, *MDB1mut*-Paro^R {*ΔatpB::dB$_M$gfp* Spec^R}

Table 6: Expected descendants from a [mt+ {*ΔatpB::dB$_M$gfp* SpecR}] X [mt- *MDB1mut*-ParoR] cross. Grey shaded strains are sensitive to either spectinomycin or paromomycin and should not survive our double selection screen.

I crossed each MDB1 variant strain with each {*ΔatpB*::*dB$_M$gfp*} variant. After 10 days of maturation, zygotes were exposed to light in TAP medium to induce germination and plated on double selective medium (containing spectinomycin and paromomycin) to kill any surviving vegetative cell. The spectinomycin kills the *MDB1mut* parent and the paromomycin kills the {*ΔatpB*::*dB$_M$gfp*} parent as well as the progeny that did not inherit the *MDB1mut* allele (Table 6). Descendant colonies started emerging after about 10 days.

The whole procedure is described in Materials and Methods p144, it is quite time consuming, and unfortunately my first attempt was unsuccessful; none of the 16 crosses had worked. I had to restart the whole process with barely any time left. Thankfully, my second attempt worked for every single cross, and I recently analysed the descendants.

Analysing the descendants

As can be seen in Table 6, cells resistant to both antibiotics may contain, in addition to the mutated MDB1 transgene, either the wild type MDB1 allele or the mutant *mdb1-1* allele. In progeny expressing both the *MDB1mut* and *MDB1* alleles, any conclusions on the code would be completely wrong. And so, we carefully checked the absence of the functional *MDB1* allele in the descendants. In strain *mdb1-1*, the deletion of a single A in exon 5 creates a new *Bsr*I restriction site while removing the *Bst*XI site present in the WT sequence. The regions of *MDB1* surrounding the mutations (Figure 91.A) of 8 descendants for each cross were amplified by PCR. *MDB1mut* sequences were not amplified since the primers hybridise to introns 3 and 5 respectively that were not kept in the *MDB1mut* artificial construct (Figure 86). Amplicons were digested by *Bsr*I. A subset of the tested descendants is presented in Figure 91.B and reveals a random distribution of the $MDB1_{WT}$ and *mdb1-1* alleles among the 8 descendants analysed for each cross, as expected from Mendelian segregation of the two *MBD1* alleles. *mdb1-1* descendants could be recovered from each cross.

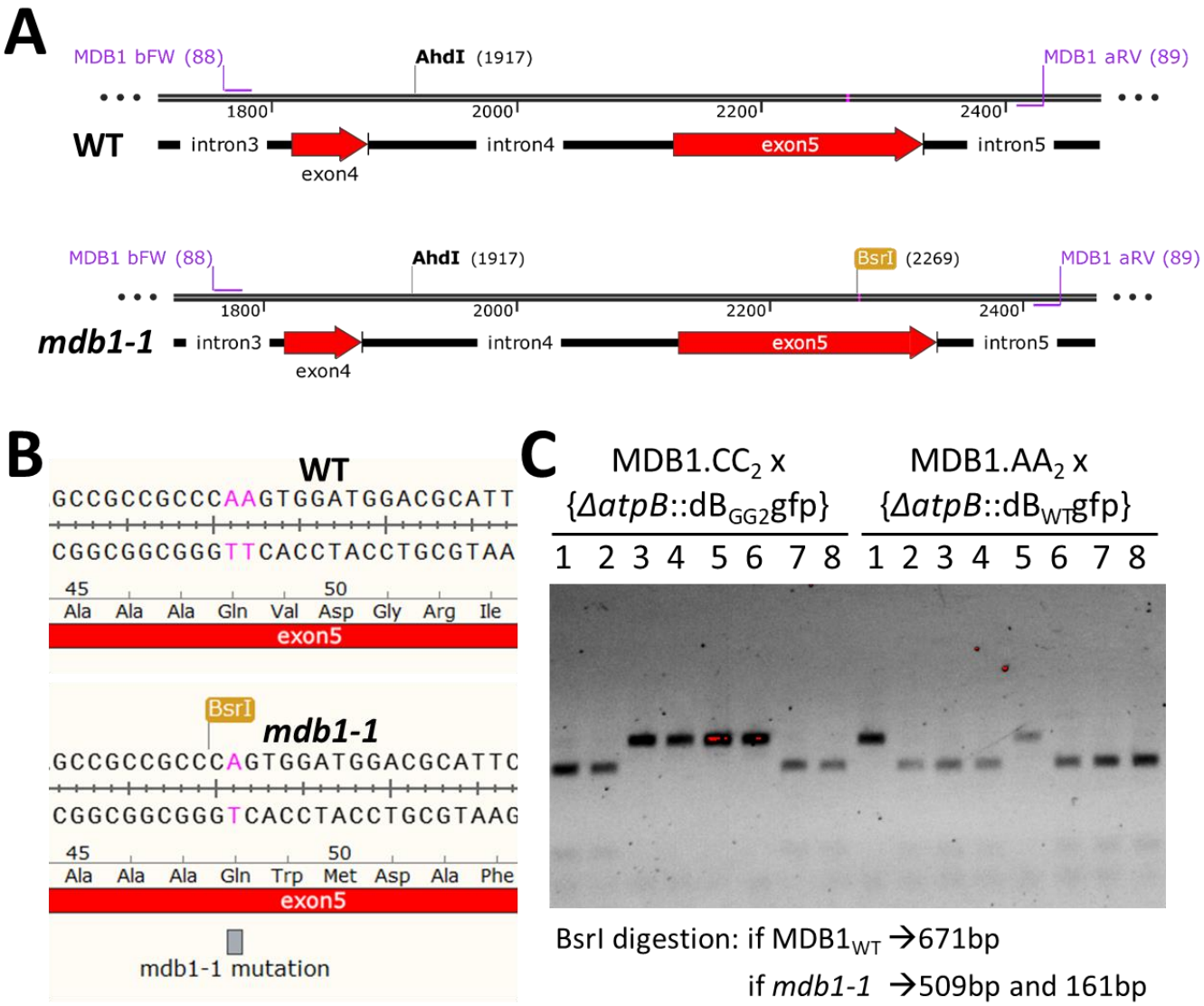


Figure 91: A. PCR amplification of MDB1 exon 5, B. a *Bsr*I restriction site is induced by the deletion of one A in *mdb1-1*. C. Digestion of subsequent amplicons with *Bsr*I. the *mdb1-1* allele A insertion creates a *Bsr*I site that does not exist otherwise in MDB1 5th exon.

COULD A PRELIMINARY EXPERIMENT HOLD PROMISES?

We recently extracted and started to analyse on RNA blots those descendants. Due to time constraints, a few of the biological replicates are missing but overall, we managed to obtain RNAs for each combination. Preliminary blots showing one descendant of each cross are presented in Figure 92.

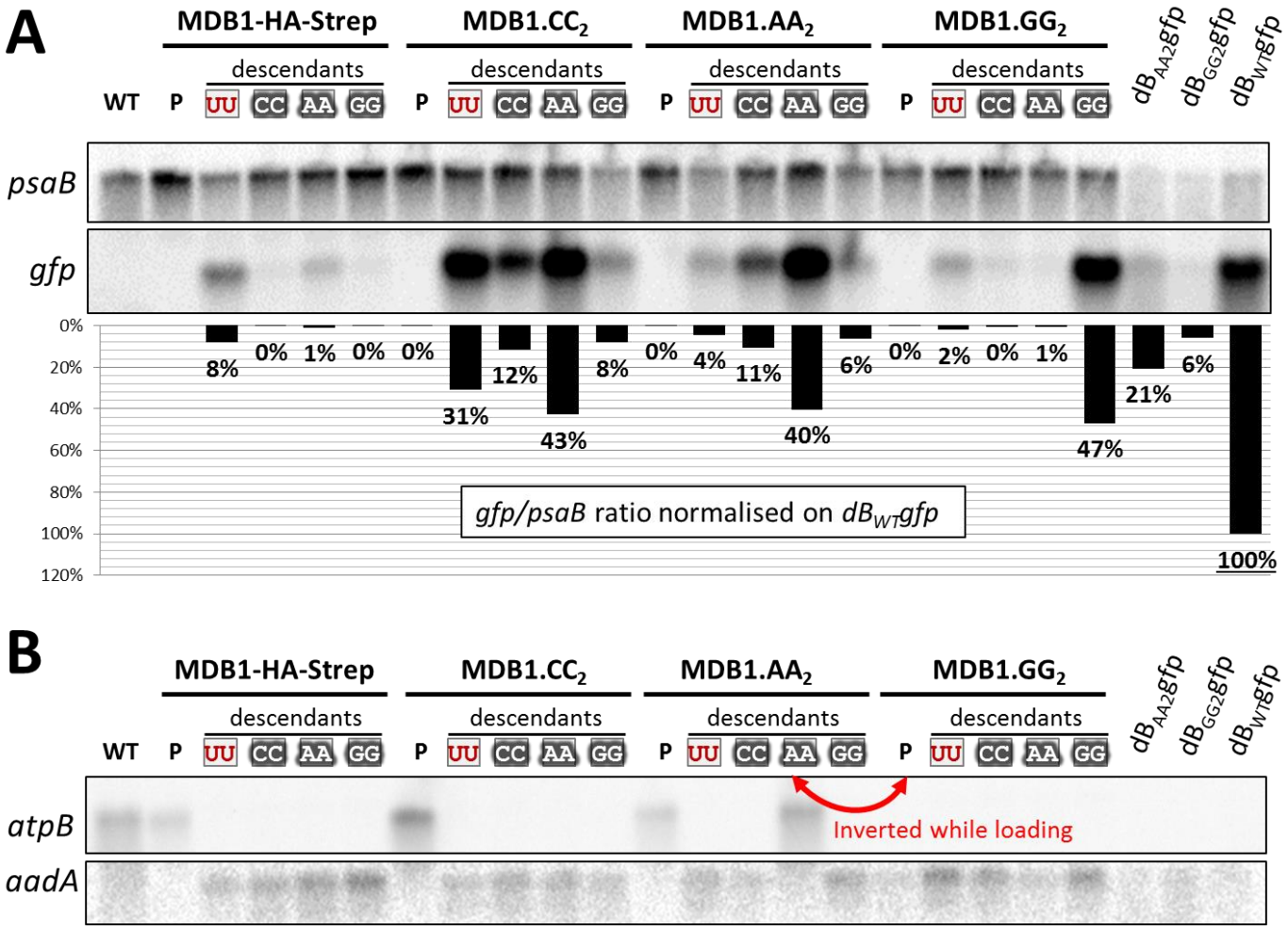


Figure 92: Preliminary RNA blots showing the first descendant of each of the *MDB1mut* x *dB$_M$gfp* cross. A. Filter was hybridised with ^{33}P labelled *gfp* and *psaB* (loading control) probes. **B.** Filter was hybridised with ^{33}P labelled *atpB* and *aadA* probes. P are the MDB1mut parents.

Do take notice that in Figure 92.A the relative levels of *gfp* mRNA between the different MDB1mut do not directly reflect the stabilising properties of the different OPR motifs, but probably rather the accumulation level of the modified MDB1 proteins. Reporter transcript levels must be compared in the same cross, and even then, we need to be careful, as unfortunately I could not yet assess the protein expression levels of each descendant. I extracted the proteins but ran out of time to perform immunoblots. These blots should allow us to check whether the accumulation of the MDB1mut proteins is the same in the descendants, as should be expected. Therefore, it is for now impossible to rule out that variations in the chimeric transcripts accumulation could stem from variations in the MDB1mut levels.

Another caution to keep in mind is that our reference *dB$_{WT}$gfp* RNA, seems affected by degradation. The long RNAs like *psaB*, which are more sensitive to degradation, are probably underestimated in our reference. This probably induces a bias in our quantification, by lowering the ratio of detected *gfp* of the descendant strains. A new RNA extract of *dB$_{WT}$gfp* for a proper reference must be added in our future experiments. However, while keeping these limitations in mind, we could tentatively draw preliminary conclusions on the recognition of nucleotides by the different OPR motifs from this blot.

First: the absence of *atpB* mRNA accumulation in the strains was verified, to ensure that the descendants are really *ΔatpB* (two samples were inverted while loading the gel). Similarly, as expected, the MDB1mut parents and the WT did not express any *gfp* mRNA (Figure 92.B).

The descendants stemming from MDB1-HA-Strep, our control MDB1, crosses with the various *dB$_{M}$gfp* chloroplast variants displayed the same pattern of chimeric transcript accumulation than in our previous experiments (Figure 82). The UU_2 mRNA is best recognised by the wild type MDB1 11th and 12th OPR repeats. The AA_2 mRNA is slightly stabilised, while the CC_2 and GG_2 mRNA are not stabilised at all. The low levels of *gfp* mRNA in this progeny, probably stem from a lower expression of our transformed MDB1 in the parent strain than the endogenous MDB1. Alternatively, the Strep or HA tags could hinder the MDB1 stabilisation ability. But this does not seem in line with the relatively strong accumulation of the chimeric transcripts allowed by our other MDB1mut, which also bear a HA and Strep tag.

MDB1.CC_2 in contrast appears to be able to stabilise to some extent all the target variants. It appears to stabilise a bit less efficiently the GG_2 mRNA than the UU_2, CC_2 and AA_2. Nonetheless, it appears able to bind the four different targets. Moreover, it does not seem to favour CC_2 compared to AA_2 and UU_2.

MDB1.AA_2 seems to stabilise the AA_2 mRNA, manages to stabilise CC_2 but barely UU2 and GG_2.

MDB1.GG_2 appears not to stabilise the CC_2 and AA_2 transcripts but stabilises the GG_2 and slightly the UU_2 ones.

Altogether this blot (Figure 92) shows that modifying the fifth and sixth residues of the OPR repeats change which target variant is best stabilised, and thus should reflect modifications in the specific recognition of nucleotides by MDB1.

However, there is a caveat. In addition to these two RNA blots we performed three more, with biological replicates Figure 93, and our story becomes considerably more complicated… And our results unreliable for now.

BIOLOGICAL REPLICATES POINT TO DIFFICULT AND RISKY INTERPRETATIONS…

Several descendants of the same crosses display highly different chimeric RNA accumulation. To be able to draw any information from these, we need first to understand the reason for this discrepancy. Is it technical? We had little time to perform the RNA extraction and RNA blots, so mistakes might have occurred. But the problem might be biological: is the protein level the same in all descendants? Did some genetic recombination or unsuspected genetic determinant induce a bias in our process? Unfortunately, I am not able to answer these questions for now.

Some of the crosses appear consistent, MDB1.CC_2 has a low affinity for the reporter mRNA in all the GG_2 descendants. MDB1.AA_2 has a low affinity for both the UU_2 and GG_2 chimeric RNA and slightly more for the CC_2. MDB1.GG_2 does not bind at all the CC_2 and AA_2 transcripts, but a bit the UU_2 ones. But if we intend to draw real and reliable conclusions on the OPR recognition code it is of the utmost importance to analyse the MDB1mut protein levels. And if the immunoblots revealed incongruence in protein accumulation we will need to amend these analyses.

I hope to succeed in formally validating or infirming the OPR code in the coming months. This study of the OPR code is described in ***ARTICLE 4***, an article in preparation, attached at the end of this manuscript.

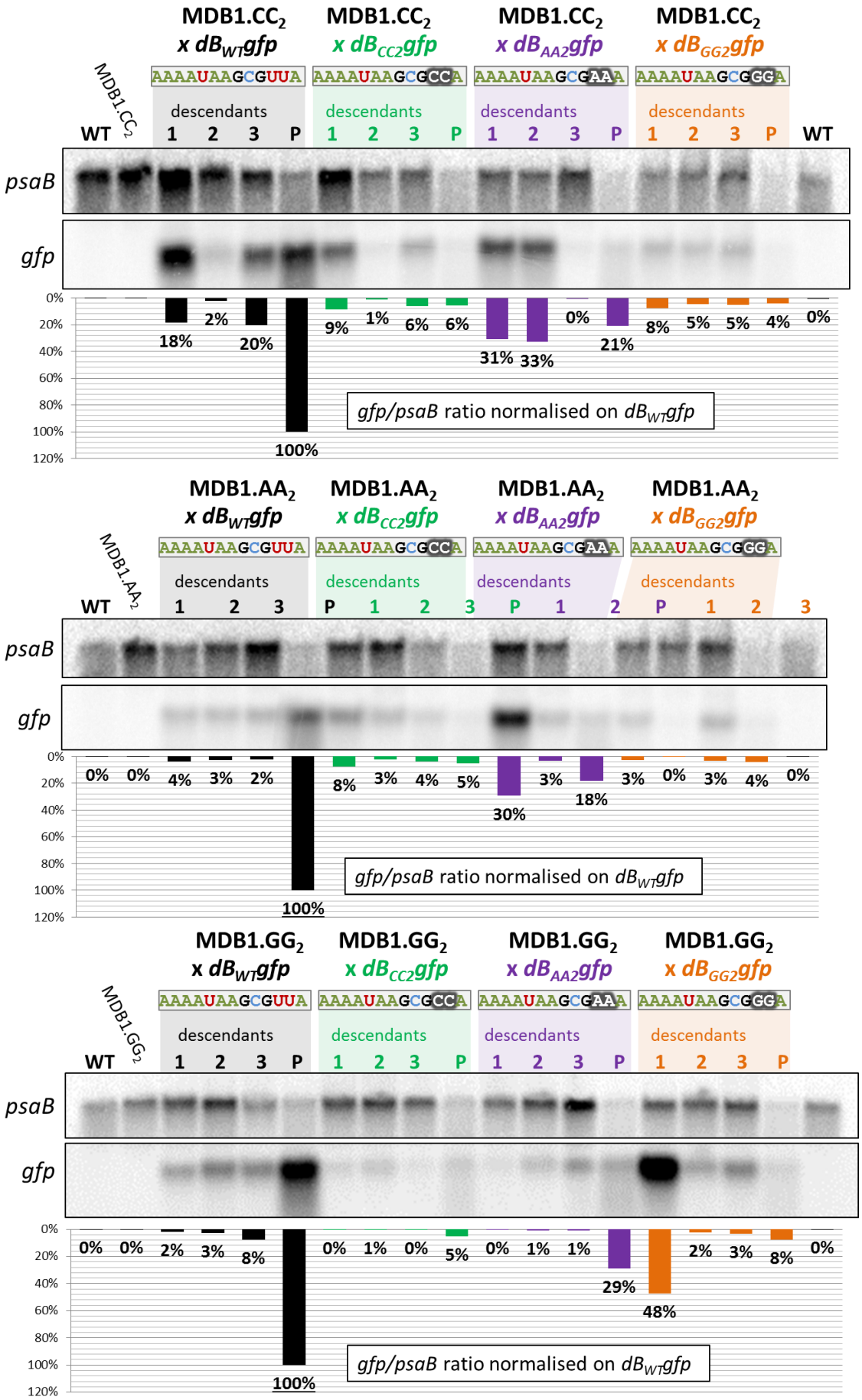


Figure 93: Preliminary RNA blots of the MDB1.CC$_2$ MDB1.AA$_2$ and MDB1.GG$_2$ descendants. Filters were hybridised with ^{33}P labelled *gfp* and *psaB* (loading control) probes. P are the *dB$_M$gfp* parents, 1; 2 and 3 the descendants of a same cross.

DISCUSSION

CHIMERIC SYSTEMS AND COMPETITION WITH *ATPB* MRNA

All chimeric reporter transcripts, either with added Spinach2 aptamers or not, are dramatically less stabilised in presence of the competing *atpB* mRNA. Could we completely exclude that this is caused by a more active transcription of the *atpB* mRNA? When present in greater quantities, it would mechanically and passively sequester much of MDB1, leaving very few for the lowly transcribed chimeric transcript.

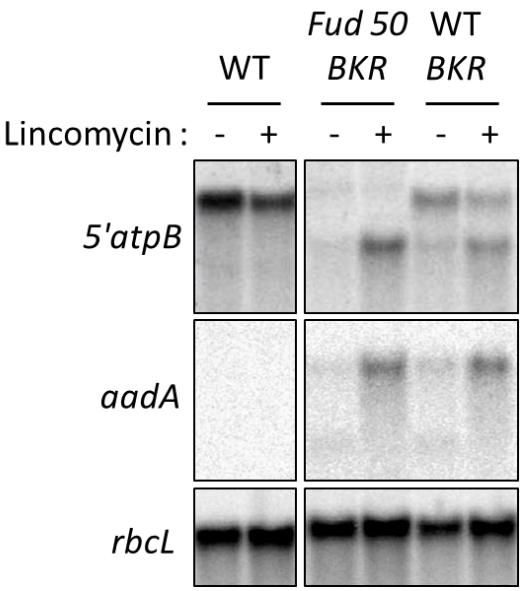


Figure 94: RNA blot of a chimeric construct bearing the entire *atpB* 5' UTR, from Yves Choquet. The highest band hybridised with the 5'atpB probe is the endogenous *atpB* transcript, the lower one the chimeric one. Lincomycin prevents chloroplast translation.

IS THE *GFP* MRNA LESS TRANSCRIBED THAN *ATPB* MRNA?

Looking at another chimeric transcript like *BKR* (*atpB 5'*UTR:*aadA*:*rbcL* 3'UTR), whose transcription is driven by a complete *atpB* 5'UTR and inserted in a locus upstream of *petA*, it appears that the endogenous *atpB* is not more accumulated than the chimera (Figure 94).

In this blot the cells were treated or not with lincomycin before RNA extraction. Lincomycin stops the chloroplast translation. This treatment was needed because the *aadA* CDS is cleaved in 5' from transcripts when translated and that irrespective of the 5'UTR used. This is possibly caused by a specific endonuclease recognising a sequence in the beginning of *aadA* CDS, maybe an NCL protein? In any case, in classical RNA blots, the chimeric *aadA* constructs are very difficult to observe and lincomycin treatment allows to recover the full length *aadA* mRNA.

This blot suggests that neither the *rbcL* 3'UTR, nor an insertion locus different from the endogenous *atpB* region significantly impact the accumulation and probably the transcription of the *BKR* chimera.

Moreover, experiments from nearly 30 years ago (Blowers *et al.*, 1990) revealed that the transcription rate of an *atpB* 5'UTR:*uidA* fusion was the same as for the endogenous *atpB*, and that the promoter fragment retained in our construct should be as efficient as the full-length *atpB* 5'UTR (Blowers *et al.*, 1990; Klein *et al.*, 1992). Altogether, this suggests that transcription of our *gfp* chimera and the endogenous *atpB* are probably similar, and that differences in accumulation of the gfp chimeras are not caused by a transcriptional defect in comparison of *atpB*.

THE *GFP* MRNA IS LIKELY LESS STABILISED IN PRESENCE OF *ATPB* MRNA

Alternatively, instead of a modified transcription rate, I suggest that additional specific sequences in the *atpB* transcript favour its stabilisation.

A putative secondary structure formed by an interaction between the 5' and 3' UTR could maybe modify the stability of the transcripts, as the 3'*rbcL* seems to hinder the 5' end maturation of a*tpB* (Chapter I). However, this model (Figure 95) seems unlikely.

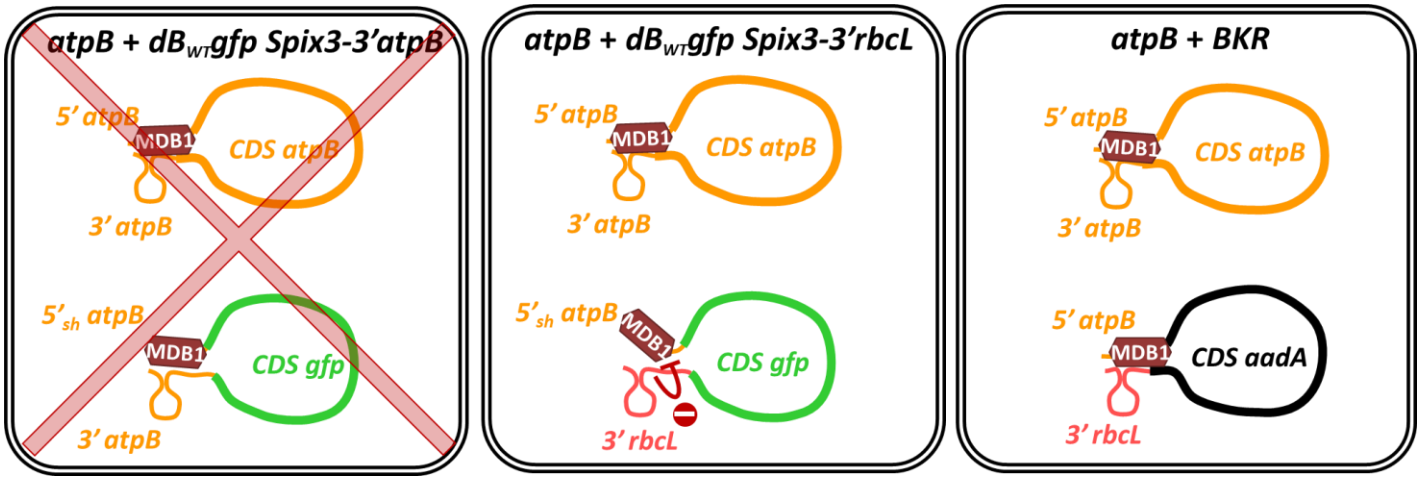


Figure 95: A tentative model based on a specific 5'$_{sh}$*atpB*/3'*rbcL* UTR destabilising interaction, does not seems likely as both atpB and rbcL 3' UTR gfp chimeric transcript are destabilised.

Indeed *atpB* 3'UTR by itself is unlikely to stabilise transcripts much better than *rbcL* 3'UTR, as Spinach2 chimeric transcripts bearing that *atpB* 3'UTR in place of *rbcL* 3'UTR are less stabilised, even in a *ΔatpB* context. Moreover, those reporter transcripts did not accumulate significantly more when in competition with the *atpB* mRNA (Figure 80). Unlike the *BKR* transcripts, that as we just saw compete well with the endogenous *atpB* mRNA, both the *dB$_{WT}$gfp.Spix3-3'rbcL* and *dB$_{WT}$gfp.Spix3-3'atpB* transcript accumulated weakly when *atpB* transcript was present. If the 3'end was a crucial actor of the stabilisation mechanism, we would have expected the chimeric transcripts bearing the *atpB* 3'UTR to be as stabilised as the *atpB* mRNA. Moreover, if the 3' UTR played a role in the competition for stabilisation factors, the {WT::*dB$_{WT}$gfp.Spix3-3'atpB*} transformants should accumulate less *atpB* mRNA and more *gfp* transcript. Surprisingly, we witnessed instead an increased accumulation for both transcripts.

Altogether, I suggest that specific sequences in the 5'UTR part that was deleted in the *gfp* chimera are important for transcript stabilisation. This stabilisation could be linked to the MDB1-mediated one, as the chimeric *BKR* transcript, driven by the complete *atpB* 5'UTR accumulates at about the same level than the endogenous *atpB* transcript in a competition (Figure 94). This could mean that unknown factor(s) could either recruit MDB1 on *atpB* transcript or anchor MDB1 by improving its affinity for its target RNA in a ternary complex for instance (Figure 96). This observation is also coherent with the lower resilience of the MDB1/*atpB* target when the end of the 5'UTR is absent

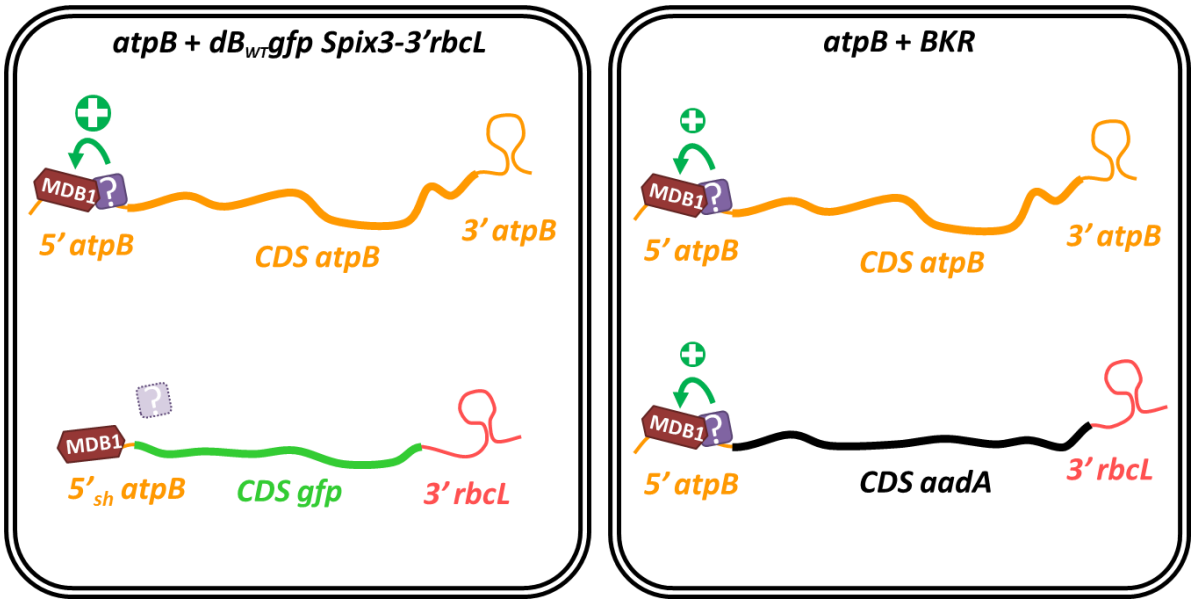


Figure 96: A more probable model explaining the different competitiveness of the BKR and *dB$_{WT}$gfp.Spix3-3'rbcL* chimera. The short *atpB* 5'UTR lacks specific sequences that improve *atpB* MDB1-mediated stabilisation.

ATPB 3'UTR MIGHT MEDIATES A DIMINUTION OF TRANSCRIPT ACCUMULATION

The surprising negative effect of *atpB* 3'UTR on the levels of *dB$_{WT}$gfp.Spix3-3'atpB* chimeric transcripts, suggests either that the *rbcL* 3'UTR stabilises better the *dB$_{WT}$gfp.Spix3-3'rbcL* chimera or that a specific endonuclease might target the 3'UTR of *atpB*, and that its presence in the chimera exposes it more to degradation than *rbcL* 3'UTR.

A look at the accumulation of chimeric transcripts bearing complete *atpB* 5'UTR with either the *atpB*, *petA* or *rbcL* 3'UTR (Figure 97), reveals that, in absence of translation, the *atpB* 3'UTR chimera is more accumulated that the others. This would suggest that *rbcL* 3'UTR might not really stabilise more efficiently our chimera. In addition, replacing *atpB* 3'UTR by *petA* or *rbcL* 3'UTR in another study (Rott *et al.*, 1998b) did not recover as much *atpB* transcript than in the WT.

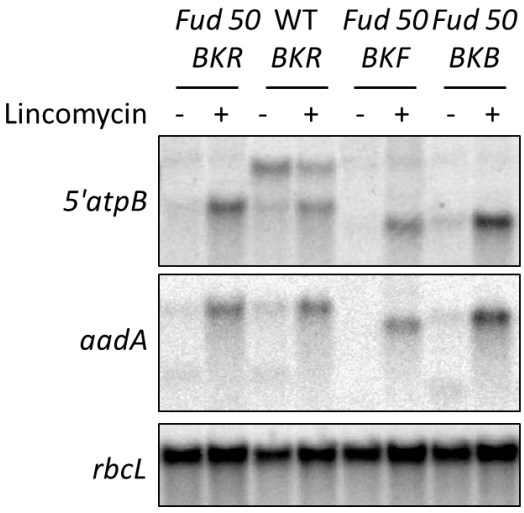


Figure 97: RNA blot of a chimeric constructs bearing the entire *atpB* 5' UTR, from Yves Choquet. The highest band hybridised with the 5'atpB probe is the endogenous *atpB* transcript, the lower one the chimeric one. Lincomycin prevents chloroplast translation.

Moreover, *atpB* half-life increases when translation is interrupted (Kato *et al.*, 2006). This effect does not transpire in Figure 94, where we look at the accumulation of the *atpB* mRNA. But this RNA accumulation reflects a combination of the transcription and decay rate of the transcript. Similarly, the authors had observed a stable level of the *atpB* transcript when translation was stopped, but an increased level in a pulse-chase experiment. This suggests that *atpB* transcription rate might adjust when a certain stock of mRNA is present.

I propose that a low abundance or low activity specific endonuclease might cleave open *atpB* 3'UTR and allow the degradation of *atpB* mRNA from its 3' end (Figure 98). The putative degradative 3'→5' process occurring in mature *atpB* mRNA 3' end described in Chapter I might be linked to this one. A family of putative specific endonucleases factors known in *C. reinhardtii* is the NCL (NCC-like), with 38 proteins. One of them might be responsible for the *atpB 3'UTR*-mediated destabilisation.

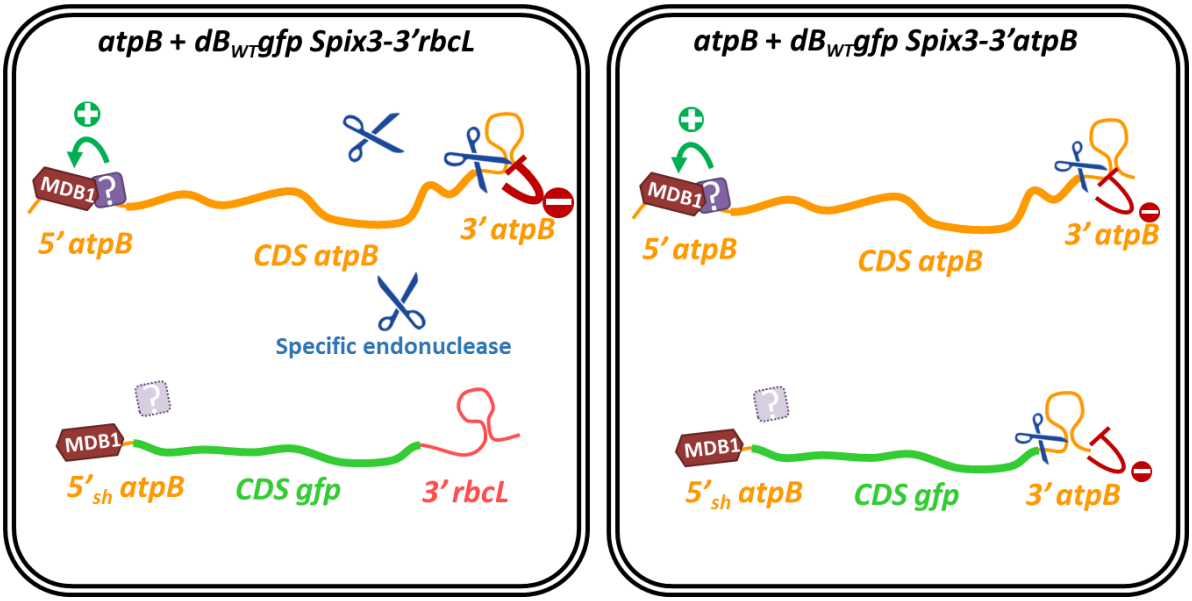


Figure 98: Putative model explaining the variation in accumulation of *atpB* mRNA in presence of the *dB$_{WT}$gfp Spix3-3'atpB or dB$_{WT}$gfp Spix3-3'rbcL* transcripts. A specific endonuclease of *atpB* 3'UTR would be distributed between the endogenous 3'*atpB* and the chimeric one.

CHIMERIC REPORTER CONSTRUCTS MIMIC AN "*IN VITRO*" SYSTEM

We confirmed that the use of a reporter chimeric transcript, with only a short part of *atpB* 5'UTR allowed us to study the interaction of MDB1 with its target sequence. The differences, compared to a control construct, in transcripts stabilisation of the variant versions between the *dB* and *dBgfp-3'rbcL* series was striking. By putting mutated target sequences in the chimeric reporter, we could confirm that the 3' half of the sequence was more critical for MDB1 binding. Moreover, we saw that the draft OPR code was coherent with the repeats studied so far: OPR repeats 11 and 12 of MDB1, with notably a glutamate in position six, recognise preferentially: UU>>AA>CC≥GG. From this it also emerges that the steric clash effect of purine versus pyrimidine might not be that crucial to the molecular recognition mechanism.

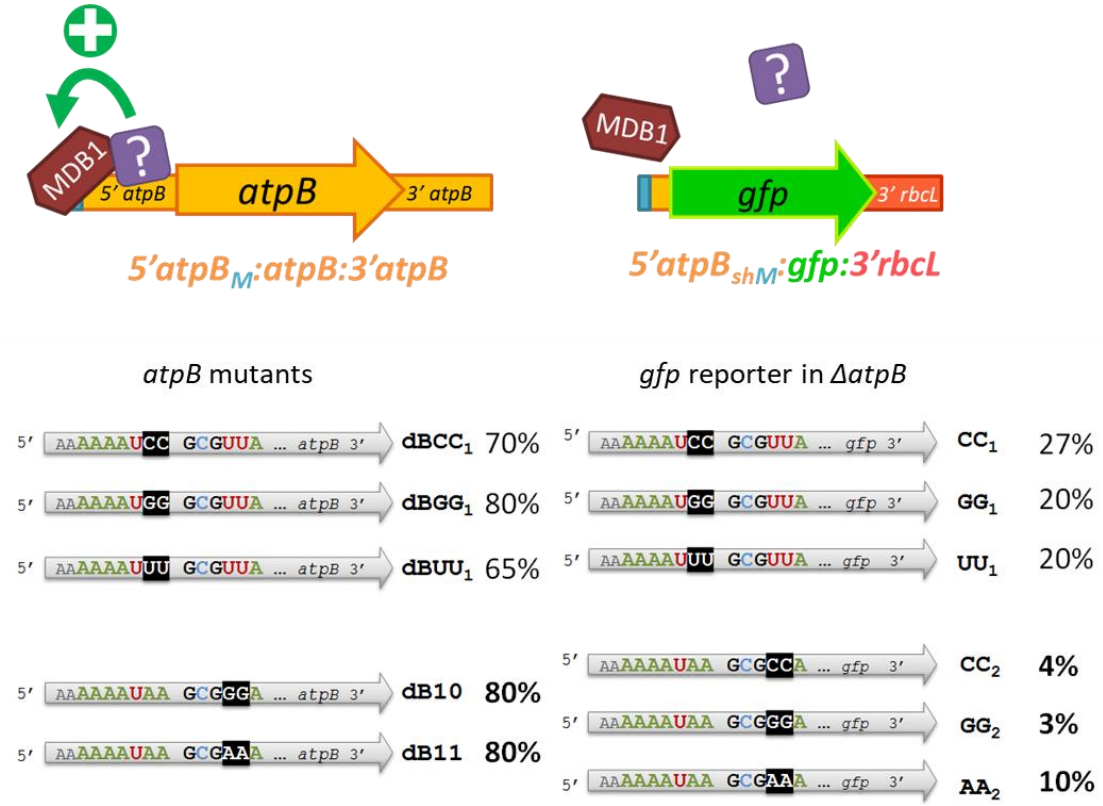


Figure 99: Comparison between some of our initial mutants (Chapter III) versus our mutated chimeric transcripts. The cartoon on top depicts a putative model of MDB1 interaction with the studied transcripts. Below are the relative accumulation levels of *atpB* MDB1 target variants compared to a control target.

VALIDATION OF THE OPR RECOGNITION CODE REMAINS ELUSIVE

Finally, we successfully established a strategy to test the molecular properties of the OPR code. Based on crosses, it should allow us to compare the accumulation of chimeric reporter genes in strain expressing a modified MDB1 at the same rate. After nuclear transformation by electroporation, our screens based first on the associated paromomycin resistance, then on the rescue of photosynthesis, let us select candidates transformants expressing sizeable level of MDB1mut. The expression of the HA-tagged MDB1 mutants was then directly checked by immunoblots. One strain expressing the highest amount of MDB1mut for each variant was selected for the subsequent crosses.

The chimeric *gfp* reporter, being linked to spectinomycin resistance, and MDB1mut, being linked to paromomycin resistance, allowed us to directly select the progeny on double selective media. Lastly, I was able to pick up those descendants expressing no endogenous MDB1 in addition to the mutated MDB1.

Unfortunately, time constraints prevent me to formally finish the validation of the OPR code for now. We performed RNA blots but many combinations remain unclear. I hope to be able to verify some of our observations on some combinations of the code by assessing the MDB1mut protein levels in the next months.

If the protein accumulation proves to be consistent, RNA blots should show the differences in nucleotide recognition induced by changes in the fifth and sixth residues of the 11th and 12th OPR repeats... Unless some unpredictable genetic effect complicates our observations, for instance: double insertions of the *MDB1mut* segregating independently in the descendants, or silencing of *MDB1mut* by epigenetic marks…

We need first to ascertain that no technical problems occurred in our RNA analysis. RNA blots must be remade; some RNA should maybe also be extracted again in case they were confusions in handling the many strains at once.

Position	Residue										
3	X	X	X	X	R, K	X	X	X	X	X	X
4	X	P	X	X	X - P	X	X	X	X	X	X
5	X	Q	X - R, K	R, K	X	R	X - R	R, Q	X	R	R
6	E	G	D	D	D	Q	Q	A	H	S	N
Recognised nucleotide	U	A	G	U	U	U	N	A	?	A	?

Not tested yet
Still unclear
Coherent
Novel observation
Validated

Table 7: The OPR code: where we stand.

Nonetheless if I were to discuss the more coherent apparent nucleotide specificities of the OPR repeat combinations we observed so far in RNA blots I would say: that U are recognised to some degree by the four MDB1mut. This is not too surprising considering that the MDB1mut transformants were selected on phototrophy recovery. Accordingly, they should be able to stabilise at least partially the endogenous *atpB* with the wild type MDB1 binding sequence, which contains UU. As was suggested by our previous experiments with the endogenous MDB1, an E in sixth position does appear to create a stronger affinity for U then A and barely any for C and G, more biological replicates must be tested to ascertain this claim.

The consistent biological replicates in our new blots suggests that having a D in the sixth position of the repeat prevents interaction with C or A. But the affinity of this motif for G remains unclear.

The use of TQ in the fifth and sixth positions of the repeats was a bit of a gamble as no clear correlation could be observed in known OPR/RNA example. And our results are similarly not clear cut; this TQ combination seems to bind any nucleotide, but to prefer U and A, rather than C and has less affinity for G. Obviously all those results need to be reproduced to draw solid conclusions.

It is very frustrating that some of our most anticipated combinations; MDB1.AA2 with *dBAA2gfp*, and MDB1.GG2 with *dBGG2gfp* are inconclusive. We cannot reject the predictions of the draft "OPR code" neither validate them for the time being.

From the experiments done so far, I would say that we did not find an OPR motif displaying a preference for C. It remains to be seen if a H or N in the sixth position induce a strong affinity for C. Otherwise, OPR proteins might bind only loosely on cytosines. Moreover, while U was tolerated by the motif with a D in sixth position (with no R or K in vicinity) C was not tolerated at all. And here, more than a question of size of the nucleotide (puric versus pyrimidic), the main determinant for the molecular interaction of this OPR motif might instead be the position of acceptor or donor atoms to form hydrogen bonds.

Another sticking point is that apparently the TQ residues combination does not show any clear preferences for certain nucleotides and appears to tolerate any of them to a large extent. This combination is found mainly in T factors. From this, I suggest that this kind of "looser" OPR motif in T factor OPR tracks might lower the binding specificity of the protein and induce the transient or light interaction of the T factor with its target mRNA.

Discussion and Conclusion

Modified from Dartmouth Electron Microscope Facility, Dartmouth College

MDB1 AND MTHI1 ARE KEY FACTORS FOR THE PLASTID ATP SYNTHASE EXPRESSION.

MDB1 is critical for *atpB* expression. It is a bona fide M factor of *atpB*; it does not only stabilise it but is also necessary for its 5' end maturation. This maturation appears to be accessory for chimeric transcripts translation. In plants *atpB* and *atpE* are expressed in the same polycistronic transcript, but in *Chlamydomonas*, it appears that both are expressed independently. No factor in common necessary for their expression has been found so far.

MTHI1, as a common essential factor controlling both *atpI* and *atpH*, ties their expression together, in the absence of a regulatory CES mechanism between those two critical subunits of the ATP synthase. The question of how the right stoichiometry of AtpH:AtpI is achieved has still not been elucidated yet. It likely does not derive directly from MTHI1 differential affinity for *atpH* and *atpI* transcripts but rather from other specific interacting factors.

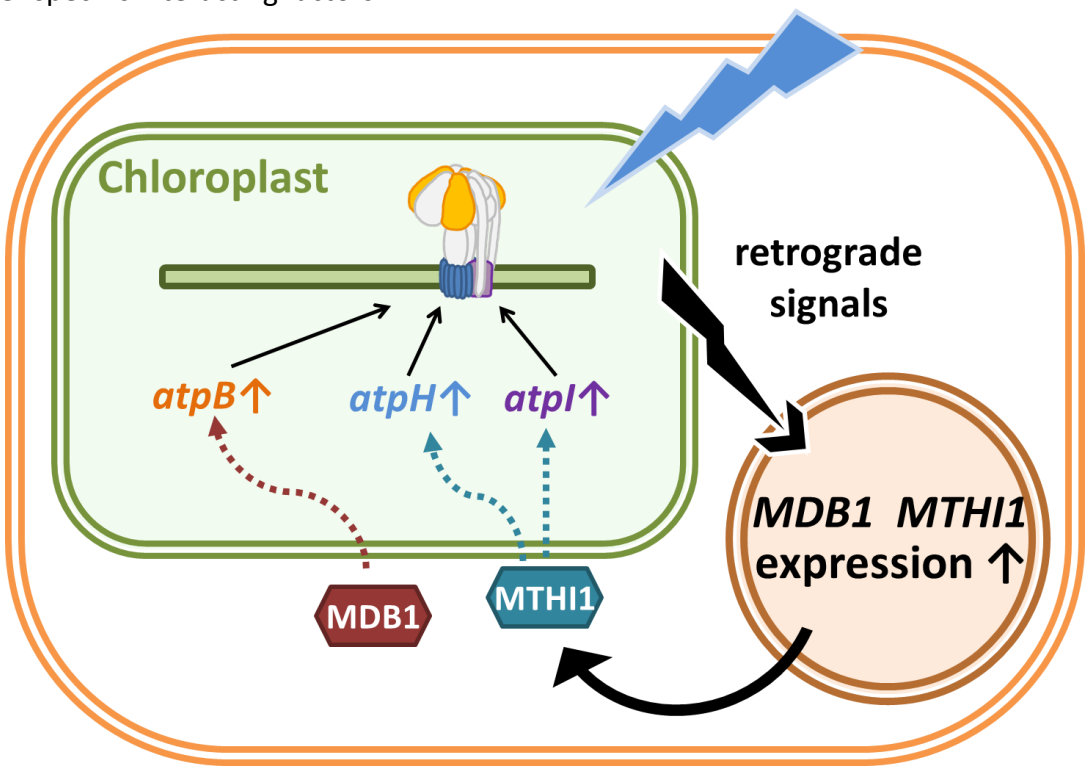


Figure 100 MDB1 and MTHI1 are part of the cross-talks between the chloroplast and the nucleus.

When the electron transport chain works at a high pace, an excessive proton concentration start building up inside the lumen. This excessive gradient needs to be dissipated by the ATP synthase.

Both factors are more expressed when the cells are exposed to light, this activation seems to rely on the bilin retrograde pathway. Following the circadian cycle, MDB1 and MTHI1 are expressed at night. Their transcript levels attain a pic just at light onset, then their expressions diminish until the next night. This expression pattern is similar to that of the nucleus-encoded ATP synthase subunits, to allow the harmonious assembly of all of ATP synthase subunits (Zones *et al.*, 2015).

THE "OPR CODE" AND *IN VIVO* REVELATIONS:

By designing a reporter construct, based on an exogenous sequence, driven by only a short part of the *atpB* 5'UTR, we could look more closely at the specificity of MDB1 for its target sequence. An observation that we made is that the 5' half of MDB1 target sequence on *atpB* appears less crucial for the binding of the M factor. This difference could stem from:

- Lower affinity of MDB1 for this part of the sequence, rendering MDB1 less sensitive to mutations in this area. As the chloroplast genome is very AT rich, the presence of A stretches, like in this part of the target sequence, is quite common. If MDB1 had a strong affinity for such sequences this might cause it to stall on other mRNA. As no other footprints of MDB1 could be observed so far (Cavaiuolo *et al.*, 2017), a lower affinity for the first part of the sequence is probable and could prevent such sterile interactions.
- Structural and functional properties of MDB1. The first OPR track of MDB1 could for instance be involved in "scanning" the mRNA, and the second track would lock on the specific sequence in the 3' part of the target. Interestingly, the part of the target recognised by the OPR motifs adjacent to the "hinge" seems quite important for the physical interaction.
- Some influences from the very upstream part of *atpB* 5'UTR. Could yet another factor modulates the binding of the first OPR track of MDB1?

Figure 101: Observations of the importance of parts of *atpB* target sequence for MDB1 binding.

To rule out the influence of yet other unpredictable *in vivo* factors in the MDB1/*atpB* interaction, it would be illuminating to perform *in vitro* experiments, with purified MDB1 and the mutated *atpB* variants. Those results would be great to contrast with the *in vivo* results.

Similar experiments with MTHI1 and *atpH* target variants would also help pinpointing the specificity of OPR M factors and mRNAs interactions.

Unfortunately, I could not yet gather the final data to assess whether the "OPR code" is consistent with *in vivo* molecular interactions. While the fifth and sixth residues of the OPR repeat appear implicated in the specificity of interaction, no conclusive observations on the OPR recognition code could be made so far. Unexpected variations in biological replicates considerably complicate our analyses.

But, had we not looked at the OPR protein/RNA interaction *in vivo*, but directly *in vitro*, we probably would not have discovered the truly resilient nature of those interactions. Those observations, that first baffled us, widened our understanding of OTAFs in *Chlamydomonas reinhardtii*. While organellar gene expression depends on crucial factors, those do not work alone and the expression of organellar mRNAs probably relies on a suite of factors, influencing each other, the M factors being the corner stones of these resilient expression edifices. Considering the growing number of M and T factors co-stabilising mRNAs in *C. reinhardtii*, those tripartite or higher order

complexes might be widespread in the chloroplast. One advantage of such a system would be to compensate moderate mutations in one of the target sequences, the other factor with its own recognised sequence would help to anchor its partner on the transcript.

This resilient nature of OPR M factors/RNA interaction *in vivo* might not be a lone case. Indeed, large scale *in vivo* studies of previously characterised PPR proteins of maize revealed unsuspected binding sites, which cannot be simply explained by the code established *in vitro* and *in silico*. This illustrates that PPR/mRNA recognition mechanisms can be more complicated than expected and that PPR interaction with mRNA are quite flexible (Rojas *et al.*, 2018). Considering that those characterised PPR have many PPR repeats (19 PPR motifs for PPR10 and 28 PPR motifs for PGR3) the fact that they do not show a stronger affinity might appear counter intuitive. But their recognition of several divergent targets might be caused by a higher tolerance for mismatches, as was observed for long designer PPR (Miranda *et al.*, 2018). Interestingly, in this study the authors observed a seemingly lower affinity of the PPR motifs at the 3' end that those at the 5' end, in contrary to our observations with MDB1. Perhaps the structure of MDB1, with its peculiar "hinge" allows its two OPR tracks to act as two short and semi-independent RNA binding units that do not display this weaker affinity in 3', quite the contrary. Moreover, both PPR10 and PGR3 collaborate with another factor, ATP4, to stabilise the 3' end of *psaJ* and of *rpl14*. It might turn out that to some extent, PPR proteins also interact with secondary factors, which would either enhance their action or tether them to transcripts that have less optimal binding sequences. This ATP4 PPR factor is implicated in various expression mechanisms (translation initiation or stabilisation) of various chloroplast mRNA: *atpA*, *atpB*, *atpE*, *atpF*, *psaJ* and *rpl14* (Zoschke *et al.*, 2012; Rojas *et al.*, 2018) but at various degrees: for example: its absence only reduces the accumulation of *atpF*, *rpl14* or *psaJ,* diminishes *atpA* translation, but is dramatic for the translation of *atpB*. Those striking multiple functions illustrate how ATP4 probably interacts with several other OTAF, like PPR10 and PGR3, and explain the observed functional redundancy of some processes. This suggests that in some cases, when a protein is absent, another one can rescue part of the function.

However, a directly redundant system is unlikely to exist in both of our OPR cases, as deletion of MTHI1 or MDB1 yield a total loss of *atpH* and *atpB* transcripts. Furthermore, extended mutation of the biding sites of those two M factor does prevent mRNA stabilisation. If the other factors were truly redundant, they would either bind on the same target, which is impossible considering that in absence of MDB1 or MTHI1 they do not, or bind on another target sequence and singlehandedly sustain some transcript accumulation, which does not seem to occur here. So, our putative secondary factors would need the M factor to exert their influence on the mRNA.

Even *in vitro*, the well-studied PPR10 was shown to perform non canonical interactions, impossible to predict following the current PPR code, with one of its target sequence, on *atpH* mRNA (Miranda, 2017). This complicates the prediction of target sequences *in vivo*. Moreover, PPR10 was found to be incapable of opening even weak secondary structures in its *atpH* target *in vitro* (McDermott *et al.*, 2018). This

could imply that other factors, such as RNA chaperones or helicases could open the secondary structure of *atpH* sequestrating the PPR10 target sequence.

OPR PROTEINS, ADVOCATES OF TEAMWORK?

The emerging picture of OPR action in *Chlamydomonas reinhardtii* is a network of actors cooperating for the expression of their chloroplastic target. For *atpH* expression, we found a secondary factor MTH2. This protein does not seem to have any discernible known domains and it is for now impossible to say whether it interacts directly with *atpH* mRNA or MTHI1. The probable implication of such a pioneer protein opens the door to unsuspected interactants implicated in chloroplast gene expression.

The search for the partners bound to MTHI1 would surely help to untangle those interactive *atpH* and *atpI* expression systems.

With our strep tagged MDB1 it would also be informative to immunoprecipitate it to pull down and characterise its putative partners. And to do that, either in the presence of the endogenous *atpB* or of our *gfp* chimeric construct, to assess whether the difference in resilience of MDB1 does stem from secondary factors or not. If so, I expect we would recover partners factors with the endogenous *atpB* mRNA. And in contrast, I assume we would not to find them with the chimeric transcript.

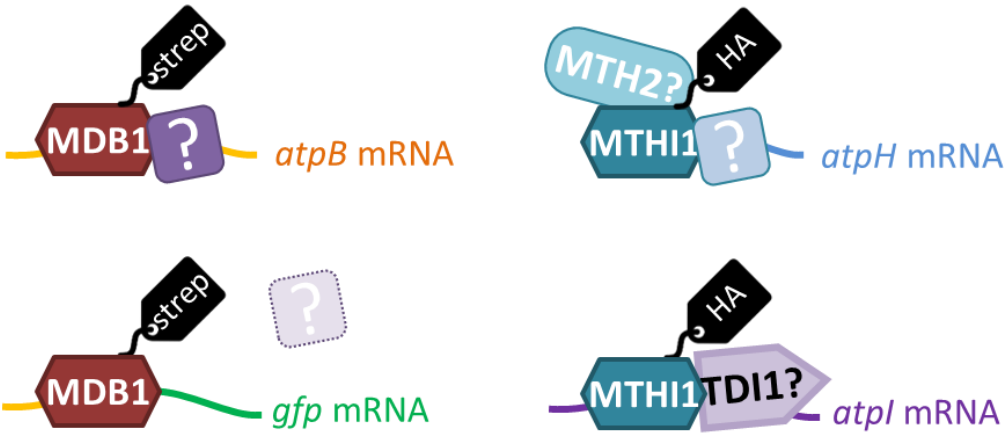


Figure 102: What we might find by immunoprecipitation of our tagged MDB1 and MTHI1.

Many of the OPR proteins of *C. reinhardtii* have not been characterised so far. Much mystery still surrounds them. In addition to their RNA binding activity they might also bear domains to interact with other proteins, other OTAFs to modulate their functions or perhaps to recruit ribosomes.

MTHI1 C terminal domain appears non-essential for its functions *in vivo*, both for *atpI* and *atpH* expression. *mthi1-1* deficient cells can be complemented with truncated MTHI1 (see ***ARTICLE 3***). This seems logical for the stabilisation function, as it is carried out by the OPR track. However, the association of MTHI1 with polysomes, and its implication in *atpI* translation even if does not bind stably on it, suggest that it might interact directly or indirectly with ribosomes. Maybe the OPR track bears by itself the ability to interact with ribosomes or other secondary factors, such as MTH2.

As the TPR proteins are α-solenoid proteins that are mostly implicated in protein/protein interactions, we could imagine that the OPR domains could interact with proteins. Therefore, OPR proteins might be implicated in unsuspected mechanisms.

OPR: AN OPPORTUNITY FOR EVOLUTIONARY PROTOTYPES

The resilience of chloroplast expression networks could allow a drift in OPR sequences. Acting as a "safety net", it would foster the emergence of new functions by ensuring that the chloroplast genes keep being expressed sufficiently. Auxiliary factors could then shift their activity.

ASA2, an OPR protein implicated in the mitochondrial ATP synthase complex of Chlorophyceae (Vazquez-Acevedo *et al.*, 2006; Murphy *et al.*, 2019) is an example of a "strange" function. Interestingly, a PPR protein (p18) is also associated with the mitochondrial ATP synthase of *Trypanosoma brucei* (Montgomery *et al.*, 2018). Is this a simple coincidence that reflects the random recruitment of those abundant proteins in the organelles to perform other functions? Or is it a true evolutionary convergence? It is conceivable that the α-solenoid structure of these proteins families, with its flexible properties could have a structural advantage in the function of the motile ATP synthase.

Finally, the NCL family (Drapier, 2002; Boulouis *et al.*, 2015) is the epitome of the dynamic evolutionary trajectory of OPR proteins. While their functions remain nebulous, their potential as rapidly evolving endonucleases could play a part in cell immunity for instance, by cleaving intruding viral RNA. Alternatively, they could have an impact in the speciation of chlamydomonales by acting as a barrier to sexual reproduction, by destroying the mRNAs of organelles genes with different alleles, leading to unviable descendants.

Considering that similar proteins, the HPR (Hillebrand *et al.*, 2018) and FASTK (Boehm *et al.*, 2017) exist in distantly related eukaryotes, like metazoans, this extended family of proteins with a RAP domain might have ancient origins.

OPR proteins have not been found so far in archaea or bacteria, except in pathogen bacteria like the intracellular *Coxiella burnetii*, *Parachlamydia acanthamoebae or Orientia tsutsugamushi*, which possibly acquired them from their host. The OPR family might take root in early eukaryotes before the divergence in bikonts and unikonts. This ancient family could have expended so much in green algae to ensure the expression of the chloroplast genome. In streptophytes the PPR proteins instead would have prospered.

It is intriguing to see so many parallels between those two protein families that appear unrelated. They seem to have emerged from convergent evolution. As the knowledge of PPR and OPR protein continues to expend, we might be able to compare them more closely and try to determine whether their different destinies in photosynthetic lineages reflects some of their differing specificities and properties or whether they expended by chance.

Materials and Methods

Modified from Dartmouth Electron Microscope Facility, Dartmouth College

MATERIALS

STRAINS

Strain	Mating type	Genotype
WT.T222	+	Wild type
WT.S24	-	Wild type
ΔatpB	+	Deletion of *atpB* and its 5'UTR
ΔatpH	+	Deletion of *atpH* and its 5'UTR
ΔatpI	+	Deletion of *atpI* and its 5'UTR
ΔpetA	+	Deletion of *petA* and its 5'UTR
mthi1-1 (ac46)		*MTHI1* mutant
mthi1-2 (II 174)	+	*MTHI1* mutant
mth2-1 (L63a)	+	Insertion in *MTH2* (Houille-Vernes *et al.*, 2011)
mth2-2	+	*MTH2* insertional mutant from Clip library (Li *et al.*, 2019)
mdb1-1 (thm24.2)	-	Deletion of one A in *MDB1* causing a frameshift
mdb1-2 (L35a)	+	Deletion of *MDB1* and 6 other genes(Houille-Vernes *et al.*, 2011)
mdb1-3 (K4.20)	+	TOC1 into *MDB1* first intron

PLASMIDS

Some of the plasmids used in this work are listed in ***ANNEX 1.***

PCR PRIMERS

The PCR primers used in this work are listed in ***ANNEX 2.***

METHODS

CULTURE CONDITIONS

Chlamydomonas reinhardtii cells were grown in Tris-acetate-phosphate medium (TAP), pH 7.2 at 25°C, under constant illumination at 5 to 10 $\mu E.m^{-2}s^{-1}$.

CROSSES

Crosses were performed according to (Harris, 1989). Strains were plated on N10 media to induce gametogenesis by nitrogen starvation. After 5 days, the cells were suspended in sterile water and the parental strains were mixed together. The cells were left to mate under high light. After 1 hour, 3 hours and 20 hours, aliquots of the crosses were deposited on TAP30 plates, then kept in the dark around 10 days for zygote development. Then, the zygotes were recovered after elimination of the vegetative cells with chloroform vapours and inoculated in TAP medium for two days under moderate light, to induce germination. Finally, those cultures were plated either on double selective media (spectinomycin and paromomycin) for the *MDB1mut* crosses, to kill any surviving vegetative cell, or on permissive media for the *K4.20* cross. Descendant colonies started emerging after about 10 days.

K4.20 progeny was selected on ATP synthase deficiency, then the presence of TOC1 and the mt were assessed by PCR. The presence of the *mdb1-1* allele was assessed by PCR amplification of the *MDB1* locus with MDB1 bFW and MDB1 aRV and subsequent digestion by *Bsr*I.

PHOTOTROPHY TEST

Droplets of liquid culture at about $1*10^6$ cells/mL are put on TAP and minimum media and grown for at least 7 days under 55 µE m^{-2} s^{-1} illumination. Growth phenotypes are then compared between the two media.

FLUORESCENCE LIVE-IMAGING

Fluorescence of live cells on plates was measured with a SpeedZen camera (Beambio). PSII maximal yield (Fv/Fm) is calculated from variable fluorescence (Fv=Fm-F_0) in dark-adapted cells using weak excitation pulses before (F_0) and after (Fm) a saturating flash

CLONING

TARGET MUTAGENESIS

Mutated *atpB* target fragments were generated first by PCR mutagenesis amplification of two mutated fragments with the primers listed in ***ANNEX2***, then assembly and amplification of the two fragments together with dBExt_RV and atpB5'FWx. The purified fragments were then digested with *Xho*I and *BseR*I and inserted in the p^{Kr}atpB plasmid at the same sites.

Mutated *atpH* target fragments were similarly produced with cemAFW and Mut-atpH-xRV, atpHext-RV and Mut-atpH-xFW. Assembled and amplified with cemAFW and atpHext_RV. The resulting fragments were digested with *EcoR*V and *EcoR*I and integrated in the p^{Kr}atpH plasmid at the same sites.

GFP CHIMERIC CONSTRUCTS

The paAKX plasmid (Wostrikoff et al, 2004) was digested by *Apa*I and *Ale*I to retrieve a 2509 bp fragment containing a spectinomycin resistance cassette (the aminoglyside 3' adenyl transferase coding sequence: *aadA* (Goldschmidt-Clermont, 1991) driven by the *psaA* 5'UTR and followed by *rbcL* 3'UTR. This cassette is also flanked by two direct repeats (a fragment of the *tet* gene conferring resistence to tetracycline (Fischer *et al.*, 1996)) of 485 bp, to create a recycling *aadA* cassette self-excising by spontaneous homologous recombination as described in (Fischer *et al.*, 1996). After a Klenow treatment, this fragment was inserted into the pWF plasmid (which contains chloroplastic sequences targeting the insertion in a neutral locus, next to *petA*) at the *Hinc*II site, giving the pWFaAKX plasmid.

A 756 bp *Azotobacter vinelandii* green fluorescence protein sequence was amplified by PCR from pGFP with the GFP-CDS_FW and GFP_CDS_RV2 primers (see primers table in ***ANNEX2***). This DNA fragment was digested by *Pst*I and *Eco*RI and integrated into the corresponding sites in the paAKRaA plasmid (Fu *et al.*, 2017) to place the *gfp* sequence in front of the *rbcL* 3'UTR, giving the pgfpRaA plasmid.
pgfpRaA was then digested by *Bam*HI and *Xho*I, the 946bp fragment was inserted into pWFaAKX at the *Xho*I and *Bgl*II sites, this yielded the pWFaAKXgfpR plasmid.

Fragments of *atpB* 5'UTR with the mutated target were obtained either: by PCR amplification of previously used plasmids p^{Kr}atpB, patpB$_{CC}$, patpB$_{TT}$, patpB$_{GG}$ with the atpB-Anton_FW and atpB_Anton_WT_RV primers, or by PCR mutagenesis with primers atpB-Anton_FW and atpB-Anton-M1_RV, atpB-Anton-M2_RV, atpB-Anton-M3_RV using p147 as template. All those 144bp amplifications products were then digested by *Xma*I and *Xho*I and inserted at the corresponding sites in pWFaAKXgfpR. The final pWFaAKX-dB(WT)gfpR, pWFaAKX-dB(CC_1)gfpR, pWFaAKX-dB(TT_1)gfpR, pWFaAKX-dB(GG_1)gfpR, pWFaAKX-dB(CC_2)gfpR, pWFaAKX-dB(AA_2)gfpR, pWFaAKX-dB(GG_2)gfpR were thus obtained.

SPINACH2 CONSTRUCTS

A triplet of consecutive Spinach2 aptamers (Strack *et al.*, 2013) sequences, separated by restriction sites, was ordered from GenScript and cut with *Mfe*I and *Pst*I and inserted in pgfpRaA to give pgfp-Spinach2x3-RaA. This plasmid was subjected to the same cloning procedure as previously described for pgfpRaA to obtain pWFaAKX-dB(WT)gfp-Spinach2x3-R. To obtain the construct with 2 spinach2, pWFaAKX-dB(WT)gfp-Spinach2x3-R was digested with *EcoR*I and *Hpa*I, and after a Klenow treatment to fill in the overhangs, was ligated. This yielded the plasmid pWFaAKX-dB(WT)gfp-Spinach2x2-R.

To design the chimeric construct with *atpB* 3'UTR instead of *rbcL* 3'UTR, a synthetic sequence was ordered and introduced into the pWFaAKX-dB(WT)gfp-Spinach2x3-R giving the pWFaAKX-dB(WT)gfp-Spinach2x3-B (GenScript).

MDB1 VARIANTS

Synthetic DNA sequences were ordered from GenScript and inserted either in the vector MDB1-HA-pJFL between the *Xho*I and *Bgl*II sites for MDB1-CC_1-HA, MDB1-GG_1-HA, and MDB1-UU_1-HA or in the vector pMDB1 -HA-Strep-JHL between the *Ns*iI and *Sna*BI sites for MDB1-CC_2-HA-Strep, MDB1-GG_2-HA-Strep and MDB1-AA_2-HA-Strep (GenScript) (see ***ANNEX 5***).

CHLOROPLAST TRANSFORMATION

Chloroplast transformation by tungsten microbeads bombardment (Boynton *et al.*, 1988) was conducted essentially as described (Kuras and Wollman, 1994) except that the cells were directly transformed on TAP-spectinomycin (100µg/mL) plates. Resulting transformants were sub-cloned on TAP-spec (500µg/mL) for several generations. Homoplasmy was assessed by PCR amplification of the construct sequence and disruption of the recipient loci.

RNA EXTRACTION AND RNA-BLOT

RNA extraction and RNA blots were performed as in (Drapier and Wollman, 1998), some with radioactive labelled probes as described. The others were performed with digoxigenin (DIG) labelled DNA probes generated by PCR (Roche) and hybridised on the nylon filter bound RNA. The probes were then bound by anti-DIG antibodies and incubated with CDP-*Star* (Roche), chemiluminescence was then detected with a Chemidoc. Transcript quantification was done using the image lab software.

PROTEIN EXTRACTION AND IMMUNO-BLOT

Immunoblots were performed on exponentially growing cells (2 x 10^6 cells/mL) according to (Kuras and Wollman, 1994). Cell extracts were loaded in 8-16% acrylamide gels (*Biorad*) or in constant 18% acrylamide 8M urea gels, on an equal chlorophyll basis. Anti-tubulin, anti-cytochrome *f*, anti-β-CF1, anti AtpH, anti-PsaD, anti-OEE2, anti-HA antibodies were used, and detected either by anti-mouse IgG or anti-rabbit IgG antibodies.

GENOMIC ANALYSIS

For each strain, we ordered a Illumina sequencing technology based NGSelect DNA data package from *Eurofins*, comprising the generation of a standard genomic library (DNA fragmentation, adapter ligation, size selection and amplification) and a data package of >5 million pair reads (2x150bp). From this raw data, we generated genomic sequences using open-source platform Galaxy (usegalaxy.org) as follows. Paired end reads raw data was converted to appropriate fastq format using FASTQ groomer and their quality was confirmed using FASTQC (maximum quality scores were well maintained all over the 150bp, not shown). The genome sequences were reconstructed using the published workflow "SNP calling on paired end data" for the mapping of the paired reads against a reference *C. reinhardtii* genome sequence (our WT strain T222+ genome was generated by Olivier Vallon). We visualized the genomes using IGV (software.broadinstitute.org).

REFERENCES

Allen, J. F. (2017) The CoRR hypothesis for genes in organelles. J Theor Biol, 434, 50-57.

Allen, J. F. and Raven, J. A. (1996) Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. Journal of molecular evolution, 42, 482-492.

Allison, L. A., Simon, L. D. and Maliga, P. (1996) Deletion of rpoB reveals a second distinct transcription system in plastids of higher plants. The EMBO journal, 15, 2802-2809.

Anthonisen, I. L., Salvador, M. L. and Klein, U. (2001) Specific sequence elements in the 5' untranslated regions of rbcL and atpB gene mRNas stabilize transcripts in the chloroplast of Chlamydomonas reinhardtii. RNA (New York, N.Y.), 7, 1024-1033.

Aphasizheva, I., Maslov, D. A., Qian, Y., Huang, L., Wang, Q., Costello, C. E. and Aphasizhev, R. (2016) Ribosome-associated pentatricopeptide repeat proteins function as translational activators in mitochondria of trypanosomes. Molecular microbiology, 99, 1043-1058.

Arnal, N., Quadrado, M., Simon, M. and Mireau, H. (2014) A restorer-of-fertility like pentatricopeptide repeat gene directs ribonucleolytic processing within the coding sequence of rps3-rpl16 and orf240a mitochondrial transcripts in Arabidopsis thaliana. The Plant journal : for cell and molecular biology, 78, 134-145.

Arndt, N. T. and Nisbet, E. G. (2012) Processes on the Young Earth and the Habitats of Early Life. Annual Review of Earth and Planetary Sciences, 40, 521-549.

Aubourg, S., Boudet, N., Kreis, M. and Lecharny, A. (2000) In Arabidopsis thaliana, 1% of the genome codes for a novel protein family unique to plants. Plant molecular biology, 42, 603-613.

Auchincloss, A. H., Zerges, W., Perron, K., Girard-Bascou, J. and Rochaix, J. D. (2002) Characterization of Tbc2, a nucleus-encoded factor specifically required for translation of the chloroplast psbC mRNA in Chlamydomonas reinhardtii. The Journal of cell biology, 157, 953-962.

Baier, T., Wichmann, J., Kruse, O. and Lauersen, K. J. (2018) Intron-containing algal transgenes mediate efficient recombinant gene expression in the green microalga Chlamydomonas reinhardtii. Nucleic acids research, 46, 6909-6919.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W. and Noble, W. S. (2009) MEME SUITE: tools for motif discovery and searching. Nucleic acids research, 37, W202-208.

Balczun, C., Bunse, A., Schwarz, C., Piotrowski, M. and Kuck, U. (2006) Chloroplast heat shock protein Cpn60 from Chlamydomonas reinhardtii exhibits a novel function as a group II intron-specific RNA-binding protein. FEBS letters, 580, 4527-4532.

Ball, S. G., Subtil, A., Bhattacharya, D., Moustafa, A., Weber, A. P., Gehre, L., Colleoni, C., Arias, M. C., Cenci, U. and Dauvillee, D. (2013) Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? The Plant cell, 25, 7-21.

Ban, T., Ke, J., Chen, R., Gu, X., Tan, M. H., Zhou, X. E., Kang, Y., Melcher, K., Zhu, J. K. and Xu, H. E. (2013) Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. The Journal of biological chemistry, 288, 31540-31548.

Barkan, A. and Goldschmidt-Clermont, M. (2000) Participation of nuclear genes in chloroplast gene expression. Biochimie, 82, 559-572.

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y. S., Bond, C. S. and Small, I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS genetics, 8, e1002910.

Barkan, A. and Small, I. (2014) Pentatricopeptide repeat proteins in plants. Annual review of plant biology, 65, 415-442.

Baross, J. A. H., S.E. (1985) Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life. Origins of Life, 327-345.

Beatty, J. T., Overmann, J., Lince, M. T., Manske, A. K., Lang, A. S., Blankenship, R. E., Van Dover, C. L., Martinson, T. A. and Plumley, F. G. (2005) An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. Proceedings of the National Academy of Sciences of the United States of America, 102, 9306-9310.

Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B. and Barkan, A. (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. Molecular and cellular biology, 28, 5337-5347.

Bell, S. A., Shen, C., Brown, A. and Hunt, A. G. (2016) Experimental Genome-Wide Determination of RNA Polyadenylation in Chlamydomonas reinhardtii. PloS one, 11, e0146107.

Berthold, P., Schmitt, R. and Mages, W. (2002) An engineered Streptomyces hygroscopicus aph 7" gene mediates dominant resistance against hygromycin B in Chlamydomonas reinhardtii. Protist, 153, 401-412.

Bietenhader, M., Martos, A., Tetaud, E., Aiyar, R. S., Sellem, C. H., Kucharczyk, R., Clauder-Munster, S., Giraud, M. F., Godard, F., Salin, B., Sagot, I., Gagneur, J., Dequard-Chablat, M., Contamine, V., Hermann-Le Denmat, S., Sainsard-Chanet, A., Steinmetz, L. M. and di Rago, J. P. (2012) Experimental relocation of the mitochondrial ATP9 gene to the nucleus reveals forces underlying mitochondrial genome evolution. PLoS genetics, 8, e1002876.

Blaby-Haas, C. E. and Merchant, S. S. (2019) Comparative and Functional Algal Genomics. Annual review of plant biology, 70, 605-638.

Blatch, G. L. and Lassle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. BioEssays : news and reviews in molecular, cellular and developmental biology, 21, 932-939.

Blowers, A. D., Ellmore, G. S., Klein, U. and Bogorad, L. (1990) Transcriptional analysis of endogenous and foreign genes in chloroplast transformants of Chlamydomonas. The Plant cell, 2, 1059-1070.

Bock, R. (2017) Witnessing Genome Evolution: Experimental Reconstruction of Endosymbiotic and Horizontal Gene Transfer. Annual review of genetics, 51, 1-22.

Boehm, E., Zaganelli, S., Maundrell, K., Jourdain, A. A., Thore, S. and Martinou, J. C. (2017) FASTKD1 and FASTKD4 have opposite effects on expression of specific mitochondrial RNAs, depending upon their endonuclease-like RAP domain. Nucleic acids research, 45, 6135-6146.

Boudreau, E., Nickelsen, J., Lemaire, S. D., Ossenbuhl, F. and Rochaix, J. D. (2000) The Nac2 gene of Chlamydomonas encodes a chloroplast TPR-like protein involved in psbD mRNA stability. The EMBO journal, 19, 3366-3376.

Bouhedda, F., Autour, A. and Ryckelynck, M. (2017) Light-Up RNA Aptamers and Their Cognate Fluorogens: From Their Development to Their Applications. International journal of molecular sciences, 19.

Boulouis, A., Drapier, D., Razafimanantsoa, H., Wostrikoff, K., Tourasse, N. J., Pascal, K., Girard-Bascou, J., Vallon, O., Wollman, F. A. and Choquet, Y. (2015) Spontaneous dominant mutations in chlamydomonas highlight ongoing evolution by gene diversification. The Plant cell, 27, 984-1001.

Boulouis, A., Raynaud, C., Bujaldon, S., Aznar, A., Wollman, F. A. and Choquet, Y. (2011) The nucleus-encoded trans-acting factor MCA1 plays a critical role in the regulation of cytochrome f synthesis in Chlamydomonas chloroplasts. The Plant cell, 23, 333-349.

Boussardon, C., Avon, A., Kindgren, P., Bond, C. S., Challenor, M., Lurin, C. and Small, I. (2014) The cytidine deaminase signature HxE(x)n CxxC of DYW1 binds zinc and is necessary for RNA editing of ndhD-1. The New phytologist, 203, 1090-1095.

Boussardon, C., Salone, V., Avon, A., Berthome, R., Hammani, K., Okuda, K., Shikanai, T., Small, I. and Lurin, C. (2012) Two interacting proteins are necessary for the editing of the NdhD-1 site in Arabidopsis plastids. The Plant cell, 24, 3684-3694.

Boynton, J. E., Gillham, N. W., Harris, E. H., Hosler, J. P., Johnson, A. M., Jones, A. R., Randolph-Anderson, B. L., Robertson, D., Klein, T. M., Shark, K. B. and et al. (1988) Chloroplast transformation in Chlamydomonas with high velocity microprojectiles. Science, 240, 1534-1538.

Bradbeer, J. W., Atkinson, Y. E., BÖRner, T. and Hagemann, R. (1979) Cytoplasmic synthesis of plastid polypeptides may be controlled by plastid-synthesised RNA. Nature, 279, 816-817.

Brinkman, F. S., Blanchard, J. L., Cherkasov, A., Av-Gay, Y., Brunham, R. C., Fernandez, R. C., Finlay, B. B., Otto, S. P., Ouellette, B. F., Keeling, P. J., Rose, A. M., Hancock, R. E., Jones, S. J. and Greberg, H. (2002) Evidence that plant-like genes in Chlamydia species reflect an ancestral relationship between Chlamydiaceae, cyanobacteria, and the chloroplast. Genome research, 12, 1159-1167.

Brzezowski, P., Wilson, K. E. and Gray, G. R. (2012) The PSBP2 protein of Chlamydomonas reinhardtii is required for singlet oxygen-dependent signaling. Planta, 236, 1289-1303.

Burger, G., Saint-Louis, D., Gray, M. W. and Lang, B. F. (1999) Complete sequence of the mitochondrial DNA of the red alga Porphyra purpurea. Cyanobacterial introns and shared ancestry of red and green algae. The Plant cell, 11, 1675-1694.

Calder, K. M. and McEwen, J. E. (1991) Deletion of the COX7 gene in Saccharomyces cerevisiae reveals a role for cytochrome c oxidase subunit VII in assembly of remaining subunits. Molecular microbiology, 5, 1769-1777.

Cavaiuolo, M., Kuras, R., Wollman, F. A., Choquet, Y. and Vallon, O. (2017) Small RNA profiling in Chlamydomonas: insights into chloroplast RNA metabolism. Nucleic acids research, 45, 10783-10799.

Cenci, U., Bhattacharya, D., Weber, A. P. M., Colleoni, C., Subtil, A. and Ball, S. G. (2017) Biotic Host-Pathogen Interactions As Major Drivers of Plastid Endosymbiosis. Trends in plant science, 22, 316-328.

Chen, L. J. and Li, H. M. (1998) A mutant deficient in the plastid lipid DGD is defective in protein import into chloroplasts. The Plant journal : for cell and molecular biology, 16, 33-39.

Choquet, Y., Goldschmidt-Clermont, M., Girard-Bascou, J., Kück, U., Bennoun, P. and Rochaix, J.-D. (1988) Mutant phenotypes support a trans-splicing mechanism for the expression of the tripartite psaA gene in the C. reinhardtii chloroplast. Cell, 52, 903-913.

Choquet, Y., Stern, D. B., Wostrikoff, K., Kuras, R., Girard-Bascou, J. and Wollman, F. A. (1998) Translation of cytochrome f is autoregulated through the 5' untranslated region of petA mRNA in Chlamydomonas chloroplasts. Proceedings of the National Academy of Sciences of the United States of America, 95, 4380-4385.

Choquet, Y. and Wollman, F. A. (2002) Translational regulations as specific traits of chloroplast gene expression. FEBS letters, 529, 39-42.

Choquet, Y. and Wollman, F. A. (2009) Chapter 29 - The CES Process. In The Chlamydomonas Sourcebook (Second Edition) (Eds, H., H. E., B., S. D. and B., W. G.) Academic Press, pp. 1027-1063.

Choquet, Y., Wostrikoff, K., Rimbault, B., Zito, F., Girard-Bascou, J., Drapier, D. and Wollman, F. A. (2001) Assembly-controlled regulation of chloroplast gene translation. Biochemical Society transactions, 29, 421-426.

Chua, N. H., Blobel, G., Siekevitz, P. and Palade, G. E. (1973) Attachment of chloroplast polysomes to thylakoid membranes in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America, 70, 1554-1558.

Chua, N. H., Blobel, G., Siekevitz, P. and Palade, G. E. (1976) Periodic variations in the ratio of free to thylakoid-bound chloroplast ribosomes during the cell cycle of Chlamydomonas reinhardtii. The Journal of cell biology, 71, 497-514.

Cline, S. G., Laughbaum, I. A. and Hamel, P. P. (2017) CCS2, an Octatricopeptide-Repeat Protein, Is Required for Plastid Cytochrome c Assembly in the Green Alga Chlamydomonas reinhardtii. Frontiers in plant science, 8, 1306.

Colcombet, J., Lopez-Obando, M., Heurtevin, L., Bernard, C., Martin, K., Berthome, R. and Lurin, C. (2013) Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. RNA biology, 10, 1557-1575.

Coquille, S., Filipovska, A., Chia, T., Rajappa, L., Lingford, J. P., Razif, M. F., Thore, S. and Rackham, O. (2014) An artificial PPR scaffold for programmable RNA recognition. Nature communications, 5, 5729.

Corliss, J. B. (1990) Hot springs and the origin of Life. Nature, 347.

Couradeau, E., Benzerara, K., Gerard, E., Moreira, D., Bernard, S., Brown, G. E., Jr. and Lopez-Garcia, P. (2012) An early-branching microbialite cyanobacterium forms intracellular carbonates. Science, 336, 459-462.

Cozens, A. L. and Walker, J. E. (1987) The organization and sequence of the genes for ATP synthase subunits in the cyanobacterium Synechococcus 6301. Support for an endosymbiotic origin of chloroplasts. Journal of molecular biology, 194, 359-383.

Daga, A., Micol, V., Hess, D., Aebersold, R. and Attardi, G. (1993) Molecular characterization of the transcription termination factor from human mitochondria. The Journal of biological chemistry, 268, 8123-8130.

Dagan, T., Roettger, M., Stucken, K., Landan, G., Koch, R., Major, P., Gould, S. B., Goremykin, V. V., Rippka, R., Tandeau de Marsac, N., Gugger, M., Lockhart, P. J., Allen, J. F., Brune, I., Maus, I., Puhler, A. and Martin, W. F. (2013) Genomes of Stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic

photosynthesis from prokaryotes to plastids. Genome biology and evolution, 5, 31-44.
Dahan, J. and Mireau, H. (2013) The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR genes. RNA biology, 10, 1469-1476.
Day, A., Schirmer-Rahire, M., Kuchka, M. R., Mayfield, S. P. and Rochaix, J. D. (1988) A transposon with an unusual arrangement of long terminal repeats in the green alga Chlamydomonas reinhardtii. The EMBO journal, 7, 1917-1927.
De Clerck, O., Bogaert, K. A. and Leliaert, F. (2012) Diversity and Evolution of Algae. 64, 55-86.
De Mia, M., Lemaire, S. D., Choquet, Y. and Wollman, F. A. (2019) Nitric Oxide Remodels the Photosynthetic Apparatus upon S-Starvation in Chlamydomonas reinhardtii. Plant physiology, 179, 718-731.
de Vries, J. and Archibald, J. M. (2017) Endosymbiosis: Did Plastids Evolve from a Freshwater Cyanobacterium? Current biology : CB, 27, R103-R105.
Derelle, E., Ferraz, C., Escande, M. L., Eychenie, S., Cooke, R., Piganeau, G., Desdevises, Y., Bellec, L., Moreau, H. and Grimsley, N. (2008) Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga Ostreococcus tauri. PloS one, 3, e2250.
Deusch, O., Landan, G., Roettger, M., Gruenheit, N., Kowallik, K. V., Allen, J. F., Martin, W. and Dagan, T. (2008) Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. Molecular biology and evolution, 25, 748-761.
Dogra, V., Rochaix, J. D. and Kim, C. (2018) Singlet oxygen-triggered chloroplast-to-nucleus retrograde signalling pathways: An emerging perspective. Plant, cell & environment, 41, 1727-1738.
Douchi, D., Qu, Y., Longoni, P., Legendre-Lefebvre, L., Johnson, X., Schmitz-Linneweber, C. and Goldschmidt-Clermont, M. (2016) A Nucleus-Encoded Chloroplast Phosphoprotein Governs Expression of the Photosystem I Subunit PsaC in Chlamydomonas reinhardtii. The Plant cell, 28, 1182-1199.
Drager, R. G., Girard-bascou, J., Choquet, Y., Kindle, K. L. and Stern, D. B. (1998) In vivo evidence for 5'→3' exoribonuclease degradation of an unstable chloroplast mRNA. The Plant Journal, 13, 85-96.
Drager, R. G., Zeidler, M., Simpson, C. L. and Stern, D. B. (1996) A chloroplast transcript lacking the 3' inverted repeat is degraded by 3'-->5' exoribonuclease activity. RNA (New York, N.Y.), 2, 652-663.
Drapier, D., Girard-Bascou, J. and Wollman, F. A. (1992) Evidence for Nuclear Control of the Expression of the atpA and atpB Chloroplast Genes in Chlamydomonas. The Plant cell, 4, 283-295.
Drapier, D., Girard-Bascou, J. , Stern, D. B. , Wollman, F.A. (2002) A dominant nuclear mutation in Chlamydomonas identifies a factor controlling chloroplast mRNA stability by acting on the coding region of the atpA transcript. The Plant Journal, 687-697.
Drapier, D., Rimbault, B., Vallon, O., Wollman, F. A. and Choquet, Y. (2007) Intertwined translational regulations set uneven stoichiometry of chloroplast ATP synthase subunits. The EMBO journal, 26, 3581-3591.
Drapier, D., Suzuki, H. , Levy, H. , Rimbault, B. , Kindle, K. L. , David B. Stern, D. B. , and and Wollman, F. A. (1998) The Chloroplast atpA Gene Cluster in Chlamydomonas reinhardtii. Plant physiology, 629–641.

Duanmu, D., Casero, D., Dent, R. M., Gallaher, S., Yang, W., Rockwell, N. C., Martin, S. S., Pellegrini, M., Niyogi, K. K., Merchant, S. S., Grossman, A. R. and Lagarias, J. C. (2013) Retrograde bilin signaling enables Chlamydomonas greening and phototrophic survival. Proceedings of the National Academy of Sciences of the United States of America, 110, 3621-3626.

Dudkina, N. V., Heinemeyer, J., Keegstra, W., Boekema, E. J. and Braun, H. P. (2005) Structure of dimeric ATP synthase from mitochondria: an angular association of monomers induces the strong curvature of the inner membrane. FEBS letters, 579, 5769-5772.

Durnford, D. G., Deane, J. A., Tan, S., McFadden, G. I., Gantt, E. and Green, B. R. (1999) A phylogenetic assessment of the eukaryotic light-harvesting antenna proteins, with implications for plastid evolution. Journal of molecular evolution, 48, 59-68.

Dyall, S. D., Brown, M. T. and Johnson, P. J. (2004) Ancient invasions: from endosymbionts to organelles. Science, 304, 253-257.

Eberhard, S., Drapier, D. and Wollman, F.-A. (2002) Searching limiting steps in the expression of chloroplast encoded proteins: relations between gene copy number, transcription, transcript abundance and translation rate in the chloroplast of Chlamydomonas reinhardtii. The Plant Journal, 31, 149-160.

Eberhard, S., Loiselay, C., Drapier, D., Bujaldon, S., Girard-Bascou, J., Kuras, R., Choquet, Y. and Wollman, F. A. (2011) Dual functions of the nucleus-encoded factor TDA1 in trapping and translation activation of atpA transcripts in Chlamydomonas reinhardtii chloroplasts. The Plant journal : for cell and molecular biology, 67, 1055-1066.

Eichler-Stahlberg, A., Weisheit, W., Ruecker, O. and Heitzer, M. (2009) Strategies to facilitate transgene expression in Chlamydomonas reinhardtii. Planta, 229, 873-883.

Elston, T., Wang, H. Y. and Oster, G. (1998) Energy transduction in ATP synthase. Nature, 391, 510-513.

Emanuelsson, O., Nielsen, H. and von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein science : a publication of the Protein Society, 8, 978-984.

Engel, B. D., Schaffer, M., Kuhn Cuellar, L., Villa, E., Plitzko, J. M. and Baumeister, W. (2015) Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography. eLife, 4.

Fan, W. H., Woelfle, M. A. and Mosig, G. (1995) Two copies of a DNA element, 'Wendy', in the chloroplast chromosome of Chlamydomonas reinhardtii between rearranged gene clusters. Plant molecular biology, 29, 63-80.

Filipovska, A. and Rackham, O. (2013) Pentatricopeptide repeats: modular blocks for building RNA-binding proteins. RNA biology, 10, 1426-1432.

Fischer, N., Stampacchia, O., Redding, K. and Rochaix, J. D. (1996) Selectable marker recycling in the chloroplast. Mol Gen Genet, 251, 373-380.

Fischer, W. W., Hemp, J. and Johnson, J. E. (2016) Evolution of Oxygenic Photosynthesis. Annual Review of Earth and Planetary Sciences, 44, 647-683.

Ford Doolittle, W. (1998) You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends in Genetics, 14, 307-311.

Fu, H. Y., Picot, D., Choquet, Y., Longatte, G., Sayegh, A., Delacotte, J., Guille-Collignon, M., Lemaitre, F., Rappaport, F. and Wollman, F. A. (2017) Redesigning the QA

binding site of Photosystem II allows reduction of exogenous quinones. Nature communications, 8, 15274.
Fuchs, G. (2011) Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? Annual review of microbiology, 65, 631-658.
Fujii, S., Bond, C. S. and Small, I. D. (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proceedings of the National Academy of Sciences of the United States of America, 108, 1723-1728.
Fujii, S., Suzuki, T., Giege, P., Higashiyama, T., Koizuka, N. and Shikanai, T. (2016) The Restorer-of-fertility-like 2 pentatricopeptide repeat protein and RNase P are required for the processing of mitochondrial orf291 RNA in Arabidopsis. The Plant journal : for cell and molecular biology, 86, 504-513.
Germain, A., Hotto, A. M., Barkan, A. and Stern, D. B. (2013) RNA processing and decay in plastids. Wiley interdisciplinary reviews. RNA, 4, 295-316.
Gibor, A. and Izawa, M. (1963) The DNA Content of the Chloroplasts of Acetabularia. Proceedings of the National Academy of Sciences of the United States of America, 50, 1164-1169.
Gilson, P. R., Maier, U.-G. and McFadden, G. I. (1997) Size isn't everything: lessons in genetic miniaturisation from nucleomorphs. Current opinion in genetics & development, 7, 800-806.
Giovannoni, S. J., Turner, S., Olsen, G. J., Barns, S., Lane, D. J. and Pace, N. R. (1988) Evolutionary relationships among cyanobacteria and green chloroplasts. Journal of bacteriology, 170, 3584-3592.
Giraud, M.-F., Paumard, P., Soubannier, V., Vaillier, J., Arselin, G., Salin, B., Schaeffer, J., Brèthes, D., di Rago, J.-P. and Velours, J. (2002) Is there a relationship between the supramolecular organization of the mitochondrial ATP synthase and the formation of cristae? Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1555, 174-180.
Goldschmidt-Clermont, M. (1991) Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker of site-directed transformation of chlamydomonas. Nucleic acids research, 19, 4083-4089.
Gray, M. W., Sankoff, D. and Cedergren, R. J. (1984) On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. Nucleic acids research, 12, 5837-5852.
Green, B. R. (2011) Chloroplast genomes of photosynthetic eukaryotes. The Plant journal : for cell and molecular biology, 66, 34-44.
Guillaumot, D., Lopez-Obando, M., Baudry, K., Avon, A., Rigaill, G., Falcon de Longevialle, A., Broche, B., Takenaka, M., Berthome, R., De Jaeger, G., Delannoy, E. and Lurin, C. (2017) Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. Proceedings of the National Academy of Sciences of the United States of America, 114, 8877-8882.
Gully, B. S., Cowieson, N., Stanley, W. A., Shearston, K., Small, I. D., Barkan, A. and Bond, C. S. (2015) The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atpH RNA. Nucleic acids research, 43, 1918-1926.
Guzman-Zapata, D., Dominguez-Anaya, Y., Macedo-Osorio, K. S., Tovar-Aguilar, A., Castrejon-Flores, J. L., Duran-Figueroa, N. V. and Badillo-Corona, J. A. (2017) mRNA imaging in the chloroplast of Chlamydomonas reinhardtii using the light-up aptamer Spinach. Journal of biotechnology, 251, 186-188.

Hahn, A., Vonck, J., Mills, D. J., Meier, T. and Kuhlbrandt, W. (2018) Structure, mechanism, and regulation of the chloroplast ATP synthase. Science, 360.

Hall, W. T. and Claus, G. (1963) Ultrastructural Studies on the Blue-Green Algal Symbiont in Cyanophora Paradoxa Korschikoff. The Journal of cell biology, 19, 551-563.

Hammani, K. and Barkan, A. (2014) An mTERF domain protein functions in group II intron splicing in maize chloroplasts. Nucleic acids research, 42, 5033-5042.

Hammani, K., Cook, W. B. and Barkan, A. (2012) RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. Proceedings of the National Academy of Sciences of the United States of America, 109, 5651-5656.

Han, K. Y., Leslie, B. J., Fei, J., Zhang, J. and Ha, T. (2013) Understanding the photophysics of the spinach-DFHBI RNA aptamer-fluorogen complex to improve live-cell RNA imaging. Journal of the American Chemical Society, 135, 19033-19038.

Harris, E. H. (1989) The Chlamydomonas Source Book: A Comprehensive Guide to Biology and Laboratory Use., Academic Press, San Diego.

Harris, E. H. (2001) Chlamydomonas as a Model Organism. Annual review of plant physiology and plant molecular biology, 52, 363-406.

Hauler, A., Jonietz, C., Stoll, B., Stoll, K., Braun, H. P. and Binder, S. (2013) RNA Processing Factor 5 is required for efficient 5' cleavage at a processing site conserved in RNAs of three different mitochondrial genes in Arabidopsis thaliana. The Plant journal : for cell and molecular biology, 74, 593-604.

Hazkani-Covo, E., Zeller, R. M. and Martin, W. (2010) Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS genetics, 6, e1000834.

Hernandez-Garcia, C. M. and Finer, J. J. (2014) Identification and validation of promoters and cis-acting regulatory elements. Plant science : an international journal of experimental plant biology, 217-218, 109-119.

Hicks, A., Drager, R. G., Higgs, D. C. and Stern, D. B. (2002) An mRNA 3' processing site targets downstream sequences for rapid degradation in Chlamydomonas chloroplasts. The Journal of biological chemistry, 277, 3325-3333.

Hillebrand, A., Matz, J. M., Almendinger, M., Muller, K., Matuschewski, K. and Schmitz-Linneweber, C. (2018) Identification of clustered organellar short (cos) RNAs and of a conserved family of organellar RNA-binding proteins, the heptatricopeptide repeat proteins, in the malaria parasite. Nucleic acids research, 46, 10417-10431.

Hohmann-Marriott, M. F. and Blankenship, R. E. (2011) Evolution of photosynthesis. Annual review of plant biology, 62, 515-548.

Holzl, G. and Dormann, P. (2007) Structure and function of glycoglycerolipids in plants and bacteria. Progress in lipid research, 46, 225-243.

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J. and Nakai, K. (2007) WoLF PSORT: protein localization predictor. Nucleic acids research, 35, W585-587.

Hosler, J. P., Wurtz, E. A., Harris, E. H., Gillham, N. W. and Boynton, J. E. (1989) Relationship between Gene Dosage and Gene Expression in the Chloroplast of Chlamydomonas reinhardtii. Plant physiology, 91, 648-655.

Houille-Vernes, L., Rappaport, F., Wollman, F. A., Alric, J. and Johnson, X. (2011) Plastid terminal oxidase 2 (PTOX2) is the major oxidase involved in chlororespiration in

Chlamydomonas. Proceedings of the National Academy of Sciences of the United States of America, 108, 20820-20825.
Ikeda, T. M. and Gray, M. W. (1999) Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria. Molecular and cellular biology, 19, 8113-8122.
Ilgu, M., Ray, J., Bendickson, L., Wang, T., Geraskin, I. M., Kraus, G. A. and Nilsen-Hamilton, M. (2016) Light-up and FRET aptamer reporters; evaluating their applications for imaging transcription in eukaryotic cells. Methods, 98, 26-33.
Im, C. S., Eberhard, S., Huang, K., Beck, C. F. and Grossman, A. R. (2006) Phototropin involvement in the expression of genes encoding chlorophyll and carotenoid biosynthesis enzymes and LHC apoproteins in Chlamydomonas reinhardtii. The Plant journal : for cell and molecular biology, 48, 1-16.
Itakura, A. K., Chan, K. X., Atkinson, N., Pallesen, L., Wang, L., Reeves, G., Patena, W., Caspari, O., Roth, R., Goodenough, U., McCormick, A. J., Griffiths, H. and Jonikas, M. C. (2019) A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America, 116, 18445-18454.
Jackson, C., Clayden, S. and Reyes-Prieto, A. (2015) The Glaucophyta: the blue-green plants in a nutshell. Acta Societatis Botanicorum Poloniae, 84, 149-165.
Jacobs, J., Marx, C., Kock, V., Reifschneider, O., Franzel, B., Krisp, C., Wolters, D. and Kuck, U. (2013) Identification of a chloroplast ribonucleoprotein complex containing trans-splicing factors, intron RNA, and novel components. Molecular & cellular proteomics : MCP, 12, 1912-1925.
Jarvis, P. and Soll, J. (2001) Toc, Tic, and chloroplast protein import. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 1541, 64-79.
Johnson, X., Wostrikoff, K., Finazzi, G., Kuras, R., Schwarz, C., Bujaldon, S., Nickelsen, J., Stern, D. B., Wollman, F. A. and Vallon, O. (2010) MRL1, a conserved Pentatricopeptide repeat protein, is required for stabilization of rbcL mRNA in Chlamydomonas and Arabidopsis. The Plant cell, 22, 234-248.
Junesch, U. and Gräber, P. (1987) Influence of the redox state and the activation of the chloroplast ATP synthase on proton-transport-coupled ATP synthesis/hydrolysis. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 893, 275-288.
Junge, W., Lill, H. and Engelbrecht, S. (1997) ATP synthase: an electrochemical ransducer with rotatory mechanics. Trends in Biochemical Sciences, 22, 420-423.
Kajava, A. V. (1998) Structural diversity of leucine-rich repeat proteins. Journal of molecular biology, 277, 519-527.
Kato, K., Ishikura, K., Kasai, S. and Shinmyo, A. (2006) Efficient translation destabilizes transcripts in chloroplasts of Chlamydomonas reinhardtii. Journal of bioscience and bioengineering, 101, 471-477.
Kato, Y. and Sakamoto, W. (2018) FtsH Protease in the Thylakoid Membrane: Physiological Functions and the Regulation of Protease Activity. Frontiers in plant science, 9, 855.
Keeling, P. J. (2010) The endosymbiotic origin, diversification and fate of plastids. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 365, 729-748.

Keeling, P. J. and Palmer, J. D. (2008) Horizontal gene transfer in eukaryotic evolution. Nature reviews. Genetics, 9, 605-618.

Kelley, D. S., Karson, J. A., Blackman, D. K., Fruh-Green, G. L., Butterfield, D. A., Lilley, M. D., Olson, E. J., Schrenk, M. O., Roe, K. K., Lebon, G. T., Rivizzigno, P. and Party, A. T. S. (2001) An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30 degrees N. Nature, 412, 145-149.

Khan, H., Parks, N., Kozera, C., Curtis, B. A., Parsons, B. J., Bowman, S. and Archibald, J. M. (2007) Plastid genome sequence of the cryptophyte alga Rhodomonas salina CCMP1319: lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. Molecular biology and evolution, 24, 1832-1842.

Khrebtukova, I. and Spreitzer, R. J. (1996) Elimination of the Chlamydomonas gene family that encodes the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Proceedings of the National Academy of Sciences of the United States of America, 93, 13689-13693.

Kim, M., Abdi, K., Lee, G., Rabbi, M., Lee, W., Yang, M., Schofield, C. J., Bennett, V. and Marszalek, P. E. (2010) Fast and forceful refolding of stretched alpha-helical solenoid proteins. Biophysical journal, 98, 3086-3092.

Klein, U., De Camp, J. D. and Bogorad, L. (1992) Two types of chloroplast gene promoters in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America, 89, 3453-3457.

Kleinknecht, L., Wang, F., Stube, R., Philippar, K., Nickelsen, J. and Bohne, A. V. (2014) RAP, the sole octotricopeptide repeat protein in Arabidopsis, is required for chloroplast 16S rRNA maturation. The Plant cell, 26, 777-787.

Kobayashi, K., Kawabata, M., Hisano, K., Kazama, T., Matsuoka, K., Sugita, M. and Nakamura, T. (2012) Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. Nucleic acids research, 40, 2712-2723.

Kobayashi, T., Yagi, Y. and Nakamura, T. (2019) Comprehensive Prediction of Target RNA Editing Sites for PLS-Class PPR Proteins in Arabidopsis thaliana. Plant & cell physiology, 60, 862-874.

Kobe, B. and Kajava, A. V. (2000) When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. Trends in Biochemical Sciences, 25, 509-515.

Kotera, E., Tasaka, M. and Shikanai, T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature, 433.

Kovar, J. L., Zhang, J., Funke, R. P. and Weeks, D. P. (2002) Molecular analysis of the acetolactate synthase gene of Chlamydomonas reinhardtii and development of a genetically engineered gene as a dominant selectable marker for genetic transformation. The Plant journal, 29, 109-117.

Kuchka, M. R., Mayfield, S. P. and Rochaix, J. D. (1988) Nuclear mutations specifically affect the synthesis and/or degradation of the chloroplast-encoded D2 polypeptide of photosystem II in Chlamydomonas reinhardtii. The EMBO journal, 7, 319-324.

Kuhlbrandt, W. (2019) Structure and Mechanisms of F-Type ATP Synthases. Annual review of biochemistry, 88, 515-549.

Kuras, R. and Wollman, F. A. (1994) The assembly of cytochrome b6/f complexes: an approach using genetic transformation of the green alga Chlamydomonas reinhardtii. The EMBO journal, 13, 1019-1027.

Lane, N., Allen, J. F. and Martin, W. (2010) How did LUCA make a living? Chemiosmosis in the origin of life. BioEssays : news and reviews in molecular, cellular and developmental biology, 32, 271-280.

Larkum, A. W., Lockhart, P. J. and Howe, C. J. (2007) Shopping for plastids. Trends in plant science, 12, 189-195.

Le Hir, H., Nott, A. and Moore, M. J. (2003) How introns influence and enhance eukaryotic gene expression. Trends in Biochemical Sciences, 28, 215-220.

Ledford, H. K., Chin, B. L. and Niyogi, K. K. (2007) Acclimation to singlet oxygen stress in Chlamydomonas reinhardtii. Eukaryotic cell, 6, 919-930.

Lee, I. and Hong, W. (2004) RAP – a putative RNA-binding domain. Trends Biochem Sci, 29, 567-570.

Lefebvre-Legendre, L., Choquet, Y., Kuras, R., Loubery, S., Douchi, D. and Goldschmidt-Clermont, M. (2015) A nucleus-encoded chloroplast protein regulated by iron availability governs expression of the photosystem I subunit PsaA in Chlamydomonas reinhardtii. Plant physiology, 167, 1527-1540.

Lemaire, C. and Wollman, F. A. (1989) The Chloroplast ATP Synthase in Chlamydomonas reinhardtii. 2. BIOCHEMICAL-STUDIES ON ITS BIOGENESIS USING MUTANTS DEFECTIVE IN PHOTOPHOSPHORYLATION. JOURNAL OF BIOLOGICAL CHEMISTRY, 264, 10235-10242.

Levine, R. P. (1960) A screening technique for photosynthetic mutants in unicellular algae. Nature, 188, 339-340.

Li, X., Patena, W., Fauser, F., Jinkerson, R. E., Saroussi, S., Meyer, M. T., Ivanova, N., Robertson, J. M., Yue, R., Zhang, R., Vilarrasa-Blasi, J., Wittkopp, T. M., Ramundo, S., Blum, S. R., Goh, A., Laudon, M., Srikumar, T., Lefebvre, P. A., Grossman, A. R. and Jonikas, M. C. (2019) A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. Nature genetics, 51, 627-635.

Lipinski, K. A., Puchta, O., Surendranath, V., Kudla, M. and Golik, P. (2011) Revisiting the yeast PPR proteins--application of an Iterative Hidden Markov Model algorithm reveals new members of the rapidly evolving family. Molecular biology and evolution, 28, 2935-2948.

Lister, D. L., Bateman, J. M., Purton, S. and Howe, C. J. (2003) DNA transfer from chloroplast to nucleus is much rarer in Chlamydomonas than in tobacco. Gene, 316, 33-38.

Loiselay, C., Gumpel, N. J., Girard-Bascou, J., Watson, A. T., Purton, S., Wollman, F. A. and Choquet, Y. (2008) Molecular identification and function of cis- and trans-acting determinants for petA transcript stability in Chlamydomonas reinhardtii chloroplasts. Molecular and cellular biology, 28, 5529-5542.

Loizeau, K., Qu, Y., Depp, S., Fiechter, V., Ruwe, H., Lefebvre-Legendre, L., Schmitz-Linneweber, C. and Goldschmidt-Clermont, M. (2014) Small RNAs reveal two target sites of the RNA-maturation factor Mbb1 in the chloroplast of Chlamydomonas. Nucleic acids research, 42, 3286-3297.

Ludwig, K. A., Shen, C.-C., Kelley, D. S., Cheng, H. and Edwards, R. L. (2011) U–Th systematics and 230Th ages of carbonate chimneys at the Lost City Hydrothermal Field. Geochimica et Cosmochimica Acta, 75, 1869-1888.

Lumbreras, V., Stevens, D. R. and Purton, S. (1998) Efficient foreign gene expression in Chlamydomonas reinhardtii mediated by an endogenous intron. The Plant Journal, 14, 441-447.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M. L., Mireau, H., Peeters, N., Renou, J. P., Szurek, B., Taconnat, L. and Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. The Plant cell, 16, 2089-2103.

Lyons, T. W., Reinhard, C. T. and Planavsky, N. J. (2014) The rise of oxygen in Earth's early ocean and atmosphere. Nature, 506, 307-315.

Mackinder, L. C., Meyer, M. T., Mettler-Altmann, T., Chen, V. K., Mitchell, M. C., Caspari, O., Freeman Rosenzweig, E. S., Pallesen, L., Reeves, G., Itakura, A., Roth, R., Sommer, F., Geimer, S., Muhlhaus, T., Schroda, M., Goodenough, U., Stitt, M., Griffiths, H. and Jonikas, M. C. (2016) A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. Proceedings of the National Academy of Sciences of the United States of America, 113, 5958-5963.

Majeran, W., Olive, J., Drapier, D., Vallon, O. and Wollman, F. A. (2001) The light sensitivity of ATP synthase mutants of Chlamydomonas reinhardtii. Plant physiology, 126, 421-433.

Malnoe, A., Wang, F., Girard-Bascou, J., Wollman, F. A. and de Vitry, C. (2014) Thylakoid FtsH protease contributes to photosystem II and cytochrome b6f remodeling in Chlamydomonas reinhardtii under stress conditions. The Plant cell, 26, 373-390.

Markham, N. R. and Zuker, M. (2008) UNAFold: software for nucleic acid folding and hybridization. Methods in molecular biology, 453, 3-31.

Martin, W. and Herrmann, R. G. (1998) Gene transfer from organelles to the nucleus: how much, what happens, and Why? Plant physiology, 118, 9-17.

Martin, W. and Kowallik, K. (1999) Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren imPflanzenreiche'. European Journal of Phycology, 34, 287-295.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M. and Penny, D. (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proceedings of the National Academy of Sciences of the United States of America, 99, 12246-12251.

Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M. and Kowallik, K. V. (1998) Gene transfer to the nucleus and the evolution of chloroplasts. Nature, 393, 162-165.

Martin, W. F., Bryant, D. A. and Beatty, J. T. (2018) A physiological perspective on the origin and evolution of photosynthesis. FEMS microbiology reviews, 42, 205-231.

Marx, C., Wunsch, C. and Kuck, U. (2015) The Octatricopeptide Repeat Protein Raa8 Is Required for Chloroplast trans Splicing. Eukaryotic cell, 14, 998-1005.

Matsuo, M., Ito, Y., Yamauchi, R. and Obokata, J. (2005) The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. The Plant cell, 17, 665-675.

Maul, J. E., Lilly, J. W., Cui, L., dePamphilis, C. W., Miller, W., Harris, E. H. and Stern, D. B. (2002) The Chlamydomonas reinhardtii plastid chromosome: islands of genes in a sea of repeats. The Plant cell, 14, 2659-2679.

Mayfield, S. P., Bennoun, P. and Rochaix, J. D. (1987) Expression of the nuclear encoded OEE1 protein is required for oxygen evolution and stability of

photosystem II particles in Chlamydomonas reinhardtii. The EMBO journal, 6, 313-318.
McDermott, J. J., Civic, B. and Barkan, A. (2018) Effects of RNA structure and salt concentration on the affinity and kinetics of interactions between pentatricopeptide repeat proteins and their RNA ligands. PloS one, 13, e0209713.
McFadden, G. I. and van Dooren, G. G. (2004) Evolution: red algal genome affirms a common origin of all plastids. Current biology : CB, 14, R514-516.
Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N. and Schuster, G. (2003) HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. The Plant cell, 15, 1480-1495.
Merendino, L., Perron, K., Rahire, M., Howald, I., Rochaix, J. D. and Goldschmidt-Clermont, M. (2006) A novel multifunctional factor involved in trans-splicing of chloroplast introns in Chlamydomonas. Nucleic acids research, 34, 262-274.
Mereschkowski, C. (1905) Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. Biol. Centralbl., 25, 593-604.
Miller, M. J. and McMahon, D. (1974) Synthesis and maturation of chloroplast and cytoplasmic ribosomal RNA in Chlamydomonas reinhardi. Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis, 366, 35-44.
Minai, L., Wostrikoff, K., Wollman, F. A. and Choquet, Y. (2006) Chloroplast biogenesis of photosystem II cores involves a series of assembly-controlled steps that regulate translation. The Plant cell, 18, 159-175.
Miranda, R. G., McDermott, J. J. and Barkan, A. (2018) RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR-RNA interactions. Nucleic acids research, 46, 2613-2623.
Miranda, R. G. R., M.; Montgomery, M. P.; Gribbin, K. P.; Barkan, A. (2017) RNA-binding specificity landscape of the pentatricopeptide repeat protein PPR10. RNA (New York, N.Y.).
Monod, C., Goldschmidt-Clermont, M. and Rochaix, J. D. (1992) Accumulation of chloroplast psbB RNA requires a nuclear factor in Chlamydomonas reinhardtii. Mol Gen Genet, 231, 449-459.
Montgomery, M. G., Gahura, O., Leslie, A. G. W., Zikova, A. and Walker, J. E. (2018) ATP synthase from Trypanosoma brucei has an elaborated canonical F1-domain and conventional catalytic sites. Proceedings of the National Academy of Sciences of the United States of America, 115, 2102-2107.
Moran, N. A. (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proceedings of the National Academy of Sciences of the United States of America, 93, 2873-2878.
Moriyama, T., Terasawa, K., Fujiwara, M. and Sato, N. (2008) Purification and characterization of organellar DNA polymerases in the red alga Cyanidioschyzon merolae. The FEBS journal, 275, 2899-2918.
Munoz-Gomez, S. A., Mejia-Franco, F. G., Durnin, K., Colp, M., Grisdale, C. J., Archibald, J. M. and Slamovits, C. H. (2017) The New Red Algal Subphylum Proteorhodophytina Comprises the Largest and Most Divergent Plastid Genomes Known. Current biology : CB, 27, 1677-1684 e1674.
Murakami, S., Kuehnle, K. and Stern, D. B. (2005) A spontaneous tRNA suppressor of a mutation in the Chlamydomonas reinhardtii nuclear MCD1 gene required for stability of the chloroplast petD mRNA. Nucleic acids research, 33, 3372-3380.

Murphy, B. J., Klusch, N., Langer, J., Mills, D. J., Yildiz, Ö. and Kühlbrandt, W. (2019) Rotary substates of mitochondrial ATP synthase reveal the basis of flexible F1-Fo coupling. Science, 364, eaaw9128.

Nisbet, E. G. C., J. R.; Van Dover, C.L. (1995) Origins of photosynthesis. Nature, 373, 479-480.

Nishimura, K., Kato, Y. and Sakamoto, W. (2017) Essentials of Proteolytic Machineries in Chloroplasts. Molecular plant, 10, 4-19.

Nowack, E. C., Melkonian, M. and Glockner, G. (2008) Chromatophore genome sequence of Paulinella sheds light on acquisition of photosynthesis by eukaryotes. Current biology : CB, 18, 410-418.

Ochoa de Alda, J. A., Esteban, R., Diago, M. L. and Houmard, J. (2014) The plastid ancestor originated among one of the major cyanobacterial lineages. Nature communications, 5, 4937.

Ohad, I., Siekevitz, P. and Palade, G. E. (1967) Biogenesis of chloroplast membranes. I. Plastid dedifferentiation in a dark-grown algal mutant (Chlamydomonas reinhardi). The Journal of cell biology, 35, 521-552.

Okuda, K. and Shikanai, T. (2012) A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. Nucleic acids research, 40, 5052-5064.

Paige, J. S., Wu, K. Y. and Jaffrey, S. R. (2011) RNA mimics of green fluorescent protein. Science, 333, 642-646.

Paila, Y. D., Richardson, L. G. L. and Schnell, D. J. (2015) New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development. Journal of molecular biology, 427, 1038-1060.

Palmer, J. (1985) Comparative Organization of Chloroplast Genomes. Annual review of genetics, 19, 325-354.

Parfrey, L. W., Lahr, D. J., Knoll, A. H. and Katz, L. A. (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proceedings of the National Academy of Sciences of the United States of America, 108, 13624-13629.

Pfalz, J., Bayraktar, O. A., Prikryl, J. and Barkan, A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. The EMBO journal, 28, 2042-2052.

Pfanzagl, B., Zenker, A., Pittenauer, E., Allmaier, G., Martinez-Torrecuadrada, J., Schmid, E. R., De Pedro, M. A. and Loffelhardt, W. (1996) Primary structure of cyanelle peptidoglycan of Cyanophora paradoxa: a prokaryotic cell wall as part of an organelle envelope. Journal of bacteriology, 178, 332-339.

Piccioni, R. G., Bennoun, P. and Chua, N. H. (1981) A nuclear mutant of Chlamydomonas reinhardtii defective in photosynthetic photophosphorylation. Characterization of the algal coupling factor ATPase. Eur J Biochem, 117, 93-102.

Pinnaduwage, P. and Bruce, B. D. (1996) In vitro interaction between a chloroplast transit peptide and chloroplast outer envelope lipids is sequence-specific and lipid class-dependent. The Journal of biological chemistry, 271, 32907-32915.

Ponce-Toledo, R. I., Deschamps, P., Lopez-Garcia, P., Zivanovic, Y., Benzerara, K. and Moreira, D. (2017) An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. Current biology : CB, 27, 386-391.

Ponce-Toledo, R. I., Lopez-Garcia, P. and Moreira, D. (2019) Horizontal and endosymbiotic gene transfer in early plastid evolution. The New phytologist, 224, 618-624.

Popot, J. L. and de Vitry, C. (1990) On the microassembly of integral membrane proteins. Annu. Rev. Biophys. Biophys. Chem. , 19: 369-403.

Preker, P. J. and Keller, W. (1998) The HAT helix, a repetitive motif implicated in RNA processing. Trends in Biochemical Sciences, 23, 15-16.

Price, D. C., Chan, C. X., Yoon, H. S., Yang, E. C., Qiu, H., Weber, A. P., Schwacke, R., Gross, J., Blouin, N. A., Lane, C., Reyes-Prieto, A., Durnford, D. G., Neilson, J. A., Lang, B. F., Burger, G., Steiner, J. M., Loffelhardt, W., Meuser, J. E., Posewitz, M. C., Ball, S., Arias, M. C., Henrissat, B., Coutinho, P. M., Rensing, S. A., Symeonidi, A., Doddapaneni, H., Green, B. R., Rajah, V. D., Boore, J. and Bhattacharya, D. (2012) Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants. Science, 335, 843-847.

Prikryl, J., Rojas, M., Schuster, G. and Barkan, A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. Proceedings of the National Academy of Sciences of the United States of America, 108, 415-420.

Proskurowski, G., Lilley, M. D., Seewald, J. S., Fruh-Green, G. L., Olson, E. J., Lupton, J. E., Sylva, S. P. and Kelley, D. S. (2008) Abiogenic hydrocarbon production at lost city hydrothermal field. Science, 319, 604-607.

Qiu, H., Price, D. C., Yang, E. C., Yoon, H. S. and Bhattacharya, D. (2015) Evidence of ancient genome reduction in red algae (Rhodophyta). J Phycol, 51, 624-636.

Rahire, M., Laroche, F., Cerutti, L. and Rochaix, J. D. (2012) Identification of an OPR protein involved in the translation initiation of the PsaB subunit of photosystem I. The Plant journal : for cell and molecular biology, 72, 652-661.

Ramel, F., Birtic, S., Ginies, C., Soubigou-Taconnat, L., Triantaphylides, C. and Havaux, M. (2012) Carotenoid oxidation products are stress signals that mediate gene responses to singlet oxygen in plants. Proceedings of the National Academy of Sciences of the United States of America, 109, 5535-5540.

Raynaud, C., Loiselay, C., Wostrikoff, K., Kuras, R., Girard-Bascou, J., Wollman, F. A. and Choquet, Y. (2007) Evidence for regulatory function of nucleus-encoded factors on mRNA stabilization and translation in the chloroplast. Proceedings of the National Academy of Sciences of the United States of America, 104, 9093-9098.

Rea, G., Antonacci, A., Lambreva, M. D. and Mattoo, A. K. (2018) Features of cues and processes during chloroplast-mediated retrograde signaling in the alga Chlamydomonas. Plant science : an international journal of experimental plant biology, 272, 193-206.

Remacle, C., Cardol, P., Coosemans, N., Gaisne, M. and Bonnefoy, N. (2006) High-efficiency biolistic transformation of Chlamydomonas mitochondria can be used to insert mutations in complex I genes. Proceedings of the National Academy of Sciences of the United States of America, 103, 4771-4776.

Ris, H. and Plaut, W. (1962) Ultrastructure of DNA-containing areas in the chloroplast of Chlamydomonas. The Journal of cell biology, 13, 383-391.

Rivals, E., Bruyere, C., Toffano-Nioche, C. and Lecharny, A. (2006) Formation of the Arabidopsis pentatricopeptide repeat family. Plant physiology, 141, 825-839.

Rivier, C., Goldschmidt-Clermont, M. and Rochaix, J. D. (2001) Identification of an RNA-protein complex involved in chloroplast group II intron trans-splicing in Chlamydomonas reinhardtii. The EMBO journal, 20, 1765-1773.

Roberti, M., Polosa, P. L., Bruni, F., Manzari, C., Deceglie, S., Gadaleta, M. N. and Cantatore, P. (2009) The MTERF family proteins: mitochondrial transcription regulators and beyond. Biochim Biophys Acta, 1787, 303-311.

Rochaix, J. D. (1996) Post-transcriptional regulation of chloroplast gene expression in Chlamydomonas reinhardtii. Plant molecular biology, 32, 327-341.

Rochaix, J. D. and Ramundo, S. (2018) Chloroplast signaling and quality control. Essays in biochemistry, 62, 13-20.

Rodriguez-Concepcion, M., D'Andrea, L. and Pulido, P. (2019) Control of plastidial metabolism by the Clp protease complex. Journal of experimental botany, 70, 2049-2058.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S. C., Roure, B., Burger, G., Loffelhardt, W., Bohnert, H. J., Philippe, H. and Lang, B. F. (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Current biology : CB, 15, 1325-1330.

Rojas, M., Ruwe, H., Miranda, R. G., Zoschke, R., Hase, N., Schmitz-Linneweber, C. and Barkan, A. (2018) Unexpected functional versatility of the pentatricopeptide repeat proteins PGR3, PPR5 and PPR10. Nucleic acids research, 46, 10448-10459.

Rojas, M., Yu, Q., Williams-Carrier, R., Maliga, P. and Barkan, A. (2019) Engineered PPR proteins as inducible switches to activate the expression of chloroplast transgenes. Nature plants, 5, 505-511.

Romani, I., Manavski, N., Morosetti, A., Tadini, L., Maier, S., Kühn, K., Ruwe, H., Schmitz-Linneweber, C., Wanner, G., Leister, D. and Kleine, T. (2015) A Member of the Arabidopsis Mitochondrial Transcription Termination Factor Family Is Required for Maturation of Chloroplast Transfer RNAIle(GAU). Plant physiology, 169, 627-646.

Rott, R., Levy, H., Drager, R. G., Stern, D. B. and Schuster, G. (1998a) 3'-Processed mRNA is preferentially translated in Chlamydomonas reinhardtii chloroplasts. Molecular and cellular biology, 18, 4605-4611.

Rott, R., Liveanu, V., Drager, R. G., Stern, D. B. and Schuster, G. (1998b) The sequence and structure of the 3'-untranslated regions of chloroplast transcripts are important determinants of mRNA accumulation and stability. Plant molecular biology, 36, 307-314.

Ruwe, H., Gutmann, B., Schmitz-Linneweber, C., Small, I. and Kindgren, P. (2019) The E domain of CRR2 participates in sequence-specific recognition of RNA in plastids. The New phytologist, 222, 218-229.

Ruwe, H. and Schmitz-Linneweber, C. (2012) Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? Nucleic acids research, 40, 3106-3116.

Sabbert, D., Engelbrecht, S. and Junge, W. (1996) Intersubunit rotation in active F-ATPase. Nature, 381, 623-625.

Sagan, L. (1967) On the origin of mitosing cells. J Theor Biol, 14, 225-274.

Sahoo, S. K., Planavsky, N. J., Kendall, B., Wang, X., Shi, X., Scott, C., Anbar, A. D., Lyons, T. W. and Jiang, G. (2012) Ocean oxygenation in the wake of the Marinoan glaciation. Nature, 489, 546-549.

Salome, P. A. and Merchant, S. S. (2019) A Series of Fortunate Events: Introducing Chlamydomonas as a Reference Organism. The Plant cell, 31, 1682-1707.

Sanchez-Baracaldo, P., Raven, J. A., Pisani, D. and Knoll, A. H. (2017) Early photosynthetic eukaryotes inhabited low-salinity habitats. Proceedings of the

National Academy of Sciences of the United States of America, 114, E7737-E7745.
Sane, A. P., Stein, B. and Westhoff, P. (2005) The nuclear gene HCF107 encodes a membrane-associated R-TPR (RNA tetratricopeptide repeat)-containing protein involved in expression of the plastidial psbH gene in Arabidopsis. The Plant journal : for cell and molecular biology, 42, 720-730.
Schmidt, G. W. and Mishkind, M. L. (1983) Rapid degradation of unassembled ribulose 1,5-bisphosphate carboxylase small subunits in chloroplasts. Proceedings of the National Academy of Sciences of the United States of America, 80, 2632-2636.
Schmitz-Linneweber, C. and Small, I. (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends in plant science, 13, 663-670.
Schulz, S., Wilkes, M., Mills, D. J., Kuhlbrandt, W. and Meier, T. (2017) Molecular architecture of the N-type ATPase rotor ring from Burkholderia pseudomallei. EMBO reports, 18, 526-535.
Schuster, G., Lisitsky, I. and Klaff, P. (1999) Polyadenylation and degradation of mRNA in the chloroplast. Plant physiology, 120, 937-944.
Schuster, G. and Stern, D. (2009) Chapter 10 RNA Polyadenylation and Decay in Mitochondria and Chloroplasts. 85, 393-422.
Schwartz, R. and Dayhoff, M. (1978) Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. Science, 199, 395-403.
Schwarz, C., Elles, I., Kortmann, J., Piotrowski, M. and Nickelsen, J. (2007) Synthesis of the D2 protein of photosystem II in Chlamydomonas is controlled by a high molecular mass complex containing the RNA stabilization factor Nac2 and the translational activator RBP40. The Plant cell, 19, 3627-3639.
Sharon, I., Alperovitch, A., Rohwer, F., Haynes, M., Glaser, F., Atamna-Ismaeel, N., Pinter, R. Y., Partensky, F., Koonin, E. V., Wolf, Y. I., Nelson, N. and Beja, O. (2009) Photosystem I gene cassettes are present in marine virus genomes. Nature, 461, 258-262.
Shen, C., Zhang, D., Guan, Z., Liu, Y., Yang, Z., Yang, Y., Wang, X., Wang, Q., Zhang, Q., Fan, S., Zou, T. and Yin, P. (2016) Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. Nature communications, 7, 11285.
Shih, P. M., Hemp, J., Ward, L. M., Matzke, N. J. and Fischer, W. W. (2017) Crown group Oxyphotobacteria postdate the rise of oxygen. Geobiology, 15, 19-29.
Shin, S. E., Lim, J. M., Koh, H. G., Kim, E. K., Kang, N. K., Jeon, S., Kwon, S., Shin, W. S., Lee, B., Hwangbo, K., Kim, J., Ye, S. H., Yun, J. Y., Seo, H., Oh, H. M., Kim, K. J., Kim, J. S., Jeong, W. J., Chang, Y. K. and Jeong, B. R. (2016) CRISPR/Cas9-induced knockout and knock-in mutations in Chlamydomonas reinhardtii. Scientific reports, 6, 27810.
Sijben-Muller, G., Hallick, R. B., Alt, J., Westhoff, P. and Herrmann, R. G. (1986) Spinach plastid genes coding for initiation factor IF-1, ribosomal protein S11 and RNA polymerase alpha-subunit. Nucleic acids research, 14, 1029-1044.
Sleep, N. H., Bird, D. K. and Pope, E. C. (2011) Serpentinite and the dawn of life. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 366, 2857-2869.
Small, I., Peeters, N., Legeai, F. and Lurin, C. (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics, 4, 1581-1590.

Small, I. D. and Peeters, N. (2000) The PPR motif – a TPR-related motif prevalent in plant organellar proteins. Trends Biochem. Sci., 25, 46-47.

Smeekens, S., Bauerle, C., Hageman, J., Keegstra, K. and Weisbeek, P. (1986) The role of the transit peptide in the routing of precursors toward different chloroplast compartments. Cell, 46, 365-375.

Smith, D. R. (2015) Mutation rates in plastid genomes: they are lower than you might think. Genome biology and evolution, 7, 1227-1234.

Smith, H. C., Gott, J. M. and Hanson, M. R. (1997) A guide to RNA editing. RNA (New York, N.Y.), 3, 1105-1123.

Soo, R. M. H., J.; Parks, D. H.; and Fischer, W. W. H., P. (2017) On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. Science, 355, 1436–1440.

Sousa, F. L., Preiner, M. and Martin, W. F. (2018) Native metals, electron bifurcation, and CO2 reduction in early biochemical evolution. Current opinion in microbiology, 43, 77-83.

Sousa, F. L., Thiergart, T., Landan, G., Nelson-Sathi, S., Pereira, I. A., Allen, J. F., Lane, N. and Martin, W. F. (2013) Early bioenergetic evolution. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 368, 20130088.

Spahr, H., Chia, T., Lingford, J. P., Siira, S. J., Cohen, S. B., Filipovska, A. and Rackham, O. (2018) Modular ssDNA binding and inhibition of telomerase activity by designer PPR proteins. Nature communications, 9, 2212.

Stampacchia, O., Girard-Bascou, J., Zanasco, J. L., Zerges, W., Bennoun, P. and Rochaix, J. D. (1997) A nuclear-encoded function essential for translation of the chloroplast psaB mRNA in chlamydomonas. The Plant cell, 9, 773-782.

Stern, D. B. and Kindle, K. L. (1993) 3'end maturation of the Chlamydomonas reinhardtii chloroplast atpB mRNA is a two-step process. Molecular and cellular biology, 13, 2277-2285.

Stern, D. B., Radwanski, E. R. and Kindle, K. L. (1991) A 3' stem/loop structure of the Chlamydomonas chloroplast atpB gene regulates mRNA accumulation in vivo. The Plant cell, 3, 285-297.

Strack, R. L., Disney, M. D. and Jaffrey, S. R. (2013) A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. Nature methods, 10, 1219-1224.

Sun, Y. and Zerges, W. (2015) Translational regulation in chloroplasts for development and homeostasis. Biochim Biophys Acta, 1847, 809-820.

Surzycki, S. J. and Shellenbarger, D. L. (1976) Purification and characterization of a putative sigma factor from Chalamydomonas reinhardi. Proceedings of the National Academy of Sciences of the United States of America, 73, 3961-3965.

Tardif, M., Atteia, A., Specht, M., Cogne, G., Rolland, N., Brugiere, S., Hippler, M., Ferro, M., Bruley, C., Peltier, G., Vallon, O. and Cournac, L. (2012) PredAlgo: a new subcellular localization prediction tool dedicated to green algae. Molecular biology and evolution, 29, 3625-3639.

Tedetti, M. and Sempéré, R. (2006) Penetration of ultraviolet radiation in the marine environment. A review. Photochemistry and photobiology, 82, 389-397.

Tice, M. M. L., D. R. (2004) Photosynthetic microbial mats in the 3,416-Myr-old ocean. Nature, 431, 549-552.

Timmis, J. N., Ayliffe, M. A., Huang, C. Y. and Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nature reviews. Genetics, 5, 123-135.

Timmis, J. N. and Scott, N. S. (1983) Sequence homology between spinach nuclear and chloroplast genomes. Nature, 305, 65-67.

Tourasse, N. J., Choquet, Y. and Vallon, O. (2013) PPR proteins of green algae. RNA biology, 10, 1526-1542.

Trachman, R. J., 3rd, Truong, L. and Ferre-D'Amare, A. R. (2017) Structural Principles of Fluorescent RNA Aptamers. Trends in pharmacological sciences, 38, 928-939.

Trosch, R., Barahimipour, R., Gao, Y., Badillo-Corona, J. A., Gotsmann, V. L., Zimmer, D., Muhlhaus, T., Zoschke, R. and Willmund, F. (2018) Commonalities and differences of chloroplast translation in a green alga and land plants. Nature plants, 4, 564-575.

Uniacke, J. and Zerges, W. (2009) Chloroplast protein targeting involves localized translation in Chlamydomonas. Proceedings of the National Academy of Sciences of the United States of America, 106, 1439-1444.

Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M. and Giegé, P. (2008) PPR336 is Associated with Polysomes in Plant Mitochondria. Journal of molecular biology, 375, 626-636.

Vaistij, F. E., Boudreau, E., Lemaire, S. D., Goldschmidt-Clermont, M. and Rochaix, J. D. (2000) Characterization of Mbb1, a nucleus-encoded tetratricopeptide-like repeat protein required for expression of the chloroplast psbB/psbT/psbH gene cluster in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences of the United States of America, 97, 14813-14818.

Van Dover, C. L., Reynolds, G. T., Chave, A. D. and Tyson, J. A. (1996) Light at deep-sea hydrothermal vents. Geophysical Research Letters, 23, 2049-2052.

Vazquez-Acevedo, M., Cardol, P., Cano-Estrada, A., Lapaille, M., Remacle, C. and Gonzalez-Halphen, D. (2006) The mitochondrial ATP synthase of chlorophycean algae contains eight subunits of unknown origin involved in the formation of an atypical stator-stalk and in the dimerization of the complex. Journal of bioenergetics and biomembranes, 38, 271-282.

Viola, S., Cavaiuolo, M., Drapier, D., Eberhard, S., Vallon, O., Wollman, F. A. and Choquet, Y. (2019) MDA1, a nucleus-encoded factor involved in the stabilization and processing of the atpA transcript in the chloroplast of Chlamydomonas. The Plant journal : for cell and molecular biology.

Walker, J. E. (2013) The ATP synthase: the understood, the uncertain and the unknown. Biochemical Society transactions, 41, 1-16.

Waltz, F., Nguyen, T. T., Arrive, M., Bochler, A., Chicher, J., Hammann, P., Kuhn, L., Quadrado, M., Mireau, H., Hashem, Y. and Giege, P. (2019) Small is big in Arabidopsis mitochondrial ribosome. Nature plants, 5, 106-117.

Wang, C., Aube, F., Quadrado, M., Dargel-Graffin, C. and Mireau, H. (2018) Three new pentatricopeptide repeat proteins facilitate the splicing of mitochondrial transcripts and complex I biogenesis in Arabidopsis. Journal of experimental botany, 69, 5131-5140.

Wang, F., Johnson, X., Cavaiuolo, M., Bohne, A. V., Nickelsen, J. and Vallon, O. (2015) Two Chlamydomonas OPR proteins stabilize chloroplast mRNAs encoding small subunits of photosystem II and cytochrome b6 f. The Plant journal : for cell and molecular biology, 82, 861-873.

Wang, F., Qi, Y., Malnoe, A., Choquet, Y., Wollman, F. A. and de Vitry, C. (2017) The High Light Response and Redox Control of Thylakoid FtsH Protease in Chlamydomonas reinhardtii. Molecular plant, 10, 99-114.

Wei, L., Derrien, B., Gautier, A., Houille-Vernes, L., Boulouis, A., Saint-Marcoux, D., Malnoe, A., Rappaport, F., de Vitry, C., Vallon, O., Choquet, Y. and Wollman, F. A. (2014) Nitric oxide-triggered remodeling of chloroplast bioenergetics and thylakoid proteins upon nitrogen starvation in Chlamydomonas reinhardtii. The Plant cell, 26, 353-372.

Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S. and Martin, W. F. (2016) The physiology and habitat of the last universal common ancestor. Nature Microbiology, 1.

Westall, F., Cavalazzi, B., Lemelle, L., Marrocchi, Y., Rouzaud, J.-N., Simionovici, A., Salomé, M., Mostefaoui, S., Andreazza, C., Foucher, F., Toporski, J., Jauss, A., Thiel, V., Southam, G., MacLean, L., Wirick, S., Hofmann, A., Meibom, A., Robert, F. and Défarge, C. (2011) Implications of in situ calcification for photosynthesis in a ~3.3Ga-old microbial biofilm from the Barberton greenstone belt, South Africa. Earth and Planetary Science Letters, 310, 468-479.

Westall, F., de Ronde, C. E., Southam, G., Grassineau, N., Colas, M., Cockell, C. and Lammer, H. (2006) Implications of a 3.472-3.333 Gyr-old subaerial microbial mat from the Barberton greenstone belt, South Africa for the UV environmental conditions on the early Earth. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 361, 1857-1875.

Williams-Carrier, R., Kroeger, T. and Barkan, A. (2008) Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. RNA (New York, N.Y.), 14, 1930-1941.

Wintermans, J. F. G. M. (1960) Concentrations of phosphatides and glycolipids in leaves and chloroplasts. Biochimica et Biophysica Acta, 44, 49-54.

Wittkopp, T. M., Schmollinger, S., Saroussi, S., Hu, W., Zhang, W., Fan, Q., Gallaher, S. D., Leonard, M. T., Soubeyrand, E., Basset, G. J., Merchant, S. S., Grossman, A. R., Duanmu, D. and Lagarias, J. C. (2017) Bilin-Dependent Photoacclimation in Chlamydomonas reinhardtii. The Plant cell, 29, 2711-2726.

Woessner, J. P., Gillham, N. W. and Boynton, J. E. (1986) The sequence of the chloroplast atpB gene and its flanking regions in Chlamydomonas reinhardtii. Gene, 44, 17-28.

Wollman, F. A. (2016) An antimicrobial origin of transit peptides accounts for early endosymbiotic events. Traffic, 17, 1322-1328.

Wostrikoff, K., Choquet, Y., Wollman, F. A. and Girard-Bascou, J. (2001) TCA1, a single nuclear-encoded translational activator specific for petA mRNA in Chlamydomonas reinhardtii chloroplast. Genetics, 159, 119-132.

Wostrikoff, K., Girard-Bascou, J., Wollman, F. A. and Choquet, Y. (2004) Biogenesis of PSI involves a cascade of translational autoregulation in the chloroplast of Chlamydomonas. The EMBO journal, 23, 2696-2705.

Wostrikoff, K. and Stern, D. (2007) Rubisco large-subunit translation is autoregulated in response to its assembly state in tobacco chloroplasts. Proceedings of the National Academy of Sciences of the United States of America, 104, 6466-6471.

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T. and Nakamura, T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. PloS one, 8, e57286.

Yagi, Y., Nakamura, T. and Small, I. (2014) The potential for manipulating RNA with pentatricopeptide repeat proteins. The Plant journal : for cell and molecular biology, 78, 772-782.

Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H. W., Wang, J., Zhu, J. K., Shi, Y. and Yan, N. (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. Nature, 504, 168-171.

Yu, Q., Barkan, A. and Maliga, P. (2019) Engineered RNA-binding protein for transgene activation in non-green plastids. Nature plants, 5, 486-490.

Zambrano, A., Fontanesi, F., Solans, A., de Oliveira, R. L., Fox, T. D., Tzagoloff, A. and Barrientos, A. (2007) Aberrant translation of cytochrome c oxidase subunit 1 mRNA species in the absence of Mss51p in the yeast Saccharomyces cerevisiae. Molecular biology of the cell, 18, 523-535.

Zerges, W. and Rochaix, J. D. (1994) THE 5' LEADER OF A CHLOROPLAST MESSENGER-RNA MEDIATES THE TRANSLATIONAL REQUIREMENTS FOR 2 NUCLEUS-ENCODED FUNCTIONS IN CHLAMYDOMONAS-REINHARDTII. Mol. Cell. Biol. , 14, 5268-5277.

Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F. and Papke, R. T. (2006) Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. Genome research, 16, 1099-1108.

Zhou, W., Lu, Q., Li, Q., Wang, L., Ding, S., Zhang, A., Wen, X., Zhang, L. and Lu, C. (2017) PPR-SMR protein SOT1 has RNA endonuclease activity. Proceedings of the National Academy of Sciences of the United States of America, 114, E1554-E1563.

Zones, J. M., Blaby, I. K., Merchant, S. S. and Umen, J. G. (2015) High-Resolution Profiling of a Synchronized Diurnal Transcriptome from Chlamydomonas reinhardtii Reveals Continuous Cell and Metabolic Differentiation. The Plant cell, 27, 2743-2769.

Zoschke, R. and Barkan, A. (2015) Genome-wide analysis of thylakoid-bound ribosomes in maize reveals principles of cotranslational targeting to the thylakoid membrane. Proceedings of the National Academy of Sciences of the United States of America, 112, E1678-1687.

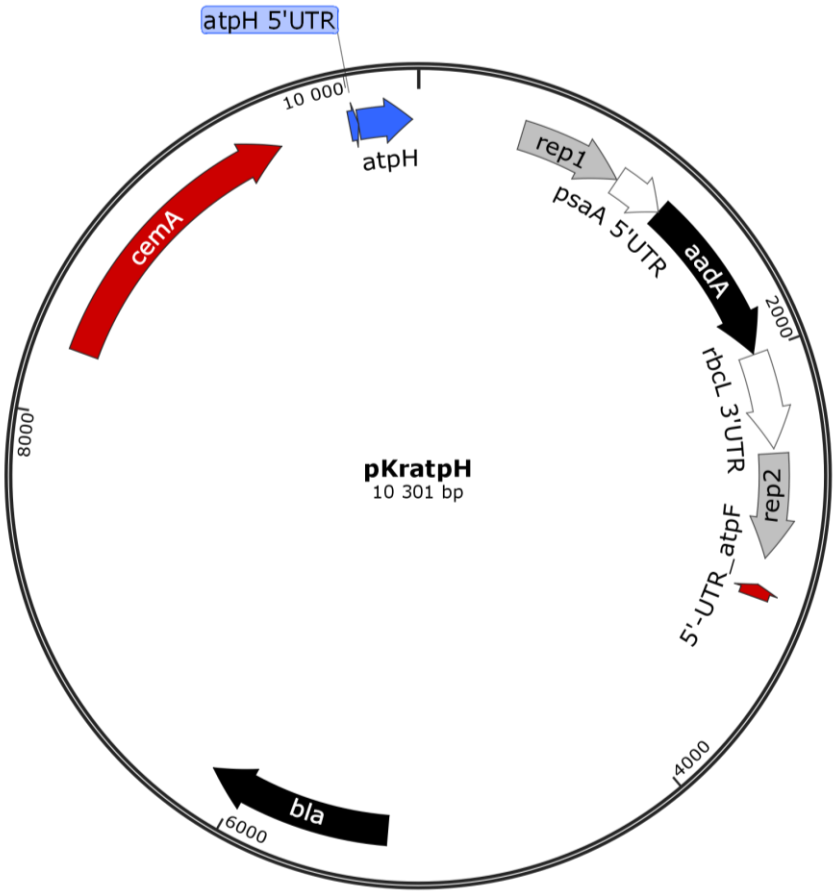
Zoschke, R., Kroeger, T., Belcher, S., Schottler, M. A., Barkan, A. and Schmitz-Linneweber, C. (2012) The pentatricopeptide repeat-SMR protein ATP4 promotes translation of the chloroplast atpB/E mRNA. The Plant journal : for cell and molecular biology, 72, 547-558.

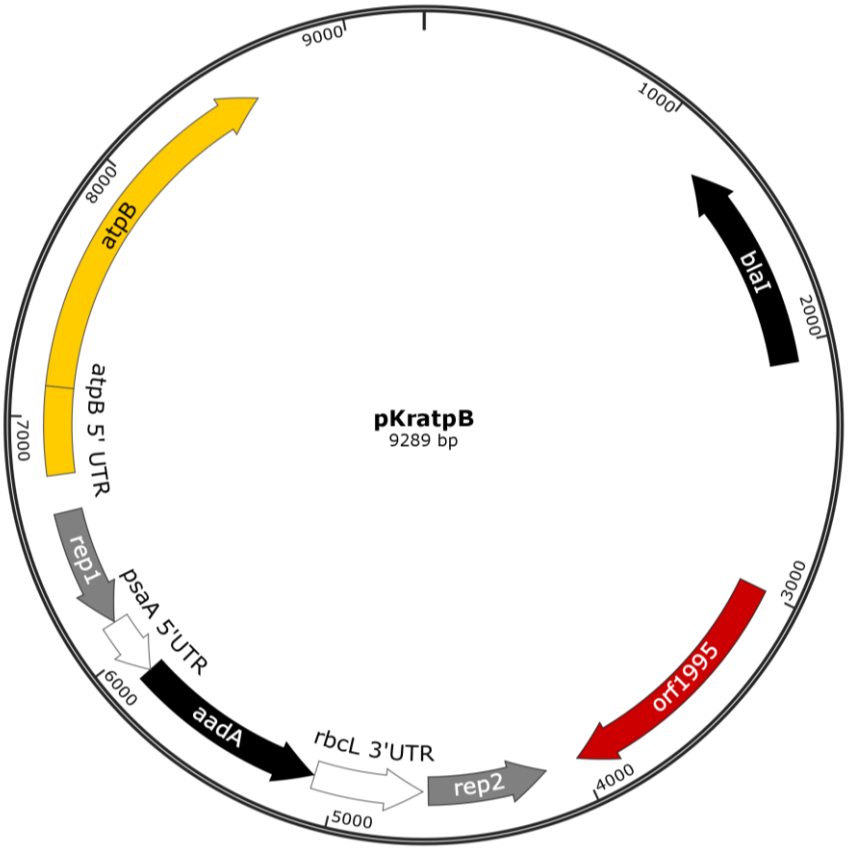
Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research, 31, 3406-3415.

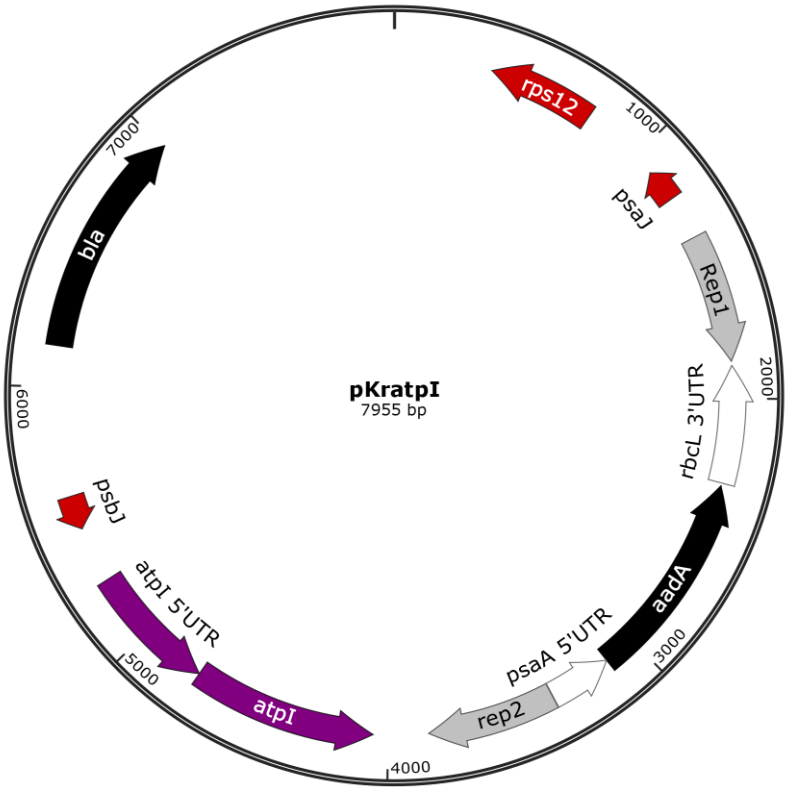
ANNEXES

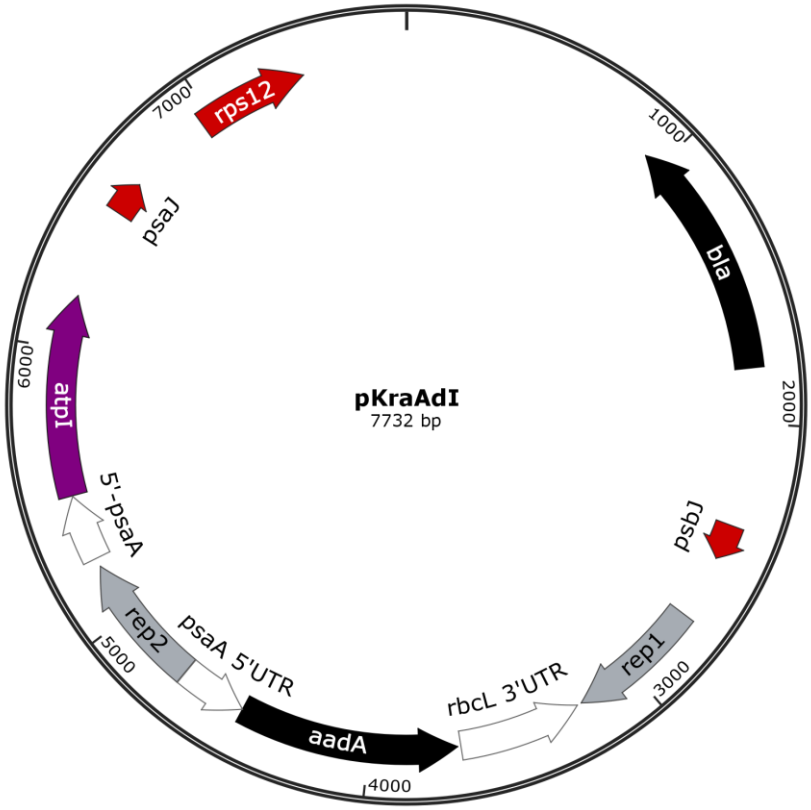
1) PLASMIDS
2) PRIMERS
3) ALIGNMENT OF MTH2
4) OVERLAPPING HPR AND OPR PROTEINS IN C. REINHARDTII
5) SYNTHETIC MDB1 FRAGMENTS
ARTICLE 1: PRELIMINARY DRAFT
ARTICLE 2: PRELIMINARY DRAFT
ARTICLE 3: SUBMITTED
ARTICLE 4: PRELIMINARY DRAFT

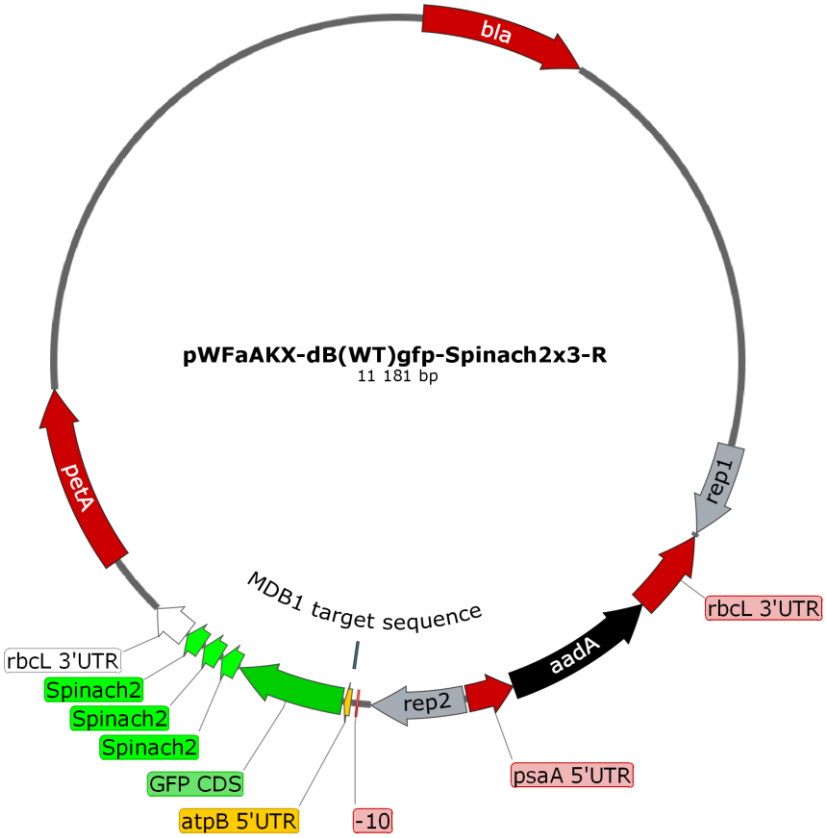
1) PLASMIDS

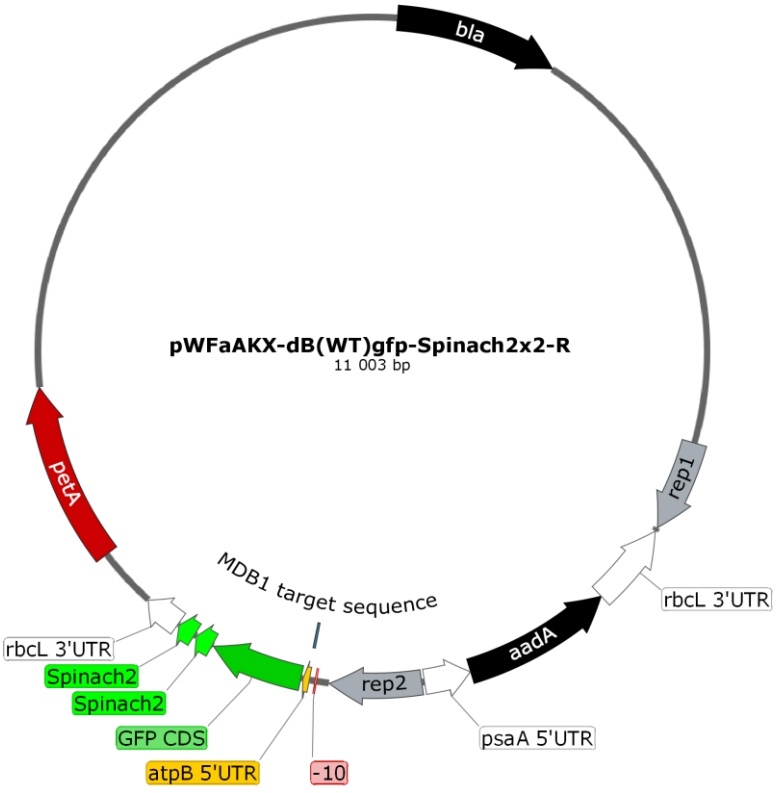

atpH 5'UTR
10 000
atpH
rep1
psaA 5'UTR
aadA
2000
rbcL 3'UTR
rep2
5'-UTR_atpF
4000
bla
6000
8000
cemA
pKratpH
10 301 bp

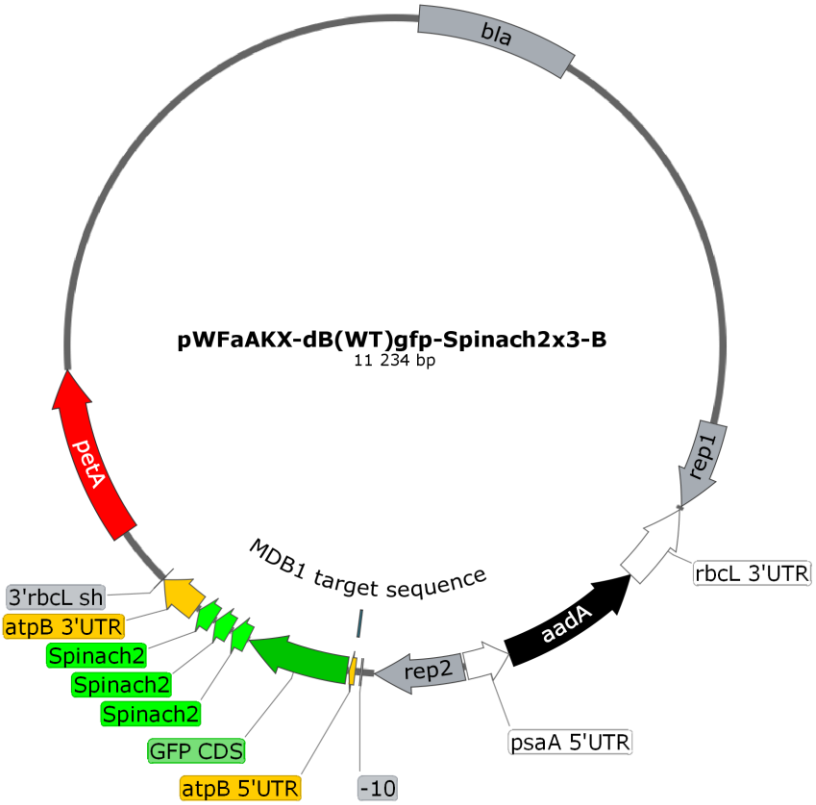

9000
1000
2000
blaI
3000
orf1995
4000
rbcL 3'UTR
rep2
5000
aadA
6000
psaA 5'UTR
rep1
7000
atpB 5' UTR
8000
atpB
pKratpB
9289 bp


pKratpI
7955 bp
rps12
psaJ
Rep1
rbcL 3'UTR
aadA
psaA 5'UTR
rep2
atpI
atpI 5'UTR
psbJ
bla
1000
2000
3000
4000
5000
6000
7000


pKraAdI
7732 bp
rps12
psaJ
bla
psbJ
rep1
rbcL 3'UTR
aadA
psaA 5'UTR
rep2
5'-psaA
atpI
1000
2000
3000
4000
5000
6000
7000


bla
pWFaAKX-dB(WT)gfp-Spinach2x3-R
11 181 bp
petA
rep1
MDB1 target sequence
rbcL 3'UTR
aadA
rbcL 3'UTR
Spinach2
Spinach2
Spinach2
rep2
GFP CDS
psaA 5'UTR
atpB 5'UTR
-10


bla
pWFaAKX-dB(WT)gfp-Spinach2x2-R
11 003 bp
petA
rep1
MDB1 target sequence
rbcL 3'UTR
aadA
rbcL 3'UTR
Spinach2
Spinach2
rep2
GFP CDS
psaA 5'UTR
atpB 5'UTR
-10


bla
pWFaAKX-dB(WT)gfp-Spinach2x3-B
11 234 bp
rep1
rbcL 3'UTR
aadA
psaA 5'UTR
rep2
-10
atpB 5'UTR
GFP CDS
Spinach2
Spinach2
Spinach2
atpB 3'UTR
3'rbcL sh
MDB1 target sequence
petA

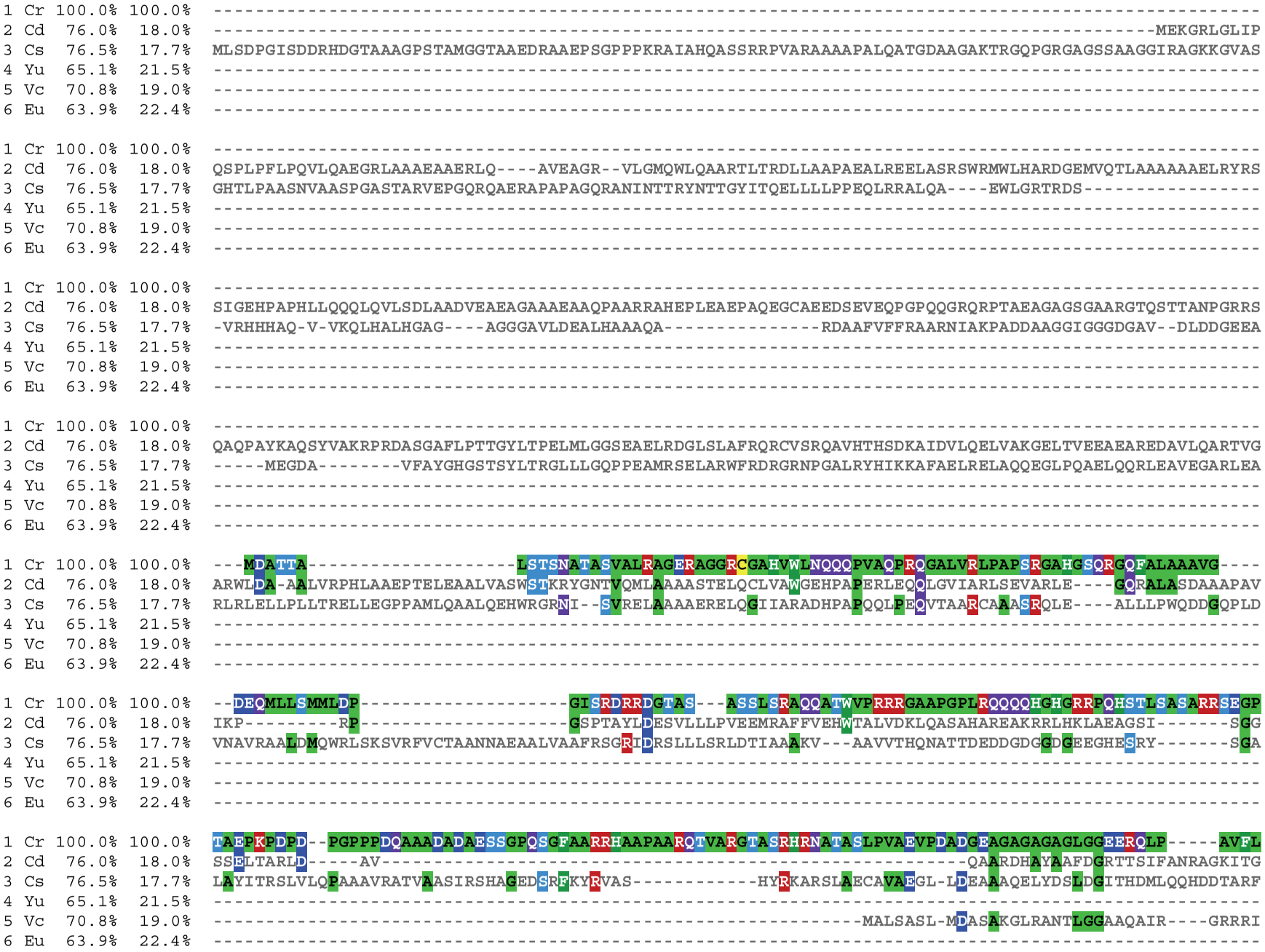
2) PRIMERS

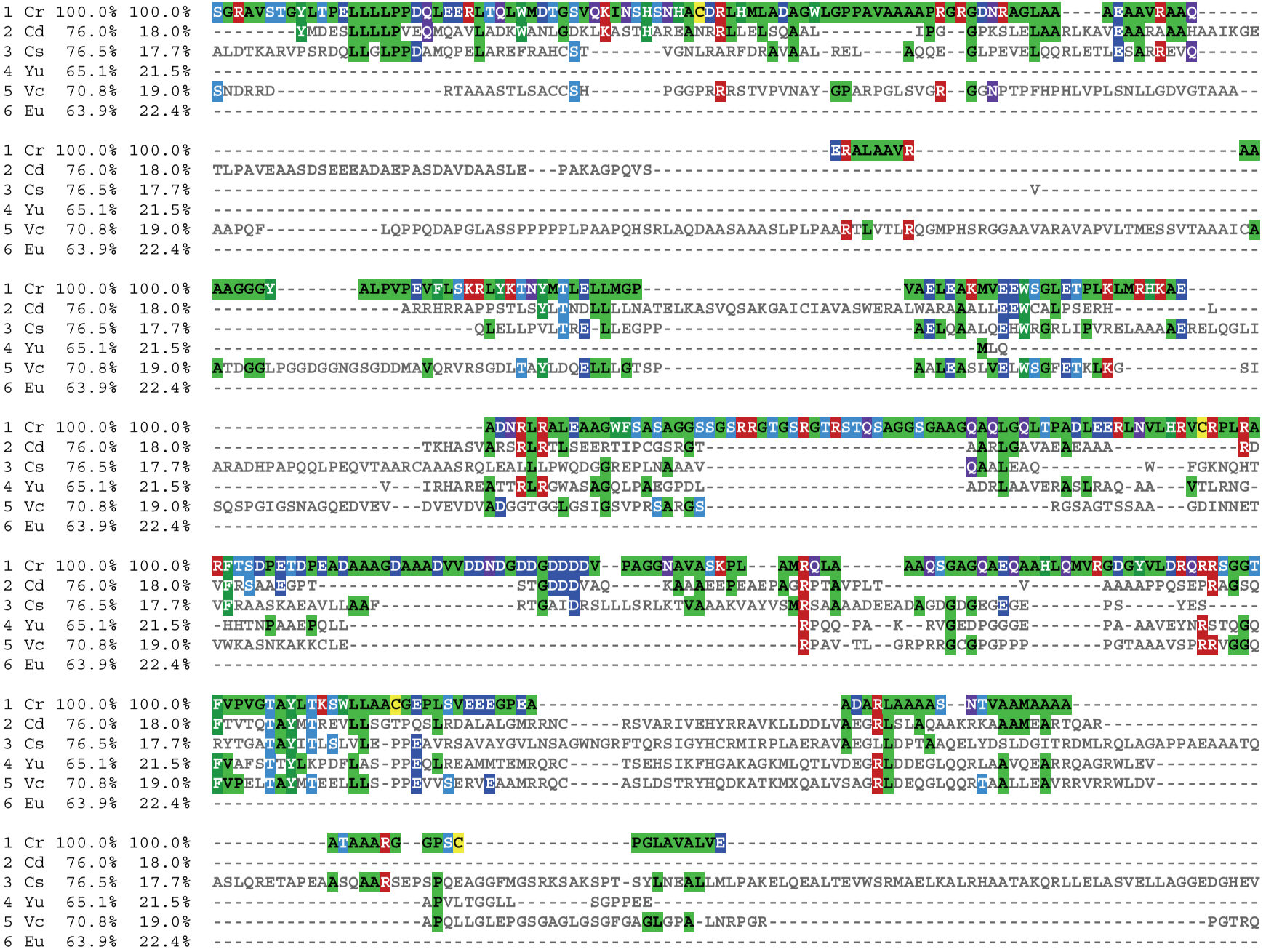
Application	Name	Sequence (5'->3')
atpH MTHI1 target mutagenesis	cemAFW	GCGAATTCCGGAAAGTCAAACAGGTATTTTCTT
	atpHext-RV	GCGTTAGCCAATACCAAACAGC
	Mut-atpH-1FW	ATTCTTTGGAAGTTATCGATTTTATTGATTCATTTAG
	Mut-atpH-1RV	TCGATAACTTCCAAAGAATATTATATTCTT
	Mut-atpH-2FW	CTTTGGTTGAAATCGATTTTATTGATTCATTTAG
	Mut-atpH-2RV	AAAATCGATTTCAACCAAAGAATATTATATTCTT
	Mut-atpH-3FW	ATTCTTTGGAACAAATCGATTTTATTGATTCATTTAG
	Mut-atpH-3RV	AAAATCGATTTGTTCCAAAGAATATTATATTCTT
atpB MDB1 target mutagenesis	atpB5'FW1	GCGCTCGAGCTTAAGTTCAAAATTCTCCACCAGCT
	atpB5'FW2	GCGCTCGAGGCTAGCTTCAAAATTCTCCACCAGCT
	atpB5'FW3	GCGCTCGAGAGATCTTTCAAAATTCTCCACCAGCT
	atpB5'FW4	GCGCTCGAGAGGCCTTTCAAAATTCTCCACCAGCT
	atpB5'FW5	GCGCTCGAGGGGCCCTTCAAAATTCTCCACCAGCT
	atpB5'FW6	GCGCTCGAGCCTAGGTTCAAAATTCTCCACCAGCT
	atpB5'FW7	GCGCTCGAGACGCGTTTCAAAATTCTCCACCAGCT
	dB1FW	ACTAAAAAAGGGGCGTTAGTGAATAATACTTTTT
	dB1RV	TCACTAACGCCCCTTTTTTAGTTTTTTCATTTAACT
	dB2FW	ACTAAAAAGCGGGCGTTAGTGAATAATACTTTTT
	dB2RV	TCACTAACGCCCGCTTTTTAGTTTTTTCATTTAACTAT
	dB3FW	ACTAAGGGGTAAGCGTTAGTGAATAATACTTTTT
	dB3RV	TCACTAACGCTTACCCCTTAGTTTTTTCATTTAACTATAT
	dB4FW	ACTAAAGGATAAGCGTTAGTGAATAATACTTTTT
	dB4RV	TCACTAACGCTTATCCTTTAGTTTTTTCATTTAACTATA
	dB5FW	ACTAAAAAATACTAGTTAGTGAATAATACTTTTT
	dB5RV	TCACTAACTAGTATTTTTTAGTTTTTTCATTTAACT
	dB6FW	ACTAAAAAATACTTGTTAGTGAATAATACTTTTT
	dB6RV	TCACTAACAAGTATTTTTTAGTTTTTTCATTTAACT
	dB7FW	ACTAAAAAATAAGAATTAGTGAATAATACTTTTTata
	dB7RV	TCACTAATTCTTATTTTTTAGTTTTTTCATTTAACT
	dB8FW	ACTAAAAAATAAGCAATAGTGAATAATACTTTTTata
	dB8RV	TCACTATTGCTTATTTTTTAGTTTTTTCATTTAACT
	dB10FW	ACTAAAAAATAAGAAAAAGTGAATAATACTTTTTATATA
	dB10RV	TCACTTTTTCTTATTTTTTAGTTTTTTCATTTAACT
	dB11FW	ACTAAAAAATAAGCGGGAGTGAATAATACTTTTTATATA
	dB11RV	TCACTCCCGCTTATTTTTTAGTTTTTTCATTTAACT
	dB12FW	ACTAAAAAATAAGCGAAAGTGAATAATACTTTTTATATA
	dB12RV	TCACTTTCGCTTATTTTTTAGTTTTTTCATTTAACT
	dBExt_RV	TTTGAAATAAGAACCTCCTCCTTCC
	atpB_Anton_FW	cgcCTCGAGAAGATGCTTTGCATCTCTAA
	atpB_Anton_WT_RV	gcgCCCGGGCCCATATAAAAAGTATTATTCACTAAC
	atpB_Anton_M1_RV	gcgCCCGGGAATTCAATATAAAAAGTATTATTCACTGGCGCTTA TTTTTTAGTTTTTTCAT
	atpB_Anton_M2_RV	gcgCCCGGGATCCATATATAAAAAGTATTATTCACTCCCGCTTA TTTTTTAGTTTTTTCAT
	atpB_Anton_M3_RV	gcgCCCGGGAGATCTATATAAAAAGTATTATTCACTTTCGCTTA TTTTTTAGTTTTTTCAT

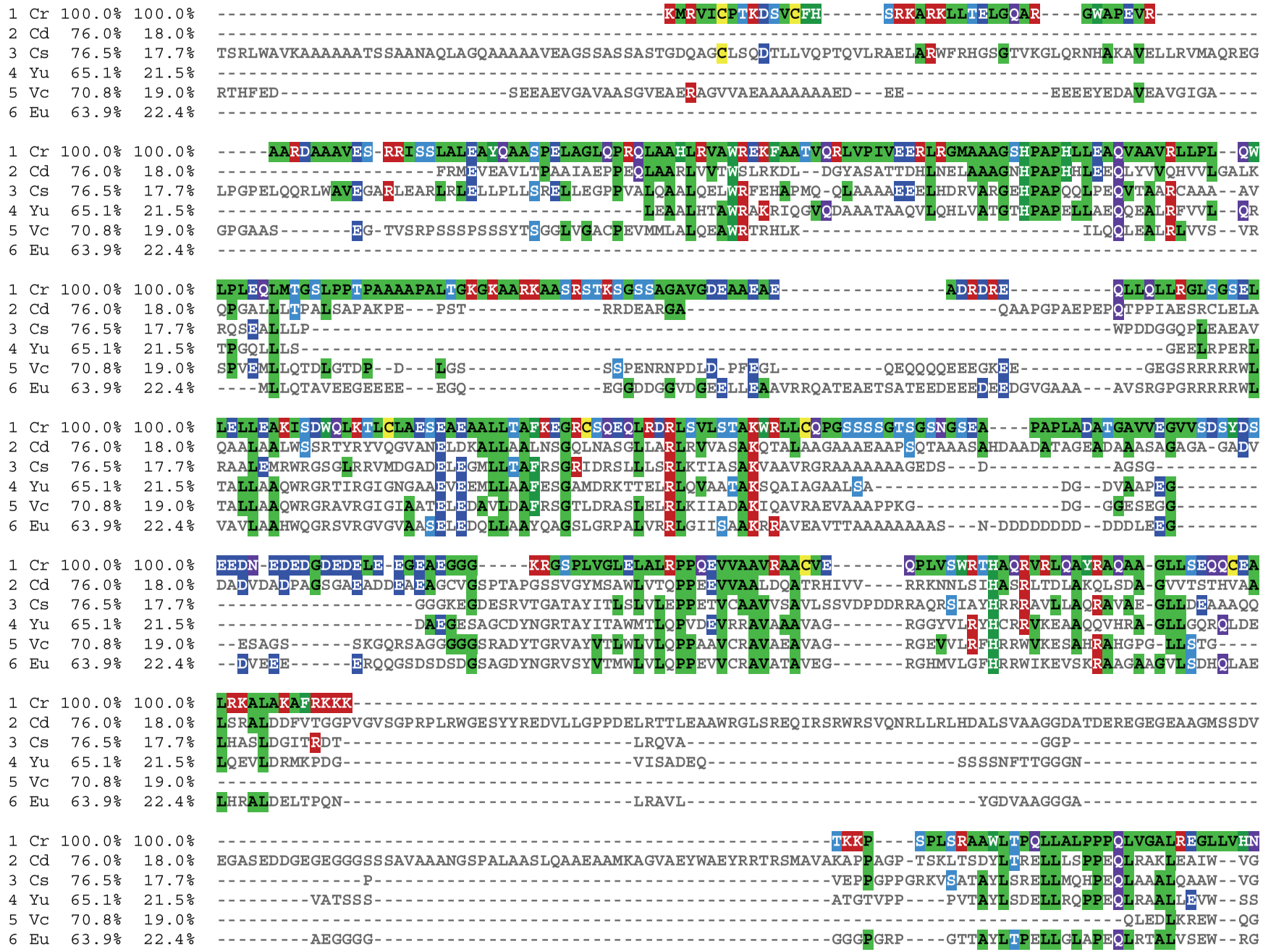
MDB1 target verification	atpB_TT_RV	AAAGTATTATTCACTAACGCAA
	atpB_GG_RV	AAAGTATTATTCACTAACGCCC
	atpB_CC_RV	AAAGTATTATTCACTAACGCGG
	atpB_CT_RV	AAAGTATTATTCACTAACGAG
	atpB_FW	ACCTCGAGTTCAAAATTCTC
	atpB_WT_FW	AGTTAAATGAAAAAACTAAAAAATAA
	atpB_RV	ATTCTTACGTATAAACCCCG
	atpBCDS_RV	TGCTGAGTTTTTAGCACGAATA
	atpBSeqRV	AAATCCACCGTTTTGTGGAA
	atpBSeqFW	GGAGACCTTCAAGCCGTACA
gfp chimeric construct	GFP-CDS-FW	CGCGAATTCGCGCTCGAGGCGCCCCGGGCCATGGGTAAAGG AGAAGAACTTTTCACTGG
	GFP-CDS-RV2	GCGCTGCAGTTATTTGTATAGTTCATCCATGCCATGTG
thm24.2^{-} phenotyping	MDB1 bFW	CTGCACTCAGGTCTTAGTCTGGC
	MDB1 aRV	CGCATCTCTTCTTTCCACGACTC
K4.20 phenotyping	TOC1 FW	CGCTACACGGCAGTAAGGAG
	MDB1 ex3 RV	GCGCCGTTTGAGCTATTTGA
atpI Stop mutagenesis	atpI Stop FW	AGAATTTTATTACATTTTTTAGATCTCTTTATTAGAAATTGCTG AAGTATCTGTA
	atpI Stop RV	TCAGCAATTTCTAATAAAGAGATCTAAAAAATGTAATAAAATT CTTACCA
	atpI 5' FW	ACTGGTCATTATTTATAGTGGT
	atpI ext RV	GCCCTTATCAAGCTTCCACATAGCGT
atpH pG homoplamy verification	atpH pG FW	CGTTCATCGCCAGCTACAGTTGC
	atpH ext RV	GCGTTAGCCAATACCAAACAGC
	cemA FW	GCGAATTCCGGAAAGTCAAACAGGTATTTTCTT
aAdI verification	aAdI test FW	GTGAAGTTTGGAAAGAAATT
	aAdI test RV	TTTGTGTTTTTGCTAAATCA
Labelling of digoxigenin PCR probes	atpH-dig-FW	AACCCTATCGTAGCTGCTGC
	atpH-dig-RV	ACTAGACCGTAAATTGTTAA
	atpB-dig-FW	CACGGTGGTGTTTCTGTATT
	atpB-dig-RV	TTACGCTTTGTGCAGAATCA
	gfp-dig-FW	TTGAATTAGATGGTGATGTT
	gfp-dig-RV	CAATTGGAGTATTTTGTTGA
	petB-dig-FW	GCTGTTATTTTAGGTATGGC
	petB-dig-RV	GATGCGTTGTAAATAGTGTT
	psbD FW	GAGCTAAACCTACAACACCA
	psbD RV	CAGTATGGCTCACTCTCTTC
	atpE-dig FW	TTTCTATTTTAACACCAGAG
	atpE-dig RV	AAGTACTTTTACAACTTGGT
atpB cRT-PCR characterisation	B_RT	ACATTGTGGCTTTACTTTCA
	B_FW	GGGAAGGAGGAGGTTCTTAT
	B_RV	CACCTTTAACTGAGAGTGTAAC
	cRT_atpB_FW0	GTAAAGCTGCTTCATTAAAA
	cRT_atpB_FW1	GTTTCTTATCACAACCATTC
	cRT_atpB_FW2	TAGCTGCTAAAGGTATTTAC
	cRT_atpB_FW3	CATTGATAACATTTTCCGTT
	atpB_dig FW	CACGGTGGTGTTTCTGTATT

3) Alignment of MTH2

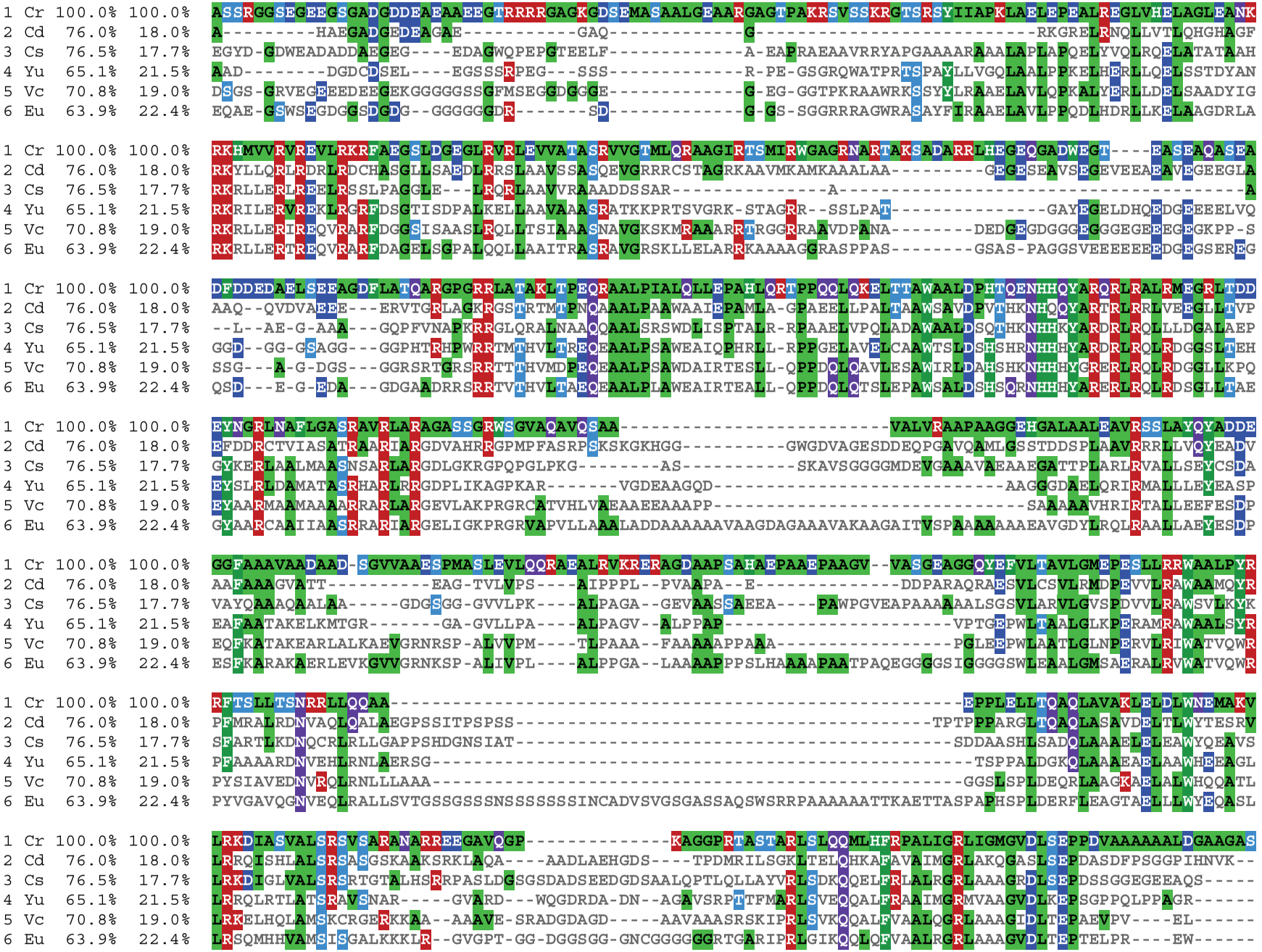
Alignment of MTH2with some of its orthologues

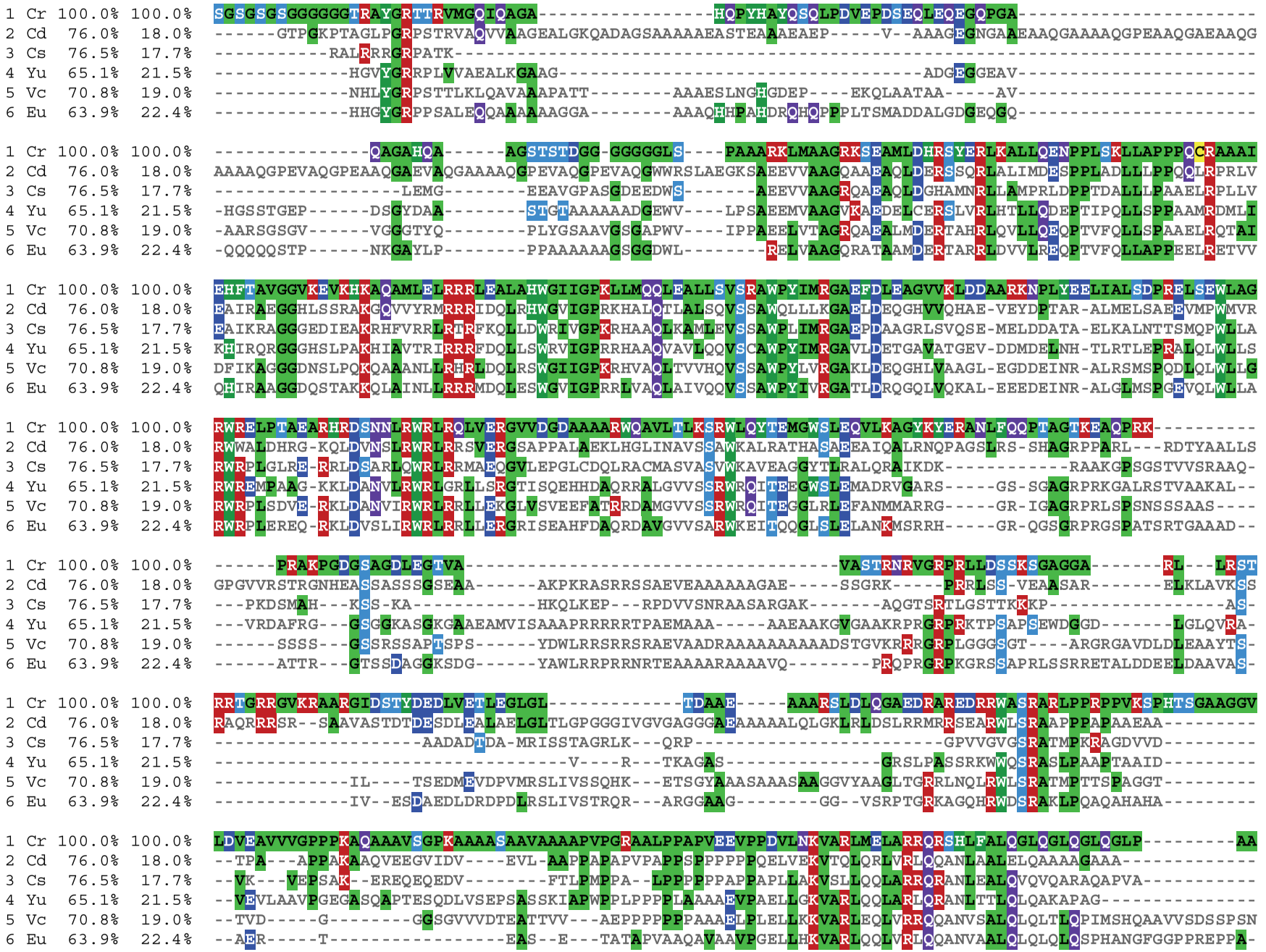


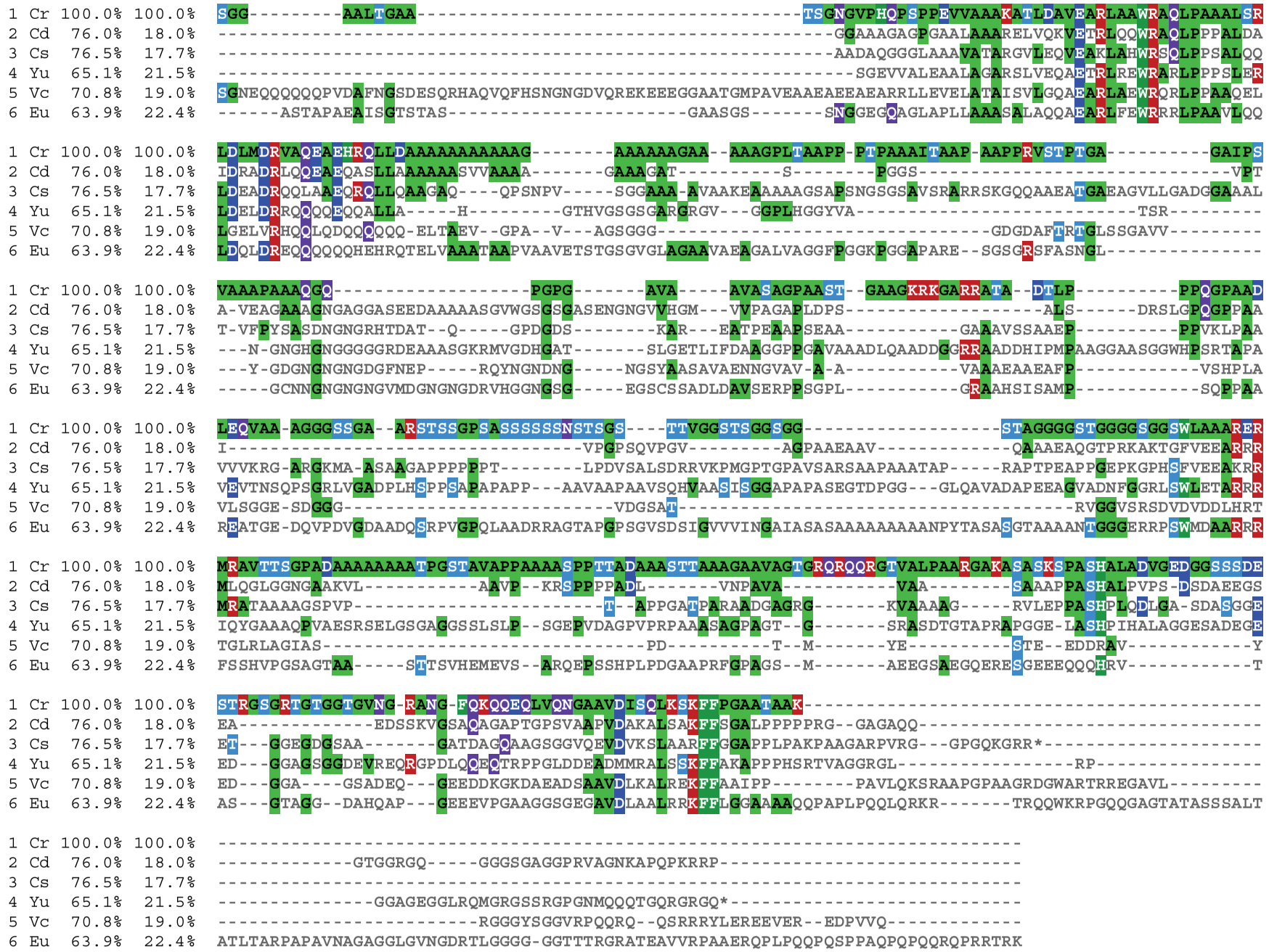

1 Cr 100.0% 100.0%
2 Cd 76.0% 18.0%
3 Cs 76.5% 17.7%
4 Yu 65.1% 21.5%
5 Vc 70.8% 19.0%
6 Eu 63.9% 22.4%


1 Cr 100.0% 100.0%
2 Cd 76.0% 18.0%
3 Cs 76.5% 17.7%
4 Yu 65.1% 21.5%
5 Vc 70.8% 19.0%
6 Eu 63.9% 22.4%


1 Cr 100.0% 100.0%
2 Cd 76.0% 18.0%
3 Cs 76.5% 17.7%
4 Yu 65.1% 21.5%
5 Vc 70.8% 19.0%
6 Eu 63.9% 22.4%


1 Cr 100.0% 100.0%
2 Cd 76.0% 18.0%
3 Cs 76.5% 17.7%
4 Yu 65.1% 21.5%
5 Vc 70.8% 19.0%
6 Eu 63.9% 22.4%


1 Cr 100.0% 100.0%
2 Cd 76.0% 18.0%
3 Cs 76.5% 17.7%
4 Yu 65.1% 21.5%
5 Vc 70.8% 19.0%
6 Eu 63.9% 22.4%


1 Cr 100.0% 100.0%
2 Cd 76.0% 18.0%
3 Cs 76.5% 17.7%
4 Yu 65.1% 21.5%
5 Vc 70.8% 19.0%
6 Eu 63.9% 22.4%

4) OVERLAPPING HPR AND OPR PROTEINS IN C. REINHARDTII

Proteins identified as HPR in *Chlamydomonas reinhardtii* from (Hillebrand et al., 2018)

Accession	Name	OPR?
20686	NCL 69	YES
20689	NCL 90	YES
21996	NCL 75	YES
22494	NCL 79	YES
35878	OPR107	YES
144190	OPR48	YES
147277	OPR100	YES
148806	OPR9	YES
151373	OPR119	YES
152682	-	NO
167067	OPR23	YES
170988	NCL 20	YES
178853	OPR68	YES
179193	NCL 98	YES
180625	NCL 112	YES
180839	NCL 84	YES
181188	NCL 91	YES
191389	OPR10	YES
196763	NCL 111	YES
196765	NCL 106	YES
287603	NCL 78	YES
287613	NCL 88	YES
288887	OPR104	YES
296752	OPR31	YES
306518	OPR23	YES
390299	OPR35	YES
402283	OPR28	YES
403175	NCL 110	YES
403647	NCL 93	YES
405879	OPR2	YES
406560	-	NO
416539	OPR5	YES
417156	OPR8	YES
419975	OPR15	YES
420894	OPR47	YES
422092	-	NO

5) Synthetic MDB1 Fragments

SYNTHETIC MDB1 FRAGMENTS

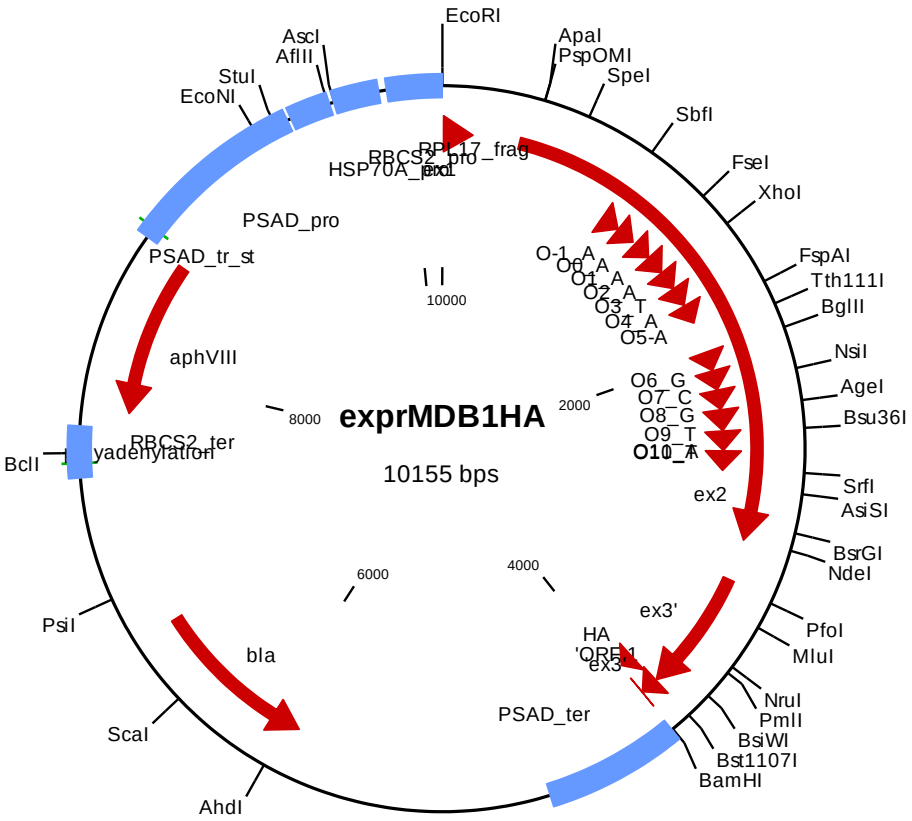
```
LOCUS       pMDB-HA-Strep_JH       10182 bp    DNA     circular SYN 18-DEC-2018
DEFINITION  Ligation of Fragment 2 into Fragment 2
ACCESSION   pMDB-HA-Strep_JH
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown
            Unclassified.
REFERENCE   1  (bases 1 to 10182)
  AUTHORS   Self
  JOURNAL   Unpublished.
COMMENT     SECID/File created by Clone Manager, Scientific & Educational
Software
COMMENT     SECNOTES|Vector molecule:  Fragment 2
               Fragment ends: BamHI and BstZ17I
               Fragment size: 10060
            Insert molecule:  Fragment 2
               Fragment ends: BstZ17I and BamHI
               Fragment size: 122
FEATURES             Location/Qualifiers
     misc_feature    659..891
                     /gene="RBCS2_ter"
                     /product="RBCS2 3' UTR and cleavage/polyadenylation
                     signal"
     misc_feature    complement(721)
                     /gene="polyadenylation"
                     /product="RBCS2 cleavage/polyadenylation signal TGTAA"
     CDS             complement(970..1773)
                     /gene="aphVIII"
     misc_feature    complement(1796..2616)
                     /gene="PSAD_pro"
                     /product="PsaD promoter and 5' UTR up to start codon;
                     transcription starts at TTGACTCG..; the first 220 nt are
                     repeated on genome 8kb upstream in tandem"
     misc_feature    complement(1844)
                     /gene="PSAD_tr_st"
                     /product="trabnscription start site of PSAD"
     misc_feature    complement(2633..2818)
                     /gene="HSP70A_pro"
                     /product="confers high probability of expression to
                     downstream RBCS2 promoter"
     misc_feature    2837..3044
                     /gene="RBCS2_pro"
     misc_feature    3079..3328
                     /gene="RPL17_frag"
                     /product="start of CDS, first intron, start of second
                     exon"
     CDS             3335..3429
                     /gene="e1"
     misc_feature    3488..3725
                     /gene="i1"
     CDS             3726..6341
                     /gene="e2"
     CDS             4419..4535
                     /gene="O0"
     CDS             4548..4658
                     /gene="O1"
```

```
     CDS             4659..4772
                     /gene="O2"
     CDS             4773..4886
                     /gene="O3"
     CDS             4887..5003
                     /gene="O4"
     CDS             5004..5117
                     /gene="O5"
     CDS             5298..5411
                     /gene="O6"
     CDS             5412..5525
                     /gene="O7"
     CDS             5526..5636
                     /gene="O8"
     CDS             5637..5760
                     /gene="O9"
     CDS             5763..5876
                     /gene="O10"
     CDS             5877..5993
                     /gene="O11"
     CDS             6000..6113
                     /gene="O12"
     misc_feature    6342..6553
                     /gene="i2"
     CDS             6554..7197
                     /gene="ex3'
     CDS             7196..7197
                     /gene="3HA'"
     CDS             7320..7324
                     /gene="'ex3"
                     /codon_start=3
     misc_feature    complement(7324..7944)
                     /gene="PSAD_ter"
                     /product="nt 2465-2470 differ from genomic sequence"
     CDS             complement(9194..10054)
                     /gene="bla"
ORIGIN
        1 ctaaattgta agcgttaata ttttgttaaa attcgcgtta aatttttgtt aaatcagctc
       61 atttttttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag aatagaccga
      121 gatagggttg agtgttgttc agtttggaa caagagtcca ctattaaaga acgtggactc
      181 caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc
      241 ctaatcaagt tttttggggt cgaggtgccg taaagcacta aatcggaacc ctaaagggag
      301 cccccgattt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa
      361 agcgaaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgctgc gcgtaaccac
      421 cacacccgcc gcgcttaatg cgccgctaca gggcgcgtcc cattcgccat tcaggctgcg
      481 caactgttgg gaagggcgat cggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg
      541 gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg
      601 taaaacgacg gccagtgagc gcgcgtaata cgactcacta tagggcgaat tgggtacccg
      661 cttcaaatac gcccagcccg cccatggaga aagaggccaa aatcaacgga ggatcgttac
      721 aaccaacaaa attgcaaaac tcctccgctt tttacgtgtt gaaaaagact gatcagcacg
      781 aaacggggag ctaagctacc gcttcagcac ttgagagcag tatcttccat ccaccgccgt
      841 tcgtcagggg gcaaggctca gatcaacgag cgcctccatt tacacggagc ggggatcgat
      901 cccaacgtcc acactgtgct gtcacccacg cgacgcaacc ctacccagcc accaacacca
      961 tcaggtccct cagaagaact cgtccaacag ccggtaaaac gccagctttt cctccgatac
     1021 cgccccatcc cacccgcgcc cgtactcccg caggaacgcc gcggaacact ccggcccgaa
     1081 ccacgggtcc tcctcgtggg ccagctcgcg cagcaccagc gcgagatcgg agtgccggtc
     1141 cgcacggccg acccgcccca cgtcgatcag cccggtcacc tcgcaggtac gagggtcgag
     1201 cagcacgttg tccgggcaca ggtgaccgtg gcaaaccgcc agatcctcgt ccgcaggccg
     1261 agtccgctcc agctcggcga gaagccgctc ccccgaccac cccttccgct cctcgtccag
     1321 atcctccaag tcgacgctcc cttcagcgac agcacgggcc gcctgcggca ccgtcaccgc
     1381 gagactgcga tcgaacggac accgctccca gtccagcgcg tgcagcgaac gagcgagccc
```

1441 cgcgagcgcc accgccacgt ccagccgctg ctcccgcggc caccgcgcac tggccggacg
1501 ccccggaacc gcttcggtga ccaaccaggc gaccctctcg tccccaccac cctccacaac
1561 acgaggtacg ggaatcccca cctccgccaa ccacaccagc cgctcagcct cacccaacaa
1621 gcccaccccg gcccccagag ctgccacctt gacaaacaac tcccgcccac caccccgaag
1681 ccgataaaca ccagcccccg aggccccatc ctccacaaca acccactcac aaccgggata
1741 ccgaccccgc agtgcacgca acgcatcgtc catgaattaa ttccgattgc tgcagttggc
1801 ttgttgtgag tagcagtggg gtcctagaat gcacaacgag tcaagagcgc aacacctaac
1861 cctggcttgc tcggcgagga aacctccccc cgagcaagcc atctcggtcg tacctccaat
1921 tcccagatcc ccctccccag cccgcaggag ccctcgcacg agtgcccgaa cgcaacaata
1981 ttgatacata atcgtccctg gcctggggga agggccgcta acgcgccggg ccgtcgcgta
2041 aataccaata atcacgccgc gtcccacttt gctctctcgc cttgcaactt aaaagcctac
2101 tgcctcgcca gatttgctcc aattgtgcta caaatgacaa ttgcgtccga atatggggcc
2161 gagcgcgtca ggaaaggtgg cactgcgatc tgctaacatg tctcggaacg acgacaagaa
2221 gcggcttttt aaggactccg agttcgggca aacatgagca tttgcctgcc ttcacgcatc
2281 ggtaggtgtg gaggcgcgcg tggagaaagg cacccgaact ggcggcgagc ccttcgaaca
2341 gccaggccgc ctgctccgcc ccttcgtctt cgcatgcgct tccaagcgat caccagcaca
2401 agggcaccgc tggcacgagt acgggttgtt gaggcatgct gagagcgcct gggtcctgtc
2461 ggtggcctag gaaagggcaa aagggctgcg gggtcggccg tgagagggag agcgtggcgg
2521 agacgtgttt ctgacgaggg ctcgtgtgac gattggtgag gcctccctcg acatgcgttc
2581 acttcctgtc aggcagacgg gcaggtgtgt gggatcgatc ctctagaagc ttggaagctc
2641 tggaagggcc gcgatggggc gcgcggcgtc cagaaggcgc catacggccc gctggcggca
2701 cccatccggt ataaaagccc gcgaccccga acggtgacct ccactttcag cgacaaacga
2761 gcacttatac atacgcgact attctgccgc tatacataac cactcagcta gcttaagatc
2821 ccatcaagct tgcatgccgg gcgcgccaga aggagcgcag ccaaaccagg atgatgtttg
2881 atggggtatt tgagcacttg caacccttat ccggaagccc cctggcccac aaaggctagg
2941 cgccaatgca agcagttcgc atgcagcccc tggagcggtg ccctcctgat aaaccggcca
3001 gggggcctat gttctttact tttttacaag agaagtcact caacggatct cttgcacgta
3061 gcgttaagcc cgaaagacgt caagtacgct cgtgagccgg agaacgcgga caagaccgcc
3121 aaggtgcgcc gtcattcagc ctgttctgct gggccaggcg tattcgcaga gctgtggggc
3181 gtgaaaaatg atgcttgatg gctctaggag ctgggggtaa atgtcggcaa gtccctgtgt
3241 tgcggcggct cactagaccg ccggcatggt cgccctctcg ctccgttcat tctgacattg
3301 caggccaagg ggtctgacct gcgtgttcga attcatgagc gggtggccgg ctcggcccgc
3361 tgcctgcgag ccgctttgtc caccgtccac gagctattca gcgccaagcg tttcatactg
3421 caggactagg ttgactcccc ggcgagggcg caaatgccat cagcaaatca gatgccaagt
3481 aacgcaggta agcgcgtggc tggggcctcg tttcaggctg cgccggggca ggggcctgta
3541 gaatggtcac caaactttgc caaagcttaa atactctcct tttggttgtt attatacgga
3601 aaaggcaaga cggagcatag cccaggggtc ttgggctgtg tccgttgagt agcgtctatc
3661 gtctgtgccg cgattgtcgc cacagtgtgt tgagcgcgtg cccgcaacac ccatatacct
3721 tgcagtcgct ccagagctat cacaacgcct acaagctcat acaccagccg gctcgtaaag
3781 cggactcggg gggccctgca gacggggcag tcgacgcccc gcccgccgag cccgcgcaag
3841 cgggacaggc cggcggggag gcgaaggggc gagggcgacg ggcagctaaa cccaagggcg
3901 cagcgagccc cacaactgga gccaccggct caaatagctc aaacggcgcc aatggcagca
3961 acggcagcca cccactgggc acgtccacgg ggccagacgc actagtgtcc ttcctgacct
4021 cgcccgactt tgccggccag ggcggctcga cctcatcaac ctctggcgcc gcggccgcag
4081 cagctacagc cacacctacc accacagcta cagccagcgg caaccccgat gaccatcagc
4141 tgcgccaaca gcaacggcac cggcagaagc aacggcagca gcaacagctc cagaccgcca
4201 caggcggcga tgccggcgag ccgcactcct cctcttccgc ctccttcggt ggcggcggtg
4261 acatgcggcc ctacagtcgc cacagcgtgg cgttgccggc gtggcagccg ccgcccagca
4321 gcctgcaggc gcgcacggcg ctcacgcgcg cgatatccac ctgccccacc tacacgcggc
4381 tgcaccagct gctgctggac aacgcactcg acttcaacgt ctaccactcc tgtgccgcgc
4441 tcagccgcgt gctcgcgctg caccggcggg gcctcagccc gcgcgagtcg cggctgttca
4501 aggagggctg ctccacgctg cagagcgtcc tgcggcggca gttgacggag ctgcacccgc
4561 gcgcggtggt ggtggcggct tacagccttg cccgcctgga gctgcccgac cgcgagctgc
4621 tggccggcct ggccgccgcc gtggagccgc agctgccggc gctgcagccg cgcggcctgg
4681 cgtcgctgct gtgggccttt gcgcggcagg gccaccagcc gccgcccaag tggatggacg
4741 cattcctcag ctgctgcgcc gcggagctgc cgcgcttcgc gccacgagag gtgagcacgc
4801 tgctgtgggg cctggcgcgg ctgcactaca aggtggcgcc ggcgcggctg cggcagctgc
4861 tggagcactc gcaggcccag atgggctctt tctgcgggcg ctcgctgtcc aacgtggtct
4921 actcgctagc gctgtctcag cagcacccgg gggaggagtg gctggcggcg gcgcaggcgc
4981 gcgcggtggc cctgggtccc agcgccttct cgccgcaggg tctgacccag atggcctggg
5041 gcctggccaa gctgggctgc ccgcccacct ccgccctgct ggacatggtg tgcgcacacg

5101 cggcggcgcg gctgccgcgc agcgccgagg agcgccgccg gctgctgcag ctgcaggcgc
5161 tgcgggaccg ctccggcttc agctcgtcga gcgaagacga cgaggtcgag gcggagggcg
5221 ctgcagccgc ttcctcttcg gggagtagga gcggccgcaa gcaacagcag cagcagcagc
5281 cgcggcggcc gctggcgccc tacaacggtc tagatctgtc gacgcttatg tacgcgctgg
5341 gcagctgggg tgcgcagccg cggcccgagg tggggcggcg cctgctgctg gcgctggagt
5401 gggagctgcc tcgcctggag gccaaccagc tgtgcaactg cgtgtgggcc tgcgcgcggc
5461 tgcggctcta cccctcgagg tcctggctgc gcgacttcta cgatgcatcc taccgccagc
5521 ttccctactt caagcccgtg gacctgagcc agtcgctgtg ggcgctggcg cggctgggcg
5581 cggcgccgcc ggaggcgtgg ctgggcggcg cgctgaaccg gctgcagcac acagccagca
5641 tgttctcacc ggtggaggtt gccaacacca tgtgggcgct ggcgaagatg ggcgtgcgcg
5701 gcgagcggct gccggcggag gtgctggcgc tgttcttcat cgccacggac cgccggctca
5761 gctcgtttaa acctcaggag ctgtgctcca tggtgtgggc gctggcgcac atgcggcggc
5821 ggcccgacaa ggagtggacg gccgagttcc tcaaggttac gtaccacaag ctgggctcca
5881 tgtcgggctg gtgcctcgcc accttggcct ggtcgctggc ggagctgcag ctcagcccgc
5941 cgcccgcctg gacgtactcc ttcgtcaacg cagcccgggc gctggcggag caggcggcgc
6001 agccgccgcc accagcggcg gcggccccca cctcggcgct gcaccagccg ggggcggagc
6061 ccccgctgcg atcgctgcgc gacctctcgc cctctgcctc cgcatctcca tcttctactg
6121 ggtcggcatt aacatatcca gccctatcag tagcagcagg cgcggcggac gccgcggggt
6181 cgggttcggg gctgtcggcc atcgacctag gtcagatcat cacgggcctg cgcaagctca
6241 actctgtaca agggctggcc aaggtcgatg acttcattgg cgaggcggag gagcggctga
6301 gggcgctgga ggcgggcagc ggagcatatg cggcgcagca ggtatgggtg gggaggcggg
6361 cggttgggcg tgtgtgttgg gagggtcggc cgtagggaga gggccacgtt aacgggcaca
6421 ggtaagcggc acgggcggcc gggggctggc aaagaggcag gcggagcagg ttgcaggatt
6481 gcaggaatca ttgggactac cagactcttt caaccctaag tgctggctgg ctatccggta
6541 tgctcgcacg caggtcgggc atttcctgtc catgtccagg aagcaggcgg ggctgcggcc
6601 ggcggcgagc cagcacgagc agcagcagga ggaggggggag tcggagccgg caggggcgga
6661 tgatgatgat gagggcgcgg gcgtcgaggt gcggcgacgc gtgtccgtgt cgggtcttgc
6721 ggaccctgac ctcagcggca gcgacgccga ggagtcggcg ccgtccgcat ctggtgcggc
6781 gcagcggcgg gcgtcagcag aggcgtcgac gtccggggtg gcgcacccgc agaagcagca
6841 gcgccggcag cgcgggagac ctggtgccgc agctgctgcc gcagcgacgt ctgccggcgc
6901 cgacgtcgcg acgccttccg agtcgccggc tgcaacgaca cgtgacgcct tgacgcgggt
6961 gcttcggacc ggcgcgcgtg ttacggtgct ggcgccgcca cccgaggtgc atctggagct
7021 cggggacgtg tctgtggagc tgccggcggg cacaggcggg ccgctgtcgt tgcgtacggc
7081 gccagccgct gccagcagca gcagcagcag tggcaagggg gccgggaacg gaagtgtggg
7141 ctcagacggc gctggcaaca gcgcggctga gatcaagcag cggttgatgg cggtgtatac
7201 ataccctac gacgttccgg actatgcgta cccctacgat gtccccgact acgctagcta
7261 cccttatgat gttcctgatt atgctttaaa ctggagccac ccccagttcg agaagtgagg
7321 atcctggcag cagctggacc gcctgtacca tggagaagag ctttacttgc cgggatggcc
7381 gatttcgctg attgatacgg gatcggagct cggaggcttt cgcgctaggg gctaggcgaa
7441 gggcagtggt gaccagggtc ggtgtggggt cggcccacgg tcaattagcc acaggaggat
7501 caggggggagg taggcacgtc gacttggttt gcgaccccgc agttttggcg gacgtgctgt
7561 tgtagatgtt agcgtgtgcg tgagccagtg gccaacgtgc cacacccatt gagaagacca
7621 accaacttac tggcaatatc tgccaatgcc atactgcatg taatggccag gccatgtgag
7681 agtttgccgt gcctgcgcgc gccccgggggg cgggggggga cgggtggggg gtaggggtc
7741 tcacgggaac agcacgctag gggtcagggg gggggggggg cgcagtttag ctgaccagcc
7801 gtgggatgat gcacgcattt gcaaggacag ggtaatcaca gcagcaacat ggtgggctta
7861 ggacagctgt gggtcagtgg acggacggca ggggagggac ggcgcagctc gggagacagg
7921 gggagacagc gtgactgtgc acatcggtac ctctagagcg gccgccaccg cggtggagct
7981 ccagcttttg ttccctttag tgagggttaa ttgcgcgctt ggcgtaatca tggtcatagc
8041 tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatacga gccggaagca
8101 taaagtgtaa agcctggggt gcctaatgag tgagctaact cacattaatt gcgttgcgct
8161 cactgcccgc tttccagtcg ggaaacctgt cgtgccagct gcattaatga atcggccaac
8221 gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgctc actgactcgc
8281 tgcgctcggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggt
8341 tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg
8401 ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc ccccctgacg
8461 agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat
8521 accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc ctgccgctta
8581 ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct
8641 gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc
8701 ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa

```
    8761 gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg
    8821 taggcggtgc tacagagttc ttgaagtggt ggcctaacta cggctacact agaaggacag
    8881 tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt
    8941 gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta
    9001 cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc
    9061 agtggaacga aaactcacgt taagggattt tggtcatgag attatcaaaa aggatcttca
    9121 cctagatcct tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa
    9181 cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg atctgtctat
    9241 ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct
    9301 taccatctgg ccccagtgct gcaatgatac cgcgagaccc acgctcaccg gctccagatt
    9361 tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggtcct gcaactttat
    9421 ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt tcgccagtta
    9481 atagtttgcg caacgttgtt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg
    9541 gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga tcccccatgt
    9601 tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt tgtcagaagt aagttggccg
    9661 cagtgttatc actcatggtt atggcagcac tgcataattc tcttactgtc atgccatccg
    9721 taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc
    9781 ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca catagcagaa
    9841 ctttaaaagt gctcatcatt ggaaaacgtt cttcggggcg aaaactctca aggatcttac
    9901 cgctgttgag atccagttcg atgtaaccca ctcgtgcacc caactgatct tcagcatctt
    9961 ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg
   10021 gaataagggc gacacggaaa tgttgaatac tcatactctt cctttttcaa tattattgaa
   10081 gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata
   10141 aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc ac
//
```



Two fragments possible: ***Xho*I-*Bgl*II (535 bp) T_6, A_7, A_8, G_9**
***Bgl*II-*Age*I 342 bp C_9, G_10**

>WT

CTCGAGAGGTGAGCACGCTGCTGTGGGGCCTGGCGCGGCTGCACTACAAGGTGGCGCCGG
CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGCGCT
CGCTGTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC
TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGCAGGGTC
TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG
ACATGGTGTGCGCACACGCGGCGGCGCGGCTGCCGCGCAGCGCCGAGGAGCGCCGCCGGC
TGCTGCAGCTGCAGGCGCTGCGGGACCGCTCCGGCTTCAGCTCGTCGAGCGAAGACGACG
AGGTCGAGGCGGAGGGCGCTGCAGCCGCTTCCTCTTCGGGGAGTAGGAGCGGCCGCAAGC
AACAGCAGCAGCAGCAGCCGCGGCGGCCGCTGGCGCCCTACAACGGTCT**AGATCT**

>AA→GG1

CTCGAGAGGTGAGCACGCTGCTGTGGGGCCTGGCGCGGCTGCACTACAAGGTGGCGCCGG
CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGCAGG
ACCTCTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC
TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCC**T**CAGGA**T**G
TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG
ACATGGTGTGCGCACACGCGGCGGCGCGGCTGCCGCGCAGCGCCGAGGAGCGCCGCCGGC
TGCTGCAGCTGCAGGCGCTGCGGGACCGCTCCGGCTTCAGCTCGTCGAGCGAAGACGACG
AGGTCGAGGCGGAGGGCGCTGCAGCCGCTTCCTCTTCGGGGAGTAGGAGCGGCCGCAAGC
AACAGCAGCAGCAGCAGCCGCGGCGGCCGCTGGCGCCCTACAACGGTCT**AGATCT**
Creation of PpuMI and Bsu36I

>AA→TT1

CTCGAGAGGTGAGCACGCTGCTGTGGGGCCTGGCGCGGCTGCACTACAAGGTGGCGCCGG
CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGCGCG
AGCT**C**TCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC
TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGCAGGAGC
TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG
ACATGGTGTGCGCACACGCGGCGGCGCGGCTGCCGCGCAGCGCCGAGGAGCGCCGCCGGC
TGCTGCAGCTGCAGGCGCTGCGGGACCGCTCCGGCTTCAGCTCGTCGAGCGAAGACGACG
AGGTCGAGGCGGAGGGCGCTGCAGCCGCTTCCTCTTCGGGGAGTAGGAGCGGCCGCAAGC
AACAGCAGCAGCAGCAGCCGCGGCGGCCGCTGGCGCCCTACAACGGTCT**AGATCT**
Loss of BanII and creation of SacI

>AA→CC1

CTCGAGAGGTGAGCACGCTGCTGTGGGGCCTGGCGCGGCTGCACTACAAGGTGGCGCCGG
CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGAACC
AGCTGTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC
TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGACGCAGC
TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG
ACATGGTGTGCGCACACGCGGCGGCGCGGCTGCCGCGCAGCGCCGAGGAGCGCCGCCGGC
TGCTGCAGCTGCAGGCGCTGCGGGACCGCTCCGGCTTCAGCTCGTCGAGCGAAGACGACG
AGGTCGAGGCGGAGGGCGCTGCAGCCGCTTCCTCTTCGGGGAGTAGGAGCGGCCGCAAGC
AACAGCAGCAGCAGCAGCCGCGGCGGCCGCTGGCGCCCTACAACGGTCT**AGATCT**
No RFLP marker; creation of a PvuII site

>AG→CT1

CTCGAGAGGTGAGCACGCTGCTGTGGGGCCTGGCGCGGCTGCACTACAAGGTGGCGCCGG
CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGCGCT
CGCTGTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC
TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGACGCAGC
TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG
ACATGGTGTGCGCACACGCGGCGGCGCGGCTGCCGCGCAGCGCCGAGGAGCGCCGCCGGC
TGCTGCAGCTGCAGGCGCTGCGGGACCGCTCCGGCTTCAGCTCGTCGAGCGAAGACGACG
AGGTCGAGGCGGAGGGCGCTGCAGCCGCTTCCTCTTCGGGGAGTAGGAGCGGCCGCAAGC
AACAGCAGCAGCAGCAGCCGCGGCGGCCGCTGGCGCCCTACAACGGTC**GAGATCT**
Loss of XbaI

WT

CTCGAGAGGTGAGCACGCTGCTGTGGGGCCTGGCGCGGCTGCACTACAAGGTGGCGCCGG
R E V S T L L W G L A R L H Y K V A P A

CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGCGCT
R L R Q L L E H S Q A Q M G S F C G R S

CGCTGTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC
L S N V V Y S L A L S Q Q H P G E E W L

TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGCAGGGTC
A A A Q A R A V A L G P S A F S P Q G L

TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG
T Q M A W G L A K L G C P P T S A L L D

ACATGGTGTGCGCACACGCGGCGGCGCGGCTGCCGCGCAGCGCCGAGGAGCGCCGCCGGC
M V C A H A A A R L P R S A E E R R R L

TGCTGCAGCTGCAGGCGCTGCGGGACCGCTCCGGCTTCAGCTCGTCGAGCGAAGACGACG
L Q L Q A L R D R S G F S S S S E D D E

AGGTCGAGGCGGAGGGCGCTGCAGCCGCTTCCTCTTCGGGGAGTAGGAGCGGCCGCAAGC
V E A E G A A A A S S S G S R S G R K Q

AACAGCAGCAGCAGCAGCCGCGGCGGCCGCTGGCGCCCTACAACGGTCT**AGATCT**
Q Q Q Q Q P R R P L A P Y N G L D

AA → GG1

CTCGAGAGGTGAGCACGCTGCTGTGGGGCCTGGCGCGGCTGCACTACAAGGTGGCGCCGG
R E V S T L L W G L A R L H Y K V A P A

CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGCGC~~T~~
R L R Q L L E H S Q A Q M G S F C G Q D
CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGCAGG

~~CG~~CT~~G~~TCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC
L S N V V Y S L A L S Q Q H P G E E W L
ACCTCTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC

TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGCAG~~CGT~~C
A A A Q A R A V A L G P S A F S P Q D L
TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCC**T**CAGGA**T**G

TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG
T Q M A W G L A K L G C P P T S A L L D
TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG

ACATGGTGTGCGCACACGCGGCGGCGCGGCTGCCGCGCAGCGCCGAGGAGCGCCGCCGGC
M V C A H A A A R L P R S A E E R R R L

TGCTGCAGCTGCAGGCGCTGCGGGACCGCTCCGGCTTCAGCTCGTCGAGCGAAGACGACG
L Q L Q A L R D R S G F S S S S E D D E

AGGTCGAGGCGGAGGGCGCTGCAGCCGCTTCCTCTTCGGGGAGTAGGAGCGGCCGCAAGC
V E A E G A A A A S S S G S R S G R K Q

AACAGCAGCAGCAGCAGCCGCGGCGGCCGCTGGCGCCCTACAACGGTCT**AGATCT**
Q Q Q Q Q P R R P L A P Y N G L D

AA → TT1

CTCGAGAGGTGAGCACGCTGCTGTGGGGCCTGGCGCGGCTGCACTACAAGGTGGCGCCGG
R E V S T L L W G L A R L H Y K V A P A

CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGCGC~~T~~
R L R Q L L E H S Q A Q M G S F C G R E
CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGCGCG

~~CG~~CTGTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC
L S N V V Y S L A L S Q Q H P G E E W L
AGCTCTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC

TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGCAGGGTC
A A A Q A R A V A L G P S A F S P Q E L
TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGCAGGAGC

TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG
T Q M A W G L A K L G C P P T S A L L D
TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG

ACATGGTGTGCGCACACGCGGCGGCGCGGCTGCCGCGCAGCGCCGAGGAGCGCCGCCGGC
M V C A H A A A R L P R S A E E R R R L

TGCTGCAGCTGCAGGCGCTGCGGGACCGCTCCGGCTTCAGCTCGTCGAGCGAAGACGACG

L Q L Q A L R D R S G F S S S S E D D E

AGGTCGAGGCGGAGGGCGCTGCAGCCGCTTCCTCTTCGGGGAGTAGGAGCGGCCGCAAGC
V E A E G A A A A S S S G S R S G R K Q

AACAGCAGCAGCAGCAGCCGCGGCGGCCGCTGGCGCCCTACAACGGTCT**AGATCT**
Q Q Q Q Q P R R P L A P Y N G L D

AA → CC1

CTCGAGAGGTGAGCACGCTGCTGTGGGGCCTGGCGCGGCTGCACTACAAGGTGGCGCCGG
R E V S T L L W G L A R L H Y K V A P A

CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGG~~CGCT~~
R L R Q L L E H S Q A Q M G S F C G N Q
CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGAACC

~~CG~~CTGTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC
L S N V V Y S L A L S Q Q H P G E E W L
AGCTGTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC

TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGCAGGGTC
A A A Q A R A V A L G P S A F S P T Q L
TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGACGCAGC

TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG
T Q M A W G L A K L G C P P T S A L L D
TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG

ACATGGTGTGCGCACACGCGGCGGCGCGGCTGCCGCGCAGCGCCGAGGAGCGCCGCCGGC
M V C A H A A A A R L P R S A E E R R R L

TGCTGCAGCTGCAGGCGCTGCGGGACCGCTCCGGCTTCAGCTCGTCGAGCGAAGACGACG
L Q L Q A L R D R S G F S S S S E D D E

AGGTCGAGGCGGAGGGCGCTGCAGCCGCTTCCTCTTCGGGGAGTAGGAGCGGCCGCAAGC
V E A E G A A A A S S S G S R S G R K Q

AACAGCAGCAGCAGCAGCCGCGGCGGCCGCTGGCGCCCTACAACGGTCT**AGATCT**
Q Q Q Q Q P R R P L A P Y N G L D

AG → CT1

CTCGAGAGGTGAGCACGCTGCTGTGGGGCCTGGCGCGGCTGCACTACAAGGTGGCGCCGG
R E V S T L L W G L A R L H Y K V A P A

CGCGGCTGCGGCAGCTGCTGGAGCACTCGCAGGCCCAGATGGGCTCTTTCTGCGGGCGCT
R L R Q L L E H S Q A Q M G S F C G R S

~~CG~~CTGTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC
L S N V V Y S L A L S Q Q H P G E E W L
AGCTGTCCAACGTGGTCTACTCGCTCGCGCTGTCTCAGCAGCACCCGGGGGAGGAGTGGC

TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGCAGGGTC
A A A Q A R A V A L G P S A F S P T Q L
TGGCGGCGGCGCAGGCGCGCGCGGTGGCCCTCGGTCCCAGCGCCTTCTCGCCGACGCAGC

TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG
T Q M A W G L A K L G C P P T S A L L D
TGACCCAGATGGCCTGGGGCCTGGCCAAGCTGGGCTGCCCGCCCACCTCCGCCCTGCTGG

ACATGGTGTGCGCACACGCGGCGGCGCGGCTGCCGCGCAGCGCCGAGGAGCGCCGCCGGC
M V C A H A A A A R L P R S A E E R R R L

TGCTGCAGCTGCAGGCGCTGCGGGACCGCTCCGGCTTCAGCTCGTCGAGCGAAGACGACG
L Q L Q A L R D R S G F S S S S E D D E

AGGTCGAGGCGGAGGGCGCTGCAGCCGCTTCCTCTTCGGGGAGTAGGAGCGGCCGCAAGC
V E A E G A A A A S S S G S R S G R K Q

AACAGCAGCAGCAGCAGCCGCGGCGGCCGCTGGCGCCCTACAACGGTCT**AGATCT**
Q Q Q Q Q P R R P L A P Y N G R D
AACAGCAGCAGCAGCAGCCGCGGCGGCCGCTGGCGCCCTACAACGGTCG**AGATCT**

GG2

ATGCATCCTACCGCCAGCTTCCCTACTTCAAGCCCGTGGACCTGAGCCAGTCGCTGTGGG
CGCTGGCGCGGCTGGGCGCGGCGCCGCCGGAGGCGTGGCTGGGCGGCGCGCTGAACCGGC
TGCAGCACACAGCCAGCATGTTCTCACCGGTG~~GA~~**C**~~GT~~**C**GCCAACACCATGTGGGCGCTGG
CGAAGATGGGCGTGCGCGGCGAGCGGCTGCCGGCGGAGGTGCTGGCGCTGTTCTTCATCG
CCACGGACCGCCGGCTCAGCTCGTTTAAACCTCAGGA**C**CTGTGCTCCATGGTGTGGGCGC
TGGCGCACATGCGGCGGCGGCCCGACAAGGAGTGGACGGCCGAGTTCCTCAAGGT***TACGTA***

Creation of AatII, EcoO109I, PpuMI, BsaHI sites

AA2

ATGCATCCTACCGCCAGCTTCCCTACTTCAAGCCCGTGGACCTGAGCCAGTCGCTGTGGG
CGCTGGCGCGGCTGGGCGCGGCGCCGCCGGAGGCGTGGCTGGGCGGCGCGCTGAACCGGC
TGCAGCACACAGCCAGCATGTTCTCACCG**CA**GG**GN**GTTGCCAACACCATGTGGGCGCTGG
CGAAGATGGGCGTGCGCGGCGAGCGGCTGCCGGCGGAGGTGCTGGCGCTGTTCTTCATCG
CCACGGACCGCCGGCTCAGCTCGTTTAAA~~CC~~**C**~~CAGG~~**GC**CTGTGCTCCATGGTGTGGGCGC
TGGCGCACATGCGGCGGCGGCCCGACAAGGAGTGGACGGCCGAGTTCCTCAAGGT***TACGTA***

Loss of AgeI, and Bsu36I Creation of PasI and EcoO109I sites

CC2

ATGCATCCTACCGCCAGCTTCCCTACTTCAAGCCCGTGGACCTGAGCCAGTCGCTGTGGG
CGCTGGCGCGGCTGGGCGCGGCGCCGCCGGAGGCGTGGCTGGGCGGCGCGCTGAACCGGC
TGCAGCACACAGCCAGCATGTTCTCACCG**ACCC**AGGTTGCCAACACCATGTGGGCGCTGG
CGAAGATGGGCGTGCGCGGCGAGCGGCTGCCGGCGGAGGTGCTGGCGCTGTTCTTCATCG
CCACGGACCGCCGGCTCAGCTCGTTTAAACCT**ACGC**AGCTGTGCTCCATGGTGTGGGCGC
TGGCGCACATGCGGCGGCGGCCCGACAAGGAGTGGACGGCCGAGTTCCTCAAGGT***TACGTA***

Loss of AgeI, and of one AlwNI of 2 and of a Bsu36I site

CC2

ATGCATCCTACCGCCAGCTTCCCTACTTCAAGCCCGTGGACCTGAGCCAGTCGCTGTGGG

CGCTGGCGCGGCTGGGCGCGGCGCCGCCGGAGGCGTGGCTGGGCGGCGCGCTGAACCGGC

TGCAGCACACAGCCAGCATGTTCTCACCG**ACNC**AGGTTGCCAACACCATGTGGGCGCTGG CC
L Q H T A S M F S P **T** **Q** V A N T M W A L
CGAAGATGGGCGTGCGCGGCGAGCGGCTGCCGGCGGAGGTGCTGGCGCTGTTCTTCATCG

CCACGGACCGCCGGCTCAGCTCGTTTAAACCT**ACNC**AGCTGTGCTCCATGGTGTGGGCGC CC
A T D R R L S S F K P T Q L C S M V W A

TGGCGCACATGCGGCGGCGGCCCGACAAGGAGTGGACGGCCGAGTTCCTCAAGGT***TACGTA***

AA2

ATGCATCCTACCGCCAGCTTCCCTACTTCAAGCCCGTGGACCTGAGCCAGTCGCTGTGGG
CGCTGGCGCGGCTGGGCGCGGCGCCGCCGGAGGCGTGGCTGGGCGGCGCGCTGAACCGGC

TGCAGCACACAGCCAGCATGTTCTCACCG**CA**GG**GN**GTTGCCAACACCATGTGGGCGCTGG AA
L Q H T A S M F S P **Q** **G** V A N T M W A L

CGAAGATGGGCGTGCGCGGCGAGCGGCTGCCGGCGGAGGTGCTGGCGCTGTTCTTCATCG

CCACGGACCGCCGGCTCAGCTCGTTTAAA~~CC~~**C**~~CAGG~~**GC**CTGTGCTCCATGGTGTGGGCGC AA
A T D R R L S S F K P Q **G** L C S M V W A

TGGCGCACATGCGGCGGCGGCCCGACAAGGAGTGGACGGCCGAGTTCCTCAAGGT***TACGTA***

GG2

ATGCATCCTACCGCCAGCTTCCCTACTTCAAGCCCGTGGACCTGAGCCAGTCGCTGTGGG
CGCTGGCGCGGCTGGGCGCGGCGCCGCCGGAGGCGTGGCTGGGCGGCGCGCTGAACCGGC

TGCAGCACACAGCCAGCATGTTCTCACCGGTG~~GA~~**C**~~GT~~**C**GCCAACACCATGTGGGCGCTGG GG
L Q H T A S M F S P V D V A N T M W A L

CGAAGATGGGCGTGCGCGGCGAGCGGCTGCCGGCGGAGGTGCTGGCGCTGTTCTTCATCG

CCACGGACCGCCGGCTCAGCTCGTTTAAACCTCAGGA**C**CTGTGCTCCATGGTGTGGGCGC GG
A T D R R L S S F K P Q **D** L C S M V W A

TGGCGCACATGCGGCGGCGGCCCGACAAGGAGTGGACGGCCGAGTTCCTCAAGGT***TACGTA***

ARTICLE 1

"DETERMINANTS FOR 5' PROCESSING AND STABILITY OF THE CHLOROPLAST ATPB MRNA IN CHLAMYDOMONAS."

This is a draft version of an article devoted to the study of the role of MDB1 in the stabilisation and maturation of the atpB mRNA. Some experiments should be performed again (e.g. The B Northern fig 6A), and complementary experiments have still to be done.

Determinants of 5' processing and stability of the chloroplast *atpB* mRNA in *Chlamydomonas*.

Marina Cavaiuolo, Stefania Viola, Domitille Jarrige, Blandine Rimbault[1], Dominique Drapier, Katia Wostrikoff, Francis-André Wollman and Yves Choquet

UMR 7141, CNRS/UPMC; Institut de Biologie Physico-Chimique, 13, rue Pierre et Marie Curie, F-75005 Paris, France.

[1]: present address:

*Corresponding author: Y. Choquet E-mail: choquet@ibpc.fr

Running title: *atpB* mRNA 5' processing in Chlamydomonas

Character count:

Keywords: RNA-RNA interaction/ mRNA processing/ mRNA stability / 5'-3' UTR interactions / chloroplast/ *Chlamydomonas reinhardtii*

Abstract (~200 words)

In chloroplasts, every post-transcriptional step of gene expression, from maturation to translation and degradation of mRNAs, is controlled by a combination of *cis*-acting RNA elements, among which RNA secondary structures, and gene-specific *trans*-acting factors. Here, we report the characterization of MDB1, a nucleus-encoded OctotricoPetpide Repeat (OPR) protein required for the stabilization of the chloroplast *atpB* mRNA in *Chlamydomonas*. MDB1 binds the 5'end of the mature *atpB* mRNA as revealed by primer extension, cRT-PCR and small RNA sequencing. The *atpB* gene is transcribed as a tri-phosphorylated precursor, whose decay is initiated by the conversion of the 5' triphosphate to a monophosphate, thereby becoming a substrate for 5' → 3' exonucleolytic degradation, blocked, in the wild-type, by MDB1. We show that interactions between the 5' and 3' UTRs of the mRNA control the correct processing of chimeric transcripts driven by the *atpB* 5'UTR. ?We propose a model for *atpB* gene expression that involves the concerted action of MDB1 and interactions between 5' and 3' UTRs.?

Introduction

Chloroplasts originated through endosymbiosis from a cyanobacterial ancestor (Keeling, 2010). Modern chloroplasts only retained from their ancestor a limited set of genes (Maul et al., 2002), whose expression is governed by nucleus-encoded Organelle Trans-Acting factors (OTAFs). These factors, identified by genetic screens of photosynthetic mutants, interact with *cis*-acting RNA sequences present in the 5'- and/or 3'-UTRs of their chloroplast targets to control their maturation, translation and decay (for reviews: (Barkan and Goldschmidt-Clermont, 2000; Choquet and Wollman, 2002; Woodson and Chory, 2008; Barkan and Small, 2014). In the model unicellular green alga *Chlamydomonas reinhardtii*, M factors (for maturation/stability) govern the stable accumulation of their target mRNAs while T factors (for translation) control their translation. According to an emerging consensus, M factors bind the 5'UTR of their target mRNA, where they act as a barriers against 5' → 3' exoribonucleases, thereby stabilising the transcripts and determining mature 5'ends (Drager et al., 1998; Vaistij et al., 2000; Pfalz et al., 2009; Yoon, 2009; Loizeau et al., 2014), reviewed in (Barkan, 2011; Barkan and Small, 2014). Most OTAFs belong to helical repeat protein families -e.g. the TPR, PPR and OPR (Tetra-, Penta- and Octo-tricoPeptide Repeat) protein families- respectively defined by the presence of tandem repeats of a degenerate motif of 34, 35 and 38 residues, each repeat interacting with a specific nucleotide, thus allowing a sequence-specific recognition of the RNA target (reviewed in: (Barkan and Small, 2014; Hammani et al., 2014).

Chimeric genes expressed in the chloroplast genome, made of 5' or 3' UTRs fused to reporter coding sequences have been instrumental to identify the target and the mode of action of these OTAFS. However, in *Chlamydomonas* chloroplasts, reporter genes did not always fully mimic the accumulation and translation patterns of the endogenous mRNA, whose 5'UTR they borrow (Ishikura et al., 1999; Kasai et al., 2003; Minai et al., 2006) others rbcL): regulatory elements may also reside in coding regions and/or 3' UTRs. Regulation of chloroplast gene expression would therefore rely on several factors and sequence elements, the interaction of which, although possibly pivotal for understanding chloroplast gene expression, is still poorly known yet.

While nuclear-encoded mRNAs harbour a 5′ cap (a 7-methylguanylate connected to the mRNA by a triphosphate) that protects them from 5'->3' exonucleolytic degradation (Furuichi et

al., 1977; Hsu and Stevens, 1993), prokaryote and organelle mRNAs do not: primary transcription products only harbour a tri-phosphorylated nucleotide at their 5'end. Moreover, at variance with bacterial mRNAs, most chloroplast RNAs undergo intercistronic cleavage and 5'end processing that release shorter mono-phosphorylated mature transcripts.

Secondary structures in the 3'UTRs of chloroplast mRNAs also are important determinants for maturation and stability (Germain et al., 2013). 3' stem-loops protect mRNAs against 3' → 5' exoribonucleases and determine the position of mature 3'ends (Stern and Gruissem, 1987; Stern et al., 1989; Stern et al., 1991; Drager et al., 1996): deletion of the inverted repeat in the Chlamydomonas *atpB* 3'UTR led to drastically reduced amounts of *atpB* transcripts that became heterogeneous in size, and to a ~60% decreased protein level (Stern et al., 1991). While 3' UTRs may also contain regulatory elements since the proper processing of *atpB* mRNA 3'end stimulates its translation (Rott et al., 1998), they have long been thought as constitutive cis elements: their thermodynamic stability certainly contributes to set the accumulation level of a transcript, but OTAFs specifically targeting the 3'UTR of a given gene have not been identified so far in Chlamydomonas, in contrast to the numerous identified OTAFs targeting the 5'UTR of chloroplast genes.

In a previous study we described a nuclear mutant of Chlamydomonas lacking accumulation of the *atpB* mRNA, encoding the chloroplast ATP synthase subunit β (Drapier et al., 1992). Here, we identified the mutated gene. While studying its role in the expression of the *atpB* mRNA, we observed that interactions between 5' and 3' UTRs affect the 5' end processing of the *atpB* transcripts.

Results

Accumulation of the chloroplast *atpB* mRNA is controlled by the OPR MDB1 protein

The non-phototrophic *thm24* mutant lacks accumulation of the *atpB* mRNA (Drapier et al., 1992). More recently, we isolated, in an insertional mutagenesis campaign with an *aphVIII* cassette (Houille-Vernes et al., 2011), a second mutant, *L35a*, displaying the same phenotype. In 10 parental di-type tetrads from a back-cross of this latter mutant with the wild type, the mutant phenotype segregated with the resistance to paromomycin suggesting a tight linkage of the mutation with the locus of *aphVIII* insertion. Whole genome sequencing of the *L35a* and *thm24* strains revealed that, in the *L35a* strain, the insertion of the cassette in chromosome 14 was associated with a 30 kb deletion (from position 1029641 to position 1060055), encompassing 7 genes models, out of which one, Cre14.g614550, encodes an OPR protein (Fig. 1A). In the *thm24* mutant strain, this region was retained but a 1 bp deletion in gene model Cre14.g614550 leads to premature translation abortion after codon 406 (Suppl. Fig. S1). That we actually identified the gene responsible for the phenotype of the two strains was confirmed as we could complement both mutants either with cosmids encompassing the deleted region or with a minigene version of Cre14.g614550 (Suppl. Fig. S2, see M&M for details on the midigene). Cre14.g614550 was thus renamed *MDB1*, according to the gene nomenclature for Chlamydomonas OTAFs, for **M**aturation/stability of complex **D** (ATP synthase) *atp**B*** mRNA. The *thm24* and *L35a* strains will be hereafter referred to as *mdb1-1* and *mdb1-2*, respectively. The *MDB1* gene contains 11 exons and encodes a protein of 1137 residues (Fig. 1A), predicted to be targeted to the chloroplast by the Predotar (Small et al., 2004), Predalgo (Tardif et al., 2012) and TargetP (Emanuelsson et al., 2001) softwares. Analysis with the FTrep program (Rahire et al., 2012) identified 13 OPR repeats within the MDB1 protein (Suppl. Fig. S3).

MDB1 binds the mature 5'end of the *atpB* mRNA to protect it from 5' → 3' exonucleases.

Because most OTAFs act on the 5'UTR of their target mRNA, we tested whether MDB1 interacts genetically with the 360 bp long *atpB* 5'UTR (Woessner et al., 1986). To this aim, we used two strains, BFFF (named *dBf* in (Drapier et al., 2007)) and BKR (named ATG12 in (Rimbault et al., 2000)), that respectively express the *5'atpB-petA-3'petA* (*BFFF*) and *5'atpB-*

aadA-3'rbcL (*BKR*) chimeras (Fig. 2A; see Table I for the description of the chimeras used in this work), in which the *petA* coding sequence and 3'UTR or the *aadA* coding sequence fused to the *rbcL* 3'UTR are respectively expressed under the control of the *atpB* promoter and 5'UTR. After transformation in the chloroplast genome of *C. reinhardtii*, the *BFFF* chimera replaced the endogenous *petA* gene, while the *BKR* chimera was inserted downstream of the *petA* gene (Fig. 2A). These strains were crossed with the *mdb1-1, mt-* mutant. In tetrad progeny from both crosses, all members inherited the chimeric genes uniparentally transmitted by the *mt+* parent. The two members that inherited the wild-type *MDB1* allele, as they accumulated the *atpB* mRNA, also accumulated the chimeric transcripts. The two other members inherited the *mdb1-1* allele: they failed to accumulate both the *atpB* mRNA and the chimeric transcripts (Fig. 2B). Thus, the *atpB* 5' UTR is sufficient to confer an MDB1-dependent stability to a downstream coding sequence.

To assess whether the MDB1 factor protein is physically associated *in vivo* with its genetically identified RNA target, we generated by complementation of the *mdb1-1* mutant, a strain expressing an HA-tagged version of MDB1 (MDB1-HA; see M & M for details), for immuno-detection or -precipitation with antibodies directed against the HA tag. The tag was inserted in a region of the protein poorly conserved among Chlamydomonadales species (Suppl. Fig S4). After immuno-precipitation of this tagged version of MDB1, RNAs, extracted from the pellet, were analysed by dot-blots, using probes against the *atpB* (and *atpA* as a negative control) 5'UTRs (Fig. 2C). A specific signal, observed in the complemented strain, but not in the wild type, with the *atpB* probe, but not with the *atpA* probe, indicated that the tagged MDB1-HA protein indeed interacts specifically, directly or indirectly, with the *atpB* 5'UTR *in vivo***.**

The stable interaction between a M factor and its target mRNA often leads to the accumulation of a footprint, a cluster with a sharp 5'end of small RNAs protected from degradation by the bound protein (Ruwe and Schmitz-Linneweber, 2012; Zhelyazkova et al., 2012; Cavaiuolo et al 2017). These footprints usually map at the 5' ends of the mature chloroplast mRNAs, thus pinpointing the binding sites of M factors. A small RNA of 20 bp mapping to the mature 5'end of the *atpB* mRNA, previously identified in sRNA-seq datasets (Cavaiuolo et al 2017), likely corresponds to the footprint of MDB1. Indeed, its accumulation was drastically reduced in the *mdb1-2* mutant, compared to the wild type (<1% of WT level: Fig. 3). By contrast, sRNAs generated from other regions of the *atpB* mRNA, which correspond to degradation

products of the 5' destabilized transcript, were less severely affected. Thus, MDB1, alone or associated with other proteins, interacts with the mature 5' end of its target mRNA.

Poly(G) cages form highly stable tertiary structures that impede the progression of exoribonucleases and have been instrumental to show that, in *Chlamydomonas* chloroplasts, M factors protect their target transcript from 5' → 3' exonucleases (Drager et al., 1998; Drager et al., 1999; Vaistij et al., 2000; Yoon, 2009). We thus inserted 18 consecutive G residues in the *atpB* 5'UTR at position -32 relative to the start codon (Fig. 4A). This modified *atpB* gene, associated with a spectinomycin resistance cassette (Goldschmidt-Clermont, 1991) for the selection of transformants, replaced, after transformation, the endogenous *atpB* gene of the *mdb1-1* strain. The resulting transformants accumulated a shorter version of the transcript (fig. 4B). We mapped its 5' end by sRNA-Seq and found a peak of sRNAs overlapping the polyG cage (fig. 4C).

One of these transformants was then crossed to our reference strain S24 mt⁻. Zygotes were germinated on TAP medium and hatched progeny spread on TAP plates. 5 progeny, randomly picked up from the plates, all contained the polyG and were resistant to spectinomycin, as expected from the uniparental inheritance of the chloroplast genome of the *mt+* parent. Two of them were similar to their *mdb1-1* {pG*atpB*} parent. The other three progeny accumulated the short *atpB* transcript and, in addition, a transcript slightly larger than the endogenous *atpB* mRNA due to the insertion of 37 bp in the *atpB* 5'UTR. They inherited the wild-type *MDB1* allele, as confirmed by the restoration of the MDB1 footprint, but also show a peak overlapping the polyG tract. They were nonetheless unable of phototrophic growth probably because the insertion of the polyG tract 32 nt upstream of the initiation codon impedes the translation of the *pG-atpB* mRNA. The shorter transcript, present in both wild-type and mutant background, was 1.5 fold more abundant than the endogenous *atpB* mRNA in the wild type or than the full-length *pG-atpB* in wild-type progeny, suggesting that only a fraction of the transcribed *atpB* mRNA is actually stabilized by MDB1, while most *atpB* transcription products are rapidly degraded.

We thus concluded that the *atpB* mRNA is degraded by 5' → 3' exonucleases, unless protected by the bound MDB1 factor, present in limiting amount.

MDB1 is required for the processing of the *atpB* transcript

Chloroplast primary transcription products are often 5' or 3' processed to yield shorter mature mRNAs. The *atpB* mRNA, as many other chloroplast transcripts, harbour two 5' ends, a minor one, tri-phosphorylated and mapping to the transcription start site (TSS: +1), and a major one, trimmed up to position +27 (in the following, positions will be given with respect to the transcription start site, unless otherwise specified) and mono-phosphorylated (Blowers et al, 1990; Woessner et al, 1986; Anthoninson et al, 2001; Cavaiuolo et al, 2017). The *atpB* 3'end was mapped at the end of an inverted repeat (Stern et al, 1991). Because M factors often stabilise the processed form of their target transcript, we mapped the *atpB* mRNA 5' end(s) by primer extension and circular RT-PCR (cRT-PCR) in the wild-type and *mdb1-1* strains (Fig. 5).

Primer extension experiments revealed both previously mapped 5' ends of the precursor (+1) and mature (+27) transcripts in the wild type, the mature form being by far the most abundant (Fig. 5A). By contrast, only the precursor transcript was detected in the *mdb1-1* mutant, indicating that the MDB1 protein is required for the processing and/or the stabilization of the mature transcript, as confirmed by S1 protection assay (Suppl. Fig S4). Despite a twice higher input of mutant vs. wild-type RNA, the precursor band remained of lower intensity, suggesting that MDB1 also stabilises somehow the precursor transcript.

Primer extension provides an estimation of the relative amount of RNA species, but not of their 5' phosphorylation state, nor of their length, downstream of the primer used for reverse transcription. By contrast, cRT-PCR, although not quantitative, discriminates between mono- and tri phosphorylated 5'ends, since only mono-phosphorylated mRNAs can self-ligate and circularize. Tri-phosphorylated transcripts cannot, unless first converted to a mono-phosphorylated form by treatment with 5'RNA polyphosphatase (RPP). cRT-PCR also allows the simultaneous determination of the 5' and 3'ends of individual RNA molecules, whether originating from intact or degraded transcripts. We used a specific primer on *atpB* CDS (+1196/TSS; see Suppl. Table ST1 for the oligonucleotides used in this study) for retro-transcription of circularized mRNAs from wild-type and *mdb1-1* strains, either mock- or RPP-treated, and the resulting cDNAs were amplified with outward-directed primers towards the 5' and 3' termini as diagrammed in Fig 5B. Because the primary and mature 5'ends of *atpB* mRNA only differ by 27 nt, cRT-PCR products are not easily sized-discriminated on a gel. Indeed, PCR yielded products of similar size (~650-bp) in both RPP- and mock-treated samples (Fig 5B).

In the *mdb1-1* mutant, where only the primary transcript should be present, no amplification was expected from the mock-treated sample. Surprisingly, a PCR product was amplified from the mock-treated sample, although of low abundance. Sequencing of gel-purified amplicons revealed only primary *atpB* 5'ends (+1), even in the mock-treated sample. Thus the *atpB* precursor is present as two mRNA species that differ by the phosphorylation state of the same *5′* end. The presence in the *mdb1-1* mutant of a mono-phosphorylated precursor mRNA, in addition to the tri-phosphorylated transcription product, suggests that 5' triphosphate removal occurs in the chloroplast, maybe as a prerequisite for further maturation in the wild type, or, in the absence of MDB1, for degradation.

Sequences of the wild-type amplicons identified a processed 5'end (+27) in the mock sample. Unexpectedly, we only detected a primary 5'end (+1) in the RPP sample. Indeed, sequencing of cRT-PCR products mostly detects the shortest ends of a transcript: downstream of the processed 5'end, the sequence of the precursor should be superposed, and, based on the relative abundance of the precursor vs. processed transcripts detected in primer extension analysis, overwhelmed by that of the *atpB* 3'UTR. Would this later be heterogeneous, the sequence may become unreadable. Nonetheless, the sequence of the precursor *atpB* mRNA was unambiguously red (Suppl. Fig S5).

To understand the origin of this apparent discrepancy, we repeated the cRT-PCR experiment on the wild-type samples, using that time a forward primer (atpBFW2: +749/TSS) located upstream of the primer used for reverse transcription. In that way, the reverse transcriptase has to amplify the circularised mRNA more than once before a PCR product can be amplified (Suppl. Fig. S6A for a schematic representation), allowing all degradation products of the *atpB* mRNA ending downstream of the primer used for retro-transcription (TSS + 1196) to contribute to PCR amplification. By contrast, in the previous PCR, only the degradation products ending downstream of the forward primer, i.e. after position TSS + 1656, contributed to the 650 bp amplicon. Amplicons were separated on a gel and those extracted from a broad band of the RPP-treated sample (Suppl. Fig. S6B) were cloned and sequenced. 11 clones, out of 20 (Suppl. Fig. S6C), displayed a precursor 5'end, a proportion again much higher than expected from the relative amount of the precursor and processed transcripts in primer extension experiments, but incompatible with the detection of only the precursor form in our first cRT-PCR experiment. The other 9 mapped to position +27 (the processed mRNA). Surprisingly, the 3' end of only 6 clones

mapped close to the previously published 3'end (Stern et al, 1991): all but 1 show a precursor 5'end. The other 14 clones ended within the *atpB* CDS, two of which being also poly-adenylated, which suggested that they are degradation products of the *atpB* mRNA. Most importantly, 13 of these truncated clones mapped upstream of the 3' primer used in our first cRT-PCR experiment (fig. 6) and would not have contributed to the sequenced 650 bp amplicon in our first cRT-PCR experiment. Among the 7 others, 6 had their 5'end at position +1 and only one originated from a processed mRNA, which explains why only the precursor transcript was found in our first cRT-PCR experiment. The 20 clones, however, and also others more severely truncated, should contribute to primer extension, explaining the discrepancy between the results of the two experimental approaches. That degradation of the *atpB* mRNA had happened during the RNA extraction or manipulation prior to ligation appears unlikely as such unspecific degradation would have targeted similarly the precursor, half of which are full-length, and the processed mRNA, all truncated in the *atpB* CDS.

5' processing of *atpB* is not the primary function of MDB1, but a consequence of its binding to the 5'UTR.

We wondered whether the CDS and/or 3'UTR could interfere with the action of MDB1 and participate to this 5' processing event. We thus used the above described *BFFF* and *BKR* chimeras, this latter being also introduced in the *Fud50* recipient strain, where the *atpB* gene is partially deleted (Woessner et al. 1984).

We first analysed the levels of the chimeric mRNAs by RNA blots and compared their amount to that of the endogenous *atpB* transcript in the wild type (Fig 6A). As already observed (Fig. 2B), the accumulation of the *BFFF* transcript was lower than that of the *atpB* transcript. Compared to *BFFF* mRNA levels, the accumulation of *BKR* chimeric transcript in the wild type was even lower and hardly detectable, as previously reported (Rimbault et al., 2000). However it accumulated to a higher level in the *Fud50* context, suggesting a competition, released in the absence of the endogenous *atpB* mRNA, between the endogenous and chimeric *atpB* 5'UTRs for some stabilisation factors, MDB1 being an obvious candidate. In strains carrying the *BKR* chimera, as in other strains expressing *aadA* chimeras (see below), hybridization with probes specific for the *atpB* 5'UTR or for the *aadA* coding sequence identified two bands (Fig. 6A, B). The higher one, detected with both probes, corresponds to the full-length chimeric mRNA. The

lower ones are typical of *aadA* chimeras and correspond to degradation products generated, upon translation of *aadA* chimeras, by an endonucleolytic cleavage, shortly after the *aadA* initiation codon (Y. Choquet, unpublish. observations, see also Fig. 2 in (Goldschmidt-Clermont, 1991)). This releases the *atpB* 5'UTR fused to the very beginning of the *aadA* CDS, of similar size for all chimeras (indicated by an asterisk *), and to the *aadA* coding sequence fused to the downstream 3'UTR, whose size varies depending on the UTR (indicated by diamonds ◊). Indeed, the levels of the chimeric *aadA*-based mRNA were much increased, while the lower bands (* and ◊) almost completely disappeared when chloroplast translation was blocked with lincomycin for 4 hr. This cleavage was observed irrespective of the 5' or 3'UTRs present in the chimeras and was thus an intrinsic property of the *aadA* coding sequence. After lincomycin treatment, the level of the BKR mRNA became comparable to that of the chimeric BFFF transcript. Still, *BKR* mRNA level remained higher in the *Fud50* context than in the wild-type background (Fig. 6B).

We then characterized the 5'ends of the *BFFF* and *BKR* chimeric transcripts by primer extension and cRT-PCR (Fig. 7). Both precursor and processed 5'ends of the chimeric transcript were detected in the BFFF strain by primer extension analysis, the mature form being by far the most abundant, as for the endogenous *atpB* (Fig 7A). cRT-PCR yielded a single PCR product of the expected size (500 bp, Fig 7B). At variance with the result observed on the *atpB* mRNA itself, sequencing of the amplicon identified only the processed 5'end in both RPP and mock samples, as expected from the relative abundance of the two mRNA species in primer extension (Fig 7B).

Surprisingly enough, for the *BKR* chimera in the wild-type, as in the Fud50 genetic context, primer extension revealed a small amount of the primary mRNA but no processed form, despite the presence of the wild-type *MDB1* allele (Fig. 7B). cRT-PCR experiments identified an amplicon of the expected size (~600-bp), irrespective of the RPP treatment or of the genetic background. After sequencing, only primary 5'ends were found, even in mock-treated samples, indicating again that the triphosphate can be removed *in vivo*. In both wild-type and *Fud50* backgrounds the 3'end of the chimeric transcripts mapped at position +84 with respect to the *rbcL* stop codon, i.e. at the major *rbcL* 3' end identified by (Goldschmidt-Clermont et al., 2008). Sequencing of amplicons with another set of primers led to the same results (not shown). Defective 5' processing of the 5'*atpB-aadA-3'rbcL* transcript did not prevent its translation since the two strains were resistant to Spectinomycin. Moreover, the accumulation of the chimeric

transcript still relied on the MDB1 factor (Fig 2). Therefore, 5' end processing is not a prerequisite for the expression of 5'*atpB*-driven transcripts, strongly suggesting that it is not the primary action of MDB1 but rather a consequence of its binding to the *atpB* mRNA.

Together these experiments showed that the correct 5' processing of *atpB* mRNA, being facilitated in the BFFF strain, compared to that in the wild type, and impaired in the BKR strain, should involve other sequence elements in the 3'UTR or CDS, alone or in combination.

The nature of 3' UTR determines 5' processing of *atpB* chimera transcripts

To further investigate this latter point, we generated alternative chimeras, always based on the *atpB* 5'UTR, fused to different CDS (from the *petA, atpB* or *rbcL* genes) followed by the *rbcL* 3'UTR (respectively the *BFR, BBR* and *BRR* chimeras, Fig. 8A). As a control we used the 5'*atpA-aadA*-3'*rbcL* chimera (*AKR* chimera, (Goldschmidt-Clermont, 1991). The *BRR* and *BFR* chimeras respectively replaced the endogenous *rbcL* gene in the wild type and the *petA* gene in the *Fud50* strain. The *BBR* chimera restored a full-length *atpB* coding sequence in strain *FuD50* while the *AKR* chimera was introduced in the wild-type chloroplast genome at a neutral site downstream of the *petA* gene.

The accumulation of these transcripts, comprising chloroplast coding sequences, was higher than that of chimeras based on the heterologous *aadA* coding sequence, as expected from their translation–induced cleavage (Fig. 6). The *BFR* mRNA was highly abundant in the *Fud50* deletion strain, MDB1 being fully available to stabilize the chimeric mRNA. The *BBR* was much less accumulated than the endogenous *atpB* mRNA, but at a still significant level. The *BRR* transcript accumulated to only ~10 % of the wild-type level of *rbcL* mRNA. Note that because of size similarity, the *atpB* and *BRR* transcript cannot be discriminated using a probe against the *atpB* 5' UTR. The *BRR* gene was not translated (Suppl Fig. S7), as a dialog between the *rbcL* 5'UTR and CDS seems to be required for efficient *rbcL* mRNA expression (Salvador et al, 1993; other refs). With the exception of the BRR strain, the accumulation of the *rbcL* mRNA was not significantly affected in strains expressing chimeric transcripts stabilised at their 3'end by the *rbcL* 3'UTR, excluding a strong competition for putative 3'*rbcL* binding factors (fig. 6A). Besides *BRR*, all other chimeras were expressed, as the BFFF and BBR strains were phototrophic and respectively accumulated the cytochrome *f* and ATP synthase subunit β (Fig. 6C), while the

AKR strtain was, as previously reported (ref), resistant to Spectinomycin. Again 5'processing is not a prerequisite for gene expression.

We determined the 5'ends of these chimeric mRNAs, either mock or RPP-treated, by cRT-PCR (Fig 8B). In all cases, amplicons of the expected size (350 bp for the *BFR, BRR* and *AKR* chimeric genes, 450-bp for the *BBR* chimera) were obtained. Additional shorter PCR products, likely corresponding to degradation intermediates were also observed. Sequencing of the amplicons revealed the presence of only a primary 5'end for the three *atpB*-driven chimeras in both RPP- and mock-treated samples (Fig 8B, left), as previously observed for the *BKR* transcript. For all amplicons, 3'end mapped to position +84 with respect to the *rbcL* stop codon. Impaired 5'processing in the presence of the *rbcL* 3'UTR seems specific to the *atpB* 5'UTR, since, at variance with the *BKR* transcript, the *AKR* transcript (Fig 8B, right) harboured a processed 5' end (+36 / TSS; (Viola et al., 2019) in both RPP- and mock-treated samples, while its 3' end mapped to position +84. These results clearly reveal a specific incompatibility between the *atpB* 5'UTR and the *rbcL* 3'UTR for the correct 5' processing of 5'*atpB*-driven transcripts, even in the presence of MDB1, and irrespective of the CDS between them.

To challenge the role of the *rbcL* 3'UTR, we constructed the *BKB* and *BKF* chimeras made of the *atpB* 5'UTR fused to the *aadA* coding sequence, followed by the *atpB* or *petA* 3' UTR (Fig 9A). These chimeras were introduced in the chloroplast genome of the *Fud50* recipient strain, downstream of the *petA* gene. Even in the *Fud50* genetic context, the *BKB* and *BKF* mRNAs were poorly accumulated, much less than the *BKR* mRNA (fig. 6) and even almost below detection threshold for the *BKF* mRNA, but regained accumulation upon lincomycin treatment (Fig. 6B). In untreated cells expressing these 5'*atpB-aadA* cassettes, the reduced transcript accumulation, caused by the translation induced cleavage, did not prevent translation, assessed indirectly by the level of resistance of the strains to antibiotics. When plated on TAP medium supplemented with increasing concentrations of spectinomycin and streptomycin, the {*Fud50* BKR}, {*Fud50* BKF}, {*Fud50* BKB}, and {*Fud50* 5_MBKB} (see below) strains all grew on antibiotic concentrations as high as 1000 µg mL^{-1} of spectinomycin plus 100 µg mL^{-1} of streptomycin (data not shown). Since poorly expressed *aadA* cassettes did not allow growth of strains on antibiotic concentrations as low as 100 $\mu g.mL^{-1}$ of spectinomycin plus 7.5 $\mu g.mL^{-1}$ of streptomycin (Choquet et al, 1998; Minai et al, 2005), we concluded that all chimeras were efficiently translated, which explains their active degradation.

cRT-PCR of the *BKB* mRNAs revealed the presence of primary and processed 5'ends in the RPP- and mock-samples respectively (Fig 10C), as already observed for the wild-type *atpB* gene. The 3' end mapped to position +89, as in the wild-type. Waiting for Marina's cRT-PCR results.

5'-3' UTR interactions may be involved in maturation of *atpB* chimera transcripts

Since *rbcL* and *petA* 3'UTRs respectively affect the 5' processing of 5'*atpB*-driven transcripts, we assessed the role of the *atpB* 3′UTR in this processes and looked for potential base-pairing between the *atpB* 5' and 3' UTRs. Interestingly, the 5' and 3' UTRs of *atpB* could base pair over 18 nt with nt 17-27 of the *atpB* 5'UTR pairing with nt 1-10 of the 3' UTR and nt 1-13 of the 5'UTR pairing with the 13-nt just downstream the stem loop in the 3'UTR, that encompass the *atpB* 3'ends and end just before the endonuclease site mapped by Stern and Kindle, (1993) (Fig 10A). To assess if disruption of this putative secondary structure would impact 5' processing or decrease transcript stability, we mutated nucleotides +2,+4,+6,+8,+10,+19 and +20 of the *atpB* 5'UTR (Fig 10B). The wild-type and mutated 5'UTRs, fused to the *aadA* CDS and followed by the *atpB* 3'UTR (*BKB* and 5_M*BKB* chimeras respectively) were inserted by transformation in the *Fud50* chloroplast genome, downstream of the *petA* gene.

RNA blot analysis did not evidence a significant decrease in the accumulation of the 5_M*BKB* transcript, compared to the *BKB* mRNA (Fig. 6B). cRT-PCR of the 5_M*BKB* mRNA revealed the presence of primary and processed 5'ends in the RPP- and mock-samples respectively (Fig 10C), as for the BKB and *atpB* mRNAs, indicating that this putative secondary structure does not play a significant role in 5'end maturation. However, at the 3'end, we observed a different cleavage site of the precursor mRNA: in the 5_MBKB RPP sample the 3'end was found at position +198 with respect to *atpB* stop codon, ~100-nt downstream of the mature 3'end as mapped here and by Stern (Stern et al, 1991).

Although we do not fully understand the molecular basis for these processing differences, mutations in the *atpB* 5'UTR alter transcript maturation of 3'ends highlighting some requirement of 5′ and 3′ UTR communication for proper gene expression. Together, our results indicated that lack of 5' processing does not prevent the accumulation and translation of chimeric transcripts, even if different combination of UTRs and CDS induce differences in the mRNA levels, as already documented (many refs).

To our knowledge, the only other example of communication between the 5' and 3' UTRs of a chloroplast gene reported to date concerns the *psbA* mRNA. The *psbA* 3'UTR, although devoid by itself of affinity for the protein complex that activates *psbA* mRNA translation, increases the affinity of this complex for the *psbA* 5' UTR, when fused in cis to it (Katz and Danon, 2002). In both cases this could provide the basis for a quality control mechanism by favouring correctly processed mRNA over the numerous transcripts that are under degradation and lack a full-length CDS.

Discussion

In our previous study of the nuclear control of the subunit β expression, we identified a nucleus-encoded factor, MDB1, required for the stable accumulation of the *atpB* mRNA. Here, we cloned the gene and further characterized the role of its protein product in *atpB* mRNA stabilisation and processing. MDB1 is an OPR protein, which further illustrates the prevalent role, in *C. reinhardtii*, of this family of nucleus-encoded helical repeat proteins in the post-transcriptional control of chloroplast gene expression.

The OPR protein MDB1 stabilises the *atpB* mRNA by targeting its 5'UTR.

When chimeras made of the *atpB* 5'UTR fused to various coding sequences and 3'UTRs (e.g. the *BFFF* and *BKR* chimeras) were introduced in the chloroplast genome of the wild-type and *mdb1* recipient strains, the chimeric mRNAs accumulated in the wild type, but not in the *mdb1* backgrounds. MDB1, thus, genetically interacts with the 5' UTR of its target gene, as do all M factors studied up to now in *Chlamydomonas.* RNA immunoprecipitation experiments further showed that it associates *in vivo* with the *atpB* 5'UTR. While this method does not allow to discriminate direct or indirect interactions with the *atpB* mRNA, OPR proteins being RNA-binding proteins, we favour a direct interaction. sRNA-Seq experiments further precise the binding site of MDB1: a small RNA footprint mapping to the first 20 nt of the *atpB* transcript whose accumulation almost vanishes in *mdb1* mutant. A point mutagenesis study (Anthonisen et al 2001) previously identified nucleotides 5-8 and 10-16 within this footprint as critical for the stability of reporter transcript made of the beginning of the *atpB* 5'UTR fused to the *uidA* reporter sequence. Moreover while the *atpB* 5'UTR is generally not conserved through evolution,, neither in sequence nor in length, even between closely related species, two short regions, one just upstream the translation initiation codon and the other surrounding this footprint, are conserved among Chlamydomonadales (Suppl. Fig. S9), further highlighting the role of this specific region. MBB1 by binding to the very 5' end of the mature *atpB* transcript would protect it from the action of 5' → 3' exonucleases, particularly active in the chloroplast, thereby stabilising it, as do other M factors (Loiselay et al, 2008; Drager et al 1998; Loizeau et al, 201?; Pfalz et al, 2011). However, as also observed in other strains defective for a M factor (Cavaiuolo et al 2017; Wang et al 2015), this footprint was not totally abolished in the *mdb1* mutant. This could be due to off-target effects: chloroplasts contain a plethora of helical repeat proteins that, in virtue of their RNA binding properties, may transiently bind sequences similar to

their own target (Hammani et al., 2009), thereby protecting to some extent the footprint of MDB1, even if this interaction is stable enough to significantly to stabilise the whole transcript.

In the absence of MDB1, the *atpB* mRNA can also be protected from 5' → 3' exo-ribonucleases by a polyG tract, which generates new 5' ends at the 5' border of the "G cage", in both wild-type and *mdb1* strains. In the wild type these pG-mRNA intermediates are more abundant than the full-length mRNA. As other chloroplast-encoded genes (Loiselay et al, 2008), the *atpB* gene is transcribed in excess over what can be stabilised by MDB1, and a significant fraction of the newly transcribed transcript is destined to degradation, as shown by the many 3'ends mapping within the *atpB* coding sequence in cRT-PCR analysis of the *atpB* mRNA in the wild type (Suppl. Fig S6).

MDB1 more specifically stabilises the processed form of the *atpB* mRNA.

Primer extension and cRT-PCR analyses indicated that the mature *atpB* transcript is a processed mRNA (this work; (Blowers et al, 1990; Woessner et al, 1986; Anthoninson et al, 2001; Cavaiuolo et al, 2017), whose accumulation depends on MDB1. Low amounts of the primary mRNA were detected in the wild type as in the mutant, while the abundant processed form was only found in the wild type. Similar results were observed with other stabilisation factors e.g. the MBB1 and NAC2 proteins, that respectively control the accumulation of the *psbB/psbH* and *psbD* processed mRNAs (Vaistij et al 2000a; Nickelsen et al. 1994).

However, our 5'end mapping experiments of the *BKR*, *BBR*, *BRR* and *BFR* mRNAs revealed the absence of processing of the *atpB* 5'UTR even in the presence of MDB1. The accumulation of the *BKR* transcript nevertheless relies on MBD1, while the accumulation of the endogenous precursor transcript is much reduced in the *mdb1* mutant, compared to the wild type. The *atpB* precursor transcript also is thus stabilized by MDB1, whose role in transcript stabilization is not limited to the mature mRNA, but concerns also the precursor.

A 5'-end-dependent pathway for degrading primary transcript in chloroplast

In these 5'processing defective strains, as in the *mdb1* mutant, we could detect the precursor transcript by cRT-PCR on mock-treated samples, showing that the triphosphorylated transcription product can be converted *in vivo* to a monophosphorylated form. Similar results were also obtained for *petA* (Suppl. fig S8) and *rbcL* mRNAs (Johnson et al., 2010) in

Chlamydomonas and for some chloroplast and mitochondrial transcripts in *Arabidopsis* (Kuhn et al., 2005; Zhelyazkova et al., 2012), suggesting that this conversion step is a general feature of organelle gene expression. It could be the first step of a 5'-end degradation pathway, as observed in bacteria (Richards et al., 2011; Luciano et al., 2012; Foley et al., 2015). In *E. coli* and *B. subtilis*, this conversion is carries out by the NUDIX hydrolase RppH (Deana et al., 2008; Richards et al., 2011), in prelude to RNA cleavage by the endonucleases RNaseE and RNaseY, respectively, and subsequent exonucleolytic trimming (Mudd et al., 1990; Shahbabian et al., 2009). mRNA decay triggered by 5' pyrophosphate removal is functionally similar to mRNA decapping in eukaryotes, which leads to mono-phosphorylated mRNAs degraded by 5' → 3' exoribonucleases (Muhlrad et al., 1994; Muhlrad and Parker, 1994). In *Arabidopsis*, 9 of the 28 encoded Nudix hydrolases are predicted to be targeted to the chloroplast. They exhibit pyrophosphohydrolase activity toward various substrates, such as ADP-Ribose, ADP-Glucose, CoA, and NADH but whether one of them can act on triphosphorylated mRNAs is still unknown (Ogawa et al., 2008; Yoshimura and Shigeoka, 2015). *Chlamydomonas* also contains 26 genes encoding putative Nudix hydrolases, out of which three to six, depending on the prediction software used, could be chloroplast localised. Future work will help elucidate if these and/or other Nudix hydrolases dephosphorylate primary transcripts 5′ ends. However, this pyrophosphatase activity should remain low in *Chlamydomonas* chloroplast, as several chloroplast transcripts, including some highly abundant (e.g. *rbcL*, *atpH*, *petA*, *psaA*), are constitutively triphosphorylated (Loiselay et al., 2008; Johnson et al., 2010; Cavaiuolo et al., 2017)(Ozawa et al, 2020), with only a tiny fraction in a mono-phosphorylated state.

Impact of 3'UTRs on processing of *atpB* 5' chimeric mRNAs

In this work, we studied the behaviour, in terms of stability and 5' processing, of a series of chimera, in which the *atpB* 5' UTR was combined with various coding sequences and 3' UTRs and showed that 3'UTRs can impact the 5' processing of these chimeric transcripts. The *rbcL* 3'UTR, in combination with four different coding sequences (*atpB*, *rbcL*, *petA*, *aadA*), prevented the processing of the *atpB* 5'UTR, but not of the *atpA* 5'UTR. Conversely, the *petA* 3'UTR still allowed the 5' processing of the *BKF* and *BFF*chimeric transcript. Processing defects were specific to *5'atpB*-driven chimeras, as the AKR mRNAs were normally processed. We prove here that 5' processing can be controlled by 3' UTR sequences. In eukaryotes as in prokaryotes,

interactions between the 5' and 3' ends of a given mRNA are often used for post-transcriptional regulation of gene expression, and govern mRNA stability, translation and degradation (Filbin and Kieft, 2016; De los Mozos et al, 2013). Circularization of the mRNA through physical bridges between 5' and 3' ends can regulate translation in eukaryotes and virus (Gallie, 1998; Filbin and Kieft, 2016). To our knowledge, the only other example of communication between the 5' and 3' UTRs of a chloroplast gene reported to date concerns the *psbA* mRNA. The *psbA* 3'UTR, although devoid by itself of affinity for the protein complex that activates *psbA* mRNA translation, increases the affinity of this complex for the *psbA* 5' UTR, when fused in cis to it (Katz and Danon, 2002). In both cases this could provide the basis for a quality control mechanism by favouring correctly processed mRNA over the (numerous) transcripts undergoing degradation and lacking a full-length CDS.

The different 3'UTRs in our chimeric mRNAs would change the overall RNA folding of the transcripts in a way that would disrupt or favour the formation of new RNA-RNA interactions that could eventually interfere whit mRNA processing. RNA folding predictions suggested a possible complementarity between the *atpB* 5' and 3'UTRs. Mutagenesis of the 5'UTR sequence involved in this base pairing did not affect 5'processing, but altered the maturation of its 3'end. The mature 3' end of the *atpB* mRNA is generated by a two step-process which involves an endonuclease cleavage downstream of a stabilising stem-loop structure followed by 3' → 5' trimming up to this stem-loop (Stern and Kindle 1993). The UCA endonuclease target site ~13 bp downstream the stem loop (Stern and Kindle 1993) is partially involved in this putative base-pairing (Fig 10A-B). We postulate that the substitutions, by destabilising this structure could prevent or attenuate the activity of the endonuclease.

Both RNA folding predictions and experimental assay revealed the presence of secondary structure in the 5'UTR of *atpB* (Anthonisen et al., 2001). In bacteria, the presence of such hairpin structure at the 5' ends of mRNA extends the half-life of transcripts by preventing 5' attacks by RppH and 5'->3' degradation (Emory et al., 1992; Arnold et al., 1998), other ref). Here, sequestering the tri-phosphorylated 5' end of the *atpB* mRNA in a secondary structure could possibly render it less accessible to an RppH-like enzyme. Binding of MDB1 could open such terminal structure making the primary 5'end of *atpB* accessible to the nucleolytic attack. Similarly, the *rbcL* 3'UTR can base-pair with the processed *atpB* 5'UTR and possibly inhibit

triphosphate removal by an RppH-like enzyme, a prerequisite for 5'->3' for efficient 5' → 3' degradation.

Material and methods

Strains, growth conditions and crosses.

WT-t222, WT S24, *Fud50* (Woessner et al 1984),), *mdb1-thm24* (Drapier et al 1992), *mdb1-L35a* (Houille-Vernes et al., 2011), L35a.011- +MDB1-HA complemented strain, and transformed strains of *Chlamydomonas*, all derived from 137c, were grown at 25°C in Tris-acetate-phosphate (TAP) medium (pH 7.2) (Harris, 1989) under continuous low light (5 to 10 µE m^{-2} s^{-1}). Crosses were performed according to Harris (1989).

Plasmid constructions and chloroplast transformation

Standard nucleic acids manipulations were performed according to (Sambrook et al., 1989). Details of DNA constructs are provided in the Supplementary Materials section. Primers used in this study are listed in Supplemental Table ST2. All constructs were sequenced before transformation into Chlamydomonas.

Genomic DNA preparation, whole genome sequencing and data analysis

were performed as in (Boulouis et al., 2015).

Transformation

Chloroplast transformation was performed by tungsten particle bombardment (Boynton et al., 1988) as described in (Kuras and Wollman, 1994). Transformants were either selected on TAP medium supplemented with spectinomycin (100 µg ml^{-1}) under continuous low light (5 to 10 µE m^{-2} s^{-1}) or on Minimum medium (60 µE m^{-2} s^{-1}) for the restoration of phototrophy. They were sub-cloned on selective medium (TAP medium supplemented with 500 µg.mL^{-1} of spectinomycin in darkness or Minimum medium under 120 µE m^{-2} s^{-1}) until they reached homoplasmy, assessed by restriction fragment length polymorphism or sequencing of specific PCR products. At least three independent transformants were analysed for each transformation and proved to be nearly identical.

Nuclear transformation by electroporation was performed as described in Raynaud et al, 2008 using the following electrical parameters: 25 mF and 1000 V cm^{-1}. Transformants were selected on Mimimum medium for the restoration of phototrophy.

RNA extraction and analysis.

Total RNA was extracted from 200 mL cultures (2-3 x 10^6 cells mL^{-1}) according to (Drapier et al., 1998). For sRNA-Seq and cRT-PCR assays, Arintricarboxylic acid was omitted in the AQE extraction buffer and RNA was subsequently treated with DNAse I (NEB). A fraction of the RNA sample was treated with RNA 5' Polyphosphatase (RPP, Epicentre) to convert 5' triphosphates to mono-phosphorylated 5' ends, then phenol-chloroform extracted. Mock-treated RNA samples went through the same process without addition of the RPP enzyme.

sRNA-Seq datasets were produced and analysed as described in (Cavaiuolo et al, 2017) and deposited in the Short Read Archive (SRA) database under the BioProject PRJNA379963.

For cRT-PCR 10 µg of RPP and mock-treated total RNAs were self-ligated using T4 RNA ligase (Fisher Scientific) for 2 h at 37°C, then ethanol precipitated. RNAs were reverse-transcribed with Superscript III (Invitrogen) using gene-specific primers (2 pmol). The cDNA was PCR-amplified with Taq PCR Core Kit (Qiagen) or Phusion High-Fidelity DNA Polymerase (Thermo Scientific) using outward-directed primers against the 5' and 3' ends of the transcript. Amplicons were gel purified and sequenced, or eventually cloned into the pCR Blunt-TOPO™ Vector using the TOPO™ zero Blunt Cloning Kit (Invitrogen).

Primer extension was performed as described in (Sturm et al., 1994). 15 µg of total RNA, mixed with 1.5 pmol of the 5′ ^{33}P-phosphorylated oligonucleotide, denatured at 75°C for 4 min were then rapidly cooled down in ethanol-dry ice. Reverse transcription was performed at 54°C for 30 min with Avian Myeloblastosis Virus Reverse Transcriptase (AMV RT). After LiCl precipitation, the reaction was run on 5% sequencing gel alongside a sequencing reaction of the same region.

RNA blots was carried out as described in (Drapier et al., 1998), using PCR-generated DNA probes labeled with digoxigenin (Sigma) or with ^{33}P-labeled DNA probes. Probes, if not otherwise specified were prepared by PCR with the oligonucleotides indicated in (Maul et al., 2002).

Ribonuclease protection experiments were performed as described in (Drapier et al., 1992) to determine the 5'end of *atpB* using the 575bp HindIII/HinfI fragment as a probe (5' region of *atpB*), while primer extensions were performed on total RNA extracts as in (Sturm et al, 1994).

RNA immunoprecipitation was done according to Boulouis et al, 2011.

Protein Preparation, Separation, and Analysis

Protein isolation, separation, and immunoblot analyses were performed on exponentially growing cells (2-3 10^6 cells·mL^{-1}) as described (Kuras and Wollman, 1994). All immunoblots were repeated at least twice and performed on three independent transformants. Cell extracts, loaded on equal chlorophyll basis, were analysed by SDS-PAGE (12-18% acrylamide gradients and 8 M urea). At least three biological replicas were performed for each experiment. Proteins were detected by ECL. Primary antibodies, diluted 100,000-fold (antibody cytochrome *f*), 50,000-fold (CF1β), 10 000-fold (OEE2) were revealed by horseradish peroxidase–conjugated antibodies against rabbit IgG (Promega). Antibodies against the β-subunit of F1/CF1, the OEE2 subunit from the photosystem II oxygen-evolving complex, and cytochrome *f* have been described (de Vitry et al., 1989; Lemaire and Wollman, 1989; Kuras and Wollman, 1994). MDB1-HA was detected by ECL using monoclonal anti HA.11 (Covance) antibodies, and horseradish peroxidase–conjugated antibody against mouse IgG (Promega). Protein accumulation (normalized to that of OEE2, as an internal standards) was, when required, quantified from ChemiTouch (Bio-Rad) scans of the membrane, using the ImageLab (v3.0) software.

Figure legends

Figure 1: Structure of the *MDB1* gene.

A) The top line shows a schematic map of the MDB1 genomic region on chromosome 14, with the relevant genes models. That encoding an OPR protein is shown in blue. The red rectangle indicates the region deleted in the *Δmdb1-L35* mutant strain. The position of the BAC sequences (PTQ4126 Chr14:1039373-1097235 and PTQ4327 Chr14:995949-105714) able to complement the *mdb1* mutations are shown as dashed lines. A diagram of the MDB1 gene is shown below, with exons represented as blue boxes. The position of the mutation in the *thm24* strain with respect to the nucleotide sequence is indicated.

B) Schematic representation of the MDB1 protein, with the position of the OPR motives shown as blue arrows. The green rectangle indicate portions of the proteins conserved in other Chlamydomonadaceae algae, while grey rectangles point to region specific to *C. reinhardtii*. The predicted secondary structure of the protein as determined by Jpred **(http://www.compbio.dundee.ac.uk/jpred/) is shown below.**

Figure 2: The *atpB* 5'UTR is the target of the MDB1 protein.

A) Schematic map of the BFFF and BKR chimeric genes.

B) RNA blot analysis of transcript accumulation among the progeny of BFFF (mt +) or BKR (mt+) crosses with the *mdb1-1* (mt-) mutant. Two members of each tetrad lacked the *atpB* and chimeric mRNAs. The asterisk indicate a co-transcript initiated at the *petA* promoter that extends up to the *rbcL* 3'UTR. The *petD* and *petA* mRNAs provide loading controls.

C) RNA immunoprecipitation analysis. Schematic representation of the *atpB* 5'UTR: the position of the transcription start site (TSS) and of the 5' post-transcriptional processing site (5PTP) are shown. The probe used for dot-blot hybridization is depicted as a blue line from position +27 to 321 of the 5'UTR. The results from HA-RIP from *mdb1-2* ::*MDB1-HA* and WT cells expressing MDB1 with or without an HA tag respectively are shown on the right. RNA was extracted before (I, input) and after immunoprecipitation with HA-antibodies (P, pellet) and hybridized with the *atpB* 5'UTR probe. Hydridisation with a probe on the *atpA* 5'UTR provides a negative control.

Figure 3: A small RNA mapping to the *atpB* 5'end is reduced in the absence of MDB1.

Distribution of small RNAs (11-44-nt) along the atpB gene in mutant strain mdb1-2 (red) compared to the wild type (blue). The horizontal arrow indicates the atpB coding sequence and its orientation on the Cp genome. The vertical arrows point to the position of the mature 5'end (+27) and transcription start site (+1). The sequence of the small RNA is displayed below. The box marks the region important for transcript stability determined by (Anthonisen et al.) from position +31 to 42 of the atpB 5' UTR. sRNA-Seq coverage is expressed in reads per million (RPM) and averaged over two biological replicates for each strain.

Figure 4: Binding of MDB1 to the 5'UTR of *atpB* protects from 5'>3' exonucleases

A) Scheme of the pG-*atpB* gene cassette construction. The pG(18) was inserted 32 nt upstream of the ATG. The pG-*atpB* was inserted along with a selectable marker cassette, in place of the endogenous *atpB*. B) pG-*atpB* mRNA accumulation. Two bands were detected, which correspond to the pG transcript and to a degradation intermediate that accumulate as a result of exoribonuclease activity (scheme is shown on the left, with the polyG cage symbolized by a red hexagon). Samples marked with an asterisc * were subjected to small RNA sequencing. C) Coverage of small RNAs in the wild-type (left) and mutant mdb1-1 strains (right) along the *atpB*-pG 5'UTR. Vertical arrows point to the 5'ends at position +1 and +27. The two horizontal arrows indicate the poly(G) tract and the first 100 nt of the *atpB* coding sequence. Coverage is expressed in RPM and averaged over two biological replicates for each strain.

Figure 5: Determination of *atpB* 5'ends in the wild type and in the *mdb1-1* strain.

A) Primer extension analysis. An end-labeled primer was annealed to total RNA from *mdb1-1* and wild type and then extended with reverse transcriptase. The extension products were run on a 6% polyacrylamide/8 M urea gel alongside a DNA sequence ladder obtained with the same primer. The positions of the mapped 5' ends in WT and mutant are indicated by arrow on the 5'UTR sequence. In A and B the -10 box of the promoter is underlined.

B) cRT-PCR analysis. Schematic representation of the *atpB* gene structure with the position of the primers used: the vertical black arrow indicates the position of the transcription start site; the blue left oriented dashed arrow indicates the primer used for reverse transcription of circularized *atpB* mRNA, while left and right-directed black arrows correspond to the PCR primers. Agarose gels showing the resulting amplicons, with a molecular weight marker on the right. RPP and mock respectively indicate RNA samples treated or not with RPP, to distinguish precursors from processed transcripts. The positions of the mapped 5' and 3' ends in wild type and *mdb1-1* along the 5'UTR or 3'UTR sequences are indicated with arrows on the right panel (for the 3' end, nucleotides are numbered from the last nucleotide of the stop codon). The 3'ends of *atpB* are underlined (Stern et al, 1991).

Figure 6: Expression of chimeric transcripts.

A) Accumulation of the chimeric mRNA assessed by RNA blots

Total RNAs from the indicated strains were hybridized with probes specific to the *atpB* 5'UTR, or the *aadA*, *rbcL* and *petA* coding sequences. For the three blots the nucleus-encoded *cblp2* transcript provides a loading control. Histograms indicate mRNA levels of 5'*atpB* chimera, *rbcL* and *petA* as ratio of the amount of RNA in mutant or chimeric strains over that observed in the wild type (both amount being normalised to the accumulation of the cblp2 transcript to correct for variations in loading).

B) Accumulation of the chimeric *aadA* transcripts is affected by the translation of the mRNA.

Total RNA of the indicated strains, treated (+) or not (-) with lincomycin for 4 hr, were hybridised with the probes indicated on the left. The position of the chimeric transcript cleavage products (* and ◊) is indicated. The red asterisk * points to a transcript recognised by the petA probe initiated at the *petA* promoter and ending at the *ycf2* 3'UTR, included in the promoter and 5'UTR fragment of atpB. The AKR strain, expressing a 5'*atpA-aadA*-3'*rbcL* cassette, was included as a control to show that the translation-induced cleavage of the chimeric mRNA does not depend on the *atpB* 5'UTR but on the *aadA* CDS. The vertical line separates non-contiguous lanes of the same gel, in order to remove irrelevant intervening samples.

C) Accumulation of the translation products of the *BFF*, *BFR* and *BRR* chimeric genes.

Whole cell proteins extracts from the indicated strains were separated by electrophoresis, blotted on nitrocellulose membrane and immuno-decorated with antibodies against the proteins indicated on the left. Three independent transformants are shown for each chimeric context. The OEE2 subunit of PSII and the Ponceau red staining of the membrane provide loading controls.

Figure 7: Determination of the 5'- and 3' ends of the BFR, BBR, BRR and AKR chimeric transcripts.

A) Schematic representation of the chimeras with the positions of the primers used for cRT-PCR. Dashed arrows shows the primers used for reverse transcription of the circularized RNAs. Same conventions than in Fig. 2A. The position of the insertion of the recycling 5*'psaA-aadA-3'rbcL* cassette is shown, with an arrow indicating the sense of transcription. For the *BBR* chimera the thick arrow represents the inverted repeat. Bs: *Bse*RI, E: *Eco*RI; X: *Xho*I; K: *Kpn*I; RV*: Eco*RV.

B) cRT-PCR results. See Fig.5C for details

Figure 8: Determination of the ends of the *BKF* chimeric transcripts.

Schematic representation of the chimeric gene structure and cRT-PCR results of BKF in A and AFFF in B. See Fig.5C for details.

Figure 9: Putative RNA:RNA interactions between the 5'UTR and 3'UTR of *atpB*.

A) RNA secondary structure predictions of the *5'atpB-3'atpB* UTR pair. The positions of the *atpB* 5'ends are indicated by arrows. The endonuclease cleavage sites detected by (Stern and Kindle 1993) are highlighted by a pink rectangle and black arrows point to the 3'ends mapped by (Stern 1991). The grey arrow indicates the major 3'end mapped in this study.

B) Schematic representation of the chimeric gene structure BKB and 5' BKB and primer used for cRT-PCR. The individual single base changes introduced into the 5' UTR sequence of the 5'BKB are shown in red.

C) cRT-PCR results. See Fig.5C for details

Supplemental Material:

Supplementary Methods: DNA constructs

Supplementary Figures:

Suppl. Fig. S1: complementation of the mdb1 mutants by either BACs or midigene
Suppl. Fig. S2: The OPR repeats within MDB1
Suppl. Fig S3: Conservation of MDB1 among Chlamydomonadales.
Suppl; fig. S4: S1 mapping of the *atpB* 5'ends in the wild-type and mdb1-1 strains.
Suppl. Fig. S5: Sequence analysis of the RPP-treated sample from wild type.
Suppl. Fig. S6: Alternative determination of the *atpB* 5' end by cRT-PCR.
Suppl. Fig. S7: Characterisation of the BRR transformants
Suppl. Fig S8: Determination of *petA* 5' and 3' ends in the Fud50 strain.
Suppl. Fig S9: Conservation of the MDB1 binding site among Chlamydomonadales

Acknowledgments:

Funding: This work was supported by basic support to Unité Mixte de Recherche 7141, from CNRS and Sorbonne Université, by Agence Nationale de la Recherche ChloroRNP: ANR-13-BSV7-0001-01), and by the "Initiative d'Excellence" program (Grant "DYNAMO," ANR-11-LABX-0011-01).

Table I: chimeras used in that work:

Name of the chimera	5'UTR	CDS	3'UTR	Insertion locus	ref
AKR	*atpA*	*aadA*	*rbcL*	*petA* (*Eco*RV)	1
pGatpB	*atpB(pG)*	*atpB*	*atpB*	*atpB*	This work
BFFF	*atpB*	*petA*	*petA*	*petA*	2
BFR	*atpB*	*petA*	*rbcL*	*petA*	This work
BRR	*atpB*	*rbcL*	*rbcL*	*rbcL*	This work
BBR	*atpB*	*atpB*	*rbcL*	*atpB*	This work
BKR	*atpB*	*aadA*	*rbcL*	*petA* (*Eco*RV)	3
BKF	*atpB*	*aadA*	*petA*	*petA* (*Eco*RV)	This work
BKB	*atpB*	*aadA*	*atpB*	*petA* (*Eco*RV)	This work
5_M*BKB*	$5'_M$*atpB*	*aadA*	*atpB*	*petA* (*Eco*RV)	This work

1: Kuras et al, 1997 ; 2: Drapier et al, 2008; 3: Rimbault et al, 2000

References

Anthonisen, I.L., Salvador, M.L., and Klein, U. (2001). Specific sequence elements in the 5' untranslated regions of rbcL and atpB gene mRNas stabilize transcripts in the chloroplast of Chlamydomonas reinhardtii. Rna **7,** 1024-1033.

Arnold, T.E., Yu, J., and Belasco, J.G. (1998). mRNA stabilization by the ompA 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. RNA **4,** 319-330.

Barkan, A. (2011). Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. Plant Physiol **155,** 1520-1532.

Barkan, A., and Goldschmidt-Clermont, M. (2000). Participation of nuclear genes in chloroplast gene expression. Biochimie **82,** 559-572.

Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol **65,** 415-442.

Boulouis, A., Drapier, D., Razafimanantsoa, H., Wostrikoff, K., Tourasse, N.J., Pascal, K., Girard-Bascou, J., Vallon, O., Wollman, F.A., and Choquet, Y. (2015). Spontaneous dominant mutations in chlamydomonas highlight ongoing evolution by gene diversification. Plant Cell **27,** 984-1001.

Boynton, J.E., Gillham, N.W., Harris, E.H., Hosler, J.P., Johnson, A.M., Jones, A.R., Randolph-Anderson, B.L., Robertson, D., Klein, T.M., and Shark, K.B. (1988). Chloroplast transformation in Chlamydomonas with high velocity microprojectiles. Science **240,** 1534-1538.

Cavaiuolo, M., Kuras, R., Wollman, F.A., Choquet, Y., and Vallon, O. (2017). Small RNA profiling in Chlamydomonas: insights into chloroplast RNA metabolism. Nucleic Acids Res **45,** 10783-10799.

Choquet, Y., and Wollman, F.A. (2002). Translational regulations as specific traits of chloroplast gene expression. FEBS Lett **529,** 39-42.

de Vitry, C., Olive, J., Drapier, D., Recouvreur, M., and Wollman, F.A. (1989). Posttranslational events leading to the assembly of photosystem II protein complex: a study using photosynthesis mutants from Chlamydomonas reinhardtii. J Cell Biol **109,** 991-1006.

Deana, A., Celesnik, H., and Belasco, J.G. (2008). The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. Nature **451,** 355-358.

Drager, R.G., Zeidler, M., Simpson, C.L., and Stern, D.B. (1996). A chloroplast transcript lacking the 3' inverted repeat is degraded by 3'-->5' exoribonuclease activity. RNA **2,** 652-663.

Drager, R.G., Higgs, D.C., Kindle, K.L., and Stern, D.B. (1999). 5' to 3' exoribonucleolytic activity is a normal component of chloroplast mRNA decay pathways. Plant J **19,** 521-531.

Drager, R.G., Girard-Bascou, J., Choquet, Y., Kindle, K.L., and Stern, D.B. (1998). In vivo evidence for 5'-->3' exoribonuclease degradation of an unstable chloroplast mRNA. Plant J **13,** 85-96.

Drapier, D., Girard-Bascou, J., and Wollman, F.A. (1992). Evidence for Nuclear Control of the Expression of the atpA and atpB Chloroplast Genes in Chlamydomonas. Plant Cell **4,** 283-295.

Drapier, D., Rimbault, B., Vallon, O., Wollman, F.A., and Choquet, Y. (2007). Intertwined translational regulations set uneven stoichiometry of chloroplast ATP synthase subunits. EMBO J **26,** 3581-3591.

Drapier, D., Suzuki, H., Levy, H., Rimbault, B., Kindle, K.L., Stern, D.B., and Wollman, F.A. (1998). The chloroplast atpA gene cluster in Chlamydomonas reinhardtii. Functional analysis of a polycistronic transcription unit. Plant Physiol **117,** 629-641.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2001). Predicting subcellular localization of proteins based on their N- terminal amino acid sequence. J.Mol.Biol.2000.Jul.21;300(4):1005-16. **300,** 1005-1016.

Emory, S.A., Bouvet, P., and Belasco, J.G. (1992). A 5'-terminal stem-loop structure can stabilize mRNA in Escherichia coli. Genes Dev **6,** 135-148.

Foley, P.L., Hsieh, P.K., Luciano, D.J., and Belasco, J.G. (2015). Specificity and evolutionary conservation of the Escherichia coli RNA pyrophosphohydrolase RppH. J Biol Chem **290,** 9478-9486.

Furuichi, Y., LaFiandra, A., and Shatkin, A.J. (1977). 5'-Terminal structure and mRNA stability. Nature **266,** 235-239.

Germain, A., Hotto, A.M., Barkan, A., and Stern, D.B. (2013). RNA processing and decay in plastids. Wiley interdisciplinary reviews. RNA **4,** 295-316.

Goldschmidt-Clermont, M. (1991). Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker of site-directed transformation of chlamydomonas. Nucleic.Acids.Res. **19,** 4083-4089.

Goldschmidt-Clermont, M., Rahire, M., and Rochaix, J.D. (2008). Redundant cis-acting determinants of 3' processing and RNA stability in the chloroplast rbcL mRNA of Chlamydomonas. Plant J **53,** 566-577.

Hammani, K., Bonnard, G., Bouchoucha, A., Gobert, A., Pinker, F., Salinas, T., and Giege, P. (2014). Helical repeats modular proteins are major players for organelle gene expression. Biochimie **100,** 141-150.

Harris, E.H. (1989). The Chlamydomonas source book: a comprehensive guide to biology and laboratory use. (San Diego: Academic Press).

Houille-Vernes, L., Rappaport, F., Wollman, F.A., Alric, J., and Johnson, X. (2011). Plastid terminal oxidase 2 (PTOX2) is the major oxidase involved in chlororespiration in Chlamydomonas. Proc Natl Acad Sci U S A **108,** 20820-20825.

Hsu, C.L., and Stevens, A. (1993). Yeast cells lacking 5'-->3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. Mol Cell Biol **13,** 4826-4835.

Ishikura, K., Takaoka, Y., Kato, K., Sekine, M., Yoshida, K., and Shinmyo, A. (1999). Expression of a foreign gene in Chlamydomonas reinhardtii chloroplast. J Biosci Bioeng **87,** 307-314.

Johnson, X., Wostrikoff, K., Finazzi, G., Kuras, R., Schwarz, C., Bujaldon, S., Nickelsen, J., Stern, D.B., Wollman, F.A., and Vallon, O. (2010). MRL1, a conserved Pentatricopeptide repeat protein, is required for stabilization of rbcL mRNA in Chlamydomonas and Arabidopsis. Plant Cell **22,** 234-248.

Kasai, S., Yoshimura, S., Ishikura, K., Takaoka, Y., Kobayashi, K., Kato, K., and Shinmyo, A. (2003). Effect of coding regions on chloroplast gene expression in Chlamydomonas reinhardtii. J Biosci Bioeng **95,** 276-282.

Katz, Y.S., and Danon, A. (2002). The 3'-untranslated region of chloroplast psbA mRNA stabilizes binding of regulatory proteins to the leader of the message. J Biol Chem **277,** 18665-18669.

Keeling, P.J. (2010). The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Lond B Biol Sci **365,** 729-748.

Kuhn, K., Weihe, A., and Borner, T. (2005). Multiple promoters are a common feature of mitochondrial genes in Arabidopsis. Nucleic Acids Res **33,** 337-346.

Kuras, R., and Wollman, F.A. (1994). The assembly of cytochrome b6/f complexes: an approach using genetic transformation of the green alga Chlamydomonas reinhardtii. EMBO J **13,** 1019-1027.

Lemaire, C., and Wollman, F.A. (1989). The chloroplast ATP synthase in Chlamydomonas reinhardtii. I. Characterization of its nine constitutive subunits. J Biol Chem **264,** 10228-10234.

Loiselay, C., Gumpel, N.J., Girard-Bascou, J., Watson, A.T., Purton, S., Wollman, F.A., and Choquet, Y. (2008). Molecular identification and function of cis- and trans-acting determinants for petA transcript stability in Chlamydomonas reinhardtii chloroplasts. Mol Cell Biol **28,** 5529-5542.

Loizeau, K., Qu, Y., Depp, S., Fiechter, V., Ruwe, H., Lefebvre-Legendre, L., Schmitz-Linneweber, C., and Goldschmidt-Clermont, M. (2014). Small RNAs reveal two target sites of the RNA-maturation factor Mbb1 in the chloroplast of Chlamydomonas. Nucleic Acids Res **42,** 3286-3297.

Luciano, D.J., Hui, M.P., Deana, A., Foley, P.L., Belasco, K.J., and Belasco, J.G. (2012). Differential control of the rate of 5'-end-dependent mRNA degradation in Escherichia coli. J Bacteriol **194,** 6233-6239.

Maul, J.E., Lilly, J.W., Cui, L., dePamphilis, C.W., Miller, W., Harris, E.H., and Stern, D.B. (2002). The Chlamydomonas reinhardtii plastid chromosome: islands of genes in a sea of repeats. Plant Cell **14,** 2659-2679.

Minai, L., Wostrikoff, K., Wollman, F.A., and Choquet, Y. (2006). Chloroplast biogenesis of photosystem II cores involves a series of assembly-controlled steps that regulate translation. Plant Cell **18,** 159-175.

Mudd, E.A., Krisch, H.M., and Higgins, C.F. (1990). RNase E, an endoribonuclease, has a general role in the chemical decay of Escherichia coli mRNA: evidence that rne and ams are the same genetic locus. Mol Microbiol **4,** 2127-2135.

Muhlrad, D., and Parker, R. (1994). Premature translational termination triggers mRNA decapping. Nature **370,** 578-581.

Muhlrad, D., Decker, C.J., and Parker, R. (1994). Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. Genes Dev **8,** 855-866.

Ogawa, T., Yoshimura, K., Miyake, H., Ishikawa, K., Ito, D., Tanabe, N., and Shigeoka, S. (2008). Molecular characterization of organelle-type Nudix hydrolases in Arabidopsis. Plant Physiol **148,** 1412-1424.

Pfalz, J., Bayraktar, O.A., Prikryl, J., and Barkan, A. (2009). Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. EMBO J **28,** 2042-2052.

Rahire, M., Laroche, F., Cerutti, L., and Rochaix, J.D. (2012). Identification of an OPR protein involved in the translation initiation of the PsaB subunit of photosystem I. Plant J **72,** 652-661.

Richards, J., Liu, Q., Pellegrini, O., Celesnik, H., Yao, S., Bechhofer, D.H., Condon, C., and Belasco, J.G. (2011). An RNA pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in Bacillus subtilis. Mol Cell **43,** 940-949.

Rimbault, B., Esposito, D., Drapier, D., Choquet, Y., Stern, D., and Wollman, F.A. (2000). Identification of the initiation codon for the atpB gene in Chlamydomonas chloroplasts excludes translation of a precursor form of the beta subunit of the ATP synthase. Mol Gen Genet **264,** 486-491.

Rott, R., Levy, H., Drager, R.G., Stern, D.B., and Schuster, G. (1998). 3'-Processed mRNA is preferentially translated in Chlamydomonas reinhardtii chloroplasts. Mol Cell Biol **18,** 4605-4611.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning. (Cold Spring Harbor Laboratory Press).

Shahbabian, K., Jamalli, A., Zig, L., and Putzer, H. (2009). RNase Y, a novel endoribonuclease, initiates riboswitch turnover in Bacillus subtilis. EMBO J **28,** 3523-3533.

Small, I., Peeters, N., Legeai, F., and Lurin, C. (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics **4,** 1581-1590.

Stern, D.B., and Gruissem, W. (1987). Control of plastid gene expression: 3' inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell **51,** 1145-1157.

Stern, D.B., Jones, H., and Gruissem, W. (1989). Function of plastid mRNA 3' inverted repeats. RNA stabilization and gene-specific protein binding. Journal of Biological Chemistry **264,** 18742-18750.

Stern, D.B., Radwanski, E.R., and Kindle, K.L. (1991). A 3' stem/loop structure of the Chlamydomonas chloroplast atpB gene regulates mRNA accumulation in vivo. Plant Cell **3,** 285-297.

Sturm, N.R., Kuras, R., Buschlen, S., Sakamoto, W., Kindle, K.L., Stern, D.B., and Wollman, F.A. (1994). The petD gene is transcribed by functionally redundant promoters in Chlamydomonas reinhardtii chloroplasts. Mol Cell Biol **14,** 6171-6179.

Tardif, M., Atteia, A., Specht, M., Cogne, G., Rolland, N., Brugiere, S., Hippler, M., Ferro, M., Bruley, C., Peltier, G., Vallon, O., and Cournac, L. (2012). PredAlgo: a new subcellular localization prediction tool dedicated to green algae. Mol Biol Evol **29,** 3625-3639.

Vaistij, F.E., Goldschmidt-Clermont, M., Wostrikoff, K., and Rochaix, J.D. (2000). Stability determinants in the chloroplast psbB/T/H mRNAs of Chlamydomonas reinhardtii. Plant J.2000.Mar;21(5):469-82. **21,** 469-482.

Viola, S., Cavaiuolo, M., Drapier, D., Eberhard, S., Vallon, O., Wollman, F.A., and Choquet, Y. (2019). MDA1, a nucleus-encoded factor involved in the stabilization and processing of the atpA transcript in the chloroplast of Chlamydomonas. Plant J.

Woessner, J.P., Gillham, N.W., and Boynton, J.E. (1986). The sequence of the chloroplast atpB gene and its flanking regions in Chlamydomonas reinhardtii. Gene **44,** 17-28.

Woodson, J.D., and Chory, J. (2008). Coordination of gene expression between organellar and nuclear genomes. Nat Rev Genet **9,** 383-395.

Yoon, B.J. (2009). Hidden Markov Models and their Applications in Biological Sequence Analysis. Current genomics **10,** 402-415.

Yoshimura, K., and Shigeoka, S. (2015). Versatile physiological functions of the Nudix hydrolase family in Arabidopsis. Biosci Biotechnol Biochem **79,** 354-366.

Zhelyazkova, P., Sharma, C.M., Forstner, K.U., Liere, K., Vogel, J., and Borner, T. (2012). The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. Plant Cell **24,** 123-136.

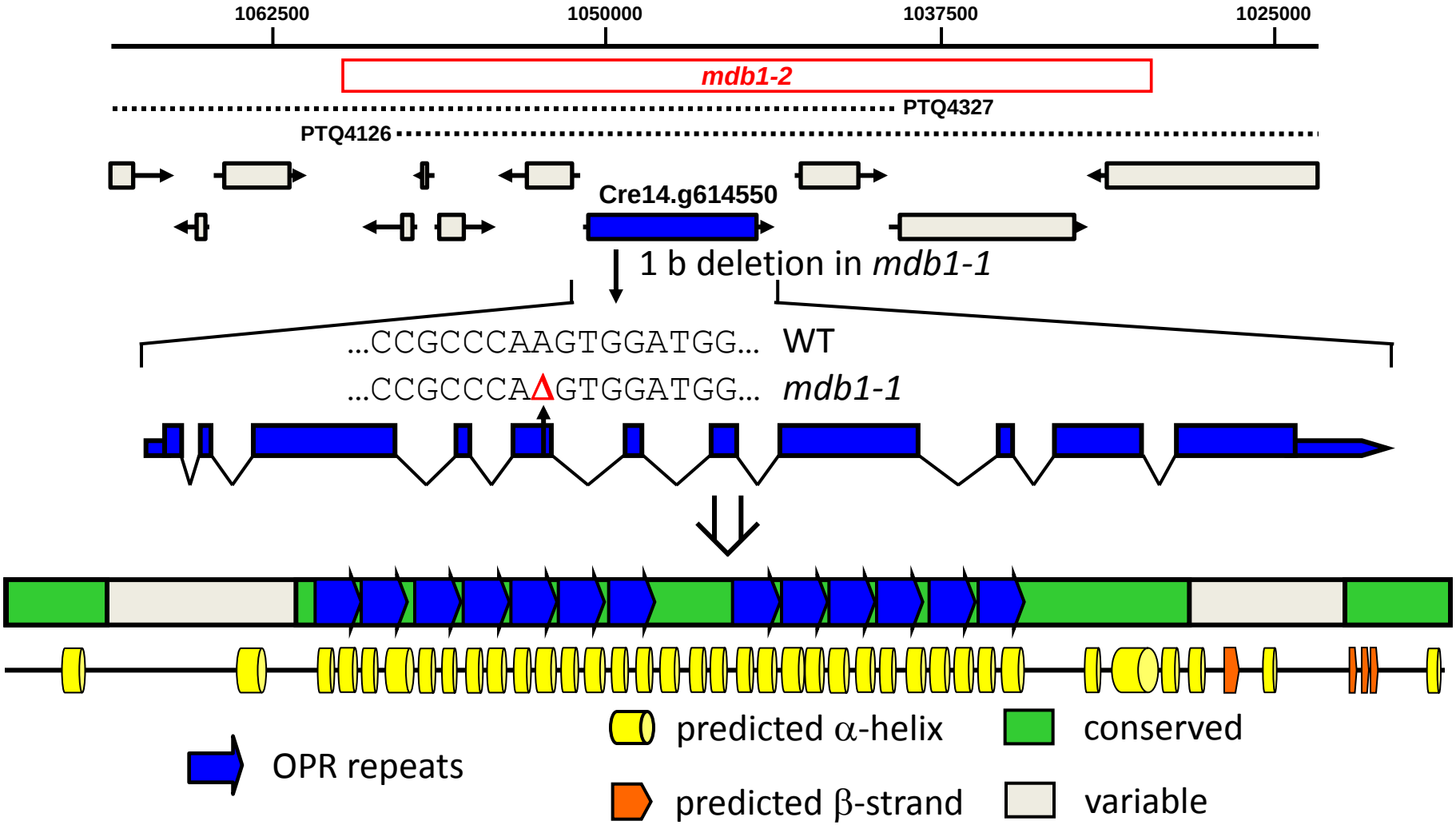


Figure 1. Structure of the MDB1 gene.

A) The top line shows a schematic map of the MDB1 genomic region on chromosome 14, with the relevant genes models. That encoding an OPR protein is shown in blue. The red rectangle indicates the region deleted in the *Δmdb1-L35* mutant strain. The position of the BAC sequences (PTQ4126 Chr14:1039373-1097235 and PTQ4327 Chr14:995949-105714) able to complement the *mdb1* mutations are shown as dashed lines. A diagram of the MDB1 gene is shown below, with exons represented as blue boxes. The position of the mutation in the *thm24* strain with respect to the nucleotide sequence is indicated.

B) Schematic representation of the MDB1 protein, with the position of the OPR motives shown as blue arrows. The green rectangle indicate portions of the proteins conserved in other Chlamydomonadaceae algae, while grey rectangles point to region specific to *C. reinhardtii*. The predicted secondary structure of the protein as determined by Jpred (http://www.compbio.dundee.ac.uk/jpred/) is shown below.

Figure 2

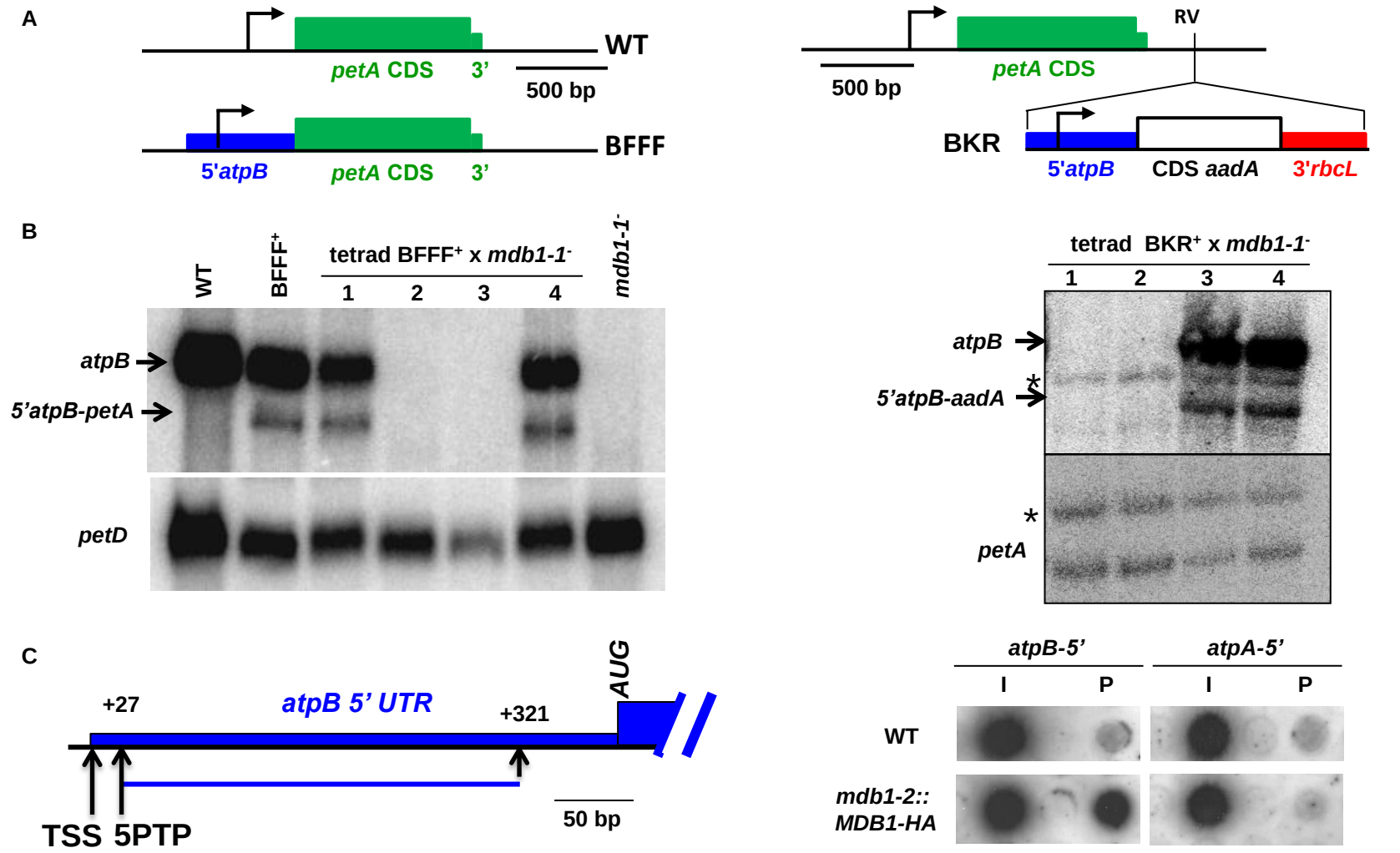


Figure 2. The *atpB* 5'UTR is the target of the MDB1 protein.

A) Schematic map of the BFFF and BKR chimeric genes. Thick rectangle symbolizes the coding sequences, thins rectangles the UTRs. Bent arrows represents promotters. Relevant restriction sites are indicated (RV: *Eco*RV)

B) RNA blot analysis of transcript accumulation among the progeny of BFFF (mt +) or BKR (mt+) crosses with the *mdb1-1* (mt-) mutant. Two members of each tetrad lacked the *atpB* and chimeric mRNAs. The asterisk indicate a co-transcript initiated at the *petA* promoter that extends up to the *rbcL* 3'UTR. The *petD* and *petA* mRNAs provide loading controls.

C) RNA immunoprecipitation analysis. Schematic representation of the *atpB* 5'UTR: the position of the transcription start site (TSS) and of the 5' post-transcriptional processing site (5PTP) are shown. The probe used for dot-blot hybridization is depicted as a blue line from position +27 to 321 of the 5'UTR. The results from HA-RIP from *mdb1-2* ::*MDB1-HA* and WT cells expressing MDB1 with or without an HA tag respectively are shown on the right. RNA was extracted before (I, input) and after immunoprecipitation with HA-antibodies (P, pellet) and hybridized with the *atpB* 5'UTR probe. Hydridisation with a probe on the *atpA* 5'UTR provides a negative control.

Figure 3

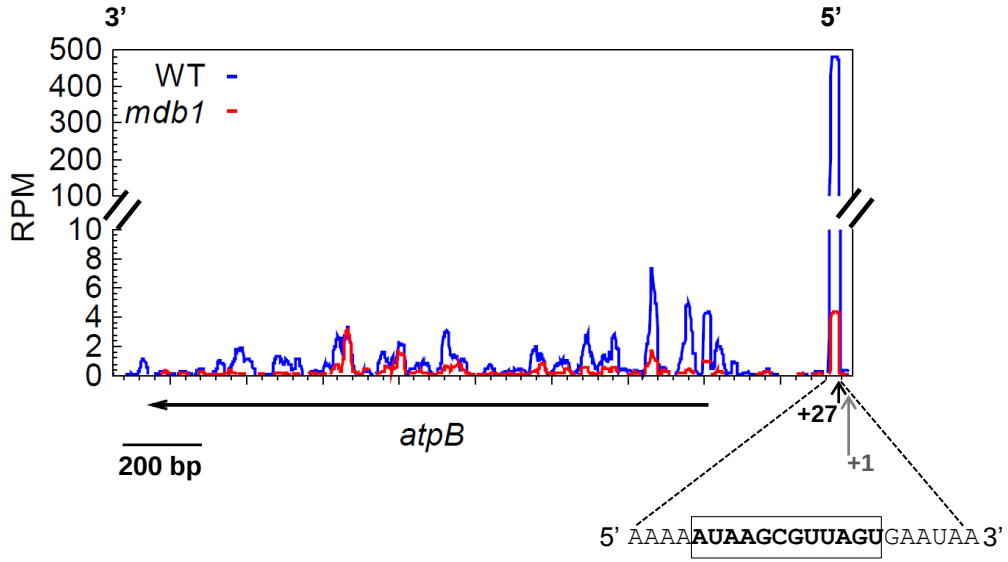


Figure 3. A small RNA mapping to the *atpB* 5'end is reduced in the absence of MDB1.
Distribution of small RNAs (11-44-nt) along the *atpB* gene in mutant strain *mdb1-2* (red) compared to the wild type (blue). The horizontal arrow indicates the *atpB* coding sequence and its orientation on the Cp genome. The vertical arrows point to the position of the mature 5'end (+27) and transcription start site (+1). The sequence of the small RNA is displayed below. The box marks the region important for transcript stability determined by (Anthonisen et al.) from position +31 to 42 of the *atpB* 5' UTR. sRNA-Seq coverage is expressed in reads per million (RPM) and averaged over two biological replicates for each strain.

Figure 4

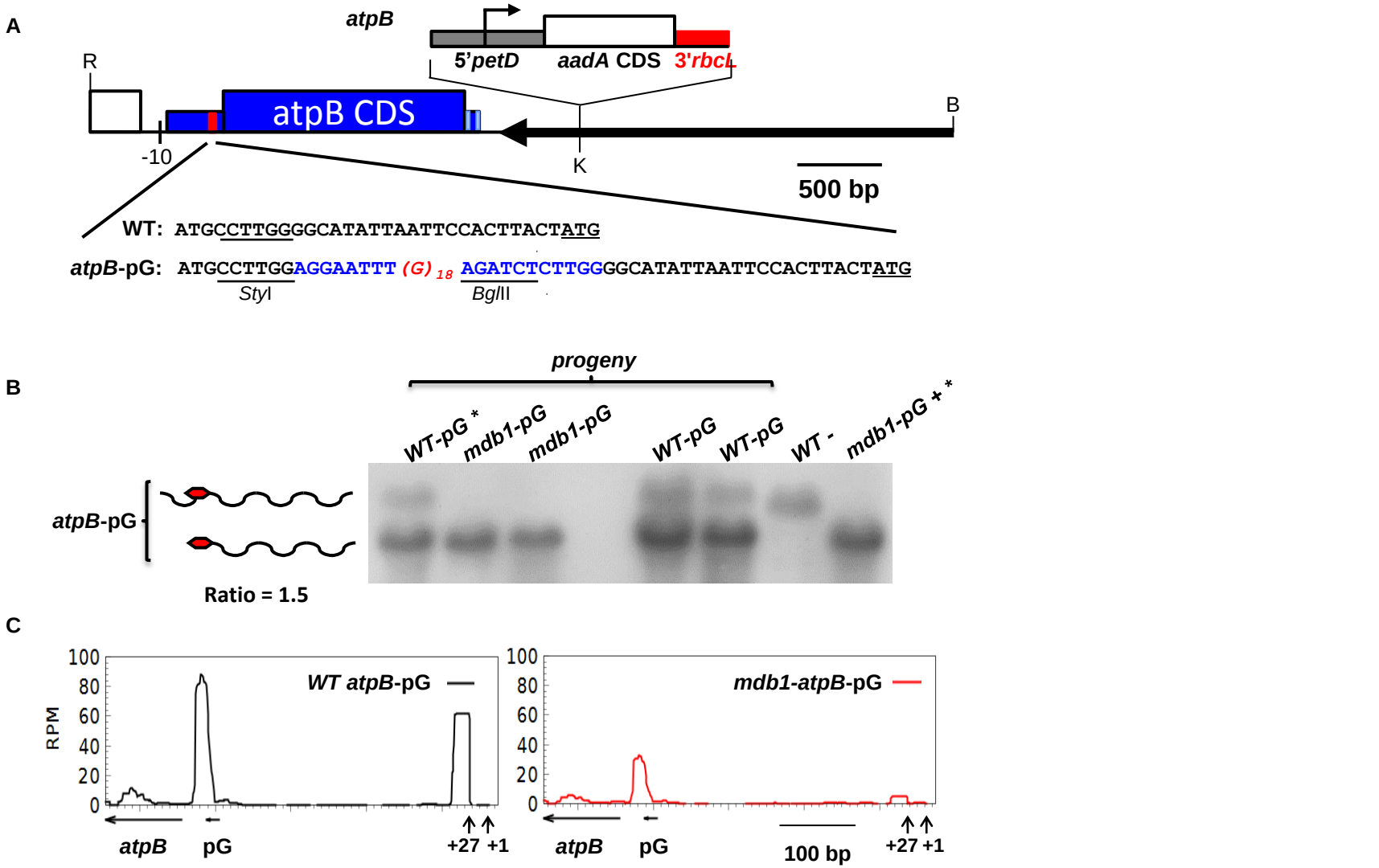


Figure 4. Binding of MDB1 to the 5'UTR of *atpB* protects from 5'>3' exonucleases

A) Scheme of the pG-*atpB* gene cassette construction. The pG(18) was inserted 32 nt upstream of the ATG. The pG-*atpB* was inserted along with a selectable marker cassette, in place of the endogenous *atpB*. B) pG-*atpB* mRNA accumulation. Two bands were detected, which correspond to the pG transcript and to a degradation intermediate that accumulate as a result of exoribonuclease activity (scheme is shown on the left, with the polyG cage symbolized by a red hexagon). Samples marked with an asterisc * were subjected to small RNA sequencing. C) Coverage of small RNAs in the wild-type (left) and mutant mdb1-1 strains (right) along the *atpB*-pG 5'UTR. Vertical arrows point to the 5'ends at position +1 and +27. The two horizontal arrows indicate the poly(G) tract and the first 100 nt of the *atpB* coding sequence. Coverage is expressed in RPM and averaged over two biological replicates for each strain.

Figure 5

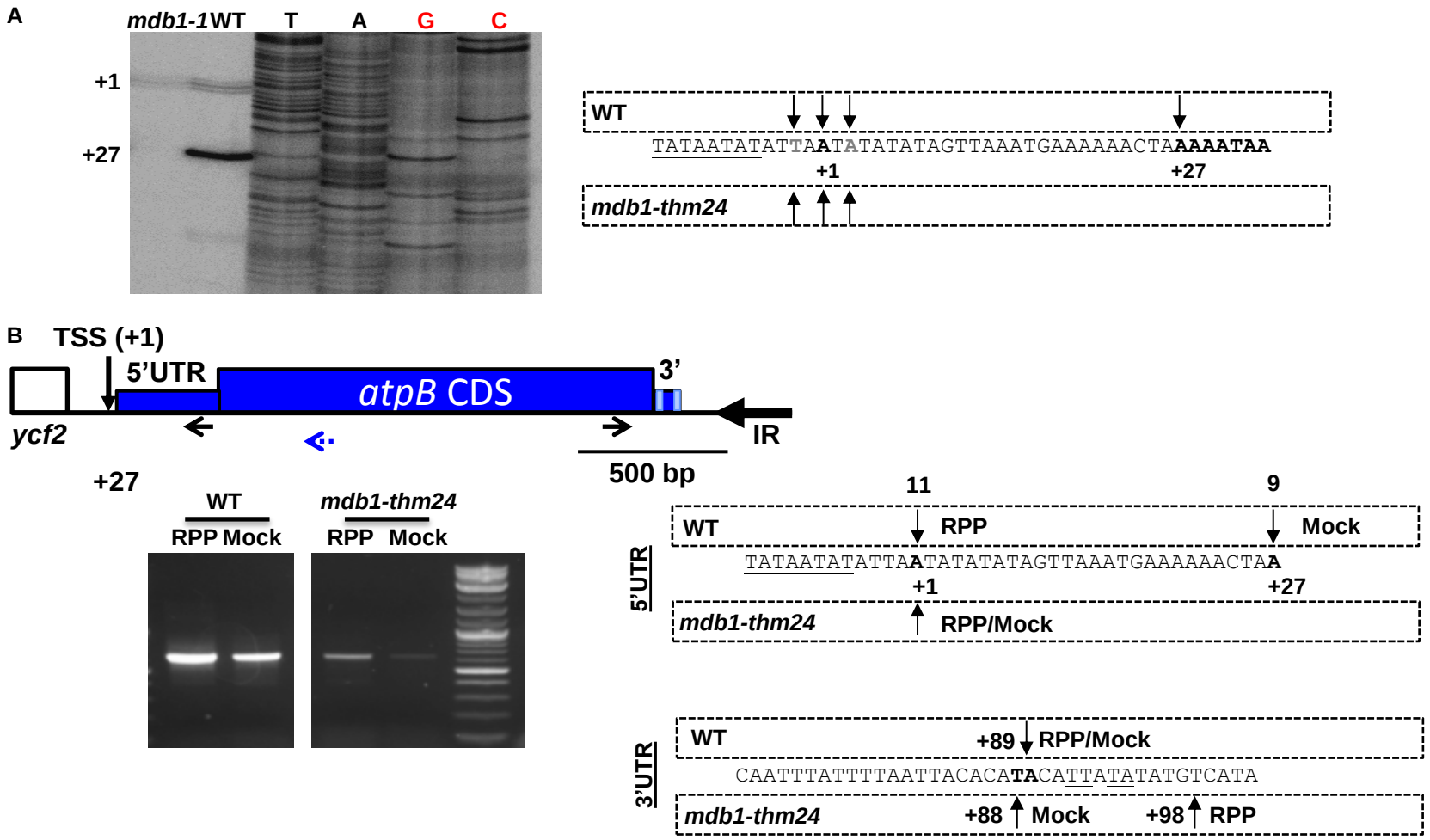


Figure 5: Determination of *atpB* 5'ends in the wild type and in the *mdb1-1* strain.
A) Primer extension analysis. An end-labeled primer was annealed to total RNA from *mdb1-1* and wild type and then extended with reverse transcriptase. The extension products were run on a 6% polyacrylamide/8 M urea gel alongside a DNA sequence ladder obtained with the same primer. The positions of the mapped 5' ends in WT and mutant are indicated by arrow on the 5'UTR sequence. In A and B the -10 box of the promoter is underlined.
B) cRT-PCR analysis. Schematic representation of the *atpB* gene structure with the position of the primers used: the vertical black arrow indicates the position of the transcription start site; the blue left oriented dashed arrow indicates the primer used for reverse transcription of circularized *atpB* mRNA, while left and right-directed black arrows correspond to the PCR primers. Agarose gels showing the resulting amplicons, with a molecular weight marker on the right. RPP and mock respectively indicate RNA samples treated or not with RPP, to distinguish precursors from processed transcripts. The positions of the mapped 5' and 3' ends in wild type and *mdb1-1* along the 5'UTR or 3'UTR sequences are indicated with arrows on the right panel (for the 3' end, nucleotides are numbered from the last nucleotide of the stop codon). The 3'ends of *atpB* are underlined (Stern et al, 1991).

Figure 6A

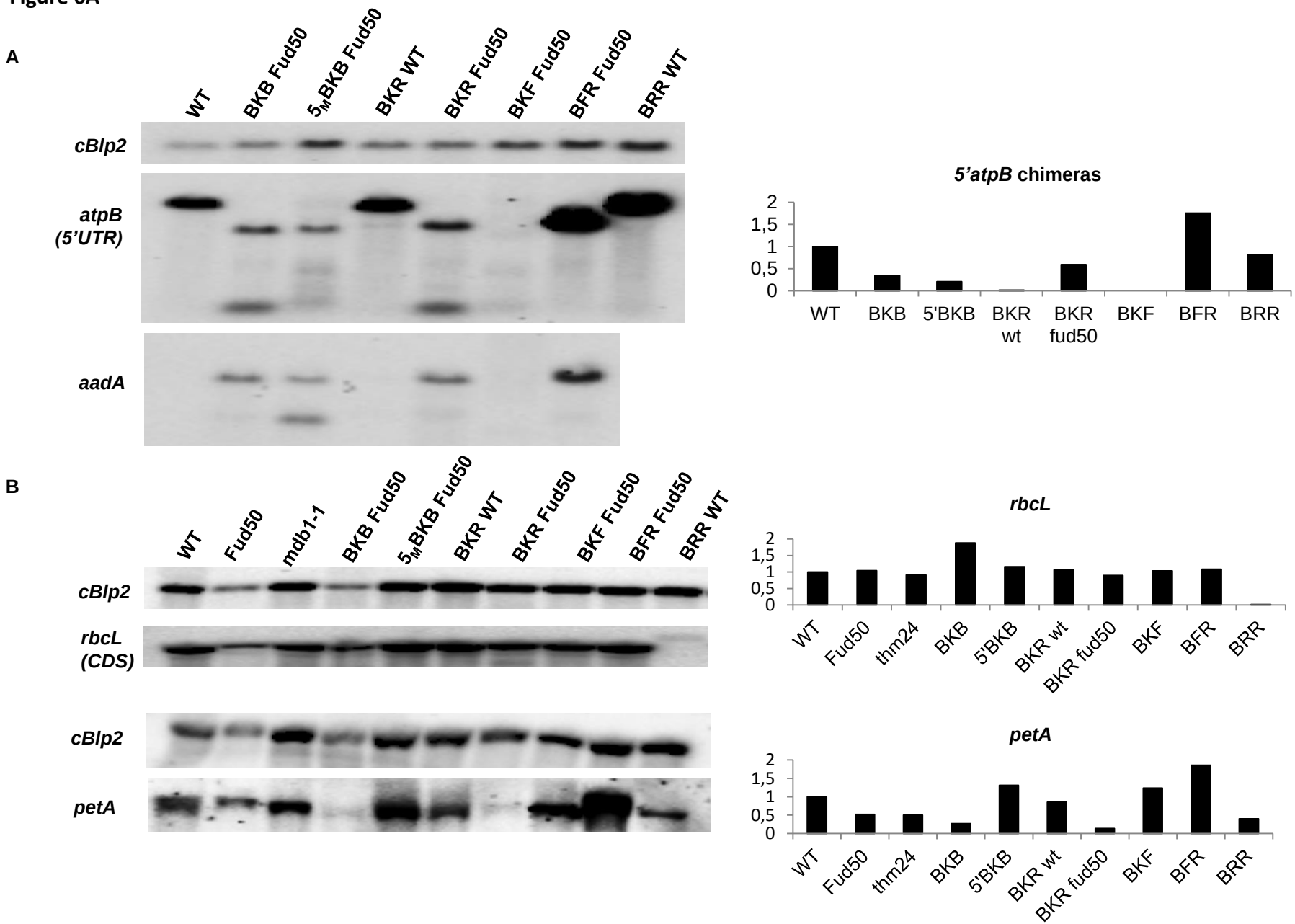


Fig. 6: Accumulation of chimeric transcripts assessed by RNA blots. Total RNAs were separated on a 1.2% agarose–formaldehyde gel were hybridized with probes specific to the *atpB* 5'UTR, or the *aadA*, *rbcL* and *petA* coding sequences. For the three blots the nucleus-encoded *cβlp2* transcript provides a loading control. Histrograms indicate mRNA levels of 5'*atpB* chimera, *rbcL* and *petA* as ratio of the mutant /chimeric strains versus WT after normalisation to the accumulation of the cβlp2 transcript.

Figure 6B

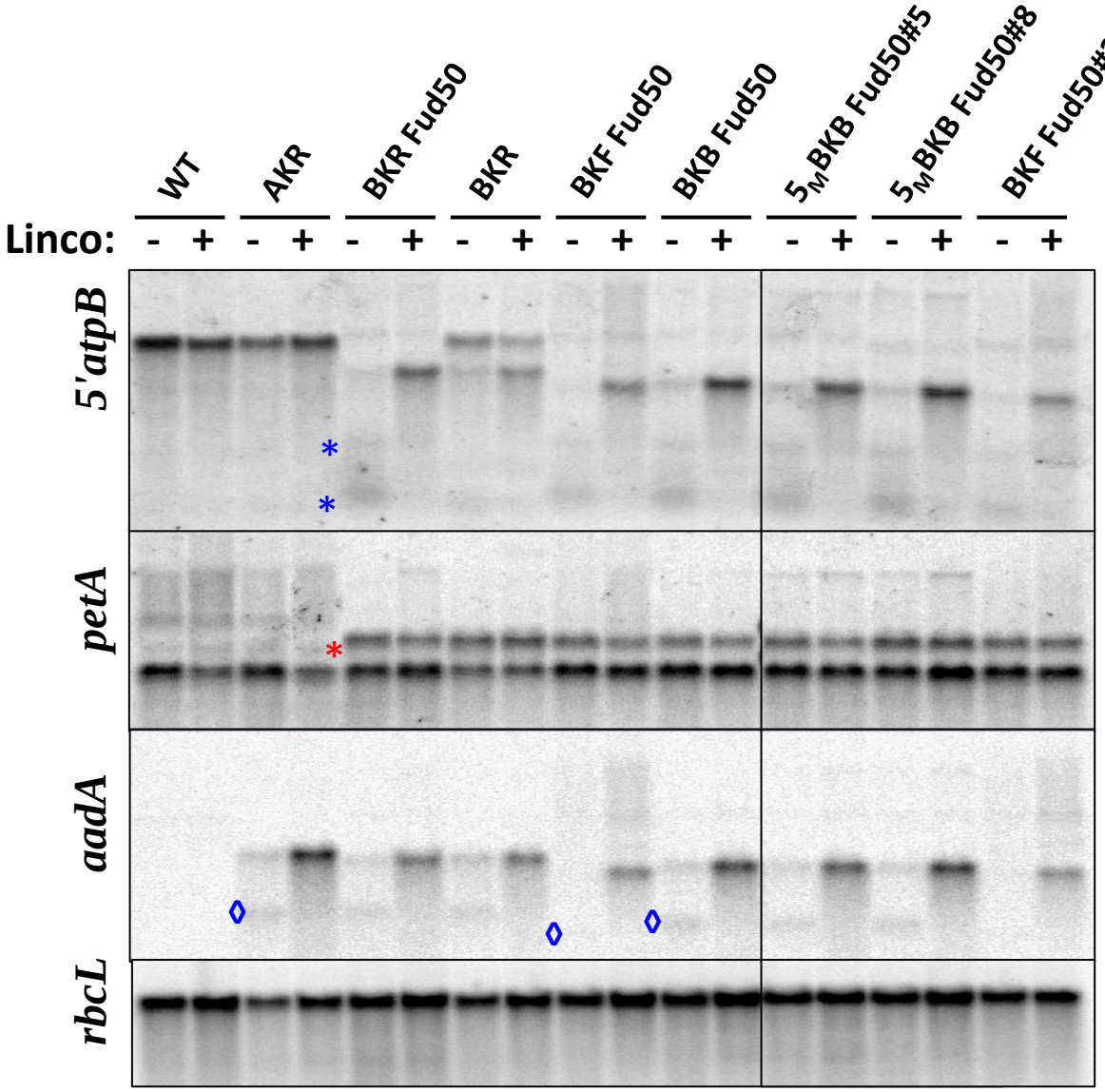


Figure 6B: Accumulation of the chimeric *aadA* transcripts is affected by the translation of the mRNA.
Total RNA of the indicated strains, treated (+) or not (-) with lincomycin for 4 hr, were hybridised with the probes indicated on the left. The position of the chimeric transcript cleavage products (* and ◊) is indicated. the red asterisk * points to a transcript initiated at the *petA* promoter and terminated at the *ycf2* 3'UTR, included in the promoter and 5'UTR fragment of atpB. The AKR strain, expressing a 5'*atpA*-*aadA*-3'*rbcL* cassette, was included to shown that the translation-induced cleavage of the chimeric mRNA does not depend on the *atpB* 5'UTR but on the *aadA* CDS. The vertical line separates non contiguous lanes of the same gel, in order to remove irrelevant intervening samples.

Figure 6C

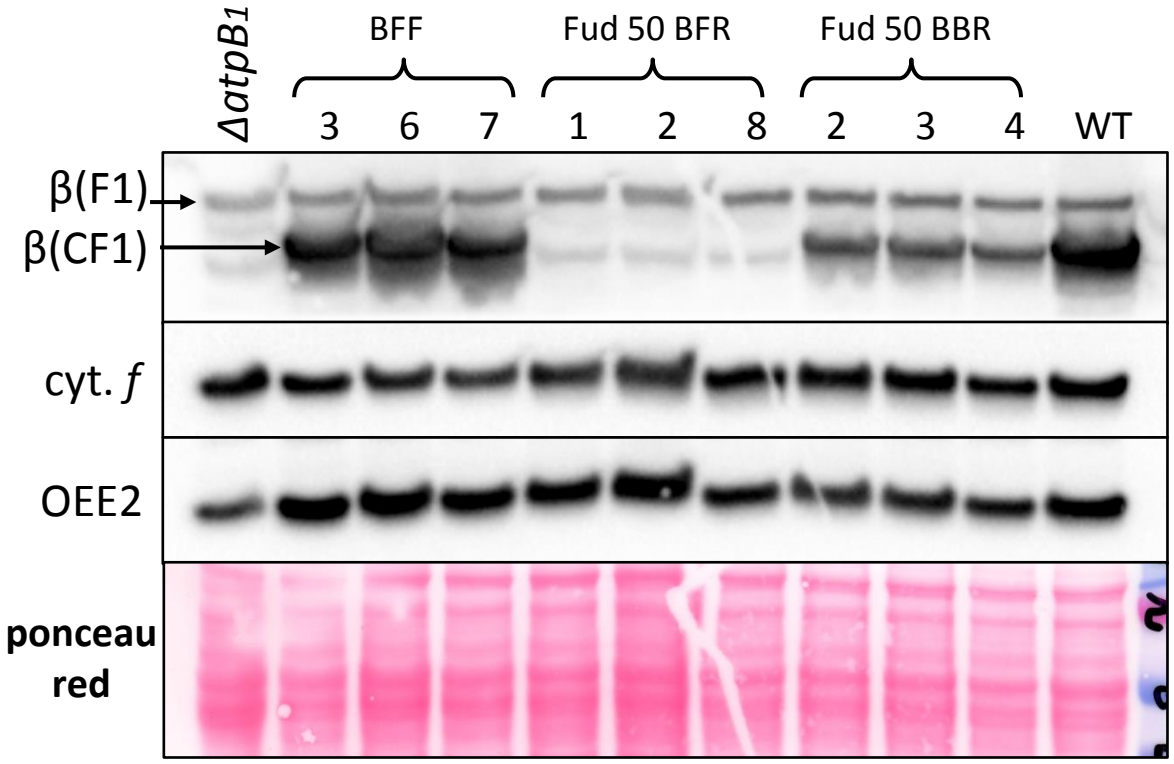


Figure 6C: Accumulation of the translation products of the chimeric genes.
Whole cell proteins extracts from the indicated strains were separated by electrophoresis, blotted on nitrocellulose membrane and immuno-decorated with antibodies against the proteins indicated on the left. Three independent transformants are shown for each chimeric context.

Figure 7

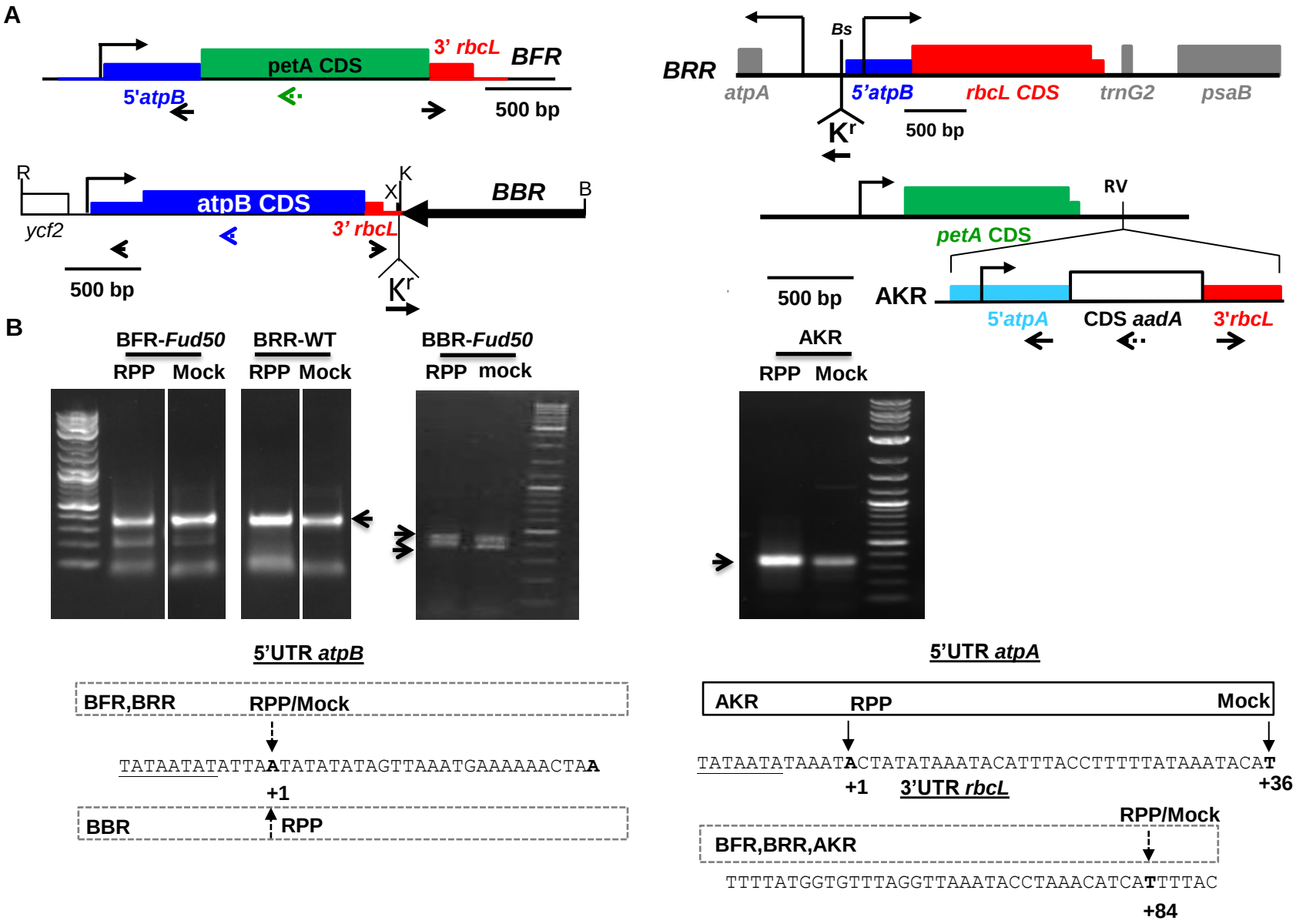


Figure 7: Determination of the 5'- and 3-ends of the *BFR*, *BBR*, *BRR* and *AKR* chimeric transcripts:.
A) Schematic representation of the chimeras with the positions of the primers used for cRT-PCR. Dashed arrows shows the primers used for reverse transcription of the circularized RNAs. Same conventions than in Fig. 2A. The position of the insertion of the recycling 5*'psaA-aadA-3'rbcL* cassette is shown, with an arrow indicating the sense of transcription. For the *BBR* chimera the thick arrow represents the inverted repeat. Bs: *Bse*RI, E: *Eco*RI; X: *Xho*I; K: *Kpn*I; RV*: Eco*RV.
B) cRT-PCR results. See Fig.5C for details

Figure 8

To be done

Figure 8. Determination of the 5’ and 3’ends of the *BKF* chimeric transcript.
Schematic representation of the chimeric gene structure and cRT-PCR results of BKF. See Fig.3C for details.

Figure 9

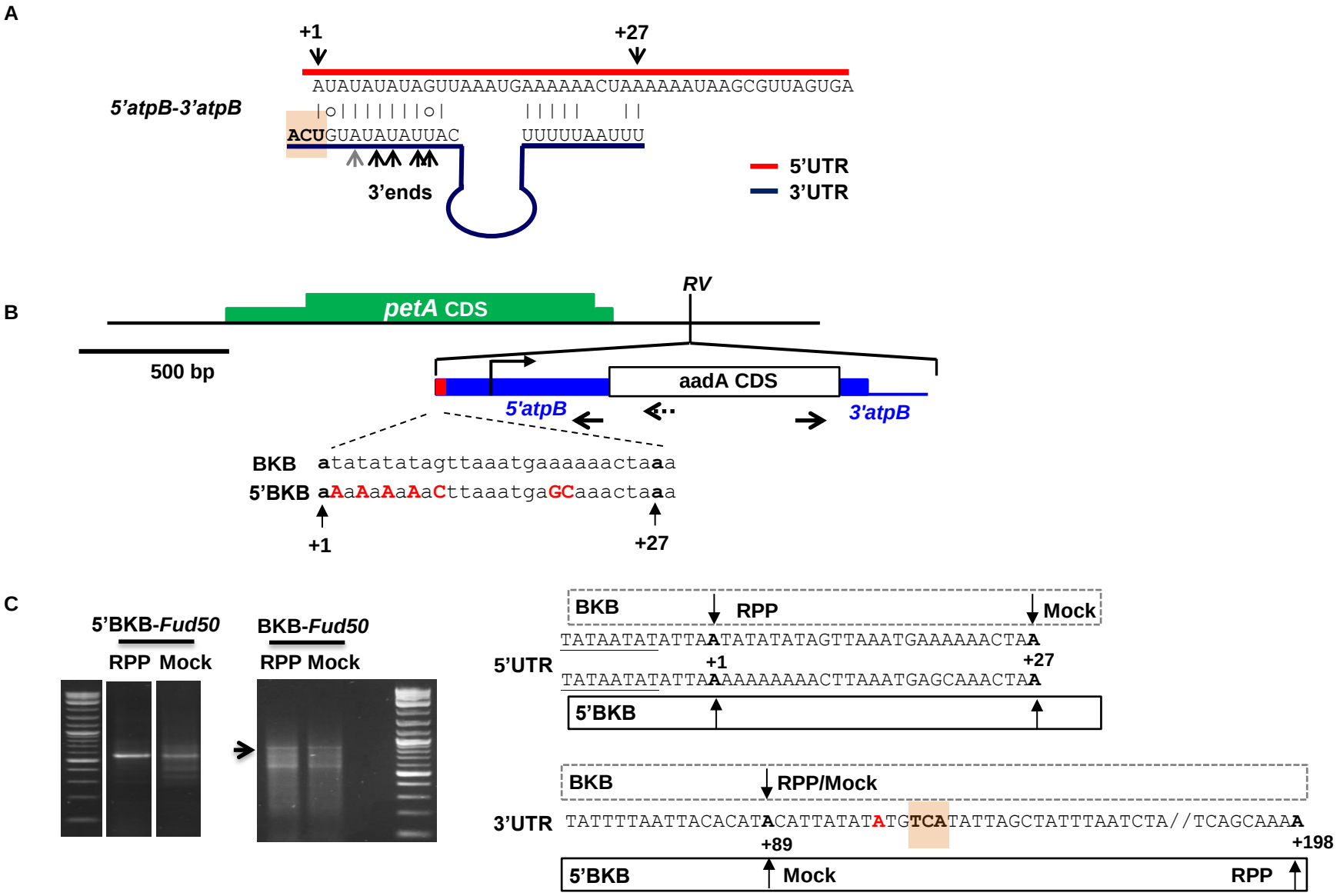


Figure 9. Putative RNA:RNA interactions between the 5'UTR and 3'UTR of *atpB*.
A) RNA secondary structure predictions of the *5'atpB-3'atpB* UTR pair. The positions of the *atpB* 5'ends are indicated by arrows. The endonuclease cleavage sites detected by (Stern and Kindle 1993) are highlighted by a pink rectangle and black arrows point to the 3'ends mapped by (Stern 1991). The grey arrow indicates the major 3'end mapped in this study.
B) Schematic representation of the chimeric gene structure BKB and 5' BKB and primer used for cRT-PCR. The individual single base changes introduced into the 5' UTR sequence of the 5'BKB are shown in red.
C) cRT-PCR results. See Fig.3C for details

Supplemental Materials

Supplementary Methods

DNA constructs

Plasmids AFFF (Choquet et al, 1998), AKR, BKR and BFFF have been already described. AKR, BKR and BFFF were respectively called pWFA in (Kuras et al., 1997), pKdBTf in (Drapier et al., 2007) and pWFB ATG123 in (Rimbault et al., 2000).

Construction of chimeras:

BKB: To construct the BKB plasmid (*5'atpB-aadA-3'atpB*) a PCR fragment was amplified from the template plasmid pcp270 (Rochaix, 1978) encompassing the whole atpG gene with primers *atpB*3'cod and *atpB*3'REV, digested with PstI and XbaI and colned into vector pWFB ATG123 (Rimbault et al., 2000) digested with the same enzymes to yield plasmid pWFAUG123B3'. Vector pWFB ATG123 was digested with NotI and the 1985 bp fragment (encompassing the sequences downstram of the petA gene up the the vector polylinker) was cloned into plasmid pWFBAUG123 linearized with NotI. This yield, after checking the right orientation of the insert, plasmid pBKB.

BFR: The pBFR plasmid (*5'atpB-petA-3'rbcL*) was created by a two-step megaprime PCR procedure (Higuchi, 1990): two pairs of primers, petA_FW1/petA_RV and rbcL_FW1/rbcL_RV, were used to amplify two partially overlapping fragments from templates pWF (Kuras and Wollman, 1994) and pBKR respectively. Amplicons were mixed and used as templates in a third PCR with the external primers petA_FW1 and rbcL_RV. The final amplicon was digested by BsrGI and StuI, and cloned into plasmid pBFFF digested with the same enzymes.

BKF: The pBKF plasmid (*5'atpB-aadA-3'petA*) was created by cloning a 386-bp PstI-BamHI fragment amplified from plasmid pWF (Kuras and Wollman, 1994) with primers petA_FW2/petA_RV2 into vector pBKB, digested by PstI and BamHI.

5'BKB: The p5MBKB plasmid (5M*atpB-aadA*-3'*atpB*) was constructed by a two-step megaprime PCR strategy from template plasmid pBKB with primer pairs petA_FW1/5B_RV1 and 5B_FW1/ dBExt_RV. The 978-bp amplicon resulting from the final PCR was digested with EcoRI and BseRI and cloned into vector pBKB digested by the same enzymes to create plasmid p5MBKB.

BRR: To generate the pBMet2R plasmid, the *atpB* promoter and 5'UTR regions was first fused to part of *rbcL* CDS sequence by overlapping PCR using the following primers: AtpBPromPmlSmaI.F/atpB-rbcL b.R and atpB-rbcL b.F/RbcL EcoNI.R, using the pWBKB1

and R15 Bam plasmid templates with the Phusion Taq polymerase (NEB) for the first PCR reactions. The resulting 814 bp fragment was further amplified using the IP-atpB Prom.F and IP-rbcL EcoNI primers, and assembled into the R15 backbone (carrying a 4.4 kb region encompassing the *rbcL* gene and its flanking sequences), previously amplified by the IP-R15 BseRI.R and IP-R15 EcoNI.F2 primers using the In-Fusion PCR Cloning kit (Clontech). Insertion of the recycling *5'psaA-aadA-3'atpB* cassette (a KpnI-SacI blunted fragment of the paAXdB plasmid described in Wietrzynski et al., in preparation) was then inserted at the *Bse*RI restriction site, yielding the pKrBRRR plasmid in which the *aadA* marker is in opposite orientation compared to *rbcL*.

BBR: To construct the pBBR plasmid (*5'atpB-atpB-3'rbcL*) a 741 bp fragment (containing the last 505 bp of the *atpB* coding sequence and the *rbcL* 3'UTR) flanked by the ClaI and KpnI sites at the 5' and 3' ends respectively, was synthetized by Genescript and cloned into the p112 vector (Woessner et al., 1986), digested with the same enzymes.. The resulting plasmid was digested with SacI and XhoI to insert a 5'psaA-driven recycling resistance cassette excised with the same enzyme from plasmid p5'aA-aadA$_{485}$ (Boulouis et al., 2015).

atpB-polyG: To construct plasmid p*atpB*-pG, plasmid p*atpB*-EP1.8 (An EcoRI-PstI fragment of the chloroplast genome of 1.8 kb encompassing the atpB 5'region, sub-cloned into PUC XX), was digested by StyI which cuts 27 bp upstream of the *atpB* translation initiation codon. It was ligated with annealed oligonucleotides D*atpB*G and R*atpB*G. The 1737 bp *Eco*RI-*Cla*I fragment of plasmid pDAAD (Rott et al., 1996) was then replaced by the *Eco*RI-*Cla*I fragment of p*atpB*-pG containing the 18 guanosine G stretch to yield plasmid pDAADG, which also contains a 5'*petD-aadA-3'rbcL* selection marker inserted into the first *Kpn*I site downstream of the *atpB* gene.

***MDB1* minigene:** We recovered from the Kazusa DNA Research Institute an EST clone (AV644102) for gene model Cre14.g614550. Sequencing of the clone confirmed the structure of the gene as shown on the Phytozome 12 browser (V56.5), but revealed that the clone lacked the first 2 exons and part of the third exon. To get a minigene version of the MDB1 gene, cosmid PTQ4126 was digested with SbfI and the 4419 bp fragment was cloned into the AV644102 EST clone digested with the same enzyme. After verification of the correct orientation of the cloned fragment, the resulting plasmid was digested with BamH1 and AclI, filled with Klenow and religated on itself to yield pMiniMDB1 that contains 1287 bp upstream of the translation initiation codon.

Tagged version of the *MDB1* gene: To allow the immune-detection of the MDB1 protein, a HA-tag was inserted after codon 95, in a coil region highly variable among species.

A PCR fragment was amplified using a two steps Megaprime procedure with the external primers MDB1XFW and MDB1XRV and the mutagenic primers MDB1_HA_FW and MDB1_HA_RV. The resulting 1578 bp amplicon was digested with BstEII and SbfI and cloned into plasmid pMiniMDB1digested with the same enzymes to create plasmid pMDB1-HA.

Legends of Supplementary Figures:

Suppl. Fig. S1: complementation of the mdb1 mutants by either BACs or minigene

A) Structure of the midigene

B) Complementation of the *mdb1* mutants by BACs (left, *mdb1-2* panel, right *mdb1-1*) or by the minigene restores the accumulation of the *atpB* mRNA and of the b subunit of chloroplast ATP synthase. The b subunit of the mitochondrial ATP synthase and the *petD* mRNA provide the respective loading controls.

Suppl. Fig. S2: The OPR repeats within MDB1

Alignment of the OPR repeats with the residues obeying the OPR consensus highlighted in grey.

Suppl fig. S3: S1 mapping of the *atpB* 5'ends in the wild-type and mdb1-1 strains.

Suppl. Fig. S4: Conservation of MDB1 among Chlamydomonadales

DNA regions encoding MDB1 orthologues were retrieved from the NCBI database by TBLASTN searches, using CrMDB1 as a query. Gene models were then predicted with the GreenGenie2 software (http://stormo.wustl.edu/GreenGenie2/; (Kwan et al., 2009)) and manually edited to include missing obvious regions of similarity, if required. Alignment of MTHI1 orthologues was performed with the MUSCLE software using default options and manually edited to improve the alignment. The position OPR repeats of the protein from *C. reinhardtii* are shown above the alignments. In the sequence from C. reinhardtii, residues highlighted in yellow were replaced by the HA tag. Additional OPR repeats found in the species-specific insertions are highlighted in yellow or green. Residues conserved in more than half of the sequences are written in red, while conservative substitutions are written in blue.

Abbreviations of species names are as follows:

Chzof: Chromochloris zofingiensis; Chasy: Chlamydomonas asymetrica; Cheus: Chlamydomonas eustigma; Chchl: Chlamydoimonas chlamydogama; ; Cheur: Chlamydomonas euryale; Tesoc: Tetrabaena socialis; Chsph: Chlamydomonas sphaeroides; Gopec: Gonium pectorale; Chsp3 Chsp3: Chlamydomonas sp. 3112; Chdeb: Chlamydomonas debaryana; ChspW: Chlamydomonas sp. WS3; EuspN: Eudorina sp. 2006-703-Eu-15; Chrei: Chlamydomonas reinhardtii; Vocar: Volvox carteri; Yauni: Yamagishiella unicocca; Dusal: Dunaliella salina; Duter: Dunaliella tertiolecta; Chapp: Chlamydomonas applanata; Chlei: Chlamydomonas leiostraca.

Suppl. Fig. S5: Sequence analysis of the RPP-treated sample from wild type.

The position of the matured (5PTE) and precursor 5'end of the *atpB* mRNA are shown. The poor resolution of the sequence downstream of the *atpB* precursor mRNA 5' end suggests heterogeneity of the *atpB* 3'end.

Suppl. Fig. S6: Alternative determination of the *atpB* 5' end by cRT-PCR.

A) Schematic representation of the two PCR strategies. The position of the various primers used are shown. The red segment at the beginning of the *atpB* mRNA represents the processed part of the transcript, only found in the precursor transcript. A hypothetical *atpB* transcript under degradation is also shown by a blue dot. It doesn't contribute to the final amplicon with the classical cRT-PCR strategy (left), but it does with the second cRT-PCR strategy (right).

B) Agarose gels showing the amplicons resulting from the second cRT-PCR strategy, with a molecular weight marker on the right. RPP and mock respectively indicate RNA samples treated or not with RPP, to distinguish precursors from processed transcripts. The band used for cloning experiment is shown as a red rectangle.

C) Schematic representation of the results from the sequence analysis of the 20 clones (details in Suppl. Table ST2). A red segment at the 5'end of the *atpB* mRNA indicates a precursor transcript, while a blue segment at the 3'end of the transcript symbolizes the presence of a polyA tail. The left part shows the contribution of the clones to the amplicon in the two cRT-PCR strategies. In the classical strategy, 7 clones (6 precursors and one processed) would have contributed to the final amplicon, a proportion explaining why its sequence revealed only the precursor form. In the second strategy, 11 precursor and 9 processed transcripts participate to the final amplicon. In the primer extension experiment, one can expect many more partially degraded *atpB* mRNA to contribute to the extension product (all those whose 3'end lies between the primers used for reverse transcription and primer extension - primer PE, above the *atpB* transcript-), resulting in a high representation of the processed form.

Suppl. Fig. S7: Characterization of the BRR transformants

A) Accumulation of Rubisco LSU subunit in the BRR transformants compared to a dilution series of wild-type proteins. Cytochrome f accumulation is shown as loading control. **D**R is a strain bearing a deletion of the chloroplast *rbcL* gene.

B) rbcL RNA accumulation in the BRR transformants compared to WT strain. *psaB* hybridization and ethidium bromide stained gel provide loading controls.

Suppl. Fig S8: Determination of *petA* 5'ends in the Fud50 strain**.**

Supplemental Fig. S9: Conservation of the MDB1 target in *atpB* 5'UTRs.

A) Alignment of *atpB* 5'UTRs

Available chloroplast genomes of Chlorophyceae were scanned for occurrences of the putative MDB1 binding site AAATAAGNGTTAG. In the reported species this sequence was found in the intergenic region upstream of atpB (and almost always only here), at a variable distance (ranging from 152 in *Dunaliella salina* to 1468 in *Palmellopsis texensis*; mean size 550) from the translation initiation codon. Sequences were aligned with the MUSCLE software, using default options, and then manually edited to improve the alignment. Residues conserved in more than half sequences are written in red, while conservative substitutions are written in blue. A putative -10 Pribnow box found upstream of the putative MDB1 binding site in some species is written in bold and underlined. It was always found at more than 10 nt of the MDB1 binding site, suggesting that the mature *atpB* mRNA is a processed transcript in these species as well.

Abbreviations of species names are as follows:

Chlei: Chlamydomonas leiostrac; Voafr: Volvox africanus; Chacu: Chariaciochoris aciminata; Chtat: Chlorococcum tatrense; Dusal: Dunaliella salina; Chrei: Chlamydomonas reinhardtii; Gopec: Gonium pectorale; Phlen: Phacotus lenticularis; Plsta: Pleodorina starrii; Yauni: Yamagishiella unicocca; Eu_sp: Eudorina sp.; Tesoc: Tetrabaena socialis; ChspU: Chlamydomonas sp. UWO 241; Botex: Borodinellopsis texensis; Loseg: Lobochlamys segnis; Mimon: Micronegla monida; Chdeb: Chlamydomonas debaryana; Chsp3: Chlamydomonas sp3212; Chsph: Chlamydomonas sphaeroides; Halac: Haematococcus lacustris; Chapp: Chlamydomonas applanata; Chsti: Chlorosarcina stigmatica; Prbot: Protosiphon botryoides; Chper: Chloromonas perforata; patex: Palmellopsis texensis.

All species belong to the Chlamydomonadale order.

Supplemental Table ST2: The ends of the clone circularized *atpB* transcripts from Suppl. Fig S6

Clone number	5'end/TSS	3'end/TSS	length	Remarks
1	+27	+1448	1065	
2	+3	+1905	1522	
3	+3	+1602	1219	
4	+27	+1314	931	
5	+27	+1465	1082	polyadenylated
6	+27	+1293	910	polyadenylated
7	+1	+1905	1522	
8	+27	+1570	1187	
9	+1	+1397	1014	
10	+27	+1905	1522	
11	+1	+1330	947	
12	+27	+1396	1013	polyadenylated
13	+1	+1927	1544	
14	+1	+1330	947	
15	+1	+1886	1503	
16	+27	+1502	1119	
17	+1	+1905	1522	
18	+1	+1697	1314	
19	+27	+1192	809	
20	+1	+1386	1003	

Supplemental References

Boulouis, A., Drapier, D., Razafimanantsoa, H., Wostrikoff, K., Tourasse, N.J., Pascal, K., Girard-Bascou, J., Vallon, O., Wollman, F.A., and Choquet, Y. (2015). Spontaneous dominant mutations in chlamydomonas highlight ongoing evolution by gene diversification. Plant Cell **27,** 984-1001.

Drapier, D., Rimbault, B., Vallon, O., Wollman, F.A., and Choquet, Y. (2007). Intertwined translational regulations set uneven stoichiometry of chloroplast ATP synthase subunits. EMBO J **26,** 3581-3591.

Higuchi, R. (1990). Recombinant PCR. In: PCR protocols: a guide to methods and application. (London, New York: Academic press).

Kuras, R., and Wollman, F.A. (1994). The assembly of cytochrome b6/f complexes: an approach using genetic transformation of the green alga Chlamydomonas reinhardtii. EMBO J **13,** 1019-1027.

Kuras, R., de Vitry, C., Choquet, Y., Girard-Bascou, J., Culler, D., Buschlen, S., Merchant, S., and Wollman, F.A. (1997). Molecular genetic identification of a pathway for heme binding to cytochrome b6. J Biol Chem **272,** 32427-32435.

Kwan, A.L., Li, L., Kulp, D.C., Dutcher, S.K., and Stormo, G.D. (2009). Improving gene-finding in Chlamydomonas reinhardtii:GreenGenie2. BMC Genomics **10,** 210.

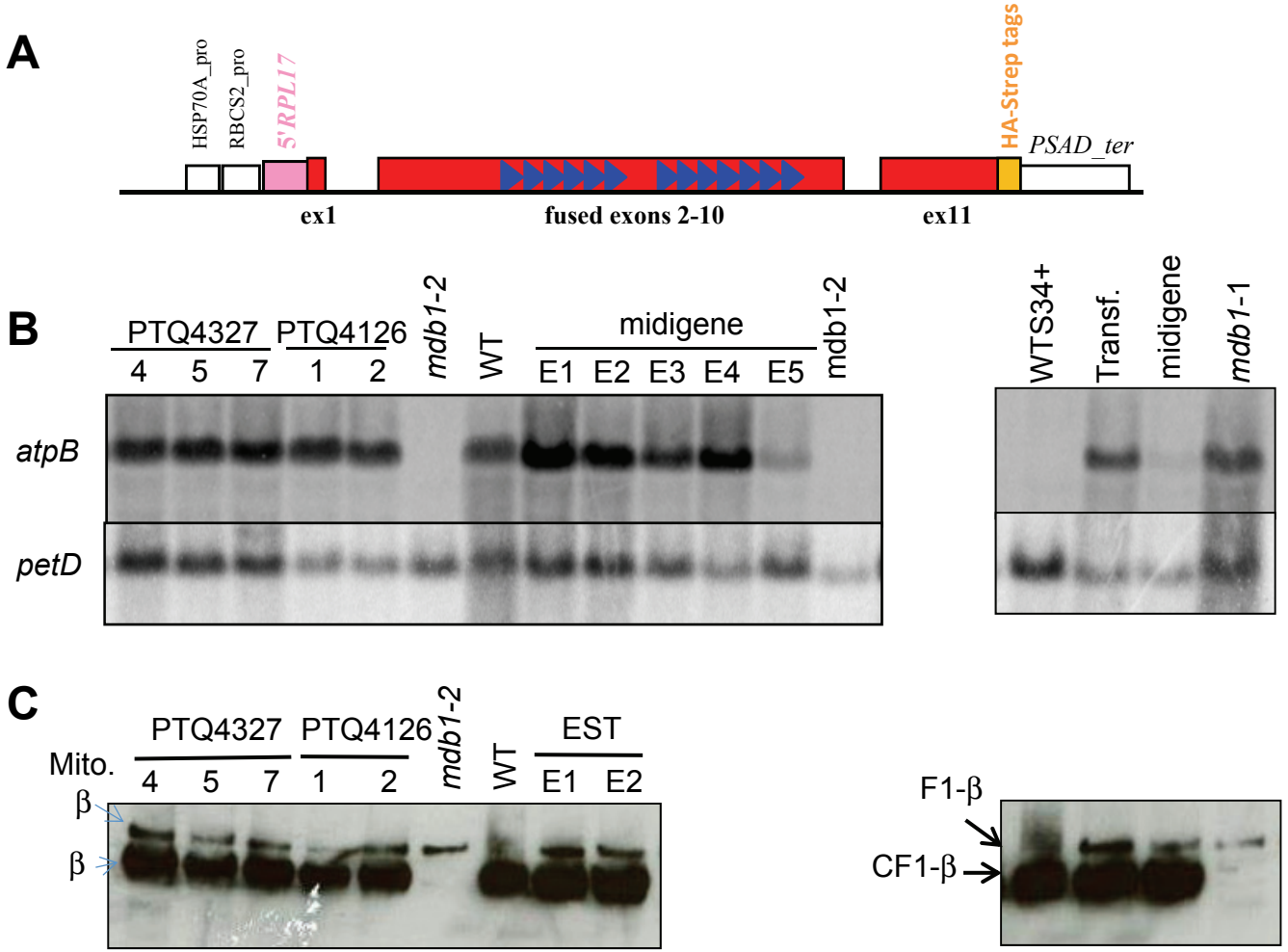
Rimbault, B., Esposito, D., Drapier, D., Choquet, Y., Stern, D., and Wollman, F.A. (2000). Identification of the initiation codon for the atpB gene in Chlamydomonas chloroplasts excludes translation of a precursor form of the beta subunit of the ATP synthase. Mol Gen Genet **264,** 486-491.

Rochaix, J.D. (1978). Restriction endonuclease map of the chloroplast DNA of Chlamydomonas reinhardii. Journal of Molecular Biology **126,** 597-617.

Rott, R., Drager, R.G., Stern, D.B., and Schuster, G. (1996). The 3' untranslated regions of chloroplast genes in Chlamydomonas reinhardtii do not serve as efficient transcriptional terminators. Mol Gen Genet **252,** 676-683.

Woessner, J.P., Gillham, N.W., and Boynton, J.E. (1986). The sequence of the chloroplast atpB gene and its flanking regions in Chlamydomonas reinhardtii. Gene **44,** 17-28.

Cavaiuolo et al, Suppl. Fig. S1

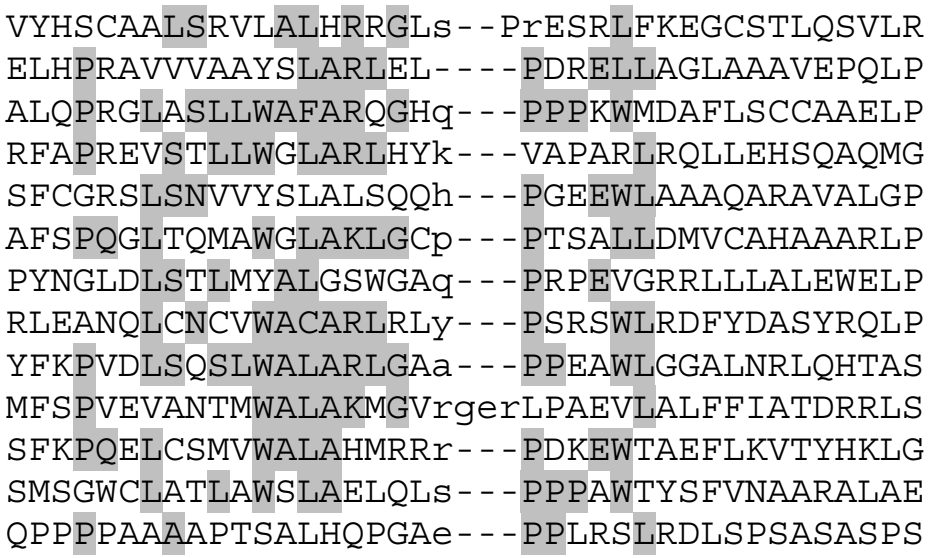


Suppl. Fig. S1: complementation of the mdb1 mutants by either BACs or minigene

A) Structure of the midigene

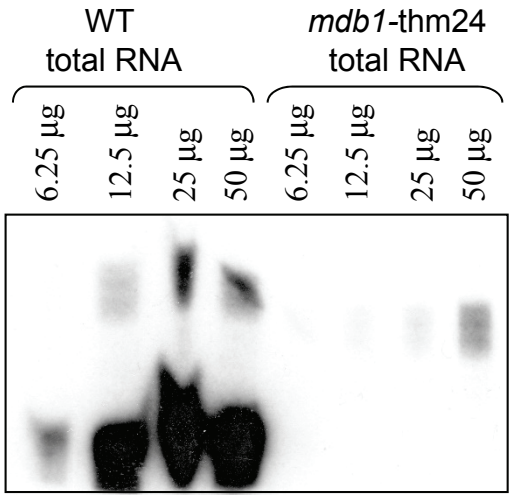
B) Complementation of the *mdb1* mutants by BACs (left, *mdb1-2* panel, right *mdb1-1*) or by the minigene restores the accumulation of the *atpB* mRNA and of the β subunit of chloroplast ATP synthase. The β subunit of the mitochondrial ATP synthase and the *petD* mRNA provide the respective loading controls.

Cavaiuolo et al, Suppl. Fig. S2
3



Suppl. Fig. S2: The OPR repeats within MDB1
Alignment of the OPR repeats with the residues obeying the OPR consensus highlighted in grey.

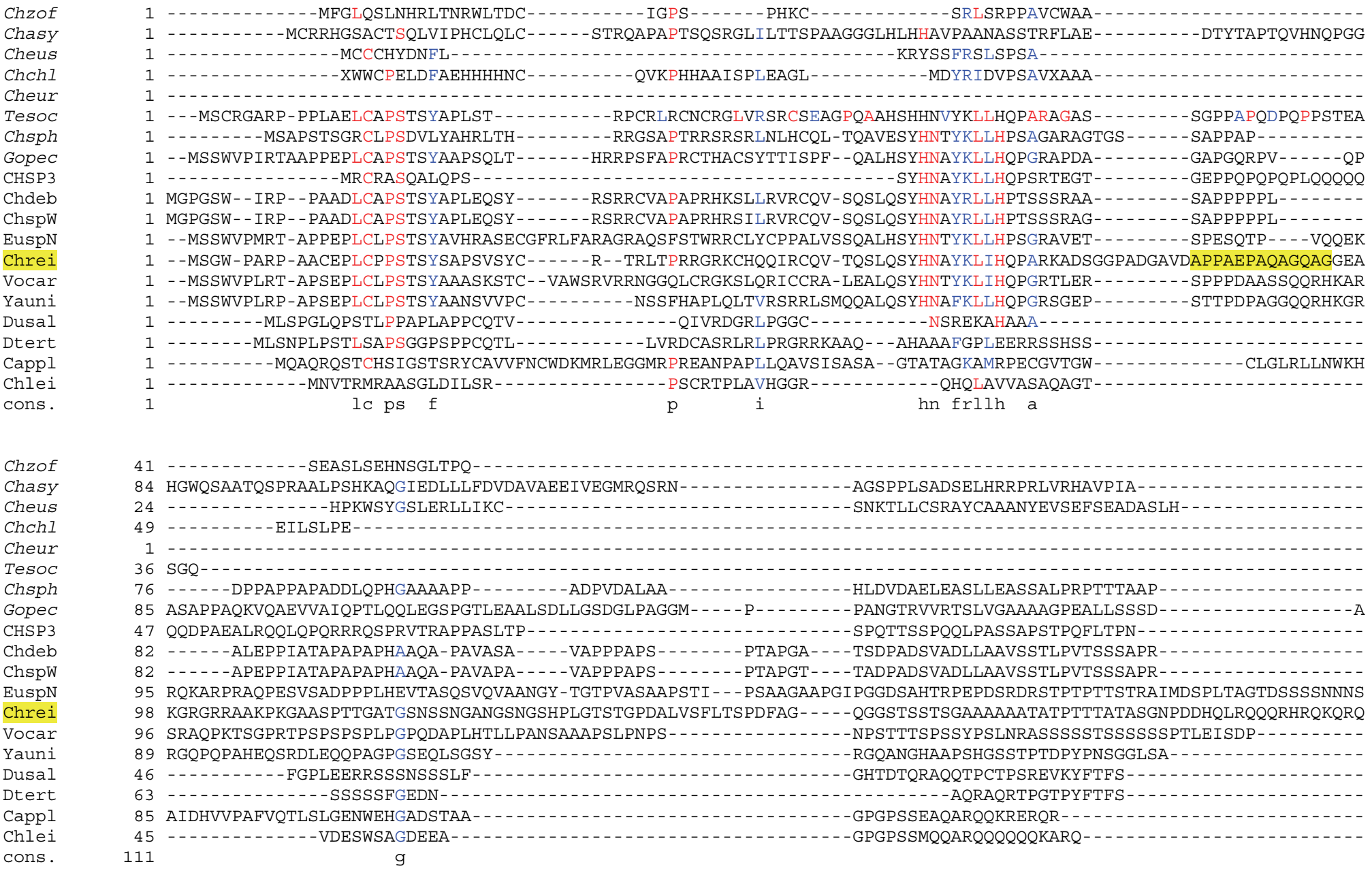
Indicate sizes,
see Blandine thesis



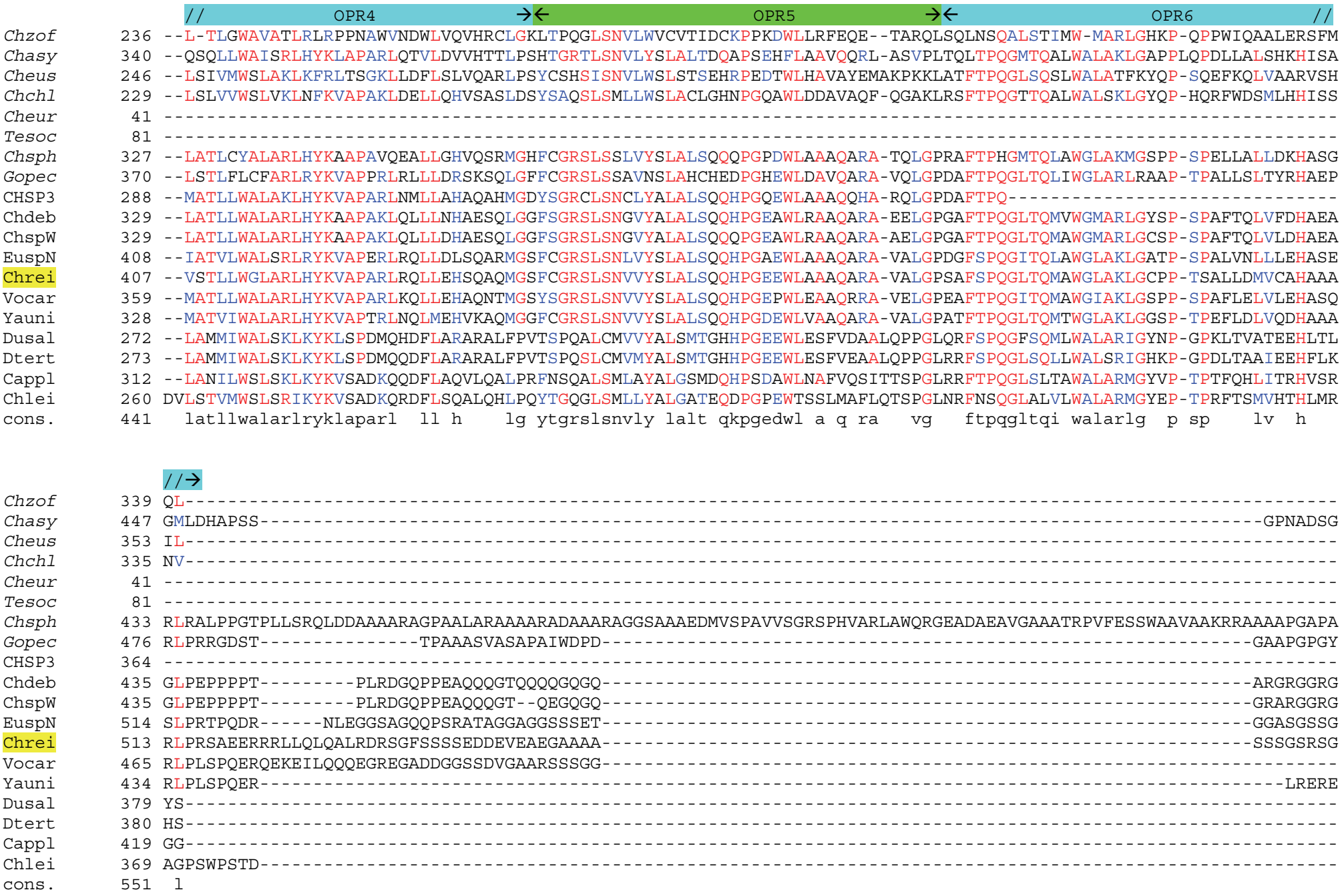
S1 nuclease protection

Suppl fig. S4: S1 mapping of the *atpB* 5'ends in the wild-type and mdb1-1 strains.

Suppl. Fig. S4: Conservation of MDB1 among Chlamydomonadales

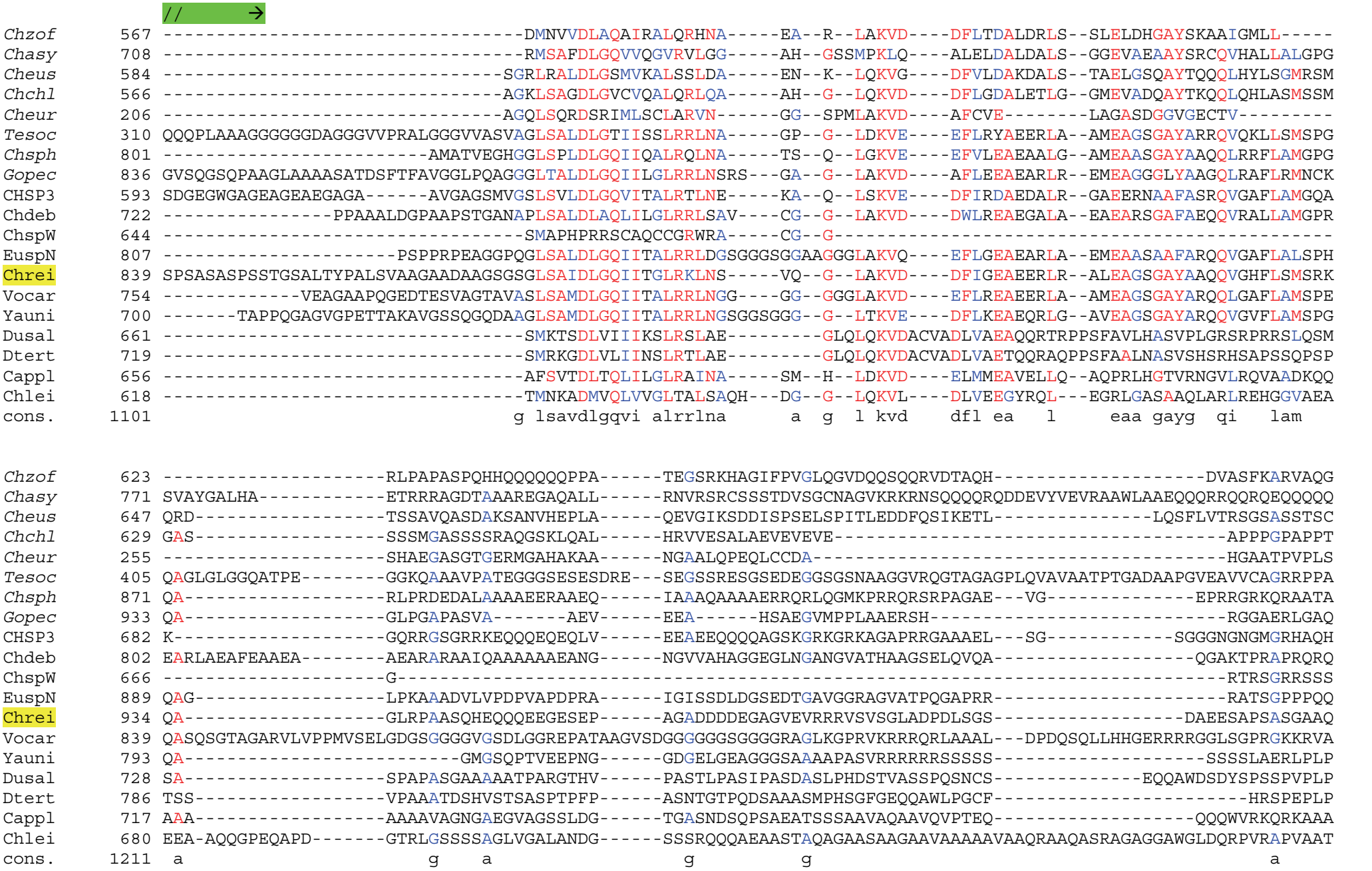


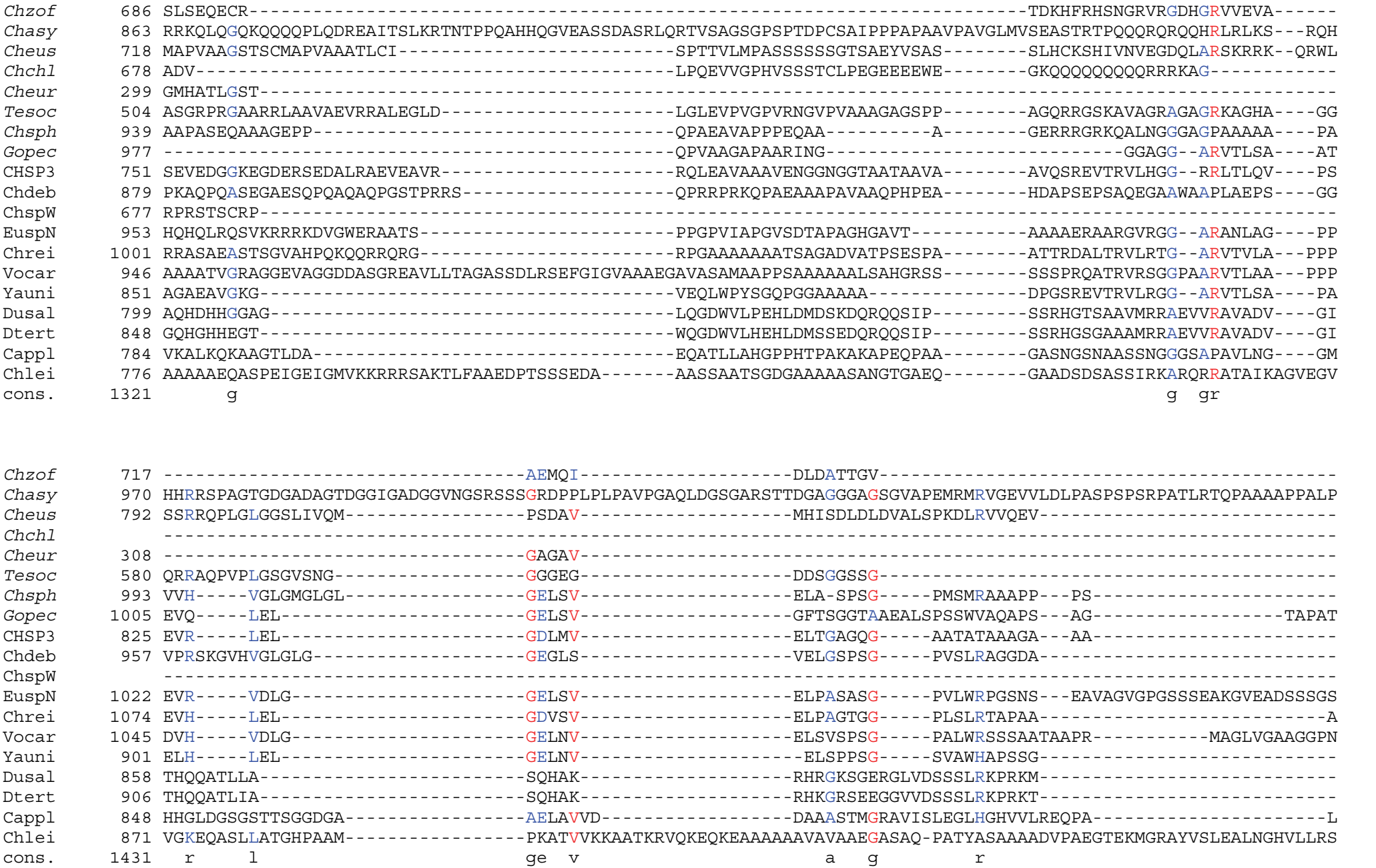

OPR1
OPR1
OPR2
OPR3
Chzof 55 ----------QQGTQQPAAQGSRSRPPAAIRPQHDGKFQRRLTAKIKHAKDLYAIRELMHIHEGCLNHIHVNAIITHVAQLADSR-QLRP--
Chzof 56 ----------QGTQQPAAQGSRSRPPAAIRPQHDGKFQRRLTAKIKHAKDLYAIRELMHIHEGCLNHIHVNAIITHVAQLADSRQ-LRP--RD
Chasy 157 ----------SPNEASSSAGANGDQQHLIQQPQVSNRRLAFAQALARASSVRELSTMFTGNASEFDSSHLVAFLTRLQELHEQHL-VRP--RE
Cheus 70 ----------VRRGTHQTARPRIEVSSWSPPQ---DDAKAWTSLINSCKNPAQLQRCVNGLQT-WNKYHLTAAFSRLASFCRG-----PDSYH
Chchl 56 ----------TKVGIRNFKKALKRPVVYT---------NTLTSRLAAVDSPQELQKVLSEAEPVLNHIHLCAAFHRASVLARRPQ-LTP--RE
Cheur 1 ----------
Tesoc 36 ----------
Chsph 135 ----------PPGGAWPPPGSSPFPGSRPYTSHSVRLHDWTPPPTALAAKTALTRALATSASYTRLHSLIGSHGSEFNTFHTTAALTRMVALHRAG--LSP--RE
Gopec 163 SASALSRKGRRSGPARGVGASTTPATPAEPDLRPLTPHTPDLTLWQPPPSSLHARTALTRAIATAPSYTRLHALIADHASDFNTHHACAALNRMLVLHRAG--FTP--RE
CHSP3 106 ----------YSILATDPLPQPAHLPSWQPPPTSLHHRTALTHAITNAPSYTALHQLIIDNAAEFNVYHTCASLSRVLALHARG--FSP--RE
Chdeb 149 ----------GGDPEKPYTRHSVSLTAWQPPPTSLQRRSALTRALLTSPSYARLHALLLDNAADFNVHHTTAALSRVLQLQREG--LSP--RE
ChspW 149 ----------GGDPEKPYTRHSVSLPSWQPPPTSLQRRSALTRALLTSPSYARLHALLLDNAADFNVHHTTAALSRVLQLQREG--LSP--RE
EuspN 201 GGGSSSNLYNGANPIGTAATAGGPPRRDAGDQPPFIPHNMYLRNWQPPATSLQSRTALTRAITTCPSYTRLHQLILDNASEFNAYHICAALSRVLKLHAAG--LTP--RE
Chrei 203 QQQLQTATGGDAGEPHSSSSASFG---GGGDMRPYSRHSVALPAWQPPPSSLQARTALTRAISTCPTYTRLHQLLLDNALDFNVYHSCAALSRVLALHRRG--LSP--RE
Vocar 179 ----------EAEDLRPYTPHTVHLPHWQPPPTSLQHRTALTRAITSCPTYTRLHQLILENAADFNCYHTCAALSRVLALHGAG--LTP--RE
Yauni 148 ----------SAAEHRPYTPHSVSLPQWQPPPTSMQHRTALTRAITTCASYTRLHQLIVDNAADFNAYHTCAALSRLMVLHTAG--LTP--RE
Dusal 88 ----------EQEGTSSEPDQPVHELQWEPPRR--QDESQLTLFITGCESTSSLLELLLHNQPRLNGIHLSAALDRTIKLYRQEA-QRPSWRT
Dtert 89 ----------EQEGMSSSQDHPAEELHWEPPSR--QDHSQLTLSITGCDSTSSLLELLLHNQPRLNGIHLSAAFDRTIKLYRQEA-QRPSWRT
Cappl 132 ----------QRGNVRPSSAPAATLEAWRAPK---LDHSLLTSLIKSSQNCSSLLELLQQYQKQLNAINVVAALHRTVELYRQGASLRP--LE
Chlei 78 ----------RRSKAEPVRTERVPMQ-WQAPR---LNNQALTMLIVEAPDCTSLLEILQQNQNNLNHINVVAALNRTTKLYHSGR-IRP--MD
cons. 221 hp i l w pp rtalt ai s s t lr lm na efn h aaltrvv l r g ltp re
Chzof 136 MQLLPALL-PSILDLVVQQI--PNY---DRSIANTLHALARLELEERALVSKLLAAAEPLLQTFSSQGLANTAWALAKMSYVPSSGICGLLFTMSGYHMVHFNPQE----
Chasy 237 ARLLYNTV-PRVMALVRSLAAAGAL--DTRSIPQTAYLLARLDMYDREAIAALEAAAEPLMGAMAPPGLASLLWALAKLDHAPPARWMEALITAAFIKLQSFKPKE----
Cheus 144 ARVASSRVLPALITRAKENL--PRL--QARSLATVAHCVGAYEYKDKELMAELAKISEQEFANFQPQGLSNLIWSFARLEVQPSQRWMDAFLQACVSSLGTFKPQE----
Chchl 127 ARIISSQTIPNLITLVRPRV--QEL--NARAVSTVAHCLGVVEHRDRDLMSDLGQRAEAIMADFTTQGLANTIWGFAKVGVQPSARWMDSFLGMVHTKLHDFRPQE----
Cheur 1 ----------KLVDALSMAAGAKMRAMAPLHLAKAAWAVARLRHAPGRRW----------
Tesoc 38 RGLVRSRC----------SEA--GPQA----------AHSHHNVYKLLHQPARAGASSGPP----------APQD----
Chsph 226 SRLFRDGC-SAMQAILRRQL--PEL--SPRCVVVAAHCLAKLQLADRELLPGLAAAVEDQLSLLQPQGLVSLLWSFAAQGHPPSPRWMQLALGTAMGRLGAYSPAD----
Gopec 269 ARLFKEGC-SAVQSALRRLV--GEL--EPRSVVAVAYGLARLELPDRELLAGLLEASQQQLSKMKPQGLASLLWSFSRLNVQPPAKWMDAFLSACASELDGFGPRD----
CHSP3 185 ARLFKEGC-SALQGVLRRQLQQSEL--HPRAVVVAAHSLAKLELPDRELLAGLAGAVEPQLHVLQAQGLSSLLWSFARQGHQPPPRWMESFLACCAADLPAFAPHH----
Chdeb 228 ARLFKEGC-STLQTVLRRQV--PDL--SPRAIVVAAHALAKLELPDRELLPALAAAVEPQLRALQPQGLSTLLWAFAAQGHQPPPKWMDSYLAAAAATLPGWGDRD----
ChspW 228 ARLFKEGC-STLQTVLRRQV--PDL--SPRAIVVAAHALAKLELPDRELLPALAAAVEPQLRALQPQGLSTLLWAFAAQGHQPPPKWMDSYLAAAAATLPGWGDRD----
EuspN 307 SRLFKEGC-STLQAIMRRQL--PEL--QPRALVVAAYSLARLELPDRELLAGLASATEPHLSALKPQGLCSLLWAFARQNHQPSPKWMDGALSACAADLDAFAPRD----
Chrei 306 SRLFKEGC-STLQSVLRRQL--TEL--HPRAVVVAAYSLARLELPDRELLAGLAAAVEPQLPALQPRGLASLLWAFARQGHQPPPKWMDAFLSCCAAELPRFAPRE----
Vocar 258 SRLFKEGC-SMLQTILRRQV--SEL--QPRALVVAAYSLARLELPDRELLGGLAAAVEPHLSALQPQGLSSLLWAFARQSHQPSPKWMDALLSAAAADLATFSPRD----
Yauni 227 SRLFKEGC-STMQTILRRQL--PEL--QPRAVVMAAYSLARLELPDRELLAGLAAAVEPHLSALKPQGLSSLIWAFARQGHQPSPKWMDALLSAVAADLEAFGPRD----
Dusal 168 ARMITGALLPLLTSELQALM--PNLANQPRVVASIAHSLGCLDVRDRDLLGGLAALAQGCMPEMSTQGLSNILWAFARCEYQPSAGWMSAYVCACRARIASFRPQE----
Dtert 169 ARMITGALLPLLTSELQTLM--PDLANQPRVVTSIAHSLGCLDVRDRDLLGGLAALAQGCMPELSTQGLANILWAFARCEYQPSAGWMSAYVCACRARIAAFRPQE----
Cappl 210 ARVVTGQILPLLLAEVRNMM--QLL--QPRAIATIVYSLGSLDVRDRELLVELAQRAEPQLADFTTQGLSNMLWAFARCGYQPPARWMDGFVTVVHAKRTRLAPQE----
Chlei 154 SRIISGQILPLLVSDVRNRM--ALF--KARGVATVAHALGSLDVRDKELFSSLASQAERQLPDFTPQGLSNMLWAFARCGYQPPTRWMDAYVSTCHAKLAQFGXAGAVHS
cons. 331 arl r gc l tllr qi el prai ah larlel drellaalaaa ep l m pqgl sllwafar ghqp rwmdafl a a l fap e


// OPR4 → ← OPR5 → ← OPR6 //
Chzof 236 --L-TLGWAVATLRLRPPNAWVNDWLVQVHRCLGKLTPQGLSNVLWVCVTIDCKPPKDWLLRFEQE--TARQLSQLNSQALSTIMW-MARLGHKP-QPPWIQAALERSFM
Chasy 340 --QSQLLWAISRLHYKLAPARLQTVLDVVHTTLPSHTGRTLSNVLYSLALTDQAPSEHFLAAVQQRL-ASVPLTQLTPQGMTQALWALAKLGAPPLQPDLLALSHKHISA
Cheus 246 --LSIVMWSLAKLKFRLTSGKLLDFLSLVQARLPSYCSHSISNVLWSLSTSEHRPEDTWLHAVAYEMAKPKKLATFTPQGLSQSLWALATFKYQP-SQEFKQLVAARVSH
Chchl 229 --LSLVVWSLVKLNFKVAPAKLDELLQHVSASLDSYSAQSLSMLLWSLACLGHNPGQAWLDDAVAQF-QGAKLRSFTPQGTTQALWALSKLGYQP-HQRFWDSMLHHISS
Cheur 41 ----------
Tesoc 81 ----------
Chsph 327 --LATLCYALARLHYKAAPAVQEALLGHVQSRMGHFCGRSLSSLVYSLALSQQQPGPDWLAAAQARA-TQLGPRAFTPHGMTQLAWGLAKMGSPP-SPELLALLDKHASG
Gopec 370 --LSTLFLCFARLRYKVAPPRLRLLLDRSKSQLGFFCGRSLSSAVNSLAHCHEDPGHEWLDAVQARA-VQLGPDAFTPQGLTQLIWGLARLRAAP-TPALLSLTYRHAEP
CHSP3 288 --MATLLWALAKLHYKVAPARLNMLLAHAQAHMGDYSGRCLSNCLYALALSQQHPGQEWLAAAQQHA-RQLGPDAFTPQ----------
Chdeb 329 --LATLLWALARLHYKAAPAKLQLLLNHAESQLGGFSGRSLSNGVYALALSQQHPGEAWLRAAQARA-EELGPGAFTPQGLTQMVWGMARLGYSP-SPAFTQLVFDHAEA
ChspW 329 --LATLLWALARLHYKAAPAKLQLLLDHAESQLGGFSGRSLSNGVYALALSQQQPGEAWLRAAQARA-AELGPGAFTPQGLTQMAWGMARLGCSP-SPAFTQLVLDHAEA
EuspN 408 --IATVLWALSRLRYKVAPERLRQLLDLSQARMGSFCGRSLSNLVYSLALSQQHPGAEWLAAAQARA-VALGPDGFSPQGITQLAWGLAKLGATP-SPALVNLLLEHASE
Chrei 407 --VSTLLWGLARLHYKVAPARLRQLLEHSQAQMGSFCGRSLSNVVYSLALSQQHPGEEWLAAAQARA-VALGPSAFSPQGLTQMAWGLAKLGCPP-TSALLDMVCAHAAA
Vocar 359 --MATLLWALARLHYKVAPARLKQLLEHAQNTMGSYSGRSLSNVVYSLALSQQHPGEPWLEAAQRRA-VELGPEAFTPQGITQMAWGIAKLGSPP-SPAFLELVLEHASQ
Yauni 328 --MATVIWALARLHYKVAPTRLNQLMEHVKAQMGGFCGRSLSNVVYSLALSQQHPGDEWLVAAQARA-VALGPATFTPQGLTQMTWGLAKLGGSP-TPEFLDLVQDHAAA
Dusal 272 --LAMMIWALSKLKYKLSPDMQHDFLARARALFPVTSPQALCMVVYALSMTGHHPGEEWLESFVDAALQPPGLQRFSPQGFSQMLWALARIGYNP-GPKLTVATEEHLTL
Dtert 273 --LAMMIWALSKLKYKLSPDMQQDFLARARALFPVTSPQSLCMVMYALSMTGHHPGEEWLESFVEAALQPPGLRRFSPQGLSQLLWALSRIGHKP-GPDLTAAIEEHFLK
Cappl 312 --LANILWSLSKLKYKVSADKQQDFLAQVLQALPRFNSQALSMLAYALGSMDQHPSDAWLNAFVQSITTSPGLRRFTPQGLSLTAWALARMGYVP-TPTFQHLITRHVSR
Chlei 260 DVLSTVMWSLSRIKYKVSADKQRDFLSQALQHLPQYTGQGLSMLLYALGATEQDPGPEWTSSLMAFLQTSPGLNRFNSQGLALVLWALARMGYEP-TPRFTSMVHTHLMR
cons. 441 latllwalarlryklaparl ll h lg ytgrslsnvly lalt qkpgedwl a q ra vg ftpqgltqi walarlg p sp lv h
//→
Chzof 339 QL----------
Chasy 447 GMLDHAPSS----------GPNADSG
Cheus 353 IL----------
Chchl 335 NV----------
Cheur 41 ----------
Tesoc 81 ----------
Chsph 433 RLRALPPGTPLLSRQLDDAAAARAGPAALARAAAARADAAARAGGSAAAEDMVSPAVVSGRSPHVARLAWQRGEADAEAVGAAATRPVFESSWAAVAAKRRAAAAPGAPA
Gopec 476 RLPRRGDST----------TPAAASVASAPAIWDPD----------GAAPGPGY
CHSP3 364 ----------
Chdeb 435 GLPEPPPPT---------PLRDGQPPEAQQQGTQQQQGQGQ----------ARGRGGRG
ChspW 435 GLPEPPPPT---------PLRDGQPPEAQQQGT--QEGQGQ----------GRARGGRG
EuspN 514 SLPRTPQDR------NLEGGSAGQQPSRATAGGAGGSSSET----------GGASGSSG
Chrei 513 RLPRSAEERRRLLQLQALRDRSGFSSSSEDDEVEAEGAAAA----------SSSGSRSG
Vocar 465 RLPLSPQERQEKEILQQQEGREGADDGGSSDVGAARSSSGG----------
Yauni 434 RLPLSPQER----------LRERE
Dusal 379 YS----------
Dtert 380 HS----------
Cappl 419 GG----------
Chlei 369 AGPSWPSTD----------
cons. 551 l


OPR7
OPR8
//
Chzof 341 ----------GTCNAHDLSTLLVALAVMDFRP--GSTWHRAVMLRAKALLPMFTIRQVSNIFWACARE------ITPQELW----------
Chasy 463 RGAQRAAAPRHRVQRYTGIDISTALYALAKLGQSPLLPHQTLVGVLRALVQMMPTMAPNHLANVTWALARSNT----PRPPGAW----------
Cheus 355 ----------LSCNSIDLATLTYAYAALRMPP--DEQLFLRLQKASLKQMHELLPSHLAKTTWAFAKLG-----LVPSDAW----------
Chchl 337 ----------RSYKGIDLATALYAFARLDVHP--SPKVRALLEHTSRQEMWHLEGAHLANILWAFAKLD-----TLPSKRW----------
Cheur 41 ----------
Tesoc 81 ----------
Chsph 543 DPAFPAFTRGPLLPAYNGVDLVILLYAMALWGCPP--QPLLARRVLLRLELELPSLGVQHLAMCLWSVARLR-----LFPSRSW----------
Gopec 510 RQLRGPYGRSKWVAAYKGMDLAVLLYVLAVWGVRL--QPAFGRQLLAAVQMKLPELEPNQLCNCLWACARLR-----LFPTRSW----------
CHSP3 364 ----------MEMPRMESNQICNCLWACARLR-----VQPPAVW----------
Chdeb 475 Q--------RRSAQRYNSQDLSMLLYSLAVWGAAP--PAALARKLMLAVQYALPRLETHQLCNCLWSCARLG-----LYPSASW----------
ChspW 473 Q--------RRSAQRYNGQDLSTLLYSLAVWGAAP--PAALARRLMLAVQYALPRLETHQLCNCIWSCARLG-----LYPSASW----------
EuspN 557 GVAAGKR--PRARGRYTGVDLSTLMYVLACFGARP--PAELGRRLLSAVQWALPSLEPNGMCNCLWACARLR-----LYPYKMW----------
Chrei 562 RKQQQQQQPRRPLAPYNGLDLSTLMYALGSWGAQP--RPEVGRRLLLALEWELPRLEANQLCNCVWACARLR-----LYPSRSW----------
Vocar 506 ----SHLQQRRRRERYSGLDLATLLYSLASLGAQP--RADLGRRLLAAVQWELPTLEANGMCNCLWACARLR-----IFPTKMW----------
Yauni 448 QQREGGEGRRRRLGRYNGVDLSTFMYSLASLGAQP--RVDLSRKLLAALQWELPSMEANQMCNCLWACARLR-----LYPHKTW----------
Dusal 381 ----------RMKGYKSIDVATTLYAVARMQLPM--SRNLLRLLLDQVDKHMFSFQPSQLANVGWALARLQLH---LRQYRQWQQSVHQREEQQQQQQQQQQQQQQQAA
Dtert 382 ----------RMQGYKSIDVSTTLYAMARMQLPM--SRNLLRLFLDQVEKHMFSFQPSQLANIGWALARLQMQ---LRQLRQWQQSAHQDTEQQQQQQQQQQQQQQQQQ
Cappl 421 ----------LQAFKPIDVATMVYAMARMGMPL--SKELHASFQQALQGAAGGMQPNQLANSTWALAHFYTMDPERAPRTSL----------
Chlei 378 -------SLPPRIAACNGIDVATIMYGMGRMHMHV--TPNLLTALLIKLKHEVPGLDPAQMSNVVWALAQFQVKNPLLRVHPSF----------
cons. 661 y gidl tllyala mg p l r ll al lp le ql n lwa arl l p w
//
OPR9
//
Chzof 404 ----------IARALRKIYHGLQDTNAQDLANSLWALAHLNV
Chasy 543 ----------VAVLLAAVANHMAQLRPADLSQLGWALASLGI
Cheus 419 ----------LRNLLSCSFQLMDIFNARDLSMFGWGLAKLGV
Chchl 401 ----------MLKALTVIYTIMPSLKPRDLAMVGWALAKLRV
Cheur 41 ----------LYRLLSESYRCLPAFEARDVAMLAWALATMRV
Tesoc 87 ----------//WALARLRA
Chsph 620 ----------LVHAYDATYRQLHAFKPQDLSMTLWALARLQA
Gopec 587 ----------LVVFFHASFRSLRLFKPGDLAQCLWALGRLQS
CHSP3 393 ----------LRTFFDASYRQLPFFKPGDLSQSLWALARLGV
Chdeb 544 ----------LREWYDASYRQLPYFKPVDLSQSLWGLARLSA
ChspW 542 ----------LRDWYDASYRQLPYFKPVDLSQSLWGLARLSA
EuspN 632 ----------LSVFFDASYRQLPYFKPVDLSQTLWALARLGA
Chrei 639 ----------LRDFYDASYRQLPYFKPVDLSQSLWALARLGA
Vocar 579 ----------LSAFFDASYRQLPYFKPVDLSQTLWALARLQA
Yauni 525 ----------LKAFFDASYRQLPYFKPVDLSQSLWALARLQA
Dusal 475 YSTDQEQQNWQGHAHYEQQHLQHVVASQSPETLLDLNQQHHPSSSSTEVHSLPSSGQQQQQQQQQQQQQQDFPLPTALLRKYLAACYGSLDRFTSLDLSMTCWSMATMRV
Dtert 476 QASQLLDQEEQQQKLQGQMHHEQQHLLHMDVSGSPEKLINSRQYNDYSSSSSSSSSTAAHSVPTAGQQQLDYSLPTSILKRYLAACYASLDRFTSLDLSMTCWALATMRV
Cappl 491 ----------LSRILHASYVRMDRFSTHDLSQLAWALAKMEV
Chlei 453 ----------LAQFFSMAYLRLDRFTHHDLAMTSWALAKMRV
cons. 771 l fy asyr l fkp dlsqslWalarl v


// OPR9 → ← OPR10 → ← OPR11 → ← //
Chzof 436 HPSAGWMSLFSAQAQAIAGQFKSQEIANTIWAYARLRTKPKA----LLSALFQGANHRLSAFKPAELSSLMWALAKLHIVPSKEWKEEFLQASYHKIGAMSPQSLSNVIW
Chasy 575 APPAPWLRAYEARVATSARLFAPREVAAVVWALSRLG---GELPGEVLAEFFDATDRRLSSFSGPELGCMVTSLARLRVQPHKEWMDEFIKVTFHKLAAMGPQELANIGW
Cheus 451 VLPESFAQRYVRRIEAVAGEFPPQEVANTLWALACFNIRPSS---VLVAHFFDATDQRLSSFKASELSHMLWALADRRCVLDAQWINEYLKVSFLRMAEFSPQGLANMIW
Chchl 433 QPPRTLLLAYVRRVEVLAGEFKPQEVANTVWALARFGLQPSS---SLLVEFFVATDQRLSSFKPVELNQMLWALAKVRTTPDKPWVEEFLKVTFHKLPDFSGQGLANMCW
Cheur 73 SLPDAFVKPLVARAEALAEEFPPQEVANTLWAFSRLGVEPSH---ALLEHFFESTDHRLSEFKPMELSQVLWALGRSRSQLERAWVNELLQVLLVRLPELSPHGLSSVVW
Tesoc 104 APPDPWLGAALHRLQSSASMFSPVEVASTMWALAKLDVSGERLPSEVLALFFIATDRRLSSFKPQELCCMAWALARMRRRPDKEWTAEFLKVSYHKLSSMNGWCLATLSW
Chsph 652 VPPTPWLAAAVSRVEAVATMFNPVEVANTLWALASLGVRGESLPGEMLALFFMATDRRLSSFKQQELSSMIWALATMMRRPDKEWAAEFLKVTYVRLPSMGGWSLATIAW
Gopec 619 HPPAEWLAAVLTRLQLTASMFSPVEVATTMWALAALGVRGQQLPGEVLALWFIATDRRLSSFRPAELVSMVWALARMGRRPDKEWSAELLKVTYHKLGAMNGWCLGVLAW
CHSP3 425 APPGPWLTAAMTRLQQTASMFQPVEVSQTMWALARLGVRGEALPSEALALFFIATDRRLSGFKPQELCAMAWALARMRRRPDKEWAAEFLKVTYHKVPSMGGWCLATLAW
Chdeb 576 VPPQPWLAAALTRLQHAASMFTPVEVASCLWALAKLGVAGERLPGEVLALFFIATDRRLSSFKPQELCSMLWALARMRRRPDKEWTAEFLKASFHKLPSMSGWCLGTLAW
ChspW 574 VPPQPWLAAALTRLQHAASMFTPVEVASCLWALAKLGVAGERLPGEVLALFFIATDRRLSSFKPQQ----------
EuspN 664 APPRAWMASVITRLHRSASEFSPVEVATTLWALARLGVRAEQLPTEVMVLFFIATDRRLSSFKPQELSSMVWGLARMRWRPDKEWTAELLKVTYHKLGTMSGRCLATIAW
Chrei 671 APPEAWLGGALNRLQHTASMFSPVEVANTMWALAKMGVRGERLPAEVLALFFIATDRRLSSFKPQELCSMVWALAHMRRRPDKEWTAEFLKVTYHKLGSMSGWCLATLAW
Vocar 611 APPPAWVASVMVRLQHSATMFSPVEVATTMWALAKLGVRGRQMPGEVLALFFIATDRRLSSFKPQELCSMILALAHMRRRPDKEWAAEFLKVTYHRLASMSGWCLATVAW
Yauni 557 LPPDPWVAAVLTRLQHTASMFTPVDVASTMWALARLGVRGDRLPAEVLALFFIATDRRLSSFKPQELCSMIWALAHMRRRPDKEWTAEFLKVTYHKLATMNGWCLATLAW
Dusal 585 TPPPSFLSVFLRRVEQ----------MGFTPDKAWTEEFLHATFHKLPVLGCQGLSNVIW
Dtert 586 APPPSFLSVFLRRVEQVGAEFAPQEVANTLWALARLGAKPPA---PVMAEFFSATDRRLSSFKPHELSSMVWALAKMGFTPDKAWTEEFLHATFHKLPGLGSQGLTNVIW
Cappl 523 QLPPPFLALFERRVEQVGANFLPQEVSNTLWAFARFGAVPSA---SVLVEFFEATDRRLSNFKTQELANMIWALAKVRSTPDKRWCEEFFKATYYRLAEFERVGLCNTIW
Chlei 485 QPPAPFLSAFLRRVEQAGAEFSPQEVSNTIWSLARFGVAPSS---GLMVEFFVATDKRLSSFKPQELSAMAWALARGGHKPDRRWSEEFLHATFHKLPEFSQQGLCNMIW
cons. 881 pp wlga v rv aslf pvevantiwalarlgvkg lp evlalffvatdrrlssfkpqel mmwalakmr rpdkew eflkvtyhkl m g la v w
// OPR12 → ← OPR13
Chzof 542 SVTE---LHLQPPPA-LYHWVHASRQCLL----------
Chasy 682 GLAE---LSCHPPGGWLYAYANAARAALP----------
Cheus 558 ALER---LHIMPPPAWLYSYVNACRTLLI----------
Chchl 540 ALAT---LDLHPTPAWLYNYVNVCRRHIE----------
Cheur 180 GFVR---MGIVPPPAWQYAYVNVCRRHMG----------
Tesoc 214 SIAE---LALAPPPAWIYSYVNAVRALAEATPMPPPAAAA----------SDAAAAEARRREWLQTGDAEEGVEAVEMEPQLEGPLDVFLAQAGGMVAAPMPAALAAAQ
Chsph 762 SLAE---LELTPPPAWTYALVNAARGLMAAAAPQPT----------PAGGAA----------
Gopec 729 SLAE---LGLAPPPAWVYSFVNAARALLEATPLPAAQAAAEGQPQELPQEQPAEAEPKNRRERRRLRQQAVAAATARRLEHEKLELAGAAYDLLSLPPLSPSPQPVVAAA
CHSP3 535 ALAE---LALVPPPAWTYSFVNAARALMTTPAPSAG----------VKEEGGAVAARA----------AAGAAAMGVTAMD
Chdeb 686 ALAELSDLGLHPPPAWIYALVNAARQRLAATPPVAV----------
ChspW 640 ---------PTPP----------
EuspN 774 SLSE---LSLVPPPAWLYSFVNATRALMNASAASQQ----------
Chrei 781 SLAE---LQLSPPPAWTYSFVNAARALAEQAAQPPP----------PAAAAPTSALHQ----------PGAEPPLRSLRDL
Vocar 721 SLAE---LSLMPPPAWIYSYVNAARALLNVPAPPPP----------
Yauni 667 SLSE---LSLLPPPAWTYCYVNAARGLLDILPLPQT----------
Dusal 635 ALAI---MPIRPPPAWLYAFVKVVSADKA----------
Dtert 693 ALAV---MPLRPPPAWLYAFVKVVSADKA----------
Cappl 630 ALGT---LSLQPTPAWMYAFVKAARLQLH----------
Chlei 592 ALPM---MGMAPTPAWLYQYVKVATPHMD----------
cons. 991 glae l l Pppawly wvnaara l


//
Chzof 567 DMNVVDLAQAIRALQRHNA EA R LAKVD DFLTDALDRLS SLELDHGAYSKAAIGMLL
Chasy 708 RMSAFDLGQVVQGVRVLGG AH GSSMPKLQ ALELDALDALS GGEVAEAAYSRCQVHALLALGPG
Cheus 584 SGRLRALDLGSMVKALSSLDA EN K LQKVG DFVLDAKDALS TAELGSQAYTQQQLHYLSGMRSM
Chchl 566 AGKLSAGDLGVCVQALQRLQA AH G LQKVD DFLGDALETLG GMEVADQAYTKQQLQHLASMSSM
Cheur 206 AGQLSQRDSRIMLSCLARVN GG SPMLAKVD AFCVE LAGASDGGVGECTV
Tesoc 310 QQQPLAAAGGGGGGDAGGGVVPRALGGGVVASVAGLSALDLGTIISSLRRLNA GP G LDKVE EFLRYAEERLA AMEAGSGAYARRQVQKLLSMSPG
Chsph 801 AMATVEGHGGLSPLDLGQIIQALRQLNA TS Q LGKVE EFVLEAEAALG AMEAASGAYAAQQLRRFLAMGPG
Gopec 836 GVSQGSQPAAGLAAAASATDSFTFAVGGLPQAGGLTALDLGQIILGLRRLNSRS GA G LAKVD AFLEEAEARLR EMEAGGGLYAAGQLRAFLRMNCK
CHSP3 593 SDGEGWGAGEAGEAEGAGA AVGAGSMVGSLSVLDLGQVITALRTLNE KA Q LSKVE DFIRDAEDALR GAEERNAAFASRQVGAFLAMGQA
Chdeb 722 PPAAALDGPAAPSTGANAPLSALDLAQLILGLRRLSAV CG G LAKVD DWLREAEGALA EAEARSGAFAEQQVRALLAMGPR
ChspW 644 SMAPHPRRSCAQCCGRWRA CG G
EuspN 807 PSPPRPEAGGPQGLSALDLGQIITALRRLDGSGGGSGGAAGGGLAKVQ EFLGEAEARLA EMEAASAAFARQQVGAFLALSPH
Chrei 839 SPSASASPSSTGSALTYPALSVAAGAADAAGSGSGLSAIDLGQIITGLRKLNS VQ G LAKVD DFIGEAEERLR ALEAGSGAYAAQQVGHFLSMSRK
Vocar 754 VEAGAAPQGEDTESVAGTAVASLSAMDLGQIITALRRLNGG GG GGGLAKVD EFLREAEERLA AMEAGSGAYARQQLGAFLAMSPE
Yauni 700 TAPPQGAGVGPETTAKAVGSSQGQDAAGLSAMDLGQIITALRRLNGSGGSGGG G LTKVE DFLKEAEQRLG AVEAGSGAYARQQVGVFLAMSPG
Dusal 661 SMKTSDLVIIIKSLRSLAE GLQLQKVDACVADLVAEAQQRTRPPSFAVLHASVPLGRSRPRRSLQSM
Dtert 719 SMRKGDLVLIINSLRTLAE GLQLQKVDACVADLVAETQQRAQPPSFAALNASVSHSRHSAPSSQPSP
Cappl 656 AFSVTDLTQLILGLRAINA SM H LDKVD ELMMEAVELLQ AQPRLHGTVRNGVLRQVAADKQQ
Chlei 618 TMNKADMVQLVVGLTALSAQH DG G LQKVL DLVEEGYRQL EGRLGASAAQLARLREHGGVAEA
cons. 1101 g lsavdlgqvi alrrlna a g l kvd dfl ea l eaa gayg qi lam
Chzof 623 RLPAPASPQHHQQQQQQPPA TEGSRKHAGIFPVGLQGVDQQSQQRVDTAQH DVASFKARVAQG
Chasy 771 SVAYGALHA ETRRRAGDTAAAREGAQALL RNVRSRCSSSTDVSGCNAGVKRKRNSQQQQRQDDEVYVEVRAAWLAAEQQQRRQQRQEQQQQQ
Cheus 647 QRD TSSAVQASDAKSANVHEPLA QEVGIKSDDISPSELSPITLEDDFQSIKETL LQSFLVTRSGSASSTSC
Chchl 629 GAS SSSMGASSSSRAQGSKLQAL HRVVESALAEVEVEVE APPPGPAPPT
Cheur 255 SHAEGASGTGERMGAHAKAA NGAALQPEQLCCDA HGAATPVPLS
Tesoc 405 QAGLGLGGQATPE GGKQAAAVPATEGGGSESESDRE SEGSSRESGSEDEGGSGSNAAGGVRQGTAGAGPLQVAVAATPTGADAAPGVEAVVCAGRRPPA
Chsph 871 QA RLPRDEDALAAAAEERAAEQ IAAAQAAAAERRQRLQGMKPRRQRSRPAGAE VG EPRRGRKQRAATA
Gopec 933 QA GLPGAPASVA AEV EEA HSAEGVMPPLAAERSH RGGAERLGAQ
CHSP3 682 K GQRRGSGRRKEQQQEQEQLV EEAEEQQQQAGSKGRKGRKAGAPRRGAAAEL SG SGGGNGNGMGRHAQH
Chdeb 802 EARLAEAFEAAEA AEARARAAIQAAAAAAEANG NGVVAHAGGEGLNGANGVATHAAGSELQVQA QGAKTPRAPRQRQ
ChspW 666 G RTRSGRRSSS
EuspN 889 QAG LPKAAADVLVPDPVAPDPRA IGISSDLDGSEDTGAVGGRAGVATPQGAPRR RATSGPPPQQ
Chrei 934 QA GLRPAASQHEQQQEEGESEP AGADDDDEGAGVEVRRRVSVSGLADPDLSGS DAEESAPSASGAAQ
Vocar 839 QASQSGTAGARVLVPPMVSELGDGSGGGGVGSDLGGREPATAAGVSDGGGGGGSGGGGRAGLKGPRVKRRRQRLAAAL DPDQSQLLHHGERRRRGGLSGPRGKKRVA
Yauni 793 QA GMGSQPTVEEPNG GDGELGEAGGGSAAAAPASVRRRRRRSSSSS SSSSLAERLPLP
Dusal 728 SA SPAPASGAAAAATPARGTHV PASTLPASIPASDASLPHDSTVASSPQSNCS EQQAWDSDYSPSSPVPLP
Dtert 786 TSS VPAAATDSHVSTSASPTPFP ASNTGTPQDSAAASMPHSGFGEQQAWLPGCF HRSPEPLP
Cappl 717 AAA AAAAVAGNGAEGVAGSSLDG TGASNDSQPSAEATSSSAAVAQAAVQVPTEQ QQQWVRKQRKAAA
Chlei 680 EEA-AQQGPEQAPD GTRLGSSSSAGLVGALANDG SSSRQQQAEAASTAQAGAASAAGAAVAAAAAVAAQRAAQASRAGAGGAWGLDQRPVRAPVAAT
cons. 1211 a g a g g a

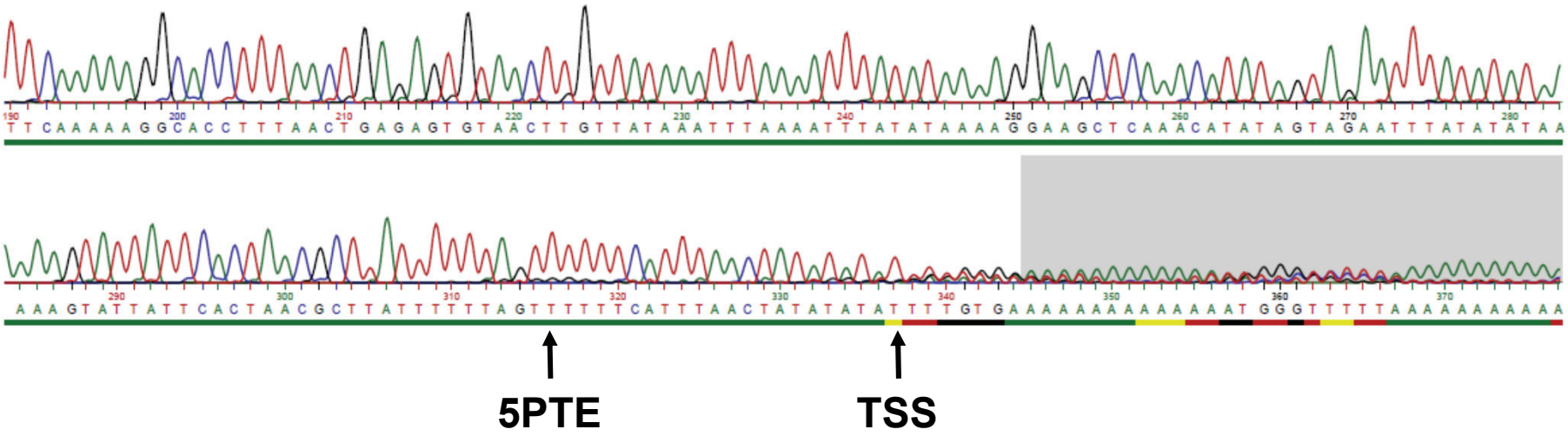

Chzof 686 SLSEQECR----------TDKHFRHSNGRVRGDHGRVVEVA------
Chasy 863 RRKQLQGQKQQQQPLQDREAITSLKRTNTPPQAHHQGVEASSDASRLQRTVSAGSGPSPTDPCSAIPPPAPAAVPAVGLMVSEASTRTPQQQRQRQQHRLRLKS---RQH
Cheus 718 MAPVAAGSTSCMAPVAAATLCI----------SPTTVLMPASSSSSSGTSAEYVSAS--------SLHCKSHIVNVEGDQLARSKRRK--QRWL
Chchl 678 ADV----------LPQEVVGPHVSSSTCLPEGEEEEWE--------GKQQQQQQQQQRRRKAG----------
Cheur 299 GMHATLGST----------
Tesoc 504 ASGRPRGAARRLAAVAEVRRALEGLD----------LGLEVPVGPVRNGVPVAAAGAGSPP--------AGQRRGSKAVAGRAGAGRKAGHA----GG
Chsph 939 AAPASEQAAAGEPP----------QPAEAVAPPPEQAA----------A--------GERRRGRKQALNGGGAGPAAAAA----PA
Gopec 977 ----------QPVAAGAPAARING----------GGAGG--ARVTLSA----AT
CHSP3 751 SEVEDGGKEGDERSEDALRAEVEAVR----------RQLEAVAAAVENGGNGGTAATAAVA--------AVQSREVTRVLHGG--RRLTLQV----PS
Chdeb 879 PKAQPQASEGAESQPQAQAQPGSTPRRS----------QPRRPRKQPAEAAAPAVAAQPHPEA--------HDAPSEPSAQEGAAWAAPLAEPS----GG
ChspW 677 RPRSTSCRP----------
EuspN 953 HQHQLRQSVKRRRKDVGWERAATS----------PPGPVIAPGVSDTAPAGHGAVT----------AAAAERAARGVRGG--ARANLAG----PP
Chrei 1001 RRASAEASTSGVAHPQKQQRRQRG----------RPGAAAAAAATSAGADVATPSESPA--------ATTRDALTRVLRTG--ARVTVLA---PPP
Vocar 946 AAAATVGRAGGEVAGGDDASGREAVLLTAGASSDLRSEFGIGVAAAEGAVASAMAAPPSAAAAAALSAHGRSS--------SSSPRQATRVRSGGPAARVTLAA---PPP
Yauni 851 AGAEAVGKG----------VEQLWPYSGQPGGAAAAA----------DPGSREVTRVLRGG--ARVTLSA----PA
Dusal 799 AQHDHHGGAG----------LQGDWVLPEHLDMDSKDQRQQSIP---------SSRHGTSAAVMRRAEVVRAVADV----GI
Dtert 848 GQHGHHEGT----------WQGDWVLHEHLDMSSEDQRQQSIP---------SSRHGSGAAAMRRAEVVRAVADV----GI
Cappl 784 VKALKQKAAGTLDA----------EQATLLAHGPPHTPAKAKAPEQPAA--------GASNGSNAASSNGGGSAPAVLNG----GM
Chlei 776 AAAAAEQASPEIGEIGMVKKRRRSAKTLFAAEDPTSSSEDA------AASSAATSGDGAAAAASANGTGAEQ--------GAADSDSASSIRKARQRRATAIKAGVEGV
cons. 1321 g g gr
Chzof 717 ----------AEMQI----------DLDATTGV----------
Chasy 970 HHRRSPAGTGDGADAGTDGGIGADGGVNGSRSSSGRDPPLPLPAVPGAQLDGSGARSTTDGAGGGAGSGVAPEMRMRVGEVVLDLPASPSPSRPATLRTQPAAAAPPALP
Cheus 792 SSRRQPLGLGGSLIVQM----------PSDAV----------MHISDLDLDVALSPKDLRVVQEV----------
Chchl ----------
Cheur 308 ----------GAGAV----------
Tesoc 580 QRRAQPVPLGSGVSNG----------GGGEG----------DDSGGSSG----------
Chsph 993 VVH-----VGLGMGLGL----------GELSV----------ELA-SPSG-----PMSMRAAAPP---PS----------
Gopec 1005 EVQ-----LEL----------GELSV----------GFTSGGTAAEALSPSSWVAQAPS---AG----------TAPAT
CHSP3 825 EVR-----LEL----------GDLMV----------ELTGAGQG-----AATATAAAGA---AA----------
Chdeb 957 VPRSKGVHVGLGLG----------GEGLS----------VELGSPSG-----PVSLRAGGDA----------
ChspW ----------
EuspN 1022 EVR-----VDLG----------GELSV----------ELPASASG-----PVLWRPGSNS---EAVAGVGPGSSSEAKGVEADSSSGS
Chrei 1074 EVH-----LEL----------GDVSV----------ELPAGTGG-----PLSLRTAPAA----------A
Vocar 1045 DVH-----VDLG----------GELNV----------ELSVSPSG-----PALWRSSSAATAAPR----------MAGLVGAAGGPN
Yauni 901 ELH-----LEL----------GELNV----------ELSPPSG-----SVAWHAPSSG----------
Dusal 858 THQQATLLA----------SQHAK----------RHRGKSGERGLVDSSSLRKPRKM----------
Dtert 906 THQQATLIA----------SQHAK----------RHKGRSEEGGVVDSSSLRKPRKT----------
Cappl 848 HHGLDGSGSTTSGGDGA----------AELAVVD----------DAAASTMGRAVISLEGLHGHVVLREQPA----------L
Chlei 871 VGKEQASLLATGHPAAM----------PKATVVKKAATKRVQKEQKEAAAAAAVAVAAEGASAQ-PATYASAAAADVPAEGTEKMGRAYVSLEALNGHVLLRS
cons. 1431 r l ge v a g r

```
Chzof      730 -----GLVGVGA-------------------------------------------
Chasy     1080 GAGEAAGPRSGAASAG-------------NAGTTCDSSPVHGRQ-----------
Cheus          ----------------------------------------------------------
Chchl          ----------------------------------------------------------
Cheur          ----------------------------------------------------------
Tesoc      609 -----GGGSNGASSAG-------------DAGDSSTAAGVGTRLLAV--------
Chsph     1029 -----AGPEAAGGGEG---------------------LAEIKRRLLAV--------
Gopec     1046 NGAEAGGVGAHNGAAGG-----------GEETGQL--AAEIKRKLMAVP-------
CHSP3      856 --ASGVGAATGAAAAA-----------GTAAGAA--ASLSQP-------------
Chdeb      994 GAAAAAGGGAGSGVDG-------------------ISAAEIRRRLLSV--------
ChspW          ----------------------------------------------------------
EuspN     1077 SSSNGHGAEVTGGAAG-------------GEDLSVSLAEIKRRLLAV--------
Chrei     1104 SSSSSSGKGAGNGSVG-------------SDGAGNSAAEIKQRLMAV--------
Vocar     1092 GDGNGPGNGSGNGVSGSRSSSSSDECEIADGGGMS--MAEIKRRLLAV--------
Yauni      929 SASNGAGGEPTAVVTG-------------GEAGAANAAEIKKRLLAV--------
Dusal      895 ---AADSEEKGDPACT-----------DLHVSNVSRISNDRLSMSLPR-------
Dtert      943 ---APDSDEKGDPACM-------------DLQMXPAKVMQLERSLQVNMFGTGVCV
Cappl      901 NGHHAPVISAVSGAAG----------------------GINAKDMLVR-------
Chlei      963 PAGAAGGNGNGNGSAT-----------GNGTGPTATDAELQPLVGVKT-------
cons.     1541     a g   g a  g                        v  rl
```

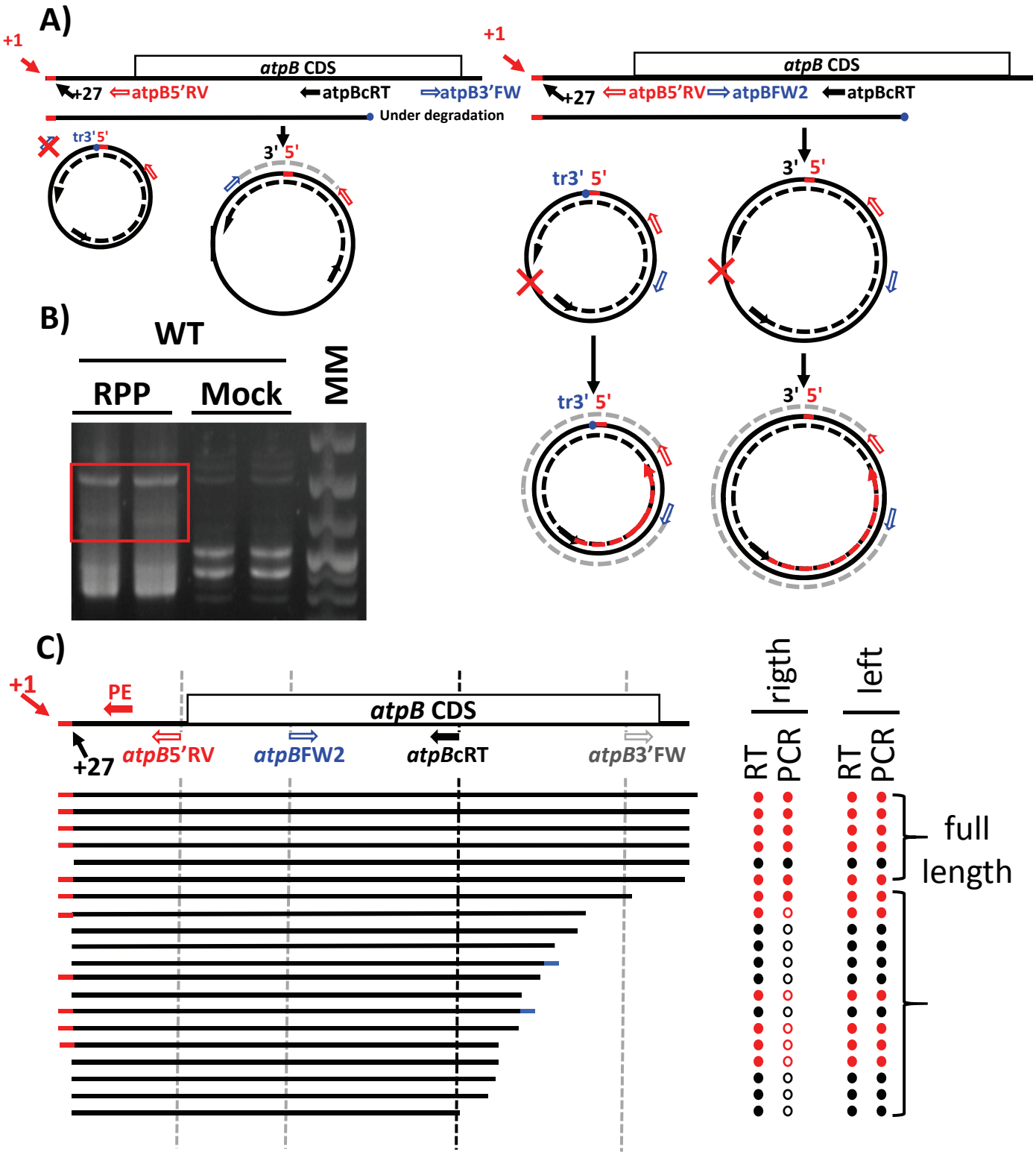
DNA regions encoding MDB1 orthologues were retrieved from the NCBI database by TBLASTN searches, using CrMDB1 as a query. Gene models were then predicted with the GreenGenie2 software (http://stormo.wustl.edu/GreenGenie2/; (Kwan et al., 2009)) and manually edited to include missing obvious regions of similarity, if required. Alignment of MTHI1 orthologues was performed with the MUSCLE software using default options and manually edited to improve the alignment. The position OPR repeats of the protein from *C. reinhardtii* are shown above the alignments. In the sequence from C. reinhardtii, residues highlighted in yellow were replaced by the HA tag. Additional OPR repeats found in the species-specific insertions are highlighted in yellow or green. Residues conserved in more than half of the sequences are written in red, while conservative substitutions are written in blue.

Abbreviations of species names are as follows:

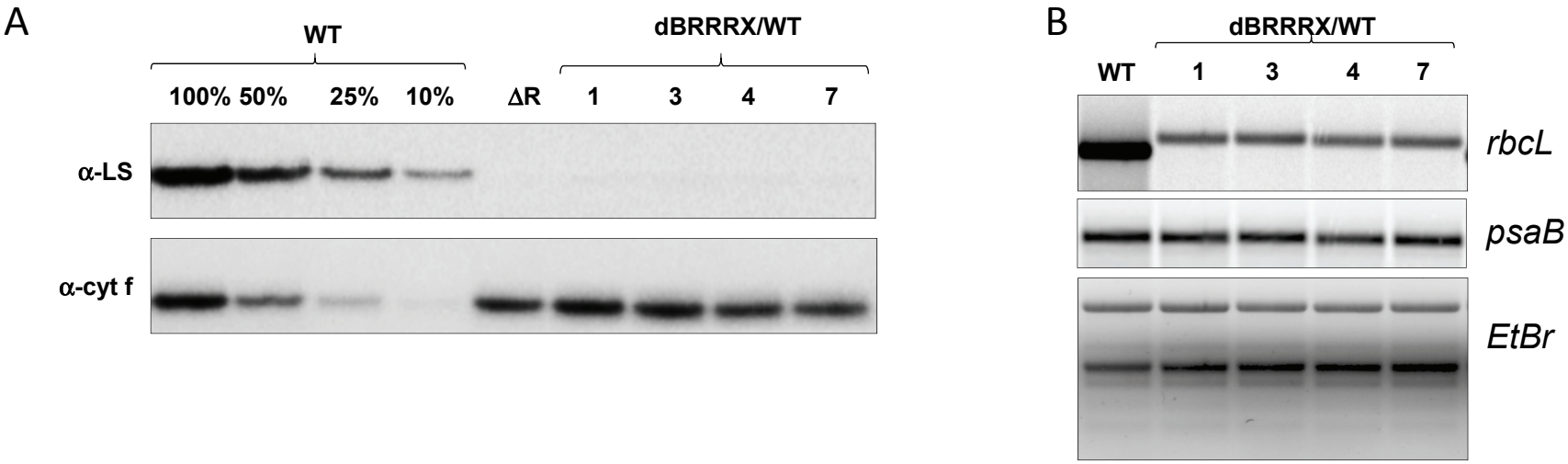
Chzof: Chromochloris zofingiensis; Chasy: Chlamydomonas asymetrica; Cheus: Chlamydomonas eustigma; Chchl: Chlamydoimonas chlamydogama; ; Cheur: Chlamydomonas euryale; Tesoc: Tetrabaena socialis; Chsph: Chlamydomonas sphaeroides; Gopec: Gonium pectorale; Chsp3 Chsp3: Chlamydomonas sp. 3112; Chdeb: Chlamydomonas debaryana; ChspW: Chlamydomonas sp. WS3; EuspN: Eudorina sp. 2006-703-Eu-15; Chrei: Chlamydomonas reinhardtii; Vocar: Volvox carteri; Yauni: Yamagishiella unicocca; Dusal: Dunaliella salina; Duter: Dunaliella tertiolecta; Chapp: Chlamydomonas applanata; Chlei: Chlamydomonas leiostraca.



Suppl. Fig. S5: Sequence analysis of the RPP-treated sample from wild type.
The position of the matured (5PTE) and precursor 5'end of the *atpB* mRNA are shown. The poor resolution of the sequence downstream of the *atpB* precursor mRNA 5' end suggests heterogeneity of the *atpB* 3'end.


A)
+1
atpB CDS
+27
atpB5'RV
atpBcRT
atpB3'FW
Under degradation
tr3'5'
3'5'
+1
atpB CDS
+27
atpB5'RV
atpBFW2
atpBcRT
tr3' 5'
3' 5'
tr3' 5'
3' 5'
B)
WT
RPP
Mock
MM
C)
+1
PE
atpB CDS
+27
atpB5'RV
atpBFW2
atpBcRT
atpB3'FW
rigth
left
RT
PCR
RT
PCR
full
length

Suppl. Fig. S6: Alternative determination of the *atpB* 5' end by cRT-PCR.
A) Schematic representation of the two PCR strategies. The position of the various primers used are shown. The red segment at the beginning of the *atpB* mRNA represents the processed part of the transcript, only found in the precursor transcript. A hypothetical *atpB* transcript under degradation is also shown by a blue dot. It doesn't contribute to the final amplicon with the classical cRT-PCR strategy (left), but it does with the second cRT-PCR strategy (right).
B) Agarose gels showing the amplicons resulting from the second cRT-PCR strategy, with a molecular weight marker on the right. RPP and mock respectively indicate RNA samples treated or not with RPP, to distinguish precursors from processed transcripts. The band used for cloning experiment is shown as a red rectangle.
C) Schematic representation of the results from the sequence analysis of the 20 clones (details in Suppl. Table ST2). A red segment at the 5'end of the *atpB* mRNA indicates a precursor transcript, while a blue segment at the 3'end of the transcript symbolizes the presence of a polyA tail. The left part shows the contribution of the clones to the amplicon in the two cRT-PCR strategies. In the classical strategy, 7 clones (6 precursors and one processed) would have contributed to the final amplicon, a proportion explaining why its sequence revealed only the precursor form. In the second strategy, 11 precursor and 9 processed transcripts participate to the final amplicon. In the primer extension experiment, one can expect many more partially degraded *atpB* mRNA to contribute to the extension product (all those whose 3'end lies between the primers used for reverse transcription and primer extension - primer PE, above the *atpB* transcript-), resulting in a high representation of the processed form.

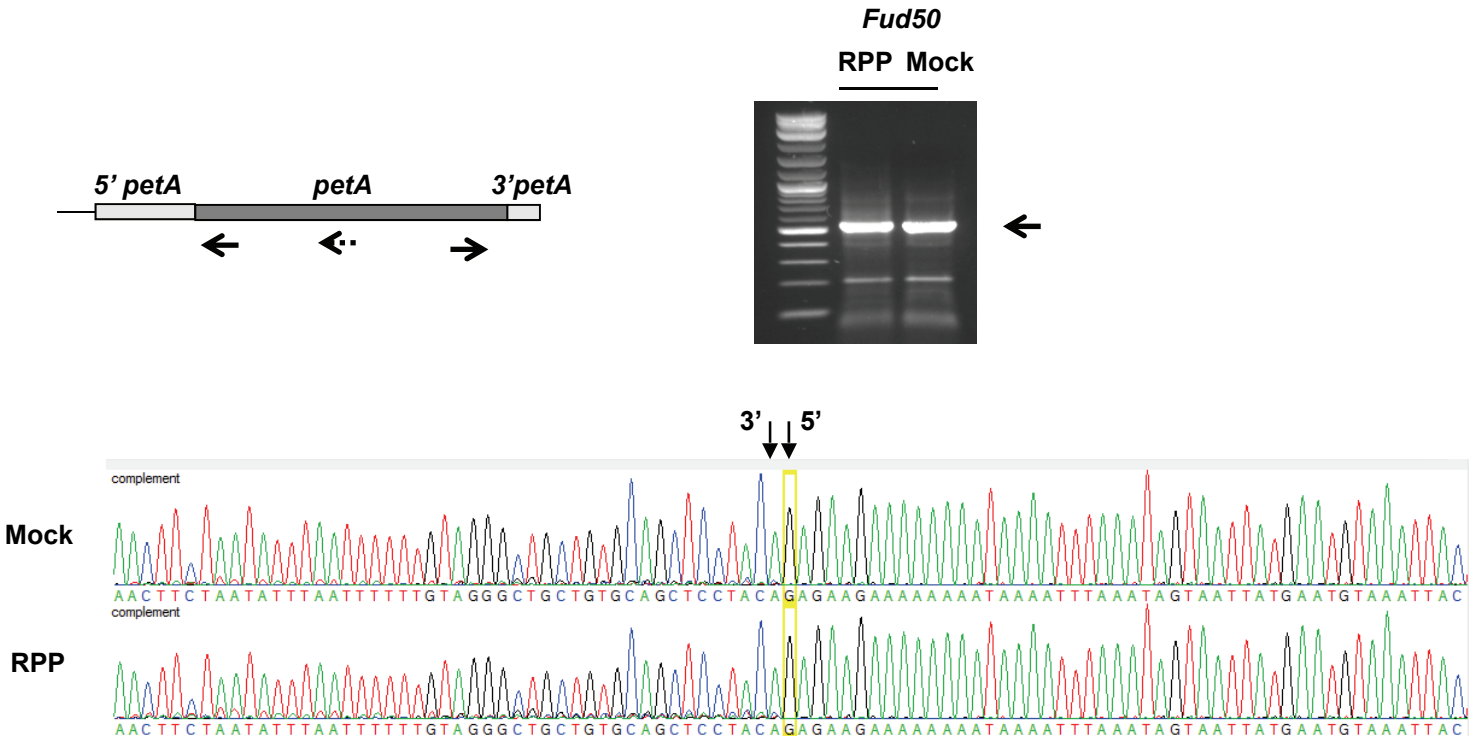


Suppl. Fig. S7: Characterization of the BRR transformants

A) Accumulation of Rubisco LSU subunit in the BRR transformants compared to a dilution series of wild-type proteins. Cytochrome f accumulation is shown as loading control. ΔR is a strain bearing a deletion of the chloroplast *rbcL* gene.

B) rbcL RNA accumulation in the BRR transformants compared to WT strain. *psaB* hybridization and ethidium bromide stained gel provide loading controls.

Suppl. Fig S8 Determination of *petA* 5'ends in the Fud50 strain




Chlei
Phlen
Chapp
Loseg
Botex
Chacu
Chsti
Mimon
Prbot
ChspU
HaspN
Chtat
Chper
Chrei
Tesoc
Yauni
Voafr
Plsta
EuspN
Gopec
Chdeb
Dusal
Patex
Chsp3
Chsph
← MDB1 BS →
Cons.
at tgtttt gaataaaaaaa g
ga taaataaaa ta
aaaAAATAAGcGTTAg t

```
Voafr  --GAATAATACT---------TATAATTTATAGGTTCTAACAATCCCCATTAGCT------
Plsta  --AAATAATGGTTAATATTAATATTATTAATAGCTTCTAACAATAC---------------
EuspN  --GAATAATACT---------TATTATTAATAACTTCTAACAATCTCCGTTAGTT------
Gopec  --GAATAATACT---------TTTTATTTATAAATTCTAACTATACTTCTTCTCT------
Chdeb  --GAATATCATA---------TAATATTATAAATTCTAATTTTTCTTTGTCTAAT------
Dusal  --GAATTTTAAT--------------TCCATAACTTTTAGAAATAAAAAAAAAAAAA----
Patex  --GAAAAA-------------ACTTATAAAAAAGGATTATTTGCCGTATAATA--------
Chsp3  --AAATTTTAC----------TTTTAAAATACAGTCAATTTAAATACACGTAAA-----
Chsph  --GAATATTACT---------TTATATTCAAAAAATGAGGTTATAGATGAGCAAT------
Cons.    aaata ta           ttttatt ataaa t ta  aatg   a
```

Supplemental Fig. S9: Conservation of the MDB1 target in *atpB* 5'UTRs.

A) Alignment of *atpB* 5'UTRs

Available chloroplast genomes of Chlorophyceae were scanned for occurrences of the putative MDB1 binding site AAATAAGNGTTAG. In the reported species this sequence was found in the intergenic region upstream of atpB (and almost always only here), at a variable distance (ranging from 152 in *Dunaliella salina* to 1468 in *Palmellopsis texensis*; mean size 550) from the translation initiation codon. Sequences were aligned with the MUSCLE software, using default options, and then manually edited to improve the alignment. Residues conserved in more than half sequences are written in red, while conservative substitutions are written in blue. A putative -10 Pribnow box found upstream of the putative MDB1 binding site in some species is written in bold and underlined. It was always found at more than 10 nt of the MDB1 binding site, suggesting that the mature *atpB* mRNA is a processed transcript in these species as well.

Abbreviations of species names are as follows:

Chlei: Chlamydomonas leiostrac; Voafr: Volvox africanus; Chacu: Chariaciochoris aciminata; Chtat: Chlorococcum tatrense; Dusal: Dunaliella salina; Chrei: Chlamydomonas reinhardtii; Gopec: Gonium pectorale; Phlen: Phacotus lenticularis; Plsta: Pleodorina starrii; Yauni: Yamagishiella unicocca; Eu_sp: Eudorina sp.; Tesoc: Tetrabaena socialis; ChspU: Chlamydomonas sp. UWO 241; Botex: Borodinellopsis texensis; Loseg: Lobochlamys segnis; Mimon: Micronegla monida; Chdeb: Chlamydomonas debaryana; Chsp3: Chlamydomonas sp3212; Chsph: Chlamydomonas sphaeroides; Halac: Haematococcus lacustris; Chapp: Chlamydomonas applanata; Chsti: Chlorosarcina stigmatica; Prbot: Protosiphon botryoides; Chper: Chloromonas perforata; patex: Palmellopsis texensis.

All species belong to the Chlamydomonadale order.

ARTICLE 2:

"PHOTOSENSITIVE MUTANTS OF THE PERIPHERAL STALK OF THE CHLOROPLAST ATP SYNTHASE IN CHLAMYDOMONAS REINHARDTII."

This is a draft version of an article devoted to the study of ATP synthase mutants. Some complementary experiments have still to be done and some parts of the article have yet to be written.

DRAFT ARTICLE:

PHOTOSENSITIVE MUTANTS OF THE PERIPHERAL STALK OF THE CHLOROPLAST ATP SYNTHASE IN CHLAMYDOMONAS REINHARDTII

Authors: Frédéric Chaux, Domitille Jarrige, Marcio Rodrigues-Azevedo, Sandrine Bujaldon, Shin-Ichiro Ozawa, Yves Choquet, Catherine de Vitry

Unité Mixte de Recherche (UMR) 7141, Centre National de la Recherche Scientifique (CNRS) and Sorbonne Université, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, F-75005 Paris, France

Keywords: ATP synthase, Chlamydomonas reinhardtii, ATPG, AtpF

Abstract:

RESULTS

FORWARD GENETIC SCREEN ON HIGH LIGHT SENSITIVITY

Proteases are major maintenance factors of the photosynthetic complexes embedded in the thylakoid. Among them, previous investigations of *ftsh1* deficient mutants (Malnoe *et al.*, 2014; Wang *et al.*, 2017) showed that hetero-oligomeric protease FtsH1/2 is of prime importance, especially for PSII repair, the mutant *ftsh1-1* accumulating inactive FtsH1 being highly photosensitive.

After insertional mutagenesis, transformants were screened for *ftsh1-1*-like light sensitive phenotype. Wild-type (WT) strain T222+ was electroporated in the presence of the hygromycin resistance cassette, and cells were plated on selection medium (TAP hygromycin 20 mg L^{-1}). To assess putative altered protease activity of transformants, we set a fast screening procedure by imaging chlorophyll fluorescence on plates with a time-resolved wide-angle camera (Johnson *et al.*, 2009) aimed at measuring the ability to maintain PSII function upon short photo-inhibitory treatment. About 2000 random transformants were transferred to fresh TAP plates and grown for one week in permissive conditions (5 µmol photon m^{-2} s^{-1}). As a control of how FTSH1 protease defect impacts fluorescence yields, we introduced the *ftsh1-1* mutant and the complemented strain C17 (*ftsh1-1*(pSL18-FTSH1), (Malnoe *et al.*, 2014)) as well as the untransformed WT.

Using the method of weak detection pulses and saturating flash (reviewed in(Baker, 2008)), we measured on each plate (i) the basal fluorescence F_0 and the maximal fluorescence F_M in the so-called dark-adapted state, as well as (ii) transient fluorescence F' and maximal fluorescence F_M' after 5 seconds low light. The ratio F_V/F_M, where $F_V=F_M-F_0$ is the so-called maximum PSII yield and depends almost solely on PSII function (*e.g.* PSII deficient mutants show F_V/F_M close to 0); in contrast, the PSII (operating) yield Y(II), defined as $(F_M'-F')/F_M'$, decreases when photosynthetic electron flow is limited downstream PSII (*e.g.* PSI deficient mutants show Y(II) close to 0 even under low light). Figure 1 exemplifies our screening procedure with F_V/F_M pictures of the plate where the mutant E236 was isolated. Under the permissive conditions of growth (left panel: TAP, low light), the *ftsh1-1* mutant shows F_V/F_M and Y(II) (not shown) similar to WT (≈0,65). We hence discarded clones with F_V/F_M or Y(II) close to 0 under these permissive conditions to focus on "*ftsh1-1*-like" phenotypes. To trigger the processes of damage and repair, the photo-inhibitory treatment then consisted in an exposure to 1000 µmol photon m^{-2} s^{-1} for 1 h. This resulted in a drastic decrease of F_V/F_M in *ftsh1-1* (≈0,05) mutant as compared to the moderate decrease in WT, C17 and most of the plated strains (≈0,30) (Fig. 1, central panel). Moreover, F_V/F_M recovered to pre-treatment values after 3 h back in low light (recovery period) in WT, C17 and most of the plated strains but not in *ftsh1-1* (Fig. 1, right panel). These changes were expected due to the central role of FtsH1 protease in PSII repair (Malnoe *et al.*, 2014) and allowed us to identify ≈20

photosensitive clones altered in the maintenance of photosynthesis, such as the mutant E236 (Fig. 1, white circle).

To identify the most robust mutant strains, selected clones were grown in liquid TAP to assess their photosensitivity several times in more controlled conditions (exponential phase, similar cell concentration, homogeneous light intensity (5 µmol photon m^{-2} s^{-1}), etc). We thus selected 6 mutant strains which reproducibly exhibited decreased tolerance to high light, as exemplified in Figure 2A. Mutant strains E236, E271, F28N and F292 showed F_V/F_M values similar to WT under the permissive growth conditions (T_0), but upon a 30 min photo-inhibitory treatment (1000 µmol photon m^{-2} s^{-1}) F_V/F_M values decreased more drastically than the WT, reaching values similar to *ftsh1-1*. However, when cells were moved back to low light, F_V/F_M values recovered faster than the *ftsh1-1*, the mutant E236 being even about twice faster than the three others (E271, F28N and F292). The two other mutant F28O and E113 showed similar trends, except that F_V/F_M values were already respectively more or less lower than the WT under growth conditions (T_0), suggesting that their photosynthetic apparatus is already under pressure in low light.

Interestingly, while WT strain rapidly reaches a low level of steady-state fluorescence (F_S) under low light, four strains (E236, E271, F28N and F292) exhibited a similar phenotype of rising fluorescence almost up to F_M (Fig. 2B). We attributed it to the gradual appearance of a bottleneck in the use of photosynthetic electron and the build-up of very high proton gradient which downregulated PSII, as previously reported in ATP synthase mutants (Johnson *et al.*, 2014).

Mutants defective in the chloroplast ATP synthase are highly sensitive to light (Majeran *et al.*, 2001). We hence compared the growth of the mutants E236, E271, F28N and F292 to that of the ATP synthase mutant *mdb1*, which lacks ATP synthase because the maturation factor MDB1 is required for the maturation of atpB transcript (Drapier *et al.*, 1992). To decipher how phototrophic growth and tolerance to high light are impacted in these mutants, they were grown either in minimal medium (MIN) or in the presence of acetate as a reduced carbon source (TAP medium), and exposed to increasing light intensities (Fig. 2C). In TAP, the E271, F28N, F292 and *mdb1* mutants grew slightly less than WT strain under moderate light (25 µmol photon m^{-2} s^{-1}) and they almost did not grow under saturating light (120 µmol photon m^{-2} s^{-1}) while they were able to grow in the dark (Suppl. Fig. 2), further confirming the photosensitive phenotypes. In MIN however, these four mutants (E271, F28N, F292 and *mdb1*) were not able to grow at all. This holds true even under weaker intensities (not shown). This suggests that photosensitivity in E271, F28N and F292 mutants originate from strong impairments of photosynthesis itself similar, to that of the *mdb1* mutant. In contrast with E271, F28N, F292 and *mdb1* mutants, the mutant E236 was less impacted under moderate light (similar to WT in TAP, moderate growth in MIN) and growth

phenotype became obvious under saturating light (decreased growth in TAP as compared to WT, no growth in MIN).

Similarities between our four mutants and *mdb1* were further observed in fluorescence kinetics from the dark-grown cells (Suppl. Fig. 2). Although all five mutants had the same growth (Suppl. Fig. 2A) and same F_V/F_M as WT (Suppl. Fig. 2C), PSII yield under low light was slightly lower than WT after 30s (Suppl. Fig. 2D) and reached almost zero after 3 min (Suppl. Fig. 2E), similar to PSII yield time-course in Figure 2B.

Altogether, these preliminary characterizations indicated that four mutants isolated by screening on photosensitivity were putatively deficient in ATP synthase. Some differences however existed: E271, F28N and F292 were obligate heterotrophs similar to *mdb1* while E236 can grow photo-autotrophically but is intolerant to high light.

Evidence for impairments of ATP synthase complex biogenesis

To gain insight on these mutants at the molecular level, we investigated the accumulation of ATP synthase by immunodetection. As controls, we used deletion mutants of chloroplastic ATP synthase genes (*ΔatpB*, *ΔatpH*, *ΔatpI*) as well as *mda1*, *mdb1* and *mrl1* nuclear mutant strains, which lack RNA-binding factors required for the maturation of chloroplast transcripts *atpA*, *atpB* (Drapier *et al.*, 1992; Viola *et al.*, 2019)(Cavaiuolo *et al* in preparation) and *rbcL* (Johnson *et al.*, 2010), respectively. The levels of all four detected ATP synthase subunits were overall much lower in all four mutants than in the WT (Fig. 3), although in distinct patterns of accumulation which may reflect distinct impairments in the biogenesis of ATP synthase. Note that the subunits α, β and ε are parts of the soluble fraction (CF_1), which can assemble in the stroma (Lemaire and Wollman, 1989a), while subunits III is part of the membrane fraction (CF_0).

First of all, no subunit III was detected in all four mutants (Fig. 3B), suggesting that none of these strains can complete the assembly of the full CF_1-CF_0 ATP synthase complex. However, as compared to WT levels, all four mutants accumulated about 1/5 of β (Fig. 3C), and α was detected only in F28N and F292 in about the same range (1/5 of WT level; Fig. 3A); this suggests that F28N and F292 do produce the catalytic hexameric head $\alpha_3\beta_3$ while impairment in E236 and E271 impacts assembly prior to this step. Nevertheless, E236 together with F292 were the only mutants were ε can be detected (Fig. 3B). This suggests that F292 can probably advance the furthest in ATP synthase assembly (likely with a full CF_1 sub-complex) as compared to other mutants. Also α and ε can be accumulated in absence of each other (in F28N and E236 respectively) while β is a pre-requisite for both α and ε (in E271; see also (Drapier *et al.*, 2007).

We then investigated whether ATP synthase impairment in our strains may be due to mutations of RNA-binding factors controlling expression of chloroplast-encoded genes. Indeed, the expression of many plastid-encoded genes relies on specific nucleus-encoded

factors to stabilize, maturate and/or translate target transcripts (Barkan and Goldschmidt-Clermont, 2000). We hence ran northern blots against transcripts of the ATP synthase plastid genes: we were able to detect *atpA*, *atpB*, *atpE*, *atpH* and *atpI* transcripts (Suppl. Fig. 2) but not atpF. In contrast with the controls (mda1 strain and the deletion mutants *ΔatpB*, *ΔatpH* and *ΔatpI*), no significant change in accumulation and no shift between monocistronic and polycistronic forms was detected in mutants F28N, F292, E236 and E271 as compared to the WT.

ATP SYNTHASE DEFECTS SEGREGATE INDEPENDENTLY OF THE ANTIBIOTIC RESISTANCE.

Looking for the causal mutations in our mutants (mt+), we first tested if their phenotype was genetically linked to the insertions of hygromycin resistance cassette. Hence we backcrossed them to the WT strain S1D2 (mt-) and followed in their progeny the segregation of growth abilities: (i) phototroph versus obligate heterotroph and (ii) hygromycin resistance versus sensitivity. We generated progenies in batches (random progeny analysis). For each mutant, instead of isolating pre-meiotic zygotes from one another as for tetrad dissection, we induced germination in a batch liquid culture and plated their progenies on TAP.

When colonies appeared, we tried to sort the progeny in advance on their ATP synthase deficient phenotype based on their different fluorescence kinetics (e.g. Fig. 2B), and we transferred them on TAP (control), MIN and hygromycin-containing TAP plates to analyse their phototroph or obligate heterotroph growth and their antibiotic resistance (Suppl. Fig. S4). Our prediction worked almost perfectly on the cross F28NxWT (Suppl. Fig. S4, top panels), yet the weakness of the fluorescence signals on small colonies partially hindered our guess for other crosses. Most importantly, in the progeny of all four crosses, we observed the presence of a significant number of clones which were (i) both non-phototrophic and hygromycin-sensitive (indicated by thick red circles) or (ii) phototrophic and hygromycin-resistant (indicated by thin blue squares). Note that phototrophic growth under moderate light was not fully impaired in E236 (dashed circle), in good agreement with other results (Fig. 2C), yet a significant decrease in growth was readily observable in about half of the progeny from the backcross with S1D2. Altogether, backcrosses show that in all four mutants the impairment of ATP synthase was genetically independent from the insertion of the antibiotic resistance cassette but rather due to other mutations. This independence precluded mutation identification by flanking sequence analysis and required other approaches for the molecular characterization of the mutations.

GENES *ATPG* AND *ATPF* ENCODING THE TWO PERIPHERAL STALK SUBUNITS ARE RESPECTIVELY ALTERED IN E236 AND *FUD18* MUTANTS

We performed whole genome sequencing to identify the nuclear mutations. The whole genomes of mutant strains E236, F28N, F292 and E271 as well as the WT strain were obtained by paired-end sequencing and mapped on the latest genome assembly (see

Material & Methods for detail). We first looked for candidate mutations in the few genes knowingly related to ATP synthase, namely nuclear-encoded subunits, assembly factors and regulators of organellar genome expression (ROGEs) dedicated to chloroplast-encoded ATP synthase genes (Ruhle and Leister, 2015). The mutant E236 (Fig. 4) harboured several features of interest in the *ATPG* gene (locus Cre11.481450): (i) only few grey rectangles as compared to the WT, indicating a decrease in read coverage, (ii) a series of coloured bars indicating wrong nucleotides as compared to WT sequence and (iii) more than twenty reads one end of which maps in Cre11.481450 and the other end at distinct genomic locations (*e.g.* chromosomes 5, 6, 7, 12, 16 as shown by coloured rectangles in mutant E236 in Figure 4A), more specifically in *TOC1* transposons. Overall, this suggests a *TOC1* insertion in the 3'-UTR of *ATPG*, about 30 bp downstream the stop codon. We confirmed this large rearrangement by PCR amplification (Fig. 4B), showing no product in mutant E236 when using a primer pair spanning the whole locus (pair A), while the end of the CDS can be well amplified in both WT and E236 mutant. In the other mutants, we found no mutations in known ATP synthase-related genes, so we will implement a genome-scale analysis in the future and decided to focus on mutant E236.

The flexible peripheral stalk of the chloroplast ATP synthase has two protein subunits often named b and b' (Hahn *et al.*, 2018), respectively named in *Chlamydomonas* CFo subunit I which is encoded by chloroplast gene *atpF* and CFo subunit II which is encoded by nuclear gene *ATPG*. Having identified an *ATPG* mutant (E236), we searched for an *atpF* mutant. On the base of the absence of CF0 subunit I synthesis, mutant *Fud18* had been previously proposed to be altered in the *atpF* chloroplast gene (Lemaire and Wollman, 1989b). To further identify *Fud18* mutation, we amplified and sequenced the *atpF* gene region in mutant *Fud18*. We confirmed *Fud18* is an *atpF* mutant and identified the mutation as a deletion of one T after the start codon of *atpF* which causes a frameshift and an abortive Stop codon (Fig. 4C). This *atpF* mutant is photosensitive (Fig. 5A) and accumulates some α subunit of the hydrophilic fraction (CF_1) but no detectable ATPG (Fig. 5B).

To investigate *ATPG* impairment in mutant E236, we aimed at complementing it with a WT gene version (native promoter and terminator sequences). Mutant cells were electroporated in the presence of the 2.5 kb PCR product obtained in the WT (Fig. 4B) or in absence of DNA (negative transformation control) then plated on MIN medium and grown under low light (positive growth control) or high light (selection). Cells from both transformations grew under low light but only transformation in the presence of *ATPG* yielded clones under high light. Some clones were randomly picked up and grown in TAP low light to analyse complementation at physiological, biochemical and functional levels (Fig. 5).

Cells from WT, mutant E236 and three complemented clones E236*::ATPG* (C1-C3) as well as two ATP synthase mutants *atpF* (*Fud18*) and *ΔatpH* were plated either on TAP or MIN media and exposed to low, moderate and high light (Fig. 5A). All strains readily grew on TAP under

low light (2 µmol photon m^{-2} s^{-1}). On MIN under moderate light (60 µmol photon m^{-2} s^{-1}), E236 grew very slowly whereas the growth of all three complemented clones was similar to that of the WT. Even under excess light (200 µmol photon m^{-2} s^{-1}), complemented clones were able to grow as well as the WT, in contrast with mutant E236, *atpF* and *ΔatpH* strains, suggesting that *ATPG* transformation restored photosynthetic performance in mutant E236 under moderate and high light.

At the biochemical level (Fig. 5B), both ATPG, for which we raised an antibody, and AtpB proteins accumulated significantly more in complemented strains than in the mutant E236, although much less than the WT. Moreover, ATP synthase function was further characterized using chlorophyll fluorescence kinetics as in Fig. 2B on cells grown in TAP under low light. Under moderate light (Fig. 5C), complemented strains exhibited kinetics similar to the WT, in contrast with mutant E236 where fluorescence increased due to the bottleneck in the use of light. However, under saturating light (Fig. 5D), complemented strains displayed fluorescence levels intermediate between that of the WT and mutant E236. This suggests that the ATP synthase level restored in complemented strains (i) allows efficient photosynthetic electron and proton transfers under light-limited conditions where E236 is already mildly affected, but (ii) becomes slightly limiting under excess light, yet not detrimental to growth (Fig. 5A). Altogether, this suggests an only limited restoration of ATP synthase accumulation by complementation with WT version of *ATPG*, which is usual in *Chlamydomonas* under native promoters but sufficient to assume that E236 strain is an *ATPG* mutant.

MATERIAL AND METHODS

STRAINS, GROWTH CONDITIONS AND TRANSFORMATION PROCEDURE

We electroporated the hygromycin resistance cassette in the presence of the CRISPR Cas 9 enzyme and RNA guides to edit genes coding for thylakoid proteases FtsH2 and EGY1. Mutants were plated on selective medium; colonies were then transferred to fresh TAP medium and grown mixotrophically in low light (10 µmol photon m^{-2} s^{-1}) for 7 days before chlorophyll fluorescence analysis and photo-inhibitory treatment (1h at 1000 µmol photon m^{-2} s^{-1}).

Depending on the gene targeted by the CRISPR-Cas9 transformations, namely *FTSH2* or *EGY1*, these mutants were analysed by Western Blot or PCR. FtsH2 (lower band), as well as FtsH1 (upper band), was readily detected by the Anti-Var2 antibody in all mutants (Suppl. Fig.1A), suggesting normal accumulation of the thylakoid FtsH protease. Likewise, the targeted *EGY1* region was readily amplified in all mutants (Suppl. Fig.1B), suggesting absence of alteration at this locus.

GENOMICS

For each strain, we ordered an Illumina sequencing technology-based NGSelect DNA data package (Eurofins, Germany), comprising the generation of a standard genomic library (DNA fragmentation, adapter ligation, size selection and amplification) and a data package of >5 million pair reads (2x150bp). From these raw data, we generated genomic sequences using open-source platform Galaxy (usegalaxy.org) as follows. Paired-end reads raw data were converted to appropriate fastq format using FASTQ groomer and their quality were confirmed using FASTQC (maximum quality scores were well maintained all over the 150bp, not shown). The genome sequences were reconstructed using the published workflow “SNP calling on paired end data” for the mapping of the paired reads against a reference *C. reinhardtii* genome sequence (that of our WT strain T222+ was generated by Olivier Vallon (personal communication) as described in ref Merchant 2017). We visualized the genomes using IGV (software.broadinstitute.org).

BIOCHEMISTRY

Whole-cell proteins were separated by SDS-PAGE 8M urea. Proteins were electrotranferred onto nitrocellulose membranes in a semi-dry blotting transfer apparatus. Heme *f* peroxidase activity was detected on blot membranes by using the chemiluminescence. For immunodetection, the blot membranes were incubated separately in primary antibodies against ATP synthase subunits, FtsH1/2 subunits, and loading control proteins as RNAse J (RNJ) and cytochrome *f*, and revealed by secondary antibody conjugated with HRP by the chemiluminescence method. WT was loaded in decreasing amount as indicated by fractions. The mutant strains *ΔatpB*, *ΔatpH*, *ΔatpI* were obtained by homologous recombination

(deletion of 5'-UTR and CDS); the nuclear mutations mdb1 and mrl1 impair AtpB and RbcL accumulation, respectively

RNA

Total RNA were extracted in phenol-chloroform, separated on agarose, transferred to membrane by capillarity and UV-crosslinked. Probes were produced by PCR in the presence of digoxigenin-UTP. After hybridization, membrane was incubated in the presence of anti-digoxigenin antibody conjugated to alkaline phosphatase and revealed by chemiluminescence in the presence of CDP-Star.

REFERENCES

Baker, N. R. (2008) Chlorophyll fluorescence: a probe of photosynthesis in vivo. Annual review of plant biology, 59, 89-113.

Barkan, A. and Goldschmidt-Clermont, M. (2000) Participation of nuclear genes in chloroplast gene expression. Biochimie, 82, 559-572.

Drapier, D., Girard-Bascou, J. and Wollman, F. A. (1992) Evidence for Nuclear Control of the Expression of the atpA and atpB Chloroplast Genes in Chlamydomonas. The Plant cell, 4, 283-295.

Drapier, D., Rimbault, B., Vallon, O., Wollman, F. A. and Choquet, Y. (2007) Intertwined translational regulations set uneven stoichiometry of chloroplast ATP synthase subunits. The EMBO journal, 26, 3581-3591.

Hahn, A., Vonck, J., Mills, D. J., Meier, T. and Kuhlbrandt, W. (2018) Structure, mechanism, and regulation of the chloroplast ATP synthase. Science, 360.

Johnson, X., Steinbeck, J., Dent, R. M., Takahashi, H., Richaud, P., Ozawa, S., Houille-Vernes, L., Petroutsos, D., Rappaport, F., Grossman, A. R., Niyogi, K. K., Hippler, M. and Alric, J. (2014) Proton gradient regulation 5-mediated cyclic electron flow under ATP- or redox-limited conditions: a study of DeltaATpase pgr5 and DeltarbcL pgr5 mutants in the green alga Chlamydomonas reinhardtii. Plant physiology, 165, 438-452.

Johnson, X., Vandystadt, G., Bujaldon, S., Wollman, F. A., Dubois, R., Roussel, P., Alric, J. and Beal, D. (2009) A new setup for in vivo fluorescence imaging of photosynthetic activity. Photosynthesis research, 102, 85-93.

Johnson, X., Wostrikoff, K., Finazzi, G., Kuras, R., Schwarz, C., Bujaldon, S., Nickelsen, J., Stern, D. B., Wollman, F. A. and Vallon, O. (2010) MRL1, a conserved Pentatricopeptide repeat protein, is required for stabilization of rbcL mRNA in Chlamydomonas and Arabidopsis. The Plant cell, 22, 234-248.

Lemaire, C. and Wollman, F. (1989a) The Chloroplast ATP Synthase in Chlamydomonas reinhardtii. 1. CHARACTERIZATION OF ITS NINE CONSTITUTIVE SUBUNITS. Journal of Biological Chemistry, 264, 10228-10234.

Lemaire, C. and Wollman, F. A. (1989b) The Chloroplast ATP Synthase in Chlamydomonas reinhardtii. 2. BIOCHEMICAL-STUDIES ON ITS BIOGENESIS USING MUTANTS DEFECTIVE IN PHOTOPHOSPHORYLATION. JOURNAL OF BIOLOGICAL CHEMISTRY, 264, 10235-10242.

Majeran, W., Olive, J., Drapier, D., Vallon, O. and Wollman, F. A. (2001) The light sensitivity of ATP synthase mutants of Chlamydomonas reinhardtii. Plant physiology, 126, 421-433.

Malnoe, A., Wang, F., Girard-Bascou, J., Wollman, F. A. and de Vitry, C. (2014) Thylakoid FtsH protease contributes to photosystem II and cytochrome b6f remodeling in Chlamydomonas reinhardtii under stress conditions. The Plant cell, 26, 373-390.

Ruhle, T. and Leister, D. (2015) Assembly of F1F0-ATP synthases. Biochim Biophys Acta, 1847, 849-860.

Viola, S., Cavaiuolo, M., Drapier, D., Eberhard, S., Vallon, O., Wollman, F. A. and Choquet, Y. (2019) MDA1, a nucleus-encoded factor involved in the stabilization and processing of the atpA transcript in the chloroplast of Chlamydomonas. The Plant journal : for cell and molecular biology.

Wang, F., Qi, Y., Malnoe, A., Choquet, Y., Wollman, F. A. and de Vitry, C. (2017) The High Light Response and Redox Control of Thylakoid FtsH Protease in Chlamydomonas reinhardtii. Molecular plant, 10, 99-114.

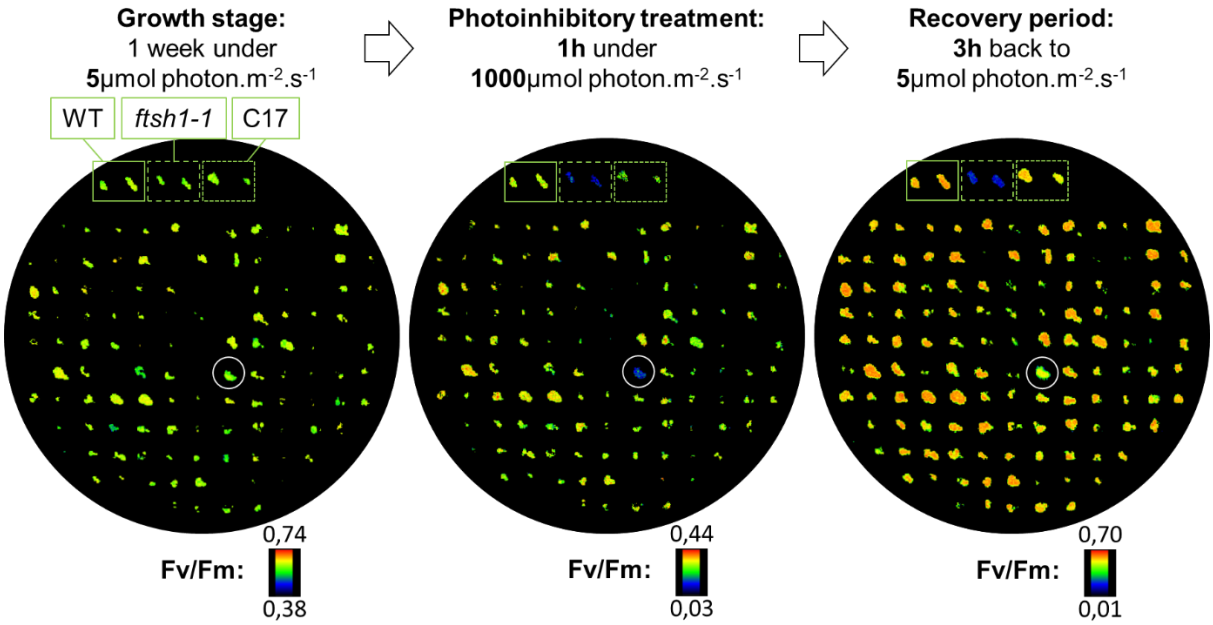


Figure 1. Mutants affected by short photoinhibitory treatment can be isolated by screening on chlorophyll fluorescence.

After transformation, colonies were transferred to fresh TAP medium and grown mixotrophically in low light (5µmol photon m^{-2} s^{-1}) for 7 days. Chlorophyll fluorescence analysis was performed before (left) and after (centre) photoinhibitory treatment of 1h at 1000µmol photon m^{-2} s^{-1}. Immediately after this, plates were put back to 5µmol photon m^{-2} s^{-1} for 3h (recovery period) before fluorescence was assessed again (right). Depicting evolution of PSII function throughout these treatments, shown are the maximum quantum efficiency of PSII photochemistry Fv/Fm (false-colour scale) for the plate from which the mutant E236 was isolated (white circle). Were introduced as controls the wild-type (WT), the *ftsh1-1* mutant and C17, the complemented strain *ftsh1-1*(pSL18-FTSH1).

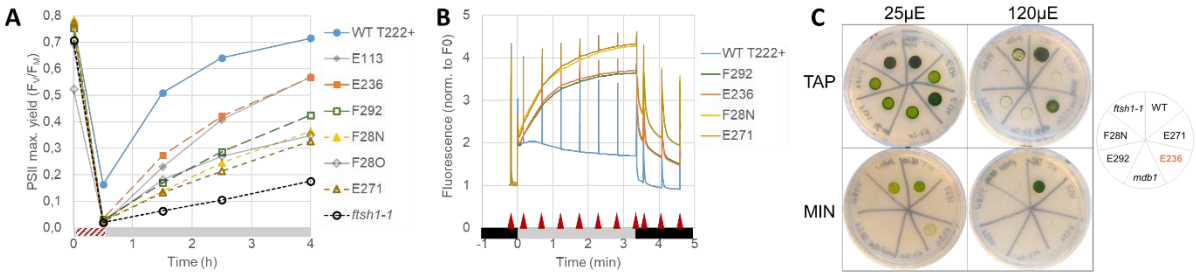


Figure 2. Fluorescence kinetics and growth tests point to ATP synthase defects in four mutants.

Cells were grown in low light (5 µmol photon m^{-2} s^{-1}) in liquid TAP. **A:** Time course of PSII maximal yield F_V/F_M upon photoinhibitory treatment (30 min 1000 µmol photon m^{-2} s^{-1}, red hatched box) and recovery (grey box). **B:** Chlorophyll fluorescence kinetics were monitored in the transition from dark (black box) to low light (grey box). Saturating pulses (red triangles) were applied to estimate PSII yield in the light. **C:** Cells were spotted onto plates containing TAP (top panels) and minimal (MIN, bottom panels) and grown for 6 days under moderate light (25 µmol photon m^{-2} s^{-1}, left panels) or saturating light (120 µmol photon m^{-2} s^{-1}, right panels). As a control, the *mdb1* mutant was introduced, which lacks ATP synthase because the maturation factor MDB1 is required for the maturation of *atpB* transcript.

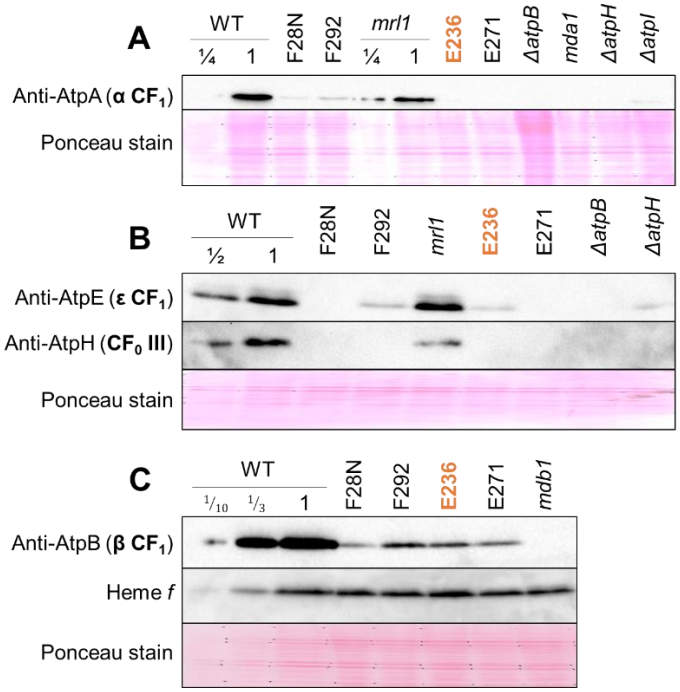


Figure 3. Western blot analysis suggests that the mutations impair ATP synthase accumulation/assembly at distinct steps

Whole-cell protein extracts were separated by SDS/urea-PAGE (**A:** 16%; **B** and **C:** 12-18%) and analysed by immunodetection using antibodies against ATP synthase subunits. Transfers were checked by Ponceau staining; heme *f* (from cytochrome. *f*) was detected by its peroxidase activity as a loading control. Controls were loaded in decreasing amount as indicated by fraction numbers.

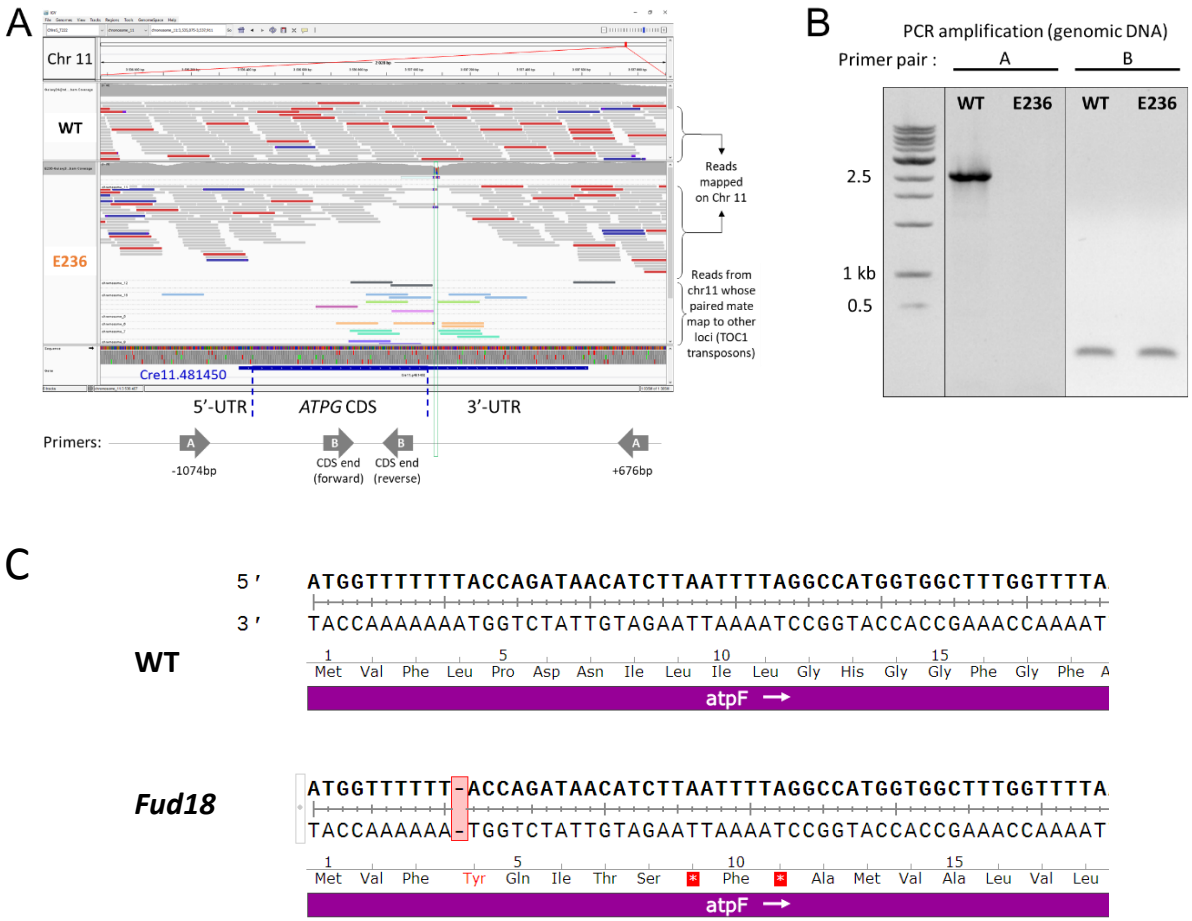


Figure 4. Genome sequencing reveals transposon insertion in the 3'-UTR of *ATPG* and the one nucleotide deletion leading to a frame-shift in *atpF*.

Paired-end sequencing reads were mapped on WT T222+ genome. **A:** Read mapping in WT (top) and E236 (bottom) centred on the Cre11.481450 locus (*ATPG*). In grey, reads with both ends mapping in neighbouring areas (red and blue reads denote interspace that are longer or shorter than the expected range, respectively); in light colours, reads with ends mapping on distinct chromosomic areas, in this case both chromosome 11 (Chr 11) and *TOC1* transposon sequences. Below are shown the gene model for *ATPG* (5'-untranslated region (UTR), coding sequence (CDS) and 3'-UTR) and the map of primers used for PCR amplification (numbers stand for the distance upstream ATG start codon (-) and downstream 3'-UTR end (+). **B:** PCR products obtained on WT and E236 genomic extracts were separated on agarose. **C:** The *Fud18* mutant has a deletion of one T after the start codon of *atpF* which causes a frameshift and an abortive Stop codon.

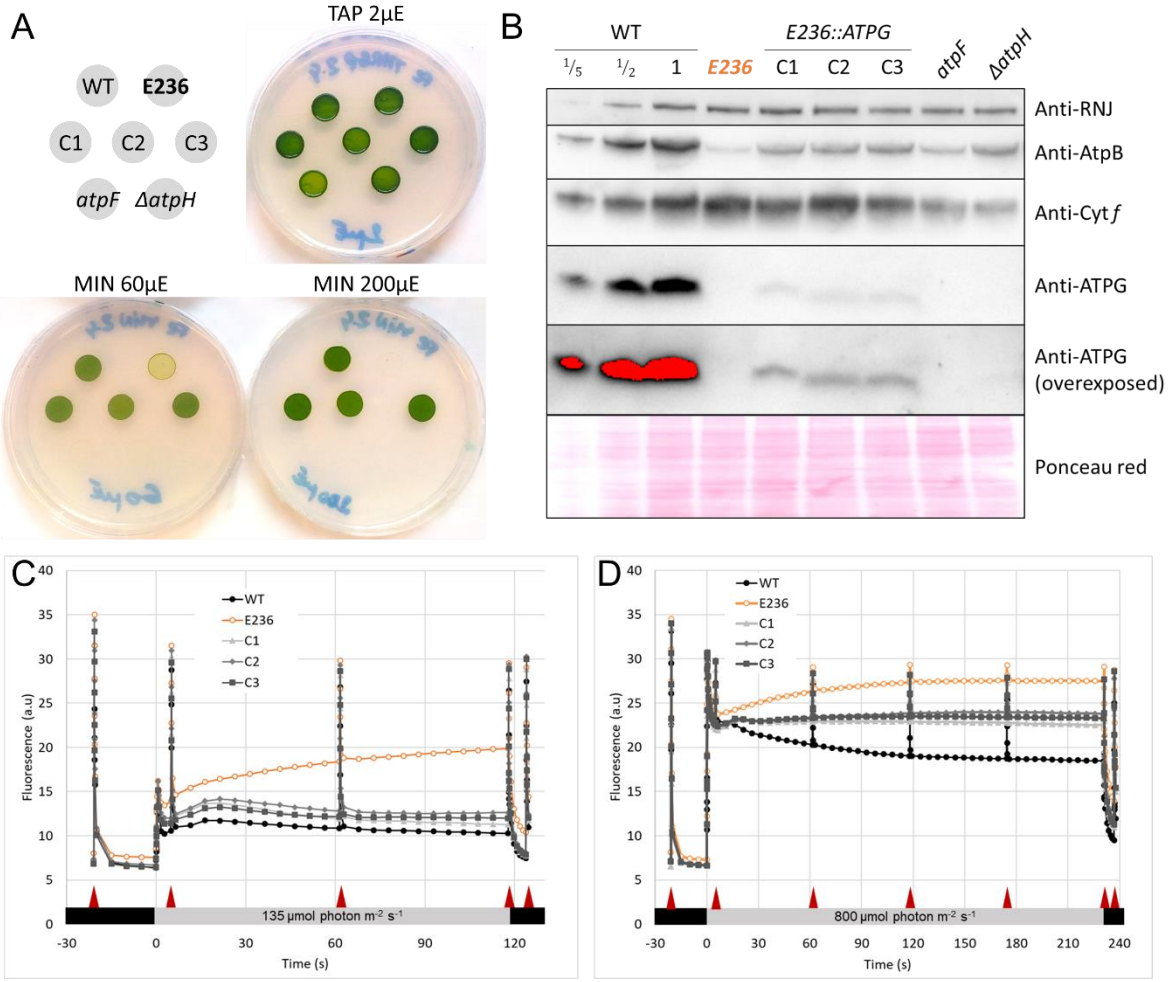
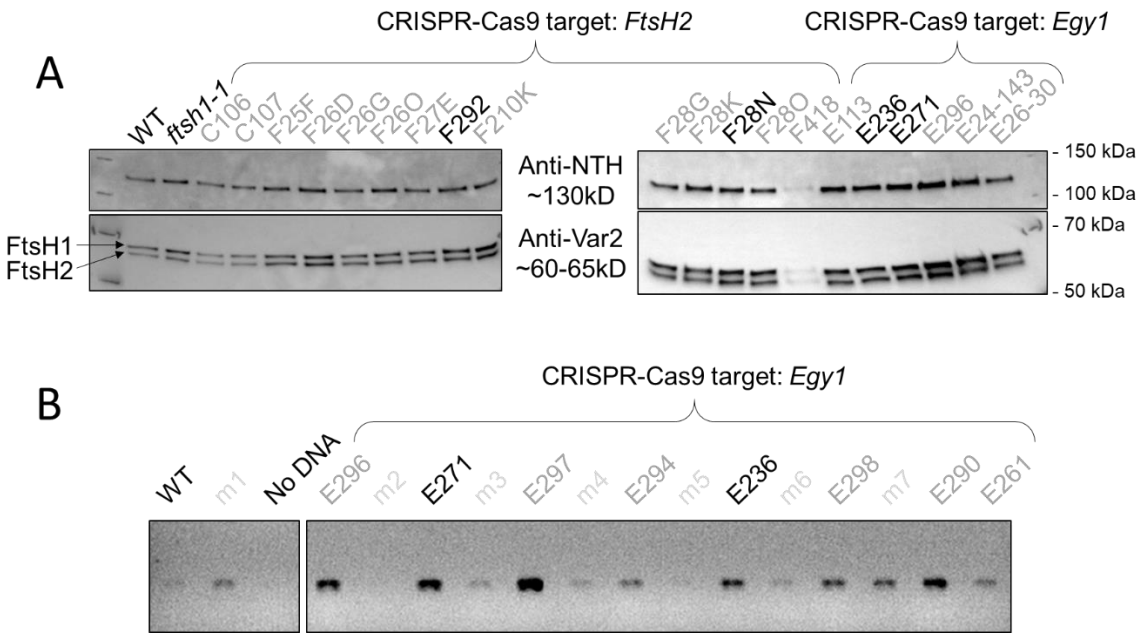


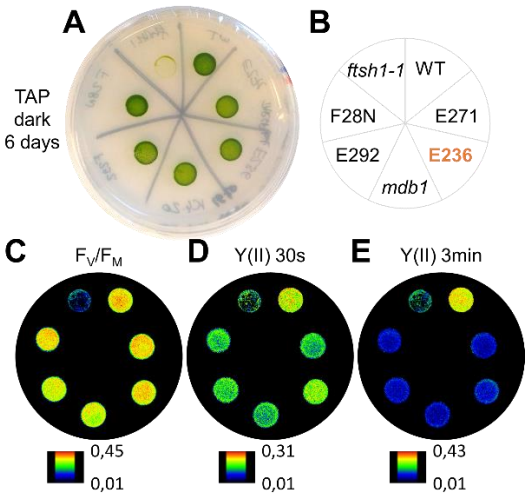
Figure 5. Growth phenotype in mutant E236 is complemented by WT version of *ATPG*, although only partially restoring ATP synthase accumulation and photosynthetic electron flow.

E236 mutant cells were electroporated in the presence of WT *ATPG* and grown on MIN in high light. Three transformants were randomly selected (C1-C3) and grown in liquid TAP under low light for the following analyses. **A:** Cells were plated on TAP medium (control, very low light: 2 µmol photon m^{-2} s^{-1}) and on MIN medium for growth under non-saturating light (60 µmol photon m^{-2} s^{-1}) or excess light (200 µmol photon m^{-2} s^{-1}). **B:** Whole-cell protein extracts were separated by SDS-PAGE and detected by immunoblot. **C:** and **D:** Fluorescence was recorded upon illumination under moderate (C: 135 µmol photon m^{-2} s^{-1}) or high (D: 800 µmol photon m^{-2} s^{-1}) levels of green actinic light. Saturating flashes (red arrows) were given before (black boxes) and during illumination (grey boxes).



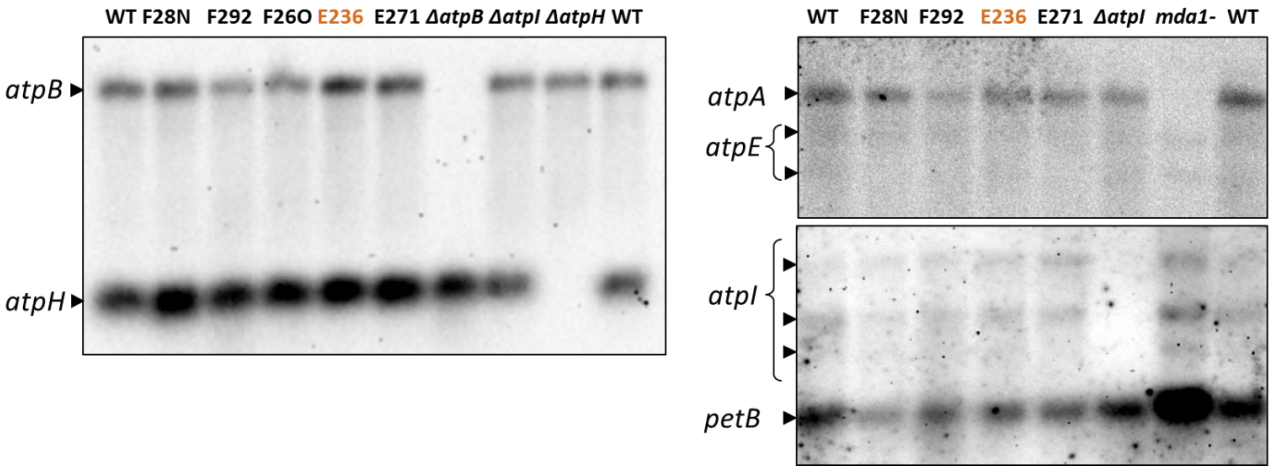
Supplemental Figure 1. Selected mutants were analysed at molecular level.

Black characters denote strains demonstrating strong and robust fluorescence phenotypes in contrast with dark grey characters. **A:** Western blot analyses using anti-Var2 (FtsH1/2) and anti-NTH (as a loading control) antibodies on whole-cell protein extracts. Photosensitive control strain *ftsh1-1* accumulates a mutated version (R420C) of FtsH1. **B:** PCR products in the CRISPR-targeted area of *EGY1*. Mutants m1-m7 were pale green clones which yielded very few DNA (when re-analyzed, they exhibited WT-like PCR products (not shown)).



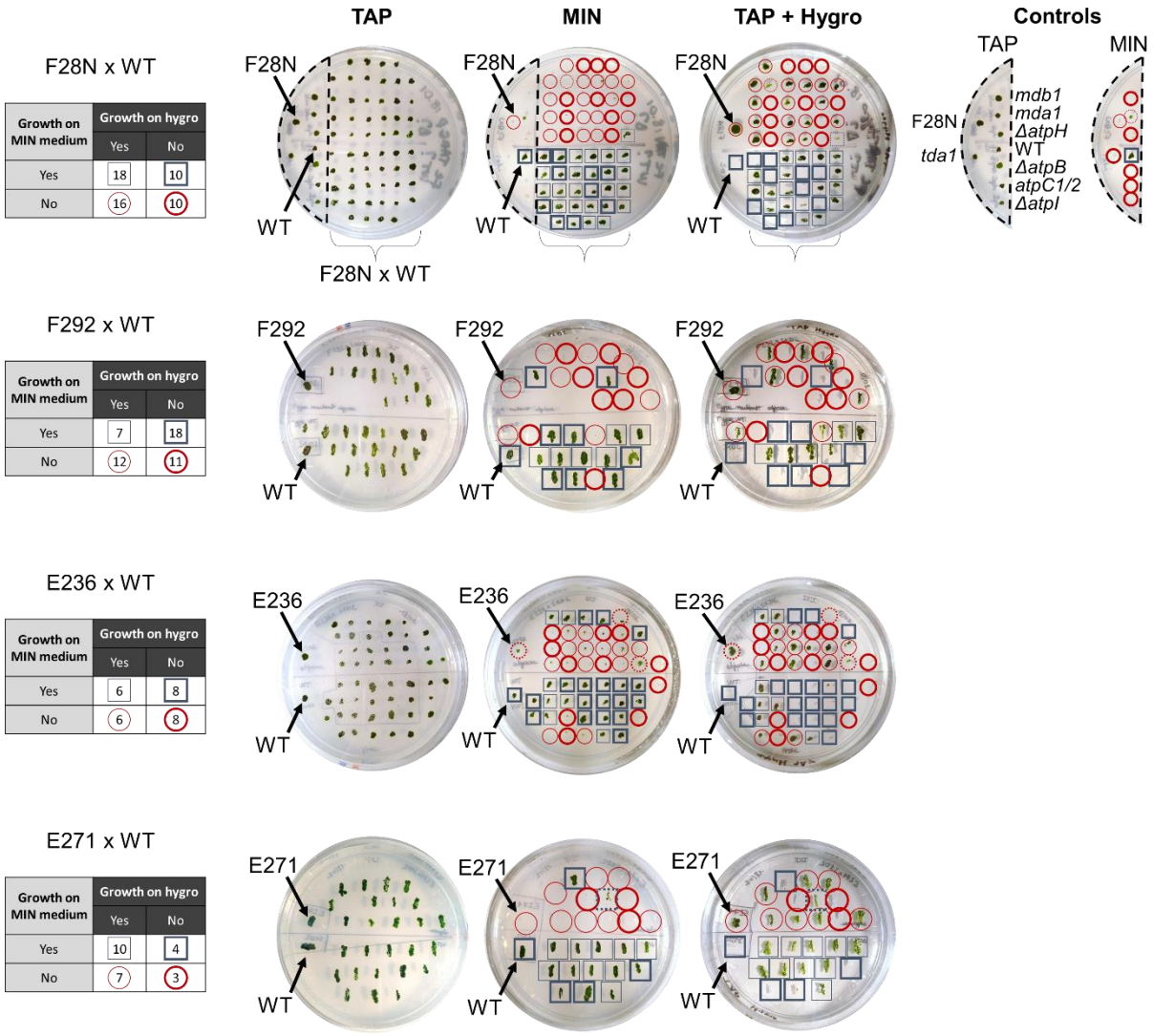
Supplemental Figure 2. Fluorescence kinetics of dark-grown cells illustrates phenotypic similarities between selected mutants and *mdb1*.

Cells were grown in low light (5 µmol photon m^{-2} s^{-1}) in liquid TAP then spotted onto TAP plates and grown in the dark. As a control was introduced the *mdb1* mutant, which lacks ATP synthase because the maturation factor MDB1 is required for the maturation of *atpB* transcript. **A:** plate after 6 days of growth. **B:** Diagram of strains on the plate. **C:** PSII maximal yield F_V/F_M. **D** and **E:** Chlorophyll fluorescence kinetics were monitored in the transition from dark to low light (green actinic light: 120 µmol photon m^{-2} s^{-1}). PSII yield Y(II) after 30 s (D) and 3 min (E) of green light.



Supplemental Figure 3. Northern blot analysis reveals no defect in *atpA*, *atpB*, *atpH, atpE* nor *atpI* transcript levels in F28N, F292, E236 and E271mutants. Three separate blots were hybridised with digoxigenin labelled probes.

Total RNA extracts were separated on agarose gels and analysed by immunodetection in the presence of specific digoxigenin-UTP probes.



F28N x WT

Growth on MIN medium	Growth on hygro	
	Yes	No
Yes	18	10
No	16	10

F292 x WT

Growth on MIN medium	Growth on hygro	
	Yes	No
Yes	7	18
No	12	11

E236 x WT

Growth on MIN medium	Growth on hygro	
	Yes	No
Yes	6	8
No	6	8

E271 x WT

Growth on MIN medium	Growth on hygro	
	Yes	No
Yes	10	4
No	7	3

Supplemental Figure 4. Absence of co-segregation between non-phototroph phenotype and hygromycin resistance in the progeny from backcrosses of ATP synthase mutants with WT. Mutants were backcrossed to the WT strain S1D2 and the progeny were plated in the same order on TAP (control), MIN medium (obligate phototrophic growth) and TAP + 20mg.L^{-1} hygromycin (resistance cassette required for growth). Clones growing on MIN are identified by squares while obligate phototrophs are in circles (dashed lines for clones in-between). Clones growing or not on TAP + hygromycin are identified by thin or thick lines, respectively. The clone count is summarized in the tables on the left of each cross. As supplementary controls, we also plated on TAP and MIN medium some ATP synthase mutants, located in the dashed semi-circle (top panels) and reported in the top right inset.

ARTICLE 3:

"THE OPR PROTEIN MTHI1 CONTROLS THE EXPRESSION OF TWO DIFFERENT SUBUNITS OF ATP SYNTHASE CFO IN CHLAMYDOMONAS REINHARDTII."

This article is under review

The OPR protein MTHI1 controls the expression of two different subunits of ATP synthase CFo in *Chlamydomonas reinhardtii*

Shin-Ichiro Ozawa[a,b], Marina Cavaiuolo[a], Domitille Jarrige[a], Richard Kuras[a], Mark Rutgers[a], Stephan Eberhard[a], Dominique Drapier[a], Francis-André Wollman[a], and Yves Choquet[a]*[1]

[a]: Unité Mixte de Recherche (UMR) 7141, Centre National de la Recherche Scientifique (CNRS) and Sorbonne Université, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, F-75005 Paris, France

[b]: Research Institute for Interdisciplinary Science, Okayama University, Okayama 700-8530, Japan

*: Corresponding author: choquet@ibpc.fr

Phone : +33 (0)1 58 41 50 75

Running title: Post transcriptional control of CFo biogenesis

One sentence summary: The OPR protein MTHI1 is a major actor in the biogenesis of the chloroplast ATP synthase that co-regulates the expression of AtpH and AtpI, the two subunits of the proton channel.

Keywords: Nucleus-encoded trans-acting factors; Chloroplast biogenesis, ATP synthase; OPR proteins; translation activation; Chlamydomonas.

ACKNOWLEDGMENTS: We thank Rachel Dent and Silvia Ramundo for their kind gifts of the CAL014.01.38 mutant and of α-Rps12 antibody, respectively, Jean-David Rochaix for kindly providing the indexed cosmid library and all members of UMR7141 for stimulating discussions and/or critical reading of the manuscript.

FUNDING: This work was supported by basic support to Unité Mixte de Recherche 7141, from CNRS and Sorbonne Université, by Agence Nationale de la Recherche (ChloroMitoCES: BLAN-NT09_451610 and ChloroRNP: ANR-13-BSV7-0001-01), and by the "Initiative d'Excellence" program (Grant "DYNAMO," ANR-11-LABX-0011-01).

[1] Address correspondence to choquet@ibpc.fr. The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Yves Choquet (choquet@ibpc.fr).

Abstract:

In *Chlamydomonas reinhardtii*, chloroplast gene expression is tightly regulated post-transcriptionally by gene-specific *trans*-acting protein factors. Here we report the identification of an OctotricoPeptide Repeat (OPR) protein, MTHI1, critical for the biogenesis of chloroplast ATP synthase CFo. At variance with most trans-acting factors characterised so far in *C. reinhardtii* that control the expression of a single gene, MTHI1 targets two distinct transcripts: it is required for the accumulation and the translation of the *atpH* mRNA, encoding a subunit of the selective proton channel but it also enhances the translation of the *atpI* mRNA, which encodes the other subunit of the channel. MTHI1 targets the 5'UTR of both *atpH* and *atpI* genes. Co-immuno-precipitation and small RNA sequencing revealed that MTHI1 binds specifically a sequence highly conserved among Chlorophyceae and the Ulvale clade of Ulvophyceae at the 5'end of the tri-phosphorylated *atpH* mRNA. A very similar sequence, located about 60 nt upstream of the *atpI* initiation codon, was also found in some Chlorophyceae and Ulvale species and is essential for *atpI* mRNA translation in *C. reinhardtii*. Such a dual targeted *trans*-acting factor, thus provides a mean to co-regulate the expression of the two proton hemi-channels.

Introduction:

In chloroplasts, photosynthetic energy conversion is performed by oligomeric protein complexes made of subunits of dual genetic origin. Indeed, due to the extensive gene transfer from the cyanobacterial ancestor of chloroplasts to the nucleus of the host cell, only some subunits of the photosynthetic apparatus are still organelle-encoded, whereas others are expressed in the nucleo-cytosol, then imported into organelles. Thus, assembly of photosynthetic protein complexes requires a tight cooperation between two genetic compartments to avoid the wasteful or even deleterious accumulation of unassembled subunits. A first level of coordination between the two genetic compartments is brought by a plethora of nucleus-encoded factors that tightly control each post-transcriptional step of chloroplast gene expression (processing, trimming, splicing, editing, stabilisation, translation activation and decay of chloroplast RNAs; reviewed in (Barkan and Goldschmidt-Clermont, 2000; Schmitz-Linneweber and Small, 2008; Woodson and Chory, 2008; Germain et al., 2013; Zoschke and Bock, 2018). Thanks to this nuclear control of chloroplast gene expression that emerged after endosymbiosis, gene expression remains commensurate in the chloroplast and nucleo-cytosol, despite a huge imbalance in gene copy number that may differ by as much as four orders of magnitude. In the unicellular green alga *Chlamydomonas reinhardtii*, nucleus-encoded factors mostly belong to two major functional classes: the M factors involved in chloroplast mRNA maturation and stabilisation and the T factors required for mRNA translation activation (Choquet and Wollman, 2002). Most of these factors pertain to helical repeat protein families, such as PPR (PentatricoPeptide Repeat), HAT (Half A Tetratricopeptide repeat), mTERF (mitochondrial TERmination Factor) or OPR (OctatricoPeptide Repeat) proteins (reviewed in Barkan and Small, 2014; Hammani et al., 2014). These proteins comprise tandem repeats of simple structural motives that fold into antiparallel α-helices and stack on each other to form a concave surface, well suited to interact with RNA molecules. Each repeat contacts, though amino acids at determined position, one specific nucleotide thereby allowing sequence specific recognition. While the PPR family has greatly expanded in land plants, with more than 450 members in Arabidopsis or rice, it remained limited in green algae (14 PPR proteins in *C. reinhardtii* (Tourasse et al., 2013), which instead expresses numerous OPR proteins (>125 in *C. reinhardtii* vs. only 1 in Arabidopsis).

Beside this nuclear control on chloroplast gene expression, other fine-tuning regulations set the synthesis of the individual subunits of a photosynthetic protein to the

stoichiometry required for their functional assembly, as shown by the pleiotropic loss of all subunits of a complex in any mutant lacking expression of one of its major subunits. Two major mechanisms account for this concerted accumulation in *C. reinhardtii* (reviewed by (Choquet and Vallon, 2000). Some subunits, particularly those encoded in the nucleus, are expressed normally but rapidly degraded when they cannot assemble, while many chloroplast-encoded subunits of the photosynthetic apparatus show an assembly-dependent regulation on their synthesis, called "control by epistasy of synthesis" or CES process (Choquet and Wollman, 2009). In absence of their assembly partners the rate of synthesis of CES subunits is dramatically reduced. In most cases, the CES process relies on a negative feedback mediated by the unassembled CES subunit on its own translation (Choquet et al., 1988; Choquet et al., 2003; Wostrikoff et al., 2004; Minai et al., 2006; Wostrikoff and Stern, 2007; Choquet and Wollman, 2009). However, the CES processes that control the biogenesis of the CF1 sector of ATP synthase, present atypical features, which account for the 3:3:1 uneven stoichiometry in the synthesis of the α, β and γ subunits (Drapier et al., 2007).

The mechanisms ensuring the 1:1:14:1 accumulation of the AtpF, AtpG, AtpH and AtpI subunits of the CFo sector have not been investigated so far. The acetate-requiring *ac46* mutant, isolated by Levine more than half-a-century ago (Levine, 1960), was latter characterised as defective for photosynthesis, because of a single nuclear mutation (Levine and Goodenough, 1970). It does not express the chloroplast-encoded AtpH subunit (Lemaire and Wollman, 1989a), because it does not accumulate the monocistronic *atpH* mRNA (Majeran et al., 2001). Beside defective expression of AtpH, this mutant also shows a strongly reduced synthesis of AtpI, another chloroplast-encoded CFo subunit (Lemaire and Wollman, 1989a), which, together with the tetra-decameric ring of AtpH subunits, forms the membrane-embedded proton channel. The mutation thus affects the *MTHI1* gene whose product is required for the **M**aturation/stability and **T**ranslation of the *atp**H*** and *atp**I*** mRNAs and the *ac46* mutation was renamed *mthi1-1*. The coupled expression of AtpH and AtpI was possibly indicative of a CES relationship. In the mitochondria of the yeast *Saccharomyces cerevisiae*, mutants lacking expression of Atp9p, the mitochondrion-encoded counterpart of AtpH, show reduced synthesis of Atp6p and Atp8p, the former being the mitochondrial equivalent of AtpI (Jean-Francois et al., 1986; Ooi et al., 1987; Payne et al., 1991; Bietenhader et al., 2012). Together, these results prompted us to investigate the expression of the *atpH* and *atpI* genes in *mthi1* mutants.

114

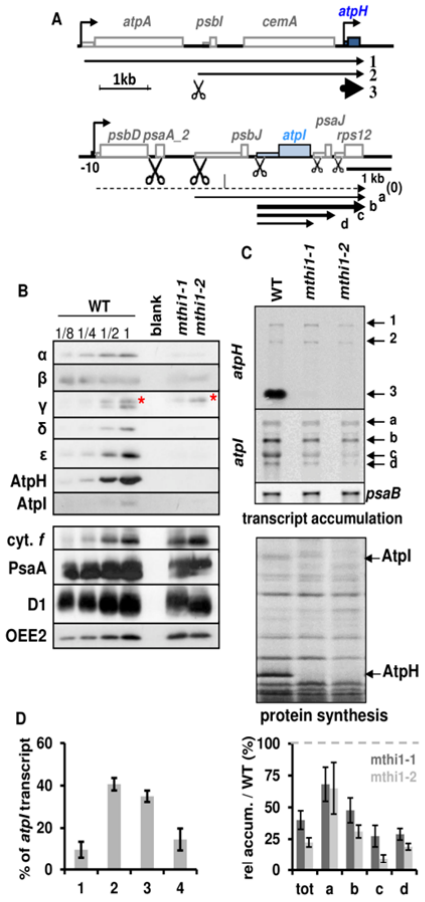
Results:

Lack of MTHI1 leads to a reduced accumulation and translation of the *atpI* mRNA.

We recovered a photosynthetic mutant, kindly provided by Rachel Dent, generated by insertional mutagenesis with a paromomycin resistance cassette (Dent et al., 2005), originally called CAL014.01.38). It shows the same phenotype than *mthi1-1*, lacks the *atpH* mRNA, hence AtpH synthesis (Fig. 1C) and accumulation of all subunits of the ATP synthase complex (Fig. 1B). In addition, it shows a strongly reduced synthesis of AtpI in -^{14}C pulse labelling experiments (Fig. 1C) and was renamed *mthi1-2*.

The reduced synthesis of AtpIt prompted us to monitor the accumulation of its transcript in *mthi1* mutants. The *atpI* gene belongs to a polycistronic transcription unit, which comprises *psbD*, the second exon of *psaA*, *psbJ, atpI*, *psaJ*, and *rps12* (Cavaiuolo et al., 2017 and Fig. 1A). As previously reported (Liu et al., 1989; Rymarquis et al., 2006; Jalal et al., 2015) and illustrated in Fig. 1A,C, the wild type displays four major *atpI* transcripts, respectively the *psbJ-atpI-psaJ-rps12*, *atpI-psaJ-rps12*, *atpI-psaJ* and *atpI* tetra-, tri-, di- and mono-cistronic transcripts. The tri and di-cistronic transcripts account for 75% of the *atpI*-containing mRNAs. In *mthi1* mutants, the accumulation of *atpI* transcripts was reduced by ~60%, the di- and mono-cistronic transcripts being the most reduced, by ~85% and ~75%, respectively (Fig. 1C,D).

To understand whether this reduced transcript level was responsible for the reduced synthesis of AtpI in *mthi1* mutants, we compared the loading of *atpI* transcripts on polysomes in the wild type and in three strains lacking expression of AtpH: Δ*atpH*, an *atpH* deletion strain (see Table I for strains constructed in that study) and *mthi1-1, -2* (Fig. 2A). Free mRNAs and dissociated 50S and 30S ribosome subunits are found in "light" fractions (10 to 6) of sucrose gradients, while transcripts found in "heavy" fractions (5 to 1) correspond to polysomes of increasing sizes (Minai et al., 2006; Eberhard et al., 2011). The distribution of the *psbD* mRNA, whose expression is unrelated to ATP synthase biogenesis, was unchanged in the three mutant strains and the wild type, with a peak centred on fractions 4-5. The distribution of the four *atpI* transcripts was similar in the wild-type and Δ*atpH* strains, with a peak centred on fraction 4. In *mthi1* mutants, the distribution of the tetra- and tri-cistronic transcripts, were similar to that in the wild type, probably because these transcripts respectively comprise 3 and 2 open reading frames in addition to the *atpI* coding sequence. In stark contrast, the two smaller transcripts were virtually absent in fractions 1-5 and mostly



Ozawa et al, Fig. 1

Figure 1: Phenotype of mthi1 mutants

A) Schematic representation of the *atpH* (top) and *atpI* (bottom) transcription units. Coding sequences are shown as thick rectangles, while 5'UTRs are depicted as thin rectangles. Bent arrows represent promoters. The major transcripts detected in panel C with probes specific to *atpH* or *atpI* are indicated. (0) stand for a precursor transcript that cannot be observed in the wild type because it is efficiently processed, but can be detected in *psaA* trans-splicing mutants (Choquet et al, 1988). Scissors indicate the position of processing events, whose efficiency symbolised by their size.

B) Pleiotropic loss of ATP synthase subunits in *mthi1* mutants.

Total cell extracts of wild-type (a dilution series is shown) and of the two *mthi1* mutants strains were probed with antibodies against the proteins indicated on the left. The accumulation of all ATP synthase subunits was dramatically decreased in the two mutant strains, while that of cyt. *f*, PsaA, D1 and OEE2, respectively used as proxies of the abundance of the cytochrome b_6f complex, photosystem I (PSI) and PSII was unaffected. The red asterisk points to a cross-reaction of the antibody, preserved in the mutant strains, against the γ subunit of mitochondrial ATP synthase.

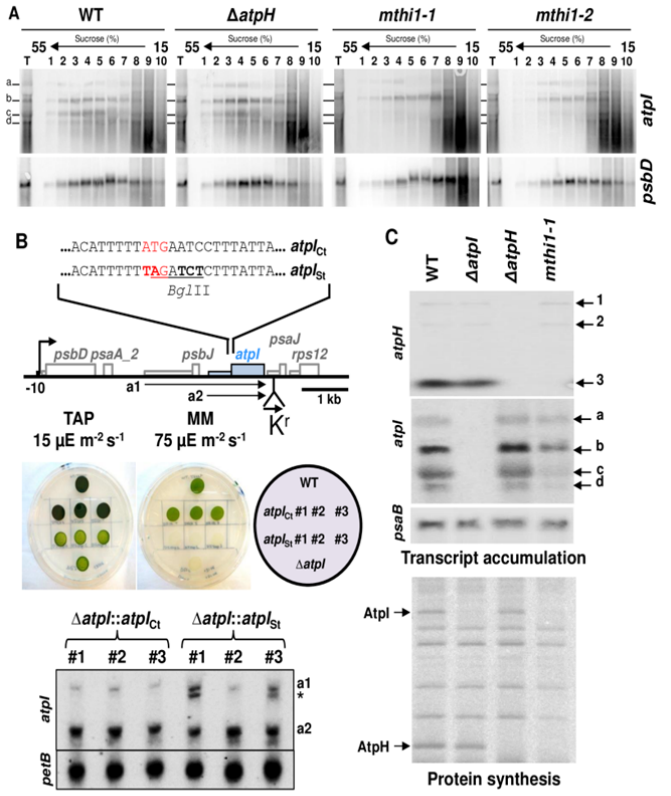
C) (Top) Accumulation of the *atpH* and *atpI* transcripts in the same strains. The *psaB* transcript provides a loading control.

(Bottom) Rate of translation of ATP synthase subunits in the same strains, assessed by 5' pulse labelling experiments in the presence of ^{14}C acetate (5 µCi.mL^{-1}) and of the cycloheximide inhibitor of cytosolic translation (10 µg.mL^{-1}). The positions of the AtpI and AtpH subunits are indicated.

D) Quantification of *atpI* transcripts amount in wild-type and mutant strains.

(Left) Relative accumulation of the four *atpI*-containing transcript in the wild type, expressed as the percentage of the total amount of *atpI* transcript. (Right) Relative abundance of each *atpI* transcript, and of the sum of them, reported to that of the same band in the wild-type (set to 100, symbolised by a grey dashed line) in the two mutants (dark grey *mthi1-1*; light grey *mthi1-2*; n=4).

148 found in fractions 7-9 (Fig. 2A). Their exclusion from polysomal fractions shows that the



Ozawa et al, Fig. 2

Figure 2: The MTHI1 factor controls the translation of the atpI mRNA

A) Loading of *atpI* mRNAs on polysomes.

Solubilized whole-cell extracts (T) from wild-type, Δ*atpH* and the two *mthi1* mutant strains, pre-treated for 10 min with CAP (200 µg mL^{-1}) were loaded on sucrose gradients. After ultracentrifugation, ten fractions were collected and the transcripts present in each fraction were analysed by RNA blots using the probes indicated on the left.

B) Defective *atpI* mRNA translation is not responsible for its decreased abundance in *mthi1* mutants.

(top) Schematic representation of the changes introduced into the *atpI* gene: mutated nucleotides are shown in bold: they change the initiation codon (written in red) to a stop codon and introduced a *Bgl*II RFLP marker (underlined).

(middle) Phototrophic growth of the *atpI*$_{St}$ and *atpI*$_{Ct}$ strains assessed on minimum medium (devoid of acetate) under 75 µE m^{-2} s^{-1}. Three independent transformants are shown. The growth of the strain on TAP medium (15 µE m^{-2} s^{-1}), as well as the growth of the wild type and of the Δ*atpI* strain are shown as controls.

(bottom) Accumulation of *atpI* transcripts, schematically depicted in panelA, in a control strain bearing the *aadA* cassette alone and in strains bearing the *aadA* cassette associated with the untranslatable *atpI*$_{St}$ gene. Three independent transformants are shown for each construct. Because of the polar effect of *aadA* cassette, co-transcripts with *atpJ* and/or *rps12* cannot be observed. The origin of the transcripts indicated by an aterisk (*) is unknown. *petB* provides a loading control.C) *atpH* and *atpI* gene expression in the wild-type, Δ*atpH*, Δ*atpI* and *mthi1-1* strains.

C) *atpH* and *atpI* gene expression in the wild-type, Δ*atpH*, Δ*atpI* and *mthi1-1* strains.

(top) accumulation of the *atpH* and *atpI* transcripts. *psaB* provides a loading control.

(bottom) Rate of translation of the *atpH* and *atpI* transcripts in the same strains, assessed as in Fig. 1B by pulse labelling experiments. The positions of the AtpI and AtpH subunits are indicated.

149 reduced synthesis of AtpI was not only due to the reduced accumulation of *atpI* transcripts nor

to an increased and rapid proteolytic disposal of AtpI in the absence of its assembly partner, AtpH. Rather, the translation of *atpI* transcripts is severely impaired in *mthi1* mutants.

To further assess the relationship between *atpI* transcript accumulation and translation, we constructed an untranslatable version of the *atpI* gene, $atpI_{St}$, whose initiation codon was replaced by an amber codon (Fig. 2B). This mutated *atpI* gene was, as all chimera and mutated genes used in this study (Table I), associated with an *aadA* cassette to select transformants for spectinomycin resistance. After transformation it replaced the endogenous *atpI* gene. Because not synthesising the AtpI subuni, transformants were unable of phototrophic growth (Fig. 2B). However, the mutated *atpI* transcripts accumulated to the same levels than in control strains transformed with an unmodified *atpI* gene, just associated with the *aadA* cassette (Fig. 2B): the reduced accumulation of *atpI* mRNA in *mthi1* mutants is not due to impaired translation but to the lack of MTHI1 that therefore does not only activate the translation of the *atpI* mRNA but also contributes to its stabilisation.

AtpI and AtpH are synthesised independently

The reduced translation of *atpI* transcripts in *mthi1* mutants could be explained in two ways: either MTHI1 is a bi-functional protein required for the stable accumulation of the *atpH* mRNA and for the translation of the *atpI* transcript or AtpI is a CES subunit, requiring the presence of AtpH to be synthesised at sustained rates, as in yeast. The similar loading of *atpI* transcripts on polysomes in wild-type and Δ*atpH* strains (Fig. 2A) strongly argues against the latter hypothesis. As a further challenge, we compared the translation of the *atpH* and *atpI* mRNAs by pulse labelling experiments in strains *mthi1-1*, Δ*atpH* and Δ*atpI*. While the synthesis of AtpI was strongly reduced in the *mthi1-1* strain, it was similar in the Δ*atpH* and wild-type strains (Fig. 2C). Conversely, AtpH was synthesized at the same level in the wild-type and Δ*atpI* strains. The two subunits are thus synthesized independently, ruling out a CES relationships and showing that MTHI1 controls the expression of two different genes, at variance with most M or T factors studied so far in Chlamydomonas.

The MTHI1 factor targets the *atpI* 5'UTR.

We studied the role of MTHI1 in *atpI* gene expression using chimeric genes. We first constructed a chimeric *atpI* gene, in which the *atpI* 5'UTR was replaced by the promoter and 5'UTR of the *psaA* gene (Fig. 3A). After transformation, this *aAdI* chimera (see footnote [8] of Table 1 for chimeras nomenclature) replaced the endogenous *atpI* gene in the wild-type and *mthi1-1* recipient strains. In a wild-type background, it was expressed at a level sufficient to

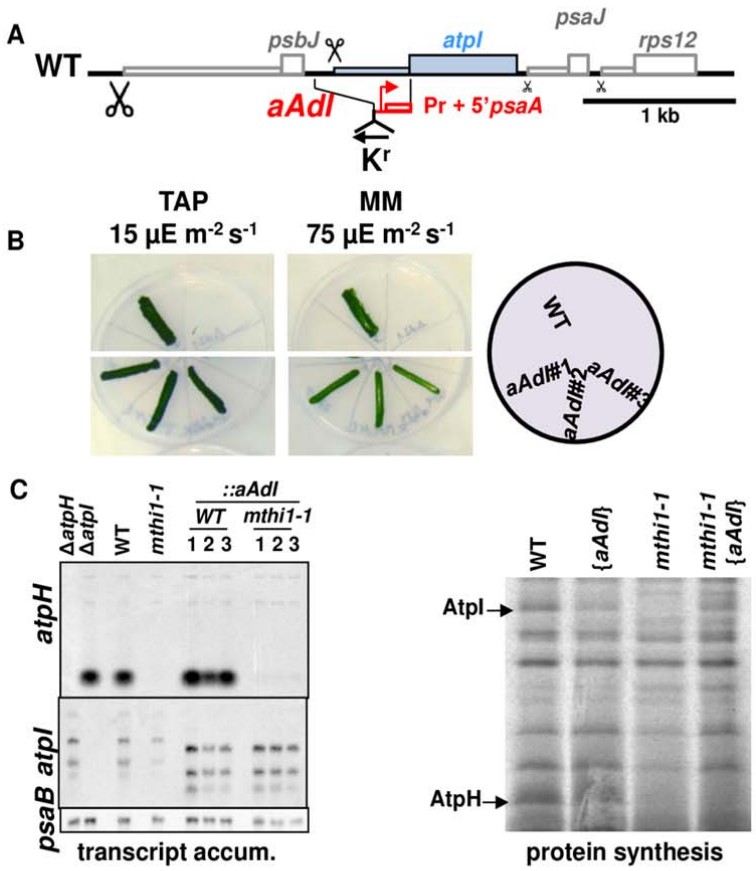


Fig. 3: the MTHI1 factor targets the atpI 5'UTR.

A) Schematic representation of the *aAdI* chimera, where the *atpI* 5'UTR had been replaced by the promoter and 5' untranslated regions of the *psaA* gene. The position of the recycling *aadA* cassette (K^r), inserted in reverse orientation with respect to *atpH*, is shown.

B) Photoautotrophic growth of the *aAdI* strain, assessed as in panel 2B.

C) (Left) *atpH* and *atpI* transcript accumulation in the wild-type and *mthi1-1* strains transformed by the *aAdI* construct, whose transcript is shorter than the endogenous *atpI* transcript, because of the small size of the *psaA* 5'UTR. The recipient strains are shown as well as the Δ*atpH* and Δ*atpI* strains for controls. Three independent transformants are shown for each genetic background. The *psaB* transcript provides a loading control. (Right) rate of AtpH and AtpI synthesis in the wild-type and *mthi1-1* strains and in the corresponding strains transformed by the *aAdI* construct, assessed as in Fig. 1B by pulse labelling experiments.

Ozawa et al, Fig. 3

184 sustain phototrophy (Fig. 3B). When introduced in the *mthi1-1* recipient strain, it did not
185 restore phototrophy in transformants that still lack accumulation of the *atpH* mRNA.
186 However, pulse labelling experiments showed a restored synthesis of the AtpI subunit (Fig.

3C). The down regulation of *atpI* mRNA translation in absence of MTHI1 thus depends on the *atpI* 5'UTR.

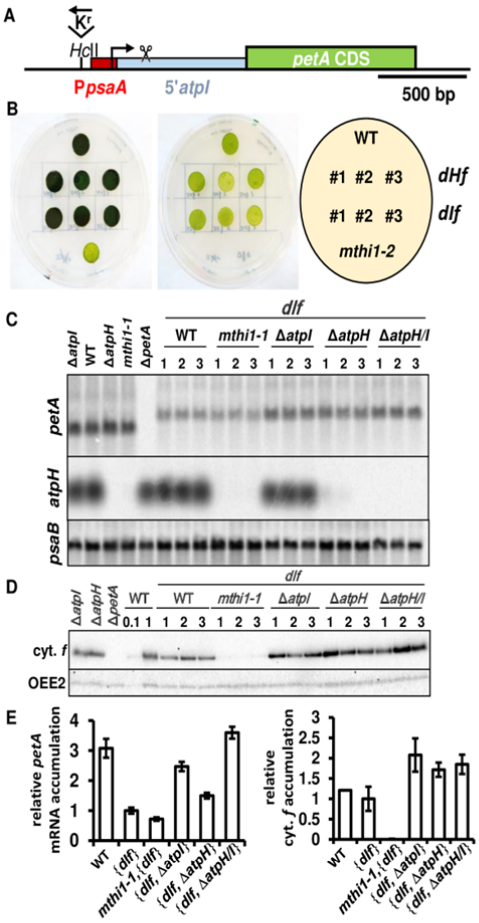
In another chimera, *dIf*, the *atpI* 5'UTR region was fused in frame to the *petA* coding sequence, previously shown to be a convenient reporter gene (Wostrikoff et al., 2004). The *atpI* 5'UTR being uncharacterized so far, we first determined its length (493 nt) by 5'RLM-RACE (Suppl. Fig. S1A). Furthermore, as the *atpI* gene is part of a polycistronic unit, with no indication for a dedicated promoter (Cavaiuolo et al., 2017, Suppl. Fig. S1B), we placed the *psaA* promoter upstream of the *atpI* 5'UTR (Fig. 4A). After transformation, the *dIf* chimera replaced the endogenous *petA* gene in the wild-type, *mthi1-1*, Δ*atpH*, Δ*atpI* and Δ*atpH/I* strains.

Transformants derived from the wild-type strain grew on minimum medium (Fig. 4B): the *atpI* 5'UTR can drive cytochrome *f* synthesis at levels sufficient to sustain phototrophic growth. However, the accumulation of both the chimeric transcript and its cytochrome *f* gene product were lower than those of the endogenous *petA* gene. When introduced in the *mthi1-1* strain, the accumulation of the chimeric transcript was further reduced, but cytochrome *f* only accumulated to trace amount: the *atpI* 5'UTR thus confers an MTHI1-dependant rate of translation to a reporter coding sequence. Similar results were obtained using the heterologous *aadA* coding sequence as a reporter (Suppl. Fig. S2).

Most interestingly, the expression of the *5'atpI-petA* chimera was increased in the deletion strains Δ*atpH*, Δ*atpI*, Δ*atpH/I* (Fig. 4C,D;F). In the Δ*atpH* strain, the accumulation of the chimeric transcript was increased 1.5-fold, compared to the wild-type background, but remained lower than that of the endogenous *petA* gene, while the accumulation of its gene product became higher than that of the endogenous cytochrome *f*. The expression of the *atpH* and *atpI* genes thus relies on a common factor present in limiting amount, possibly MTHI1. The accumulation of the chimeric mRNA was further increased (by 2.5 fold) when the *atpI* gene was deleted as it was in a strain deleted for both *atpH* and *atpI* genes. The chimeric and endogenous *atpI* transcripts compete for the binding of some factors in limiting amount.

The MTHI1 factor targets the *atpH* 5'UTR to stabilise the transcript and activate its translation.

We similarly identified the target of MTHI1 within the *atpH* mRNA. The *dHf* chimeric gene, made of the *atpH* promoter and 5'UTR fused in frame to the *petA* coding sequence (Fig. 5A) was introduced by transformation in the chloroplast genome of the wild-type, *mthi1-1*, Δ*atpH*, Δ*atpI* and Δ*atpH-atpI* recipient strains. In a wild-type background,



Ozawa et al, Fig. 4

Fig. 4: the atpI 5'UTR is sufficient to confer a MTHI1-dependant translation to a reporter gene.

A) Schematic map of the *dIf* construct inserted instead of the endogenous *petA* gene. The red rectangle indicates the *psaA* promoter region placed upstream of the *psbJ-atpI* intergenic fragment (in light blue) that was chosen long enough to include the *atpI* processing site. The scissors above the intergenic region indicate the position of the 5'end of the processed *atpI* mRNA. The position of the recycling selection cassette, upstream of the chimeric *petA* gene and in reverse orientation with respect to this latter, is shown.

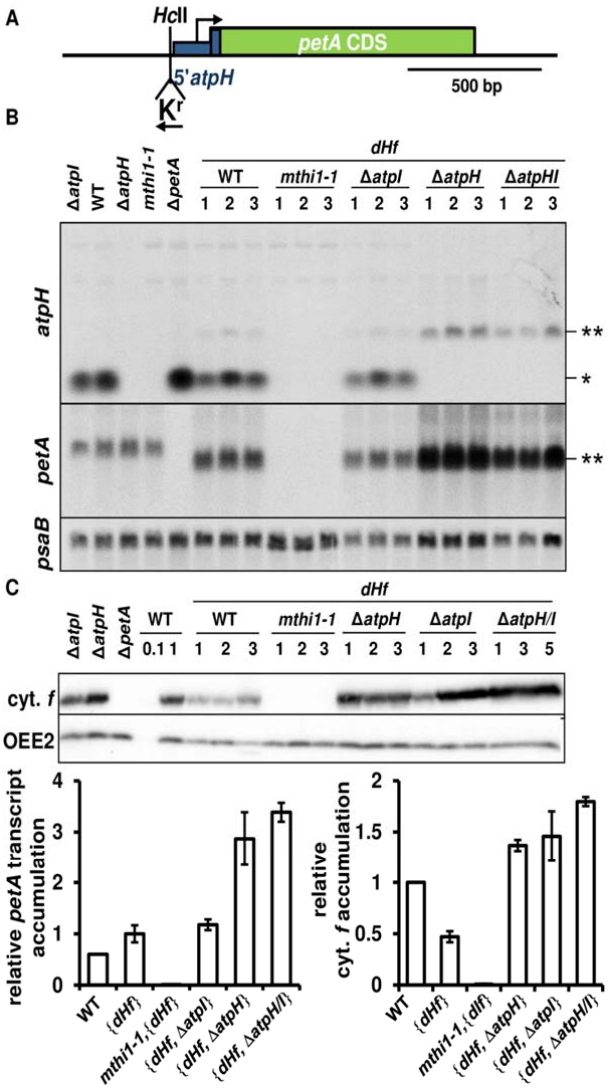
B) Photoautotrophic growth of the *dHf* (see Fig. 5) and *dIf* chimeric strains (three independent transformants) assessed as in fig. 2B. The growth of the wild type and of the *mthi1-2* strains are shown as controls.

C) Accumulation of the *petA* transcript, either endogenous or chimeric, in the chloroplast genome of the wild-type, *mthi1-1*, Δ*atpH*, Δ*atpI*, and Δ*atpH/I* recipient strains, shown aside, as well as a Δ*petA* strain for comparison. Three independent transformants are shown for each genetic context. The accumulation of the *atpH* mRNA in the same strains is also shown, while that of the *psaB* mRNA provides a loading control.

D) Accumulation of cytochrome *f* in the same strains (loading control: OEE2).

E) Quantification of the *petA* transcript (left) and cytochrome *f* (right) in transformed strains shows a competition between the chimera and the endogenous *atpI* gene for the expression of 5'*atpI*-driven genes. Value for the *dIf* transcript in the wild type recipient strain is set to 1; n=6.

221 transformants were phototrophic: the *atpH* 5'UTR allows the expression of cytochrome *f* (Fig.



Ozawa et al, Fig. 5

Fig. 5: the MTHI1 factor targets the atpH 5'UTR.

A) Schematic representation of the *dHf* chimera, with the position of the recycling *aadA* cassette (in reverse orientation with respect to the *petA* gene) shown. The blue thick rectangle represents the first 25 nt of the *atpH* coding sequence fused in frame with the *petA* coding sequence, added to the construct to improve the expression of the chimera.

B) Accumulation of the *atpH* and *petA* transcripts in the wild type, *mthi1-1*, Δ*atpH,* Δ*atpI* and Δ*atpH/atpI* strains carrying the *dHf* chimera instead of the endogenous *petA* gene. Unstransformed wild-type, Δ*atpH*, Δ*atpI*, Δ*petA* and *mthi1-1* strains are shown as controls. Asterisk indicates the position of the *atpH* mRNA, while the double asterisk points to a cross-reaction of the probe that comprises the *atpH* 5'UTR with the *dHf* chimeric transcript. Three independent transformants are shown for each genetic context. The *psaB* transcript provides a loading control.

C) Cyt. *f* accumulation in the same strains, with OEE2 as a loading control.

D) Quantification of the relative accumulation of the *petA* transcript (Left) and of cyt. *f* (Right) in the same strains. Values for *dHf* transformed in the wild-type strain are set to 1; n=6

222 4B). The *dHf* chimeric transcript accumulated to 150 % of the endogenous *petA* transcript

level, but its protein product was twice less abundant than the endogenous cytochrome *f* (Fig. 5B,C,D). In the *mthi1-1* background, the chimeric *petA* mRNA did not accumulate and cytochrome *f* was totally absent (Fig. 5B,C).

Deletion of the *atpH* gene increased the expression of the chimera at the transcript and cytochrome *f* levels by 3 and 1.5 fold respectively, compared to the wild-type background (Fig. 5B,C,D). The chimeric transcript competes with the endogenous *atpH* mRNA for MTHI1 binding. Deletion of the *atpI* gene increased only moderately the accumulation of the chimeric transcript but stimulated its translation, while the simultaneous deletion of the two genes increased the accumulation of both the chimeric transcript and its cytochrome *f* gene product. Again, these observations were confirmed using the *aadA* coding sequence as a heterologous reporter gene (Suppl. Fig. S3). We noted that a dicistronic *petA-aadA* transcript accumulated to the same level in the four progeny but was not expressed in the *mthi1* offspring (Suppl. Fig. S3 B,C), suggesting that MTHI1 could also be required for the translation of the *atpH* mRNA.

To address this point, we constructed a modified *atpH* gene, whose transcript is stabilised independently of the presence of MTHI1 thanks to the insertion, immediately after the *atpH* transcription start site, of a polyG cage, a very stable secondary structure impeding the progression of 5' → 3' exoribonucleases (Vreken and Raue, 1992; Drager et al., 1996; Drager et al., 1998). This modified pG*atpH* gene (Fig. 6A) replaced the endogenous *atpH* gene in wild-type and *mthi1-2* strains and we monitored its expression in transformants. Those recovered from the wild-type strain were phototrophic (Fig. 6B) and accumulated similar amounts of the *atpH* transcript (Fig. 6C) and of AtpH (Fig. 6D) as the control strain: the polyG cage at the beginning of the *atpH* transcript did not prevent its translation. Transformants derived from the *mthi1-2* strain recovered, although reduced, accumulation of the *atpH* mRNA but were nevertheless unable of phototrophic growth. They lacked accumulation of the AtpH subunit (Fig. 6D), probably because the synthesis of the AtpI subunit was still impaired in the *mthi1* background. To overcome this issue, we replaced the *atpI* gene of the *mthi1-2* {*pGatpH*} strain by its chimeric *aAdI* version, whose expression does not depend on the presence of MTHI1 (Fig. 2B,C). Despite the restored expression of AtpI, the *mthi1-2* {*aAdI, pGatpH*} transformants were still unable of photosynthetic growth and lacked accumulation of the AtpH subunit (Fig. 6B,D). Thus, MTHI1 beside stabilising the *atpH* mRNA, is also required for its translational activation.

Characterisation of the MTHI1 protein.

Ozawa et al, Fig. 6

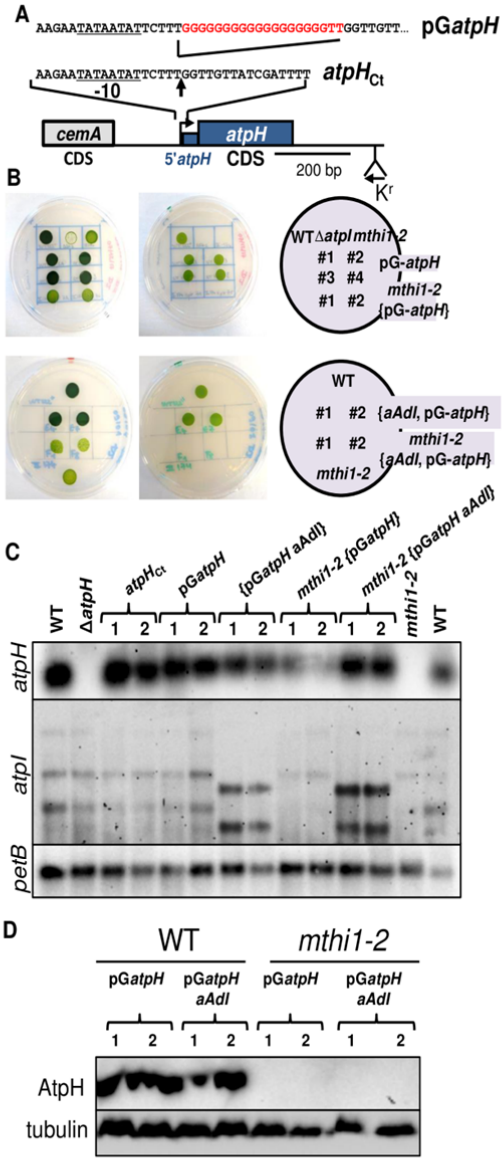


Fig. 6: MTHI1 is required for the translation of the atpH gene

A) Schematic map of the pG*atpH* construct with a zoom to the region surrounding the *atpH* transcription start site, indicated by a vertical arrow, where the polyG tract was inserted. The *atpH* promoter is underlined, and the position of the recycling *aadA* cassette is shown. A construct carrying the selection cassette at the same position but devoid of the polyG insertion was used as control ($atpH_{Ct}$). To avoid any polar effect on the expression of the downstream located *atpF* gene (co-transcribed with *atpH*), all experiments were performed after excision of the recycling *aadA* cassette.

B) Phototrophic growth of the pG*atpH*, *mthi1-2* {pG*atpH*}, {*aAdI* pG*atpH*} and *mthi1-2* {*aAdI* pG*atpH*} strains (two independent transformants each) assessed as in Fig. 2B. Growth of the wild type and of the *mthi1-2* and Δ*atpI* strains are shown as controls.

C) Accumulation of the *atpH* and *atpI* transcripts in the wild-type strain transformed by the $atpH_{Ct}$ and pG*atpH* constructs and in the *mthi1-2* strain transformed with the pG*atpH* gene. The *aAdI* construct then replaced the endogenous *atpI* gene in the resulting pG*atpH* and *mthi1-2* {pG*atpH*} strains. Two independent transformants are shown for each genetic background. The *petB* transcript provides a loading control.

D) Accumulation of the AtpH subunit in strains expressing the polyG construct. Tubulin provides a loading control.

256 We cloned the *MTHI1* gene, by complementation of an *mthi1-1, arg7*, *cw15* strain with

an indexed library of cosmids (see Suppl. Fig. S4 for details). Evidence that the *MTHI1* gene actually corresponds to gene model Cre17.g734564 came from the complementation of both *mthi1-1* and *mthi1-2* mutations by an EST clone (AV629671) obtained from Kazusa DNA Research Institute. Sequencing of the *MTHI1* region revealed that the translation initiation codon was substituted by an AUU codon in strain *mthi1-1*, while insertion of a C residue after codon 138 yielded to premature translation abortion after codon 188 in strain *mthi1-2* (Suppl. Fig. S5B).

The *MTHI1* gene encodes a protein of 828 residues (Fig. 7A, suppl. Fig. S5C), predicted to be targeted to the chloroplast by the Predalgo and TargetP programs (Tardif et al., 2012; Almagro Armenteros et al., 2019). Prediction of secondary structure by the Scratch protein predictor software (http://scratch.proteomics.ics.uci.edu/) suggested that the mature MTHI1 protein potentially comprises two different domains. Following a predicted chloroplast targeting peptide of 48 residues, the N-terminal domain (up to residue 566) contains pairs of α-helices (Fig. 7A; Suppl. Fig. S5C), 9 of which are typical OPR repeats (Fig. 7B). The C-terminal domain harbours mainly coiled-coil or intrinsically disordered sequences with no obvious motifs, but several stretches of A and Q residues (Suppl. Fig. S5C), as in other Chlamydomonas M and T factors (Boudreau et al., 2000; Auchincloss et al., 2002; Raynaud et al., 2007).

BLAST searches found orthologues of MTHI1 in green algae (Suppl. Fig. S6). The region of similarity was restricted to the N-terminal, OPR-containing part of the protein, while the C-terminal tail was highly variable in length and sequence, even between the most closely related species. Thus, fusing a HA tag for immuno-detection at the C-terminus of MTHI1 should not be deleterious for its function. Indeed, we could still complement the *mthi1-1* mutation with a tagged version of *MTHI1*, was the tag inserted in genomic (g transformants) or cDNA (c transformants) constructs (Fig. 8A). The tagged genomic construct, including 4280 bp upstream of the translation initiation codon, i.e. presumably the whole *MTHI1* promoter, allowed a higher accumulation of the tagged protein than the tagged cDNA construct (compare clones g6 and g9 with clones c in Fig. 8B).

We over-expressed the MTHI1 protein and raised an antibody against the mature protein to compare the accumulation of the tagged protein in transformants with that of the endogenous protein in the wild type. Despite the higher accumulation of MTHI1 in clones transformed with the genomic construct, the *atpH* mRNA was not more accumulated than in wild type (Fig. 8B,E). Either the C-terminal tag somehow decreases protein activity, or other factors limit the abundance of the *atpH* mRNA. As expected from the requirement of MTHI1

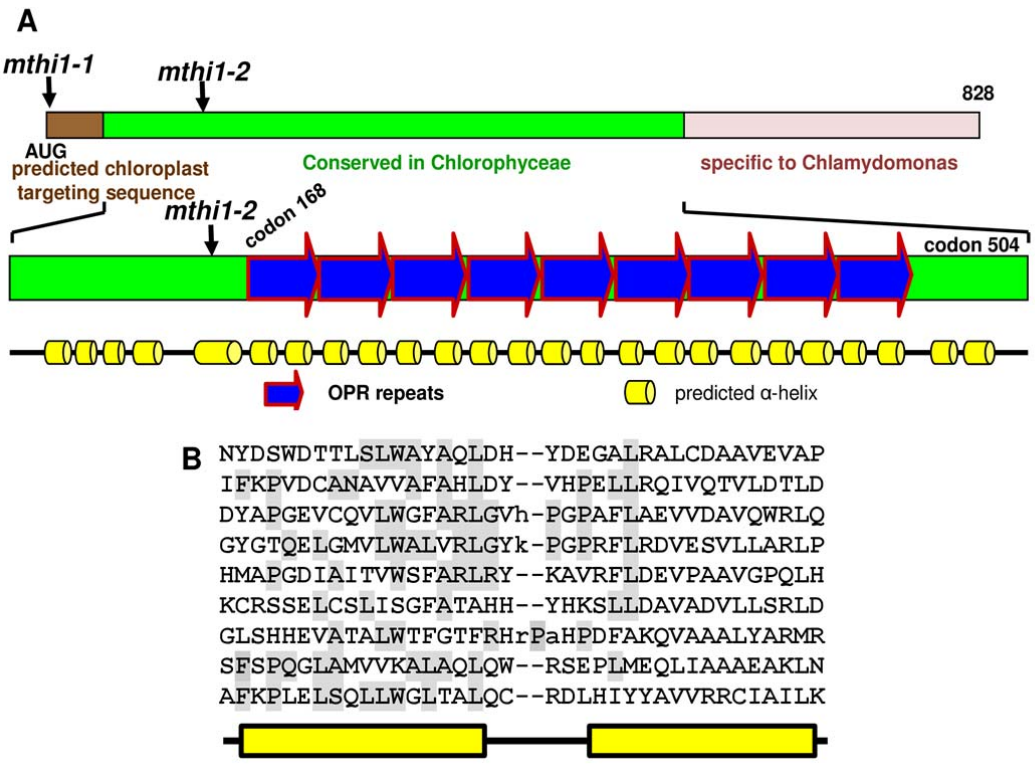


Fig. 7: the MTHI1 protein.

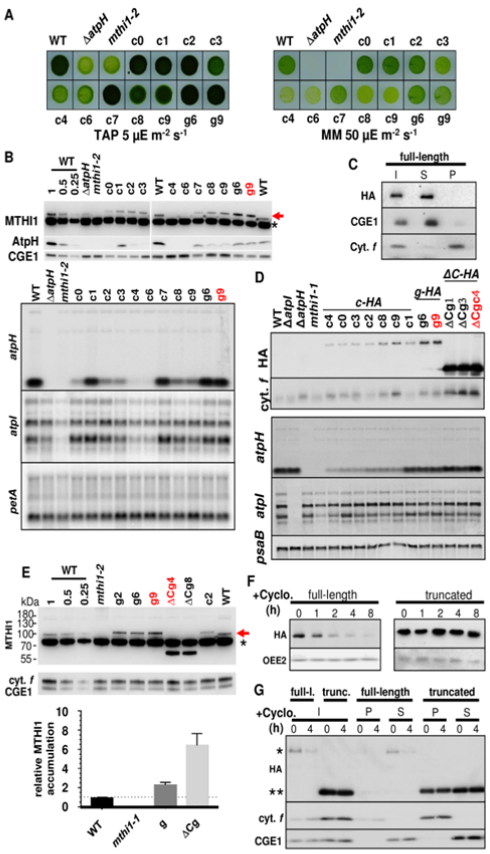
A) Schematic representation of the MTHI1 protein. The brown rectangle depicts the chloroplast transit peptide as predicted by the ChloroP program. The green rectangle indicates the region of the protein conserved in other Chlorophyceae species (see Supp. Fig S9), while the pink rectangle points to a rapidly evolving and disordered region. The position of the two *mthi1* mutations is shown. Blue arrows represent the OPR repeats, whose sequence is shown in panel (B), with the amino acid residues obeying the OPR consensus shaded in grey. The lower scheme shows the predicted secondary structure of this region.

Ozawa et al, Fig. 7

291 for the accumulation of the *atpH* mRNA, the levels of MTHI1 and of the *atpH* mRNA were

292 correlated (Fig. 8D,B): the clones (e.g. c4 or c6) accumulating less MTHI1, - undetectable

293 with the antibody against the whole protein (Fig. 8B), but detectable with that against the HA



Ozawa et al, Fig. 8

Fig.8: complementation of the mthi1-1 mutant strain.

A) complementation of the *mthi1* strain with a tagged version of the *MTHI1* gene, either the tagged cDNA (c clones) or the genomic construct (g clones) restores phototrophy, assessed by plating the cells on Minimum medium plates as in Fig. 2B. The growth of the wild-type and of the Δ*atpH* and *mthi1-2* strains is shown as control.

B) Accumulation of the MTHI1 protein (red arrow), either endogenous or tagged, of the AtpH subunit (top) and of the *atpH* and *atpI* transcripts (bottom) detected in the same strains with an antibody against the MTHI21 protein. Note the larger size of the tagged protein, compared to the endogenous one, due to the insertion of the triple HA tag. CGE1and cyt. *f* or *petA* mRNA are shown as the respective loading controls in protein and RNA blots. The name of the clone used for further analysis of MTHI1 in the next figures is written in red. (asterisk: cross-contaminant).

C) MTHI1 is a soluble protein.

Cellular extract (I) from the complemented strain g9 was separated into soluble (S) and insoluble (pellet: P) fractions by ultracentrifugation and equal volumes of each fraction were probed with antibodies against the HA tag and against GrpE and cytochrome *f* as controls for the purity of the fractions.

D) The C-terminal domain of MTHI1 is dispensable for its function

Accumulation of the tagged MTHI1 protein, probed with an antibody against the HA tag, and of the *atpH* and *atpI* transcripts in *mthi1-1* strain complemented with the tagged versions of the *MTHI1* gene, either the tagged cDNA (*c* clones), the genomic construct (g clones) or its C-terminally truncated version (ΔCg clones). All transformants were selected for recovery of phototrophy on Minimum medium plates. Overaccumulation of the truncated MTHI1 protein does not lead to an increased abundance of the *atpH* transcripts. Cyt. *f* and *psaB* are shown as the respective loading controls in protein and RNA blots. The name of the clones used for further analysis of MTHI1 in the next figures is written in red.

E) Deletion of the C-terminal domain results in higher abundance of MTHI1.

(top) Accumulation of MTHI1 protein (red arrow) in *mthi1-1* strains complemented with either the tagged *MTHI1* cDNA (*c*), the tagged genomic construct (*g*) or its C-terminally truncated version (ΔCg), probed with an antibody against the MTHI1 protein. The name of the clones used for further analysis of MTHI1 is written in red. (asterisk: cross-contaminant).

(bottom) quantification of MTHI1 accumulation, in the strains shown in top panel, normalised to that of cytochrome *f* and reported to the accumulation of MTHI1 in wild-type cells, set to 1 (symbolised by a dashed line). Error bars represent SD, n = 3.

F) The C-terminal domain of MTHI1 contributes to its high turn-over.

Stability of full-length MTHI1 or of its C-terminally truncated version, assessed by immunoblots of cells treated with cycloheximide for the indicated times. Accumulation of OEE2 in the same samples is shown as a loading control.

G) Differential solubility of the full-length MTHI1 protein and of its C-ter truncated version.

Cellular extracts of transformants expressing the full-length (*) and the truncated (**) versions of the tagged MTHI1 protein, treated with cycloheximide for 0 or 4 hours (Input panel, left), were fractionated into soluble (S) and membrane (P) fractions and analysed as in panel F. Distribution of CGE1 and cyt. *f* are shown to assess the purity of the fractions.

294 tag (c4 in Fig. 8E) - also accumulated *atpH* mRNA levels just above the detection threshold.

295 They were however able of phototrophic growth (Fig. 8A), confirming some restoration of
296 ATP synthase.
297 We used one clone complemented with the tagged genomic construct (g9) to study
298 MTHI1 intra-organelle localisation and found it exclusively soluble (Fig. 8C).
299
300 **The C-terminal tail is dispensable for the main function of the MTHI1 factor.**
301 The poor conservation of the C-terminal domain raised the question of its function. We
302 thus constructed a truncated version of the gene, lacking residues 573-797, i.e. most C-
303 terminal domain, but still containing the HA tag. This truncated MTHI1 could still
304 complement the *mthi1-1* mutation. As shown on Fig. 8D,E, it accumulated to much higher
305 levels than the full-length protein, but did not proportionally increase the abundance of the
306 *atpH* or *atpI* mRNAs: either part of the truncated MTHI1 protein was not involved in mRNA
307 stabilisation or other factors became limiting. To understand the origin of this differential
308 accumulation, we compared the stability of the full-length and truncated MTHI1 by following
309 their decay in complemented strains incubated with cycloheximide, an inhibitor of cytosolic
310 translation (Fig. 8F). MTHI1 was short-lived, with a half-life of about 1 hour. Most
311 interestingly, its truncated version remained stable over the 8 hr of the experiment: the C-
312 terminus tail apparently controls the half-life of the whole protein. We repeated fractionation
313 experiments on these complemented strains, treated with cycloheximide for 4 hr. In total cell
314 extracts, the level of MTHI1-HA was strongly reduced upon cycloheximide treatment, while
315 that of its truncated version was insensitive to the antibiotic. After fractionation into soluble
316 and insoluble fractions, the full-length MTHI1 was almost exclusively found in the soluble
317 fraction. By contrast, its truncated version was significantly found in the pellet, most probably
318 as large aggregates that fell down during ultracentrifugation. Both the aggregated and soluble
319 populations were stable over 4 hr (Fig. 8G).
320
321 **MTHI1 belongs *in vivo* to a large complex that also contains the *atpH* and/or *atpI***
322 **mRNA**
323 We used size exclusion chromatography to investigate whether MTHI1 belongs to a
324 high molecular mass complex, as do almost all *trans*-acting factors studied so far (Boudreau
325 et al., 2000; Vaistij et al., 2000b; Auchincloss et al., 2002; Dauvillee et al., 2003; Perron et al.,
326 2004; Schwarz et al., 2007; Johnson et al., 2010; Boulouis et al., 2011). Soluble extract from
327 clone g9 was fractionated on a Superose 6 column, optimal for separating protein complexes
328 in the 5 to 5,000 kD range. As shown in Fig. 9A, MTHI1 belongs to complexes ranging from

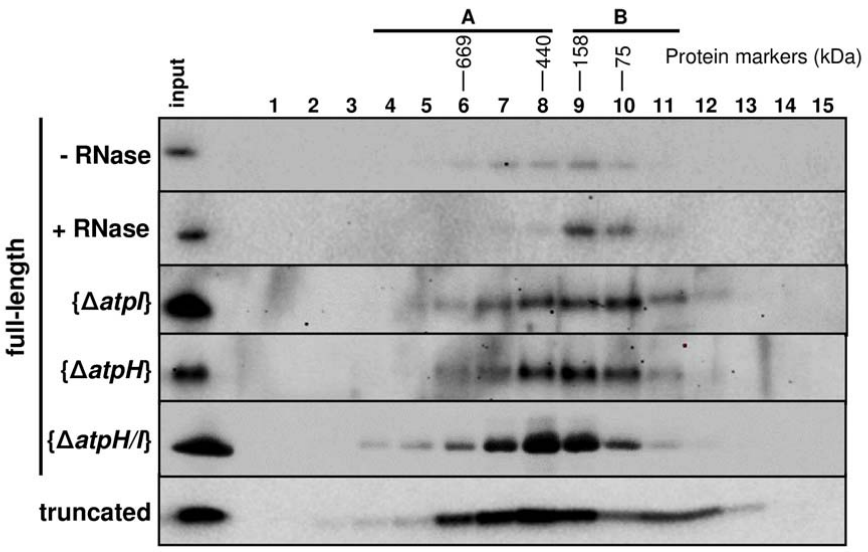


Fig. 9: MTHI1 belong to a high molecular weight complex that interacts with the atpH and atpI transcripts.

Soluble extracts from strains listed at the left of the figure were fractionated on a Superose 6 10/300 HR column and probed with an antibody against the HA tag. Molecular masses of the complexes found in each fraction were estimated by comparison with standards of the HMW gel filtration calibration kit (GE Healthcare).

Ozawa et al, Fig. 9

75 (fraction 10) to >700 kD (fraction 5), peaking in fractions 8-9 (150 to 450 kD). As no special care was taken to preserve the integrity of the RNA moiety, RNAs, if retained by MTHI1, were probably restricted to fragments closely surrounding its binding site and only account for a minor increase in molecular mass. When the supernatant was treated with

RNase, prior to loading on the column, MTHI1 presented a sharper distribution in slightly lighter fractions 9 and 10, consistent with a monomeric state. Thus, mRNAs appear responsible for the distribution of MTHI1 in high molecular mass ribonucleoprotein complexes. We analysed the distribution of MTHI1 in complemented strains lacking the *atpH* mRNA, the *atpI* mRNA or both. Upon deletion of either *atpH* or *atpI* the distribution of MTHI1 remained unchanged and centred on fractions 8-10, the deletion of both *atpH* and *atpI* genes shifted the distribution of MTHI1 to larger complexes, centred on fraction 8, but extending to still heavier fractions. In the absence of its two RNA targets, MTHI1 undergoes conformational changes, possibly making aggregates. A similar behaviour had been already reported for MCA1 and TCA1 in the absence of their *petA* mRNA target (Boulouis et al., 2011). This behaviour, however, is opposite to that observed upon RNase treatment. Aggregation of MTHI1 in the absence of its RNA target was corroborated by the distribution pattern of the truncated MTHI1. Partially found in the pellet after ultracentrifugation, it presented a bimodal distribution with a first peak in fraction 11-12, likely corresponding to monomers and a broad peak in fractions 8 to 2, with a maximum in fraction 8.

MTHI1 interacts with the *atpH* mRNA.

To investigate the interaction of MTHI1 with the *atpH* and *atpI* transcripts, we sequenced the small RNA fraction (size range: 11-44 nt, sRNA-Seq) of RNA samples prepared from the wild type or *mthi1-1* strains, since the interaction of an M factor on its target transcript tends to generate of a footprint that pinpoints its binding site (Cavaiuolo et al., 2017; Ruwe and Schmitz-Linneweber, 2012; Zhelyazkova et al., 2012). Fig. 10A shows the normalized coverage of RNAs along the *atpH* and *atpI* loci. Fig. 10B shows the coverage of sRNAs (11-44 nt) over the same loci. A prominent peak of small RNAs of about 19-21 nt in length with a sharp 5'end was found in the wild type at the very beginning of the *atpH* mRNA. This peak was only seen after treatment of the RNA samples with RNA PolyPhosphatase (RPP), an enzyme that removes the pyrophosphate moiety of tri-phosphorylated transcription products. Most, if not all, monocistronic *atpH* mRNA is thus transcribed from the *atpH* promoter and does not result from the processing of the large precursor transcribed from the *atpA* promoter. In the *mthi1-1* mutant, the amplitude of that peak was reduced by more than 5-fold (Fig. 10B).

We used the g9 strain complemented with the tagged version of MTHI1 to immuno-precipitate it with an antibody against the HA tag (Fig. 11A). RNAs, extracted from the pulled-down material, were analysed by dot-blots (Fig.11B). We observed a signal with a

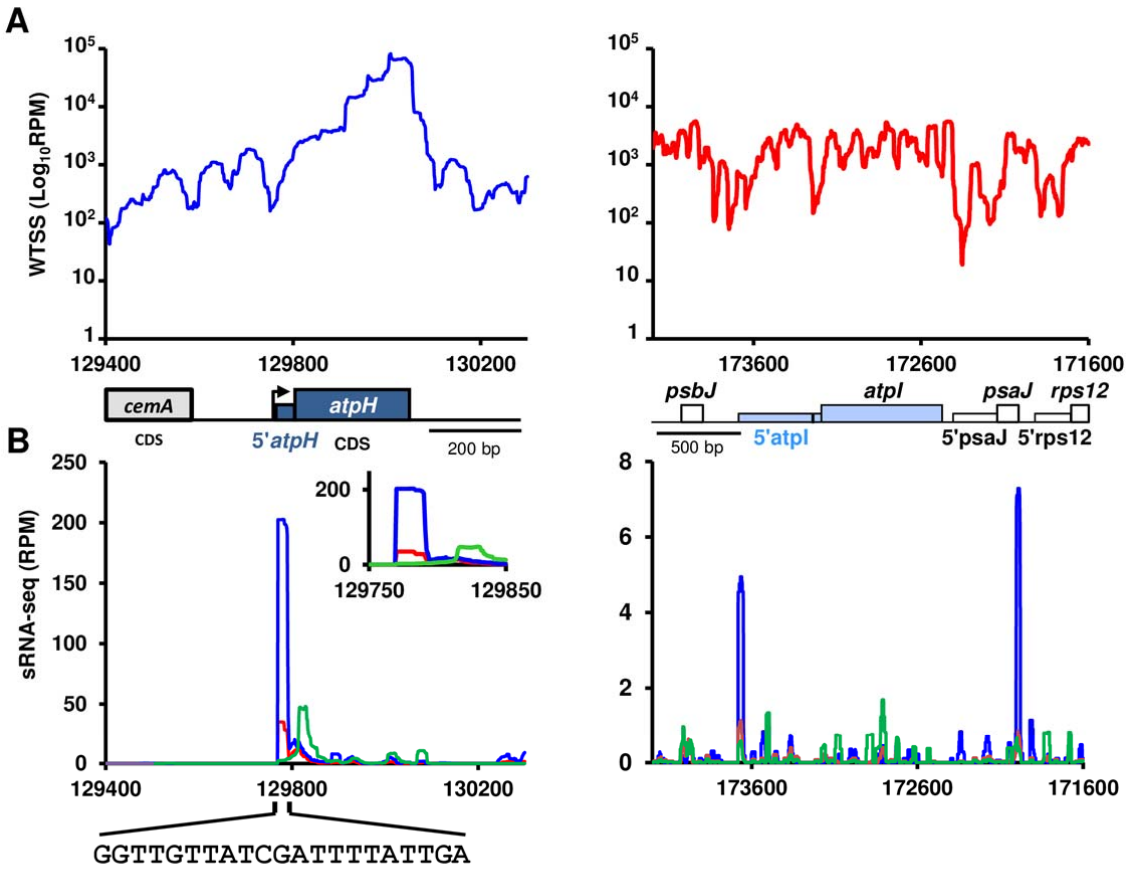


Fig. 10: Transcriptional profile of the atpH and atpI genes.

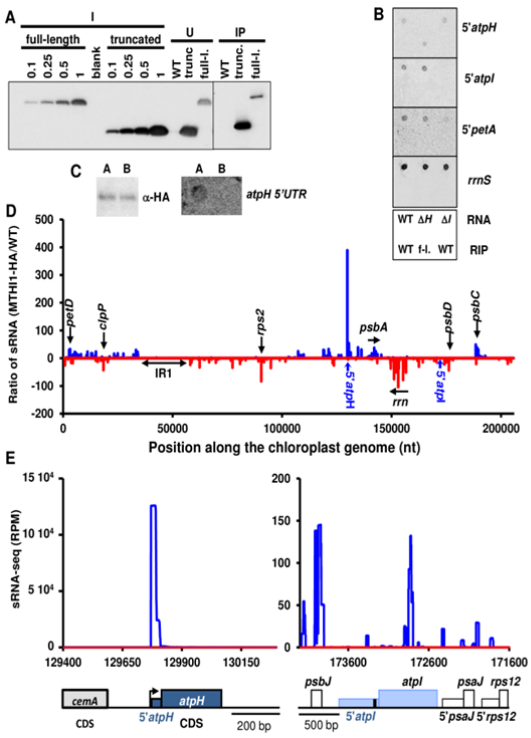
A) Coverage, normalised as RPM (log scale) of pooled bi-directional and directional wild type WTSS along the *atpH* and *atpI* loci. Positions of relevant genes and 5'UTRs are shown below. The black bar in the *atpI* 5'UTR shows the position of the MTHI1 target (see below). Redrawn from the data in Cavaiuolo et al, 2017.

B) sRNA mapping at the 5'end of the *atpH* mRNA are the footprint of MTHI1.

Coverage, normalised as RPM, of pooled sRNA-Seq along the same loci: mock- (green) versus RPP-treated (blue) wild type sRNA-Seq libraries compared to RPP-treated libraries of the *mthi1-1* mutant (red). Coverage is averaged over two biological replicates. Only reads mapping to the coding strand are shown. The inset for *atpH* shows a zoom to the *atpH* 5'UTR and the sequence of the footprint is shown at the bottom. For *atpI* a zoom to the 5'UTR region (coding strand only) is shown in Suppl. Fig. S7A. Note the very different values on y-axes of the two graphs.

Ozawa et al, Fig. 10

probe specific for *atpH* 5'UTR, but not with *rrnS* or 5'*petA* probes used as specificity controls, nor when the same procedure was applied to the the wild type, devoid of HA-tagged MTHI1. Thus MTHI1 interacts specifically, directly or indirectly, with the *atpH* 5'UTR. By contrast, no signal could be detected with an 5'*atpI* probe.



Ozawa et al, Fig. 11

Fig. 11: The MTHI1 protein interacts specifically with the atpH 5'UTR.

A) The full-length and truncated MTHI1 proteins were immuno-precipitated from a soluble cellular extract with an antibody against the HA tag (I: input; U: unbound; IP: immuno-precipitate). Immunoprecipitation of a cellular extract from the wild-type strain is shown as a negative control. The apparent slower migration of the immuno-precitated proteins in due to a "smiling" effect in the migration of the gel from which the composite figure (indicated by a vertical line) was made, in order to remove irrelevant samples.

B) RNAs extracted from immuno-precipitates were analysed by dot-blot hybridised to the probes indicated on the rigth. The lower panel shows the disposition of the samples on the filter. Top line: RNA extracted from the wild type, Δ*atpH*, and Δ*atpI* strains (without immuno-precipitation), as a control for the specificity of the probes. Bottom line: immuno-precipitated RNA from the wild type, and from complemented strains expressing the tagged MTHI1 (g9).

C) Fractions from size exclusion chromatography of a cellular extract from a strain expressing the full-length MTHI1 (first line in fig. 9), indicated by the bars A (fractions 4 to 8) and B (9 to 11), were pooled, immuno-precipitated with an antibody against the HA tag and analysed with the same antibody for the MTHI1 content of the immuno-precipitated fractions. Their RNA content was extracted and analysed by dot blot with a probe specific of the *atpH* 5'UTR, which detected a (weak) signal in pooled fractions A, further analysed by deep sRNA-Seq.

D) Ratio of normalised sRNA coverage in MTHI1 RIP samples along the chloroplast genome.

Differential enrichment was calculated as the ratio of the coverage of sRNAs at each nucleotide position in the above-defined fraction A of the MDFH1-HA sample to (that in the wild-type control sample +1). Blue curve: sRNAs mapping to the + strand, red curve sRNA mapping to the - strand. Most enriched genome positions are shown on the graph, as well as the positon of the *atpH* and *atpI* 5'UTRs. A zoom to the inverted repeat is shown in Suppl. Fig S7C.

E) Coverage of immuno-precipitated RNA (normalized as RPM) over the *atpH* and *atpI* loci, schematically depicted at the bottom of the panel. Blue curve: MTHI1-RIP sample; red curve: WT-RIP sample (negative control). A zoom to the *atpI* 5'UTR is shown in supplemental Fig. S7B. Note the very different values of the y axes in the two graphs.

Ozawa et al, Fig. 11, continued

371 We sequenced sRNAs from the immuno-precipitated samples. To increase specificity,

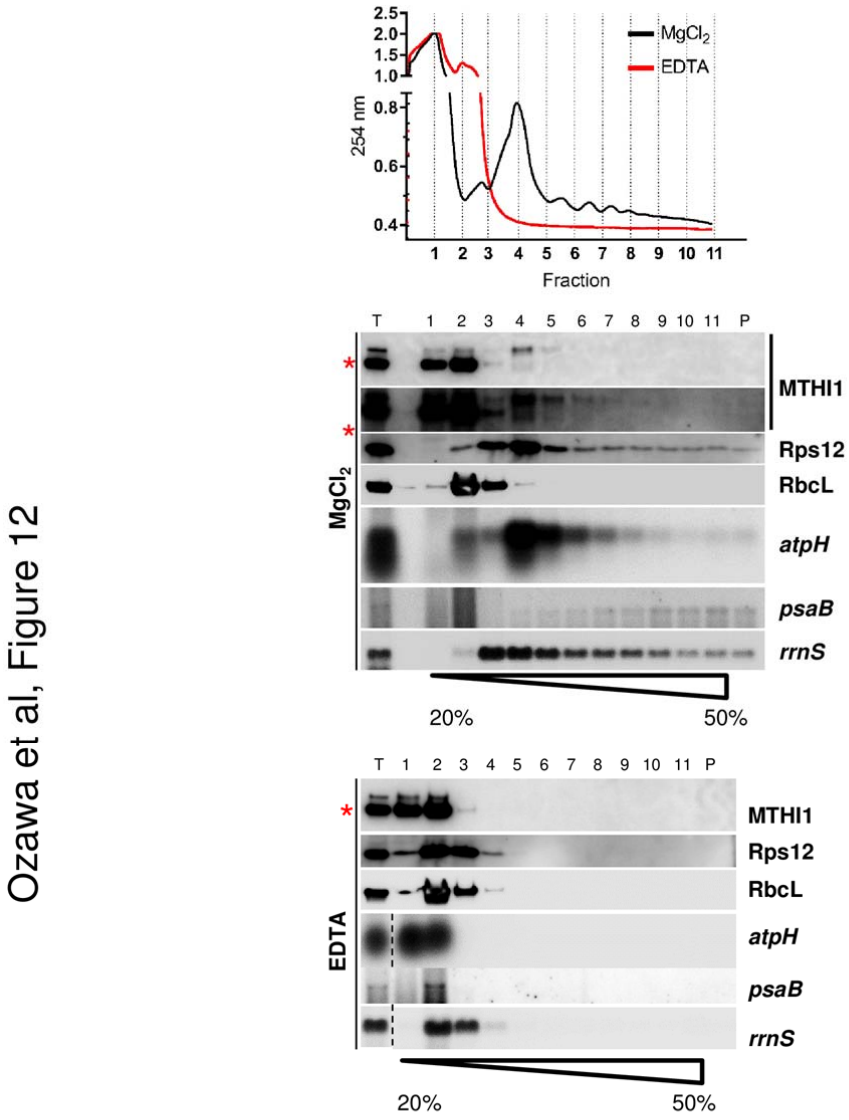


Figure 12: MTHI1 interacts with polysomes.

Distribution of MTHI1, Rps12 and RbcL proteins and of *atpH*, *psaB* and *rrnS* (16S rRNA) transcripts in wild-type cells along a sucrose gradient. For the gradient in the presence of MgCl2, an overexposed blot immuno-decorated with the antibody against the MTHI1 protein is shown. T represents the total protein and RNA extracts, P the pellet fraction. The top panel shows the UV absorbance profile along the gradient, and the lower panel, the distribution of the same proteins and transcripts in samples treated with EDTA to dissociate the ribosomes. Note that the *atpH* mRNA, encoding a short polypeptide, is not heavily loaded with ribosomes and does not penetrate deep in the gradient. (red asterisk: cross-contamination).

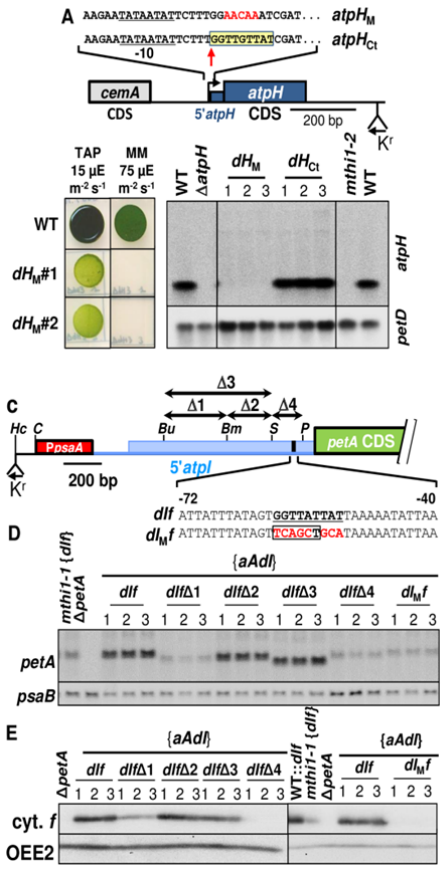
372 MTHI1 complexes were first purified by size exclusion chromatography. MTHI1-HA was
373 then immuno-precipitated independently from pooled fractions 3-8 and 9-10 (Fig. 9). Only
374 RNAs extracted from fractions 3-8 gave rise to a 5'*atpH* signal in dot-blot experiment (Fig.
375 11C) and were used for library construction. The *atpH* mRNA being tri-phosphorylated

(Cavaiuolo et al., 2017; Fig. 10B), all samples were RPP-treated before construction of the library. Fig. 11D displays the ratio of normalised sRNA coverage (expressed as RPM) in the strain complemented with the tagged MTHI1 versus that in the wild type, plotted along the chloroplast genome. In the MTHI1-HA sample, sRNAs were strongly enriched at the 5'end of the *atpH* 5'UTR, as better shown in Fig. 11E. By contrast, sRNAs were not enriched around the *atpI* 5'UTR when compared to the wild type sample (Fig. 11D,E), in agreement with the absence of *atpI* signal in RIP experiments (Fig. 11B). However, at variance with the lack of signal in dot-blots hybridized to a *rrnS* probe, sRNA mapping to the ribosomal operon were somehow enriched in the MTHI1-HA RIP library (Suppl. Fig. S7C). To solve this discrepancy, we looked to the possible association of MTHI1 to ribosomes along a sucrose gradient (Fig. 12). Being a short CDS, *atpH* only accommodates a limited number of ribosomes and does not migrate deep in the gradient. The distribution of MTHI1 parraleled that of the *atpH* mRNA: both were found in polysomal fractions 4 to 8, as shown by the UV absorbance profile and by their dissociation upon EDTA treatment. Thus MTHI1 remains bound to the *atpH* transcript when engaged in translation.

Identification of the targets of MTHI1.

To gain more information on the target of the 9 OPR repeats-containing MTHI1 protein, we looked at the conservation of the small (40 bp) *atpH* 5'UTR, well conserved among green algae: the 9 first nucleotides (GGTTGTTAT) of the *atpH* transcript were strongly conserved in almost all Chlorophyceae, in Pedinophyceae and in the Ulvale clade of Ulvophyceae (Suppl. Fig. S8A and Dataset DS1), but not in Trebouxiophyceae nor in Prasinophytes. To test whether this sequence, often localised downstream of a putative Pribnow -10 box, corresponds to the target of MTHI1, we mutated it into the poorly related GG**AACAA**AT sequence (Fig. 13A). After introduction of this mutated gene into the chloroplast genome, transformants lost phototrophy and failed to accumulate the *atpH* transcript (Fig 13B), suggesting that MTHI1 could not bind and protect the transcript anymore.

MTHI1 also targeting the *atpI* transcript, we searched for occurrence of this motif in the *atpI* 5'UTR. At variance with the *atpH* 5'UTR, the long *atpI* 5'UTR (493 bp in *C. reinhardtii*) is not conserved in Chlorophyta, except, in some Chlorophyceae, Pedinophyceae and in the Ulvale clade, for a stretch of ~60 nt upstream of the initiation codon (Suppl. Fig. S8B and Dataset DS1). Strikingly, this conserved stretch starts by a GGTT(A/G)TTAT motif. We tested its significance by introducing deletions or mutations in the *atpI* 5'UTR (Fig. 13C). To facilitate the characterisation of the resulting mutants, mutations were introduced in the



Ozawa et al, Figure 13

Fig. 13: Validation of the putative MTHI1 targets.

A) Schematic map of the dH_M construct with a zoom to the region of the MTHI1 binding site, highlighted in a yellow box. Mutated nucleotides are written in red. The *atpH* transcription start site is indicated by a vertical arrow. The *atpH* promoter is underlined, and the position of the recycling *aadA* cassette is shown. The control construct ($atpH_{Ct}$) carries the selection cassette but no mutation in the *atpH* gene.

B) (Left) Phototrophic growth of the dH_M strain (two independent transformants), assessed as in Fig. 2B. Growth of the wild type is shown as a control. (Right) Accumulation of the *atpH* transcript in the wild-type transformed by the dH_{Ct} and dH_M constructs. Three independent transformants are shown for each genetic background. The *petD* transcript provides a loading control.

C) Schematic representation of the 5'*atpI* 5'UTR region in the mutant *dIf* series.

The red rectangle represents the *psaA* promoter region and the blue line shows the *psbJ-atpI* intergenic fragment inserted in the construct (larger than the *atpI* 5'UTR, to allow the processing of the chimeric transcript). The blue rectangle symbolises the processed *atpI* 5'UTR, with the target of MTHI1 shown in black. Relevant restriction sites *Bu* (*Bsu*36I), *Bm*, (*Bsm*I), *S* (*Sna*BI), *P* (*Pf*MI), *Hc* (*Hinc*II, where the selection cassette was inserted) are indicated. Arrows above the map indicate the position of the deletions, while the lower insert shows the mutation introduced in the MTHI1 binding site (underlined) in the dI_Mf strain, with mutated nucleotides shown in red. A *Pvu*II site introduced as a RFLP marker is boxed.

D) Accumulation of the chimeric *petA* transcript in the {*aAdI*} strain transformed with the indicated *dIf* variants. Three independent transformants are shown for each construct. The *psaB* mRNA is shown as a loading control.

E) Accumulation of the chimeric cytochrome *f* in the same strains, and in the {*aAdI*}, *mthi1-1* {*dIf*} and Δ*petA* strains as controls. Immuno-detection of OEE2 provides a loading control.

409 *atpI* 5'UTR of the *dIf* reporter gene (Fig. 4A). Moreover, to avoid recombination between the

5'UTRs of the endogenous *atpI* gene and of the chimeras, the latters were introduced in the chloroplast genome of the {*aAdI*} strain, devoid of the *atpI* 5'UTR (Fig. 3A). A deletion of 168 bp in the *atpI* 5' UTR (Δ1) strongly decreased the accumulation of the chimeric transcript (Fig. 13D) and of its cytochrome *f* gene product (Fig. 13E). The deletion of the next 129 bp, either alone (Δ2) or together with the upstream 168 nt (Δ3), did not alter the accumulation of the chimeric mRNA nor of its gene product (Fig. 13D,E), suggesting that antagonistic regulatory elements at the beginning and in the middle of the *atpI* 5'UTR fine-tune the expression of the *atpI* gene, as already observed in other 5'UTRs (Costanzo and Fox, 1993; Sakamoto et al., 1994). Deletion of 86 bp encompassing the GGTTATTAT motif (Δ4) decreased the accumulation of the chimeric transcript that remained, however, more abundant than in strains carrying the Δ1 deletion, but totally abolished its translation. Mutation of this motif to T**CAGCTGC**A, leaving the rest of the UTR unaltered, led to the same decreased accumulation of the chimeric transcript than in *mthi1* mutants, and prevented cytochrome *f* expression, confirming its importance for *atpI* mRNA translation.

425

Discussion

MTHI1, a major actor of CFo biogenesis.

Here we show that MTHI1 has a dual role in controlling the expression of AtpH and AtpI, the two subunits of the selective proton channel and, therefore, is a major actor in the biogenesis of the CFo sector of chloroplast ATP synthase.

MTHI1 is required for the stable accumulation of the monocistronic *atpH* mRNA. Being an OPR protein, it likely binds directly its RNA target, as was shown for the OPR factor TAB1 (Rahire et al., 2012). As do PPR proteins in plants (reviewed in (Barkan and Small, 2014), in Chlamydomonas, OPR proteins are involved in all post-transcriptional steps of chloroplast gene expression: Maturation/stabilisation (Murakami et al., 2005; Kleinknecht et al., 2014; Wang et al., 2016; Cavaiuolo et al., 2017; Viola et al., 2019), translation activation (Auchincloss et al., 2002; Eberhard et al., 2011; Rahire et al., 2012; Lefebvre-Legendre et al., 2015), splicing (Rivier et al., 2001; Balczun et al., 2005; Merendino et al., 2006; Marx et al., 2015; Reifschneider et al., 2016). As other M factors (Loiselay et al., 2008; Wang et al., 2015; Cavaiuolo et al., 2017), MTHI1 binds to the very 5'end of its target mRNA to protect it from 5' → 3' exonucleases, whose action can alternatively be impaired by addition of a polyG cage at the beginning of the transcript. The 9 OPR repeats-containing MTHI1 protein interacts with the first nine nucleotides (GGTTGTTAT) of the *atpH* mRNA, highly conserved among Chlorophyceae, Pedinophyceae, Nephroselmidophyceae and in the Ulvale clade of the polyphyletic (Sun et al., 2016) Ulvophyceae class and whose mutation prevents *atpH* mRNA accumulation. This interaction results in a specific footprint, co-immuno-precipitated with the MTHI1 protein and highly reduced, although not totally abolished, in the *mthi1-1* mutant. Whether this is due to the leakiness of the *mthi1-1* mutant, which reverts to some extent when platted on Minimum medium, to a low affinity of other OPR proteins for *atpH* mRNA or to an intrinsic stability of tri-phosphorylated transcripts that are poor substrates for 5' → 3' exonucleases (Richards et al., 2011; Luciano et al., 2012; Foley et al., 2015) remains to be determined. The monocistronic *atpH* mRNA, transcribed from its own promoter, does not result from the processing of precursors transcribed from the *atpA* promoter. Although accumulating to wild-type level in *mthi1* mutants, they could not translate the AtpH subunit in the absence of the MTHI1 *atpH* translation activator. In the wild type, the AtpH subunit is probably not synthesised either from these precursors despite the presence of MTHI1, as the target of MTHI1 is sequestered within a stable secondary structure (Fig. 14A), probably preventing the binding of OPR proteins, as do secondary structures for

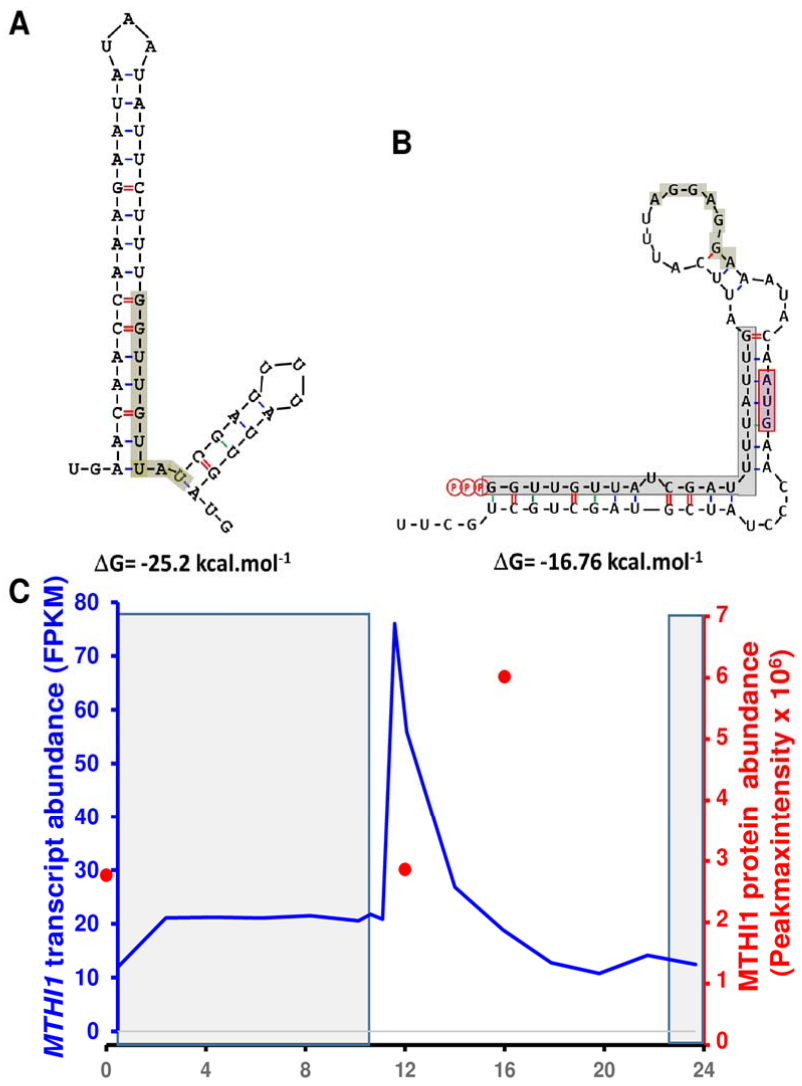


Fig. 14: Modulation of MTHI1 action.

A) Secondary structure sequestering the MTHI binding site in the precursor RNA transcribed from the *atpA* promoter.

The lowest energy structure calculated at 25°C by RNA Folding Form (M-Fold: http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi; Zuker, 2003) for the region surrounding the *atpH* 5' end in the precursor transcript initiated at the *atpA* promoter. The MTHI1 binding site is yellow-shaded.

B) Secondary structure of the *atpH* 5'UTR in the *dHf* chimera, sequestering the initiation codon.

The lowest energy structure calculated at 25°C by M-Fold for the transcribed region of *atpH* inserted upstream of the *petA* gene in chimera *dHf*. The footprint of MDH1 is grey-shaded, while the *atpH* initiation codon is pink-shaded and the Shine-Dalgarno yellow-shaded.

C) Variations of MTHI1 transcript and protein accumulation over the circadian cycle.

Redrawn from the data in Strenkert et al, 2019. The dark period is indicated by the shaded area. Blue line shows the accumulation of the *MTHI1* transcript over time (expressed as Fragments Per Kilobase Million (FPKM), the red dots show the accumulation of the MTHI1protein at the indicated time points (expressed as Peakmaxintensity).

460 PPR proteins (Kindgren et al., 2015; Zoschke et al., 2016; Miranda et al., 2017; McDermott et

al., 2018).

The fate of *trans*-acting factors during translation remains poorly known. Most of those that have been studied are not found in polysomal fractions (Boudreau et al., 2000; Auchincloss et al., 2002; Dauvillee et al., 2003; Schwarz et al., 2007; Viola et al., 2019). Either their association does not resist the polysome preparation procedure or they dissociate from their target mRNA upon translation, raising the question of the stability of translated mRNAs (Kato et al., 2006; Viola et al, 2019). MTHI1, however, remains associated with the *atpH* mRNA loaded on polysomes, while sRNAs derived from the ribosomal cluster were enriched in MTHI1 RIP samples (even if *rrnS* signal was not observed in dot-blots). This unique behaviour may favour the re-initiation of *atpH* mRNA translation, whose rate of translation in exponentially growing cells is higher than that of most other photosynthetic transcripts.

MTHI1 also contributes to the stabilisation of the *atpI* mRNA, while strongly enhancing its translation. However, we did not detect any specific footprint within the *atpI* 5'UTR, nor did we find evidence for a binding of MTHI1 that would resist RIP experiments, whether analysed by dot-blot or deep sequencing. We previously failed similarly to observe a footprint diagnostic of an interaction of the translation activator TCA1 with its *petA* 5' UTR target (Cavaiuolo et al, 2017), despite experimental evidence for TCA1 interaction with this RNA region (Loiselay et al., 2008; Boulouis et al., 2011). T factors, here MTHI1, probably interact only transiently with their target transcript, here the *atpI* 5'UTR, to promote translation. However, mutating the putative MTHI1 binding site within the *atpI* 5'UTR destabilised the 5'*atpI-petA* chimeric transcript as in *mthi1* mutants, and totally prevented the synthesis of a reporter protein, highlighting its importance for the expression of the *atpI* gene.

PPR10 and MTHI1: an example of convergent evolution.

The mode of action of *Cr*MTHI1 strikingly resembles that of *Zm*PPR10, even though the two proteins are not evolutionary related, as they belong to different protein families (OPR vs. PPR). The maize PPR10 protein targets the *atpI-atpH* intergenic region to stabilise the transcripts of these adjacent and co-transcribed genes, by respectively protecting them from 3' → 5' and 5' → 3' exonucleases (Pfalz et al., 2009). The binding of PPR10 generates a footprint matching the overlapping ends of the *atpI* and *atpH* transcripts (Zhelyazkova et al., 2012). In addition, PPR10 activates the translation of the *atpH* mRNA by opening a secondary structure that would otherwise sequester the Shine-Dalgarno sequence (Prikryl et al., 2011). MTHI1 may similarly activate the translation of the *atpH* mRNA by opening a

secondary structure sequestering the *atpH* initiation codon (Fig. 14B). However, at variance with MTHI1, PPR10 is not involved in *atpI* mRNA translation activation (Zoschke et al., 2013).

The two target genes of MTHI1 are widely separated on the chloroplast genome

By targeting two genes, widely separated on the chloroplast genome, MTHI1 appears unusual when compared to other factors characterised so far in *C. reinhardtii*. They target a single chloroplast transcript to allow its stable accumulation (Kuchka et al., 1989; Drapier et al., 1992; Drager et al., 1998; Boudreau et al., 2000; Loiselay et al., 2008; Johnson et al., 2010; Wang et al., 2015; Cavaiuolo et al., 2017) or activate its translation (Rochaix et al., 1989; Stampacchia et al., 1997; Wostrikoff et al., 2001; Auchincloss et al., 2002; Dauvillee et al., 2003; Raynaud et al., 2007; Schwarz et al., 2007; Eberhard et al., 2011; Lefebvre-Legendre et al., 2015; Cavaiuolo et al., 2017). A recent genome-wide ribosome profiling study performed on the Chlamydomonas *nac2* (*mbd1-nac2*) mutant, defective for the accumulation of the *psbD* mRNA, only detected very limited changes in chloroplast gene expression, most of which were attributed to PSII deficiency, rather than to the absence of NAC2 *per se* (Trosh et al, 2018). The only exception so far is the MBB1 factor, required for the stable accumulation of the *psbB* mRNA, coding for the CP47 core antenna of PSII, as for the correct processing and translation of the co-transcribed *psbH* mRNA, encoding another PSII subunit (Monod et al., 1994; Vaistij et al., 2000b; Vaistij et al., 2000a; Loizeau et al., 2014). In both cases, these bifunctional factors target two subunits in tight interaction in the assembled complex (Komenda et al., 2005; Boehm et al., 2011; Murphy et al., 2019), whose synthesis is highly interdependent in other organisms (Jean-Francois et al., 1986; Ooi et al., 1987; Payne et al., 1991; Komenda, 2005; Bietenhader et al., 2012). Such bifunctional factors would thus provide a mechanism alternative to the CES process for a co-regulated expression of closely interacting subunits.

The landscape of the nuclear control of chloroplast gene expression in Chlamydomonas appears widely different from that in vascular plants. Land plants *trans*-acting factors show a looser specificity. When targeting a polycistronic transcript, they may define both the 3'end of the upstream transcript and the overlapping 5'end of the downstream transcript (Pfalz et al., 2009; Zhelyazkova et al., 2012). Moreover, they often bind similar sequences in different transcription units, often coding for subunits of different protein complexes. The maize protein CRP1 activates the translation of both *petA* and *psaC* transcripts and is also required for the processing of *petB* and *petD* monocistronic RNAs in maize as in Arabidopsis (Barkan

et al., 1994; Fisk et al., 1999; Schmitz-Linneweber et al., 2005; Ferrari et al., 2017). The maize PPR10 protein, in addition to its role in *atpI* and *atpH* expression, also controls the accumulation of the monocistronic transcripts of the adjacent *rpl23* and *psaJ* genes (Pfalz et al., 2009; Prikryl et al., 2011; Zhelyazkova et al., 2012). A recent genome-wide ribosome profiling study revealed an even more complex situation by highlighting the unexpected versatility of several PPR proteins in plants, since PPR10 also stabilises the monocistronic *psaI* mRNA, while PGR3 binds to the *rpl14-rps8* intergenic region to stabilise the *rpl14* mRNA at its 3'end and to stimulate *rps8* translation (Rojas et al., 2018).

The paradoxical specificity of *trans*-acting factors in C. reinhardtii.

The high specificity of *trans*-acting factors in *C. reinhardtii* appears paradoxical, as, for example, the GTT(G/A)TTAT target of MTHI1 is not restricted to the *atpH* or *atpI* mRNAs but is found several times (3 times for GTTGTTAT, in *atpH*, *rpoC2* and *rpoC1* transcripts; twice for GGTTATTAT, in *atpI* and *rps3* transcripts; 11 times for the more degenerated GGTTNTTAT motif) in the chloroplast transcriptome of *C. reinhardtii*. These extra motives do not lead, however, to footprints nor to sRNAs enrichment in MTHI1-RIP samples, which suggests that the affinity of MTHI1 for its GGTTGTTAT target remains moderate and requires additional determinants, presently unknown, for its strong interaction with the *atpH* 5'UTR. This interaction leads to the formation of an abundant footprint, whereas that with a very similar motif in the *atpI* 5'UTR does not. This is unlikely to result from a differential affinity of MTHI1 for the GGTT**G**TTAT vs. GGTT**A**TTAT sequence: changing one for the other in the 5'UTR of the *dHf* and *dIf* chimeras did not lead to noticeable changes in cytochrome *f* expression, while changing GGTT**G**TTAT to GGTT**A**TTAT in the *atpH* transcript did not modify its expression, neither at the RNA nor at the protein levels (D. Jarrige, Y. Choquet, unpublish. res.).

The correlated abundances of MTHI1 and *atpH* mRNA in a series of transformants argues for MTHI1 being limiting for the expression of *atpH*. The stimulated expression of the *dHf* and *dIf* chimera in the absence of the *atpI* or *atpH* genes, respectively, suggests that the two genes share some common factors, MTHI1 being a likely candidate. However, the deletion of the abundant *atpH* mRNA, stoichiometrically bound to its stabilisation factor, should release much more MTHI1 protein than the deletion of the *atpI* gene, whose mRNA, 10-fold less abundant (Cavaiuolo et al., 2017), interacts only transiently with its translation activator. Still, deleting the *atpI* gene stimulates much more the expression of the *dIf* chimera

than deleting *atpH*: other factor(s) specific to the 5'UTR of the *atpI* mRNA should be limiting for *atpI* expression.

The interaction of several factors assembled in a complex on a target 5'UTR may, despite a moderate specificity/affinity of each of them for its target, lead to a strong cooperative interaction, much more stable than that between any two components taken separately. An *atpH*-specific factor interacting with both the *atpH* 5'UTR and MTHI1 could tether it on the *atpH* 5' end, but not on other occurrences of the same motif. A weak affinity of an *atpI*-specific factor for MTHI1 may similarly results in a transient, but still specific, interaction with the *atpI* 5'UTR. Such cooperative interactions prevail for the few chloroplast genes whose expression has been studied in detail in *C. reinhardtii*. MCA1 and TCA1 are respectively strictly required for the accumulation and translation of the *petA* transcript, whose translation is nevertheless reduced 10 fold in the absence of MCA1, while its stability is decreased by 85% in the absence of TCA1 (Wostrikoff et al., 2001; Raynaud et al., 2007; Loiselay et al., 2008). The two factors form a ternary complex with the *petA* mRNA (Boulouis et al., 2011) and the absence of any of them weakens the interaction between the other two. Similarly, the NAC2 stabilisation factor of the *psbD* transcript recruits the RB40 protein to activate the translation of the *psbD* mRNA, despite the poor specificity of this later for U-rich regions (Schwarz et al., 2007). Last, MDA1 and TDA1, respectively required for the accumulation and translation of the *atpA* mRNA, also form a complex assembled onto the *atpA* mRNA (Viola et al., 2019).

Such a "dually footed" mechanism could favour the high plasticity of nucleo-chloroplastic interactions observed in Chlorophyceae: despite a mutation in its target, a *trans*-acting factor would, through its interaction with other factors, remain in contact with it, allowing the selection of compensatory mutations over time. It also helps to understand the recycling of M factors: once the target mRNA degraded, the complex will dissociate, and because of the moderate affinity of the M factor for its target, the footprint sRNA will be released, rather than trapped, allowing the protein to interact with newly synthesised mRNAs.

The co-regulation of *atpH* and *atpI*: an ancestral situation

The joint control of *atpH* and *atpI* expression can be traced back during evolution. In *Escherichia coli*, the *unc* operon organisation facilitates the concerted expression of ATP synthase subunits, even if additional translational controls are required to set their contrasted stoichiometry. Cyanobacteria, including *Gloeomargarita lithophora*, the extant free-living cyanobacterium most closely related to the ancestor of chloroplasts (Ponce-Toledo et al.,

2017), partially retained this gene organisation, with ATP synthase subunits now encoded by two distinct operons: *atpI-atpH-atpG-atpF-atpD-atpA-atpC* and *atpB-atpE* (for the sake of clarity, cyanobacterial genes are named here as their chloroplast counterparts, rather than by their true name: e.g. the *atpE* locus of Gloeomargarita encoding subunit C (AtpH) is nevertheless named *atpH*). This ancestral organisation was largely preserved in Archeplastidia: while the genes encoding subunits γ, δ, AtpI and ATPG may have been relocated to the nucleus in some species, those remaining in the chloroplast still belong to two transcription units (*atpI)-atpH-(atpG)-atpF-(atpD)-atpA* and *atpB-atpE*. A noticeable exception are the Chlorophyceae in which *atp* genes are shuffled around the chloroplast genome (Dataset DS1), raising the question of their co-regulation.

In the Ulvale clade of Ulvophyceae and in Pedinophyceae, the *atpI* and *atpH* genes, although adjacent on the chloroplast genome, share in their 5'UTR a sequence similar to the MTHI1 binding site (Dataset DS1 and Suppl. Fig. 8A,B). This suggest an ancestral situation that placed the expression of the two genes under the control of an orthologue of MTHI1, paving the way for their separation in Chlorophyceae. This sequence possibly appeared early during evolution in the common ancestor of Pedinophyceae, Ulvales and Chlorophyceae, together with the appearance of an efficient processing system that generates, in green algae chloroplasts, independent monocistronic transcripts from the polycistronic transcription units that are remnants of the ancestral cyanobacterial operons.

MTHI1 is conserved in Chlorophyceae

In Chlorophyceae, the conservation of the MTHI1 target goes along with the conservation of the MTHI1 protein, since all sequenced genomes, with the exception of *Coelastrella*, encode an orthologue of MTHI1. The region of similarity is restricted to the OPR-containing N-terminal part of the protein. Even this "conserved" region evolves rapidly, with multiple species-specific insertions, some of which interrupt the OPR repeats (Suppl. Fig. S6). Strikingly, the two Ulvale genomes presently available each encode an OPR protein with 9 OPR repeats (Suppl. Fig. S6), which are the mutual best hits of *Cr*MTHI1 and are predicted, based on a preliminary version of the OPR code (Manuscript in preparation), to recognise the GGTTGTTAT sequence. These OPR are shorter than their Chlorophycean orthologues as they lack the disordered C-terminal extension.

Downstream of this conserved region, all chlorophycean MTHI1 orthologues possess a C-terminal tail, rich in stretches of identical residues, mostly A, S, Q, and R, predicted to be essentially a random coil. These tails are not conserved in sequence nor in length and do not

show similarity to other proteins in databases, suggesting that they have no specific functions. Indeed, in *C. reinhardtii*, the tail appears dispensable for the major function of the protein, as are also the N-terminal tails of TCA1 (Raynaud et al., 2007), NAC2 (Boudreau et al., 2000), RAA1 (Merendino et al., 2006) and TDA1 (Eberhard et al, 2011) or the C-terminal tail of MRL1 (Johnson et al., 2010). These tails could result from the introduction of "junk" GC-rich DNA within permissive regions of the genes or from the loss of Stop codons upon mutations in GC rich regions, extending progressively the coding sequence. However, while the full-length MTHI1 factor is short-lived with a half-life of about 1 hr, its C-ter truncated version is stable over 8 hr. Possibly, these tails, a common feature of *trans*-acting factors in Chlorophyceae, modulate the stability of the proteins, althrough the proteolytic process controlled by these tails is unknown.

The accumulation of the short-lived full-length MTHI1 protein thus depends on changes in the abundance of the *MTHI1* transcript, as occurs over the circadian cycle (Fig. 14C). MTHI1, being limiting for the expression of *atpH* and *atpI*, would couple the expression of these genes to that of the nucleus-encoded ATP synthase subunits, whose transcripts show a similar pattern of expression (Fig. 8F in Zones et al., 2015). Thus it behaves as genuine regulator of ATP synthase biogenesis.

648

METHODS

Strains, Media, Culture Conditions, and Chemicals

Wild-type t222^{+} (derived from 137c: *nit1 nit2*), mutants, and transformed strains of *C. reinhardtii*, were grown at 25°C in Tris-acetate-phosphate (TAP) medium, pH 7.2 (Harris, 1989), under continuous light (5-10 µE m^{-2} s^{-1}; white LED, whose emission spectrum is shown in Suppl. Fig. S9), unless otherwise specified. Crosses were performed according to (Harris, 1989).

Constructs and Nucleic Acid Manipulations

Standard nucleic acid manipulations were performed according to (Sambrook et al., 1989). Primers used in that study are listed in Suppl. Table ST1. All DNA constructs were sequenced before transformation in Chlamydomonas. Details of the DNA constructs are given in the Suppl. Method section.

RNA Isolation and Analysis

RNA extraction and RNA gel blot analysis were performed as described (Drapier et al., 2002) with ^{33}P-labelled probes derived from coding sequences (Eberhard et al., 2002). Transcript accumulation was quantified from PhosphorImager scans of the blots, as described by (Choquet et al., 2003). In Figs. 2B and 6C, probes amplified with primers listed in Table ST1 were digoxigenin-labelled, using DIG-dUTP, the antidigoxigenin Fab fragment and CDP Star reagent (Roche, Basel, Switzerland). Signal was acquired in ChemiTouch (BioRad, Hercules, CA, USA) and analysed with the ImageLab software (v 3.0, BioRad). Transcriptomic analyses were performed as described in (Cavaioulo et al, 2017). In Fig. 2, polysome analyses were performed as described in (Minai et al., 2006; Eberhard et al., 2011). In Fig. 12 we used a modified protocol adapted from (Trosch et al., 2018): Cell cultures were grown to mid-logarithmic phase (2-3 x 10^6 cells ml^{-1}) and supplemented with 100 µg ml^{-1} chloramphenicol 15 min before harvesting. Cell pellets were resuspended in polysome buffer (20 mM Tris pH 8.0, 25 mM KCl, 50 mM β-mercaptoethanol, 0.5 mg ml^{-1} heparin, 100 µg ml^{-1} chloramphenicol, 0.2 M sucrose, 1% (v/v) Triton X-100, 1 x Protease Inhibitor Cocktail (Roche)), with or without $MgCl_2$ (25 mM). Cells were broken with a French Press and cell lysates were centrifuged at 10,000 g for 15 min at 4°C to remove cell debris. EDTA samples were prepared without $MgCl_2$ and supplemented with 20 mM EDTA. $MgCl_2$ and EDTA supernatants were loaded on a 20 – 50% (w/v) continuous sucrose gradient. The 20% and 50% sucrose solutions were prepared in a buffer containing 20 mM Tris pH 8.0, 25 mM KCl,

5 mM β-mercaptoethanol, 0.5 mg ml^{-1} heparin and 100 µg ml^{-1} chloramphenicol and either supplemented with 25 mM $MgCl_2$ or 1 mM EDTA. Sucrose gradients were centrifuged at 38,000 rpm for 150 min in a SW41 Ti rotor (Beckman). 11 fractions were collected and the pellet was resuspended in 1.1 ml solution containing 5 mM EDTA and 0.1% SDS.

Transformation Experiments (listed in Table I)

Chloroplast transformation was achieved by tungsten particle bombardment (Boynton et al., 1988) as described in (Kuras and Wollman, 1994) using a home-made helium gun. Transformants were selected on TAP-Spec (100 µg mL^{-1}) and subcloned on TAP-Spec (500 µg mL^{-1}) until they reached homoplasmy, assessed as described in Table 1. For each transformation, at least four independent transformants were analysed. Phenotypic variations between independent transformants proved negligible.

Nuclear transformation of *mthi1* strains was performed by electroporation, as described by (Raynaud et al., 2007), with the following parameters: 10 mF/1200 $V \cdot cm^{-1}$. Transformants were selected for phototrophy on minimum medium (Harris, 1989) under high light (150 $\mu E \cdot m^{-2} \cdot s^{-1}$).

Protein Preparation, Separation, and Analysis

^{14}C pulse-labelling experiments in the presence of cycloheximide (10 µg mL^{-1}), protein isolation, separation, and immunoblot analyses were performed on exponentially growing cells (2-3 10^6 cells$\cdot mL^{-1}$) as described (Kuras and Wollman, 1994). Immunoblots were repeated at least twice and performed on three independent transformants. Cell extracts, loaded on equal chlorophyll basis, were analysed by SDS-PAGE (12-18% acrylamide gradients and 8 M urea). At least three biological replicas were performed for each experiment. Proteins were detected by ECL. Primary antibodies, diluted 100,000-fold (antibodies against cytochrome *f*, D1, and PsaA), 50,000-fold (CF1β, tubulin subunit α), 10 000-fold (AtpH, CGE1, RbcL, Rps12 and ATP synthase subunit γ), 5 000-fold (ATP synthase subunits δ, ε, and AtpI), 2 500-fold (AadA) were revealed by horseradish peroxidase–conjugated antibodies against rabbit IgG (#W401B, Promega). Antibodies against the OEE2 subunit from the photosystem II oxygen-evolving complex, the β-subunit of F1/CF1, cytochrome *f*, and CGE1 have been respectively described in (de Vitry et al., 1989; Atteia et al, 1992; Lemaire and Wollman, 1989b; Kuras and Wollman, 1994; Schroda et al., 2001). Antibody against Rps12 was kindly provided by S. Ramundo (Ramundo et al., 2013), antibodies against D1 (#AS05 084), PsaA (#AS06 172), RbcL (#AS03 037), AadA

(#AS09 580) and the ATP synthase subunits γ (AS08 312), δ (#AS10 1590), ε (#AS10 1586), AtpH (#AS09 591), and AtpI (#AS10 1583) were purchased from Agrisera and used according to manufacturer's instructions. Antibody against the α subunit of tubulin was purchased from Sigma (#MABT868). MTHI1-HA was detected by ECL using monoclonal anti HA.11 (# MMS-101R, Covance) antibodies, and horseradish peroxidase–conjugated antibody against mouse IgG (#W402B, Promega). Protein accumulation (normalised to that of OEE2 or β F1, as internal standards) was, when required, quantified from ChemiTouch (Bio-Rad) scans of the membrane, using the ImageLab (v3.0) software. For immuno-chase, cytosolic translation was arrested by supplementing cells grown in TAP medium (2-3 x 10^6 cells mL^{-1}) with cycloheximide (final concentration 10 μg mL^{-1}) at t = 0 and aliquots were taken at the indicated time points.

Gel Filtration Experiments on Soluble Cellular Extracts

Size exclusion chromatography was performed according to (Boulouis et al., 2011) with minor modifications. Cells from a 600-mL culture (2-3 10^6 cells·mL^{-1}) were centrifuged, resuspended in 3 mL of breaking buffer (5 mM HEPES-KOH, pH 7.8, 20 mM KCl, 10% glycerol, 0.5 g·L^{-1} heparin, and 5 x Roche protease inhibitors in DEPC treated water), broken with a French press at 6,000 p.s.i. and centrifuged at 346,000 g for 20 min to pellet membranes, debris and unbroken cells. 500 μL of the supernatant were loaded on a Superose 6 10/300 HR column (GE healthcare, USA). Chromatography was performed on Biologic DuoFlow (Biorad) chromatography system and protein elution, monitored on UV channel of the QuadTec device, was performed at a rate of 200 μL·min^{-1}, at 4°C, with a buffer containing 80 mM Tricine-KOH, pH 7.8, 200 mM KCl, 20 mM ε-aminocaproic acid, and 0.1 x Roche protease inhibitors. Sixteen 1 mL fractions, eluted 16 mL after void volume (8 mL), were collected and concentrated by centrifugation on Amicon Ultra-15 filter units (cutoff: 30 kD) at 4,500 g for 20 min. Fraction volumes were adjusted to 100 μL, out of which 20 μL were loaded on 8 % acrylamide gels containing 8 M urea. Fraction 16 (lower molecular mass) lacked protein and was not loaded on the gels. For RNase treatments, stromal preparations, prepared in breaking buffer lacking heparin, were incubated at 4 °C with 2500 U·mL^{-1} of RNase I and 625 U mL^{-1} of RNaseI for 45 min under gentle and continuous shaking, prior to loading on the column. For further analysis by co-immunoprecipitations, the indicated fractions were pooled, concentrated on Amicon Ultra-15 filter and adjusted to 1 mL with Lysis buffer before co-IP.

Coimmunoprecipitations

were performed according to (Boulouis et al., 2011) with minor modifications. Cells from a 400-mL culture (2 10^6 cells·mL^{-1}) were centrifuged, resuspended in 2 mL of lysis buffer (20 mM HEPES-KOH, pH 7.2, 150 mM NaCl, 10 mM KCl, 1 mM $MgCl_2$, 10% glycerol, and 2 x Roche protease inhibitors in DEPC treated water), broken by a French press at 6,000 p.s.i and centrifuged at 34,000g for 30 min to pellet membranes and debris. 500 µL of supernatant, supplemented with 0.2% Tween 20 were incubated for 1 h at 4°C in the presence of 20 µL of anti-HA-tag magnetic beads (Medical Biological Laboratories international, Japan), pre-equilibrated with Lysis buffer supplemented with 0.2% Tween 20. Beads were then washed three times with washing buffer (150 mM NaCl, 20 mM HEPES-KOH, pH 7.2, 10% glycerol, 0.2% Tween 20, 1 x Roche protease inhibitors) and twice more with 10 mM Tris-HCl, pH 7.5. Bound proteins were detached by boiling the beads for 2 min in the presence of 2% SDS, while for RNA purification, immunoprecipitation beadswere resuspended in 250 µL AE buffer (50 mM Na-acetate pH 5.2, 10 mM EDTA) and extracted with phenol/ chloroform/IsoAmyl Alcohol (25:24:1) before ethanol precipitation in the presence of 2 µL GlycoBlue (Invitrogen, USA).

Two-Step Centrifugation Procedure

Cells from 400-mL culture (2-3 10^6 cells·mL^{-1}) were centrifuged, resuspended in breaking buffer (final volume of 4 mL), broken by French press (6,000 p.s.i.) and centrifuged at 2,100 g for 5 min to remove unbroken cells, starch, and large debris. One mL of the supernatant (Input I) was ultracentrifuged at 272,000g for 30 min. The supernatant (S) was recovered and the pellet (P) was resuspended in 1mL of breaking buffer. After spectroscopic determination of chlorophyll concentration in the input fraction, equal volumes of the S and P samples were loaded on gel and analysed by immunoblot.

Accession Numbers: Sequence data from this article can be found in the NCBI Dabases under the accession numbers indicated in Dtaset DS1 for *atpH* and *atpI* 5'UTRs and for *MTHI1* coding sequences; *petA*, FJ423446.1; *psbD*, X04147.1; OEE2, M15187.1; *atpA*, X60298.1; CGE1: AAK96224.1; Rps12: AAC16329.1; RbcL: ASF83644.1; PsaB: P09144.4; *rrnS*: J01395.1; MG052656.1. *atpB*: M13704.1; PsaA: 1310243A; PsbA: 1102190A.

Figure legends:

Figure 1: Phenotype of mthi1 mutants

A) Schematic representation of the *atpH* (top) and *atpI* (bottom) transcription units. Coding sequences are shown as thick rectangles, while 5'UTRs are depicted as thin rectangles. Bent arrows represent promoters. The major transcripts detected in panel C with probes specific to *atpH* or *atpI* are indicated. (0) stand for a precursor transcript that cannot be observed in the wild type because it is efficiently processed, but can be detected in *psaA* trans-splicing mutants (Choquet et al, 1988). Scissors indicate the position of processing events, whose efficiency symbolised by their size.

B) Pleiotropic loss of ATP synthase subunits in *mthi1* mutants.

Total cell extracts of wild-type (a dilution series is shown) and of the two *mthi1* mutants strains were probed with antibodies against the proteins indicated on the left. The accumulation of all ATP synthase subunits was dramatically decreased in the two mutant strains, while that of cyt. *f*, PsaA, D1 and OEE2, respectively used as proxies of the abundance of the cytochrome b_6f complex, photosystem I (PSI) and PSII was unaffected. The red asterisk points to a cross-reaction of the antibody, preserved in the mutant strains, against the γ subunit of mitochondrial ATP synthase.

C) (Top) Accumulation of the *atpH* and *atpI* transcripts in the same strains. The *psaB* transcript provides a loading control.

(Bottom) Rate of translation of ATP synthase subunits in the same strains, assessed by 5' pulse labelling experiment in the presence of ^{14}C acetate (5 µCi.mL^{-1}) and of the cycloheximide inhibitor of cytosolic translation (10 µg.mL^{-1}). The positions of the AtpI and AtpH subunits are indicated.

D) Quantification of *atpI* transcripts amount in wild-type and mutant strains.

(Left) Relative accumulation of the four *atpI*-containing transcript in the wild type, expressed as the percentage of the total amount of *atpI* transcript. (Right) Relative abundance of each *atpI* transcript, and of the sum of them, reported to that of the same band in the wild-type (set to 100, symbolised by a grey dashed line) in the two mutants (dark grey *mthi1-1*; light grey *mthi1-2*; n=4).

Figure 2: The MTHI1 factor controls the translation of the atpI mRNA

A) Loading of *atpI* mRNAs on polysomes.

Solubilised whole-cell extracts (T) from wild-type, Δ*atpH* and the two *mthi1* mutant strains, pre-treated for 10 min with CAP (200 µg mL^{-1}) were loaded on sucrose gradients. After ultracentrifugation, ten fractions were collected and the transcripts present in each fraction were analysed by RNA blots using the probes indicated on the left.

B) Defective *atpI* mRNA translation is not responsible for its decreased abundance in *mthi1* mutants.

(top) Schematic representation of the changes introduced into the *atpI* gene. Mutated nucleotides are shown in bold: they change the initiation codon (written in red) to a stop codon and introduced a *Bgl*II RFLP marker (underlined).

(middle) Phototrophic growth of the $atpI_{St}$ and $atpI_{Ct}$ strains assessed on minimum medium (devoid of acetate) under 75 µE m^{-2} s^{-1}. Three independent transformants are shown. The growth of the strain on TAP medium (15 µE m^{-2} s^{-1}), as well as the growth of the wild type and of the Δ*atpI* strain are shown as controls.

(bottom) Accumulation of *atpI* transcripts, schematically depicted in panelA in a control strain bearing the *aadA* cassette alone and in strains bearing the *aadA* cassette associated with the untranslatable $atpI_{St}$ gene. Three independent transformants are shown for each construct. Because of the polar effect of *aadA* cassette, co-transcripts with *atpJ* and/or *rps12* cannot be observed. The origin of the transcripts indicated by an aterisk (*) is unknown. *petB* provides a loading control.

C) *atpH* and *atpI* gene expression in the wild-type, Δ*atpH*, Δ*atpI* and *mthi1-1* strains.

(top) Accumulation of the *atpH* and *atpI* transcripts. *psaB* provides a loading control. (bottom) Rate of translation of the *atpH* and *atpI* transcripts in the same strains, assessed as in Fig. 1B by pulse labelling experiments. The positions of the AtpI and AtpH subunits are indicated.

Fig. 3: the MTHI1 factor targets the atpI 5'UTR.

A) Schematic representation of the *aAdI* chimera, where the *atpI* 5'UTR had been replaced by the promoter and 5' untranslated regions of the *psaA* gene. The position of the recycling *aadA* cassette (K^r), inserted in reverse orientation with respect to *atpH*, is shown.

B) Photoautotrophic growth of the *aAdI* strain, assessed as in panel 2B.

C) (Left) *atpH* and *atpI* transcript accumulation in the wild-type and *mthi1-1* strains transformed by the *aAdI* construct, whose transcript is shorter than the endogenous *atpI* transcript, because of the small size of the *psaA* 5'UTR. The recipient strains are shown as well as the Δ*atpH* and Δ*atpI* strains for controls. Three independent transformants are shown

for each genetic background. The *psaB* transcript provides a loading control. (right) Rate of AtpH and AtpI synthesis in the wild-type and *mthi1-1* strains and in the corresponding strains transformed by the *aAdI* construct, assessed as in Fig. 1B by pulse labelling experiments.

Fig. 4: the atpI 5'UTR is sufficient to confer a MTHI1-dependant translation to a reporter gene.

A) Schematic map of the *dIf* construct inserted instead of the endogenous *petA* gene. The red rectangle indicates the *psaA* promoter region placed upstream of the *psbJ-atpI* intergenic fragment (in light blue) that was chosen long enough to include the *atpI* processing site. The scissors above the intergenic region indicate the position of the 5'end of the processed *atpI* mRNA. The position of the recycling selection cassette, upstream of the chimeric *petA* gene and in reverse orientation with respect to this latter, is shown.

B) Photoautotrophic growth of the *dHf* (see Fig. 5) and *dIf* chimeric strains (three independent transformants) assessed as in fig. 2B. The growth of the wild type and of the *mthi1-2* strains are shown as controls.

C) Accumulation of the *petA* transcript, either endogenous or chimeric, in the chloroplast genome of the wild-type, *mthi1-1*, Δ*atpH*, Δ*atpI*, and Δ*atpH/I* recipient strains, shown aside, as well as a Δ*petA* strain for comparison. Three independent transformants are shown for each genetic context. The accumulation of the *atpH* mRNA in the same strains is also shown, while that of the *psaB* mRNA provides a loading control.

D) Accumulation of cytochrome *f* in the same strains (loading control: OEE2).

E) Quantification of the *petA* transcript (left) and cytochrome *f* (right) in transformed strains shows a competition between the chimera and the endogenous *atpI* gene for the expression of 5'*atpI*-driven genes. Value for the *dIf* transcript in the wild type recipient strain is set to 1; n=6.

Fig. 5: the MTHI1 factor targets the atpH 5'UTR.

A) Schematic representation of the *dHf* chimera, with the position of the recycling *aadA* cassette (in reverse orientation with respect to the *petA* gene) shown. The blue thick rectangle represents the first 25 nt of the *atpH* coding sequence fused in frame with the *petA* coding sequence, added to the construct to improve the expression of the chimera.

B) Accumulation of the *atpH* and *petA* transcripts in the wild type, *mthi1-1*, Δ*atpH,* Δ*atpI* and Δ*atpH/I* strains carrying the *dHf* chimera instead of the endogenous *petA* gene. Unstransformed wild-type, Δ*atpH*, Δ*atpI*, Δ*petA* and *mthi1-1* strains are shown as controls.

Asterisk indicates the position of the *atpH* mRNA, while the double asterisk points to a cross-reaction of the probe that comprises the *atpH* 5'UTR with the *dHf* chimeric transcript. Three independent transformants are shown for each genetic context. The *psaB* transcript provides a loading control.

C) Cyt. *f* accumulation in the same strains, with OEE2 as a loading control.

D) Quantification of the relative accumulation of the *petA* transcript (left) and of cyt. *f* (right) in the same strains. Values for *dHf* transformed in the wild-type strain are set to 1; n=6.

Fig. 6: MTHI1 is required for the translation of the atpH gene

A) Schematic map of the pG*atpH* construct with a zoom to the region surrounding the *atpH* transcription start site, indicated by a vertical arrow, where the polyG tract was inserted. The *atpH* promoter is underlined, and the position of the recycling *aadA* cassette is shown. A construct carrying the selection cassette at the same position but devoid of the polyG insertion was used as control ($atpH_{Ct}$). To avoid any polar effect on the expression of the downstream located *atpF* gene (co-transcribed with *atpH*), all experiments were performed after excision of the recycling *aadA* cassette.

B) Phototrophic growth of the pG*atpH*, *mthi1-2* {pG*atpH*}, {*aAdI* pG*atpH*} and *mthi1-2* {*aAdI* pG*atpH*} strains (two independent transformants each) assessed as in Fig. 2B. Growth of the wild type and of the *mthi1-2* and Δ*atpI* strains are shown as controls.

C) Accumulation of the *atpH* and *atpI* transcripts in the wild-type strain transformed by the $atpH_{Ct}$ and pG*atpH* constructs and in the *mthi1-2* strain transformed with the pG*atpH* gene. The *aAdI* construct then replaced the endogenous *atpI* gene in the resulting pG*atpH* and *mthi1-2* {pG*atpH*} strains. Two independent transformants are shown for each genetic background. The *petB* transcript provides a loading control.

D) Accumulation of the AtpH subunit in strains expressing the polyG construct. Tubulin provides a loading control.

Fig. 7: the MTHI1 protein.

A) Schematic representation of the MTHI1 protein. The position of the two *mthi1* mutations is shown. The brown rectangle depicts the chloroplast transit peptide as predicted by the ChloroP program. The green rectangle indicates the region of the protein conserved in other Chlorophyceae species (see Suppl. Fig S6), while the pink rectangle points to a rapidly evolving and disordered region. The lower scheme shows the predicted secondary structure of

916 the conserved region. Blue arrows represent the OPR repeats, whose sequence is shown in
917 panel (B), with the amino acid residues obeying the OPR consensus shaded in grey.
918
919 ***Fig.8: complementation of the mthi1-1 mutant strain.***
920 A) Complementation of the *mthi1* strain with with a tagged version of the *MTHI1* gene,
921 either the tagged cDNA (c clones) or the genomic construct (g clones), restores phototrophy,
922 assessed by plating the cells on Minimum Medium plates as in Fig. 2B. The growth of the
923 wild-type, Δ*atpH* and *mthi1-2* strains is shown as control.
924 B) Accumulation of the MTHI1 protein (red arrow), either endogenous or tagged, of the
925 AtpH subunit (top) and of the *atpH* and *atpI* transcripts (bottom) detected in the same strains
926 with an antibody against the MTHI21 protein. Note the larger size of the tagged protein,
927 compared to the endogenous one, due to the insertion of the triple HA tag. CGE1and cyt. *f* or
928 *petA* mRNA are shown as the respective loading controls in protein and RNA blots. The name
929 of the clone used for further analysis of MTHI1 in the next figures is written in red. (asterisk:
930 cross-contaminant).
931 C) MTHI1 is a soluble protein.
932 Cellular extract (I) from the complemented strain g9 was separated into soluble (S) and
933 insoluble (pellet: P) fractions by ultracentrifugation and equal volumes of each fraction were
934 probed with antibodies against the HA tag and against GrpE and cytochrome *f* as controls for
935 the purity of the fractions.
936 D) The C-terminal domain of MTHI1 is dispensable for its function
937 Accumulation of the tagged MTHI1 protein, probed with an antibody against the HA
938 tag, and of the *atpH* and *atpI* transcripts in *mthi1-1* strain complemented with the tagged
939 versions of the *MTHI1* gene, either the tagged cDNA (*c-HA*), the genomic construct (*g-HA*) or
940 its C-terminally truncated version (Δ*C-HA*). All transformants were selected for recovery of
941 phototrophy on MM plates. Overaccumulation of the truncated MTHI1 protein does not lead
942 to an increased abundance of the *atpH* transcripts. Cyt. *f* and *psaB* are shown as the respective
943 loading controls in protein and RNA blots. The name of the clones used for further analysis of
944 MTHI1 in the next figures are written in red.
945 E) Deletion of the C-terminal domain results in higher abundance of MTHI1.
946 (top) Accumulation of MTHI1 protein (red arrow) in *mthi1-1* strains complemented
947 with either the tagged *MTHI1* cDNA (*c*), the tagged genomic construct (*g*) or its C-terminally
948 truncated version (ΔCg), probed with an antibody against the MTHI1 protein. The name of
949 the clones used for further analysis of MTHI1 is written in red. (asterisk: cross-contaminant).

(bottom) quantification of MTHI1 accumulation, in the strains shown in top panel, normalised to that of cytochrome *f* and reported to the accumulation of MTHI1 in wild-type cells, set to 1. Error bars represent SD, n = 3.

F) The C-terminal domain of MTHI1 contributes to its high turn-over**.**

Stability of full-length MTHI1 or of its C-terminally truncated version, assessed by immunoblots in a culture treated with cycloheximide for the indicated times. Accumulation of OEE2 in the same samples is shown as a loading control.

G) Differential solubility of the full-length MTHI1 protein and of its C-ter truncated version.

Cellular extracts of transformants expressing the full-length (*) and the truncated (**) versions of the tagged MTHI1 protein, treated with cycloheximide for 0 or 4 hours (Input panel, left), were fractionated into soluble (S) and membrane (P) fractions and analysed as in panel D. Distribution of CGE1 and cyt. *f* are shown to assess the purity of the fractions.

Fig. 9: MTHI1 belong to a high molecular weight complex that interacts with the atpH and atpI transcripts.

Soluble extracts listed at the left of the figure were fractionated on a Superose 6 10/300 HR column and probed with an antibody against the HA tag. Molecular masses of the complexes found in each fraction were estimated by comparison with standards of the HMW gel filtration calibration kit (GE Healthcare).

Fig. 10: Transcriptional profile of the atpH and atpI genes.

A) Coverage, normalised as RPM (log scale) of pooled bi-directional and directional wild type WTSS along the *atpH* and *atpI* loci. Positions of the relevant genes and 5'UTR are shown below. The black bar in the *atpI* 5'UTR shows the position of the MTHI1 target (see below). Redrawn from the data in Cavaiuolo et al, 2017.

B) sRNA mapping at the 5'end of the *atpH* mRNA are the footprint of MTHI1.

Coverage, normalised as RPM, of pooled sRNA-Seq along the same loci: mock- (green) versus RPP-treated (blue) wild type sRNA-Seq libraries compared to RPP-treated libraries of the *mthi1-1* mutant (red). Coverage is averaged over two biological replicates. Only reads mapping to the coding strand are shown. The inset for *atpH* shows a zoom to the *atpH* 5'UTR and the sequence of the *atpH* footprint is shown. For *atpI* a zoom to the 5'UTR region (coding strand only) is shown in Suppl. Fig. S9A. Note the very different values on y-axes of the two graphs.

Fig. 11: The MTHI1 protein interacts specifically with the atpH 5'UTR.

A) The full-length and truncated versions of the MTHI1 protein were immuno-precipitated from a soluble cellular extract with an antibody against the HA tag (I: input; U: unbound; IP: immuno-precipitate). Immunoprecipitation of a cellular extract from the wild-type strain is shown as a negative control. The apparent slower migration of the immuno-precitated proteins in due to a "smiling" effect in the migration of the gel from which the composite figure (indicated by a vertical line) was made.

B) RNA extracted from immuno-precipitates were analysed by dot-blot hybridised to the probes indicated on the rigth. The lower panel shows the disposition of the samples on the filter. Top line: RNA extracted from the wild type, Δ*atpH*, and Δ*atpI* strains (without immuno-precipitation), as a control for the specificity of the probes. Bottom line: immuno-precipitated RNA from the wild-type and from a complemented strain expressing the tagged MTHI1 (g9).

C) Fractions from size exclusion chromatography of a cellular extract from a strain expressing the full-length MTHI1 (first line in fig. 9), indicated by the bars A (fractions 4 to 8) and B (9 to 11), were pooled, immuno-precipitated with an antibody against the HA tag and analysed with the same antibody for the MTHI1 content of the immuno-precipitated fractions. Their RNA content was extracted and analysed by dot blot with a probe specific of the *atpH* 5'UTR, which detected a (weak) signal in pooled fractions A, further analysed by deep sRNA-Seq.

D) Ratio of normalised sRNA coverage in MTHI1 RIP samples.

Differential enrichment was calculated as the ratio of the coverage at each nucleotide position in the MTHI1-HA sample to (that in the wild-type control sample +1). Blue curve: sRNAs mapping to the + strand, red curve sRNA mapping to the - strand. Most enriched genome positions are shown on the graph, as well as the position of the *atpH* and *atpI* 5'UTRs.

E) Coverage of immuno-precipitated RNA (normalized as RPM) over the *atpH* and *atpI* loci, schematically depicted at the bottom of the panel. The black bar in the *atpI* 5'UTR shows the position of the MTHI1 target (see below). Blue curve: MTHI1-RIP sample; red curve: WT-RIP sample (negative control). A zoom to the *atpI* 5'UTR is shown in supplemental Fig. S7B. Note the very different values of the y axes in the two graphs.

Figure 12: MTHI1 interacts with polysomes.

Distribution of MTHI1, Rps12 and RbcL proteins and of *atpH*, *psaB* and *rrnS* (16S rRNA) transcripts in wild-type cells along a sucrose gradient. For the gradient in the presence of $MgCl_2$, an overexposed blot immuno-decorated with the antibody against the MTHI1 protein is shown. T represents the total protein and RNA extracts, P the pellet fraction. The top panel shows the UV absorbance profile along the gradient, and the lower panel, the distribution of the same proteins and transcripts in samples treated with EDTA to dissociate the ribosomes. Note that the *atpH* mRNA, encoding a short polypeptide, is not heavily loaded with ribosomes and does not penetrate deep in the gradient. (red asterisk: cross-contamination.

Fig. 13: Validation of the putative MTHI1 targets.

A) Schematic map of the dH_M construct with a zoom to the region of the MTHI1 binding site, highlighted in a yellow box. Mutated nucleotides are written in red. The *atpH* transcription start site is indicated by a vertical arrow. The *atpH* promoter is underlined, and the position of the recycling *aadA* cassette is shown. The control construct ($atpH_{Ct}$) carries the selection cassette but no mutation in the *atpH* gene.

B) (left) Phototrophic growth of the dH_M strain (two independent transformants), assessed as in Fig. 2B. Growth of the wild type is shown as a control. (right) Accumulation of the *atpH* transcript in the wild-type transformed by the dH_{Ct} and dH_M constructs. Three independent transformants are shown for each genetic background. The *petD* transcript provides a loading control.

C) Schematic representation of the 5'*atpI* 5'UTR region in the mutant *dIf* series.

The red rectangle represents the *psaA* promoter region and the blue line shows the *psbJ-atpI* intergenic fragment inserted in the construct (larger than the *atpI* 5'UTR, to allow the processing of the chimeric transcript). The blue rectangle symbolises the processed *atpI* 5'UTR, with the target of MTHI1 shown in black. Relevant restriction sites *Bu* (*Bsu*36I), *Bm*, (*Bsm*I), *S* (*Sna*BI), *P* (*Pfl*MI), *Hc* (*Hinc*II, where the selection cassette was inserted) are indicated. Arrows above the map indicate the position of the deletions, while the lower insert shows the mutation introduced in the MTHI1 binding site (underlined) in the dI_Mf strain, with mutated nucleotides shown in red. A *Pvu*II site introduced as a RFLP marker is boxed.

D) Accumulation of the chimeric *petA* transcript in the {*aAdI*} strain transformed with the indicated *dIf* variants. Three independent transformants are shown for each construct. The *psaB* mRNA is shown as a loading control.

E) Accumulation of the chimeric cytochrome *f* in the same strains, and in the {*aAdI*}, *mthi1-1* {*dIf*} and Δ*petA* strains as controls. Immuno-detection of OEE2 provides a loading control.

Fig. 14: Modulation of MTHI1 action.

A) Secondary structure sequestering the MTHI binding site in the precursor RNA transcribed from the *atpA* promoter.

Lowest energy structure calculated at 25°C by RNA Folding Form (M-Fold: http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi; Zuker, 2003) for the region surrounding the *atpH* 5' end in the precursor transcript initiated at the *atpA* promoter. The MTHI1 binding site is yellow-shaded.

B) Secondary structure of the *atpH* 5'UTR in the *dHf* chimera, sequestering the initiation codon.

Lowest energy structure calculated at 25°C by M-Fold for transcribed *atpH* sequences inserted upstream of the *petA* gene in chimera *dHf*. The footprint of MDH1 is grey-shaded, while the *atpH* initiation codon is pink-shaded and the Shine-Dalgarno yellow-shaded.

C) Variations of MTHI1 transcript and protein accumulation over the circadian cycle.

Redrawn from the data in Strenkert et al, 2019. The dark period is indicated by the shaded area. Blue line shows the accumulation of the *MTHI1* transcript over time (expressed as Fragments Per Kilobase Million (FPKM), the red dots show the accumulation of the MTHI1protein at the indicated time points (expressed as Peakmaxintensity).

Supplemental Fig. S1: in support of the cloning strategy in Fig. 4A.

Supplemental Fig. S2: MTHI1 is required for the translation of 5'*atpI*-driven genes. (In support of Fig. 4)

Supplemental Fig. S3: MTHI1 targets the *atpH* 5'UTR. (In support of Fig. 5)

Supplemental Fig. S4: Cloning of the *MTHI1* gene. (In support of Fig. 7)

Supplemental Fig. S5: the *MTHI1* locus (In support of Fig. 7)

Supplemental Fig. S6: Conservation of the MTHI1 sequence among Chlorophyceae.

Supplemental Fig. S7: SRNAs coverage (normalised as RPM) over the *atpI* 5'UTR and along the inverted repeat (In support of Figs. 10B and 11E).

Supplemental Fig. S8: Conservation of the MTHI1 target in *atpH* and *atpI* 5'UTRs.

Supplemental Fig. S9: Emission spectrum of the white led used to grow C. reinhardtii. (In support of the M&M section).

Supplemental Fig. S10: Sequence of the recoded *MTHI1* gene. (In support of the M&M section).

Supplemental Fig. S11: Purification and quantification of the recombinant EcMTHI1. (In support of the M&M section).

Dataset DS1: Chlorophyta species analysed for conservation of the *atpH* and *atpI* 5'UTRs (xls file).

Supplemental Table ST1: oligonucleotides used in this work

AUTHOR CONTRIBUTIONS

Y.C. designed research. S.-I. O., M. C., D. J., R. K., M. R., S. E., D. D. and Y. C. performed research, M. C. contributed to bioinformatics analysis. F.-A. W., S.-I. O. and Y. C. analysed data. S.-I. O., F.-A.W., and Y.C. wrote the article.

1099 Table I: Transformation realised during that study

1100

A: chloroplast transformation		
plasmid	Recipient strain[1]	Transformed strain[2]
pK^{r}Δ*atpH*	wild type	Δ*atpH*[3,5]
	mthi1-1	Δ*H mthi1*[3,7]
	mthi1-2	Δ*H mthi1-2*[7]
pK^{r}Δ*atpI*	wild type	Δ*atpI*[3, 5]
	mthi1-2	Δ*I mthi1*[6]
	MTHI1-HA	Δ*I* MTHI1-HA[3]
	Δ*atpH*[4]	Δ*H/I*[3,6]
	Δ*H mthi1*[4]	Δ*H/I mthi1*[3,6]
pK^r*5'psaA-atpI*	wild type	*aAdI*[3,6]
	mthi1-2	*mthi1-2* {*aAI*}[1,7]
p*atpI*$_{St}$K^r	wild type	*atpI*$_{St}$[4,7]
p*atpI*$_{Ct}$K^r	wild type	*atpI*$_{Ct}$[4,7]
pK^r*dIf*	wild type	*dIf*[4,6]
	mthi1-1	*mthi1-1*{*dIf*}[4,6]
	Δ*atpH*[4]	{Δ*H*, *dIf*}[4,6]
	Δ*atpI*[4]	{Δ*I*, *dIf*}[4,6]
	Δ*H/I*[4]	*mthi1-1*{Δ*H/I*, *dIf*}[4,6]
	aAdI[4]	{*aAdI*, *dIf*}[4,7]
pK^r*dIf*Δ1,	*aAdI*[2]	*Δ1*[4,7]
pK^r*dIf*Δ2,		*Δ2*[4,7]
pK^r*dIf*Δ3		*Δ3*[4,7]
pK^r*dIf*Δ4		*Δ4*[4,7]
pK^r*dIf*ΔT		*ΔT*[4,7]
pK^r*dHf*	wild type	*dHf*[4,6]
	mthi1-	*mthi1-1* {*dHf*}[4,7]
	Δ*atpH*[4]	{Δ*H*, *dHf*}[4,7]
	Δ*atpI*[4]	{Δ*I*, *dHf*}[4,7]

A: chloroplast transformation		
plasmid	Recipient strain[1]	Transformed strain[2]
pK^r*dHf*	Δ*H/I*[4]	{Δ*H/I*, *dHf*}[4,7 8]
pWF*dHK*	wild type	*dHK*[4,7]
pWF*dIK*	wild type	*dIK*[4,7]
pG*atpH* K^r	wild type	pG*atpH*[3,4,7]
	mthi1-2	*mthi1* {pG*atpH*}[3,4,7]
	aAdI[4]	{*aAdI*, pG*atpH* }[3,,74]
	mthi1 {*aAdI*}[4]	*mthi1* {*aAdI*, pG*atpH* }[3,4,7]
p*atpH*$_{Ct}$	wild type	*atpH*$_{Ct}$[3,4,7]
B: nuclear transformation		
Plasmid [9]	Recipient strain [10]	Transformed strain[10]
gMTHI1-HA	*mthi1-1*	MTHI1-HA (g clones)
	mthi1-2	MTHI1-HA
cMTHI1-HA	*mthi1-1*	MTHI1-HA (c clones)
	mthi1-2	MTHI1-HA
gMTHI1-HA_ΔC	*mthi1-1*	ΔCg clones

[1] All recipient strains were spectinomycin sensitive. Transformed strains were selected for resistance to spectinomycin (100 µg·mL_1) under low light (5 µE·m_2·s_1) and subcloned in darkness on TAP-spectinomycin (500 µg.mL^{-1}) until they reached homoplasmy.

[2] Transformed strains are named by their genotype. By convention, the chloroplast genotype is indicated between accolades for strains containing more than one mutation and follows, when required, the nuclear genotype.

[3] These strains were initially selected for spectinomycin resistance due to the presence of the recycling spectinomycin resistance cassette (Kr). Once homoplasmic with respect to the ATP synthase mutation, they were grown on TAP medium for several generations to allow the spontaneous loss of the recycling cassette, according to Fischer et al (1996), but not that of the ATP synthase transgene.

[4] They, therefore, became spectinomycin sensitive again and could be used as a recipient strain in a new round of transformation experiments based on selection for spectinomycin resistance.

[5] Homoplasmy was deduced from the loss of phototrophic growth capacity.

[6] Homoplasmy was assessed by RNA gel blot experiments.

[7] Homoplasmy was assessed by RFLP of specific PCR products.

[8] Chimeras are named as follows: the first two letters indicate the origin of the 5'UTR, based on the nomenclature for chloroplast genes in *C. reinhardtii* (the first letter indicates the complex: A for PSI - *ps**a**-*, B for PSII – *ps**b**-*, C for **c**ytochrome b_6f, D for ATP synthase, R for RuBisCO, the second letter indicates the gene whose 5'UTR was borrowed: i.e. for *aA* for the 5'UTR of *ps**aA***). The next two letters indicate the CDS used in the chimera, based on the same nomenclature. For historical reasons, the *petA* CDS is designed as *f* for cytochrome ***f***, instead of *cA*, the *aadA* CDS is designed as K. Unless required, the 3'UTR is not mentioned and is usually that following the CDS, or the 3'*rbcL* UTR downstream of the *aadA* CDS.

Thus the full description of the *dHf* chimera would be *atpH* 5'UTR-*petA* CDS-*petA* 3'UTR, inserted at the *petA* locus, in replacement of the endogenous *petA* gene. The *aAdI* chimera comprises the *psaA* 5'UTR-*atpI* CDS-*atpI* 3UTR chimera, substituting the endogenous *atpI* gene at the *atpI* locus. A schematic map of all chimeras is also provided in the figures.

[9] Plasmid DNA was linearised before transformation upstream of the *MTHI1* gene by *Xba*.

[10] All recipient strains were nonphotosynthetic, and transformants were selected for photoautotrophy on minimum medium (Harris, 1989) under high light (100 $\mu E.m^{-2}.s^{-1}$).

Supplemental Material

Supplementary figures:

Supplemental Fig. S1: in support of the cloning strategy in Fig. 4A.

Supplemental Fig. S2: MTHI1 is required for the translation of 5'*atpI*-driven genes. (In support of Fig. 4)

Supplemental Fig. S3: MTHI1 targets the *atpH* 5'UTR. (In support of Fig. 5)

Supplemental Fig. S4: Cloning of the *MTHI1* gene. (In support of Fig. 7)

Supplemental Fig. S5: the *MTHI1* locus (In support of Fig. 7)

Supplemental Fig. S6: Conservation of the MTHI1 sequence among Chlorophyceae.

Supplemental Fig. S7: SRNAs coverage (normalised as RPM) over the *atpI* 5'UTR and along the inverted repeat (In support of Figs. 10B and 11E).

Supplemental Fig. S8: Conservation of the MTHI1 target in *atpH* and *atpI* 5'UTRs.

Supplemental Methods:

DNA constructs

Deletions

Deletion of the *atpH* gene.

Deletion of the *atpI* gene

Construction of reporter genes

5'*atpH*-driven reporter genes

pGatpH

5'*atpI*-driven reporter genes

Construction of reporter genes driven by modified *atpI* 5'UTRs

5'*psaA*-driven *atpI*

$AtpI_{St}$

MTHI1 constructs.

MTHI1 recoding.

MTHI1 overexpression, purification and immunisation.

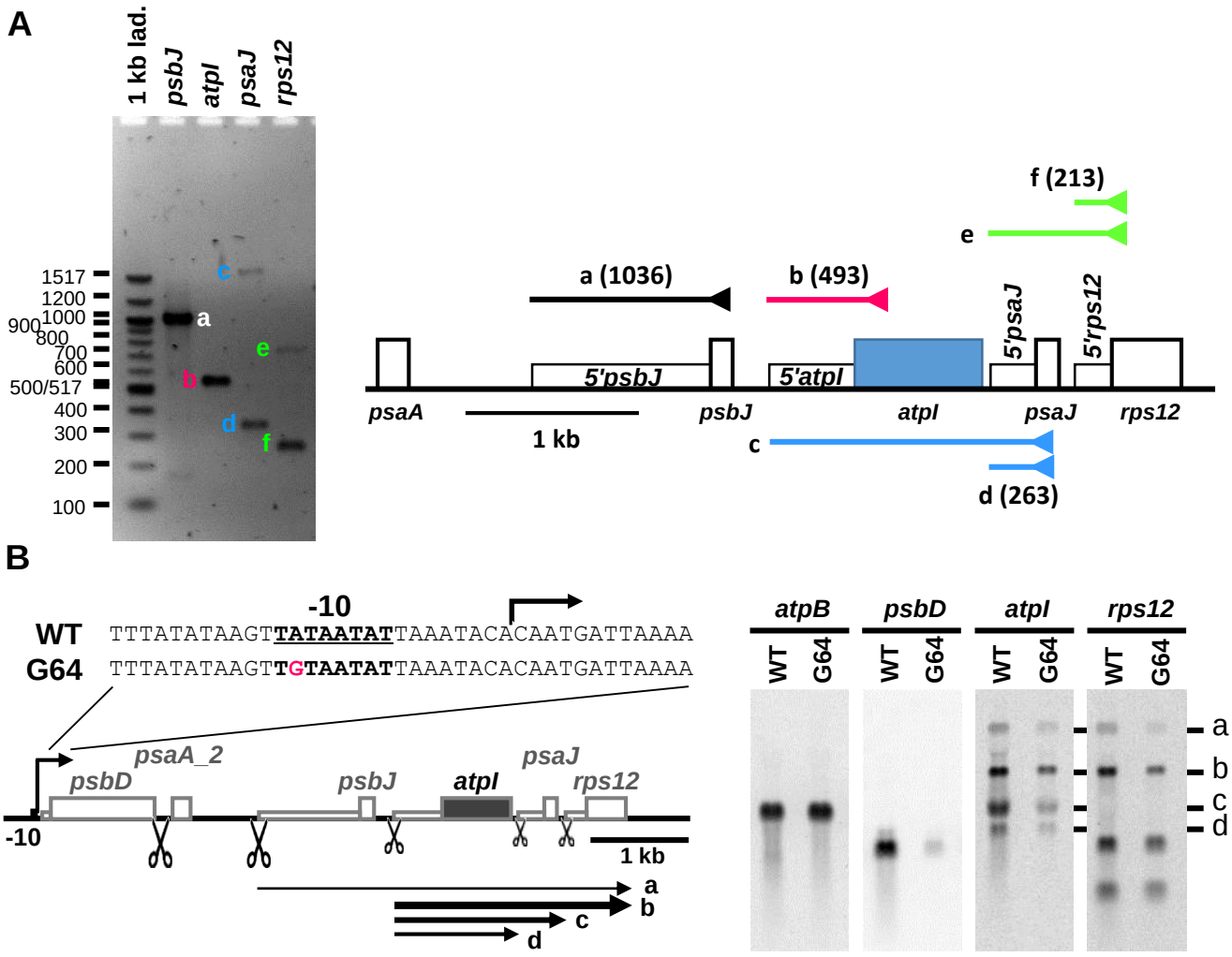
Supplemental Fig. S9: Emission spectrum of the white led used to grow C. reinhardtii. (In support of the M&M section).

Supplemental Fig. S10: Sequence of the recoded *MTHI1* gene. (In support of the M&M section).

Supplemental Fig. S11: Purification and quantification of the recombinant *Ec*MTHI1. (In support of the M&M section).

Supplemental Dataset DS1: Chlorophyta species analysed for conservation of the *atpH* and *atpI* 5'UTRs (xls file).

Supplemental Table ST1: oligonucleotides used in this work



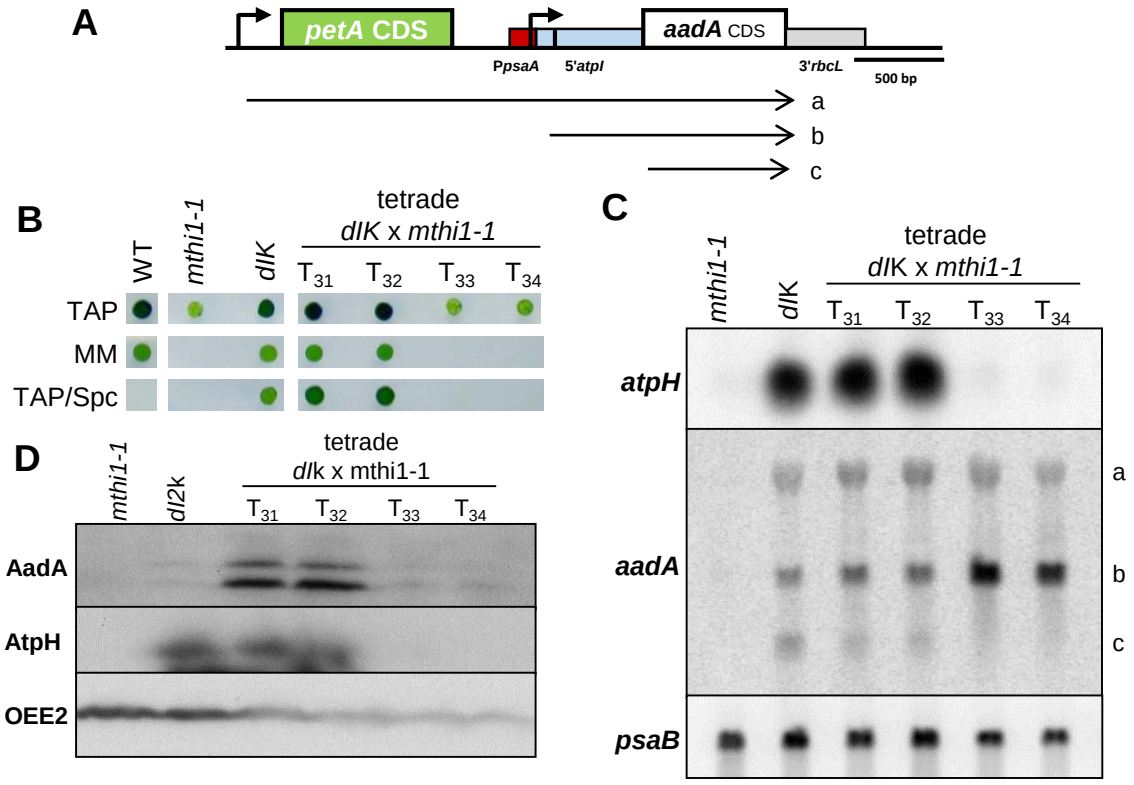
Supplemental Fig. S1: (in support of the cloning strategy in Fig. 4A)

A) Length of 5'UTRs within the *atpI* polycistronic unit.

(left) Sequence of the PCR amplicons from 5'RLM-RACE using the gene-specific primers (see Suppl. Table S2), schematically depicted by arrowheads, led to the size of 5'UTRs indicated in panel. (right) Schematic map of the *atpI* transcription unit. CDS are shown as thick rectangles, while 5'UTRs are represented by thin rectangles. The PCR amplicons detected in panel A are indicated with the length of the corresponding 5'UTRs indicated (in bp) between parentheses.

B) The *atpI* gene does not have a strong dedicated promoter

(left) Schematic representation of the *atpI* transcription unit, with a zoom to the *psbD* -10 promoter region mutated in the G64 mutant strain. As a result, *psbD* transcription is reduced 10 fold (Klinkert et al., 2005). (right) Accumulation of the *psbD*, *atpI*, and *rps12* transcripts in the G64 mutant strain. The mutation of the *psbD* promoter impacts the accumulation of the *psbD* mRNA but also that of the *atpI* mRNA, which is decreased 5 fold, compared to the wild type strain. This shows that this latter doesn't have a strong dedicated promoter. The *rps12* mRNA accumulates to 40% of the wild type level, probably because of RNA-stabilization effects compensate to some extent the reduction of transcription.

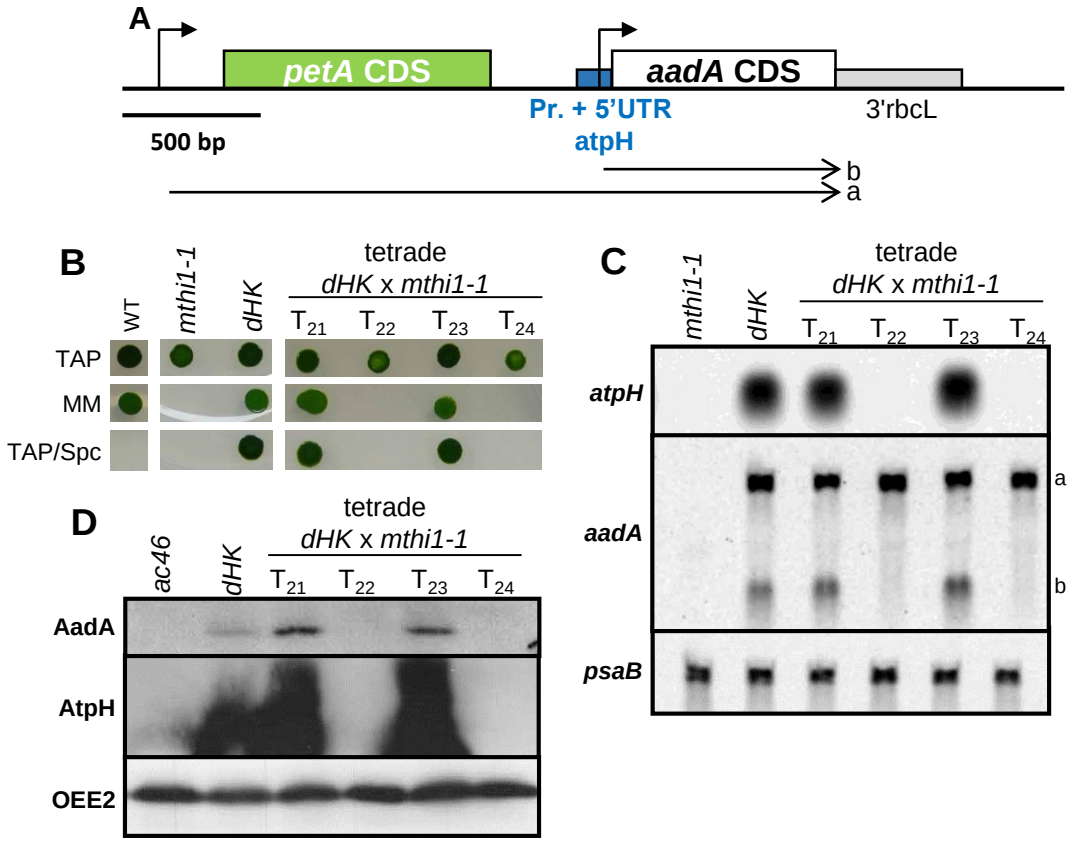


<u>Supplemental Fig. S2: MTHI1 is required for the translation of 5'*atpI*-driven genes</u> (in support of Fig. 4).

A) Map of the 5'*atpI-aadA-3'rbcl* cassette (*dIK*), inserted in a neutral site downstream of the *petA* gene and introduced by biolistic transformation in the chloroplast genome of the wild-type (mt^+) strain. Bent arrows symbolise promoters. The red rectangle represents the *psaA* promoter region inserted upstream of the *psbJ-atpI* intergenic region shown as a pale blue rectangle. The black line within this region indicates the processed *atpI* 5'end. Transcripts detected with the *aadA* specific probe are schematically depicted.

B) *dI*K transformants were recovered on TAP-spectinomycin plates and crossed with the *mthi1-1* (mt^-) strain. Thanks to the uniparental inheritance of the chloroplast genome from the *mt+* parent, all progeny inherited the chimeric gene, while the *mthi1* mutation showed mendelian segregation. Two progeny (members 1 and 2 in the representative tetrad shown in panel B-D) inherited a wild-type nuclear genome, grew phototrophically (B) and accumulated wild-type levels of the *atpH* mRNA (C) and of the AtpH subunit (D). They were resistant to spectinomycin (B) because they accumulated the monocistronic chimeric transcript (b in panel A) and expressed immuno-detectable amount of the AadA protein (D). *psaB* mRNA and OEE2 are shown as loading controls in panels C and D, respectively. The other two members of the tetrads (3 and 4) inherited the *mthi1* mutation, as shown by the lack of *atpH* transcript and by their failure to grow on minimum medium. These two progeny accumulated increased levels of the monocistronic form of the chimeric 5'*atpI-aadA* transcript. Indeed the translation of the *aadA* cassette, severely impaired in the *mthi1* background, leads to the cleavage of the chimeric transcript, shortly after the *aadA* initiation codon (transcript c, ~800 bp) (Y. Choquet, unpublish. res., see also Fig. 2 in Goldschmidt-Clermont, 1991). These two progeny, nevertheless, fail to synthesise significant amounts of the AadA protein (D) and were sensitive to spectinomycin (B). Together these results demonstrate that the 5'UTR of the *atpI* gene is sufficient to confer an MTHI1-dependent expression to 5'*atpI*-driven transcripts.

Ozawa et al, Suppl. Fig S2

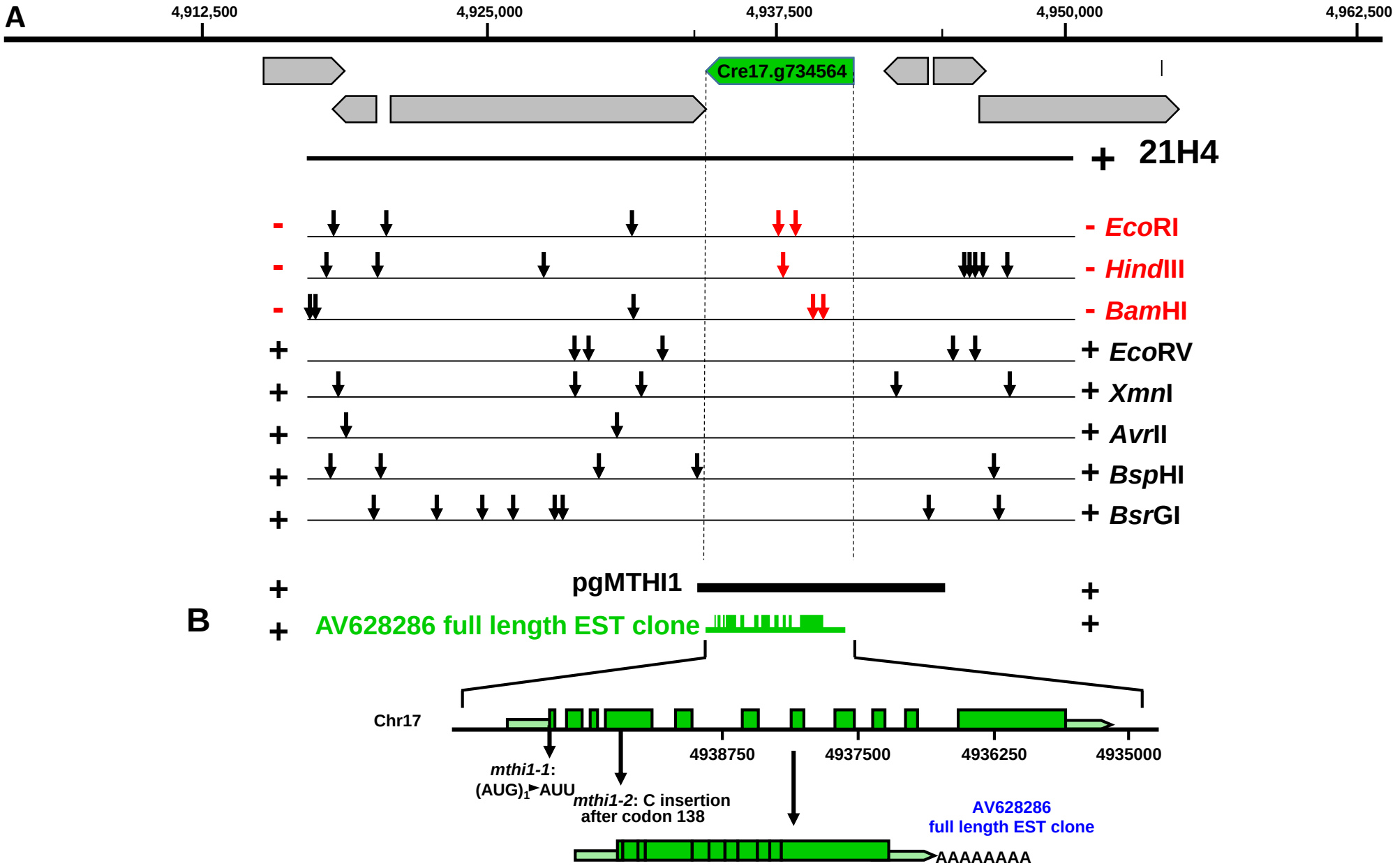


Supplemental Fig. S3: MTHI1 targets the *atpH* 5'UTR (in support of Fig. 5).

A) The 5'*atpH-aadA*-3'*rbcL* chimera (A) was inserted at a neutral site downstream of the *petA* gene and introduced by biolistic transformation in the chloroplast genome of the wild-type (*mt*$^+$) strain. Bent arrows symbolise promoters. The blue rectangle represents the *atpH* promoter and 5'untranslated regions, fused in frame to the *aadA* coding sequence. Transcripts detected with an *aadA* specific probe are schematically depicted.

B) *dH*K transformants were recovered on TAP-spectinomycin plates and crossed with the *mthi1-1* (*mt*$^-$) strain. Two members of the resulting tetrad (progeny 1 and 3 in the representative tetrad shown in panel B-D) inherited the wild-type *MTHI1* allele, grew photoautotrophically (B) and accumulated wild-type levels of the *atpH* mRNA (C) and of the AtpH subunit (D). They were resistant to spectinomycin because they accumulated the monocistronic chimeric dHK transcript (transcript (b) in panel A) and expressed immuno-detectable levels of the AadA protein (D). The other two members of the tetrad (2 and 4) inherited the *mthi1* mutation, as shown by the lack of *atpH* transcript (C) and by their failure to grow on minimum medium (B). These two progeny also failed to accumulate the monocistronic form of the chimeric 5'*atpH-aadA* transcript (C) and were sensitive to spectinomycin (B). This demonstrates that the 5'UTR of the *atpH* gene is sufficient to confer an MTHI1-dependent stability to a 5'*atpH*-driven transcript.

However, a dicistronic *petA-aadA* transcript (transcript (a) in panel A), most likely stabilised by the *petA* stabilisation factor MCA1 (Raynaud et al., 2007; Loiselay et al., 2008), accumulated to the same level in the four members of the tetrad. The *aadA* coding sequence present in this dicistronic transcript was nevertheless not expressed in the *mthi1* progeny as those were sensitive to spectinomycin and lacked accumulation of immuno-detectable AadA protein (D).



Ozawa et al, Suppl. Fig. S4

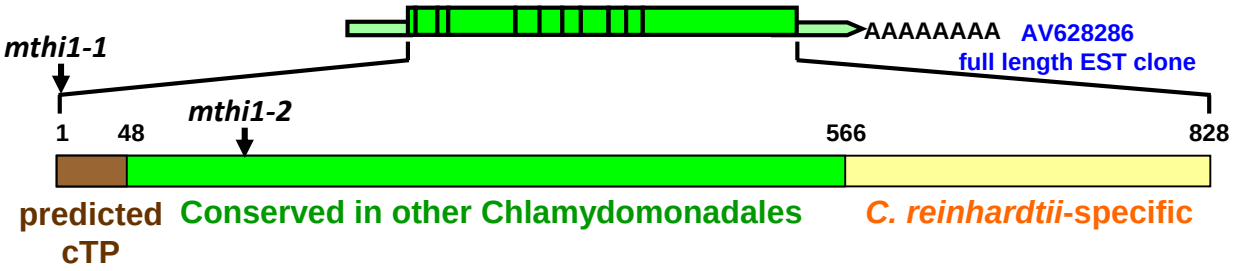
Supplemental Fig. S4: Cloning of the *MTHI1* gene.

A) The *MTHI1* gene was cloned as in (Raynaud et al, 2007) by complementation of a *mthi1-1, arg7*, *cw15* mutant strain with an indexed library of cosmids (Depège et al 2003) kindly provided by Pr J.-D. Rochaix. The cosmid vector backbone (Purton et al, 1994) included the *Arg7* gene (Debuchy et al, 1989). Phototrophic colonies, selected on minimal medium (120 $\mu E.m^{-2}.s^{-1}$), became visible after ~2 weeks. Selection for arginine prototrophy provided a control of transformation efficiency. One pool yielded ~10 photoautotrophic transformants. From this pool, cosmid 21H4 was isolated that complemented the *mthi1* phototrophic defect. Its ~33.5 kb genomic insert corresponded to nt 4917085-4950512 from chromosome 17. Complementation with cosmid digests with restriction enzymes listed on the right (**+**: complementation of the muitation; **-**: absence of complementation) further restricted the region required for complementation to that corresponding to gene model Cre17.g734564, as only the *Eco*RI, *Hind*III, *Bam*HI restriction enzymes (written in red) cutting within this gene model prevented the complementation of the mutant phenotype. This chromosomic localisation, however, is erroneous. Indeed the *ac46* mutant (*mthi1-1*) has been previously mapped to the complementation group XVI/XVII, which was later shown to correspond to chromosome 15 (Dutcher et al, 1991, Kathir et al, 2003). Crosses confirmed that the *mthi1-1* mutation was linked to the CytC1 molecular marker on chromosome 15. It is of note that the *MTHI1* gene was localised on chromosome 15 in the version 4.0 of the Chlamydomonas genome and was moved to chromosome 17 in version 5.5.

We constructed from cosmid 21H4 the pgMTHI1 plasmid with a 10 679 bp genomic insert (chromosome17:4933979-4944653), encompassing Cre17.g734164 (see Material and methods) capable to restore the phototrophic growth of both *mthi1-1* and *mthi1-2* mutant strains.

B) One EST clone (AV629671) of this gene model was obtained from Kazusa DNA Research Institute and sequenced using appropriate primers. It contained the full-length coding sequence for MTHI1, as an in-frame stop codon is located 6 nucleotides upstream of the initiation codon. Sequence comparison with the genomic scaffold showed that *MTHI1* is composed of 11 exons. A polyA tail was found 424 bp downstream of the stop codon and 15 nt downstream of the TGTAA poly-adenylation consensus signal (Silflow, 1998)

A



B

361 CATCAGCACTCGCGCGGC-AGCGACGTCGACCGCGCCCGCGCCGCGGCGGAGCTGCGGCCAGC **WT**
H Q H S R G S D V D R A R A A A E L R P A
361 CATCAGCACTCGCGCGGC**C**AGCGACGTCGACCGCGCCCGCGCCGCGGCGGAGCTGCGGCCAGC ***mthi1-2***
H Q H S R G **Q R R R P R P R R G G A A A S**

GGTGGAGGCGCTCACCAAACGCATGCACCAGCTCATAGGCAACTATGACTCCTGGGACACGACGCTC **WT**
V E A L T K R M H Q L I G N Y D S W D T T L S
GGGTGAGGCGCTCACCAAACGCATGCACCAGCTCATAGGCAACTA**TGA**CTCCTGGGACACGACGCTC ***mthi1-2***
G G G A H Q T H A P A H R Q L -ST

C

>*Cre*MTHI1

MKSLAAHNNTTTTVLYRSSSLLLPVPCTGRGGPTSTSGRNCSAPVSARAGARRGVH**NGAA**
AVHPGGLDLLDTAESSAEQLTPRRLLNRRIKSCLSPAQLAGLVLSEVGNFDQQNASHALS
RLAKMYRGRRRNSHQHSRGSDVDRARAAAELRPAVEALTKRMHQLIGNYDSWDTTLSLWA
YAQLDHYDEGALRALCDAAVEVAPIFKPVDCANAVVAFAHLDYVHPELLRQIVQTVLDTL
DDYAPGEVCQVLWGFARLGVHPGPAFLAEVVDAVQWRLQGYGTQELGMVLWALVRLGYKP
GPRFLRDVESVLLARLPHMAPGDIAITVWSFARLRYKAVRFLDEVPAAVGPQLHKCRSSE
LCSLISGFATAHHYHKSLLDAVADVLLSRLDGLSHHEVATALWTFGTFRHRPAHPDFAKQ
VAAALYARMRSFSPQGLAMVVKALAQLQWRSEPLMEQLIAAAEAKLNAFKPLELSQLLWG
LTALQCRDLHIYYAVVRRCIAILKDPAHPHYRTMTHHRVVNSVLGSCQQLGYVPWTLIDF
AESKGIRVRQPDILSSRDEDDEGVPYSHQQQQHADAVEGCGHDRAEQPWGASTGSSSSSS
RRHQRCAEEEALWAEAERAHSQQVAAGNSSSDAAMASAPDAVVLLEQGLIPHVSSSAAAD
ASHEAAAVHAAAQGEYRALQQPKPQPLAMLTERGSRHATGMIVLAGAAAVVAGEGVSAGD
AEQQSAMSAPNVAQLQESAPAAAALDGSNSGSNGAKVLSPRPRLGSARRGGPVVAGDASP
KGASAHVAVPVDSAAPSGARARALFSDPRRDSPYNVGMVAATPLTFQR*

Supplemental Fig. S5: the *MTHI1* locus (in support of Fig. 6)

A) Schematic representation of the MTHI1 protein showing its three major domains, as well as the position of the two mutations.

B) The mutation in strain *mthi1-2* results in premature translation abortion.

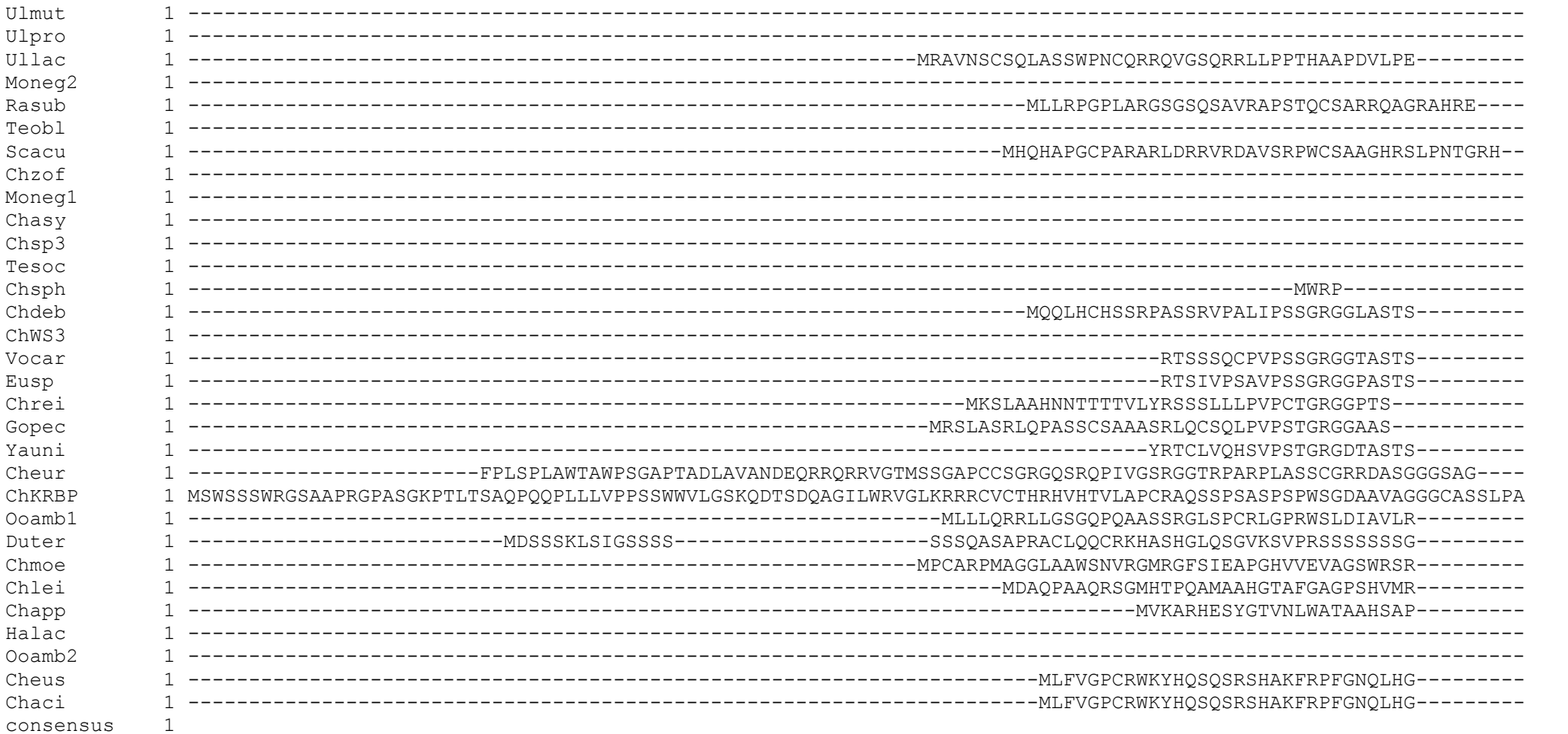
Partial sequence of the *MTHI1* cDNA (with translation) in the wild-type and mthi1-2 strains, with nucleotides numbered from the first A of the initiation codon. The inserted nucleotide is written in red.

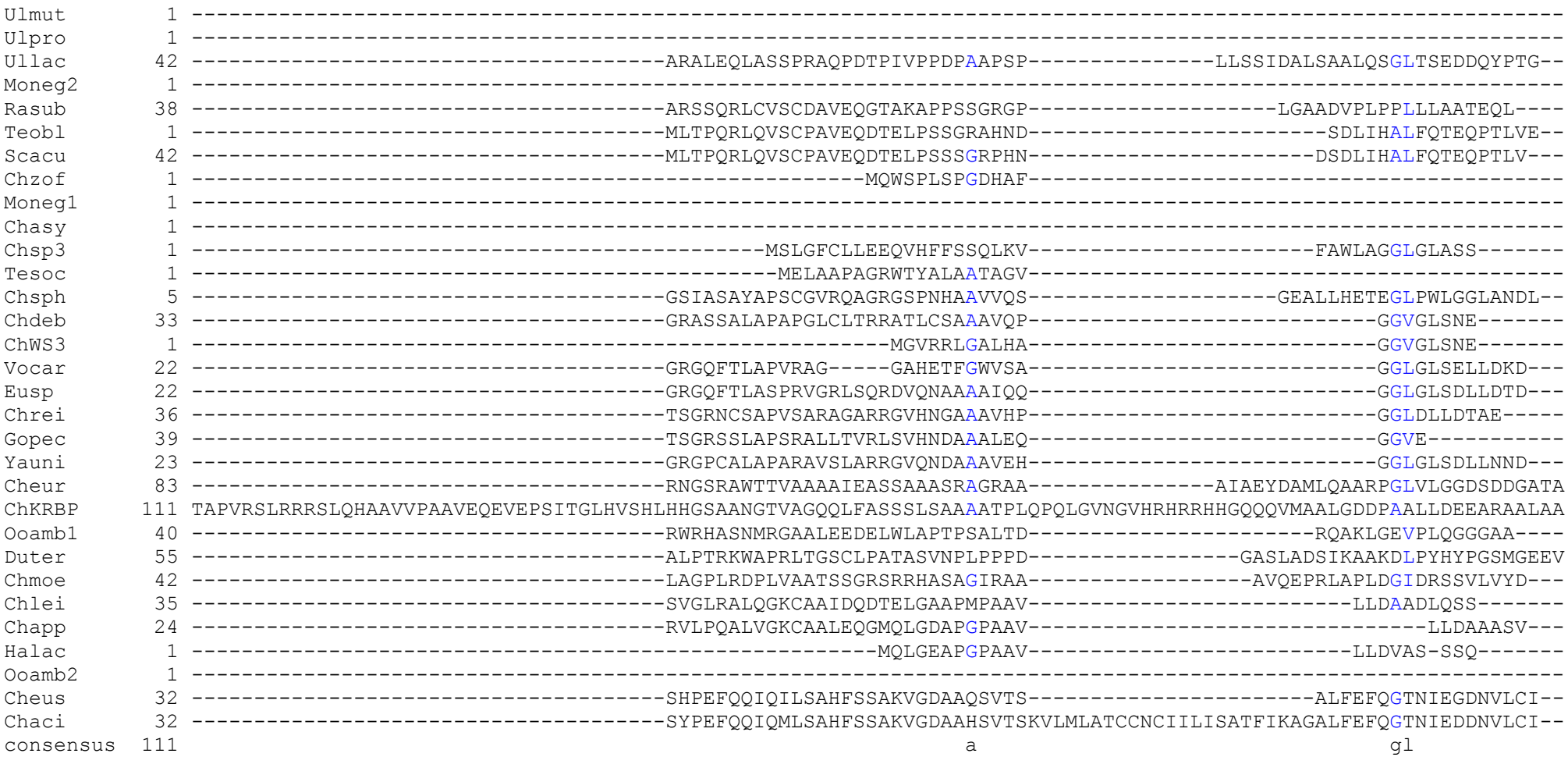
C) Sequence of the MTHI1 protein.

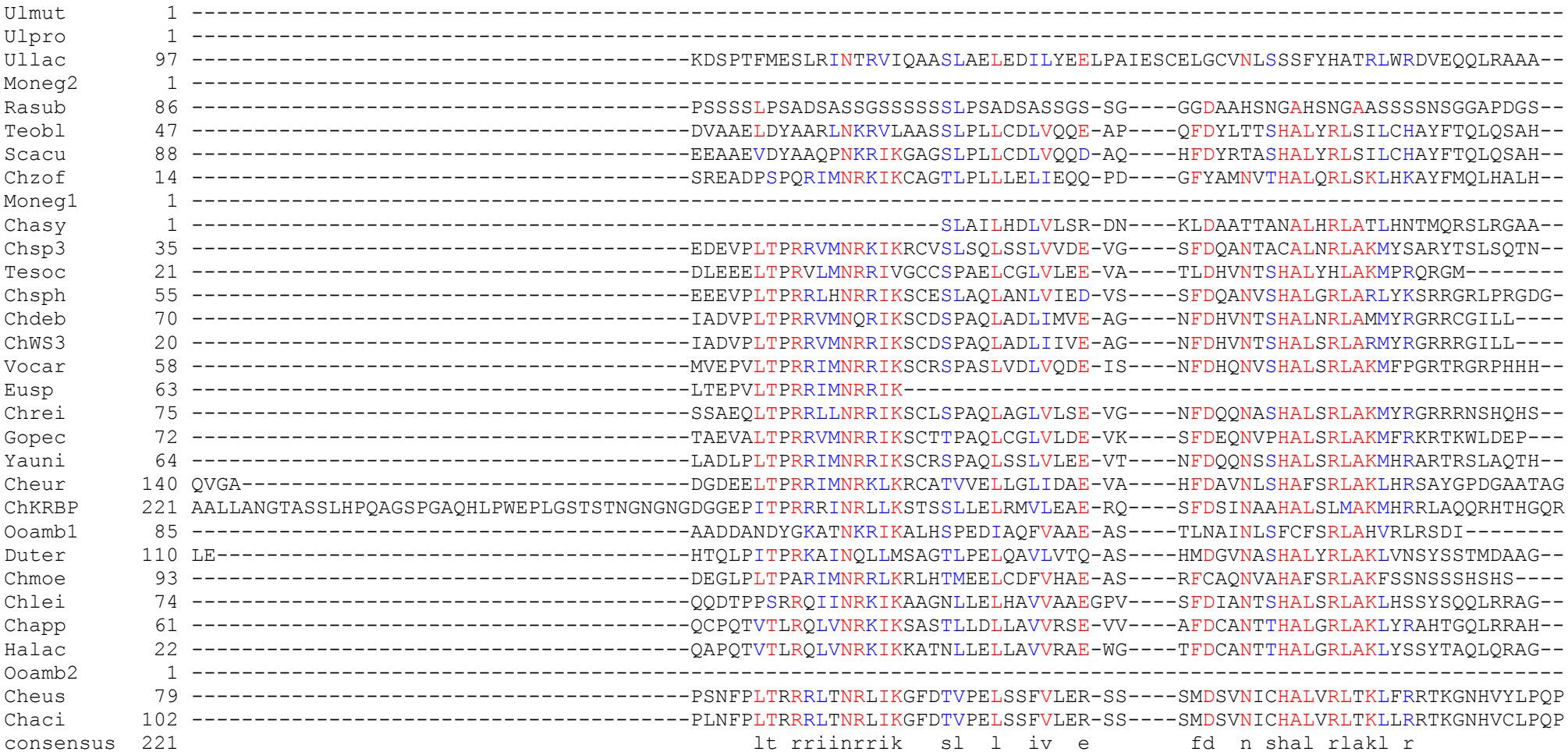
The predicted chloroplast transit peptide is written in blue, the residue encoded by the mutated codon is written in red. The OPR repeats listed in Fig. 6B are alternatively underlined and boxed. The *C. reinhardtii*-specific sequence is written in grey, with stretches of identical residues shaded.

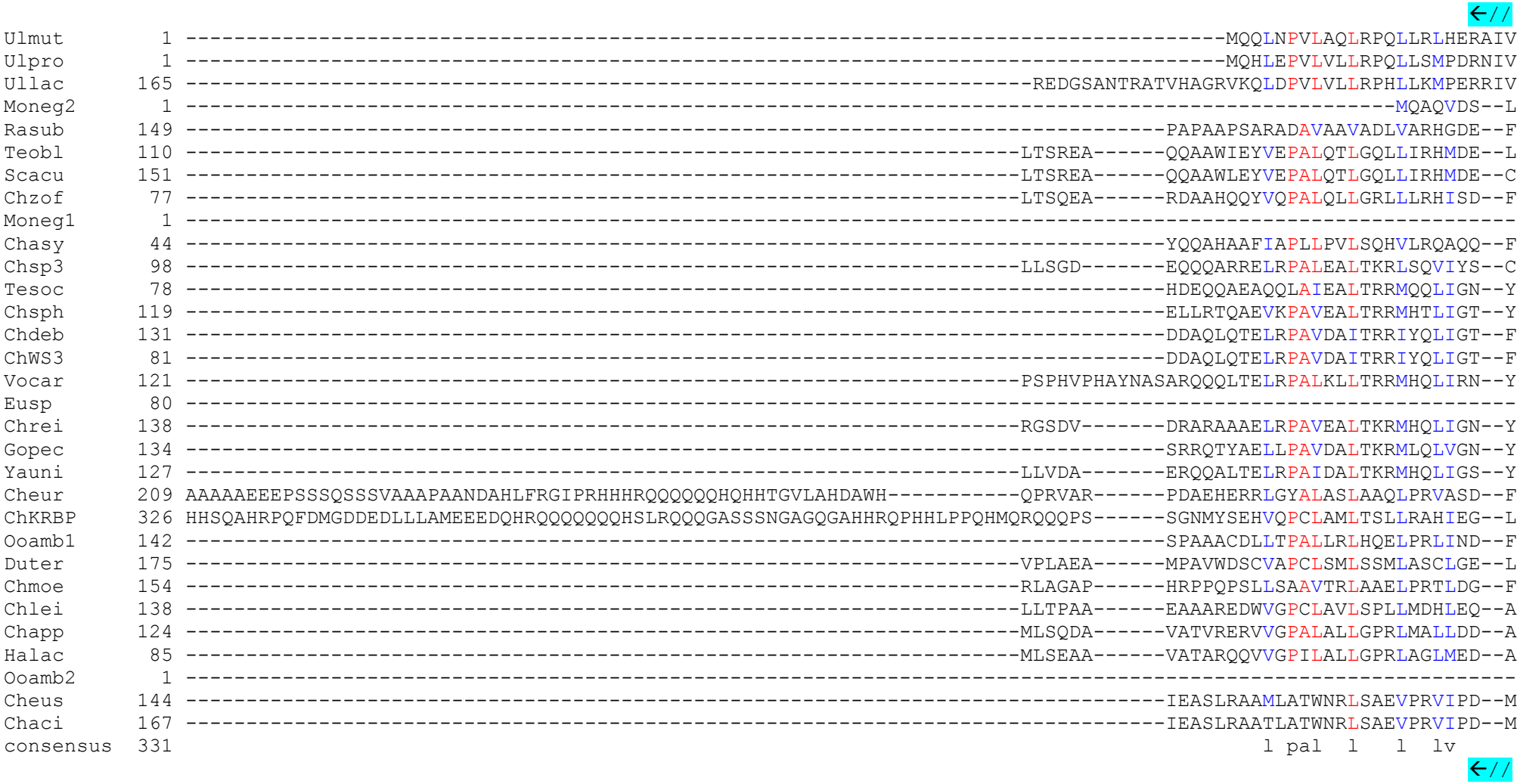
Ozawa et al, Suppl. Fig. S5

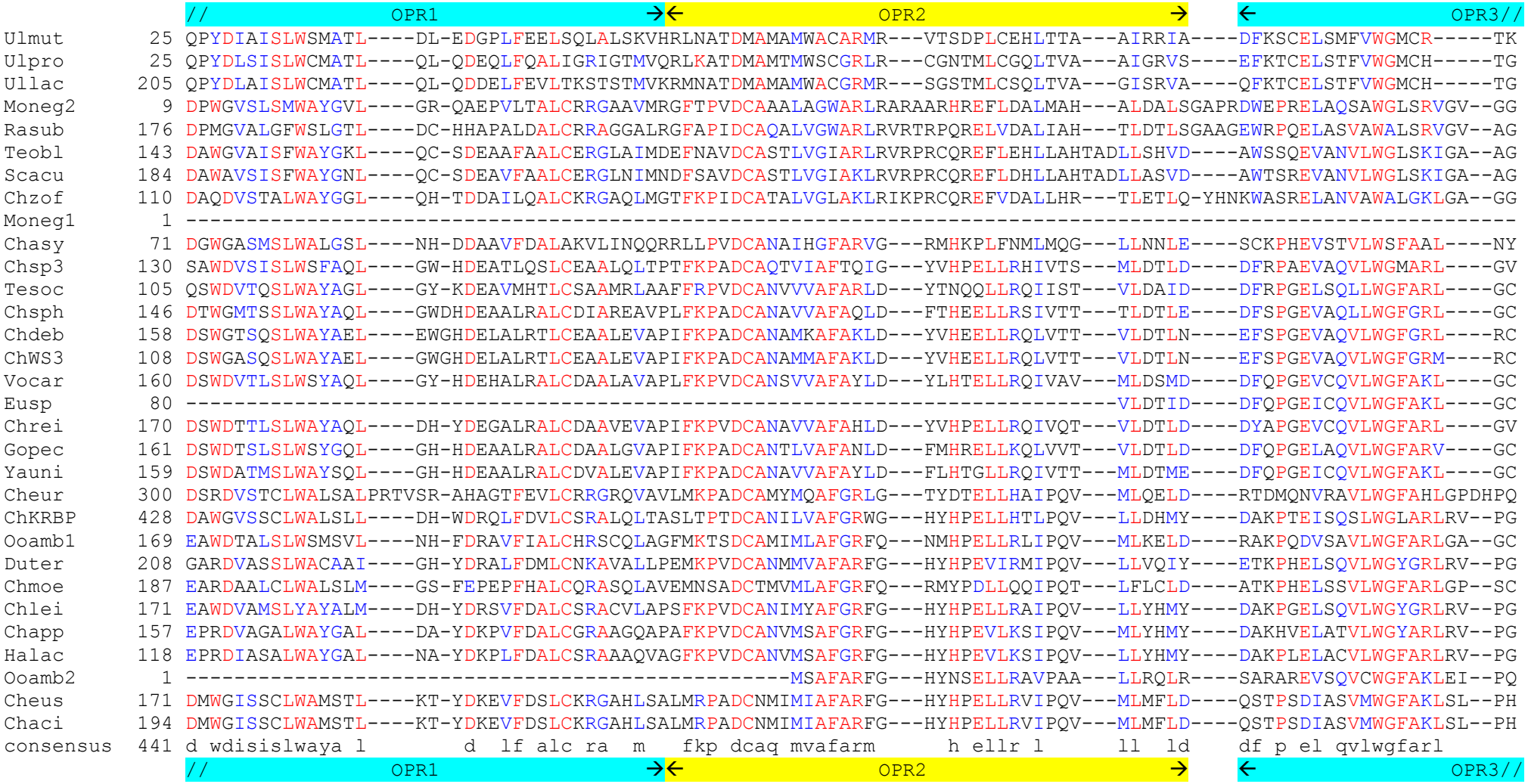
Ozawa et al, Suppl. Fig S6

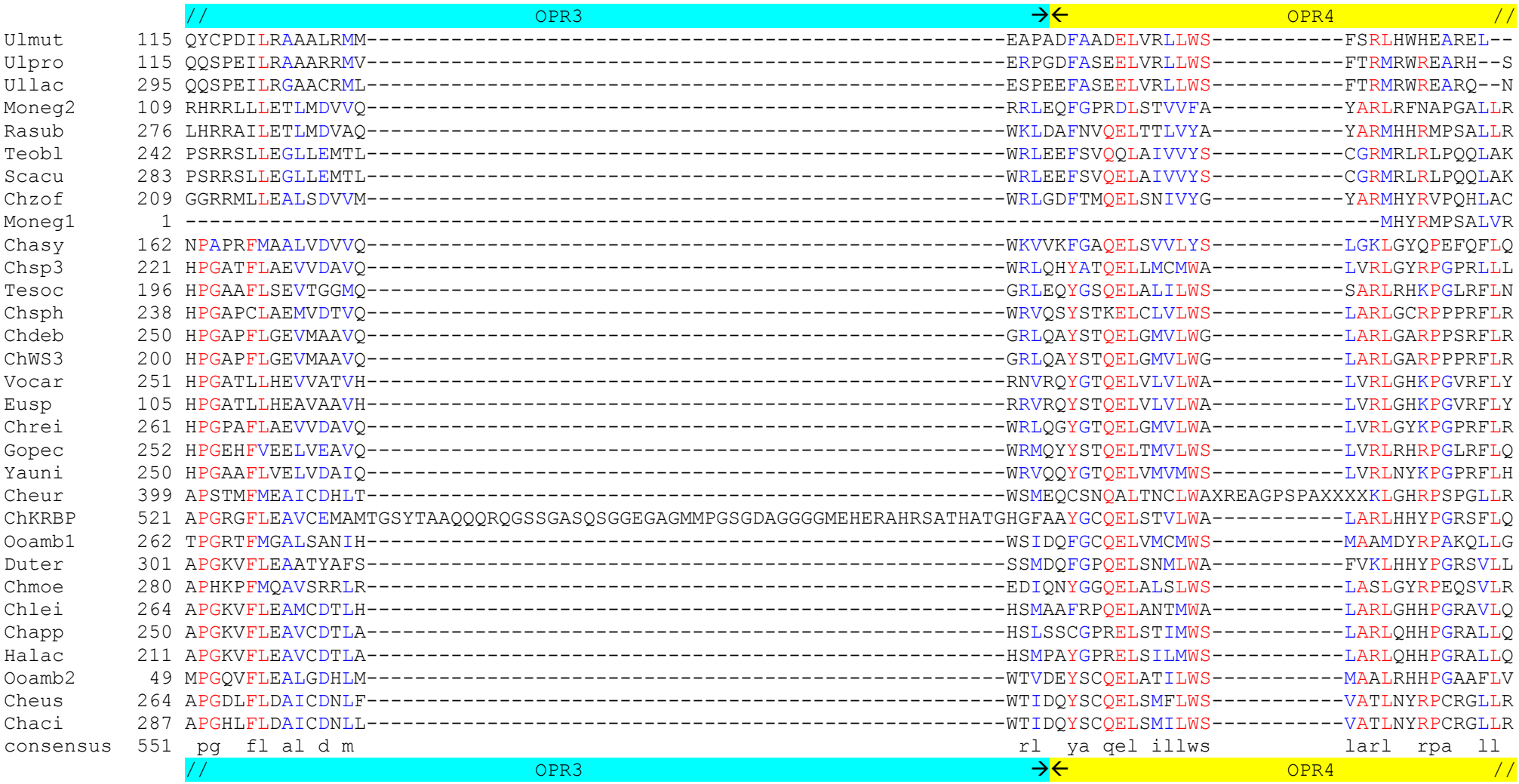



Ulmut 1
Ulpro 1
Ullac 42 ARALEQLASSPRAQPDTPIVPPDPAAPSP LLSSIDALSAALQSGLTSEDDQYPTG
Moneg2 1
Rasub 38 ARSSQRLCVSCDAVEQGTAKAPPSSGRGP LGAADVPLPPLLLAATEQL
Teobl 1 MLTPQRLQVSCPAVEQDTELPSSGRAHND SDLIHALFQTEQPTLVE
Scacu 42 MLTPQRLQVSCPAVEQDTELPSSSGRPHN DSDLIHALFQTEQPTLV
Chzof 1 MQWSPLSPGDHAF
Moneg1 1
Chasy 1
Chsp3 1 MSLGFCLLEEQVHFFSSQLKV FAWLAGGLGLASS
Tesoc 1 MELAAPAGRWTYALAATAGV
Chsph 5 GSIASAYAPSCGVRQAGRGSPNHAAVVQS GEALLHETEGLPWLGGLANDL
Chdeb 33 GRASSALAPAPGLCLTRRATLCSAAAVQP GGVGLSNE
ChWS3 1 MGVRRLGALHA GGVGLSNE
Vocar 22 GRGQFTLAPVRAG GAHETFGWVSA GGLGLSELLDKD
Eusp 22 GRGQFTLASPRVGRLSQRDVQNAAAAIQQ GGLGLSDLLDTD
Chrei 36 TSGRNCSAPVSARAGARRGVHNGAAAVHP GGLDLLDTAE
Gopec 39 TSGRSSLAPSRALLTVRLSVHNDAAALEQ GGVE
Yauni 23 GRGPCALAPARAVSLARRGVQNDAAAVEH GGLGLSDLLNND
Cheur 83 RNGSRAWTTVAAAAIEASSAAASRAGRAA AIAEYDAMLQAARPGLVLGGDSDDGATA
ChKRBP 111 TAPVRSLRRRSLQHAAVVPAAVEQEVEPSITGLHVSHLHHGSAANGTVAGQQLFASSSLSAAAATPLQPQLGVNGVHRHRRHHGQQQVMAALGDDPAALLDEEARAALAA
Ooamb1 40 RWRHASNMRGAALEEDELWLAPTPSALTD RQAKLGEVPLQGGGAA
Duter 55 ALPTRKWAPRLTGSCLPATASVNPLPPPD GASLADSIKAAKDLPYHYPGSMGEEV
Chmoe 42 LAGPLRDPLVAATSSGRSRRHASAGIRAA AVQEPRLAPLDGIDRSSVLVYD
Chlei 35 SVGLRALQGKCAAIDQDTELGAAPMPAAV LLDAADLQSS
Chapp 24 RVLPQALVGKCAALEQGMQLGDAPGPAAV LLDAAASV
Halac 1 MQLGEAPGPAAV LLDVAS-SSQ
Ooamb2 1
Cheus 32 SHPEFQQIQILSAHFSSAKVGDAAQSVTS ALFEFQGTNIEGDNVLCI
Chaci 32 SYPEFQQIQMLSAHFSSAKVGDAAHSVTSKVLMLATCCNCIILISATFIKAGALFEFQGTNIEDDNVLCI
consensus 111 a gl


Ulmut 1
Ulpro 1
Ullac 97 KDSPTFMESLRINTRVIQAASLAELEDILYEELPAIESCELGCVNLSSSFYHATRLWRDVEQQLRAAA--
Moneg2 1
Rasub 86 PSSSSLPSADSASSGSSSSSSLPSADSASSGS-SG----GGDAAHSNGAHSNGAASSSSNSGGAPDGS--
Teobl 47 DVAAELDYAARLNKRVLAASSLPLLCDLVQQE-AP----QFDYLTTSHALYRLSILCHAYFTQLQSAH--
Scacu 88 EEAAEVDYAAQPNKRIKGAGSLPLLCDLVQQD-AQ----HFDYRTASHALYRLSILCHAYFTQLQSAH--
Chzof 14 SREADPSPQRIMNRKIKCAGTLPLLLELIEQQ-PD----GFYAMNVTHALQRLSKLHKAYFMQLHALH--
Moneg1 1
Chasy 1 SLAILHDLVLSR-DN----KLDAATTANALHRLATLHNTMQRSLRGAA--
Chsp3 35 EDEVPLTPRRVMNRKIKRCVSLSQLSSLVVDE-VG----SFDQANTACALNRLAKMYSARYTSLSQTN--
Tesoc 21 DLEEELTPRVLMNRRIVGCCSPAELCGLVLEE-VA----TLDHVNTSHALYHLAKMPRQRGM--------
Chsph 55 EEEVPLTPRRLHNRRIKSCESLAQLANLVIED-VS----SFDQANVSHALGRLARLYKSRRGRLPRGDG-
Chdeb 70 IADVPLTPRRVMNQRIKSCDSPAQLADLIMVE-AG----NFDHVNTSHALNRLAMMYRGRRCGILL----
ChWS3 20 IADVPLTPRRVMNRRIKSCDSPAQLADLIIVE-AG----NFDHVNTSHALSRLARMYRGRRRGILL----
Vocar 58 MVEPVLTPRRIMNRRIKSCRSPASLVDLVQDE-IS----NFDHQNVSHALSRLAKMFPGRTRGRPHHH--
Eusp 63 LTEPVLTPRRIMNRRIK----------
Chrei 75 SSAEQLTPRRLLNRRIKSCLSPAQLAGLVLSE-VG----NFDQQNASHALSRLAKMYRGRRRNSHQHS--
Gopec 72 TAEVALTPRRVMNRRIKSCTTPAQLCGLVLDE-VK----SFDEQNVPHALSRLAKMFRKRTKWLDEP---
Yauni 64 LADLPLTPRRIMNRRIKSCRSPAQLSSLVLEE-VT----NFDQQNSSHALSRLAKMHRARTRSLAQTH--
Cheur 140 QVGA----------DGDEELTPRRIMNRKLKRCATVVELLGLIDAE-VA----HFDAVNLSHAFSRLAKLHRSAYGPDGAATAG
ChKRBP 221 AALLANGTASSLHPQAGSPGAQHLPWEPLGSTSTNGNGNGDGGEPITPRRRINRLLKSTSSLLELRMVLEAE-RQ----SFDSINAAHALSLMAKMHRRLAQQRHTHGQR
Ooamb1 85 AADDANDYGKATNKRIKALHSPEDIAQFVAAE-AS----TLNAINLSFCFSRLAHVRLRSDI--------
Duter 110 LE----------HTQLPITPRKAINQLLMSAGTLPELQAVLVTQ-AS----HMDGVNASHALYRLAKLVNSYSSTMDAAG--
Chmoe 93 DEGLPLTPARIMNRRLKRLHTMEELCDFVHAE-AS----RFCAQNVAHAFSRLAKFSSNSSSHSHS----
Chlei 74 QQDTPPSRRQIINRKIKAAGNLLELHAVVAAEGPV----SFDIANTSHALSRLAKLHSSYSQQLRRAG--
Chapp 61 QCPQTVTLRQLVNRKIKSASTLLDLLAVVRSE-VV----AFDCANTTHALGRLAKLYRAHTGQLRRAH--
Halac 22 QAPQTVTLRQLVNRKIKKATNLLELLAVVRAE-WG----TFDCANTTHALGRLAKLYSSYTAQLQRAG--
Ooamb2 1
Cheus 79 PSNFPLTRRRLTNRLIKGFDTVPELSSFVLER-SS----SMDSVNICHALVRLTKLFRRTKGNHVYLPQP
Chaci 102 PLNFPLTRRRLTNRLIKGFDTVPELSSFVLER-SS----SMDSVNICHALVRLTKLLRRTKGNHVCLPQP
consensus 221 lt rriinrrik sl l iv e fd n shal rlakl r


←//
Ulmut 1 ----------MQQLNPVLAQLRPQLLRLHERAIV
Ulpro 1 ----------MQHLEPVLVLLRPQLLSMPDRNIV
Ullac 165 ----------REDGSANTRATVHAGRVKQLDPVLVLLRPHLLKMPERRIV
Moneg2 1 ----------MQAQVDS--L
Rasub 149 ----------PAPAAPSARADAVAAVADLVARHGDE--F
Teobl 110 ----------LTSREA------QQAAWIEYVEPALQTLGQLLIRHMDE--L
Scacu 151 ----------LTSREA------QQAAWLEYVEPALQTLGQLLIRHMDE--C
Chzof 77 ----------LTSQEA------RDAAHQQYVQPALQLLGRLLLRHISD--F
Moneg1 1 ----------
Chasy 44 ----------YQQAHAAFIAPLLPVLSQHVLRQAQQ--F
Chsp3 98 ----------LLSGD-------EQQQARRELRPALEALTKRLSQVIYS--C
Tesoc 78 ----------HDEQQAEAQQLAIEALTRRMQQLIGN--Y
Chsph 119 ----------ELLRTQAEVKPAVEALTRRMHTLIGT--Y
Chdeb 131 ----------DDAQLQTELRPAVDAITRRIYQLIGT--F
ChWS3 81 ----------DDAQLQTELRPAVDAITRRIYQLIGT--F
Vocar 121 ----------PSPHVPHAYNASARQQQLTELRPALKLLTRRMHQLIRN--Y
Eusp 80 ----------
Chrei 138 ----------RGSDV-------DRARAAAELRPAVEALTKRMHQLIGN--Y
Gopec 134 ----------SRRQTYAELLPAVDALTKRMLQLVGN--Y
Yauni 127 ----------LLVDA-------ERQQALTELRPAIDALTKRMHQLIGS--Y
Cheur 209 AAAAAEEEPSSSQSSSVAAAPAANDAHLFRGIPRHHHRQQQQQQHQHHTGVLAHDAWH----------QPRVAR------PDAEHERRLGYALASLAAQLPRVASD--F
ChKRBP 326 HHSQAHRPQFDMGDDEDLLLAMEEEDQHRQQQQQQQHSLRQQQGASSSNGAGQGAHHRQPHHLPPQHMQRQQQPS------SGNMYSEHVQPCLAMLTSLLRAHIEG--L
Ooamb1 142 ----------SPAAACDLLTPALLRLHQELPRLIND--F
Duter 175 ----------VPLAEA------MPAVWDSCVAPCLSMLSSMLASCLGE--L
Chmoe 154 ----------RLAGAP------HRPPQPSLLSAAVTRLAAELPRTLDG--F
Chlei 138 ----------LLTPAA------EAAAREDWVGPCLAVLSPLLMDHLEQ--A
Chapp 124 ----------MLSQDA------VATVRERVVGPALALLGPRLMALLDD--A
Halac 85 ----------MLSEAA------VATARQQVVGPILALLGPRLAGLMED--A
Ooamb2 1 ----------
Cheus 144 ----------IEASLRAAMLATWNRLSAEVPRVIPD--M
Chaci 167 ----------IEASLRAATLATWNRLSAEVPRVIPD--M
consensus 331 l pal l l lv
←//


// OPR1 → ← OPR2 → ← OPR3//
Ulmut 25 QPYDIAISLWSMATL----DL-EDGPLFEELSQLALSKVHRLNATDMAMAMWACARMR---VTSDPLCEHLTTA---AIRRIA----DFKSCELSMFVWGMCR-----TK
Ulpro 25 QPYDLSISLWCMATL----QL-QDEQLFQALIGRIGTMVQRLKATDMAMTMWSCGRLR---CGNTMLCGQLTVA---AIGRVS----EFKTCELSTFVWGMCH-----TG
Ullac 205 QPYDLAISLWCMATL----QL-QDDELFEVLTKSTSTMVKRMNATDMAMAMWACGRMR---SGSTMLCSQLTVA---GISRVA----QFKTCELSTFVWGMCH-----TG
Moneg2 9 DPWGVSLSMWAYGVL----GR-QAEPVLTALCRRGAAVMRGFTPVDCAAALAGWARLRARAARHREFLDALMAH---ALDALSGAPRDWEPRELAQSAWGLSRVGV--GG
Rasub 176 DPMGVALGFWSLGTL----DC-HHAPALDALCRRAGGALRGFAPIDCAQALVGWARLRVRTRPQRELVDALIAH---TLDTLSGAAGEWRPQELASVAWALSRVGV--AG
Teobl 143 DAWGVAISFWAYGKL----QC-SDEAAFAALCERGLAIMDEFNAVDCASTLVGIARLRVRPRCQREFLEHLLAHTADLLSHVD----AWSSQEVANVLWGLSKIGA--AG
Scacu 184 DAWAVSISFWAYGNL----QC-SDEAVFAALCERGLNIMNDFSAVDCASTLVGIAKLRVRPRCQREFLDHLLAHTADLLASVD----AWTSREVANVLWGLSKIGA--AG
Chzof 110 DAQDVSTALWAYGGL----QH-TDDAILQALCKRGAQLMGTFKPIDCATALVGLAKLRIKPRCQREFVDALLHR---TLETLQ-YHNKWASRELANVAWALGKLGA--GG
Moneg1 1 ----------
Chasy 71 DGWGASMSLWALGSL----NH-DDAAVFDALAKVLINQQRRLLPVDCANAIHGFARVG---RMHKPLFNMLMQG---LLNNLE----SCKPHEVSTVLWSFAAL----NY
Chsp3 130 SAWDVSISLWSFAQL----GW-HDEATLQSLCEAALQLTPTFKPADCAQTVIAFTQIG---YVHPELLRHIVTS---MLDTLD----DFRPAEVAQVLWGMARL----GV
Tesoc 105 QSWDVTQSLWAYAGL----GY-KDEAVMHTLCSAAMRLAAFFRPVDCANVVVAFARLD---YTNQQLLRQIIST---VLDAID----DFRPGELSQLLWGFARL----GC
Chsph 146 DTWGMTSSLWAYAQL----GWDHDEAALRALCDIAREAVPLFKPADCANAVVAFAQLD---FTHEELLRSIVTT---TLDTLE----DFSPGEVAQLLWGFGRL----GC
Chdeb 158 DSWGTSQSLWAYAEL----EWGHDELALRTLCEAALEVAPIFKPADCANAMKAFAKLD---YVHEELLRQLVTT---VLDTLN----EFSPGEVAQVLWGFGRL----RC
ChWS3 108 DSWGASQSLWAYAEL----GWGHDELALRTLCEAALEVAPIFKPADCANAMMAFAKLD---YVHEELLRQLVTT---VLDTLN----EFSPGEVAQVLWGFGRM----RC
Vocar 160 DSWDVTLSLWSYAQL----GY-HDEHALRALCDAALAVAPLFKPVDCANSVVAFAYLD---YLHTELLRQIVAV---MLDSMD----DFQPGEVCQVLWGFAKL----GC
Eusp 80 ----------VLDTID----DFQPGEICQVLWGFAKL----GC
Chrei 170 DSWDTTLSLWAYAQL----DH-YDEGALRALCDAAVEVAPIFKPVDCANAVVAFAHLD---YVHPELLRQIVQT---VLDTLD----DYAPGEVCQVLWGFARL----GV
Gopec 161 DSWDTSLSLWSYGQL----GH-HDEAALRALCDAALGVAPIFKPADCANTLVAFANLD---FMHRELLKQLVVT---VLDTLD----DFQPGELAQVLWGFARV----GC
Yauni 159 DSWDATMSLWAYSQL----GH-HDEAALRALCDVALEVAPIFKPADCANAVVAFAYLD---FLHTGLLRQIVTT---MLDTME----DFQPGEICQVLWGFAKL----GC
Cheur 300 DSRDVSTCLWALSALPRTVSR-AHAGTFEVLCRRGRQVAVLMKPADCAMYMQAFGRLG---TYDTELLHAIPQV---MLQELD----RTDMQNVRAVLWGFAHLGPDHPQ
ChKRBP 428 DAWGVSSCLWALSLL----DH-WDRQLFDVLCSRALQLTASLTPTDCANILVAFGRWG---HYHPELLHTLPQV---LLDHMY----DAKPTEISQSLWGLARLRV--PG
Ooamb1 169 EAWDTALSLWSMSVL----NH-FDRAVFIALCHRSCQLAGFMKTSDCAMIMLAFGRFQ---NMHPELLRLIPQV---MLKELD----RAKPQDVSAVLWGFARLGA--GC
Duter 208 GARDVASSLWACAAI----GH-YDRALFDMLCNKAVALLPEMKPVDCANMMVAFARFG---HYHPEVIRMIPQV---LLVQIY----ETKPHELSQVLWGYGRLRV--PG
Chmoe 187 EARDAALCLWALSLM----GS-FEPEPFHALCQRASQLAVEMNSADCTMVMLAFGRFQ---RMYPDLLQQIPQT---LFLCLD----ATKPHELSSVLWGFARLGP--SC
Chlei 171 EAWDVAMSLYAYALM----DH-YDRSVFDALCSRACVLAPSFKPVDCANIMYAFGRFG---HYHPELLRAIPQV---LLYHMY----DAKPGELSQVLWGYGRLRV--PG
Chapp 157 EPRDVAGALWAYGAL----DA-YDKPVFDALCGRAAGQAPAFKPVDCANVMSAFGRFG---HYHPEVLKSIPQV---MLYHMY----DAKHVELATVLWGYARLRV--PG
Halac 118 EPRDIASALWAYGAL----NA-YDKPLFDALCSRAAAQVAGFKPVDCANVMSAFGRFG---HYHPEVLKSIPQV---LLYHMY----DAKPLELACVLWGFARLRV--PG
Ooamb2 1 ----------MSAFARFG---HYNSELLRAVPAA---LLRQLR----SARAREVSQVCWGFAKLEI--PQ
Cheus 171 DMWGISSCLWAMSTL----KT-YDKEVFDSLCKRGAHLSALMRPADCNMIMIAFARFG---HYHPELLRVIPQV---MLMFLD----QSTPSDIASVMWGFAKLSL--PH
Chaci 194 DMWGISSCLWAMSTL----KT-YDKEVFDSLCKRGAHLSALMRPADCNMIMIAFARFG---HYHPELLRVIPQV---MLMFLD----QSTPSDIASVMWGFAKLSL--PH
consensus 441 d wdisislwaya l d lf alc ra m fkp dcaq mvafarm h ellr l ll ld df p el qvlwgfarl
// OPR1 → ← OPR2 → ← OPR3//

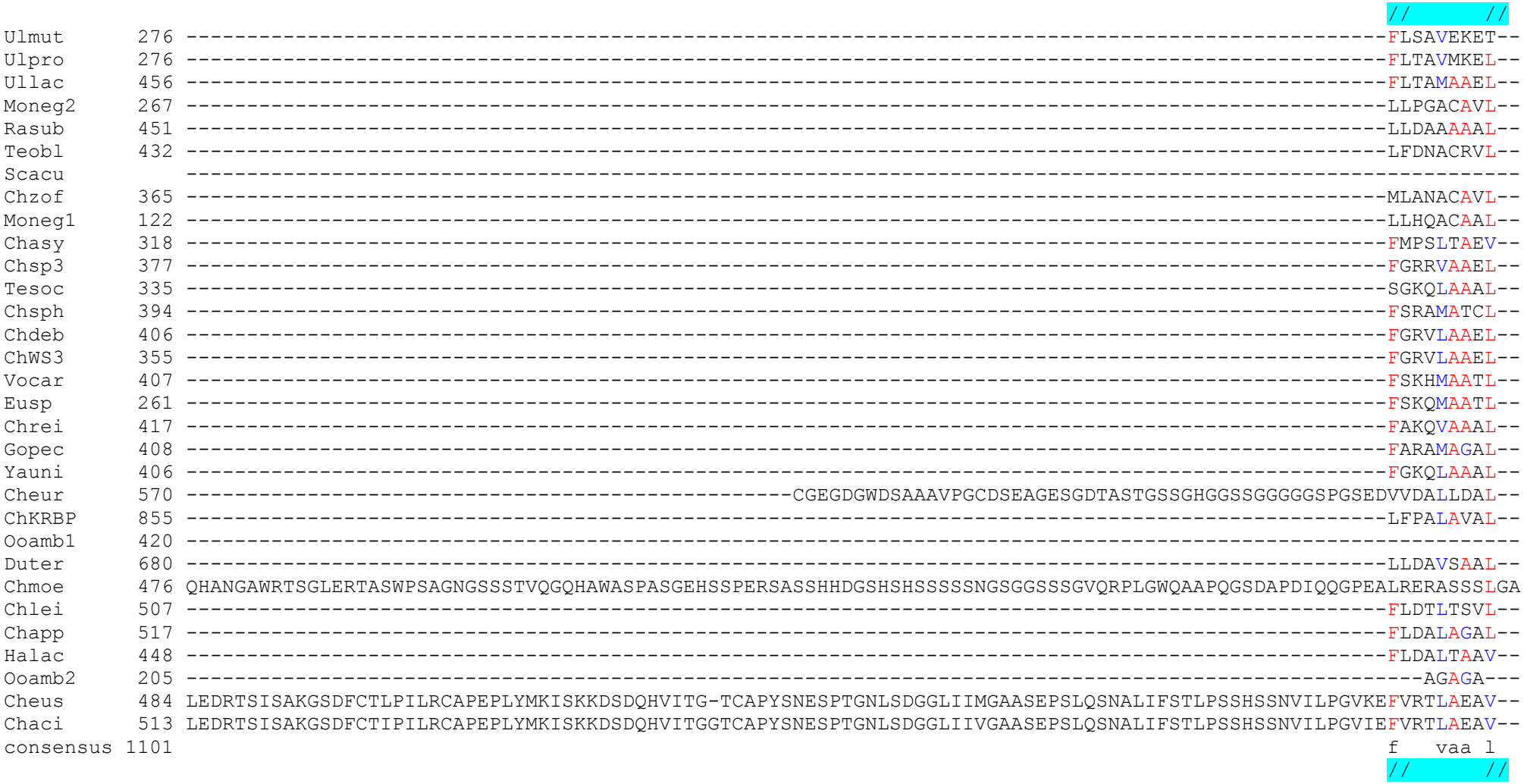

// OPR3 → ← OPR4 //
Ulmut 115 QYCPDILRAAALRMM----------EAPADFAADELVRLLWS----------FSRLHWHEAREL--
Ulpro 115 QQSPEILRAAARRMV----------ERPGDFASEELVRLLWS----------FTRMRWREARH--S
Ullac 295 QQSPEILRGAACRML----------ESPEEFASEELVRLLWS----------FTRMRWREARQ--N
Moneg2 109 RHRRLLLETLMDVVQ----------RRLEQFGPRDLSTVVFA----------YARLRFNAPGALLR
Rasub 276 LHRRAILETLMDVAQ----------WKLDAFNVQELTTLVYA----------YARMHHRMPSALLR
Teobl 242 PSRRSLLEGLLEMTL----------WRLEEFSVQQLAIVVYS----------CGRMRLRLPQQLAK
Scacu 283 PSRRSLLEGLLEMTL----------WRLEEFSVQELAIVVYS----------CGRMRLRLPQQLAK
Chzof 209 GGRRMLLEALSDVVM----------WRLGDFTMQELSNIVYG----------YARMHYRVPQHLAC
Moneg1 1 ----------MHYRMPSALVR
Chasy 162 NPAPRFMAALVDVVQ----------WKVVKFGAQELSVVLYS----------LGKLGYQPEFQFLQ
Chsp3 221 HPGATFLAEVVDAVQ----------WRLQHYATQELLMCMWA----------LVRLGYRPGPRLLL
Tesoc 196 HPGAAFLSEVTGGMQ----------GRLEQYGSQELALILWS----------SARLRHKPGLRFLN
Chsph 238 HPGAPCLAEMVDTVQ----------WRVQSYSTKELCLVLWS----------LARLGCRPPPRFLR
Chdeb 250 HPGAPFLGEVMAAVQ----------GRLQAYSTQELGMVLWG----------LARLGARPPSRFLR
ChWS3 200 HPGAPFLGEVMAAVQ----------GRLQAYSTQELGMVLWG----------LARLGARPPPRFLR
Vocar 251 HPGATLLHEVVATVH----------RNVRQYGTQELVLVLWA----------LVRLGHKPGVRFLY
Eusp 105 HPGATLLHEAVAAVH----------RRVRQYSTQELVLVLWA----------LVRLGHKPGVRFLY
Chrei 261 HPGPAFLAEVVDAVQ----------WRLQGYGTQELGMVLWA----------LVRLGYKPGPRFLR
Gopec 252 HPGEHFVEELVEAVQ----------WRMQYYSTQELTMVLWS----------LVRLRHRPGLRFLQ
Yauni 250 HPGAAFLVELVDAIQ----------WRVQQYGTQELVMVMWS----------LVRLNYKPGPRFLH
Cheur 399 APSTMFMEAICDHLT----------WSMEQCSNQALTNCLWAXREAGPSPAXXXKLGHRPSPGLLR
ChKRBP 521 APGRGFLEAVCEMAMTGSYTAAQQQRQGSSGASQSGGEGAGMMPGSGDAGGGGMEHERAHRSATHATGHGFAAYGCQELSTVLWA----------LARLHHYPGRSFLQ
Ooamb1 262 TPGRTFMGALSANIH----------WSIDQFGCQELVMCMWS----------MAAMDYRPAKQLLG
Duter 301 APGKVFLEAATYAFS----------SSMDQFGPQELSNMLWA----------FVKLHHYPGRSVLL
Chmoe 280 APHKPFMQAVSRRLR----------EDIQNYGGQELALSLWS----------LASLGYRPEQSVLR
Chlei 264 APGKVFLEAMCDTLH----------HSMAAFRPQELANTMWA----------LARLGHHPGRAVLQ
Chapp 250 APGKVFLEAVCDTLA----------HSLSSCGPRELSTIMWS----------LARLQHHPGRALLQ
Halac 211 APGKVFLEAVCDTLA----------HSMPAYGPRELSILMWS----------LARLQHHPGRALLQ
Ooamb2 49 MPGQVFLEALGDHLM----------WTVDEYSCQELATILWS----------MAALRHHPGAAFLV
Cheus 264 APGDLFLDAICDNLF----------WTIDQYSCQELSMFLWS----------VATLNYRPCRGLLR
Chaci 287 APGHLFLDAICDNLL----------WTIDQYSCQELSMILWS----------VATLNYRPCRGLLR
consensus 551 pg fl al d m rl ya qel illws larl rpa ll
// OPR3 → ← OPR4 //

```
                //      //
Ulmut      159 GVHQSVVIEL----------------------------------------------------------------------------------------------------
Ulpro      159 DVHARIVAEL----------------------------------------------------------------------------------------------------
Ullac      339 HVHERIVAEL----------------------------------------------------------------------------------------------------
Moneg2     155 -IQDRVLRSI----------------------------------------------------------------------------------------------------
Rasub      322 -ISGKLSRAL----------------------------------------------------------------------------------------------------
Teobl      288 ITHHIAVQA-----------------------------------------------------------------------------------------------------
Scacu      329 -ITHHIAGHL----------------------------------------------------------------------------------------------------
Chzof      255 -ISNRVGEHL----------------------------------------------------------------------------------------------------
Moneg1      12 -IANKVAHNT----------------------------------------------------------------------------------------------------
Chasy      208 QMERIVLSRL----------------------------------------------------------------------------------------------------
Chsp3      267 DVERVLLQRL----------------------------------------------------------------------------------------------------
Tesoc      242 EMEDVLLQRL----------------------------------------------------------------------------------------------------
Chsph      284 EVESCLLSRL----------------------------------------------------------------------------------------------------
Chdeb      296 DTETVLLQRL----------------------------------------------------------------------------------------------------
ChWS3      246 DTETVLLQRL----------------------------------------------------------------------------------------------------
Vocar      297 DVETTLLQRL----------------------------------------------------------------------------------------------------
Eusp       151 DMETTLLQRL----------------------------------------------------------------------------------------------------
Chrei      307 DVESVLLARL----------------------------------------------------------------------------------------------------
Gopec      298 LAEGTLLQRL----------------------------------------------------------------------------------------------------
Yauni      296 DVETTLLQRL----------------------------------------------------------------------------------------------------
Cheur      456 AAEARLLAAQ----------------------------------------------------------------------------------------------------
ChKRBP     620 HAESTMVRRLRVATVA-----------------------------------------------------------------------------SRQSFAAPVVQRPQQEQQQQ
Ooamb1     308 AAERRLVQLAAQ----------------------------------------------------------------------------------------------------
Duter      347 RSETLTLRRLISSRSRFVPATRASTTTNTTNNTAAASPSIPNSVPTSHTQQPPLSQSHQHVSPPSLPPHYGGAGVPAAAGEQPHLQRNGSSLASSSSSSSSGMHTSADQG
Chmoe      326 AAEQHLLAIA----------------------------------------------------------------------------------------------------
Chlei      310 RSESLMLARL----------------------------------------------------------------------------------------------------
Chapp      296 RSEALTLRRL----------------------------------------------------------------------------------------------------
Halac      257 RSEALTLRRL----------------------------------------------------------------------------------------------------
Ooamb2      95 RMEHKLVERA----------------------------------------------------------------------------------------------------
Cheus      310 KCESVLIERL----------------------------------------------------------------------------------------------------
Chaci      333 KCESVLIERL----------------------------------------------------------------------------------------------------
consensus  661  ve  ll rl
                //      //
```

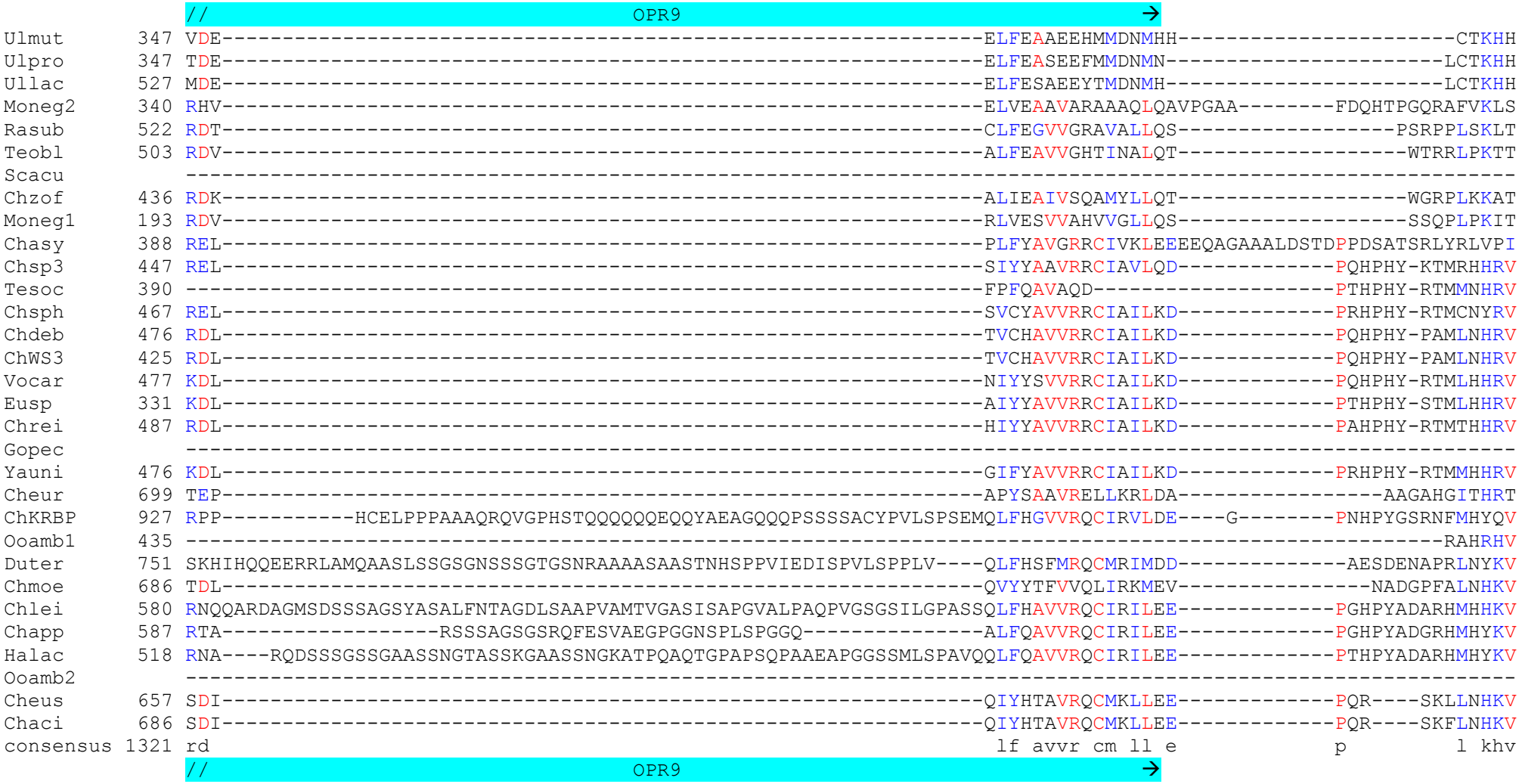
```
          //                                                                                                          //
Ulmut      169 --------------------------------------------------------------------------------------------------------------
Ulpro      169 --------------------------------------------------------------------------------------------------------------
Ullac      349 --------------------------------------------------------------------------------------------------------------
Moneg2     164 --------------------------------------------------------------------------------------------------------------
Rasub      331 --------------------------------------------------------------------------------------------------------------
Teobl      297 --------------------------------------------------------------------------------------------------------------
Scacu      338 --------------------------------------------------------------------------------------------------------------
Chzof      264 --------------------------------------------------------------------------------------------------------------
Moneg1      21 --------------------------------------------------------------------------------------------------------------
Chasy      218 --------------------------------------------------------------------------------------------------------------
Chsp3      277 --------------------------------------------------------------------------------------------------------------
Tesoc      252 --------------------------------------------------------------------------------------------------------------
Chsph      294 --------------------------------------------------------------------------------------------------------------
Chdeb      306 --------------------------------------------------------------------------------------------------------------
ChWS3      256 --------------------------------------------------------------------------------------------------------------
Vocar      307 --------------------------------------------------------------------------------------------------------------
Eusp       161 --------------------------------------------------------------------------------------------------------------
Chrei      317 --------------------------------------------------------------------------------------------------------------
Gopec      308 --------------------------------------------------------------------------------------------------------------
Yauni      306 --------------------------------------------------------------------------------------------------------------
Cheur      466 --------------------------------------------------------------------------------------------------------------
ChKRBP     656 HLPQAPASSPGGNQGAPAQEQHQHQRLDSANGVPVQHTPLGAAHVPPPPPPPPPHGSKAPGGPQVLPA------------------------APIHQHQHQQPSVHSHAG
Ooamb1     320 --------------------------------------------------------------------------------------------------------------
Duter      457 LGMDGNSATLTAASNSSHHAQGAGHSSNNENGRGPPHLHSATQPNPTLTPPLPARLTVGAAGSQDIASPHTIDQGSSSSSTTTTTTTTATSFDATAAAHLSQDGLPPIPG
Chmoe      336 --------------------------------------------------------------------------------------------------------------
Chlei      320 -------------GPAPERRRSDDSEEQAEHGGSSNGVPYANGVHSEAAQEAGSDAPQLNGS--------TQHSLDHLLGAGSAQVVASHTSEAASSSR-----------
Chapp      306 -------------GPAPDVNGLQPLQHAAEQGEVAAAGACVEGEGADEGGEVRLGLPSENGAEYATAPLSGEPSLQHLNGHANGHAHGHHMNGQVNGTAMHHAPELLVGL
Halac      267 -------------GPVPDVAGLQVLAPAAA-GQVDEV-ALLEGD----------LVPQHLGT--------------HLNGHVNVHANG-HVNGHVNGHA-HHAPQLLAGF
Ooamb2     105 --------------------------------------------------------------------------------------------------------------
Cheus      320 --------------------------------------------------------------------------------------------------------------
Chaci      343 --------------------------------------------------------------------------------------------------------------
consensus  771
          //                                                                                                          //
```

```
               // OPR4 →← OPR5 →← OPR6 //
Ulmut      169 ------ARPERIVKLVPQDFANVMWALARMHVRVGPRELDALAEAAAAQLPRFKEQ----------------------------------EVANLVYGYGAQRHQHDGLL
Ulpro      169 ------ARPERAAKLLPQDFSNTLWSLAQMKVYAEKPDLDALADAATANLHRFKEQ----------------------------------ELANLVYGYGAQRHWHEGLM
Ullac      349 ------ARPERVSKLLPQDFANTMWSMAQMQVFADPKDLDVLAEAAAGQIHRFKEQ----------------------------------ELANLVYGYGAQRHWNEGLM
Moneg2     164 ------------DTAAPLDMALLLWGFARLGFAPREELMRRLGASVVRQLGAFKPR----------------------------------ETSILLYSYGALSHNDPLVI
Rasub      331 ------------HRLGPQDAAMLLWGCARLGFRPPPSLLDALPGALAPRVAEFSPA----------------------------------EISALLYGFGALRAPAPALF
Teobl      297 ------------QNMNPLDAAHVMWGCAKVDFTPGATLLERLPGVMVPRLSEFKPQVGCGEYSKGSQGYLRCILMLDGVCGFCDVLLGGMEVCMLLYGYGHLRQCSPVLL
Scacu      338 ------------QELNPLDAAHVMWGCAKLDFTPGAALLERLPGVMVGRLSEFKPQ----------------------------------EVCNL---------------
Chzof      264 ------------LLLKPQDASNLMWALAKLHFKPNRRLLEELPVAIVDRLSEFKPQ----------------------------------EVCNLLYGYGCMRYHHPLLM
Moneg1      21 ------------DLLSPQDAANMMWGFARLAFKPGPPLLDRLPLAVAGRLHEFKPQ----------------------------------EICNLLYGYGVLRHHHPLLM
Chasy      218 ------------PHLRPEDIAQSAWGLAKLRYKA-VKLLDALPAFVIPQLHLYKVE----------------------------------EMLLLTYAYSAPRHYHRTLT
Chsp3      277 ------------PLLAPVDVAMATWALARLRYKA-VRLLDELPAAVVPELRSYGNR----------------------------------ELCGLLTGYATAKHYHKTLL
Tesoc      252 ------------PHLSPSDSCVAVWSFAHLGYKA-VRFLDEVPQSLGPQLHSCRNS----------------------------------ELCALISGFATAYHYHRVLL
Chsph      294 ------------PHSAPGDVCVAVWAFSRLRYKA-VRLLDEAPGVLGPQLHSCSSR----------------------------------ELIGLITGFATSHHYHKVLL
Chdeb      306 ------------GYLSPADMCMGLWSLSRLRYKP-VRLLDEMPAALGPMLRSCQTR----------------------------------ELLGIVSAFATAHHYHKVLL
ChWS3      256 ------------GHLSPADMCMGVWSLSRLRYKP-VRLLDEMPAALGPVLRSCQTR----------------------------------ELLGIVSAFATAHHYH-VLL
Vocar      307 ------------PHMAPPDLSLSVWTFAHLRYKA-VRLLDEVPAAVGPQLRNCSNA----------------------------------ELCSLVSGFGTAHHYHRMLL
Eusp       161 ------------PHMAPVDLSLSVWTFGHLRYKA-VRLLDEVPAAVGPQLRNCSNA----------------------------------ELCSLVSGFATAHHYHRVLL
Chrei      317 ------------PHMAPGDIAITVWSFARLRYKA-VRFLDEVPAAVGPQLHKCRSS----------------------------------ELCSLISGFATAHHYHKSLL
Gopec      308 ------------PHMAPVDVCVSVWTFAHLRYKA-VRLLDEVPQAIAPQLHACKNS----------------------------------ELCALVSAFATAHHFHRVML
Yauni      306 ------------PHMAPVDLCLSVWSMARLRYKA-VRLLDEVPQAVGPQLRNCSNA----------------------------------ELCSLVSGYATAHHYHRVLL
Cheur      466 ------------EQLRPVDVAHALWAFGRLRFRA-ARLMSELPPLLAGELEAYAPE----------------------------------ELSCVLFGYTHVRHYSPLLL
ChKRBP     742 YRPPSLRSAPDPRMLVPQDVSLTLWAFGRMRFKA-VHLLDELPLCLPAWLPAFKPS----------------------------------EVCALLAGYANARHYHRGVL
Ooamb1     320 ------------QQLNPQDIGNSLWAFGRMRYRA-SSLLASLPQRILQQLDRFTAQ----------------------------------EMSCLLYGYTQARTYHRALM
Duter      567 RKGPPLAVSQPLQPLVPQDVGNLLWAYGKLRFKG-ARLLDELPLYLGSWLHEFRTP----------------------------------DLCCLLAGYTNARHYHRGVL
Chmoe      336 ------------PHLSSQDVGNALWALGRMRYRA-TDLLDALPTHVVGRLQDFRPQ----------------------------------ELSCLLFGYTHTRHLSPHLL
Chlei      398 ----AQQQRPPMQRMVPQDVSNMLWAYGRMRYKA-AALLDTLPLHLGRWLHAWSVS----------------------------------DLCCLLVGYTHARHYHRGVL
Chapp      403 PAATAGRPPPPLQRMSAQDMGTLLWAYGRLRYKA-APLLDALPLHLGPWLRAWTTT----------------------------------DLMCLLVGYTHARHYHRGVL
Halac      336 PPIPGAR--PPLQRMTPQDMGNLLWAYGRLRYKA-APLLDALPLHLGAWLHVWTPT----------------------------------DLCCLLVGYTHARHYHRGVL
Ooamb2     105 ------------PLMSPQDIGNVLWALGRMQYKA-TALLDMLPLLVGRRLAEFKPM----------------------------------ELSCLLYAYAAARHYHAALL
Cheus      320 ------------PSLKCQDIANSLWALGRLHCKP-VRLLDELPLHFGRRLHEFKPQ----------------------------------ELSCVIFSYAQSRHFHCGLL
Chaci      343 ------------PSLKCQDIANSLWALGRLHCKP-VRLLDELPLHFGRRLHEFKPQ----------------------------------ELSCVMFSYAQSRHFHCGLL
consensus  881              kl pqDma  mWa arlryka  rlldelp  va  lh fk                                    el  lv gyg arhyhr ll
               // OPR4 →← OPR5 →← OPR6 //
```

```
                //          →   ←                    OPR7                          //
Ulmut       239 SAAAAELLLRRE--RFQDQELAMFLWAVGMLGT-A------------------P--GGEL----------------------------------------------------
Ulpro       239 TAAAAEVLLRKE--RFQDKELCMFLWAVGTLGV-A------------------P--GGEL----------------------------------------------------
Ullac       419 TAAAAEVLLRKE--RFQDKELCMFLWAVGVLGV-A------------------P--GGEL----------------------------------------------------
Moneg2      228 EAVAAAAAARLD--RFGTQELVAAIWALGALRG-A------------------GGAAAAT---------------------------------------------------
Rasub       395 EAAAAEFSGRLS--GFSSQELSMTVWAFGATSF-AAASAAAGRGSPRGPAAALPP--AAAA----------------------------------------------------
Teobl       395 EGVAAACSSRLH--EFSSQDLCVTLWSYGMVQY-S------------------P--ADTT----------------------------------------------------
Scacu           ----------------------------------------------------------------------------------------------------------------
Chzof       328 EAVAKACAPRLH--EFSTQDLVLTIWSYGMVKH-M------------------P--SDPA----------------------------------------------------
Moneg1       85 DAAARASAPRLA--EFSTQDIVLTIWAYGAVRH-M------------------P--SDPA----------------------------------------------------
Chasy       281 DAIADALLPTLS--QLSSQDVVMVAWSLANFRH-R------------------P--ANPA----------------------------------------------------
Chsp3       340 EEVAGTVVHRLP--SLTHQEVCVVLWAYGLFRH-R------------------P--SQPD----------------------------------------------------
Tesoc       315 DEASDVILGRLD--TFSHHE--------------------------------------PD----------------------------------------------------
Chsph       357 DEVAEVLLTRLE--ALTSHEVAAALWCYGVFHH-R------------------P--ASPE----------------------------------------------------
Chdeb       369 DDVAEVLVERME--DMTHHDVAVVLWTYGVFHH-R------------------P--AHPD----------------------------------------------------
ChWS3       318 DDVAEVLVERME--DMTHHEVAVVLWTYGVFHH-R------------------P--AHPD----------------------------------------------------
Vocar       370 DAVAGVVLCRIE--TLTHQEVCVVLWTYGTFRH-R------------------P--SNAD----------------------------------------------------
Eusp        224 EAVSGVVLSRLD--SLTHQEVCVVLWTYGTFRH-R------------------P--THAD----------------------------------------------------
Chrei       380 DAVADVLLSRLD--GLSHHEVATALWTFGTFRH-R------------------P--AHPD----------------------------------------------------
Gopec       371 EAVAAVAVPRLE--TISARDVAVLLWTYGAFHH-R------------------P--AHPD----------------------------------------------------
Yauni       369 DAVAGVVLSRLD--TLSHQEVCVVLWTYGTFRH-R------------------P--AHAD----------------------------------------------------
Cheur       529 DAAAPLLVARASAGQLSQQDVVIALWAYGIFGH-R------------------PGGTGGG---------------------------------------------------
ChKRBP      817 EALAPLLGGHLR--RLGQTELVVTLWAYGLFQH-H------------------P-VSEPA----------------------------------------------------
Ooamb1      383 DAAAPVITSRVS--LMSCQDIVIALWAYGIFAL-K------------------P--APAD----------------------------------------------------
Duter       642 EVLAPLLMARIH--TLNVSELVVVLWSYGIFQH-R------------------P-LSEPR----------------------------------------------------
Chmoe       399 GAAAPVLTQKSA--QMSEQDAIIALWAYGIFAHPH------------------PEHTSQG------VAGHA------ADHSLPTSNSSRWQQQQQQQQQQAAVATGAEQ
Chlei       469 DAVAPLLASRAH--ELTLSELVVPFWAYGIFQHKP------------------P--TCPG----------------------------------------------------
Chapp       478 DAVAALLQHRLP--QLGLSELVIVLWAYGIFQH-H------------------PTTTAPG---------------------------------------------------
Halac       409 DAVAPLLMHRLP--ELGLNHLVTVLWAYGIFQH-N------------------PAATAPG---------------------------------------------------
Ooamb2      168 DAAAPLLMQRAP--YMSHQDLIMAMHSYGIFGH-N------------------P--RLPG----------------------------------------------------
Cheus       383 EAAAQVIVNRVP--QMSHQDVVMALWVFGIFGH-QLKIPNIAVTSGANSFQKCQSFAKTENVTVEQSELMLK-----GQRTNLTDAEQNQRRQISVAEGQFPVAMKRDKG
Chaci       406 EAAAQVIVNRVP--QMSHQDVVMALWVFGIFGH-QLKIPNIAVTSGASSFQKSQSFAKTENVTVEQSELILKRSLLKGQRTNLTDAEQNQRRQISVAEGQFPVAMRRDKG
consensus   991 eava vll rl    ls qel m lw ygmf h r                  p    p
                //          →   ←                    OPR7                          //
```


//
//
Ulmut 276 FLSAVEKET--
Ulpro 276 FLTAVMKEL--
Ullac 456 FLTAMAAEL--
Moneg2 267 LLPGACAVL--
Rasub 451 LLDAAAAAL--
Teobl 432 LFDNACRVL--
Scacu
Chzof 365 MLANACAVL--
Moneg1 122 LLHQACAAL--
Chasy 318 FMPSLTAEV--
Chsp3 377 FGRRVAAEL--
Tesoc 335 SGKQLAAAL--
Chsph 394 FSRAMATCL--
Chdeb 406 FGRVLAAEL--
ChWS3 355 FGRVLAAEL--
Vocar 407 FSKHMAATL--
Eusp 261 FSKQMAATL--
Chrei 417 FAKQVAAAL--
Gopec 408 FARAMAGAL--
Yauni 406 FGKQLAAAL--
Cheur 570 CGEGDGWDSAAAVPGCDSEAGESGDTASTGSSGHGGSSGGGGGSPGSEDVVDALLDAL--
ChKRBP 855 LFPALAVAL--
Ooamb1 420
Duter 680 LLDAVSAAL--
Chmoe 476 QHANGAWRTSGLERTASWPSAGNGSSSTVQGQHAWASPASGEHSSPERSASSHHDGSHSHSSSSSNGSGGSSSGVQRPLGWQAAPQGSDAPDIQQGPEALRERASSSLGA
Chlei 507 FLDTLTSVL--
Chapp 517 FLDALAGAL--
Halac 448 FLDALTAAV--
Ooamb2 205 AGAGA---
Cheus 484 LEDRTSISAKGSDFCTLPILRCAPEPLYMKISKKDSDQHVITG-TCAPYSNESPTGNLSDGGLIIMGAASEPSLQSNALIFSTLPSSHSSNVILPGVKEFVRTLAEAV--
Chaci 513 LEDRTSISAKGSDFCTIPILRCAPEPLYMKISKKDSDQHVITGGTCAPYSNESPTGNLSDGGLIIVGAASEPSLQSNALIFSTLPSSHSSNVILPGVIEFVRTLAEAV--
consensus 1101 f vaa l
//
//


// → ← OPR8 → ← OPR9 //
Ulmut 285 ----------QGRAH--RLKPL---GISNTMKAFAKLRHLPQAE----FIDCMSEAAVAIMDDFSMAELSNLLWAYSQLGW---
Ulpro 285 ----------HSRVK--TLRPL---SMSNAMKAFAKLRFLPPAE----FMDSMSEVAVASLEEFNMSELSNLLWAYAQLGW---
Ullac 465 ----------HSRAH--SLKPL---SISNAMKAFAKLRYLPEAE----FMDSMSEASVDSLHHFNMSELSNLLWAYAQLGW---
Moneg2 276 ----------LARARARRLAPG---HVAVGLKGLARAGFRPPAA----LTSELCACALENLASFKPVELCHLMWALARLGA---
Rasub 460 ----------LARRR--RLLPA---QIAMAAKALARCGHAPPPG----FMDEMASLSLERLPAFKPIEMCHLLWAYASLGY---
Teobl 441 ----------LSRAN--RLLPL---ELTMAIKGFARAGYQPPPE----FMRQVAQLAMQKLHQFSPMEYSQLLWAYAALGY---
Scacu ----------
Chzof 374 ----------QSRSQ--WLLPI---QISIIMKGFAKIGYQPPTA----FMAELANVALNKIQLFKPVELCQLLWGYAHLGY---
Moneg1 131 ----------LARRR--RLLPM---QVAMVLKAFAKIGFQPPPA----VMAELGALALEKLPAFKPVELCHLMWAYARLGY---
Chasy 327 ----------YKRLR--TFKPV---GLVLMAKAFSRLGWRSDT-----MLKELTRLAEQRLADFSLEELAHLVYGLSTLQW---
Chsp3 386 ----------HGRLG--AMTPQ---GLAMVAKGLAQLQWRSGP-----LMAQLVAAAEQKLTGFKPLELSQLLWGLSSLQC---
Tesoc 344 ----------YMRMR--HFTPQ---GLAMVVKALAQLQWRSEP-----LMAELIVAAELKLLAFNT----------
Chsph 403 ----------YNRMP--SFAPQASAGLAMIAKALAQLQWRSEP-----VLAELVMAAEAKIQGFKPLEMSQLLWGLTQLQC---
Chdeb 415 ----------YGRMG--GFSAQ---GLAMVAKALAQLQWRSEP-----LLVELAAAAATKIQAFKPLEMSQLLWGLTALQC---
ChWS3 364 ----------YGRMG--GFSAQ---GLAMVAKALAQLQWRSEP-----LLVELAAAAATKIQAFKPLEMSQLLWGLTALQC---
Vocar 416 ----------YSRMP--YFAPQ---GLAMIVKALAQLQWRSEP-----LLVALMAAAETKMNGFKPLELSQLLWGLTALDC---
Eusp 270 ----------YSRMP--YFAPQ---GLAMIVKALAQLQWRSEP-----LMVALMAAAETKLNAFKPLEMSQLLWGLSSLEC---
Chrei 426 ----------YARMR--SFSPQ---GLAMVVKALAQLQWRSEP-----LMEQLIAAAEAKLNAFKPLELSQLLWGLTALQC---
Gopec 417 ----------GARMR--QFSPQ---GLAIVCKALAQLQWRSEP-----LMAELVAAAEAQLPGFK----------
Yauni 415 ----------YSRMA--HFTPQ---GLAMIAKALAQLQWRSEP-----LLMELIEAAEPKLQAFKPLEMSQLLWGLSALQC---
Cheur 628 ----------LARGGVFGVAQLPTVALANVVKALASVRRPNGSARTAALAAAVAGAAERRIEDFSPCEMSNMLYGLALLRY---
ChKRBP 864 ----------HARLRHEGARPQ---AYALVYKACANLQFRPEP-----LLQALVQAATARLPDFRPDEMANLLYGISHLGS---
Ooamb1 420 ----------AETAAGAVAALDPAT----------
Duter 689 ----------VRQLRPGTYMQ----TYSLVIKAYANLRYWPES-----VLSKMVENSEARINSFRPDEMAALLYGLSHLAV---
Chmoe 586 STSAAAPAAQSNGNGGGSSSSSSVEEMVHACIRVVRAGNVA--NLRPV---ALGNMVKALASLRASGADVSA--VVTAAAKAAVAKAPDFQPAEVANLLYGLSLMRY---
Chlei 516 ----------VERLR--GARPQ---TYSMLAKACANLRYTPDV-----LLHQMALGAARRVGEMRPDEMAALLYGMSHLAAQAV
Chapp 526 ----------AAQLP--AARPQ---TYALVVKACANLRLAPEE-----LLAEVAAGAAGRVHEMRPDEMAGLLYGLSHLAA---
Halac 457 ----------QGELR--HARPQ---TYALVAKACANLRVAPEG-----LLADLAQGAAGRVGEMRPDEMALLLYGLSHVAA---
Ooamb2 ----------
Cheus 592 ----------QSRLR--YLNPQ---GIANAIKALGCLRVSIIPEVQA-LARAAAKSAEAHIEDYRPDEISNLLYGLSLLGY---
Chaci 621 ----------QSRLR--YLNPQ---GIANAIKALGCLRVSIIPEVQA-LARAAAKSAEAHIEDYRPDEISNLLYGLSLLGY---
consensus 1211 rvh pq giavimkala l f ep li l aa rm fkp el nllwgls l
// → ← OPR8 → ← OPR9 //


// OPR9 →
Ulmut 347 VDE----------ELFEAAEEHMMDNMHH----------CTKHH
Ulpro 347 TDE----------ELFEASEEFMMDNMN----------LCTKHH
Ullac 527 MDE----------ELFESAEEYTMDNMH----------LCTKHH
Moneg2 340 RHV----------ELVEAAVARAAAQLQAVPGAA--------FDQHTPGQRAFVKLS
Rasub 522 RDT----------CLFEGVVGRAVALLQS----------PSRPPLSKLT
Teobl 503 RDV----------ALFEAVVGHTINALQT----------WTRRLPKTT
Scacu ----------
Chzof 436 RDK----------ALIEAIVSQAMYLLQT----------WGRPLKKAT
Moneg1 193 RDV----------RLVESVVAHVVGLLQS----------SSQPLPKIT
Chasy 388 REL----------PLFYAVGRRCIVKLEEEEQAGAAALDSTDPPDSATSRLYRLVPI
Chsp3 447 REL----------SIYYAAVRRCIAVLQD----------PQHPHY-KTMRHHRV
Tesoc 390 ----------FPFQAVAQD----------PTHPHY-RTMMNHRV
Chsph 467 REL----------SVCYAVVRRCIAILKD----------PRHPHY-RTMCNYRV
Chdeb 476 RDL----------TVCHAVVRRCIAILKD----------PQHPHY-PAMLNHRV
ChWS3 425 RDL----------TVCHAVVRRCIAILKD----------PQHPHY-PAMLNHRV
Vocar 477 KDL----------NIYYSVVRRCIAILKD----------PQHPHY-RTMLHHRV
Eusp 331 KDL----------AIYYAVVRRCIAILKD----------PTHPHY-STMLHHRV
Chrei 487 RDL----------HIYYAVVRRCIAILKD----------PAHPHY-RTMTHHRV
Gopec ----------
Yauni 476 KDL----------GIFYAVVRRCIAILKD----------PRHPHY-RTMMHHRV
Cheur 699 TEP----------APYSAAVRELLKRLDA----------AAGAHGITHRT
ChKRBP 927 RPP----------HCELPPPAAAQRQVGPHSTQQQQQQEQQYAEAGQQQPSSSSACYPVLSPSEMQLFHGVVRQCIRVLDE----G--------PNHPYGSRNFMHYQV
Ooamb1 435 ----------RAHRHV
Duter 751 SKHIHQQEERRLAMQAASLSSGSGNSSSGTGSNRAAAASAASTNHSPPVIEDISPVLSPPLV----QLFHSFMRQCMRIMDD----------AESDENAPRLNYKV
Chmoe 686 TDL----------QVYYTFVVQLIRKMEV----------NADGPFALNHKV
Chlei 580 RNQQARDAGMSDSSSAGSYASALFNTAGDLSAAPVAMTVGASISAPGVALPAQPVGSGSILGPASSQLFHAVVRQCIRILEE----------PGHPYADARHMHHKV
Chapp 587 RTA----------RSSSAGSGSRQFESVAEGPGGNSPLSPGGQ----------ALFQAVVRQCIRILEE----------PGHPYADGRHMHYKV
Halac 518 RNA----RQDSSSGSSGAASSNGTASSKGAASSNGKATPQAQTGPAPSQPAAEAPGGSSMLSPAVQQLFQAVVRQCIRILEE----------PTHPYADARHMHYKV
Ooamb2 ----------
Cheus 657 SDI----------QIYHTAVRQCMKLLEE----------PQR----SKLLNHKV
Chaci 686 SDI----------QIYHTAVRQCMKLLEE----------PQR----SKFLNHKV
consensus 1321 rd lf avvr cm ll e p l khv
// OPR9 →

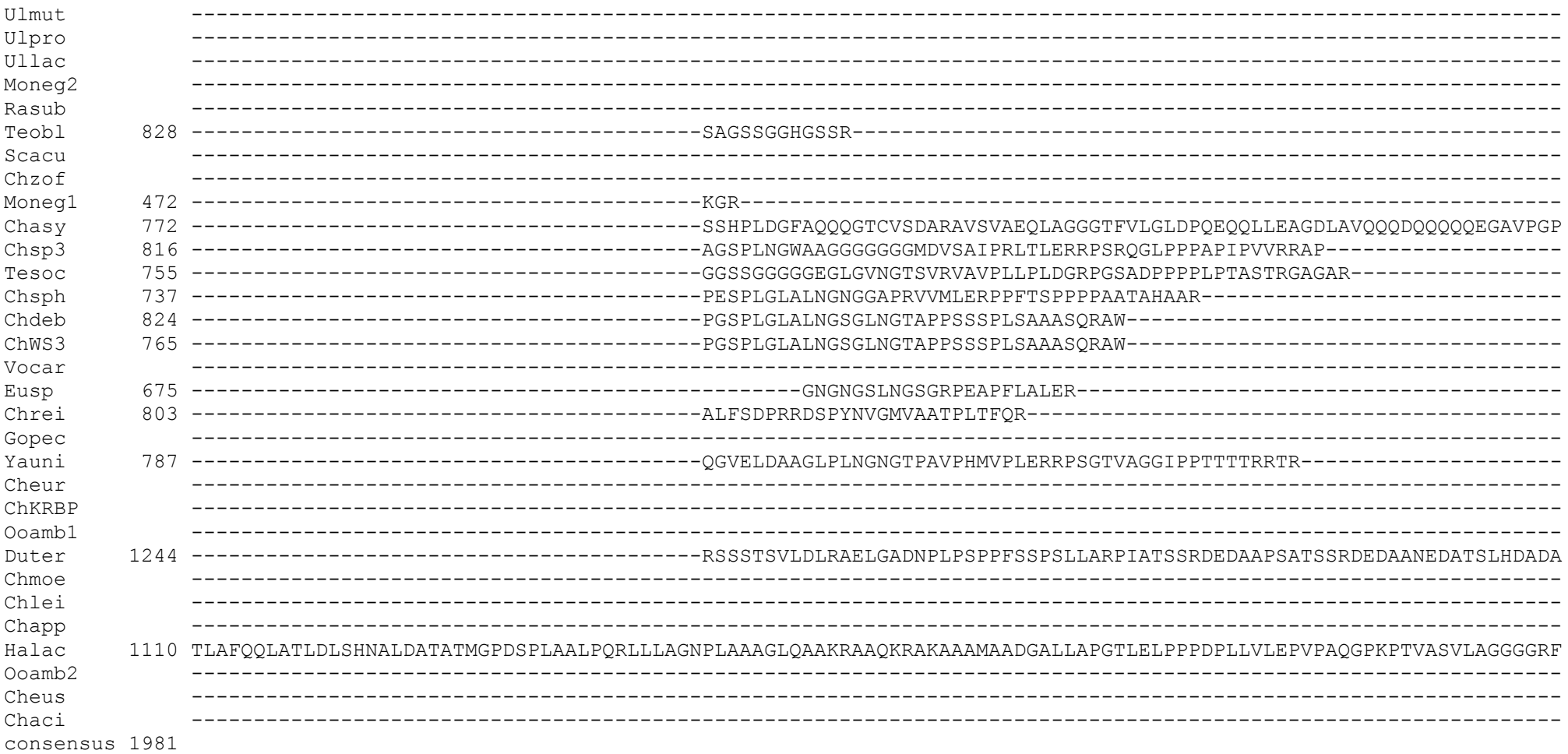
```
Ulmut      371 ILSVLNSFKASGHLCHRLVTVARNHGFYV----------------------------------------------------------------------
Ulpro      371 VSSIANSFKTSGYLCHRLVTVARNNGFHV----------------------------------------------------------------------
Ullac      551 VSSIANSFKASGYLCHRMVTVARHHGFYI----------------------------------------------------------------------
Moneg2     379 LDTVVWAAQRLGYFPAELIEAAGARGVRVK----------------------------------------------------VWRRFKTRFDLGAAP--
Rasub      551 VDTIVWASQRVGFWPQALIDTAEMRGIYVGWGNLAHGS--------------------------------------------GAWTDDGGEAEEGE--
Teobl      531 VDTIIWSCERVGFWPQSLVDTAEMRGVFVKTGSRGAAA----------------------------------------------HEEMEVISGLPQQTD--
Scacu          ----------------------------------------------------------------------------------------------------
Chzof      464 IDTIVFSCQLIGFWPQMLVDMAEMRGIYVRQNADSWNQ----------------------------------------------------------------
Moneg1     221 VDTIVWAAQQVGFWPQLLIDTAEMRGVYVNAGSYSGGP------------------------------------------------ARAAAAAEAFAAAERA--
Chasy      435 INSIVSSCQAVDYVPWTLIDYAESKGMRIKSPVQEDQE--------------------------------------------DEEEQQQEEEQEAAFRAADTAAAR--
Chsp3      480 LNSVIASCQHAGYVPWALVDFAESKGIRVK-QPGNLSS----------------------------------------------RDEEDEGVLAPPFPS--
Tesoc      413 INSVIVSCQGLGYVPWTLIDFAESKGIRVK-QPENLAS----------------------------------------------RDDDDEGLLTMADAELV
Chsph      500 VNSVLGSCQVLGYVPWTLVDFAESRGIRVR-QPDNLAS----------------------------------------------RDDEDDAL---------
Chdeb      509 VNSVIGSCQQLGYVPWTLLDFAESKGIRVK-QPGILSS----------------------------------------------RDEEDEALAPVLVAA--
ChWS3      458 VNSVIGSCQQLGYVPWTLLDFAESKGIRVK-QPEILSS----------------------------------------------RDEEDEALAPVTVAA--
Vocar      510 VNSVISSCQQLGYVPWTLIDFAESKGIRVK-EPDILSS----------------------------------------------RDEEDEGMAALLTAT--
Eusp       364 VNSVISSCQQLGYVPWTLIDFAESKGIRVK-EPDILSS----------------------------------------------RDEDDEGMLVQAPAP--
Chrei      520 VNSVLGSCQQLGYVPWTLIDFAESKGIRVR-QPDILSS----------------------------------------------RDEDDEGVPYSHQQQ--
Gopec          ----------------------------------------------------------------------------------------------------
Yauni      509 VNSVIVSCQQLEYVPWTLIDFAESKGIRVK-QPDILSS----------------------------------------------RDEDDEGMLASHPLA--
Cheur      729 LNSIVHSCVSAGYTPWTLIEAAEMRGLRLHSNPSVLQP----------------------------------------------RW-----LSPPRVAP--
ChKRBP    1014 LNSIIFSCLRVGYTPWALIDFAESKGMRLTTPPSTGGG------------------------------------------------------------S--
Ooamb1     441 VSQEDPQREPHAQEPLSLSDSRHVS------------------------------------------------------------------------AP--
Duter      843 LNSIVFSCLRVGQSPWMLIDYAESRGIRLI-QPGVAMAKLNLEREIIAKAAVEALTGSAPASPLTGTEAEAMKAKKRNRGGRRLRAMQRRLAAESEEERQTAAAEEAA--
Chmoe      717 LNSVVHSCISVGYTPWVLIECAEVRGIRVSGNSTTLTP----------------------------------------------RWQEGPPGFRARSMA--
Chlei      677 LNSIVFSCVRVGYTPWTLIDFAESKGIRII-QPHAAYA----------------------------------------------MLQTRRAAAAAARSSG--
Chapp      651 LNSIVFSCLRVGYTPWTLIDFAESRGVRII-QPHAAMG--------------------------------------------QLSEQLAAARIERTTPAGSPPPAR--
Halac      611 LNSIVFSCLRVGYAPWTLIDFAESRGVRII-QPHAAMG--------------------------------------------QLGAQLVQARHERHMPAGSPPLAR--
Ooamb2         ----------------------------------------------------------------------------------------------------
Cheus      687 LNSIVASCVSVGYIPWTLVESGEMHGMRIK-NPKTVVP----------------------------------------------HWQKLHSTYLAKKRD--
Chaci      716 LNSIIASCVSVGYIPWTLVESGEMHGIRIK-NPETVVP----------------------------------------------HWQKLHSTYLAKKRD--
consensus 1431 insvl sc  lgy pwslid ae hgvrvk np                                                    h
```


Ulmut
Ulpro
Ullac
Moneg2 424 --GDLPEEEVRGAAPGDGAGAGAVAWEEIVRPGMDLPDDFPSPERRAASPRAA----------
Rasub 603 --GEGEADEGAHAAASAGGGAAAEERRRQQQQWQQEQQEEEPGDELDEQQQQQRR----------LFAATDPAGAEGDAAEAAVVQQQQGQRRPPPQ
Teobl 584 --GSSDEEDEDSSSSSSSLGVEAGAAESSSSSSGGLNGFGDQLGAMVASAMSSHEDEEAAQLQQE----------QQQLRQLQRTEALAAAAAASSNNHQQSAAAAA
Scacu
Chzof 502 --AQNLEPLEIDGIGPPDATVMAADLQPHPSNTN----------
Moneg1 275 --GSGDAGHGGGDEGSSGSGDGGSGSSVSSEASSGVNGSSGGGDPGSSGQGHPLLLE----------HRQHGDEAVRWPAALNGHLHNPQQQQQQQQQQ
Chasy 497 --AAVPISLADRIAAAAGARAAAPSSNVDKSGRAAVPAPAAAAAAAHAAVPSSSADKSGRAVAATAAQRGNSSVQPDAILRESGSRQQQQQQSSQQQQQKQQQKQQQRQQI
Chsp3 532 --TSAPSSSSSSGSSNGVPHSGSSSSSGSSSYVSGLSGGYGSGSSSSSPVSAHEYYGSRRSHVARWADQEEEEVVAVEEEAEGEEEEAEWVEEVVVFEGDEDEGLGAGSSR
Tesoc 467 DGSEGPGTEEGDAGEGQGSGQGSSGEEDGEGRERPVAAARRRRPRPRDPAVLGEEGEGAA----------ESEAAAADGRGGGLGGVGGAAAGEQRLGRPES
Chsph 545 --ALGPGGPGAAVADEGQPERAVEEAEDEGAGEVVAEVEARGDGV----------ERRRGGRHGEWVESGAEEQAHEAARAAFRPPP
Chdeb 561 --PSAADDASEAAREGHEGSAAAAAAVSMSLASLVAAVAADAAAPSAAAAAAQEAAAAEPQAERR----------RNLFGRQEGSTGRHGRNHNQRHDHGHRHRHSN
ChWS3 510 --PSAAGDASEGSAAAAAAVSMSLASLAAAVAADAAAPLTAAQEAAAAEPQAERR----------RNLFGRQEGSTGRHGRNHNHRHDHGHRQRHSH
Vocar 562 --ATSPAARAAGDVAAAAEAAATAAPRRQKSLNQEANDGAVEEE----------EEEEEDDVGTSMACDLAAGLGAGQRQRGNPPD
Eusp 416 --SSPSAANGR--DRPAGTPAAVVAWSEEEVEEEVVVVGRSPAMDHNDGRQGQEQEECDDVGHSAAAALATSLVSGRRLQPSTASDGLRVLEQQPEAQHAQHERDPALGS
Chrei 572 --QHADAVEGCGHDRAEQPWGASTGSSSSSSRRHQRCAEEEALWAEAERAHSQQVAAGNS----------SSDAAMASAPDAVVLLEQGLIPHVSSSAAADA
Gopec
Yauni 561 --MMAEPLLSSELSPPSWVLEELQVAEPDGASGRGEAEAALGAARRSPRVRE----------VTDGEDGAGLDRLQARSTGEHPWGEEGEFDVG
Cheur 777 --QPGPRPERTRVRRMPLFIRTAGVVEHMEEHSEQQTDELA----------ERYCMQAAGVQKHDAAQHAHQGWHASLPSACG
ChKRBP 1053 --SSSNGDAQDAGSRPGQAGSGASRPLQAGARRQQQQQQQEAVAGPAP----------PKAFVTLPGQQSASLQTGPAPPEFDVKPGRAR
Ooamb1 468 --TSMQGDDDDGGCATHGLGRRPSLGEEAPSSSSSSSSSR----------
Duter 950 --AAAAEESAAASVALAGQAAAATRAEGANVPPPRAHVTLPGEAVSGHWLSASGRAGSGRLSHRRGPARRPH------SNTPCTTTSTDSGSSSSSSSQTRARTLFGSLD
Chmoe 770 --APSLLRHGRAAQAPPSPPSGLGPGEGGSTGMAPGAGGVGSSTG----------
Chlei 730 --LAGAGSTGAAGSPTPARTTLPGQAEEAEEEGHGRQQSRPGRQRGRRRGPRSTDSSGSSRF----------AQLQSELAASSQQGGEQQHLNGWSGSQQESLP
Chapp 712 --VSHPGQPDRCMLPGASPHPGAERLHRARSRRRSVQQLHGQDELLSDDAHVLPAGNGLRQQRLL----------WREHDSKAGQAVNGAGVNGHAPAGLVRPPPPE
Halac 672 --VTLPGAEDSAGLMPAEAASVMQLVRTPPSRSRPLPFKPAAGPRMNRSRSTAVRNATRMMGPSPSGTPPARPWLETNSTLGSARSAQVQVVHTGLTPAPSIRMLPQSPS
Ooamb2
Cheus 739 --FKLSFPNDT----------
Chaci 768 --FKLSFPNDT----------
consensus 1541

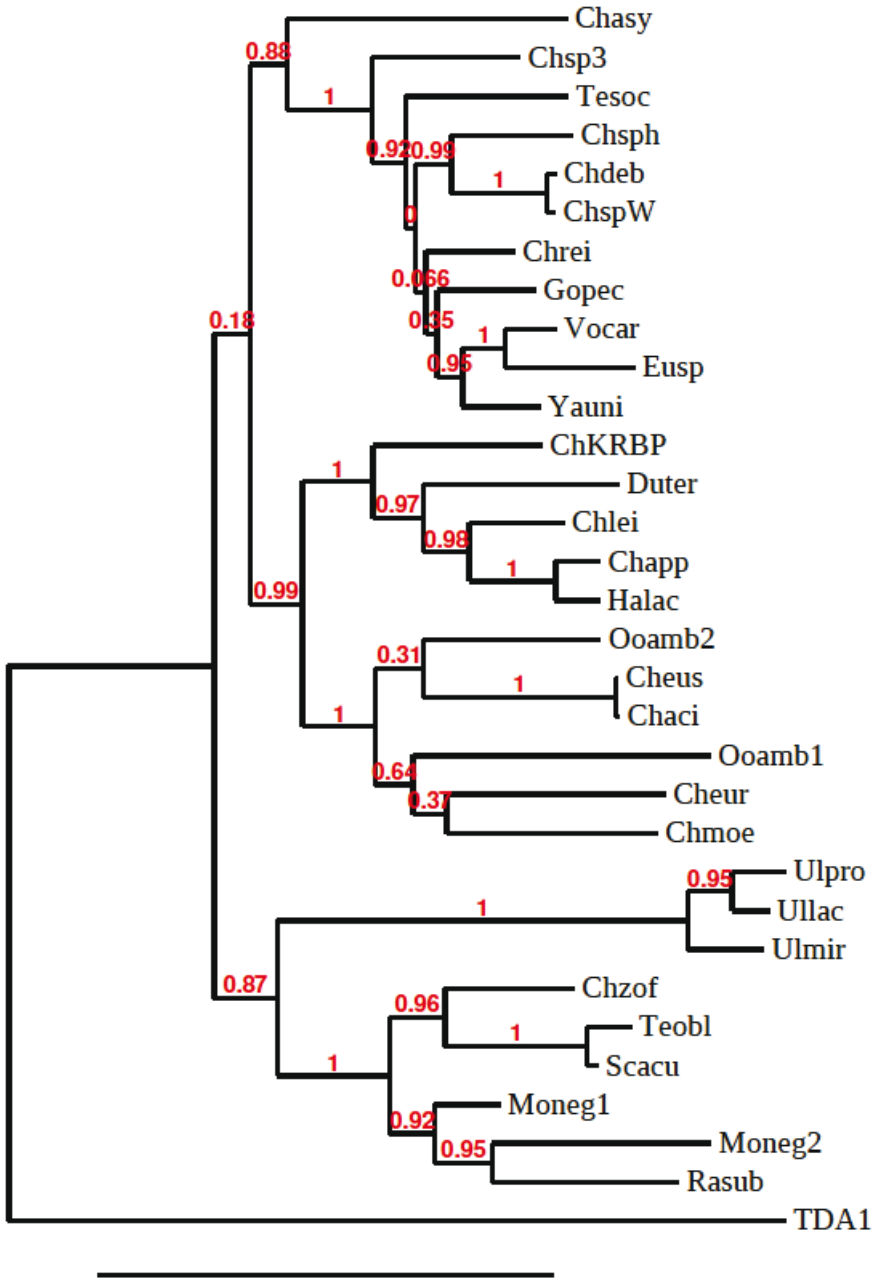

Ulmut
Ulpro
Ullac
Moneg2
Rasub 688 LSLHRVQQRAQVARHQIDQQRHSDAGDPPG
Teobl 679 GAAAAAAAVHSAAALQPVQPP
Scacu
Chzof
Moneg1 362 QQQQQQQQLQQPHRPQLQPPLGPGQR
Chasy 605 GHEHVQLAPLPPPRQPSSANSKPRPARLLSPQPHIAL
Chsp3 640 GRGGSGAAWASAGVNGYGAELRPPFAHPEVNGQAAQQQYLQQHPQQEG
Tesoc 559 QEEGLSGSRRGGAAAGLELPGGEGRQQQQPELGDGGLSSGVDRGGPGAAGGDADGLGPRASGSGPQMLYPRRGNHVLGPRPRGRPAPPVDDAATALPDAAAAEAAASSSS
Chsph 620 VELTLEDAAAIAAAAGAAADGEQPPPSAP
Chdeb 656 QH--PHPPRQAPSMLQLETNGEPVAAHQPPAAILDQDQTGHD
ChWS3 595 QHQHPHPPRQAPSMLQMETNGEPVAAHQPPAAILDQDQTGHD
Vocar 636 QLLRLQPQERPLQLHEGDDDD
Eusp 522 GQERRRHPVDAPTPLDQHAPLRVGIPPHH
Chrei 662 SHEAAAVHAAAQGEYRALQQPKPQPLAML
Gopec
Yauni 643 AEAVAEVALAADGSRQGRRPEQLLMLQPHGEQL
Cheur 848 THSVHAASSARVSLLEHVATP
ChKRBP 1131 RRQRTTLARERPRLFAMTD
Ooamb1 506
Duter 1052 GEEGVRGAERADGRWFARFVADGGELVPKPAGPSDDDGSSSSSSSSSSATVPVSPQQVDSRAWE----GPDSDGQGTGASARQGASQTEAPQQHGGQQGMSGPHEQQQQL
Chmoe
Chlei 822 IRAPRPLLWPEAREQELNGAAVNGV
Chapp 807 GEGEGEGELERAGVVRLDD
Halac 780 LTRPRAPEEGPLVPAGLAADTLQVPLHLIGEEFVDAPAGAQPFMERMVSNLGSPQRAASLGTPKALGSPASGRGGVRGPSVPPLPGAALSRMATAGVGTSAATGPATGAS
Ooamb2
Cheus
Chaci
consensus 1651

```
Ulmut          ----------------------------------------------------------------------------------------------------
Ulpro          ----------------------------------------------------------------------------------------------------
Ullac          ----------------------------------------------------------------------------------------------------
Moneg2         ----------------------------------------------------------------------------------------------------
Rasub       718 --PGAGPNSAAALATDARTLLQQRQQQQQQQQRQPPTVLRSNGRAYGAPAARLWGSG----------------------------------------
Teobl       742 QAPQLQQHLHKQQQVGSSSYSPQQQQQQQQQQQQQLRLSLLQPPRRMPQQQQHLPGVT----------------------------------------
Scacu          ----------------------------------------------------------------------------------------------------
Chzof          ----------------------------------------------------------------------------------------------------
Moneg1      388 --RPLPVSLSLHQWNHQAQLAQHQSRQQLHSQGGDGHPAAAAAAAAAAAAAGAAWRPS----------------------------------------
Chasy       684 QHPELGQPLGQQQAQEPVELQPPPFLAPFGSAARSAAQPPSAPHDAPPVGGVAWQPQQ----------------------------------------
Chsp3       730 AGAGAHSGQGQGQQLQQDTLRSHPLPHDDSSDGPVRVQLSVGEAQLLLEPQSQPQQQ----------------------------------------
Tesoc       669 SSNGAGVRRSLKEGAGGRAPRQRQQRQAEGEGPAPPLQRREDIRELSELDSPEGVAA----------------------------------------
Chsph       660 RIPAGQQPLGSRGAVSVSAAANYRQAAPAAAARSNGATNGLHGHAADELNGHHDGDA----------------------------------------
Chdeb       738 LLRPRPRGNPASARRATGSHRPIAPPPPDAHVTVPGAAVNGNGYNGASANGHGHGPA----------------------------------------
ChWS3       679 LLRPRPRGNPASARRATGSHRPIAPPPPDAHVTVPGAAANGNGYNGASANGHGHGPA----------------------------------------
Vocar       661 MAEGDAAGASSAATRTTETAATAAPVVPFLQITSSSGSGSADAADVAAITAAAAPQ----------------------------------------
Eusp        593 KAGRRAAAGAPTAAAAAQSAAADPQGEPVGSVQETQPQEQQGGRHRPHRHGAAPGA----------------------------------------
Chrei       733 ----------------AQLQESAPAAAALDGSNSGSNGAKVLSPRPRLGSARRGGPV----------------------------------------
Gopec          ----------------------------------------------------------------------------------------------------
Yauni       701 MLEADGGVGSRRGGHPSQHHHTRGSPAPGPGEPGLGRSVKMLTPRPRLRKQPAARPA----------------------------------------
Cheur       869 ----------------VSADKQGSVAEELCSLAGKPVMLLMGNMETLVFYSDNV----------------------------------------
ChKRBP     1150 ------------------------PSDGGDGSGAASSFNDGGSSSRSGGHHVMDLA----------------------------------------
Ooamb1      506 ------------------------------------------HLRGKRCWAPGPGPG----------------------------------------
Duter      1158 ESWPLHDVRRASLECGGAGITGPPLDSSVSVPPPDPASPSPHSSQQPCQSGTPHSPP----------------------------------------
Chmoe          ----------------------------------------------------------------------------------------------------
Chlei       856 LVRPRPVPLPLPSSNGVHTNGAAVQLDAAPDAKAGLRGVSFGRKAPPMLHVNPDGSSSTL----------------------------------------
Chapp       826 ------------------------VDGGVVDGGALKGVSFGRTAPPMLHVAGPGSV----------------------------------------
Halac       890 SPLAKGLGSPVWVGTGGGGDARAFADQGGGVVPVVADPHELAGMMRVRGAMRGEGEGLVEQPVVMGPKTRRAGGGRLYFRPMQGHMPRAVNAGGTGLSVGRVREGLARRM
Ooamb2         ----------------------------------------------------------------------------------------------------
Cheus          ----------------------------------------------------------------------------------------------------
Chaci          ----------------------------------------------------------------------------------------------------
consensus  1761
```

```
Ulmut          ----------------------------------------------------------------------------------------------------
Ulpro          ----------------------------------------------------------------------------------------------------
Ullac          ----------------------------------------------------------------------------------------------------
Moneg2         ----------------------------------------------------------------------------------------------------
Rasub      773 -EEPLERRALREAGSIDLL----------------------------------------------------------------------------------
Teobl      799 -SAAAAPAMLHNRPVATMPPAGNSSSSSFS------------------------------------------------------------------------
Scacu          ----------------------------------------------------------------------------------------------------
Chzof          ----------------------------------------------------------------------------------------------------
Moneg1     443 -RKVARMLESVGYGDDDVPNALDPSVEPVG------------------------------------------------------------------------
Chasy      742 -QQQQAELVVTSSPSTLLGLRGHPVGSGNAG------------------------------------------------------------------------
Chsp3      787 -QQEQVGSDLGGHHGQDNAVLQAAQGVQIP--------------------------------------------------------------------------
Tesoc      726 -GDGGGGLPPPPLGLSRGPTANGSGEVVIV--------------------------------------------------------------------------
Chsph      717 ----------AASSGPSASLQQLAQGMEVP--------------------------------------------------------------------------
Chdeb      795 -VNGISAVLPTTNGLDAVSVGIEPQGVEVP--------------------------------------------------------------------------
ChWS3      736 -VNGISAVLTTTNGLDAVSAGIEPQGVEVP--------------------------------------------------------------------------
Vocar          ----------------------------------------------------------------------------------------------------
Eusp       649 -RSSCLPGPPPPLPLQPLGVLTNGNGN------------------------------------------------------------------------
Chrei      774 -VAGDASPKGASAHVAVPVDSAAPSGARAR----------------------------------------------------------------------
Gopec          ----------------------------------------------------------------------------------------------------
Yauni      758 -GHALPPEDISEQQVDAADEVAMPEAQALQ----------------------------------------------------------------------
Cheur          ----------------------------------------------------------------------------------------------------
ChKRBP    1182 -GGGVAPTAHPHHQATAAEAMGAAAAP----------------------------------------------------------------------
Ooamb1     521 -AATCSS-------------------------------------------------------------------------------------------
Duter     1215 -LSSLHTSIPSSRAKSSDSSSSTSSSARRS-------------------------------------------------------------------
Chmoe          ----------------------------------------------------------------------------------------------------
Chlei          ----------------------------------------------------------------------------------------------------
Chapp      858 -A----------------------------------------------------------------------------------------------------
Halac     1000 IAAELHTLGPQPFKRKYEPRGATPEGVKLDGFLLLEAGDDELPEAIAAARVVGRHVASVEVADMAYFQGLRVLDVSDNRLESMEGLAPLGGLRALSLSLNRLQGPGTWDP
Ooamb2         ----------------------------------------------------------------------------------------------------
Cheus          ----------------------------------------------------------------------------------------------------
Chaci          ----------------------------------------------------------------------------------------------------
consensus 1871
```


Ulmut
Ulpro
Ullac
Moneg2
Rasub
Teobl 828 SAGSSGGHGSSR
Scacu
Chzof
Moneg1 472 KGR
Chasy 772 SSHPLDGFAQQQGTCVSDARAVSVAEQLAGGGTFVLGLDPQEQQLLEAGDLAVQQQDQQQQQEGAVPGP
Chsp3 816 AGSPLNGWAAGGGGGGGMDVSAIPRLTLERRPSRQGLPPPAPIPVVRRAP
Tesoc 755 GGSSGGGGGEGLGVNGTSVRVAVPLLPLDGRPGSADPPPPLPTASTRGAGAR
Chsph 737 PESPLGLALNGNGGAPRVVMLERPPFTSPPPPAATAHAAR
Chdeb 824 PGSPLGLALNGSGLNGTAPPSSSPLSAAASQRAW
ChWS3 765 PGSPLGLALNGSGLNGTAPPSSSPLSAAASQRAW
Vocar
Eusp 675 GNGNGSLNGSGRPEAPFLALER
Chrei 803 ALFSDPRRDSPYNVGMVAATPLTFQR
Gopec
Yauni 787 QGVELDAAGLPLNGNGTPAVPHMVPLERRPSGTVAGGIPPTTTTRRTR
Cheur
ChKRBP
Ooamb1
Duter 1244 RSSSTSVLDLRAELGADNPLPSPPFSSPSLLARPIATSSRDEDAAPSATSSRDEDAANEDATSLHDADA
Chmoe
Chlei
Chapp
Halac 1110 TLAFQQLATLDLSHNALDATATMGPDSPLAALPQRLLLAGNPLAAAGLQAAKRAAQKRAKAAAMAADGALLAPGTLELPPPDPLLVLEPVPAQGPKPTVASVLAGGGGRF
Ooamb2
Cheus
Chaci
consensus 1981

```
Ulmut          --------------------------------------------------
Ulpro          --------------------------------------------------
Ullac          --------------------------------------------------
Moneg2         --------------------------------------------------
Rasub          --------------------------------------------------
Teobl          --------------------------------------------------
Scacu          --------------------------------------------------
Chzof          --------------------------------------------------
Moneg1         --------------------------------------------------
Chasy      841 TNGVVKVRGVSCTPVESVAGTHNIDEQHQTSF----------------------
Chsp3          --------------------------------------------------
Tesoc          --------------------------------------------------
Chsph          --------------------------------------------------
Chdeb          --------------------------------------------------
ChWS3          --------------------------------------------------
Vocar          --------------------------------------------------
Eusp           --------------------------------------------------
Chrei          --------------------------------------------------
Gopec          --------------------------------------------------
Yauni          --------------------------------------------------
Cheur          --------------------------------------------------
ChKRBP         --------------------------------------------------
Ooamb1         --------------------------------------------------
Duter     1313 AAEPCAMKIPHSRVMRMRRDEDATSLHGEGAKSPLPLPAAQRPQPGTTFCYFLM
Chmoe          --------------------------------------------------
Chlei          --------------------------------------------------
Chapp          --------------------------------------------------
Halac     1220 VHIHEPLGRRGVEKAATVAAVAAVPEEVSSAAASHNPKP---------------
Ooamb2         --------------------------------------------------
Cheus          --------------------------------------------------
Chaci          --------------------------------------------------
consensus 2091
```


Chasy
Chsp3
Tesoc
Chsph
Chdeb
ChspW
Chrei
Gopec
Vocar
Eusp
Yauni
ChKRBP
Duter
Chlei
Chapp
Halac
Ooamb2
Cheus
Chaci
Ooamb1
Cheur
Chmoe
Ulpro
Ullac
Ulmir
Chzof
Teobl
Scacu
Moneg1
Moneg2
Rasub
TDA1
0.88
1
0.92
0.99
1
0
0.066
0.18
0.35
1
0.95
1
0.97
0.98
1
0.99
0.31
1
1
0.64
0.37
0.95
1
0.87
0.96
1
1
0.92
0.95

3.

Supplemental Fig. S6: Conservation of MTHI1 among green algae.

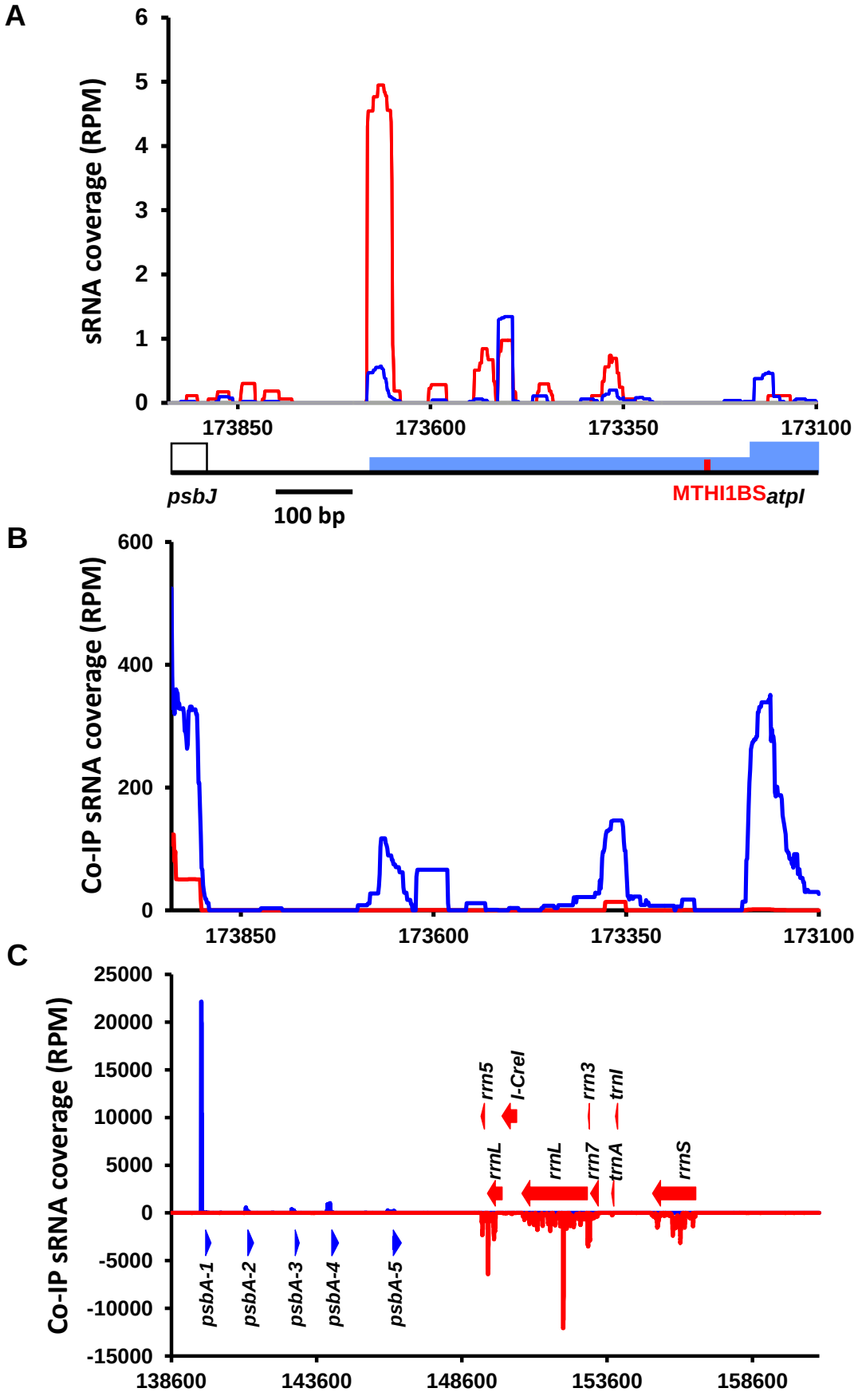
A) DNA sequences encoding MTHI1 orthologues were retrieved from the JGI phytozome (v12), the NCBI databases, and the MMETSP re-assemblies database (Keeling et al., 2014) by tBLASTn using *Cr*MTHI1 as a query. Gene models were then predicted using the Greengenie2 software (http://stormo.wustl.edu/GreenGenie2/; (Kwan et al., 2009)) and manually edited to include obvious missing regions of similarity, if required. Alignment of MTHI1 orthologues performed with the MUSCLE software using default options and manually edited to improve the alignment. OPR repeats of the protein from *C. reinhardtii* are shown above the alignment. Residues conserved in more than half of the sequences are written in red, while conservative substitutions are written in blue.

Abbreviations of species names are as follows:

Ulmut Ulva mutabilis; Ulpro: Ulva prolifera; ULLac: Ulva lactuca; Moneg: Monoraphidium neglectum; Rasub: Raphidocelis subcapitata; Teobl: Tetradesmus obliquus; Scqua: Scenedesmus acutus; Scqua: Scenedesmus quadricauca; Chzof: Chromochloris zofingiensis; Chasy: Chlamydomonas asymetrica; Chsp3: Chlamydomonas sp. 32112; Tesoc: Tetrabaena socialis; Chsph: Chlamydomonas sphaeroides; Chdeb: Chlamydomonas debaryana; ChspW: Chlamydomonas sp. WS3; Vocar: Volvox carteri; EuspN: Eudorina sp.; Chrei: Chlamydomonas reinhardtii; Gopec: Gonium pectorale; Yauni: Yamagishiella unicocca; Cheur: Chlamydomonas euryale; ChKRBP: ; Ooamb: Oophila amblystomatis; Duter: Dunaliella tertiolecta; Chmoe: Chlamydomonas moewusi; Chlei: Chlamydomonas leiostraca; Chapp: Chlamydomonas applanata; Halac: Haematococcus lacustris; Cheus: Chlamydomonas eustigma; Chaci: Chlamydomonas acidophila.

B) Phylogeny of MTHI1 orthologues.

The phylogenic tree was constructed on the Phylogeny.fr platform (Dereeper et al., 2008), including the following steps: Sequences were aligned with MUSCLE (v3.7) (Edgar, 2004) configured for highest accuracy (MUSCLE with default settings) and cured with Gblocks (v0.91b), using relaxed parameters. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program (v3.0) (Guindon et al., 2010) with reliability for internal branch assessed using the aLRT test (SH-Like) (Anisimova and Gascuel, 2006) and visualized using TreeDyn (v198.3) (Chevenet et al., 2006). The OPR protein TDA1 (Eberhard et al., 2011) was taken as outgroup.



Ozawa et al, Suppl. Fig. S7

Supplemental Fig. S7: sRNAs coverage (normalised as RPM) over the *atpI* 5'UTR and along the inverted repeat (in support of Figs. 10 and 11).

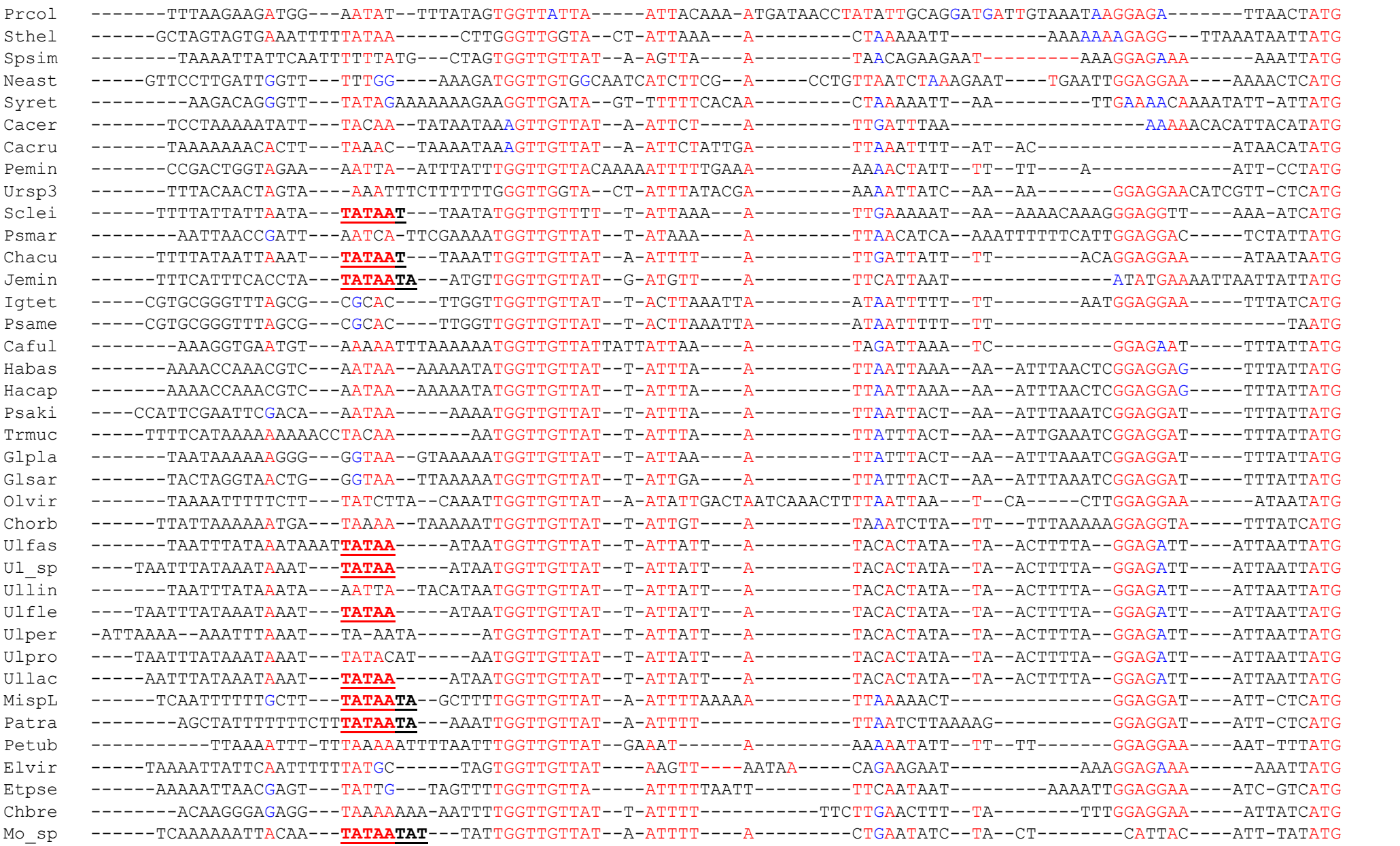
A) Coverage of sRNAs mapping to the coding strand along the *atpI* 5'UTR, schematically depicted at the bottom. The red bar symbolises the putative MTHI1 binding site. Blue line: RPP-treated wild-type sample; red line: RPP-treated *mthi1-1* sample.

B) Coverage of sRNAs immuno-precipitated with the MTHI1 protein over the *atpI* 5'UTR. Red line: sRNA coverage in the wild-type negative control; blue line sRNA coverage in the MTHI1-RIP sample. All samples were RPP-treated. Note the absence of coverage in correspondence of the MTHI1 putative binding site.

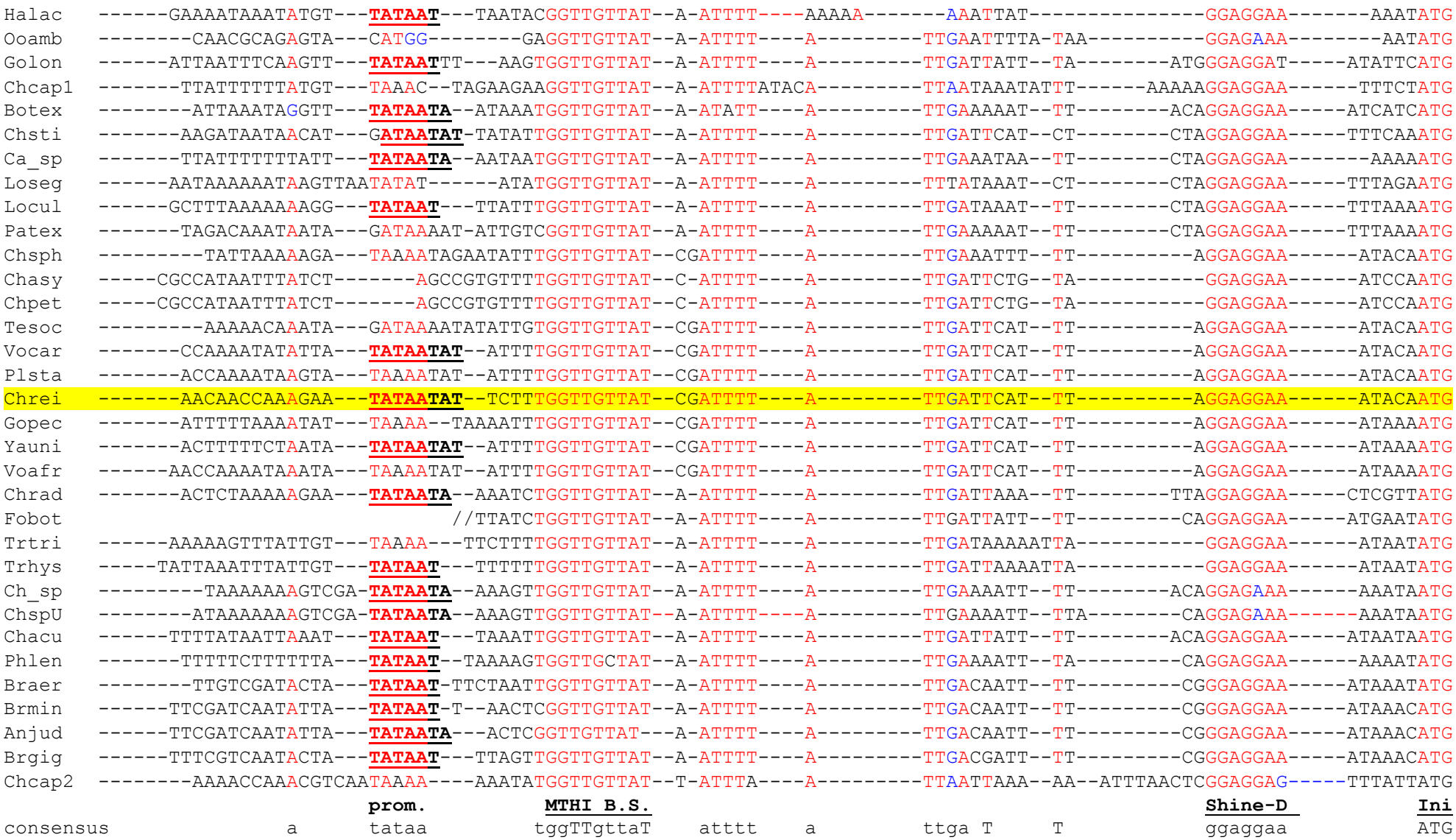
C) sRNA coverage of RPP-treated MTHI1 sample within the inverted repeat. Blue line: sRNAs mapping to the plus strand; Red line: sRNAs mapping to the minus strand. The position of genes within the inverted repeat is shown. The blue peak corresponds to the *psbA* 5'UTR.

Ozawa et al, Suppl. Fig. S8

A Alignment of *atpH* 5'UTRs




Chniv -------TTTTTTTTTTTTG---TATAAT--TGATTTTGGTTGTTAT--A-ATTTT--------------TAGATTTAATATT--------AAAGAAGGAA----ATTTTTTATG
Chros ------TCTTTTTTTTTGTG---TATAAT--TATTTTTGGTTGTTAT--A-ATTTT--------------TAGATTTAATATT--------AAAGAAGGAA-----TTTTTTATG
Halae -------CTTTAAAATAATA---TAAAA--TTAGGGTAGGTTGTTAT--A-ATTTTAGA-A---------TCAAGCGTA----------------AAAACAA---------AATATG
Rasub -------CTTTTTTTCGTTT---TATAATA--ATCTTTGGTTGTTAT--A-ATTTT----A---------CTGATTATA--TA-----------GGAGGAA-----ATCTAAATG
Moneg ---------TTTTATTTTTT-TATATAAT--TAAATTTGGTTGTTAT--A-ATTTT----A---------CTGATTATA--TA-----------GGAGGAA-----ATCTCAATG
Scqua -------TTTTATTTTTTTA---TATAAT--TAAATTTGGTTGTTAT--A-ATTTT----A---------CTGATTATA--TA-----------GGAGGAA-----ATCTCAATG
Kiape -------ATATATATATATA---TATAAT--TCTATTTGGTTGTTAT--A-ATTTT----A---------CTGATTATA--AA-----------GGAGGAA----ATC-TCAATG
Oeang ------ATTATTTTTTATGT---TATAATA--ATTAATGGTTGTTAT--T-ATTTT----A---------TTACATTTT------AAATATATAGGAGGAA-----ATCTAAATG
Oecard -------GTCTTTTTTATGT---TATAATA--ATTAATGGTTGTTAT--T-ATTTT----A---------TTACTTTT------AAATATATAGGAGGAA----ATC-TAAATG
Oecaro -------TATTTTTTTATGT---TATAATA--ATTAATGGTTGTTAT--T-ATTTT----A---------TTACTTTT--TA----AAATATAGGAGGAA-----ATCTAAATG
Oogig ------TTTGGTCTTTATAA--TTATGATA----ATTATGTTGTTGT--A-ATTTT----A---------TTGATATTT--TT--T-----A--GGAGGAA----TTATAATATG
Myhom ------TCTTATCTCAAGTAAAATAAAAGTT------GAGTTGTTAT--A-ACTAT----A---------TTGATTATT--TT---------TAGGAGGAA------TTAAAATG
Myjur -------CGCTGTTTCAT-A---TAAACTCT-TTAATGAGTTGTTAT--A-ACTAT----A---------TTGATTATT--TT---------TAGGAGGAA------TTAAAATG
Chper ------ATAATATAAAATGT---TATAAT---TTTATTGGTTGTTAT--A-ATTTT----A---------TTGATTTTTTCTT---------ATGGAGGAA------TATATATG
Chtat -----ATAATATAAAAATGT---TATAAT---TTGTTTGGTTGTTAT--A-ATTTT----A---------TTGATTTTTTCTT---------ATGGAGGAA------TATATATG
Prbot ------AAAGAAATAAATGT---TATAATAT--TTTTTGGTTGTTAT--A-ATTTT----A---------TTGATTTTTTCTT---------ATGGAGGAA------TATATATG
Stplu ------TAAAAATATAATGT---TATATT---TATGTTGGTTGTTAT--A-ATTTT----A---------TTGATTTTT--TC-----------GGAGGAA------TATATATG
Chere -----AAAAAATAAAAATGT---TATAATAC---ATTTGGTTGTTAT--A-ATTTT----A---------TTGATTTTT--TT---------ATGGAGGAA------TTATTATG
Chlei -------AATAAAAATATAG---TATAAT--GAACAATGGTTGTTAT--A-ATTTT----A---------TTGATATTA--TT---------ATGGAGGAA------TATTTATG
Chapp -------CATGAATGTATAG---TATAATAAT--AAATGGTTGTTAT--A-ATTTT----A---------TTGATCTTT--TT---------ATGGAGGAA------TTTTTATG
Dusal ------TTTATTTTATATGT---TAAAC---TATATTTGGTTGTTAT--A-ATTTT----A---------TTGATCTTA--TT---------ATGGAGGAA------TTTTTATG
DusaS ------TTTATTTTATATGT---TAAAC---TATATTTGGTTGTTAT--A-ATTTT----A---------TTGATCTTA--TT---------ATGGAGGAA------TTTTTATG
Cosp1 ------TCAAAAAATTCCAA---TATAATA---GAACTGGTTGTTAT----AATTT----A---------CTGATTATAAATA----------------------TATAAATATG
Cosp2 ------TTACCAAATGAAAA---TAAAATA---ATACAGGTTGTTAT--A-ACTTT----A---------CTGATTAAT--TT-------------AAAAA-------TAATATG
Cosai -------TAAAACACACGCA---AATAAAAT-GGTACTGGTTGTTAT--A-ACTTT----ACTGA-----TTAATTTTA-------------------------AATAAATATG
Teobl ------TAAAAAAAATATAA---TATAATAT---TAATGGTTGTTAT--A-ATTTT----A---------CTGATTATTATTC----------GGAGGAA-----ATAAATATG
Haret ------TGAATAAAAAATTA---TATAATAT---TTATGGTTGTTAT--A-ATTTT----A---------CTGATTATTTTTC----------GGAGGAA-----ATACATATG
HaspM -----TTGAATAAAAAATTA---TATAATAT---TTATGGTTGTTAT--A-ATTTT----A---------CTGATTATT--TT---------TCGGAGGAA-----ATACATATG
Pepec ------TTAGAAAAATATAA---TATAATA---GAATTGGTTGTTAT--A-ATTTT----A---------CTGATTATT--TT---------TCGGAGGAA----ATA-AATATG
Neaqu ------TTAATAATATATGA---TATAATAT---TTTTAGTTGTTAT--A-ATTTT----A---------CTGATTTTT--TA--------CTAGGAGGAA-----ACATATATG
Chinc -------AATAAAAGTTTTG---TTAAA--CTTCTTTTGATTGTTAT--A-ATTTT----A---------CTGATTATT--TA--------CAAGGAGGAA-----ATTCATATG
Peang ------AAAAACATTTAAGT---TATACT---TGAAGCGGTTGTTAT--A-ATTTT----A---------CTGATTTTT--TA--------CTAGGAGGAA-----ATTCATATG
Hyret ------AAAAACATTTAAGT---TATACT--TTAAAGCGGTTGTTAT--A-ATTTT----A---------CTGATTTTT--TA--------CTAGGAGGAA-----ATTCATATG
Pedup ------AAAAACATTTAAGT---TATACT---TGAAGCGGTTGTTAT--A-ATTTT----A---------CTGATTTTT--TA--------CTAGGAGGAA-----ATTCATATG
Sttet ------AAAACCATTTAAGT---TAAAAT---GAAAACGGTTGTTAT--A-ATTTT----A---------CTGATTTTT--TA--------CTAGGAGGAA-----ATTCATATG
Lagra ------AAAAAGATTTAAGT---TATACT---TAAAGCGGTTGTTAT--A-ATTTT----A---------CTGATTTTT--TA--------CTAGGAGGAA-----ATTCATATG
Pssch ------AACCTTAAATAAGC---TAAAA---TTAAAATGGTTGTTAT--A-ATTTT----A---------TTGATTATT-----------AACGGAAGGA----AACAAACATG
Ps-sp ------AAAAACATTTAAGT---TATACT---TAAAGTGGTTGTTAT--A-ATTTT----A---------CTGATTTTT--TA--------CTAGGAGGAA-----ATTCATATG
Psint ------AAAAACATTTAAGT---TATACT---TAAAGCGGTTGTTAT--A-ATTTT----A---------CTGATTTTT--TA--------CTAGGAGGAA-----ATTCATATG
Psbor ------AATCAGTAATATTA---TATAATA---AAATTGGTTGTTAT--A-ATTTT----A---------CTGATTTTT--TA--------CTAGGAGGAA-----ATTCATATG
PsspC -----AAAAAACATTTAAGT---TATACT---TAAAGTGGTTGTTAT--A-ATTTT----A---------CTGATTTTT--TA--------CTAGGAGGAA----ATT-CATATG
Chzof ---------TAAAAACAATTTGTTAAAAT---TATAATGGTTGTTAT--A-ACTTT----A---------TTGAATATT--TA--------ACCGGAGGAA-----ACAAATATG

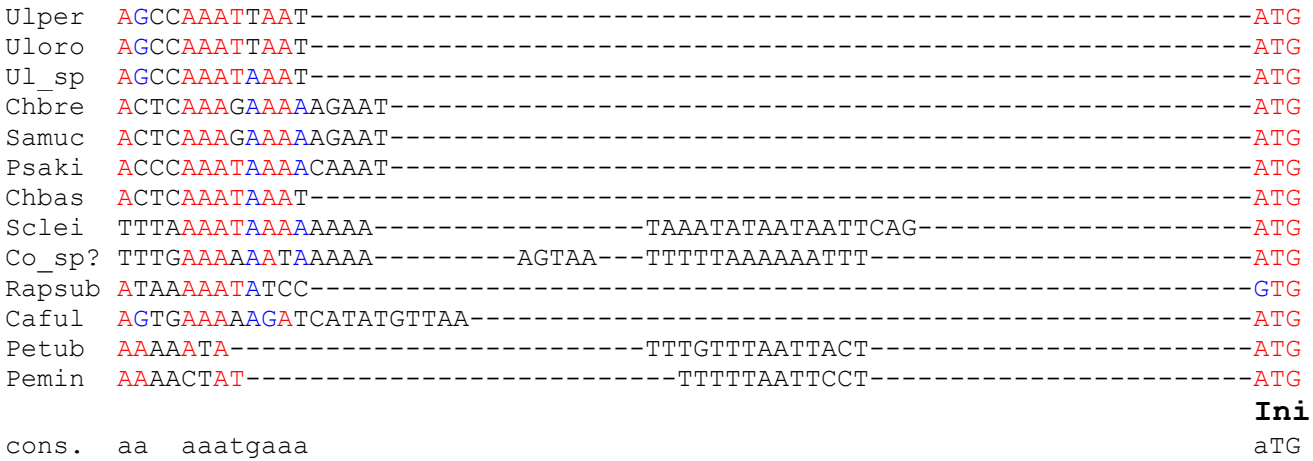

Halac ------GAAAATAAATATGT---TATAAT---TAATACGGTTGTTAT--A-ATTTT----AAAAA-------AAATTAT---------------GGAGGAA-------AAATATG
Ooamb --------CAACGCAGAGTA---CATGG---------GAGGTTGTTAT--A-ATTTT-----A---------TTGAATTTTA-TAA----------GGAGAAA--------AATATG
Golon ------ATTAATTTCAAGTT---TATAATTT---AAGTGGTTGTTAT--A-ATTTT----A---------TTGATTATT--TA---------ATGGAGGAT-----ATATTCATG
Chcap1 -------TTATTTTTTATGT---TAAAC--TAGAAGAAGGTTGTTAT--A-ATTTTATACA---------TTAATAAATATTT------AAAAAGGAGGAA------TTTCTATG
Botex --------ATTAAATAGGTT---TATAATA--ATAAATGGTTGTTAT--A-ATATT----A---------TTGAAAAAT--TT--------ACAGGAGGAA-----ATCATCATG
Chsti -------AAGATAATAACAT---GATAATAT-TATATTGGTTGTTAT--A-ATTTT----A---------TTGATTCAT--CT--------CTAGGAGGAA-----TTTCAAATG
Ca_sp -------TTATTTTTTTATT---TATAATA--AATAATGGTTGTTAT--A-ATTTT----A---------TTGAAATAA--TT--------CTAGGAGGAA-------AAAAATG
Loseg ------AATAAAAAATAAGTTAATATAT------ATATGGTTGTTAT--A-ATTTT----A---------TTTATAAAT--CT--------CTAGGAGGAA-----TTTAGAATG
Locul ------GCTTTAAAAAAAGG---TATAAT---TTATTTGGTTGTTAT--A-ATTTT----A---------TTGATAAAT--TT--------CTAGGAGGAA-----TTTAAAATG
Patex -------TAGACAAATAATA---GATAAAAT-ATTGTCGGTTGTTAT--A-ATTTT----A---------TTGAAAAAT--TT--------CTAGGAGGAA-----TTTAAAATG
Chsph ---------TATTAAAAAGA---TAAAATAGAATATTTGGTTGTTAT--CGATTTT----A---------TTGAAATTT--TT----------AGGAGGAA------ATACAATG
Chasy -----CGCCATAATTTATCT-------AGCCGTGTTTTGGTTGTTAT--C-ATTTT----A---------TTGATTCTG--TA----------GGAGGAA------ATCCAATG
Chpet -----CGCCATAATTTATCT-------AGCCGTGTTTTGGTTGTTAT--C-ATTTT----A---------TTGATTCTG--TA----------GGAGGAA------ATCCAATG
Tesoc ---------AAAAACAAATA---GATAAAATATATTGTGGTTGTTAT--CGATTTT----A---------TTGATTCAT--TT----------AGGAGGAA------ATACAATG
Vocar -------CCAAAATATATTA---TATAATAT--ATTTTGGTTGTTAT--CGATTTT----A---------TTGATTCAT--TT----------AGGAGGAA------ATACAATG
Plsta -------ACCAAAATAAGTA---TAAAATAT--ATTTTGGTTGTTAT--CGATTTT----A---------TTGATTCAT--TT----------AGGAGGAA------ATACAATG
Chrei -------AACAACCAAAGAA---TATAATAT--TCTTTGGTTGTTAT--CGATTTT----A---------TTGATTCAT--TT----------AGGAGGAA------ATACAATG
Gopec -------ATTTTTAAAATAT---TAAAA--TAAAATTTGGTTGTTAT--CGATTTT----A---------TTGATTCAT--TT----------AGGAGGAA------ATAAAATG
Yauni -------ACTTTTTCTAATA---TATAATAT--ATTTTGGTTGTTAT--CGATTTT----A---------TTGATTCAT--TT----------AGGAGGAA------ATAAAATG
Voafr ------AACCAAAATAAATA---TAAAATAT--ATTTTGGTTGTTAT--CGATTTT----A---------TTGATTCAT--TT----------AGGAGGAA------ATAAAATG
Chrad -------ACTCTAAAAAGAA---TATAATA--AAATCTGGTTGTTAT--A-ATTTT----A---------TTGATTAAA--TT--------TTAGGAGGAA-----CTCGTTATG
Fobot   //TTATCTGGTTGTTAT--A-ATTTT----A---------TTGATTATT--TT---------CAGGAGGAA-----ATGAATATG
Trtri ------AAAAAGTTTATTGT---TAAAA---TTCTTTTGGTTGTTAT--A-ATTTT----A---------TTGATAAAAATTA----------GGAGGAA------ATAATATG
Trhys -----TATTAAATTTATTGT---TATAAT---TTTTTTGGTTGTTAT--A-ATTTT----A---------TTGATTAAAATTA----------GGAGGAA------ATAATATG
Ch_sp ---------TAAAAAAAGTCGA-TATAATA--AAAGTTGGTTGTTAT--A-ATTTT----A---------TTGAAAATT--TT--------ACAGGAGAAA-----AAATAATG
ChspU --------ATAAAAAAAGTCGA-TATAATA--AAAGTTGGTTGTTAT--A-ATTTT----A---------TTGAAAATT--TTA-------CAGGAGAAA------AAATAATG
Chacu ------TTTTATAATTAAAT---TATAAT---TAAATTGGTTGTTAT--A-ATTTT----A---------TTGATTATT--TT--------ACAGGAGGAA-----ATAATAATG
Phlen -------TTTTTCTTTTTTA---TATAAT--TAAAAGTGGTTGCTAT--A-ATTTT----A---------TTGAAAATT--TA---------CAGGAGGAA------AAAATATG
Braer --------TTGTCGATACTA---TATAAT-TTCTAATTGGTTGTTAT--A-ATTTT----A---------TTGACAATT--TT---------CGGGAGGAA-----ATAAATATG
Brmin ------TTCGATCAATATTA---TATAAT-T--AACTCGGTTGTTAT--A-ATTTT----A---------TTGACAATT--TT---------CGGGAGGAA-----ATAAACATG
Anjud ------TTCGATCAATATTA---TATAATA---ACTCGGTTGTTAT---A-ATTTT----A---------TTGACAATT--TT---------CGGGAGGAA-----ATAAACATG
Brgig ------TTTCGTCAATACTA---TATAAT---TTAGTTGGTTGTTAT--A-ATTTT----A---------TTGACGATT--TT---------CGGGAGGAA-----ATAAACATG
Chcap2 --------AAAACCAAACGTCAATAAAA----AAATATGGTTGTTAT--T-ATTTA----A---------TTAATTAAA--AA--ATTTAACTCGGAGGAG-----TTTATTATG
prom.
MTHI B.S.
Shine-D
Ini
consensus a tataa tggTTgttaT atttt a ttga T T ggaggaa ATG

B Alignment of *atpI* 5'UTRs

```
Bobra   GAGCGTGCAAAAATTGACTCTTC--AGGTTATAAT-----------ATACCT-------------------------------------------------------CTTTGCGGA
Halac   AAAAAAGTAAAGAAAAACTAAAG--GGGTAATTAA--AGAACATAAAAGTAGACAATTACA-----------------------------------------------ATTTTTAAG
ChspU   AAAAAAGTCGATATAATAAAGT--TGGTTGTTAT-----------AATTTTATTGAAAA-----------------------------------------------TTTTACAGG
Rorot   AATGCAGTGAAATTTTTTTTTAA--TGGTTGTTAT-----------------------------------------------------------------------TCTTATTAA
Haret   -----TGAAAAATGTTGAACAAT--TGGTTATTAT---------TCAAATTCAACATTTT-----------------------------------------------TCACTTAAA
Ha_sp   AGAAACTTAAGAATATTTCTTTT--TGGTTGTTAAT--------TGAAATTCAACATTTT-----------------------------------------------TCACTTAAA
Chrad   AGTCTTTTTAAAAGTTTTAGTAG--TAGTTTTTAT---------A-AGAAGCTTTTAGTAGCAGCGTAGCTGGATAAAAAGAAACTTTATTACTTTAAAAAAACAGTTTTTTTTAT
Chsp3   GGGAACTGATTTTAATTTTTAGA--TTGTTATTAT---------TAAAAAATATTTTTA-----------------------------------------------TTTTGCCAA
Chsph   CTTATATTTCAAATGAGTGATAA--AGGTTATTAT---------TAAAATACATTTCTC-----------------------------------------------TCTTTCCAA
Tesoc   CAGGTAAACATATTTTTTTTATT--TGGTTTTTAT---------A-AAATAGATTTGTAT-----------------------------------------------CTTTCCAAC
Chdeb   AAAAGACTGTAAAAAATTTTTTA--TGGTTATTAT---------T-AAAAATATTTTTC-------------------------------------------------TCTTTCCAA
ChspW   AAAAGACTGTAAAAAATTTTTTA--TGGTTATTAT---------T-AAAAATATTTTTC-------------------------------------------------TCTTTCCAA
Gopec   AATTTAATACAAATGCAATCTAA--TCGTTTTTAT---------T-AAATACATTTTTA-----------------------------------------------TCTTTCCAA
Chrei   TAATAACTGGTCATTATTTATAG--TGGTTATTAT---------T-AAAAATATTAATA-----------------------------------------------TCTTTCCAA
EuspN   TTTATATTTAGTACTTAAGGCAT--TGGTTTTTAT---------AAAATATATTTAATATCTTTCCAATAAT-------------------------------TTTGGTAAG
Plsta   TTTTTATTGAACATATAAGCGAC--TGGTTATTAT---------T-AAATATATTTCATAT-------------------------------------------CTTTCCAAT
Vocar   TTTTTATTTGGTACCTAAGCCAC--CGGTTATTAT---------T-AAATATATTTAGTAT-------------------------------------------TTTTCCAAT
Voafr   TTTTTATTTGGTACCTAAACCAC--CGGTTATTAT---------T-AAATATATATACGAT-------------------------------------------TTTTCCAAT
Teobl   AAAATTCAAAAAAAGTAATTTTT--TGGTTGTTAA---------T-AAAGAATAAACCTTTTTATTTCAATAATTTTTGAAATGTAAAGC-------------TTATTCAAT
Chniv   AAAATTTAAACTAGTTTCTTTTT--TGTTTATTAA-----------AAGTTAAAAAAGTCTAATTAAAAAAAAACAGTTTT-------------------------CTTTTATAG
Difra   GATTTTTTTTCACAATAAATTAT--TGGTTGTTAA---------AAAAATGGATTGTTGCTGTTAAAATAAAAATTTTCAGCAATCTCAT---------------TTTCGTAAA
Chatm   -TTCTTAGATTAAGAACAGAGAG-TGTGTTATTAT-----------------------------------------------------------------------TGGTTAAGC
Neaqu   --TAGTGACCTGATTTCCTTTTTTGAAGTTGTTAA-----------TTTTCTATTTTCTTTTCTAT-----------------------------------------TATTATAGA
Cacer   CTAAAAATATTTACAATATAATA--AAGTTGTTAT-----------AATTCTATT-------------------------------------------------GATTTAAAA
Codec   CAGAGAGTTGCATTTGGATCGGA--GAGTTGTTATTATAGAAAATAAAAAATTCT-------------------------------------------------TATTTCAA-
Codsp   -AAAGTTTTGAATTTAGTTCAGA--TAGTTGTTATTATAGAAAATAAAAT-----------------------------------------------------TATCAGAAA
Pepec   AATGAAAAAAAGTAATTTTATTT--TGGTTGTTAA---------TAAAGAATAAATTTTCTCTTCTCAAAGAAATATTGAAGGGGTCGATTTATTCAA---------TTTTTAAAA
Kiape   TTAAATAATTTACAAATAAAACC--AGGTTGTTAA---------AAAACAAACAATTTTTTAACAGAAGAT-----------------------------------CTGTTTGTG
MispL   ATGGTAATTTTTTATTAAGCAATG--GGGTTGTCAATTCAATTGGA-AAAAATGCC-----------------------------------------------TTTTATGGG
Moneg   TTTTTTAATAAATATTAAAAAAA--TGGTTGTTAA-----------AAGGTAAACAATTTTTTATAAAT-------------------------------------TTGTTTATA
Scqua   TTTTTTAATAAATATTAAAAAAA--TGGTTGTTAA-----------AAGGTAAACAATTTTTTATAAAT-------------------------------------TTGTTTATA
Glpla   ATTTAAAAAATAAATTTGCTTTA--TGGTTGTTAA-----------AAAGTTTTTAAAAAA-------------------------------------------TTTTCAAAA
Psmar   AATAATTCAAAAAATGTTTGATA--TGGTTGTTAA-----------AAAG---------------------------------------------------TTTTAAAAA
Ulfas   AAAAATATGAAATAAAACTAAAA--TGGTTGTTAA-----------AAAGGTAAT-------------------------------------------------TTTTATAAA
Ulfle   AAAAATATGAAATAAAACTAAAA--TGGTTGTTAA-----------AAAGGTAAT-------------------------------------------------TTTTATAAA
Ullin   AAATATGAAATAAAAATAGAAAA--TGGTTGTTAA-----------AAAGGTAAT-------------------------------------------------TTTTATAAA
Ulper   AAAAATATGAAATAAAACTAAAA--TGGTTGTTAA-----------AAAGGTAAT-------------------------------------------------TTTTATAAA
Uloro   AAATTTTAAATAAAAATAGAAAA--GTGTTGTTAA-----------AAAGGTAAT-------------------------------------------------TTTTATAAA
Ul_sp.  AACAAGGGGTTTAAAAAAAAAAA--TGGTTGTTAA-----------AAAGGTAAT-------------------------------------------------TTTTATAAA
Chbre   CCGTTTCGGGACAAAGCGTATTA--TGGTTGTTAA-----------AAAGTTTTTAATAAT-----------------------------------------------TTTTAGAAA
Samuc   CCGTTTCGGGACAAAGCGTATTA--TGGTTGTTAA-----------AAAGTTTTTAATAAT-----------------------------------------------TTTTAGAAA
Psaki   TTTAAAAAATTAAGCTACAATTA--TGGTTGTTAA-----------AAAGTTTTTAATAAT-----------------------------------------------TTTTAGAAA
Chbas   GTTTCTACCACTAAAAAAAATTA--TGGTTGTTAA---------A-AAGTTTTTACTAAT-----------------------------------------------TTTTAGAAA
```

```
Sclei   TGCTTTTTCTCTAAGAAAAAGTT--TAGTTGTTTT-----------AAAATTTTTCCCTTTGTTGAGTGAAAAATTATTAAAAAATTCACTTGTA----------TTTAATAAG
Co_sp   TTTGCTTTTTAAATCAAATAAAA--TCGTTGTTAA---------T-AACTGTATAAAAATAGTTGTTAAACATATGTTTGTAA----------------------TTTGTTCAA
Rasub   AAAAATTTCAGAAATAAAAAAAT--TGGTTGTTAA---------A-AAACGGATA-------------------------------------------------GTTTTTGAA
Caful   TCATAAATTGAAAATTTTATTAA--TGGTTGTTAA-----------AAAAATATTAA-----------------------------------------------AAATTTAAA
Petub   TTTAAATTTTTGATTAATTTTTT--TGGTTGTTAC-----------AAAAAT---------------------------------------------------TTTTGAAAA
Pemin   TTTTTTATTACTAAAATTAAAGT--TGGTTGTTAC---------AAAAA-----------------------------------------------------TTTTTGAAA
                                 MTHI1 BS
cons.   aa aatatgagaaaaaaataagg  tggTtgTtaa           aaaaatatt                                                 tttt  aaa

Bobra  GGAGCTATGAACATTT-------------GAGGCCTCTTTGTCGTGTTTG----------------------ATG
Halac  AATTAAACAATT----------------------CTGTTTTAAAATATAAA----------------------ATG
Ch_sp  AGAAAAAATA-------------------------------------------------------------ATG
Rorot  AACTGGTGAAAAGAAACCATATTTTAATAACATTTAAAATTCAATATATACATATAAAATATAAAAAAATTATG
Haret  TAAAATATTATTTAAA-------------TAGCTATTTCAAAAATTT-----------------------ATG
Ha_sp  TAAAATATTATTTAAA-------------TAGCTATTTCAAAAATTT-----------------------ATG
Chrad  AGGAAGGCTATAATTGATAAACTGAATACTATTTTCATTAAAATATT-----------------------ATG
Chsp3  TAATTGGGTAAA----------------------TTTTTTTTTATTGT-----------------------ATG
Chsph  TAATTGGTTAGCTAGA----------CGTTACACTATTTAAAAAAATATGA--------------------ATG
Tesoc  AATTAGGTGAGATTTAAAAAAAAAAATATT-----------------------------------------ATG
Chdeb  AAAGTGGTAAGA----------------------CTTTTTTTTAATATATT--------------------ATG
ChspW  AAAGTGGTAAGA----------------------CTTTTTTTTAATATATT--------------------ATG
Gopec  CAATTGGTAAGAAG--------------------TTTATTACAATTTT-----------------------ATG
Chrei  TAATTGGTAAGAAT--------------------TTTATTACATTTTT-----------------------ATG
EuspN  AAGTTTAAAAAAAAAATAT-------------------------------------------------------ATG
Plsta  ATTTTTGGTAAGAAG------------------TTTCTTAAAAATAT-----------------------ATG
Vocar  AATTTTGGTAAGAAG------------------TTTATTAAAAATAT-----------------------ATG
Voafr  AATTTTGGTAAGAAG------------------TTTATTAAAAATAT-----------------------ATG
Teobl  TAATAAAATATAATAA---------AGTAG---TTATTCAAAAATTT-----------------------ATG
Chniv  TACGGTGTAATAGATAAACTAAAAAAGCAAAGCTTACTTAAAATTTTT-----------------------ATG
Difra  TAAAAAAAAAAAATGA-----------AAGAGATTTTTAAA-----------------------------ATG
Chatm  AAATAGACCATTTAAAAAAGGG-----------TATCTTAGAATT-----------------------ATG
Neaqu  ATTGAAAAAAAAAAAAA--------------ATTTTTATTTTAAATAA-----------------------ATG
Cacer  AAACACATTACAT-------------------------------------------------------ATG
Codec  -------------------------------------------------------------------ATG
Co_sp  TATTAAATAAA-------------------------------------------------------ATG
Pepec  TATTAAAAAATAG----------------------TTTTTCAAAAATTT-----------------------ATG
Kiape  AATTGAATTATAATAA----AAAAGAAAATAAATAATCGAATTATTT-----------------------ATG
MispL  ACTATAAACAATAGAACA-------------------------------------------------------ATG
Moneg  AATGAAATTACAAGAAAAAAAGAAAATAA-----------------------------------------ATG
Scqua  AATGAAATTACAAGAAAAAAAGAAAATAAATGATTGAATTATTT-----------------------ATG
Glpla  ACTCAAGTAAAAAGAAT-------------------------------------------------------ATG
Psmar  ATTTAGAAAAACCCAAATCAATAT-------------------------------------------------ATG
Ulfasc AGCCAAATTAAT-------------------------------------------------------ATG
Ulfle  AGCCAAATTAAT-------------------------------------------------------ATG
Ullin  AGCCAAATTAAT-------------------------------------------------------ATG
```



Supplemental Fig. S8: Conservation of the MDH1 target in *atpH* and *atpI* 5'UTRs.

A) Alignment of the *atpH* 5'UTRs

The ~100 nts upstream of the *atpH* initiation codon (accession numbers listed in Suppl. Table S1) were recovered in the NCBI databases and searched for the occurrence of the TGGTTGTTAT motif. Only the sequences showing at least 8 matches with this motif within a range of 70 nt from the *atpH* AUG codon were retained for the alignment performed with the MUSCLE software, using default options, and then manually edited to improve the alignment. Residues conserved in more than 79 sequences (out of 113) are written in red, while conservative substitutions are written in blue. A putative -10 Pribnow box found a few nucleotide upstream of the putative MDH1 binding site (written in bold and underlined) in many species suggest that this later corresponds the 5'end of the *atpH* transcript, as it does in *C. reinhardtii*. A putative Shine-Dalgarno sequence (6-10 nt upstream of the initiation codon) is also well conserved in most species. The sequence of *C. reinhardtii atpH* 5'UTR is highlighted in yellow.

Abbreviations of species names are as follows:

Syret: Symbiochloris reticulata; Sclei: Schizomeris leibleinii; Sthel: Stigeoclonium helveticum; Prcol: Prasinoderma coloniale; Neast: Nephroselmis astigmatica; Ulfas: Ulva fasciata; Ul_sp: Ulva sp. UNA000 71828; Ullin: Ulva linza; Ulper: Ulva pertusa; Ulpro: Ulva prolifera; Ulfle: Ulva flexuosa; Ullac: Ulva lactuca; Cacer: Carteria cerasiformis; Cacru: Carteria crucifera; Jemin: Jenufa minuta; Etpse: Ettlia pseudoalveolaris; Igtet: Ignatius tetrasporus; Psame: Pseudocharacium americanum; Ooamb: Oophila amblystomatis; Chper: Chloromonas perforata; Stplu: Stephanosphaera pluvialis; Chlei: Chlamydomonas leiostraca; Chapp: Chlamydomonas applanata; Dusal: Dunaliella salina; Halae: Hafnomonias laevis; Chbre: Chlorosarcina brevispinosa; Olvir: Oltmannsiellopsis viridis; Mo_sp: Monoraphidium-species; Chniv: Chlamydomonas nivalis; Oecard: Oedogonium cardiacum; Oecaro Oedocladium carolinianum; Pssch: Pseudomuriella schumacherensis; Oogig: Oogamochlamys gigantea; Myhom: Mychonastes homosphaera; Myjur: Mychonastes jurisii; Chzof: Chromochloris zofingiensis; Golon: Golenkinia longispicula; Cosp1: Coelastrella sp. M60; Cosp2: Coelastrella sp. UTEX B 3026; Haret: Hariotina reticulata; Teobl: Tetradesmus obliquus; Pepec: Pectinodesmus pectinatus; Neaqu: Neochloris aquatica; Chinc: Chlorotetraedron incus; Hyret: Hydrodictyon reticulatum; Pedup: Pediastrum duplex; Halac: Haematococcus lacustris; Loseg: Lobochlamys segnis; Locul: Lobochlamys culleus; Chcap1: Chlorogonium capillatum; Phlen: Phacotus lenticularis; Chrad: Chloromonas radiata; Braer: Bracteacoccus aerius; Brmin: Bracteacoccus minor; Anjud: Ankyra judayi; Brgig: Bracteacoccus giganteus; Ca-sp: Carteria sp. SAG 8-5; Trtri: Treubaria triappendiculata; Moneg: Monoraphidium neglectum; Scqua: Scenedesmus

quadricauca; Kiape: Kirchneriella aperta; Chaci: Chlamydomonas acidophila; Chsph: Chlamydomonas sphaeroides; Chasy: Chlamydomonas asymetrica; Chpet: Chlamydomonas peterfii; Tesoc: Tetrabaena socialis; Vocar: Volvox carteri; Plsta: Pleodorina starrii; Chrei: Chlamydomonas reinhardtii; Gopec: Gonium pectorale; Psmar: Pseudoneochloris marina; Chcap2: Chamaetrichon capsulatum; Chbas: Chamaetrichon basiliensis; Psaki: Pseudendoclonium akinetum; Trmuc: Trichosarcina mucosa; Glpla: Gloeotilopsis planctonica; Glsar: Gloeotilopsis sarcinoidea; Yauni: Yamagishiella unicocca; Eu_sp: Eudorina sp.; Tuaki, Tupiella akineta; Hacap, Hazenia capsulata; Peang: Pediastrum angulosum; Petub: Pedinomonas tuberculata; Pemin: Pedinomonas minor; Sttet: Stauridium tetras; Rasub: Raphidocelis subcapitata; Psint: Pseudopediastrum integrum; Psbor: Pseudopediastrum boryanum; Haret: Hariotina reticulata; Lagra: Lacunastrum gracillimum; Chacu: Characiochloris acuminata; Botex: Borodinellopsis texensis; Chtat: Chlorococcum tatrense; Chros: Chloromonas rosae; Chere: Chlorosarcinopsis eremi; Chsti: Chlorosarcina stigmatica; Patex: Palmellopsis texensis; Prbot: Protosiphon botryoides; MispL: Microspora sp. UTEXLB472; Patra: Parallela transversalis; Elvir: Elakatothrix viridis; Trhys: Trochiscia hystrix; Chorb: Chaetopeltis orbicularis; Oeang: Oedogonium angustistomum; Ursp3: Uronema sp. CCAP334/1; Cosai: Coelastrella saipanensis; PsspC: Pseudopediastrum sp. CL0201VA; Voafr: Volvox africanus; Caful: Capsosiphon fulvescens; HaspM: Hariotina sp. MMOGRB0030F; Fobot: Follicularia botryoides; ChspU: Chlamydomonas sp. UWO 241; Spsim: Spermatozopsis similis; EuspN.: Eudorina sp. NIES-3984; Chsp3: Chlamydomonas sp. 3112; ChspW: Chlamydomonas sp. WS3

B) Alignment of the *atpI* 5'UTRs

The ~200 nt upstream of the *atpI* initiation codon (accession numbers listed in Suppl. Table ST1) were recovered in the NCBI databases and searched for the occurrence of the TGGTT(G/A)TTAT motif. Only sequences showing at least 7 matches with this motif within a range of 100 nt from the *atpI* AUG codon were retained for the alignment, performed with the MUSCLE software using default options, and then manually edited to improve the alignment. Residues conserved in more than 60 % of the sequences are written in red, while conservative substitutions are written in blue. Same abbreviations of species name than in panel A.

Legend of Supplementary Figures

Supplemental Fig. S1: (In support of the cloning strategy in Fig. 4A)

A) Length of 5'UTRs within the *atpI* polycistronic unit.

(Left) Sequencing of PCR amplicons from 5'RLM-RACE using the gene-specific primers (see Suppl. Table ST2), schematically depicted by arrowheads, led to the size of 5' UTRs indicated in panel. (Right) Schematic map of the *atpI* transcription unit. CDS are shown as thick rectangles, while 5'UTRs are represented by thin rectangles. The PCR amplicons detected in panel A are indicated with the length of the corresponding 5'UTRs indicated (in bp) between parentheses.

B) The *atpI* gene does not have a strong dedicated promoter

(Left) Schematic representation of the *atpI* transcription unit, with a zoom to the *psbD* -10 promoter region mutated in the G64 mutant strain. As a result, *psbD* transcription is reduced 10 fold (Klinkert et al., 2005). (Right) Accumulation of the *psbD*, *atpI*, and *rps12* transcripts in the G64 mutant strain. The mutation of the *psbD* promoter impacts the accumulation of the *psbD* mRNA but also that of the *atpI* mRNA, which is decreased 5 fold, compared to the wild type strain. This shows that this latter doesn't have a strong dedicated promoter. The *rps12* mRNA accumulates to 40% of the wild type level, probably because of RNA-stabilization effects compensate to some extent the reduction of transcription.

Supplemental Fig. S2: MTHI1 is required for translation of 5'*atpI*-driven genes. (In support of Fig. 4)

A) Map of the 5'*atpI-aadA-3'rbcl* cassette (*dlK*), inserted in a neutral site downstream of the *petA* gene and introduced by biolistic transformation in the chloroplast genome of the wild-type (mt^+) strain. Bent arrows symbolise promoters. The red rectangle represents the *psaA* promoter region inserted upstream of the *psbJ-atpI* intergenic region shown as a pale blue rectangle. The black line within this region indicates the processed *atpI* 5'end. Transcripts detected with the *aadA* specific probe are schematically depicted.

B) *dl*K transformants were recovered on TAP-spectinomycin plates and crossed with the *mthi1-1* (mt^-) strain. Thanks to the uniparental inheritance of the chloroplast genome from the *mt*+ parent, all progeny inherited the chimeric gene, while the *mthi1* mutation showed mendelian segregation. Two progeny (members 1 and 2 in the representative tetrad shown in panel B-D) inherited the wild-type *MTHI1* allele, grew phototrophically (B) and accumulated wild-type levels of the *atpH* mRNA (C) and of the AtpH subunit (D). They were resistant to spectinomycin (B) because they accumulated the monocistronic chimeric transcript (b in panel A) and expressed immuno-detectable amount of the AadA protein (D). *psaB* mRNA and OEE2 are shown as loading controls in panels C and D, respectively. The other two members of the tetrads (3 and 4) inherited the *mthi1* mutation, as shown by the lack of *atpH* transcript and by their failure to grow on minimum medium. These two progeny accumulated increased levels of the monocistronic form of the chimeric 5'*atpI-aadA* transcript. Indeed the translation of the *aadA* cassette, severely impaired in the *mthi1* background, leads to the cleavage of

the chimeric transcript, shortly after the *aadA* initiation codon (transcript c, ~800 bp) (Y. Choquet, unpublish. res., see also Fig. 2 in Goldschmidt-Clermont, 1991). These two progeny, nevertheless, fail to synthesise significant amounts of the AadA protein (D) and were sensitive to spectinomycin (B). Together these results demonstrate that the 5'UTR of the *atpI* gene is sufficient to confer an MTHI1-dependent expression to 5'*atpI*-driven transcripts.

Supplemental Fig. S3: MTHI1 targets the *atpH* 5'UTR. (In support of Fig. 5)

A) The 5'*atpH-aadA*-3'*rbcL* chimera (A) was inserted at a neutral site downstream of the *petA* gene and introduced by biolistic transformation in the chloroplast genome of the wild-type (mt^+) strain. Bent arrows symbolise promoters. The blue rectangle represents the *atpH* promoter and 5'untranslated regions, fused in frame to the *aadA* coding sequence. Transcripts detected with an *aadA* specific probe are schematically depicted.

B) *dH*K transformants were recovered on TAP-spectinomycin plates and crossed with the *mthi1-1* (mt^-) strain. Two members of the resulting tetrad (progeny 1 and 3 in the representative tetrad shown in panel B-D) inherited the wild-type *MTHI1* allele, grew photophically (B) and accumulated wild-type levels of the *atpH* mRNA (C) and of the AtpH subunit (D). They were resistant to spectinomycin because they accumulated the monocistronic chimeric dHK transcript (transcript (b) in panel A) and expressed immuno-detectable levels of the AadA protein (D). The other two members of the tetrad (2 and 4) inherited the *mthi1* mutation, as shown by the lack of *atpH* transcript (C) and by their failure to grow on minimum medium (B). These two progeny also failed to accumulate the monocistronic form of the chimeric 5'*atpH-aadA* transcript (C) and were sensitive to spectinomycin (B). This demonstrates that the 5'UTR of the *atpH* gene is sufficient to confer an MTHI1-dependent stability to a 5'*atpH*-driven transcript.

However, a dicistronic *petA-aadA* transcript (transcript (a) in panel A), most likely stabilised by the *petA* stabilisation factor MCA1 (Raynaud et al., 2007; Loiselay et al., 2008), accumulated to the same level in the four members of the tetrad. The *aadA* coding sequence present in this dicistronic transcript was nevertheless not expressed in the *mthi1* progeny as those were sensitive to spectinomycin and lacked accumulation of immuno-detectable AadA protein (D).

Supplemental Fig. S4: Cloning of the *MTHI1* gene. (In support of Fig. 7).

A) The *MTHI1* gene was cloned as in (Raynaud et al, 2007) by complementation of a *mthi1-1, arg7*, *cw15* mutant strain with an indexed library of cosmids (Depège et al 2003) kindly provided by Pr J.-D. Rochaix. The cosmid vector backbone (Purton et al, 1994) included the *Arg7* gene (Debuchy et al, 1989). Phototrophic colonies, selected on minimal medium (120 $\mu E.m^{-2}.s^{-1}$), became visible after ~2 weeks. Selection for arginine prototrophy provided a control of transformation efficiency. One pool yielded ~10 photoautotrophic transformants. From this pool, cosmid 21H4 was isolated that complemented the *mthi1* phototrophic defect. Its ~33.5 kb genomic insert corresponded to nt 4917085-4950512 from chromosome 17. Complementation with cosmid digests with restriction enzymes listed on the right (**+**: complementation of the

mutation; -: absence of complementation) further restricted the region required for complementation to that corresponding to gene model Cre17.g734564, as only the *Eco*RI, *Hind*III, *Bam*HI restriction enzymes (written in red) cutting within this gene model prevented the complementation of the mutant phenotype. This chromosomic localisation, however, is erroneous. Indeed the *ac46* mutant (*mthi1-1*) has been previously mapped to the complementation group XVI/XVII, which was later shown to correspond to chromosome 15 (Dutcher et al, 1991). Crosses confirmed that the *mthi1-1* mutation was linked to the CytC1 molecular marker on chromosome 15. It is of note that the *MTHI1* gene was localised on chromosome 15 in the version 4.0 of the Chlamydomonas genome and was moved to chromosome 17 in version 5.5.

We constructed from cosmid 21H4 the pgMTHI1 plasmid with a 10 679 bp genomic insert (chromosome17:4933979-4944653), encompassing Cre17.g734564 capable to restore the phototrophic growth of both *mthi1-1* and *mthi1-2* mutant strains.

B) One EST clone (AV629671) of this gene model was obtained from Kazusa DNA Research Institute and sequenced using appropriate primers. It contained the full-length coding sequence for MTHI1, as an in-frame stop codon is located 6 nucleotides upstream of the initiation codon. Sequence comparison with the genomic scaffold showed that *MTHI1* is composed of 11 exons. A polyA tail was found 424 bp downstream of the stop codon and 15 nt downstream of the TGTAA poly-adenylation consensus signal (Silflow, 1998).

Supplemental Fig. S5: the *MTHI1* locus. (In support of Fig. 6).

A) Schematic representation of the MTHI1 protein showing its three major domains, as well as the position of the two mutations.

B) The mutation in strain *mthi1-2* results in premature translation abortion.

Partial sequence of the *MTHI1* cDNA (with translation) in the wild-type and mthi1-2 strains, with nucleotides numbered from the first A of the initiation codon. The inserted nucleotide is written in red. The mutation in strain *mthi1-2* results in premature translation abortion.

C) Sequence of the MTHI1 protein.

The predicted chloroplast transit peptide is written in blue, the residue encoded by the mutated codon is written in red. The OPR repeats listed in Fig. 7B are alternatively underlined and boxed. The *C. reinhardtii*-specific C-ter tail is written in grey, with stretches of identical residues shaded.

Supplemental Fig. S6: Conservation of MTHI1 among green algae. (In support of Figs. 7 and 8D,E,F)

A) DNA sequences encoding MTHI1 orthologues were retrieved from the JGI phytozome (v12), the NCBI databases, and the MMETSP re-assemblies database (Keeling et al., 2014) by tBLASTn using *Cr*MTHI1 as a query. Gene models were then predicted using the Greengenie2 software (http://stormo.wustl.edu/GreenGenie2/; (Kwan et al., 2009)) and manually edited to include obvious missing regions of similarity, if required. Alignment of MTHI1 orthologues performed with the MUSCLE software using default options and manually edited to improve the alignment. OPR

repeats of the protein from *C. reinhardtii* are shown above the alignment. Residues conserved in more than half of the sequences are written in red, while conservative substitutions are written in blue.

Abbreviations of species names are as follows:

Ulmut Ulva mutabilis; Ulpro: Ulva prolifera; ULLac: Ulva lactuca; Moneg: Monoraphidium neglectum; Rasub: Raphidocelis subcapitata; Teobl: Tetradesmus obliquus; Scqua: Scenedesmus acutus; Scqua: Scenedesmus quadricauca; Chzof: Chromochloris zofingiensis; Chasy: Chlamydomonas asymetrica; Chsp3: Chlamydomonas sp. 32112; Tesoc: Tetrabaena socialis; Chsph: Chlamydomonas sphaeroides; Chdeb: Chlamydomonas debaryana; ChspW: Chlamydomonas sp. WS3; Vocar: Volvox carteri; EuspN: Eudorina sp.; Chrei: Chlamydomonas reinhardtii; Gopec: Gonium pectorale; Yauni: Yamagishiella unicocca; Cheur: Chlamydomonas euryale; ChKRBP: ; Ooamb: Oophila amblystomatis; Duter: Dunaliella tertiolecta; Chmoe: Chlamydomonas moewusi; Chlei: Chlamydomonas leiostraca; Chapp: Chlamydomonas applanata; Halac: Haematococcus lacustris; Cheus: Chlamydomonas eustigma; Chaci: Chlamydomonas acidophila.

B) Phylogeny of MTHI1 orthologues.

The phylogenic tree was constructed on the Phylogeny.fr platform (Dereeper et al., 2008), including the following steps: Sequences were aligned with MUSCLE (v3.7) (Edgar, 2004) configured for highest accuracy (MUSCLE with default settings) and cured with Gblocks (v0.91b), using relaxed parameters. The phylogenetic tree was reconstructed using the maximum likelihood method implemented in the PhyML program (v3.0) (Guindon et al., 2010) with reliability for internal branch assessed using the aLRT test (SH-Like) (Anisimova and Gascuel, 2006) and visualized using TreeDyn (v198.3) (Chevenet et al., 2006). The OPR protein TDA1 (Eberhard et al., 2011) was taken as outgroup.

Supplemental Fig. S7: sRNAs coverage (normalised as RPM) over the *atpI* 5'UTR and along the inverted repeat. (In support of Figs. 10B and 11C,D).

A) Coverage of sRNAs mapping to the coding strand along the *atpI* 5'UTR, schematically depicted at the bottom. The red bar symbolises the putative MTHI1 binding site. Blue line: RPP-treated wild-type sample; red line: RPP-treated *mthi1-1* sample.

B) Coverage of sRNAs immuno-precipitated with the MTHI1 protein over the *atpI* 5'UTR. Red line: sRNA coverage in the wild-type negative control; blue line sRNA coverage in the MTHI1-RIP sample. All samples were RPP-treated. Note the absence of coverage in correspondence of the MTHI1 putative binding site.

C) sRNA coverage of RPP-treated MTHI1 sample within the inverted repeat. Blue line: sRNAs mapping to the plus strand; Red line: sRNAs mapping to the minus strand. The position of genes within the inverted repeat is shown. The blue peak corresponds to the *psbA* 5'UTR.

Supplemental Fig. S8: Conservation of the MDH1 target in *atpH* and *atpI* 5'UTRs.

A) Alignment of the *atpH* 5'UTRs

The ~100 nt upstream of the *atpH* initiation codon (accession numbers listed in Suppl. Table S1) were recovered in the NCBI databases and searched for the occurrence of the TGGTTGTTAT motif. Only the sequences showing at least 8 matches with this motif within a range of 70 nt from the *atpH* AUG codon were retained for the alignment performed with the MUSCLE software, using default options, and then manually edited to improve the alignment. Residues conserved in more than 79 sequences (out of 113) are written in red, while conservative substitutions are written in blue. A putative -10 Pribnow box found a few nucleotide upstream of the putative MDH1 binding site (written in bold and underlined) in many species suggest that this later corresponds the 5'end of the *atpH* transcript, as it does in *C. reinhardtii*. A putative Shine-Dalgarno sequence (6-10 nt upstream of the initiation codon) is also well conserved in most species. The sequence of *C. reinhardtii atpH* 5'UTR is highlighted in yellow.

Abbreviations of species names are as follows:

Syret: Symbiochloris reticulata; Sclei: Schizomeris leibleinii; Sthel: Stigeoclonium helveticum; Prcol: Prasinoderma coloniale; Neast: Nephroselmis astigmatica; Ulfas: Ulva fasciata; Ul_sp: Ulva sp. UNA000 71828; Ullin: Ulva linza; Ulper: Ulva pertusa; Ulpro: Ulva prolifera; Ulfle: Ulva flexuosa; Ullac: Ulva lactuca; Cacer: Carteria cerasiformis; Cacru: Carteria crucifera; Jemin: Jenufa minuta; Etpse: Ettlia pseudoalveolaris; Igtet: Ignatius tetrasporus; Psame: Pseudocharacium americanum; Ooamb: Oophila amblystomatis; Chper: Chloromonas perforata; Stplu: Stephanosphaera pluvialis; Chlei: Chlamydomonas leiostraca; Chapp: Chlamydomonas applanata; Dusal: Dunaliella salina; Halae: Hafnomonias laevis; Chbre: Chlorosarcina brevispinosa; Olvir: Oltmannsiellopsis viridis; Mo_sp: Monoraphidium-species; Chniv: Chlamydomonas nivalis; Oecard: Oedogonium cardiacum; Oecaro Oedocladium carolinianum; Pssch: Pseudomuriella schumacherensis; Oogig: Oogamochlamys gigantea; Myhom: Mychonastes homosphaera; Myjur: Mychonastes jurisii; Chzof: Chromochloris zofingiensis; Golon: Golenkinia longispicula; Cosp1: Coelastrella sp. M60; Cosp2: Coelastrella sp. UTEX B 3026; Haret: Hariotina reticulata; Teobl: Tetradesmus obliquus; Pepec: Pectinodesmus pectinatus; Neaqu: Neochloris aquatica; Chinc: Chlorotetraedron incus; Hyret: Hydrodictyon reticulatum; Pedup: Pediastrum duplex; Halac: Haematococcus lacustris; Loseg: Lobochlamys segnis; Locul: Lobochlamys culleus; Chcap1: Chlorogonium capillatum; Phlen: Phacotus lenticularis; Chrad: Chloromonas radiata; Braer: Bracteacoccus aerius; Brmin: Bracteacoccus minor; Anjud: Ankyra judayi; Brgig: Bracteacoccus giganteus; Ca-sp: Carteria sp. SAG 8-5; Trtri: Treubaria triappendiculata; Moneg: Monoraphidium neglectum; Scqua: Scenedesmus quadricauca; Kiape: Kirchneriella aperta; Chaci: Chlamydomonas acidophila; Chsph: Chlamydomonas sphaeroides; Chasy: Chlamydomonas asymetrica; Chpet: Chlamydomonas peterfii; Tesoc: Tetrabaena socialis; Vocar: Volvox carteri; Plsta: Pleodorina starrii; Chrei: Chlamydomonas reinhardtii; Gopec: Gonium pectorale; Psmar: Pseudoneochloris marina; Chcap2: Chamaetrichon capsulatum; Chbas: Chamaetrichon basiliensis; Psaki: Pseudendoclonium akinetum; Trmuc: Trichosarcina mucosa; Glpla: Gloeotilopsis planctonica; Glsar: Gloeotilopsis sarcinoidea; Yauni: Yamagishiella unicocca; Eu_sp: Eudorina sp.; Tuaki, Tupiella akineta; Hacap, Hazenia capsulata; Peang: Pediastrum angulosum; Petub: Pedinomonas tuberculata; Pemin: Pedinomonas minor; Sttet: Stauridium tetras; Rasub: Raphidocelis subcapitata; Psint:

Pseudopediastrum integrum; Psbor: Pseudopediastrum boryanum; Haret: Hariotina reticulata; Lagra: Lacunastrum gracillimum; Chacu: Characiochloris acuminata; Botex: Borodinellopsis texensis; Chtat: Chlorococcum tatrense; Chros: Chloromonas rosae; Chere: Chlorosarcinopsis eremi; Chsti: Chlorosarcina stigmatica; Patex: Palmellopsis texensis; Prbot: Protosiphon botryoides; MispL: Microspora sp. UTEXLB472; Patra: Parallela transversalis; Elvir: Elakatothrix viridis; Trhys: Trochiscia hystrix; Chorb: Chaetopeltis orbicularis; Oeang: Oedogonium angustistomum; Ursp3: Uronema sp. CCAP334/1; Cosai: Coelastrella saipanensis; PsspC: Pseudopediastrum sp. CL0201VA; Voafr: Volvox africanus; Caful: Capsosiphon fulvescens; HaspM: Hariotina sp. MMOGRB0030F; Fobot: Follicularia botryoides; ChspU: Chlamydomonas sp. UWO 241; Spsim: Spermatozopsis similis; EuspN.: Eudorina sp. NIES-3984; Chsp3: Chlamydomonas sp. 3112; ChspW: Chlamydomonas sp. WS3

B) Alignment of the *atpI* 5'UTRs

The ~200 nt upstream of the *atpI* initiation codon (accession numbers listed in Suppl. Dataset DS1) were recovered in the NCBI databases and searched for the occurrence of the TGGTT(G/A)TTAT motif. Only sequences showing at least 7 matches with this motif within a range of 100 nt from the *atpI* AUG codon were retained for the alignment, performed with the MUSCLE software using default options, and then manually edited to improve the alignment. Residues conserved in more than 60 % of the sequences are written in red, while conservative substitutions are written in blue. Same abbreviations of species name than in panel A.

Supplementary methods

DNA constructs

Plasmids p-520 and P-70, respectively containing a 7.8 PstI fragment of the chloroplast genome encompassing the 3'end of *atpA*, *psbI*, *cemA*, *atpH*, *atpF* and *rps11* cloned in Bluscript pBSKS+ vector or a 4.9 kb EcoR1 fragment of the chloroplast genome encompassing the *psbJ*, *atpI*, *psaJ* and *rps12* genes cloned into Puc8, were obtained from the Chlamydomonas Resource Center (http://chlamycollection.org/).

Deletion of the *atpH* gene.

To remove unwanted restriction sites, plasmid P-520 was first digested by *Sac*I and *Nco*I, blunted with T4 DNA pol and religated on itself, then digested with *Pac*I and *Xho*I, blunted with T4 DNA pol and religated on itself to yield plasmid p-520Sh that only contains a 4916 bp insert.

A 789 DNA fragment upstream of the *atpH* coding sequence was amplified from template p520 using primers *cemA*_RI and $atpH_{Del}$, digested by *Eco*RI and *Eco*RV and cloned into plasmid p-520Sh digested by the same enzymes to create plasmid p∆*atpH*. The recycling *psaA*-driven *aadA* cassette (Boulouis et al, 2015), excised from plasmid p5'aA-$aadA_{485}$ by digestion with *Sac*I and *Xho*I, was cloned into plasmid p∆*atpH*, digested by the same enzymes (restriction sites introduced when designing primer *atpH*_del) to yield plasmid pK^r∆*atpH*.

Deletion of the *atpI* gene

A 1013 bp DNA fragment was amplified by two step megaprime PCR (Higuchi, 1990): primers psbJ_FW/atpIDel_RV and atpI_RV/atpIDel_FW allowed the amplification from plasmid p-70 of two partially overlapping amplicons that were mixed and used as templates in a third PCR with the external primers psbJ_FW and atpI_RV. In the final amplicon, the whole *atpI* 5'UTR and CDS were deleted and replaced by a short MCS. After digestion by *Cla*I and *Hpa*I, this amplicon was cloned into plasmid P-70 digested with the same enzymes, yielding plasmid p∆*atpI*. The recycling *psaA*-*aadA* cassette, excised from plasmid *p5'aA*-$aadA_{485}$ by digestion with *Sac*I and *Xho*I, was cloned into plasmid p∆*atpI*, digested by the same enzymes (resriction sites introduced when designing primer atpIDel_FW and atpIDel_RV) to yield plasmid pK^r∆*atpI*.

Construction of reporter genes

5'*atpH*-driven reporter genes

The *atpH* promoter and 5'UTR regions, PCR-amplified from the template plasmid P-520 using primers $atpH_{prom}$ and $atpH_{ATG}$, were digested by *Eco*RV and *Nco*I and cloned into the pWF*aAK* vector digested by the same enzymes to yield plasmid pWF*dH*K.

The promoter and a slightly extended 5'UTR region of *atpH* were similarly amplified using primers PCR $atpH_{prom}$ and $atpH_{ATG}2$, digested by *Hinc*II and *Nco*I and cloned into plasmid p*aAf* (Wostrikoff et al., 2004), digested by the same enzyme to yield plasmid p*dHf*. The recycling *psaA*-driven *aadA* cassette, excised from plasmid *p5'aA-aadA*$_{485}$ by digestion with *Sac*I and *Kpn*I and blunted with T4 DNA Polymerase, was cloned into plasmid p*dHf* digested with *Hinc*II to yield plasmid pK^r*dHf*.

pGatpH

The *pGatpH* construct was created by a two-step PCR procedure, using the external primers cemA-FW and atp_RV and template plasmid $p520_{Sh}$. The final amplicon, carrying the poly(G) track (909 bp), was digested with *Eco*RI and *Eco*RV and cloned into plasmid p520Sh digested with the same enzymes to create plasmid patpH-*pG*. The recycling *5'aA-aadA*$_{485}$ was then cloned into plasmid patpH-*pG* digested with *Eco*RV to yield plasmid pK^r atpH-*pG*.

5'*atpI*-driven reporter genes

The *atpI* 5'UTR was amplified from the template plasmid P-70 using oligonucleotides $atpI_{ATG}$ and $atpI5'FW_{prom}$. The resulting 637 bp amplicon was digested by *Nsi*I and *Nco*I and cloned into vectors p*aAf* or pWF*aAK* digested with the same enzymes to yield plasmid p*dIf* or pWF*dIK*, respectively. In these constructs, the *Cla*I-*Nsi*I fragment from the *psaA* 5'UTR and promoter regions provides a promoter to drive the expression of the promoter-less *atpI* 5'UTR. The recycling *psaA-aadA* cassette, excised from plasmid *p5'aA-aadA*$_{485}$ by digestion with *Sac*I and *Kpn*I and blunted with T4 DNA Polymerase, was then cloned into plasmid p*dIf* digested with *Hinc*II to yield plasmid pK^r*dIf*.

Construction of reporter genes driven by modified *atpI* 5'UTRs

Plasmid pK^r*dIf* was digested with either *Bsu*96I and *Bsm*I, *Bsm*I and *Sna*BI, *Bsu*96I and *Sna*BI, or *Sna*BI and *Pfl*MI, blunted by T4 DNA Pol treatment and religated on itself to yield respectively plasmids pK^r*dIf*Δ1, pK^r*dIf*Δ2, pK^r*dIf*Δ3, and pK^r*dIf*Δ4*.* The putative target of MTHI1 within the *atpI* 5'UTR was also modified by megaprime PCR: primers atpI5'_FW/atpI$_{Tar}$_RV and atpITar_FW/atpI5'_RV allowed the amplification from plasmid p-70 of two partially overlapping amplicons that were mixed and used as templates in a third PCR with the external primers atpI5'_FW and atpI5'_RV. This 996 bp final amplicon was digested with *SnaB*I and *Pml*I and cloned into the pK^r*dIf* vector digested by the same enzymes to create plasmid pK^r*dIf*ΔT.

5'*psaA*-driven *atpI*

To remove unwanted restriction sites, the P-70 vector was cut with *Cla*I and *Nde*I, blunted with Klenow enzyme and religated on itself to yield plasmid p70_CN. This plasmid was then digested with *Eco*RI and *Sex*AI, filled with Klenow and religated on itself to generate plasmid P70$_{Sh}$.

The *atpI* 5'UTR was then deleted from this plasmid by two steps megaprime PCR procedure: primers *psbJ*_FW/*atpI*Chim_RV and *atpI*-Chim_FW/*atpI*5'_RV allowed the amplification from plasmid P-70 of two partially overlapping amplicons that were mixed and used as templates in a third PCR with the external primers *psbJ*_FW and *atpI*5'_RV. This 743 bp amplicon was digested with *Kpn*I and *Bst*BI and cloned into the P70$_{Sh}$ vector digested with the same enzymes to create plasmid p*atpI*Δ5'.

To generate plasmid p5'*psaA-atpI*_K^r, the promoter and 5'UTR regions of the *psaA* gene were then amplified from the template plasmid ps1A1 (Kuck et al, 1987) using primers *psaA*$_{prom}$ and *psaA*$_{ATG}$. The resulting 270 bp fragment amplicon was digested with *Cla*I and *Nco*I and cloned into vector p*atpI*Δ5', digested with the same enzymes to yield plasmid p5'*psaA-atpI*. This plasmid was digested with *Sma*I (a restriction site introduced in the psaA$_{prom}$ primer) and ligated with the recycling 5'a*A-aadA* cassette to yield plasmide*p*K^r*5'psaA-atpI.*

AtpI$_{St}$

An untranslatable version of the *atpI* gene was constructed by a two-step megaprime PCR procedure, using the external primers atpI_FW and atpI_RV2 and the mutagenic primers atpI$_{St}$_FW and atpI$_{St}$_RV on the template plasmid p70$_{Sh}$. The final

1069 bp amplicon was digested with *Kpn*I and *Bsu*36I and the resulting 654 bp fragment was cloned into p520Sh digested with the same enzymes to create p*atpI*$_{St}$. The recycling 5'a*A*-*aadA* resistance cassette was the cloned into the *Hpa*I site of this vector to yield plasmid p*atpI*$_{St}$K^{r}.

***MTHI1* constructs.**

We constructed a vector encompassing the genomic sequence of the *MTHI1* gene by digesting the 21H4 cosmid by *Eco*RV and *Xho*I, isolating the 10679 bp subfragment that was cloned into pBluescriptII SK- digested by *Xho*I and *Ale*I to create plasmid p*gMTHI1*. A triple HA tag was fused to the C-terminus of the protein by megaprime PCR, using the mutagenic primers MTHI1$_{HA}$FW and MTHI1$_{HA}$RV and the external primers MTHI1FW5 and MTHI1RV5. The resulting 1252 bp amplicon was digested with *Sfr*I and *Spe*I and cloned into the pgMTHI1 vector digested by the same enzymes to create plasmid p*gMTHI1-HA*. To remove most of the C-terminal domain of the protein, but keep the triple HA tag, a 921 bp PCR product was amplified from template p*gMTHI1-HA* with primers MTHI1DelC_FW and MTHI1DelC_RV, digested by *Hind*III and *Srf*I, and cloned into plasmid p*gMTHI1* digested by the same enzymes to generate plasmid p*gMTHI1*ΔC.

The AV629671 EST clone containing a full length cDNA cloned into the pBluescriptII SK- vector, was obtained from the Kazusa DNA Research Institute (Asamizu et al, 2000). The triple HA tag was introduced in this plasmid by cloning the 350 bp fragment recovered from the digestion of the above mentioned 1252 bp PCR fragment with *Fsp*AI and *Bst*1107I into the AV629671 vector digested by the same enzymes to yield plasmid pcMTHI1-HA.

***MTHI1* recoding.**

The *MTHI1* CDS (Cre17.g734564.t1.1) was codon optimised for expression in *E. coli* (Ec) and synthesised by GeneCust (Suppl. Fig. S10). The synthesized *EcRDH1* gene lacks the first 147 bp and contains instead 5'-ATGGCGATTGCAATTGGAATTCAT-3' which derives from the bacterial *araB* gene (ECK0064). Ec*RDH1* was cloned into vector pET28a (Novagen) using the *Nco*I and *Hind*III restriction sites to yield plasmid pET28a-EcRDH1. To introduce the hexa-histidine tag at the N-terminus, primers ecRDH1-F and ecRDH1-R were used to amplify a 573-bp DNA fragment from plasmid pET28a-ecRDH1, digested with *Nco*I

and *Age*I, and cloned into pET28a-ecRDH1 digested with the same enzymes using the *NEBuilder HighFidelity DNA Assembly Cloning* (New England Biolabs) strategy resulting in pET28a-6His-ecRDH1.

MTHI1 overexpression, purification and immunisation.

For antibody generation, MTHI1 was overexpressed in *E. coli* BL21 for 16h at 15°C and purified by nickel-nitrilotriacetic acid agarose under denaturing conditions as described earlier (Muranaka et al., 2016) (Suppl. Fig S11). After elution, fractions were concentrated on Amicon Ultra 50 kDa filters units and washed several times with 6 M urea and 20 mM Tris-HCl pH 8. For antibody generation, 0.5 mg of purified protein was used for an 88-days rabbit immunization protocol (Covalab, France).

Supplemental Fig. S9: Emission spectrum of the white led used to grow *C. reinhardtii*. (In support of the M&M section)

Measured with a S2000 Fibber Optics spectrophotometer (Ocean Optics, Inc).

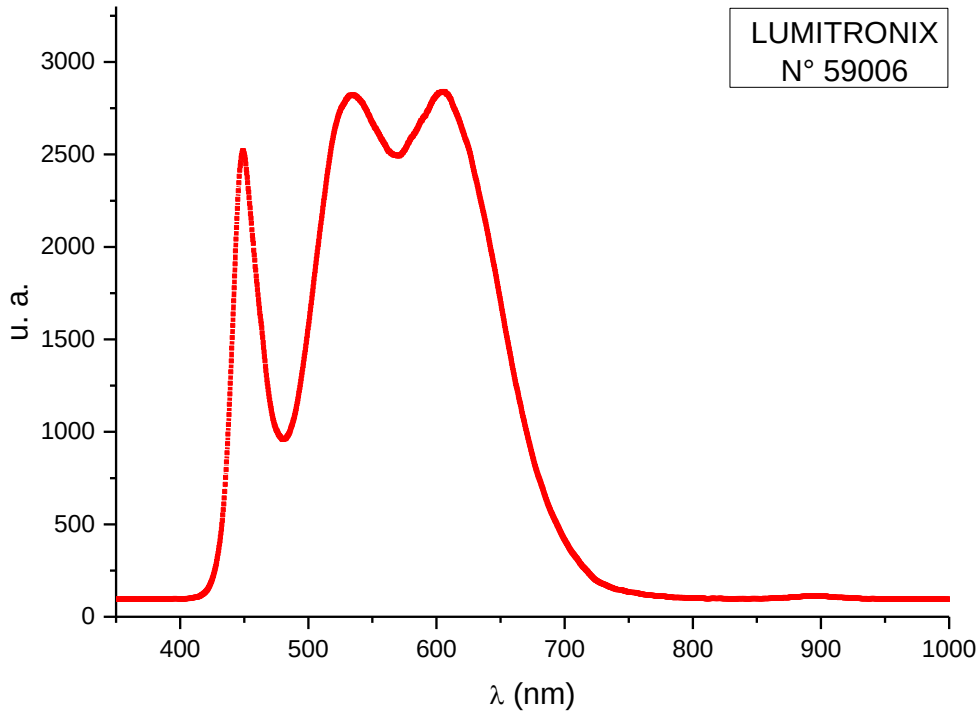
Supplemental Fig. S10: Sequence of the recoded *EcMTHI1* gene. (In support of the M&M section)

Restriction sites introduced for cloning purpose are written in colour.

Supplemental Fig. S11: Purification and quantification of the recombinant EcMTHI1. (In support of the M&M section)

A) Purification of EcMTHI1: I: whole cell lysate; S: supernatant after centrifugation to remove cell debris; FT: Flow Through across the Nickel column; W: wash fraction; E1 → E5 elution fraction. The right panel show the immuno-detection of MDH1 in the E2 → E4 fraction with an antibody against the HA tag.

B) Quantification of *Ec*MTHI1, by comparison with known amounts of a BSA standard

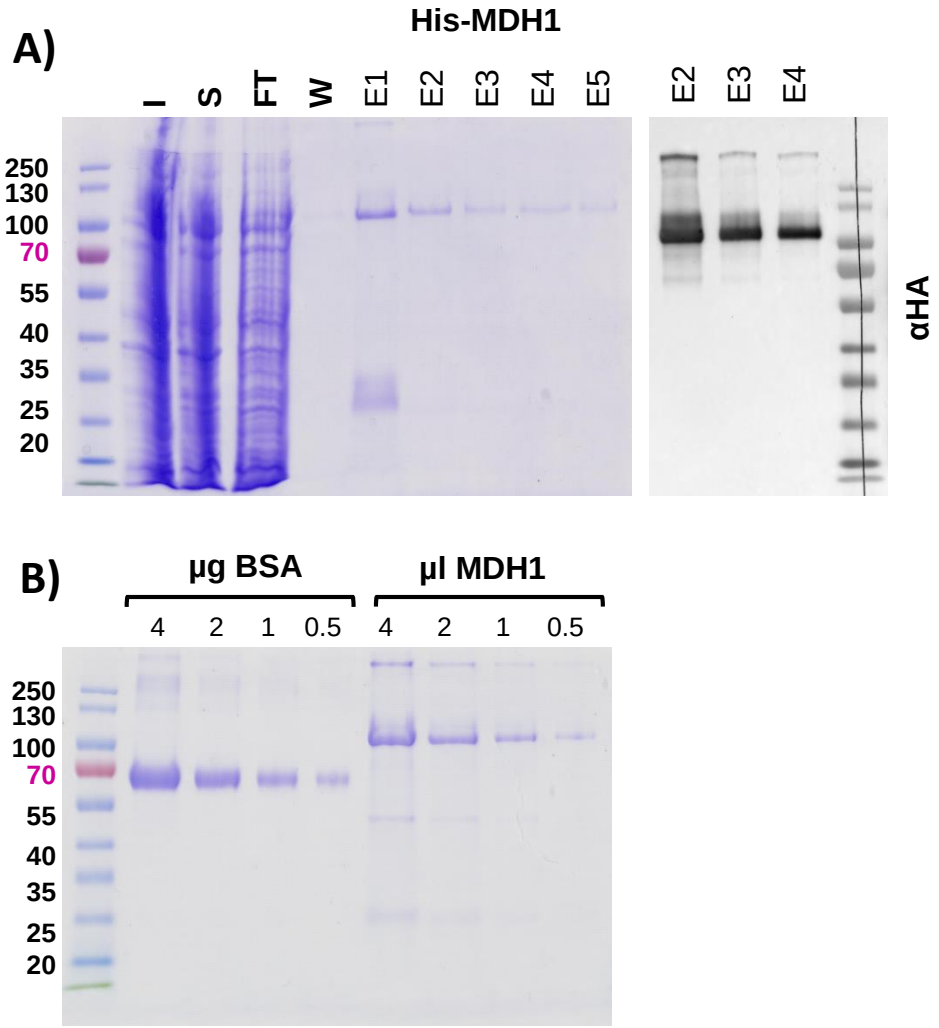


Supplemental Fig. S9: Emission spectrum of the white led used to grow *C. reinhardtii*. Measured with a S2000 Fiber Optics spectrophotometer (Ocean Optics, Inc).

ecMTHI1-HA 2455bp

CTCGAGCCATGGCGATTGCAATTGGAATTCATGGTGCGCGTCGCGGCGTTCATAACGGTGCAGCTGCTGT
TCACCCGGGCGGTCTGGACCTGCTGGATACTGCGGAATCTTCTGCAGAACAGCTGACCCCGCGCCGTCTG
CTCAACCGTCGTATCAAATCTTGCCTGTCCCCGGCGCAGCTGGCGGGTCTGGTGCTGAGCGAAGTTGGTA
ACTTCGATCAGCAGAATGCTTCCCACGCTCTGTCTCGCCTCGCCAAAATGTACCGCGGCCGTCGTCGTAA
CTCTCATCAGCACAGCCGTGGTAGCGACGTAGACCGTGCGCGTGCTGCGGCCGAGTTGCGCCCGGCTGTA
GAAGCGTTAACTAAACGTATGCACCAACTGATCGGTAACTACGATTCCTGGGATACCACTCTGTCGCTCT
GGGCTTATGCTCAGCTCGATCACTATGACGAAGGTGCTCTGCGTGCTCTGTGCGACGCGGCGGTTGAAGT
TGCGCCGATTTTCAAACCGGTCGATTGTGCGAACGCTGTTGTTGCCTTCGCGCATCTGGACTACGTTCAC
CCGGAACTGCTGCGCCAGATTGTTCAGACCGTGTTGGACACGCTGGACGATTACGCGCCTGGTGAAGTGT
GCCAGGTGCTCTGGGGCTTTGCCCGTCTGGGCGTTCATCCGGGACCGGCGTTCCTGGCGGAAGTCGTTGA
TGCGGTACAGTGGCGTCTGCAAGGCTACGGAACCCAGGAGCTGGGGATGGTGCTGTGGGCATTAGTCCGC
CTGGGTTACAAACCGGGCCCGCGTTTCCTGCGTGACGTTGAATCCGTGCTGCTCGCACGTCTGCCGCACA
TGGCACCGGGCGATATCGCGATTACCGTGTGGTCGTTTGCTCGTCTGCGTTACAAAGCAGTACGCTTCCT
GGATGAAGTGCCGGCGGCCGTGGGTCCGCAGCTGCACAAATGCCGTTCAAGCGAACTGTGCTCTCTGATC
TCCGGTTTCGCTACAGCTCACCACTATCACAAGTCTCTGCTGGACGCTGTGGCTGACGTTCTGCTGTCTC
GCCTGGATGGCCTGAGCCACCACGAAGTTGCGACTGCTCTCTGGACCTTCGGCACCTTCCGCCACCGTCC
GGCGCACCCGGATTTCGCGAAACAGGTTGCGGCAGCGCTGTACGCACGCATGCGCAGCTTTAGCCCGCAG
GGCCTGGCGATGGTGGTTAAAGCGCTGGCTCAGCTGCAGTGGCGCTCCGAGCCGCTGATGGAACAGCTGA
TCGCAGCTGCCGAGGCGAAGCTGAACGCCTTCAAACCGCTGGAACTCTCTCAGCTGCTGTGGGGCCTGAC
TGCACTGCAGTGCCGTGATCTGCATATCTACTACGCCGTGGTTCGTCGCTGCATCGCGATCCTGAAGGAT
CCGGCCCACCCGCACTATCGCACCATGACCCACCACCGCGTTGTGAACAGCGTGCTGGGCTCTTGTCAGC
AGCTGGGTTATGTTCCGTGGACCCTGATCGATTTCGCAGAATCTAAAGGGATCCGCGTACGTCAGCCAGA
CATCCTGTCTTCCCGTGATGAAGATGACGAAGGCGTACCATACTCCCATCAGCAGCAACAGCACGCGGAT
GCGGTGGAAGGCTGTGGCCACGATCGCGCGGAACAGCCGTGGGGCGCGTCTACCGGGTCCTCATCCTCTT
CTTCCCGCCGTCACCAGCGTTGTGCGGAAGAAGAAGCGCTGTGGGCTGAGGCGGAGCGCGCCCACTCACA
GCAGGTCGCGGCAGGCAACTCTAGCAGCGATGCTGCAATGGCGTCTGCTCCAGATGCGGTGGTGCTTCTG
GAACAGGGTCTTATCCCGCACGTTTCCTCTAGCGCTGCAGCCGACGCATCTCATGAAGCGGCTGCGGTTC
ACGCAGCCGCTCAGGGTGAATACCGCGCGCTGCAGCAACCGAAGCCGCAGCCGCTGGCGATGCTGACCGA
GCGTGGCTCTCGTCACGCCACCGGCATGATTGTTCTGGCAGGAGCAGCGGCCGTTGTTGCGGGTGAGGGT
GTGTCTGCAGGCGACGCAGAACAGCAATCCGCTATGTCTGCTCCGAACGTGGCCCAGCTGCAGGAAAGCG
CGCCGGCCGCTGCGGCACTGGACGGCAGCAACAGCGGGTCTAACGGGGCAAAAGTTCTGTCTCCGCGTCC
GCGTCTGGGCTCCGCCCGTCGTGGCGGCCCGGTTGTTGCAGGTGACGCGTCCCCGAAAGGCGCCTCTGCT
CACGTTGCAGTGCCGGTAGATTCCGCGGCGCCGTCTGGTGCCCGTGCGCGCGCTCTGTTTTCTGACCCGC
GCCGTGACAGCCCGTACAACGTGGGTATGGTTGCCGCGACCCCGCTGACCTTTCAGCGTTATCCGTATGA
CGTTCCGGACTATGCGTACCCGTACGACGTGCCGGACTACGCCAGCTACCCGTACGATGTTCCGGACTAC
GCGTAAAGCTTACTAGTCTCGAG

<u>Supplemental Fig. S10:</u> Sequence of the recoded *EcMTHI1* gene.
Restriction sites introduced for cloning purpose are written in colour.



<u>Supplemental Fig. S11:</u> Purification and quantification of the recombinant EcMTHI1

A) Purification of *Ec*MTHI1: I: whole cell lysate; S: supernatant after centrifugation to remove cell debris; FT: Flow Through across the Nickel column; W: wash fraction; E1 → E5 eluted fraction. The right panel shows the immuno-detection of MDH1 in the E2-E4 fractions with an antibody against the HA tag.

B) Quantification of *Ec*MTHI1, by comparison with known amounts of a BSA standard.

Supplemental Table ST1: Oligonucleotide used during that study.

name	Sequence[1]	note
Oligonucleotides used for DNA constructs, listed by order of appearance in the DNA construct section		
cemA_RI	ACAGCGTGTTATTGGTGCAA	
atpH$_{Del}$	CGC**GATATC***GAGCTC*GCG**CTCGAG**TCCTCCTAAATGAATCAATAAAATCGA	
psbJFW	CAGTTGGCTATGCCTCAACTCAC	
atpIDel_RV	**GCATGC**TCG*AGTATG*AGCTCCATGGTCATATCCTATGGATTGATGCAAAAGTCT	
atpIDel_FW	ATGA**CCATGG**AGCTCATA**CTCGAG**CATGCTGAAGCTTTAGCAGATCACCACTAATCTT	
atpI_RV	***GCGGAATTC***AGCAATTACAGGTGCTGTTGA	
atpH$_{prom}$	GCG***GTTGAC****ATGCAT*TACTTTAAATGGGAATCCTTTC	
atpH$_{ATG}$	GCG***CCATGG***AAGTTGCAGCTACGATAGGGTTCAT	
atpH$_{ATG}$2	GCG***CCATGG***TCATTGTATTTCCTCCTAAATGA	
cemAFW	GCGAATTCCGGAAAGTCAAACAGGTATTTTCTT	
atpH_RV	GCGTTAGCCAATACCAAACAGC	
pGatpH_FW	TTGGGGGGGGGGGGGGGGGGTTGGTTGTTATCGATTTTATTGATTCA	
pGatpH_RV	AACCCCCCCCCCCCCCCCCCAAAGAATATTATATTCTTTGGTTGTTTCA	
atpI$_{ATG}$	CTA**CCATGG**ATACTTCAGCAATTTCTAATAAAGG	
atpI5'_FW	GCG**ATGCAT**GCCCTTATCAAGCTTCCACATA	
atpI$_{Tar}$_FW	TCATTATTTATAGTTCAGCTGCAGAAAAATATTAATATCTTTCCAATAATTGGTAAGA	
atpI$_{Tar}$_RV	AGATATTAATATTTTTCTGCAGCTGAACTATAAATAATGACCAGTTATTATCT	
atpI5'_RV	CCAAAAAGTCGGAAGCTTAATG	
atpI$_{Chim}$_FW	AGCAGT**ATCGAT***ATC***CCATGG**TTATGAATCCTTTATTAGAAATTGCTGAAG	
atpI$_{Chim}$_RV	TCATAA**CCATGG**GAT**ATCGAT**ACTGCTATTTTCATAACAAATATATAA	
psaA$_{Prom}$	CGC**ATCGAT**A**CCCGGG**TACGAATACACATATGGTAAAAAAT	
psaA$_{AUG}$	GCG**CCATGG**TCATGGATTTCTCCTTATAATAACA	
atpI_FW	GCCCTTATCAAGCTTCCACATAGCGT	
atpI_RV2	ACGCCGCTGGTTCTACATAGCG	
atpISt_FW	AGAATTTTATTACATTTTTTAGATCTCTTTATTAGAAATTGCTGAAGTATCTGTA	
atpISt_RV	TCAGCAATTTCTAATAAAGAGATCTAAAAAATGTAATAAAATTCTTACCA	
MTHI1$_{HA}$_FW	**TACGATGTCCCCGACTACGCTAGCTACCCTTATGATGTTCCTGATTATGCT**TGAAGGCAGCTCCCAGGTTGACA	
MTHI1HA_RV	**TAGCTAGCGTAGTCGGGGACATCGTACGGGTACGCGTAGTCCGGAACGTCGTAGGGAT**AGCGTTGAAACGTTAGCGGCGTGG	
MTHI1_RV5	TCACCGCCCTCCAGCAGCTCCG	

name	Sequence[1]	note
Oligonucleotides used for DNA constructs, listed by order of appearance in the DNA construct section		
MTHI1_FW5	CAAGTGGCGGCGGGCAACAGCA	
$MTHI1_{DelC}$_FW	GATCGCGGCGGCCGAGGCCAAGC	
$MTHI1_{DelC}$_RV	GCCCGGGCTCCCTGCTGCTGTTGGTGCGAGTACGGCACG	
ecMTHI1-FW	ACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCACCATCATCACCACAGCGCGATTGCAATTGGAATTCATG	
ecMTHI1-RV	AACAGCGTTCGCACAATCGACCGGTTTGAAAATCGGCG	
Oligonucleotides used to sequence the MTHI1 cDNA and the MTHI1 gene in mthi1 the mutants		position / initiation codon
MTHI1_FW1	AGGTTTCAGCACGCGGAATGTGGG	-1416
MTHI1_RV1	AGACAGCGTCGTGTCCCAGGAGTC	+769
MTHI1_FW2	TACGCCACGGCGCCTGTTAAATCG	+439
MTHI1_RV2	GTATGGAGCATTCGAGACGCCCGC	+2422
MTHI1_FW3	GCGCATCACTACCACAAGGTGCGT	+2328
MTHI1_RV3	CTGCCGAGGAAGACACATGGGGGA	+4188
MTHI1_FW4	TCGGACACTCATCCACCCCCACAC	+3712
MTHI1_RV4	TTGCGCCACCTTACCCTACTCCCC	+5745
MTHI1_FW5	CAAGTGGCGGCGGGCAACAGCA	+4014
MTHI1_RV5	TCACCGCCCTCCAGCAGCTCCG	+5250
Oligonucleotides used for the 5'LRM-RACE identification of the 5'ends of the genes from the *atpI* transcription unit		
psbJ-5'-RACE	CAACAAGCCATAGAGGGATACGTCCAGTA	
psbJ_nested-GSP	AAACGCTCTCCAATTGATTTACAACCTTGC	
psbJ_L	GAAAGGGGAAGAGAACGTCCTTCGGAGAAT	
atpI-5'-RACE	CCTAATTCCCAGTAATAGTGCTGACCTACA	
atpI_nested-GSP	GATACTTCAGCAATTTCTAATAAAGGATTC	
atpI_L	GAGACACTCAAATATGGAGTTCCTAATACTGC	
psaJ-5'-RACE	CACCGTGAGGGTAAATAAATGTGACTGA	
psaJ_nested-GSP	ACCAAATAGTAGCAATTACAGGTGCTGTTG	
psaJ_L	GCATTAAGCAATTCACTCTTTACGCATAACT	
rps12-5'-RACE	CCCTAGCAACTTTACGAAGTGCAGAGTTTG	
rps12_nested-GSP	ACGAATTAATTGTTGAATTGTAGGCATAAAAGC	

[1] : restriction sites added in the oligonucleotide sequence for cloning or RFLP analysis purposes are written in Bold or underlined. In the mutagenic oligonucleotides, nucleotides that differ from the WT Chlamydomonas sequence are indicated in red.

References:

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Systematic biology **55**, 539-552.

Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y., and Tabata, S. (2000). Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii. DNA Res **7,** 305-307.

Boulouis, A., Drapier, D., Razafimanantsoa, H., Wostrikoff, K., Tourasse, N.J., Pascal, K., Girard-Bascou, J., Vallon, O., Wollman, F.A., and Choquet, Y. (2015). Spontaneous dominant mutations in chlamydomonas highlight ongoing evolution by gene diversification. Plant Cell **27,** 984-1001.

Chevenet, F., Brun, C., Banuls, A.L., Jacq, B., and Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics **7**, 439.

Debuchy, R., Purton, S., and Rochaix, J.D. (1989). The argininosuccinate lyase gene of *Chlamydomonas reinhardtii*: an important tool for nuclear transformation and for correlating the genetic and molecular maps of the *ARG7* locus. EMBO J, **8**, 2803-2809.

Depege, N., Bellafiore, S., and Rochaix, J.D. (2003). Role of chloroplast protein kinase Stt7 in LHCII phosphorylation and state transition in Chlamydomonas. Science **299**, 1572-1575.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., and Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res **36**, W465-469.

Dutcher, S.K., Power, J., Galloway, R.E., and Porter, M.E. (1991). Reappraisal of the genetic map of *Chlamydomonas reinhardtii*. J Hered **82**, 295-301.

Eberhard, S., Loiselay, C., Drapier, D., Bujaldon, S., Girard-Bascou, J., Kuras, R., Choquet, Y., and Wollman, F.A. (2011). Dual functions of the nucleus-encoded factor TDA1 in trapping and translation activation of atpA transcripts in *Chlamydomonas reinhardtii* chloroplasts. Plant J **67,** 1055-1066.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**, 1792-1797.

Goldschmidt-Clermont, M. (1991). Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker of site-directed transformation of chlamydomonas. Nucleic.Acids.Res. **19,** 4083-4089.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology **59**, 307-321.

Higuchi, R. (1990). Recombinant PCR. In: PCR protocols: a guide to methods and application. (London, New York: Academic press).

Kathir, P., LaVoie, M., Brazelton, W.J., Haas, N.A., Lefebvre, P.A., and Silflow, C.D. (2003). Molecular map of the Chlamydomonas reinhardtii nuclear genome. Eukaryot Cell **2,** 362-379.

Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaulot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., and Worden, A.Z. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol **12,** e1001889.

Klinkert, B., Schwarz, C., Pohlmann, S., Pierre, Y., Girard-Bascou, J., and Nickelsen, J. (2005). Relationship between mRNA levels and protein accumulation in a chloroplast promoter-mutant of *Chlamydomonas reinhardtii*. Mol Genet Genomics **274,** 637-643.

Kwan, A.L., Li, L., Kulp, D.C., Dutcher, S.K., and Stormo, G.D. (2009). Improving gene-finding in Chlamydomonas reinhardtii:GreenGenie2. BMC Genomics **10,** 210.

Loiselay, C., Gumpel, N.J., Girard-Bascou, J., Watson, A.T., Purton, S., Wollman, F.A., and Choquet, Y. (2008). Molecular identification and function of cis- and trans-acting determinants for *petA* transcript stability in *Chlamydomonas reinhardtii* chloroplasts. Mol Cell Biol **28**, 5529-5542.

Muranaka, L.S., Rutgers, M., Bujaldon, S., Heublein, A., Geimer, S., Wollman, F.A., and Schroda, M. (2016). TEF30 Interacts with Photosystem II Monomers and Is Involved in the Repair of Photodamaged Photosystem II in Chlamydomonas reinhardtii. Plant Physiol **170,** 821-840.

Purton, S., and Rochaix, J.D. (1994). Complementation of a *Chlamydomonas reinhardtii* mutant using a genomic cosmid library. Plant Mol Biol **24**, 533-537.

Ranum, L.P., Thompson, M.D., Schloss, J.A., Lefebvre, P.A., and Silflow, C.D. (1988). Mapping flagellar genes in Chlamydomonas using restriction fragment length polymorphisms. Genetics **120**, 109-122.

Raynaud, C., Loiselay, C., Wostrikoff, K., Kuras, R., Girard-Bascou, J., Wollman, F.A., and Choquet, Y. (2007). Evidence for regulatory function of nucleus-encoded factors on mRNA stabilization and translation in the chloroplast. Proc Natl Acad Sci U S A **104**, 9093-9098.

Silflow, C.D. (1998). Organization of the nuclear genome. In The molecular biology of chloroplasts and mitochondria in Chlamydomonas, J.-D. Rochaix, M. Goldschmidt-Clermont, and S. Merchant, eds (Dordrecht: Kluwer Academic Publishers), pp. 25-40.

Wostrikoff, K., Girard-Bascou, J., Wollman, F.A., and Choquet, Y. (2004). Biogenesis of PSI involves a cascade of translational autoregulation in the chloroplast of Chlamydomonas. EMBO J **23,** 2696-2705.

ARTICLE 4:

"IN VIVO STUDIES OF THE OPR RECOGNITION CODE REVEAL THE RESILIENT NATURE OF M FACTOR AND MRNA INTERACTIONS IN THE CHLOROPLAST OF CHLAMYDOMONAS REINHARDTII."

This is a draft version that will be completed when the final MDB1mut x dBMgfp-3'rbcL experiment are done. We hope to submit it in the coming months.

DRAFT ARTICLE:

IN VIVO STUDIES OF THE OPR RECOGNITION CODE REVEAL THE RESILIENT NATURE OF M FACTOR AND MRNA INTERACTIONS IN THE CHLOROPLAST OF *CHLAMYDOMONAS REINHARDTII.*

Authors: **Domitille Jarrige and Yves Choquet.**

Unité Mixte de Recherche (UMR) 7141, Centre National de la Recherche Scientifique (CNRS) and Sorbonne Université, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, F-75005 Paris, France

*: Corresponding author: choquet@ibpc.fr

Phone : +33 (0)1 58 41 50 75

Keywords: Nucleus-encoded trans-acting factors; OPR proteins, recognition code, mRNA stability, chloroplast, *Chlamydomonas reinhardtii*, protein/RNA

Abstract to be written once the MDB1mut x *dB$_M$gfp-3'rbcL* blots are done.

INTRODUCTION

Following endosymbiosis around 1.5 billion years ago, the cyanobacterial ancestor of the chloroplast underwent deep modifications. Obsolete gene loss and extensive gene transfer events to the host nucleus led to a drastic reduction of the plastid genome (Martin *et al.*, 1998; Timmis *et al.*, 2004; Moustafa and Bhattacharya, 2008). Many genes remaining in the chloroplast are still expressed in polycistronic transcription units, but the original cyanobacterial operon organisation of the genes has been largely lost. In the green alga *Chlamydomonas reinhardtii*, the chloroplast genome has been completely shuffled and the genes encoding the subunits of a same photosynthetic protein complex are most often separated (Maul *et al.*, 2002), while genes co-transcribed in polycistrons may contribute to different functions. To tune the expression of the various subunits of the plastid complexes, the chloroplast relies on post-transcriptional mechanisms (Rochaix, 1996; Choquet and Wollman, 2002). Moreover, due to gene transfer, many essential chloroplast polypeptides are now encoded in the nuclear genome and imported back into the plastid. To ensure the chloroplast viability the expression of chloroplast and nucleus encoded subunits needs to be synchronised.

Transcripts in the chloroplast stroma are subjected to a high exonuclease activity. To be stabilised, matured and translated plastid transcripts need nuclear gene products (Barkan and Goldschmidt-Clermont, 2000) the organellar trans acting factors (OTAF). OTAF are gene specific factors, able to recognise and bind specific organellar mRNA to promote their expression, either by maturation, stabilisation, editing, splicing, translation initiation… Among OTAFs the PPR (pentatricopeptide) protein family, which was defined twice independently twenty years ago (Aubourg *et al.*, 2000; Small and Peeters, 2000), has been extensively studied in land plants (Lurin *et al.*, 2004; Colcombet *et al.*, 2013; Barkan and Small, 2014) PPR proteins bear tandem repeats of a degenerate 35 amino acids motif, which fold into two antiparallel α-helices (Ringel *et al.*, 2011) and stack into a super helix, with the first helix of each repeat inside the groove. This motif was predicted to interact with the mRNA thanks to positive residues forming a continuous surface inside the groove (Small and Peeters, 2000). Further understanding of the mRNA/PPR motif interactions was achieved by molecular, computational and structural studies in the past decade (Prikryl *et al.*, 2011; Kobayashi *et al.*, 2012; Yagi *et al.*, 2013; Yin *et al.*, 2013; Ban *et al.*, 2013; Gully *et al.*, 2015) and a PPR recognition code was established *in vitro* and *in silico* (Barkan *et al.*, 2012; Yagi *et al.*, 2013), linking the nature of residues at specific positions in the first helix with the affinity for a specific nucleotide. This code was demonstrated experimentally (Barkan *et al.*, 2012), by introducing mutations at key positions of the 6^{th} and 7^{th} repeats of the maize protein PPR10 and looking *in vitro* at its binding affinity for a range of target RNA varying at the 6^{th} and 7^{th} nucleotides. Specific nucleotide recognition was thus established for several combinations of key residues.

In *Chlamydomonas reinhardtii* most OTAFs are octotricopeptide repeat (OPR) proteins. Those proteins are characterised by tandem repeats of a degenerated 38 amino-acid motif. OPR proteins are scarce in land plants and have been mostly studied in *C. reinhardtii*. They are implicated in translation initiation (Zerges and Rochaix, 1994; Stampacchia *et al.*, 1997; Auchincloss *et al.*, 2002; Eberhard *et al.*, 2011; Rahire *et al.*, 2012), RNA stabilisation and maturation (Drager *et al.*, 1998; Wang *et al.*, 2015; Viola *et al.*, 2019) and possibly RNA degradation (Boulouis *et al.*, 2015). While the OPR motif is not homologous to the PPR motif, OPR repeats are predicted to fold similarly into two antiparallel α-helices and to form an α-solenoid structure when stacked together. This has been confirmed by a recent single-particle cryo-electron microscopy analysis of *Polytomella sp.* mitochondrial ATP synthase, which revealed the structure of ASA2, an OPR protein (Murphy *et al.*, 2019). As for the PPR repeat, the first helix of the OPR repeat also lay in the groove of the super-helix. This first helix is more conserved than the second one, hinting that it is more important for the function of the motif. However, specific positions in this conserved helix are highly variable. By comparing the occurrence of those variable residues with the known target nucleotides of OPR proteins, we have established a draft OPR recognition code. In the present study, we took advantage of *Chlamydomonas reinhardtii* suitability to mutagenesis both in nucleus and chloroplast to decipher the OPR recognition code *in vivo,* with all the potentials physiological factors that may play a role in the OPR protein/mRNA interaction.

We worked with two OPR stabilisation factors: MDB1, an 1137 amino-acid protein, bearing 13 OPR motifs, required for the stabilisation of the chloroplast *atpB* mRNA, encoding subunit β of the chloroplast ATP synthase (Drapier *et al.*, 1992). Moreover, the addition of a poly-G track at the 5' end of *atpB* mRNA restored its accumulation in absence of MDB1, suggesting that this OPR factor is implicated in *atpB* transcript stabilisation by preventing its degradation by 5'→3' exonucleases (Cavaiuolo et al, in preparation). A foot-print in *atpB* mRNA 5' region, showing a specific protection from nucleases, disappeared in a mutant devoid of MDB1 (Cavaiuolo *et al.*, 2017). The 13 nucleotides putative binding site of MDB1 in that *atpB transcript* foot-print will be called the target sequence in this study.

We also studied the MTHI1 factor, a 9 OPR repeat protein described in much detail in (papier Shin). MTHI1 is a dual stabilisation and translation factor of the plastid *atpH* mRNA. MTHI1 9 nucleotides target sequence on *atpH* mRNA was also revealed by a foot-print (Cavaiuolo *et al.*, 2017).

RESULTS

Establishment of a draft OPR code

To be expended upon (table1)

Mutated *atpB* transcripts still accumulated except in one mutant strain

Point mutations were introduced in the MDB1 target sequence of *atpB* (fig 1). The aim was to test the importance of different parts of the target as well as the effect of steric clashes; when a pyrimidine is replaced with a larger purine. Those mutated targets were then inserted at the *atpB* locus along a spectinomycine resistance cassette in the chloroplast of *ΔatpB*$_1$ cells, to avoid recombination of the target with the endogenous *atpB*. After homoplasmisation, all variants were still phototrophic, bar dB12. Accumulation levels of the mutated *atpB* mRNA in transformed cells were assessed by RNA blot. 3 independent transformants analysed for each target variant proved almost identical. A representative transformant for each variant is shown in fig1. Most variants kept a high accumulation of *atpB* mRNA, compared to the control construct (dBct). Whether big or small (4 to 2 nucleotides), mutations in the 5' half of the target did not strongly destabilise the *atpB* mRNA, which accumulated to roughly 60-75 % of the control level. Mutations in the central region had a stronger effect but still did not prevent *atpB* mRNA accumulation. Mutations in the 3' half of the target had also weak effects, except for the dB12 mutant (4 contiguous nucleotides substituted, with the introduction of three steric clashes), which saw a sharp drop in *atpB* transcript accumulation (1% of the dBct level), in agreement with the loss of phototrophic growth.

Thus, the MDB1-mediated stabilisation of the *atpB* transcript was quite resilient, paradoxically suggesting that the *atpB* mRNA/MDB1 interaction is poorly sequence specific. However, in previous studies of *mdb1* mutants (Drapier *et al.*, 1992) MDB1 absence had no effect on the expression of other chloroplast genes and alternative MDB1 footprints could not be identified (Cavaiuolo *et al.*, 2017).

Comparisons between a previous study (Anthonisen *et al.*, 2001) and our results did confirm the relative importance of the 3' half of the target compared to the 5' part (fig2). But strikingly, the effects of our point mutations on transcript accumulation were far weaker than those observed by the authors. In that study, a chimeric construct was used, instead of the whole *atpB*, bearing only the first 72 nt of the *atpB* 5' UTR (i.e. including MDB1 footprint), fused to the bacterial *uidA* coding sequence and followed by the *psaB* 3' UTR. Those major differences might indicate that other regions of the *atpB* transcript are implicated in its stabilisation. This could perhaps involve other unknown auxiliary factors, specific to *atpB*, able to interact with *atpB* mRNA and to enhance MDB1 binding specificity on its target transcript in a tripartite complex.

Mutated *atpH* transcripts remained stable in most cases

To determine whether *atpB* resilient stabilisation was a lone case in *C. reinhardtii* we decided to attempt similar experiments on *atpH*. Point mutations were created in *atpH* MTHI1 target sequence (fig3), a spectinomycin selection cassette was added and the constructions were transformed into a *ΔatpH* strain in place of the deleted *atpH* locus.

Transcript accumulation of the MTHI1 *atpH* target variants was estimated by RNA blot (fig3). The control target sequence strains ΔdH_{Ct}, accumulated the *atpH* transcript stably and could perform photosynthesis. While both ΔdH_1 and ΔdH_2 transformed strains did accumulate less *atpH* transcripts than the WT or the control strain, they were still able to grow photo-autotrophically. The mutation in the 5' part of the *atpH* target seemed to cause more effect on *atpH* recognition by MTHI1 than the mutation in the 3'part. The ΔdH_3 transformed strains, in which the target is nearly unrecognisable, accumulated very low levels of *atpH* mRNA, equivalent to that of the *mthi1-2* strain, and were not phototrophic. As in *atpB* case, to disrupt *atpH* transcript stabilisation by its M factor, large mutations were needed.

MTH2 an auxiliary factor of MTHI1?

MTH2 (Cre10.g461700) is a dispensable gene implicated in *atpH* mRNA stabilisation; it encodes a putative protein of 3219 amino acids, with no discernible domains. This protein has orthologues in closely related Chlamydomonaceae. In an insertional mutant of MTH2 L63a *atpH* mRNA accumulated to about one tenth of its wild-type level. In another independent insertional mutant of MTH2: *mth2-2*, similarly the *atpH* transcript accumulated also to 10% of the WT.T222+ level (fig4). Those mutants managed to grow on minimum media, indicating that translation of *atpH* could still occur even in absence of MTH2. This dispensable MTH2 could be an auxiliary factor of MTHI1. To investigate further this possible involvement in the interaction between the MTHI1 OPR factor and its target *atpH* mRNA, *mth2-2* strains were transformed with either the ΔdH_{Ct} or the ΔdH_2 constructs. After achieving homoplasmy, growth test of the strains were done, on TAP or minimum medium. All the tested strains were still able to grow on minimum medium, albeit very slowly for the *mth2-2* {dH_2} strains. Total RNA was extracted from the strains and loaded on RNA blots (fig4). The *mth2-2* {dH_{Ct}} accumulated similar levels of *atpH* transcript as *mth2-2*, highlighting the implication of MTH2 in *atpH* stabilisation. Surprisingly considering their ability to grow photo-autotrophically, no detectable amount of *atpH* mRNA was found in the *mth2-2* {dH_2} transformants, the accumulation level was probably under the detection threshold in this blot. (Blot to be remade)

A new strategy to study the code with MDB1: a chimeric approach

Chimeric mRNA bearing MDB1 target sequence are far more easily destabilised

To see whether we could achieve similar results by alleviating the influence of potential auxiliary factors of MDB1, we designed a new chimeric reporter, made of the same 72bp of the *atpB* 5'UTR than in (Anthoninsen et al, 2001), fused to the coding sequence of *Azotobacter vinelandii* green fluorescence protein (GFP) and followed by the *rbcL* 3'UTR. As described in Fig.5, different point mutations were introduced in the MDB1 target sequence. These chimeras, associated with an *aadA* cassette for the selection of transformants, were

inserted in a neutral site, downstream of the *petA* gene, and introduced in the chloroplast genome of the WT and/or *ΔatpB* strains.

Because the chimera does not contain the *atpB* sequences required for translation initiation, the *gfp* sequence could only be used to probe transcript levels in RNA-blot experiments. Constructs bearing three consecutive Spinach2 (Strack *et al.*, 2013) light-up aptamers between the *gfp* CDS and the 3'*rbcL* were constructed, to try to visualise *in vivo* the chimeric RNA in the cells. However, while the mRNA did accumulate similarly as the simple *gfp* chimera, the fluorescence signals proved too low to detect in cells (data not shown).

When introduced in the chloroplast genome of the Δ*atpB* recipient strain, the control constructs (with or without Spinach2) with a wild-type target sequence were strongly accumulated, showing that the missing part of *atpB* 5'UTR is not necessary for MDB1-mediated stabilisation. And as expected, Spinach2 constructs failed to accumulate in an *mdb1* mutant context. (Fig.7).

By comparison, mRNA levels of mutated chimeras were quite decreased (Fig.5). Compared to the control transformed strain, the CC_1, TT_1 and GG_1 point mutations led to a 70-80 % decrease of transcript accumulation, while the CC_2, AA_2 and GG_2 mutations induced a drastic drop in transcript abundance. In the CC_2 and GG_2 variants, accumulation of the *gfp* mRNA was nearly abolished, at around 3-4%. Again, mutations in the 3'part of the MDB1 target sequence had a stronger effect than those in the 5' part. The absence of most of *atpB* sequence had a sizeable impact on MDB1-mediated transcript stability.

Effects of atpB presence on the chimeric RNA accumulation levels

The control reporter was also transformed in a WT.T222+ strain, to assess the impact of the presence of the endogenous *atpB* mRNA on the *gfp* reporter accumulation. As shown in Fig. 6 and Fig.7, the accumulation of the chimeric *gfp* mRNA was reduced about ten-fold in the presence of the endogenous *atpB*. In contrast, *atpB* mRNA accumulation was only slightly affected by the presence of the chimera, and remained at ~75 % of its level in the wild-type strain. Those results could indicate that *atpB* was more stabilised in a competition with the chimera.

We then transformed the CC_1, TT_1 and GG_1 chimeric variants into the WT.T222+ stain. Strikingly, at variance with the control chimera, the mutated chimeric transcript could not be detected (fig.8). Upon competition with the endogenous *atpB* mRNA, mutations in the target completely abolished the interaction with MDB1, hence the absence of accumulation of the chimeric transcripts.

Two main hypotheses (fig.9) could explain why MDB1 is more sensitive to modifications of its specific target sequence in absence of the whole *atpB* sequence. The first hypothesis would be that auxiliary factors recognising other regions of *atpB* mRNA could be implicated in maintaining MDB1 on its target, possibly by forming a ternary complex with *atpB* mRNA

and MDB1. This could also explain why the endogenous *atpB* is favoured by MDB1: this factor might also recruit MDB1 on *atpB* transcript or keep MDB1 sequestrated on its canonical target. Another hypothesis is that other regions of the *atpB* transcript are directly implicated in the stabilisation process. It was shown that the 3'UTR of *C. reinhardtii* chloroplastic mRNA could be implicated in its translation (Rott *et al.*, 1998) and the maturation of its 5'UTR (cf Cavaiuolo article *atpB).* Notably, *rbcL* 3'UTR prevented the 5'terminus maturation of 5'*atpB* chimeric constructs, leading to a reduced accumulation and expression of the chimeric transcripts.

Effects of the 3' end of atpB

To test whether the 3'end was responsible for MDB1 destabilisation on the *gfp* reporters, another construct containing as previously described: 5'$atpB_{SH}$, then the *gfp* CDS, three Spinach2 and terminated by *atpB* 3'UTR was designed. The construct was transformed into the WT.T222+ and $\Delta atpB_1$ strains. Accumulation of *atpB* and *gfp* transcripts was assessed by RNA blot (Fig.7). The *gfp* construct bearing the 3'*atpB* was not significantly more stabilised when in competition with the endogenous *atpB* transcript. This suggests that the 3'end is not the crucial part granting more stability to *atpB* mRNA in a competition. Surprisingly, the chimeric transcript in the $\Delta atpB_1$ strain accumulated slightly less when it was terminated by a 3'*atpB* instead of a 3'*rbcL*. Moreover, *atpB* mRNA was even more accumulated (to nearly wild-type levels) when the chimera in competition had a 3'*atpB*. Those results suggest that the 3' end of *atpB* mRNA might instead be a target for specific degradation and that the presence of a chimeric transcript with the same 3' end might alleviate that pressure on the endogenous transcript.

MDB1 mutagenesis

Mutations in MDB1 sequence were introduced to modify the putative key residues at the fifth and sixth positions in the OPR repeat 6 and 7 or 11 and 12 according to the draft OPR recognition code (Fig.10). To confront those modified MDB1 proteins with the various target mRNA we transformed those mutations into the nucleus genome of $thm24.2^-$, a knockout strain of *MDB1*. Transformants were tested for the recovery of photo-autotrophy by fluorescence imaging and the levels of tagged MDB1 protein was assessed by immunoblot, as the insertion was random the expression levels of tagged MDB1 in independent transformants were highly different. A transformant expressing the highest amount of MDB1 was picked for each mutation. Those MDB1 mutants were then crossed with the $\Delta atpB_1^+$ strains carrying the chloroplastic chimeras. As the chloroplast genome in *C. reinhardtii* is only transmitted by the mating type + parent, the progeny was plated and selected on paromycin and spectinomycin resistance, to kill parental cells and checked for the presence of *MDB1* knock-out allele $thm24.2^-$.

RNA blots

To be performed

Discussion

Conclusions on OPR code

To be written once the MDB1mut x dBMgfp-3'rbcL blots are done.

Conclusion on potential secondary factors

Specific sequences in the 5'UTR part that was deleted in the *gfp* chimera are important for the transcript stabilisation. This stabilisation could be linked to the MDB1-mediated one, as the chimeric transcripts, driven by the complete *atpB* 5'UTR accumulates at about the same level than the endogenous *atpB* transcript in a competition (ref article1). This could mean that unknown factor(s) could either recruit MDB1 on *atpB* transcript or anchor MDB1 by improving its affinity for its target RNA in a ternary complex for instance. This observation is also coherent with the lower resilience of the MDB1/*atpB* target when the end of the 5'UTR is absent. While organellar gene expression depends on crucial factors, those do not work by themselves and organellar mRNA are probably expressed by a suite of factors, influencing each other, the M factors being the corner stone of those resilient expression systems. Considering the growing number of M and T factors co-stabilising mRNA in *C. reinhardtii (ref)*, those tripartite or higher order complexes might be widespread in the chloroplast.

Methods

Strains, media and growth conditions

Chlamydomonas reinhardtii wild-type, mutant and transformed strains, derived from 137c, were grown in Tris-acetate-phosphate medium (TAP), pH 7.2 at 25°C, under constant illumination at 5 to 10 $\mu E.m^{-2}s^{-1}$. Strains are listed in the (table)

Crosses

Crosses were performed according to (Harris, 1989). Descendants were selected on paromycin (5µg/mL) supplemented TAP plates, and then on TAP-spectinomycin plates (500µg/mL). The presence of the *thm24.2*$^{-}$ *MDB1* allele was assessed by PCR amplification of *MDB1* with MDB1 bFW and MDB1 aRV and subsequent digestion by BsrI.

Nucleic acid manipulations and plasmid construction

Standard nucleic acid manipulations were performed according to (Sambrook *et al.*, 1989). Primers and plasmids used in this study are listed in the Supplementary materials section. Every DNA construct was sequenced before transformation in *C. reinhardtii*.

Target variants of *atpH*

Three steps PCR mutagenesis was performed on patpHKX with the cemAFW and Mut-atpH-*RV primers (see table) on one hand and atpHext-RV and Mut-atpH-*FW on the other. The two products were gel purified and combined by mega priming with cemAFW and atpHext-RV. A cemAFW and atpHext-RV PCR was also performed on patpHKX to recover the wild-

type MTHI1 target. Final products were digested with EcoRv and EcoRI and integrated back into the patpHKX vector at the same sites. The final plasmids were then used for chloroplast transformation.

Target variants of *atpB*

Three steps PCR mutagenesis was performed on p147 with on one hand dBExt-RV and dB*FW and on the other atpB5'FWx and dB*RV. The two different PCR products were then purified and put together for mega priming with dBExt-RV and atpB5'FWx. The final product was then digested by XhoI and BseRI and cloned into p147 at the same sites. Those constructions were targeted to the endogenous *atpB* locus.

gfp chimeric constructs

The paAKX plasmid (Wostrikoff et al, 2004) was digested by *Apa*I and *Ale*I to retrieve a 2509 bp fragment containing a spectinomycin resistance cassette (the aminoglyside 3' adenyl transferase coding sequence: *aadA* (Goldschmidt-Clermont, 1991) driven by the *psaA* 5'UTR and followed by *rbcL* 3'UTR. This cassette is also flanked by two direct repeats (a fragment of the *tet* gene conferring resistence to tetracyclin (Fischer *et al.*, 1996)) of 485 bp, to create a recycling *aadA* cassette self-excising by spontaneous homologous recombination as described in (Fischer *et al.*, 1996). After a Klenow treatment this fragment was inserted into the pWF plasmid (which contains chloroplastic sequences targeting the insertion in a neutral locus, next to *petA*) at the HincII site, giving the pWFaAKX plasmid.

A 756 bp *Azotobacter vinelandii* green fluorescence protein sequence was amplified by PCR from pGFP with the GFP-CDS_FW and GFP_CDS_RV2 primers (cf tab primers). This DNA fragment was digested by *Pst*I and *Eco*RI and integrated into the corresponding sites in the paAKRaA plasmid (Fu *et al.*, 2017) to place the *gfp* sequence in front of the *rbcL* 3'UTR, giving the pgfpRaA plasmid.

pgfpRaA was then digested by *Bam*HI and *Xho*I, the 946bp fragment was inserted into pWFaAKX at the *Xho*I and *Bgl*II sites, this yielded the pWFaAKXgfpR plasmid.

Fragments of *atpB* 5'UTR with the mutated target were obtained either: by PCR amplification of previously used plasmids p^{Kr}atpB, patpB$_{CC}$, patpB$_{TT}$, patpB$_{GG}$ with the atpB-Anton_FW and atpB_Anton_WT_RV primers, or by PCR mutagenesis with primers atpB-Anton_FW and atpB-Anton-M1_RV, atpB-Anton-M2_RV, atpB-Anton-M3_RV using p147 as template. All those 144bp amplifications products were then digested by *Xma*I and *Xho*I and inserted at the corresponding sites in pWFaAKXgfpR. The final pWFaAKX-dB(WT)gfpR, pWFaAKX-dB(CC_1)gfpR, pWFaAKX-dB(TT_1)gfpR, pWFaAKX-dB(GG_1)gfpR, pWFaAKX-dB(CC_2)gfpR, pWFaAKX-dB(AA_2)gfpR, pWFaAKX-dB(GG_2)gfpR were thus obtained.

A triplet of consecutive Spinach2 aptamers (Strack *et al.*, 2013) sequences, separated by restriction sites, was ordered from Genscript and cut with *Mfe*I and *Pst*I and inserted in pgfpRaA to give pgfp-Spinach2x3-RaA. This plasmid was subjected to the same cloning

procedure as previously described for pgfpRaA to obtain pWFaAKX-dB(WT)gfp-Spinach2x3-R. To obtain the construct with 2 spinach2, pWFaAKX-dB(WT)gfp-Spinach2x3-R was digested with *EcoR*I and *Hpa*I, and after a Klenow treatment to fill in the overhangs, was ligated. This yielded the plasmid pWFaAKX-dB(WT)gfp-Spinach2x2-R.

To design the chimeric construct with *atpB* 3'UTR instead of *rbcL* 3'UTR, a synthetic sequence was ordered and introduced into the pWFaAKX-dB(WT)gfp-Spinach2x3-R giving the pWFaAKX-dB(WT)gfp-Spinach2x3-B (GenScript).

MDB1 variants

Synthetic DNA sequences were ordered from Genscipt and inserted either in the vector MDB1-HA-pJFL between the XhoI and BglII sites for MDB1-CC1-HA, MDB1-GG1-HA, and MDB1-UU1-HA or in the vector pMDB1 -HA-Strep-JHL between the NsiI and SnaBI sites for MDB1-CC2-HA-Strep, MDB1-GG2-HA-Strep and MDB1-AA2-HA-Strep (GenScript).

Chloroplast transformation

Chloroplast transformation by tungsten microbeads bombardment (Boynton *et al.*, 1988) was conducted essentially as described (Kuras and Wollman, 1994) except that the cells were directly transformed on TAP-spectinomycin (100µg/mL) plates. Resulting transformants were sub-cloned on TAP-spec (500µg/mL) for several generations. Homoplasmy was assessed by PCR amplification of the construct sequence and disruption of the recipient loci.

Nucleus transformation

After linearization of the transformation plasmids by ScaI or AhdI, nucleus transformation was conducted in CHES buffer as described (Onishi and Pringle, 2016) with the following parameters: 600V, 25µF and 1000Ω. Transformants were selected on paromycin (5µg/mL) supplemented TAP plates.

RNA analysis

RNA extraction and RNA gels were performed as in (Drapier and Wollman, 1998). Digoxigenin (DIG) labelled DNA probes were generated by PCR (Roche) and hybridised on the nylon filter bound RNA. The probes were then bound by anti-DIG antibodies and incubated with CDP-*Star* (Roche), chemiluminescence was then detected with a Chemidoc. Transcript quantification was done using the image lab software.

Protein analysis

Immunoblots were performed on exponentially growing cells (2 x 10^6 cells/mL) according to (Kuras and Wollman, 1994). Cell extracts were loaded in 8-16% acrylamide gels (Biorad), on an equal chlorophyll basis. Anti-tubulin, anti-cytochrome *f* , anti-β-CF1, anti-HA antibodies were used. And detected either by anti-mouse IgG or anti-rabbit IgG antibodies.

Fluorescence live-imaging

Fluorescence of live cells on plates was measured with a SpeedZen camera (Beambio).

Bibliographie

Anthonisen, I. L., Salvador, M. L. and Klein, U. (2001) Specific sequence elements in the 5' untranslated regions of rbcL and atpB gene mRNas stabilize transcripts in the chloroplast of Chlamydomonas reinhardtii. RNA (New York, N.Y.), 7, 1024-1033.

Aubourg, S., Boudet, N., Kreis, M. and Lecharny, A. (2000) In Arabidopsis thaliana, 1% of the genome codes for a novel protein family unique to plants. Plant molecular biology, 42, 603-613.

Auchincloss, A. H., Zerges, W., Perron, K., Girard-Bascou, J. and Rochaix, J. D. (2002) Characterization of Tbc2, a nucleus-encoded factor specifically required for translation of the chloroplast psbC mRNA in Chlamydomonas reinhardtii. The Journal of cell biology, 157, 953-962.

Ban, T., Ke, J., Chen, R., Gu, X., Tan, M. H., Zhou, X. E., Kang, Y., Melcher, K., Zhu, J. K. and Xu, H. E. (2013) Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. The Journal of biological chemistry, 288, 31540-31548.

Barkan, A. and Goldschmidt-Clermont, M. (2000) Participation of nuclear genes in chloroplast gene expression. Biochimie, 82, 559-572.

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y. S., Bond, C. S. and Small, I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS genetics, 8, e1002910.

Barkan, A. and Small, I. (2014) Pentatricopeptide repeat proteins in plants. Annual review of plant biology, 65, 415-442.

Boulouis, A., Drapier, D., Razafimanantsoa, H., Wostrikoff, K., Tourasse, N. J., Pascal, K., Girard-Bascou, J., Vallon, O., Wollman, F. A. and Choquet, Y. (2015) Spontaneous dominant mutations in chlamydomonas highlight ongoing evolution by gene diversification. The Plant cell, 27, 984-1001.

Boynton, J. E., Gillham, N. W., Harris, E. H., Hosler, J. P., Johnson, A. M., Jones, A. R., Randolph-Anderson, B. L., Robertson, D., Klein, T. M., Shark, K. B. and et al. (1988) Chloroplast transformation in Chlamydomonas with high velocity microprojectiles. Science, 240, 1534-1538.

Cavaiuolo, M., Kuras, R., Wollman, F. A., Choquet, Y. and Vallon, O. (2017) Small RNA profiling in Chlamydomonas: insights into chloroplast RNA metabolism. Nucleic acids research, 45, 10783-10799.

Choquet, Y. and Wollman, F. A. (2002) Translational regulations as specific traits of chloroplast gene expression. FEBS letters, 529, 39-42.

Colcombet, J., Lopez-Obando, M., Heurtevin, L., Bernard, C., Martin, K., Berthome, R. and Lurin, C. (2013) Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. RNA biology, 10, 1557-1575.

Drager, R. G., Girard-bascou, J., Choquet, Y., Kindle, K. L. and Stern, D. B. (1998) In vivo evidence for 5'→3' exoribonuclease degradation of an unstable chloroplast mRNA. The Plant Journal, 13, 85-96.

Drapier, D., Girard-Bascou, J. and Wollman, F. A. (1992) Evidence for Nuclear Control of the Expression of the atpA and atpB Chloroplast Genes in Chlamydomonas. The Plant cell, 4, 283-295.

Drapier, D., Suzuki, H. , Levy, H. , Rimbault, B. , Kindle, K. L. , David B. Stern, D. B. , and and Wollman, F. A. (1998) The Chloroplast atpA Gene Cluster in Chlamydomonas reinhardtii. Plant physiology, 629–641.

Eberhard, S., Loiselay, C., Drapier, D., Bujaldon, S., Girard-Bascou, J., Kuras, R., Choquet, Y. and Wollman, F. A. (2011) Dual functions of the nucleus-encoded factor TDA1 in trapping and translation activation of atpA transcripts in Chlamydomonas reinhardtii chloroplasts. The Plant journal : for cell and molecular biology, 67, 1055-1066.

Fischer, N., Stampacchia, O., Redding, K. and Rochaix, J. D. (1996) Selectable marker recycling in the chloroplast. Mol Gen Genet, 251, 373-380.

Fu, H. Y., Picot, D., Choquet, Y., Longatte, G., Sayegh, A., Delacotte, J., Guille-Collignon, M., Lemaitre, F., Rappaport, F. and Wollman, F. A. (2017) Redesigning the QA binding site of Photosystem II allows reduction of exogenous quinones. Nature communications, 8, 15274.

Goldschmidt-Clermont, M. (1991) Transgenic expression of aminoglycoside adenine transferase in the chloroplast: a selectable marker of site-directed transformation of chlamydomonas. Nucleic acids research, 19, 4083-4089.

Gully, B. S., Cowieson, N., Stanley, W. A., Shearston, K., Small, I. D., Barkan, A. and Bond, C. S. (2015) The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atpH RNA. Nucleic acids research, 43, 1918-1926.

Harris, E. H. (1989) The Chlamydomonas Source Book: A Comprehensive Guide to Biology and Laboratory Use., Academic Press, San Diego.

Kobayashi, K., Kawabata, M., Hisano, K., Kazama, T., Matsuoka, K., Sugita, M. and Nakamura, T. (2012) Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. Nucleic acids research, 40, 2712-2723.

Kuras, R. and Wollman, F. A. (1994) The assembly of cytochrome b6/f complexes: an approach using genetic transformation of the green alga Chlamydomonas reinhardtii. The EMBO journal, 13, 1019-1027.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M. L., Mireau, H., Peeters, N., Renou, J. P., Szurek, B., Taconnat, L. and Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. The Plant cell, 16, 2089-2103.

Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M. and Kowallik, K. V. (1998) Gene transfer to the nucleus and the evolution of chloroplasts. Nature, 393, 162-165.

Maul, J. E., Lilly, J. W., Cui, L., dePamphilis, C. W., Miller, W., Harris, E. H. and Stern, D. B. (2002) The Chlamydomonas reinhardtii plastid chromosome: islands of genes in a sea of repeats. The Plant cell, 14, 2659-2679.

Moustafa, A. and Bhattacharya, D. (2008) PhyloSort: a user-friendly phylogenetic sorting tool and its application to estimating the cyanobacterial contribution to the nuclear genome of Chlamydomonas. BMC evolutionary biology, 8, 6.

Murphy, B. J., Klusch, N., Langer, J., Mills, D. J., Yildiz, Ö. and Kühlbrandt, W. (2019) Rotary substates of mitochondrial ATP synthase reveal the basis of flexible F1-Fo coupling. Science, 364, eaaw9128.

Onishi, M. and Pringle, J. R. (2016) Robust Transgene Expression from Bicistronic mRNA in the Green Alga Chlamydomonas reinhardtii. G3, 6, 4115-4125.

Prikryl, J., Rojas, M., Schuster, G. and Barkan, A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. Proceedings of the National Academy of Sciences of the United States of America, 108, 415-420.

Rahire, M., Laroche, F., Cerutti, L. and Rochaix, J. D. (2012) Identification of an OPR protein involved in the translation initiation of the PsaB subunit of photosystem I. The Plant journal : for cell and molecular biology, 72, 652-661.

Ringel, R., Sologub, M., Morozov, Y. I., Litonin, D., Cramer, P. and Temiakov, D. (2011) Structure of human mitochondrial RNA polymerase. Nature, 478, 269-273.

Rochaix, J. D. (1996) Post-transcriptional regulation of chloroplast gene expression in Chlamydomonas reinhardtii. Plant molecular biology, 32, 327-341.

Rott, R., Liveanu, V., Drager, R. G., Stern, D. B. and Schuster, G. (1998) The sequence and structure of the 3'-untranslated regions of chloroplast transcripts are important determinants of mRNA accumulation and stability. Plant molecular biology, 36, 307-314.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning. , Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Small, I. D. and Peeters, N. (2000) The PPR motif – a TPR-related motif prevalent in plant organellar proteins. Trends Biochem. Sci., 25, 46-47.

Stampacchia, O., Girard-Bascou, J., Zanasco, J. L., Zerges, W., Bennoun, P. and Rochaix, J. D. (1997) A nuclear-encoded function essential for translation of the chloroplast psaB mRNA in chlamydomonas. The Plant cell, 9, 773-782.

Strack, R. L., Disney, M. D. and Jaffrey, S. R. (2013) A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. Nature methods, 10, 1219-1224.

Timmis, J. N., Ayliffe, M. A., Huang, C. Y. and Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nature reviews. Genetics, 5, 123-135.

Viola, S., Cavaiuolo, M., Drapier, D., Eberhard, S., Vallon, O., Wollman, F. A. and Choquet, Y. (2019) MDA1, a nucleus-encoded factor involved in the stabilization and processing of the atpA transcript in the chloroplast of Chlamydomonas. The Plant journal : for cell and molecular biology.

Wang, F., Johnson, X., Cavaiuolo, M., Bohne, A. V., Nickelsen, J. and Vallon, O. (2015) Two Chlamydomonas OPR proteins stabilize chloroplast mRNAs encoding small subunits of photosystem II and cytochrome b6 f. The Plant journal : for cell and molecular biology, 82, 861-873.

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T. and Nakamura, T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. PloS one, 8, e57286.

Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H. W., Wang, J., Zhu, J. K., Shi, Y. and Yan, N. (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. Nature, 504, 168-171.

Zerges, W. and Rochaix, J. D. (1994) THE 5' LEADER OF A CHLOROPLAST MESSENGER-RNA MEDIATES THE TRANSLATIONAL REQUIREMENTS FOR 2 NUCLEUS-ENCODED FUNCTIONS IN CHLAMYDOMONAS-REINHARDTII. Mol. Cell. Biol. , 14, 5268-5277.

Position	Residue										
3	x	x	x	x	R,K	x	x	x	x	x	x
4	x	P	x	x	x - P	x	x	x	x	x	x
5	x	Q	X - R, K	R, K	x	R	X - R	R, Q	x	R	R
6 (% occurrence)	E (22,24)	G (15,36)	D (14,66)			Q (10,02)		A (9,11)	H (7,68)	S (7,34)	N (7,09)
Recognised nucleotide	U	A	G	U	U	U	?	A	?	A	?

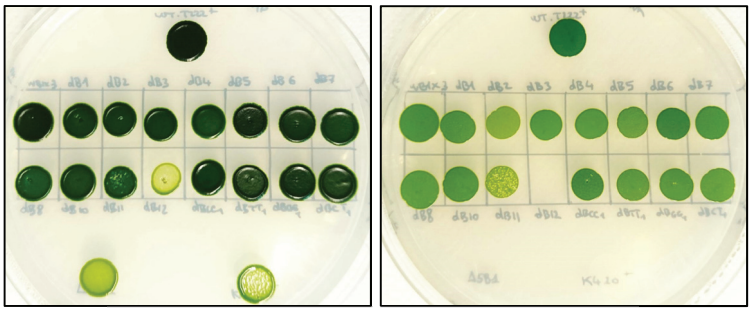
Table 1: Draft OPR recognition code.

A

5'atpB$_M$:atpB:3'atpB

B

WT							
dBCt	dB1	dB2	dB3	dB4	dB5	dB6	dB7
dB8	dB10	dB11	dB12	dBCC$_1$	dBTT$_1$	dBGG$_1$	dBCT$_1$
ΔatpB1				K4.20			



TAP medium | min medium

C

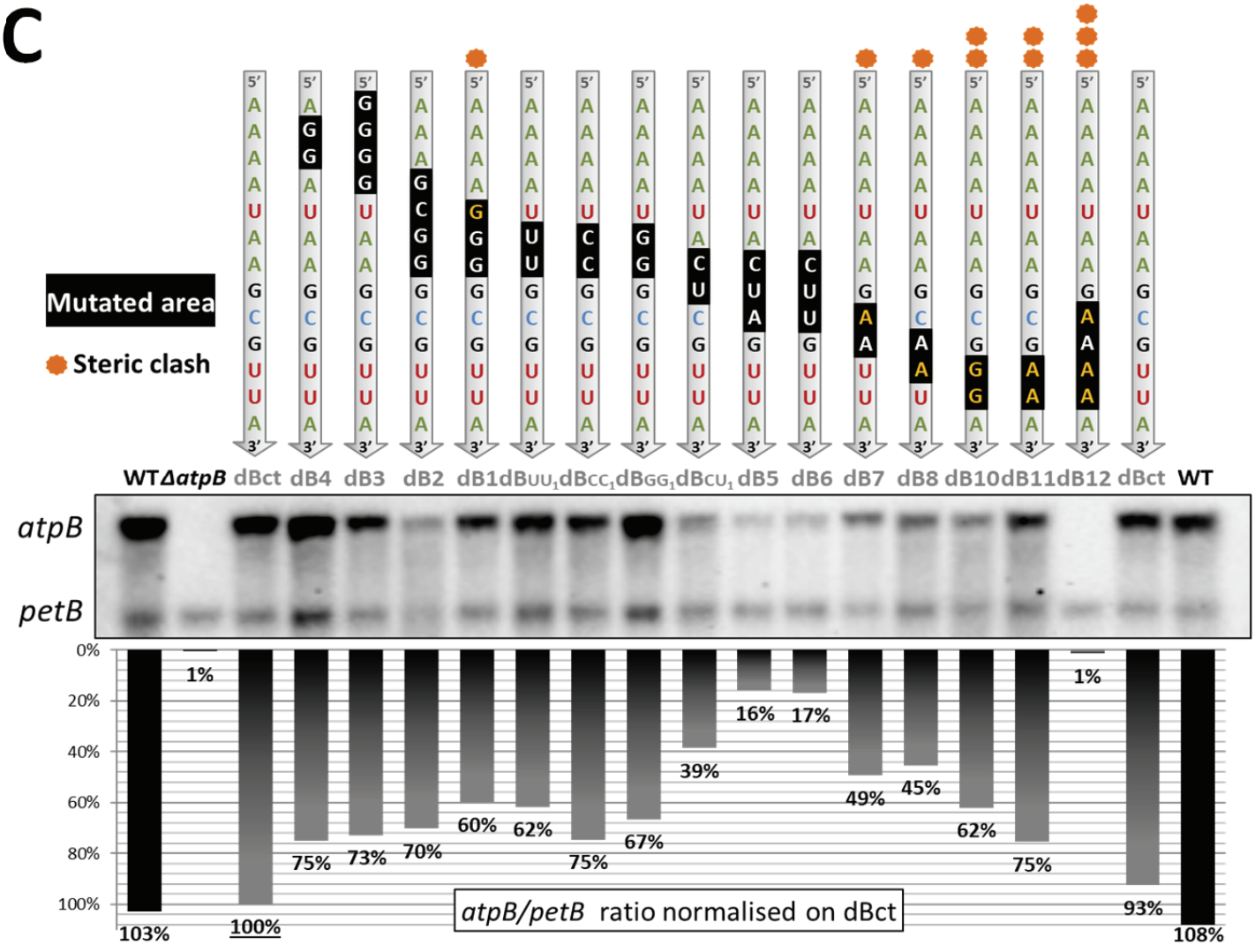


Figure 1: A. Cartoon of the *5'atpB$_M$:atpB:3'atpB* construct, transformed into the chloroplast of the *ΔatpB1* strain. **B. Growth phenotypes of the mutants, a table aside shows the placement of the strains.** Droplets of liquid culture of the strains were put on TAP and minimum media and grown for 12 days under 55 µE.m^{-2}.s^{-1} illumination. **C. RNA blot of *ΔatpB* strains transformed with the mutated *atpB* MDB1 target sequence.** Corresponding mutations are depicted on top of each blot lane. Mutated nucleotides are depicted in black squares. An orange dot denotes the introduction of a steric clash in the mutant target. *atpB* and *petA* (loading control) mRNA quantifications were performed with the image lab software and normalised on dBCt levels. The ratio of *atpB* on *petA* transcripts is depicted under the blot.

$5'atpB_{M}$:atpB:3'atpB | $5'atpB_{shM}$:uidA:3'psaB | $5'atpB_{shM}$:gfp:3'rbcL

Target sequence	
AAAAUAA GCGUUA	100%
AGGAUAA GCGUUA	75%
GGGGUAA GCGUUA	73%
AAAGCGG GCGUUA	70%
AAAAGGG GCGUUA	60%
AAAAUUU GCGUUA	62%
AAAAUCC GCGUUA	75%
AAAAUGG GCGUUA	67%
AAAAUAC UCGUUA	39%
AAAAUAC UAGUUA	16%
AAAAUAC UUGUUA	17%
AAAAUAA GAAUUA	49%
AAAAUAA GCAAUA	45%
AAAAUAA GCGGGA	62%
AAAAUAA GCGAAA	75%
AAAAUAA GAAAAA	1%

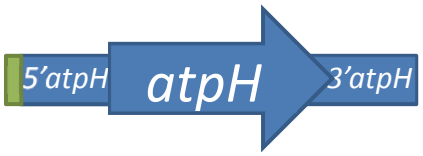
Target sequence	
AAAAUAA GCGUUA	100%
AAAGUAA GCGUUA	23%
AAAACAA GCGUUA	18%
AAAAUCA GCGUUA	41%
AAAAUAC GCGUUA	15%
AAAAUAA ACGUUA	325%
AAAAUAA GGGUUA	5%
AAAAUAA GCAUUA	1%
AAAAUAA GCGCUA	0%
AAAAUAA GCGUCA	9%
AAAAUAA GCGUUC	2%

Target sequence	
AAAAUAA GCGUUA	100%
AAAAUCC GCGUUA	29%
AAAAUUU GCGUUA	22%
AAAAUGG GCGUUA	22%
AAAAUAA GCGCCA	4%
AAAAUAA GCGAAA	11%
AAAAUAA GCGGGA	3%

From (Anthonisen *et al.*, 2001)

Figure 2 : Comparison of the accumulation levels of *atpB* transcripts bearing mutations in the MDB1 target and those of chimeras either from (Anthonisen *et al.*, 2001) or the present study with similar mutations. Cartoon depicting the corresponding mRNA used are above the target mutations.

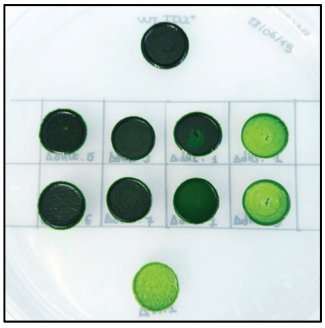
A



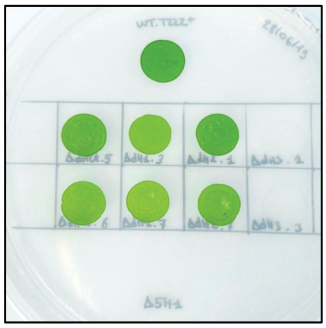
5'atpH$_M$:atpH:3'atpH

B

WT			
ΔdH$_{Ct}$2	ΔdH$_1$2	ΔdH$_2$1	ΔdH$_3$1
ΔdH$_{Ct}$3	ΔdH$_1$3	ΔdH$_2$3	ΔdH$_3$3
Δ*atpH*1			



TAP medium



min medium

C

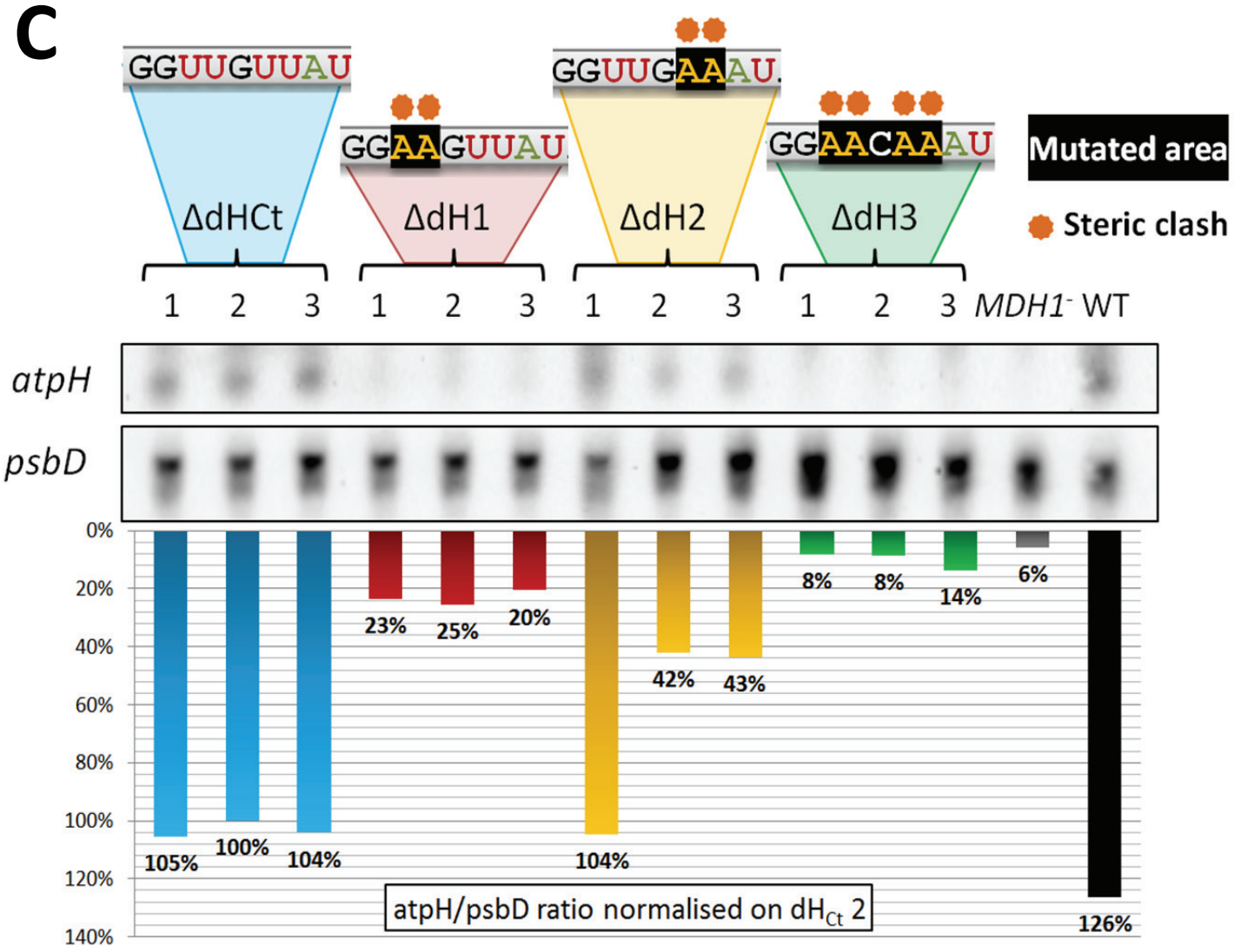


Figure 2: To be redone A. Cartoon of the *5'atpH$_M$:atpH:3'atpH* construct, transformed into the chloroplast of the *ΔatpH* strain. **B. Growth phenotypes of the mutants**, a table on the left shows the placement of the strains. Droplets of liquid culture of the strains were put on TAP and minimum media under 55µE illumination for 8 days. **C. RNA blot of *ΔatpH* strains transformed with the mutated *atpH* MDB1 target sequence.** Corresponding mutations are depicted on top of each blot lane. Mutated nucleotides are depicted in black squares. An orange dot denotes the introduction of a steric clash in the mutant target. *atpH* and *psbD* (loading control) mRNA quantifications were performed with the image lab software and normalised on ΔdHCt mean levels. The ratio of *atpH* on *psbD* transcripts is depicted under the blot.

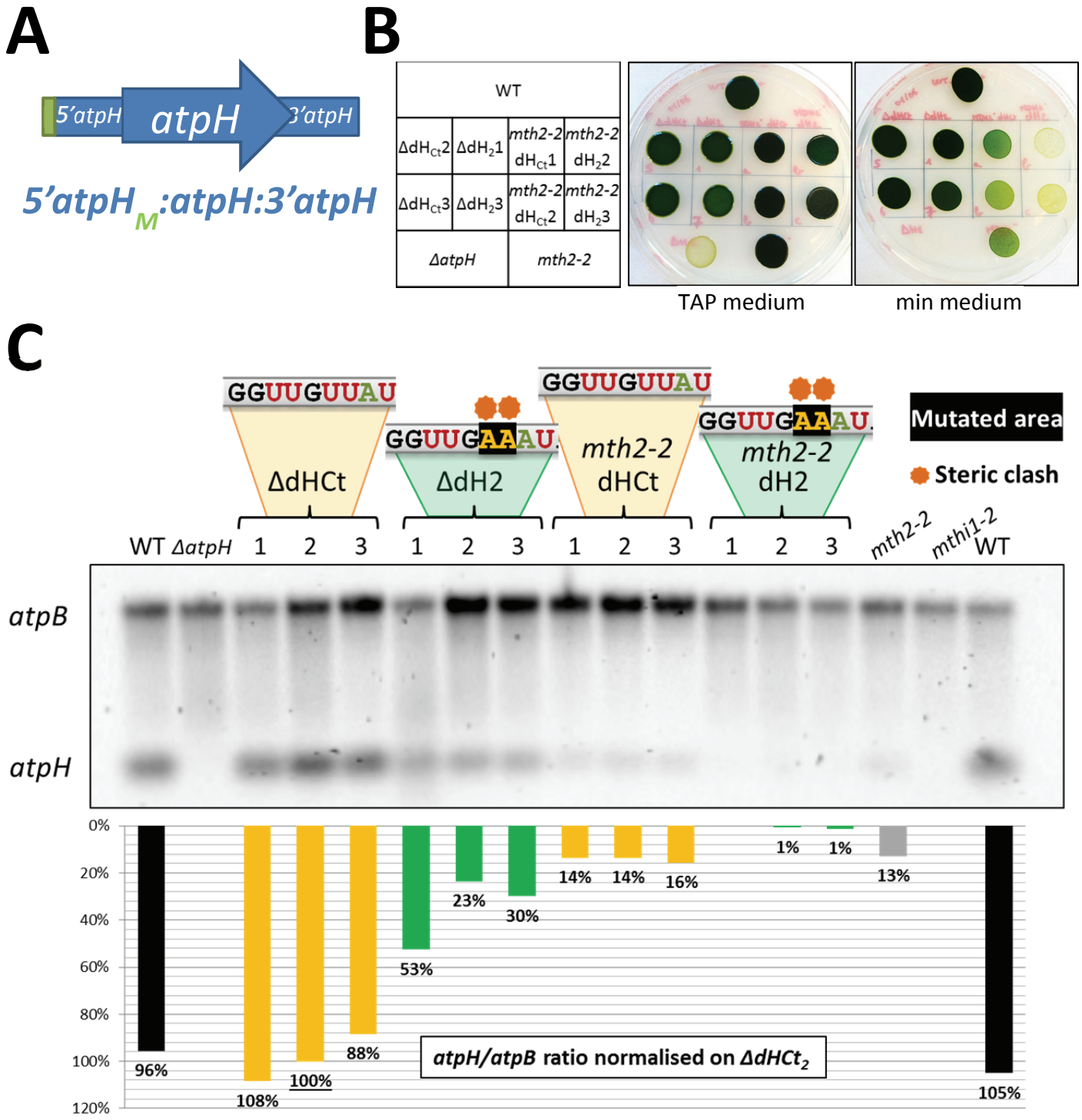


Figure 3: To be redone A. Cartoon of the *5'atpH$_M$:atpH:3'atpH* construct, transformed into the chloroplast of the *ΔatpH* and *mth2-2* strains. **B. Growth phenotypes of the mutants**, a table on the left shows the placement of the strains. Droplets of liquid culture of the strains were put on TAP and minimum media under 55 µE.m^{-2}.s^{-1} illumination. **C. RNA blot of *ΔatpH* and *mth2*$^-$ strains transformed with the mutated *atpH* MTHI1 target sequences.** Corresponding mutations are depicted on top of each blot lane. Mutated nucleotides are depicted in black squares. An orange dot denotes the introduction of a steric clash in the mutant target. *atpH* and *atpB* (loading control) mRNA quantifications were performed with the image lab software and normalised on ΔdHCt2 levels. The ratio of *atpH* on *atpB* transcripts is depicted under the blot.

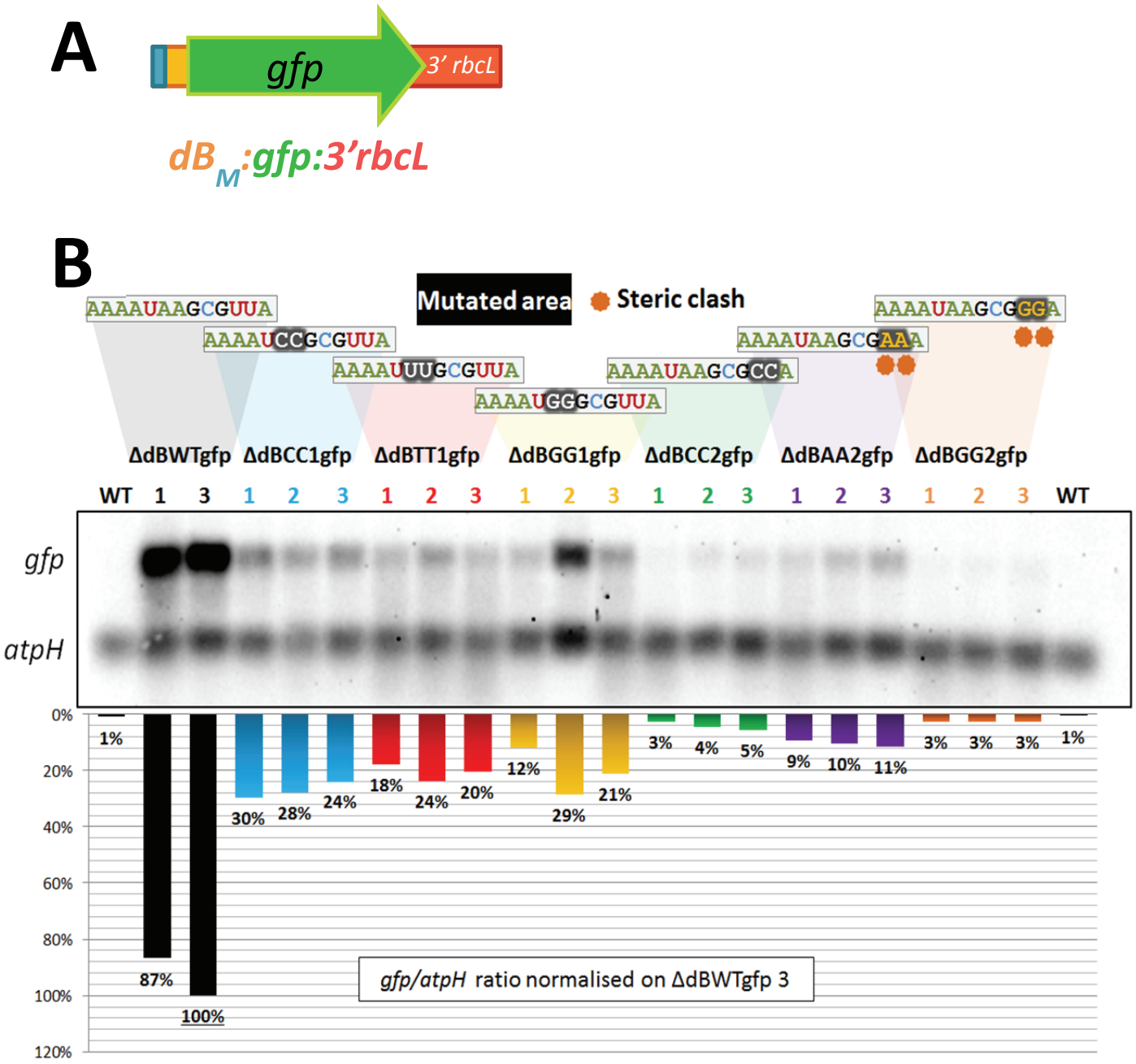


Figure 4: A. Cartoon of the chimeric construct *dB$_M$:gfp:3'rbcL*, transformed into the chloroplast of *ΔatpB* cells. **B. RNA blot of *ΔatpB* transformants chimeric constructs bearing *atpB* MDB1 target variants.** Filter was hybridised with gfp and atpH dig-dUTP labelled probes. Transcript quantifications were done with ImageLab, and normalised on *ΔatpB*::*dB$_{WT}$*gfp 3 levels. Ratio of *gfp/atpH* transcripts is depicted under the corresponding lanes, the mutations are on top. Two technical repeats were made and give the same results.

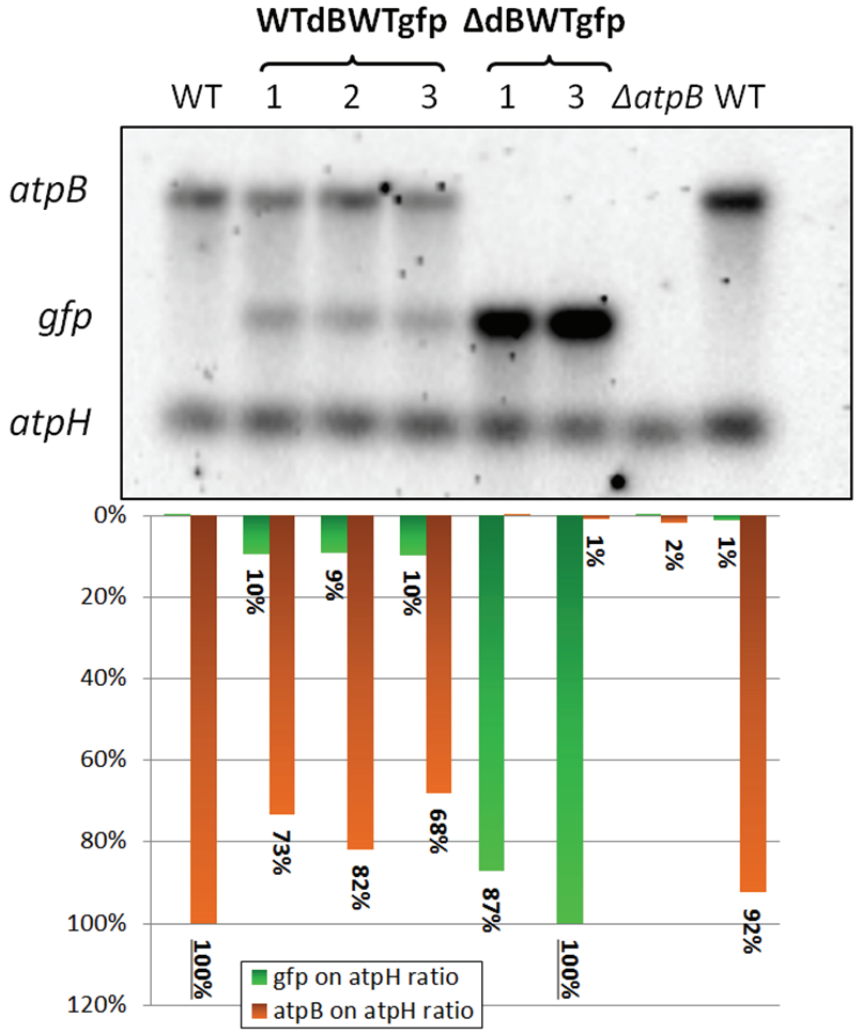


Figure 5: RNA blot of chimeric constructs transformed either in WT.T222+ or *ΔatpB* strains. *gfp*, *atpB* and *atpH* (loading control) mRNA quantifications were performed with the image lab software, and normalised on either ΔdBWTgfp3 for *gfp* or WT levels for *atpB*. Reference levels are underlined. The ratios of *gfp* and *atpB* on *atpH* transcripts are depicted under the blot.

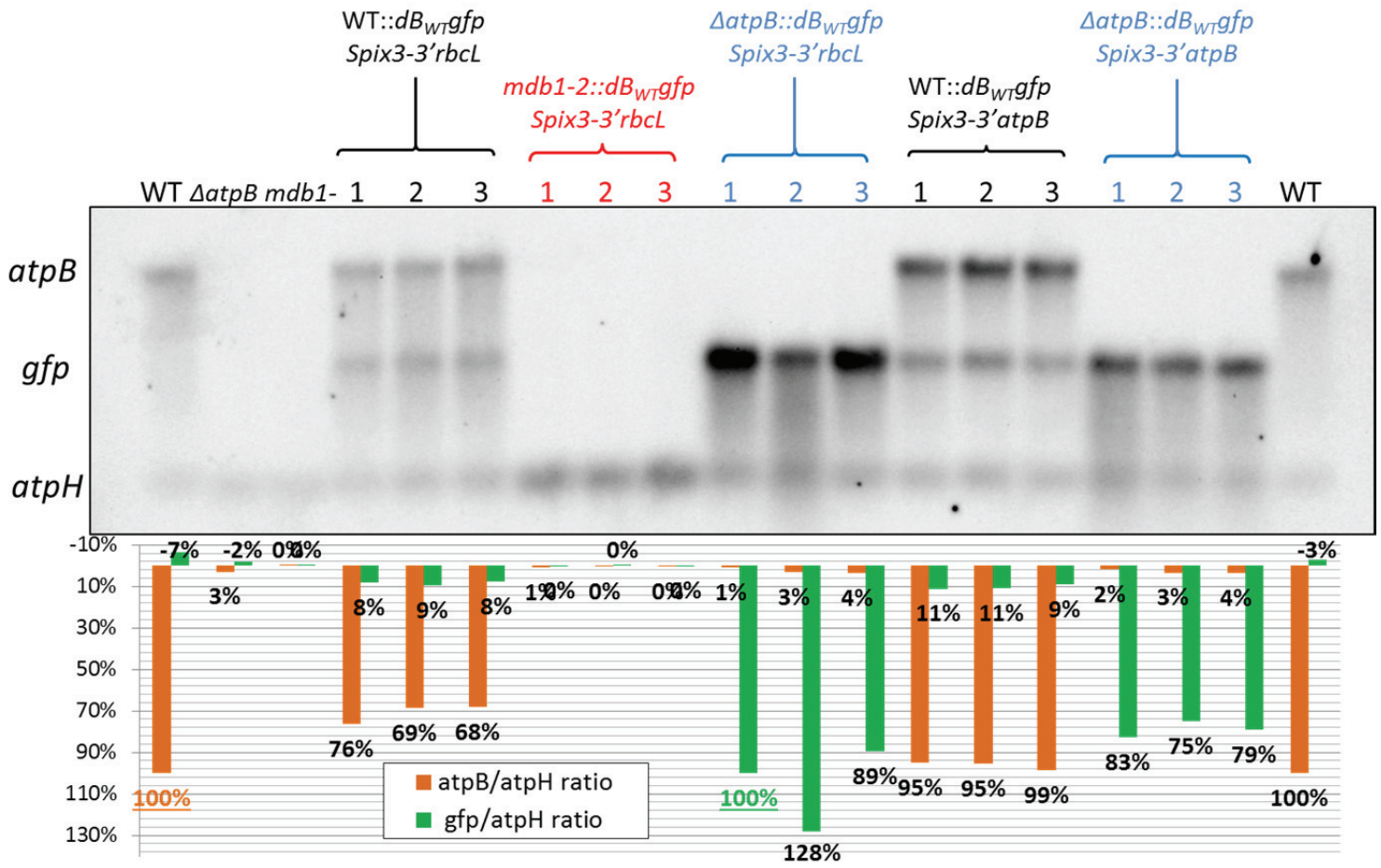


Figure 6: RNA blot of wild-type target sequence chimeric constructs transformed either in WT.T222+ or *ΔatpB* or *mdb1-2* strains. *gfp*, *atpB* and *atpH* (loading control) transcripts levels are probed. mRNA quantifications were performed with the image lab software, and normalised on either *ΔatpB dBWTgfp Spix3-3'rbcL* 1 for *gfp* or WT levels for *atpB*. Reference levels are underlined. The ratios of *gfp* and *atpB* on *atpH* transcripts are depicted under the blot.

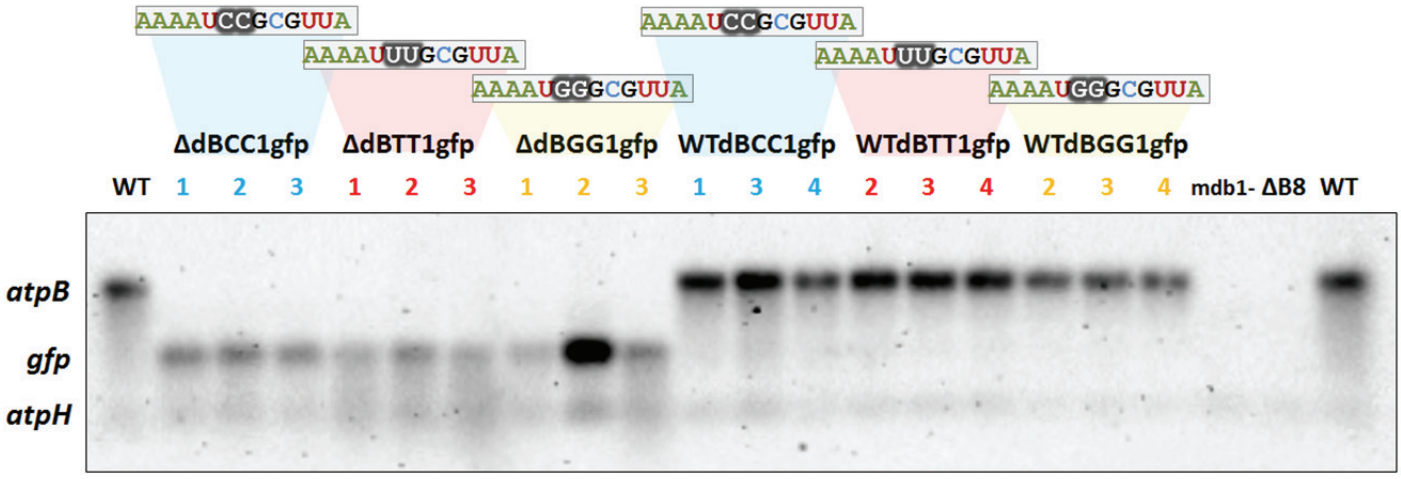


Figure 7: RNA blot of chimeric constructs transformed either in WT.T222+ or *ΔatpB* strains. Filter was hybridised with *atpB*, *gfp* and *atpH* dig-dUTP labelled probes. Four technical repeats displayed the same patterns.

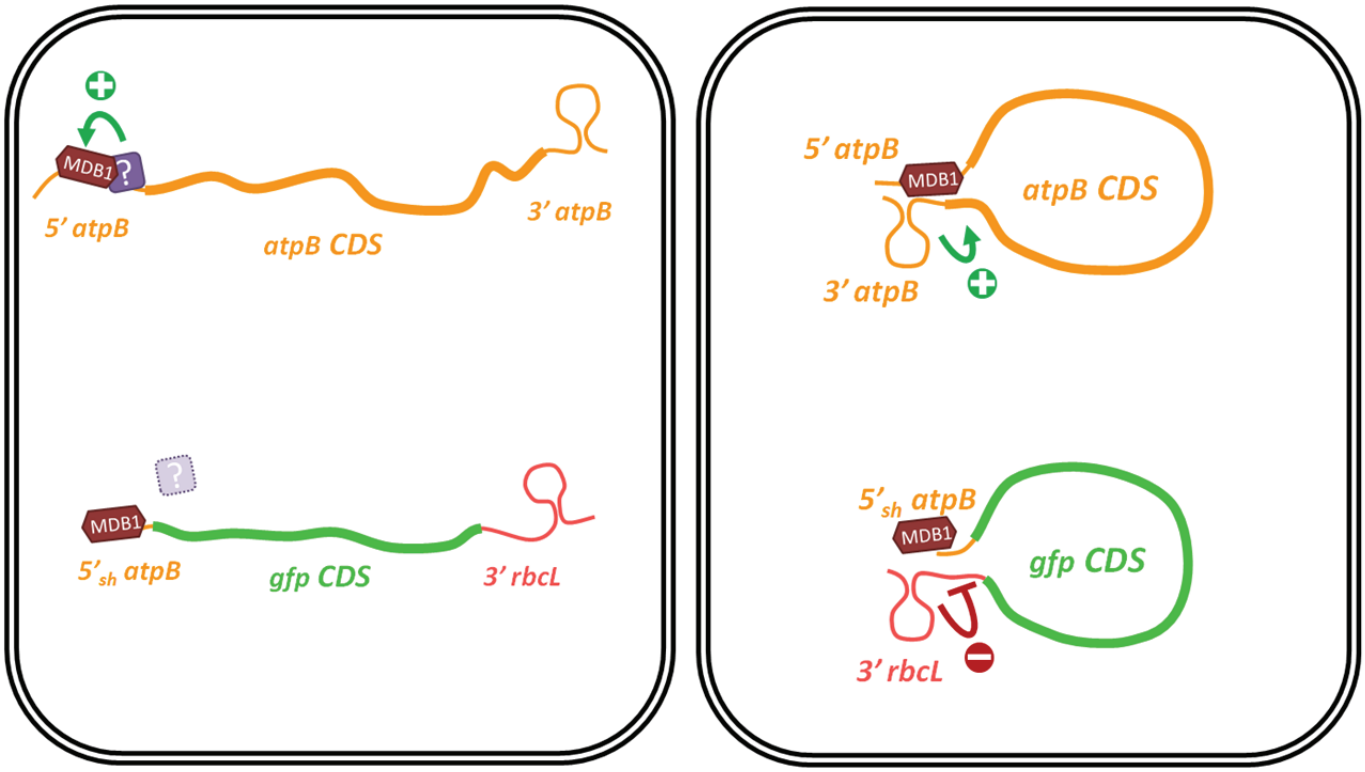


Figure 8: Two main hypotheses to explain how the mutated transcripts could be less destabilised in the presence of the whole atpB sequence. Auxiliary factors might limit *atpB* transcript destabilisation by maintaining MDB1 on its target sequence or other regions of the atpB transcript such its 3'end could be implicated in the stabilisation process.

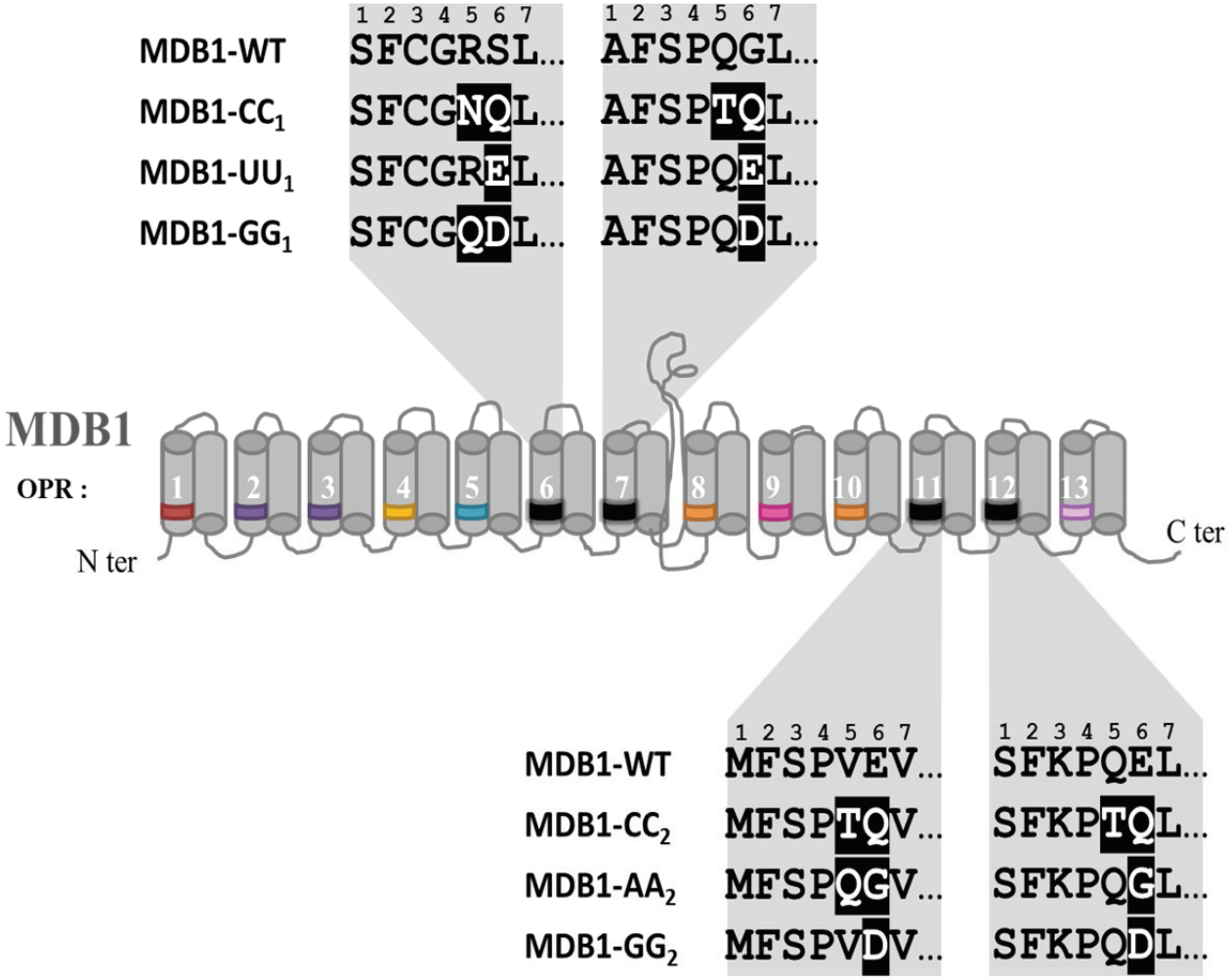


Figure 9: Modification inserted in MDB1, 6th and 7th or 11th and 12th OPR motif, they were chosen following the draft OPR code, for the CC1 and CC2 variant, one of the "unreadable" residue combination was picked at random.

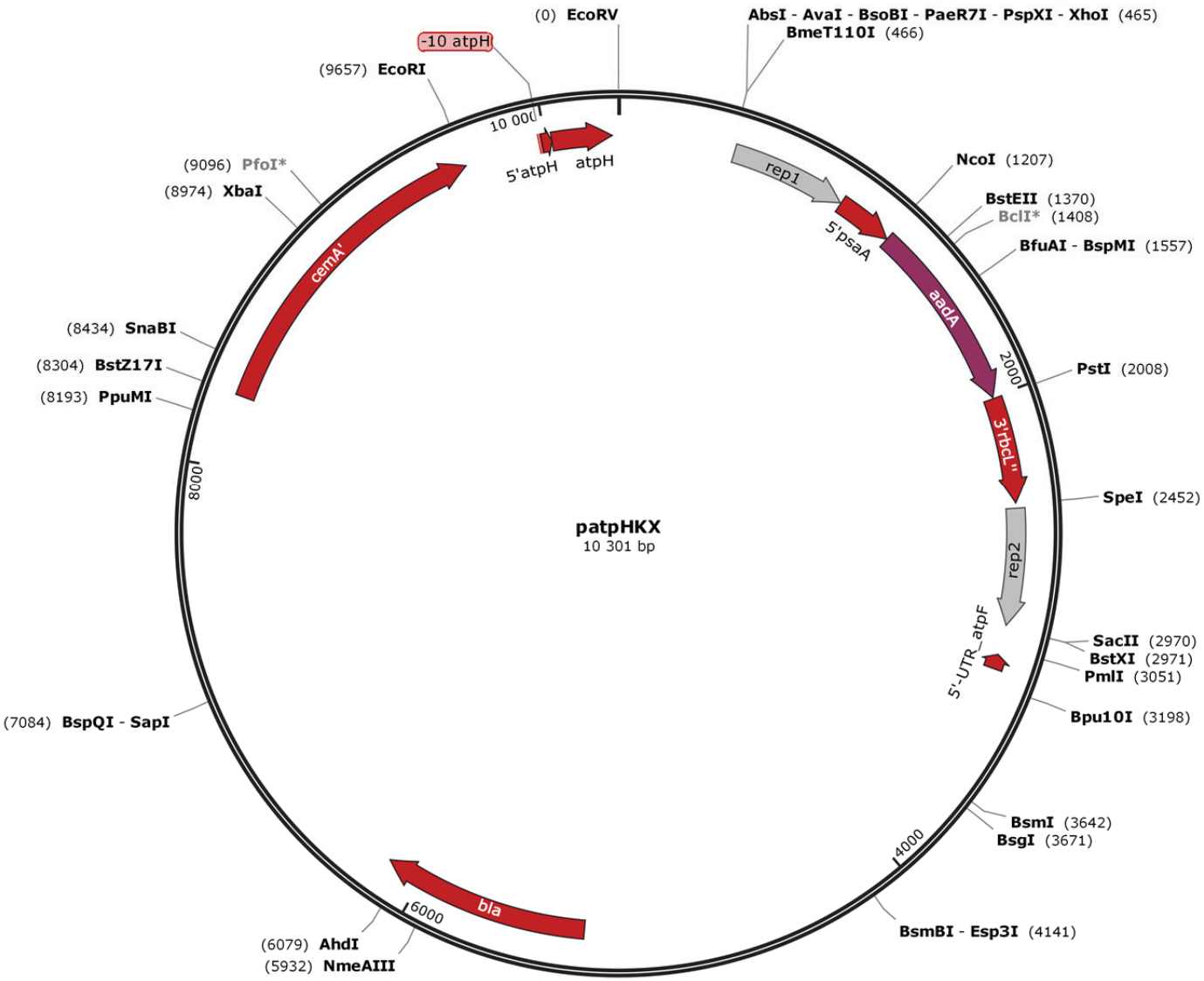
Application	Name	Sequence (5'->3')
atpH MDH1 target mutagenesis	cemAFW	GCGAATTCCGGAAAGTCAAACAGGTATTTTCTT
	atpHext-RV	GCGTTAGCCAATACCAAACAGC
	Mut-atpH-1FW	ATTCTTTGGAAGTTATCGATTTTATTGATTCATTTAG
	Mut-atpH-1RV	TCGATAACTTCCAAAGAATATTATATTCTT
	Mut-atpH-2FW	CTTTGGTTGAAATCGATTTTATTGATTCATTTAG
	Mut-atpH-2RV	AAAATCGATTTCAACCAAAGAATATTATATTCTT
	Mut-atpH-3FW	ATTCTTTGGAACAAATCGATTTTATTGATTCATTTAG
	Mut-atpH-3RV	AAAATCGATTTGTTCCAAAGAATATTATATTCTT
atpB MDB1 target mutagenesis	atpB5'FW1	GCGCTCGAGCTTAAGTTCAAAATTCTCCACCAGCT
	atpB5'FW2	GCGCTCGAGGCTAGCTTCAAAATTCTCCACCAGCT
	atpB5'FW3	GCGCTCGAGAGATCTTTCAAAATTCTCCACCAGCT
	atpB5'FW4	GCGCTCGAGAGGCCTTTCAAAATTCTCCACCAGCT
	atpB5'FW5	GCGCTCGAGGGGCCCTTCAAAATTCTCCACCAGCT
	atpB5'FW6	GCGCTCGAGCCTAGGTTCAAAATTCTCCACCAGCT
	atpB5'FW7	GCGCTCGAGACGCGTTTCAAAATTCTCCACCAGCT
	dB1FW	ACTAAAAAAGGGGCGTTAGTGAATAATACTTTTT
	dB1RV	TCACTAACGCCCCTTTTTTAGTTTTTTCATTTAACT
	dB2FW	ACTAAAAAGCGGGCGTTAGTGAATAATACTTTTT
	dB2RV	TCACTAACGCCCGCTTTTTAGTTTTTTCATTTAACTAT
	dB3FW	ACTAAGGGGTAAGCGTTAGTGAATAATACTTTTT
	dB3RV	TCACTAACGCTTACCCCTTAGTTTTTTCATTTAACTATAT
	dB4FW	ACTAAAGGATAAGCGTTAGTGAATAATACTTTTT
	dB4RV	TCACTAACGCTTATCCTTTAGTTTTTTCATTTAACTATA
	dB5FW	ACTAAAAAATACTAGTTAGTGAATAATACTTTTT
	dB5RV	TCACTAACTAGTATTTTTTAGTTTTTTCATTTAACT
	dB6FW	ACTAAAAAATACTTGTTAGTGAATAATACTTTTT
	dB6RV	TCACTAACAAGTATTTTTTAGTTTTTTCATTTAACT
	dB7FW	ACTAAAAAATAAGAATTAGTGAATAATACTTTTTata
	dB7RV	TCACTAATTCTTATTTTTTAGTTTTTTCATTTAACT
	dB8FW	ACTAAAAAATAAGCAATAGTGAATAATACTTTTTata
	dB8RV	TCACTATTGCTTATTTTTTAGTTTTTTCATTTAACT
	dB10FW	ACTAAAAAATAAGAAAAAGTGAATAATACTTTTTATATA
	dB10RV	TCACTTTTTCTTATTTTTTAGTTTTTTCATTTAACT
	dB11FW	ACTAAAAAATAAGCGGGAGTGAATAATACTTTTTATATA
	dB11RV	TCACTCCCGCTTATTTTTTAGTTTTTTCATTTAACT
	dB12FW	ACTAAAAAATAAGCGAAAGTGAATAATACTTTTTATATA
	dB12RV	TCACTTTCGCTTATTTTTTAGTTTTTTCATTTAACT
	dBExt_RV	TTTGAAATAAGAACCTCCTCCTTCC
	atpB_Anton_FW	cgcCTCGAGAAGATGCTTTGCATCTCTAA
	atpB_Anton_WT_RV	gcgCCCGGGCCCATATAAAAAGTATTATTCACTAAC
	atpB_Anton_M1_RV	gcgCCCGGGAATTCAATATAAAAAGTATTATTCACTGGCGCTTA TTTTTTAGTTTTTTCAT
	atpB_Anton_M2_RV	gcgCCCGGGATCCATATATAAAAAGTATTATTCACTCCCGCTTA TTTTTTAGTTTTTTCAT
	atpB_Anton_M3_RV	gcgCCCGGGAGATCTATATAAAAAGTATTATTCACTTTCGCTTA TTTTTTAGTTTTTTCAT
MDB1 target verification	atpB_TT_RV	AAAGTATTATTCACTAACGCAA
	atpB_GG_RV	AAAGTATTATTCACTAACGCCC
	atpB_CC_RV	AAAGTATTATTCACTAACGCGG

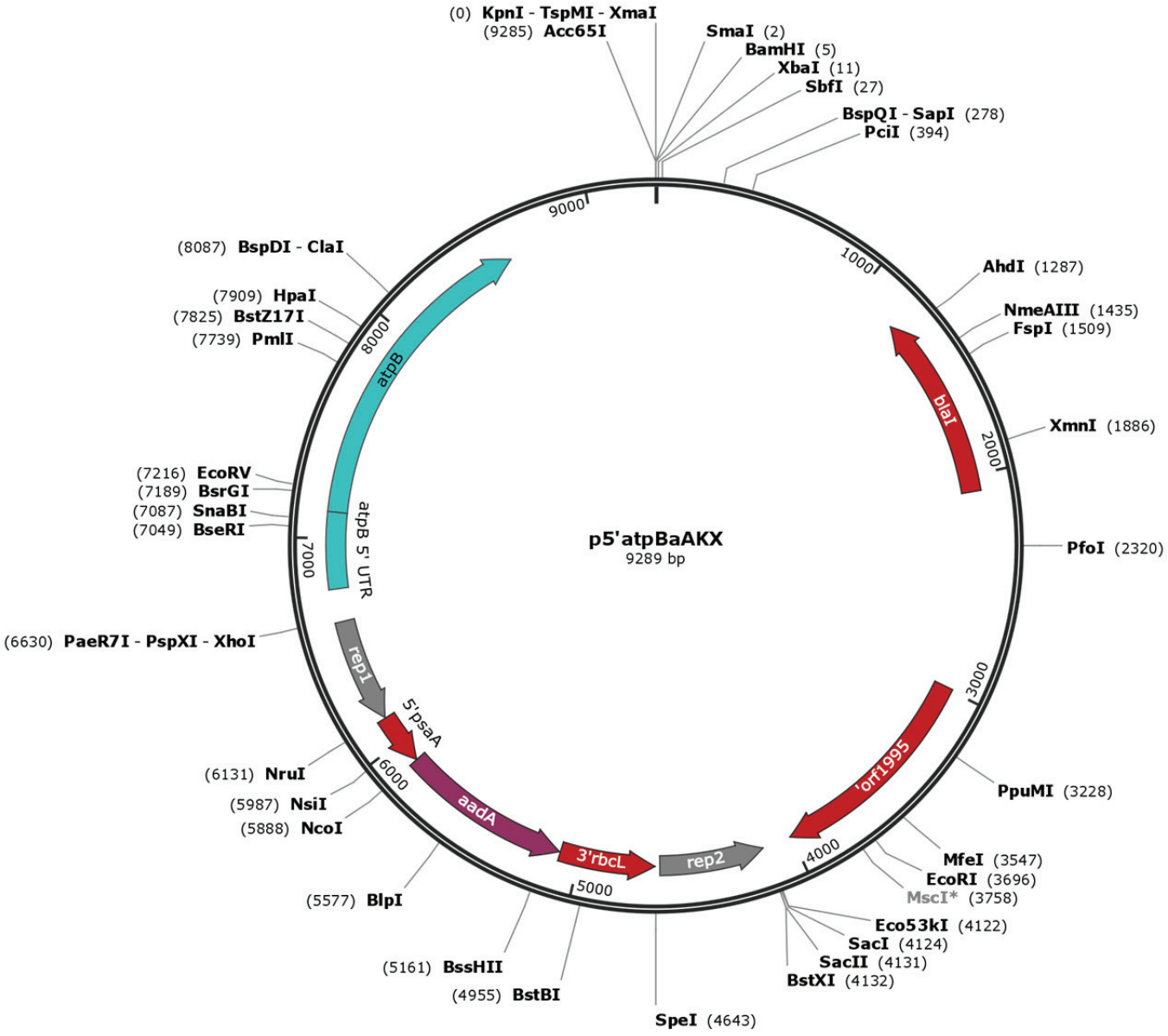
	atpB_CT_RV	AAAGTATTATTCACTAACGAG
	atpB_FW	ACCTCGAGTTCAAAATTCTC
	atpB_WT_FW	AGTTAAATGAAAAAACTAAAAAATAA
	atpB_RV	ATTCTTACGTATAAACCCCG
	atpBCDS_RV	TGCTGAGTTTTTAGCACGAATA
	atpBSeqRV	AAATCCACCGTTTTGTGGAA
	atpBSeqFW	GGAGACCTTCAAGCCGTACA
gfp chimeric construct	GFP-CDS-FW	CGCGAATTCGCGCTCGAGGCGCCCCGGGCCATGGGTAAAGG AGAAGAACTTTTCACTGG
	GFP-CDS-RV2	GCGCTGCAGTTATTTGTATAGTTCATCCATGCCATGTG
thm24.2^{-} phenotyping	MDB1 bFW	CTGCACTCAGGTCTTAGTCTGGC
	MDB1 aRV	CGCATCTCTTCTTTCCACGACTC
Labelling of digoxigenin PCR probes	atpH-dig-FW	AACCCTATCGTAGCTGCTGC
	atpH-dig-RV	ACTAGACCGTAAATTGTTAA
	atpB-dig-FW	CACGGTGGTGTTTCTGTATT
	atpB-dig-RV	TTACGCTTTGTGCAGAATCA
	gfp-dig-FW	TTGAATTAGATGGTGATGTT
	gfp-dig-RV	CAATTGGAGTATTTTGTTGA
	petB-dig-FW	GCTGTTATTTTAGGTATGGC
	petB-dig-RV	GATGCGTTGTAAATAGTGTT
	psbD FW	GAGCTAAACCTACAACACCA
	psbD RV	CAGTATGGCTCACTCTCTTC

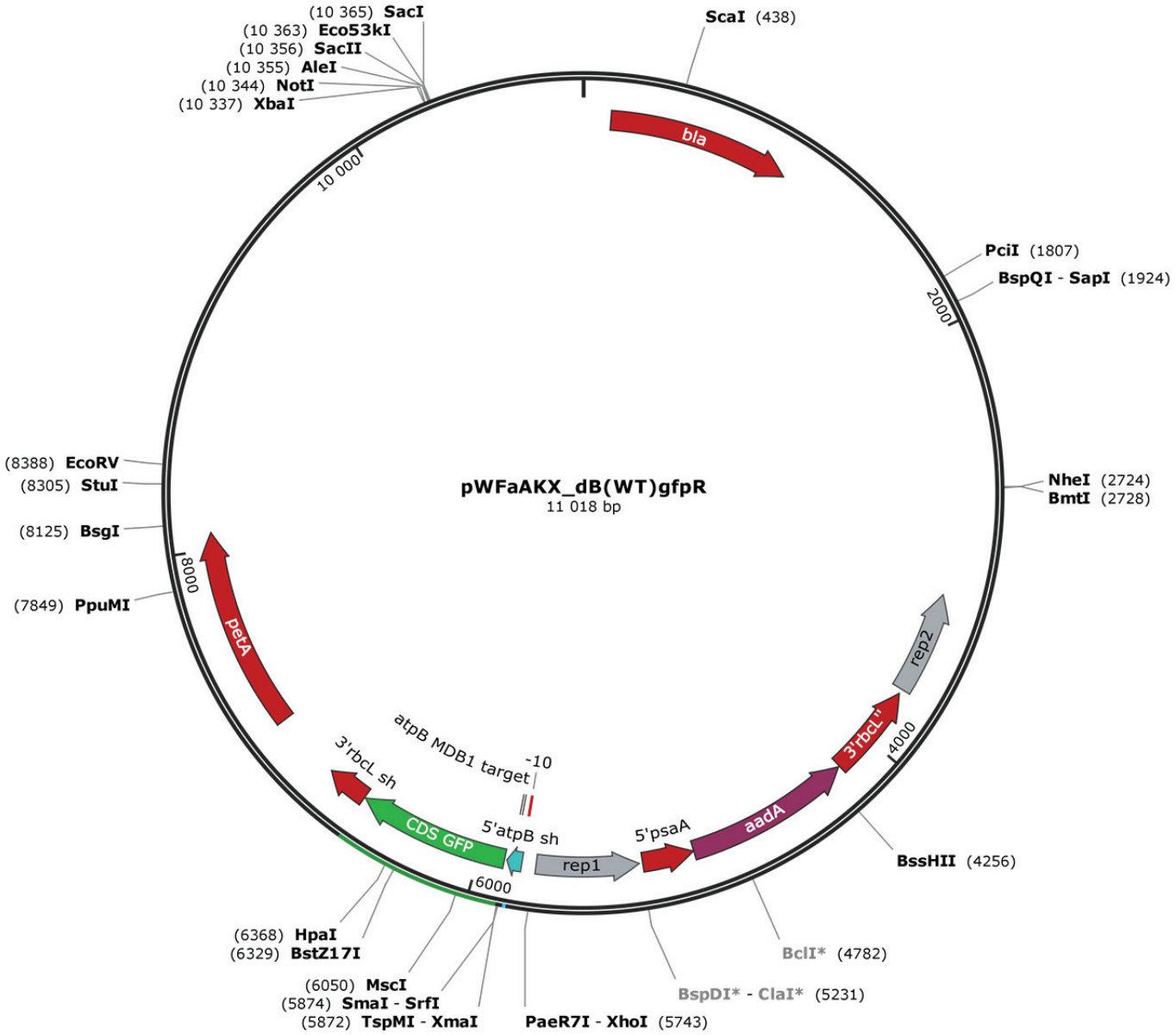
Table 2: Primers used in this study

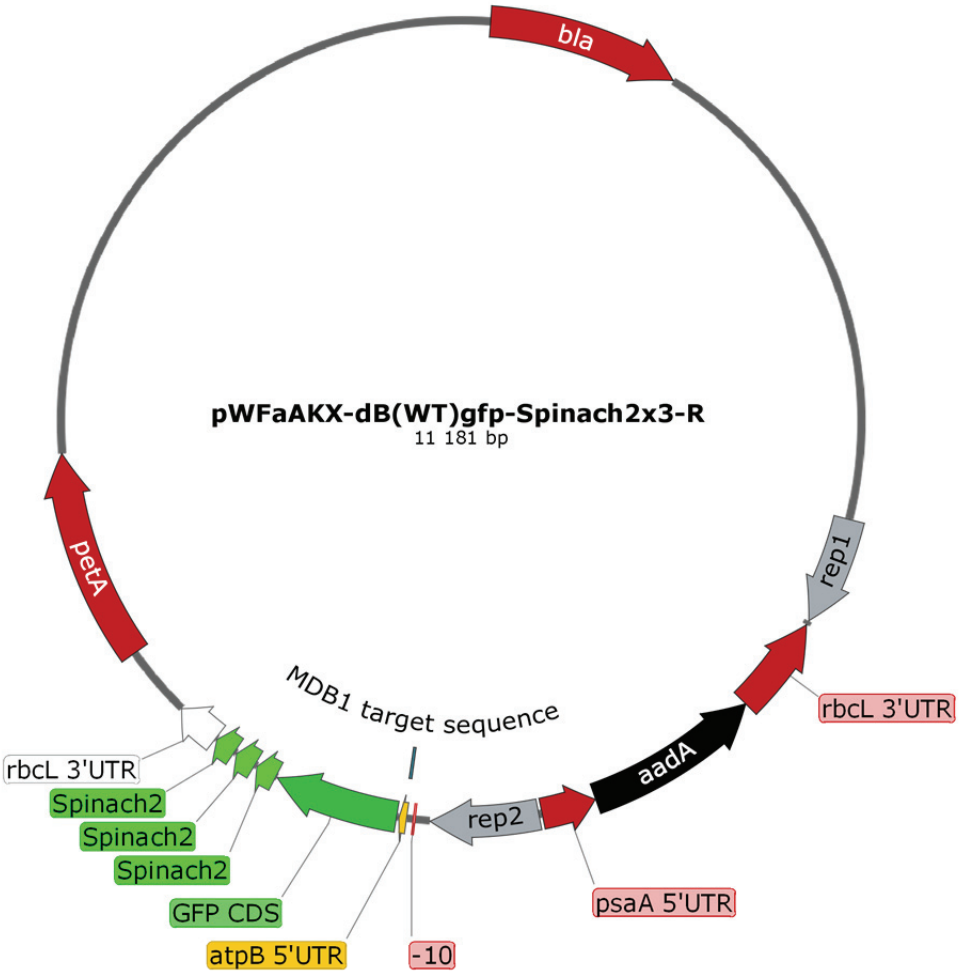
Strain	Mating type	Genotype
WT.T222	+	Wild type
WT.S24	-	Wild type
ΔatpB	+	Deletion of *atpB* and its 5'UTR
ΔatpH	+	Deletion of *atpH* and its 5'UTR
mthi1-2 (II 174)	+	*MTHI1* mutant
mth2-2	+	*MTH2* insertional mutant from Clip library (Li *et al.*, 2019)
mdb1-1 (thm24.2)	-	Deletion of one A in *MDB1* causing a frameshift
mdb1-2 (L35a)	+	Deletion of *MDB1* and 6 other genes(Houille-Vernes *et al.*, 2011)

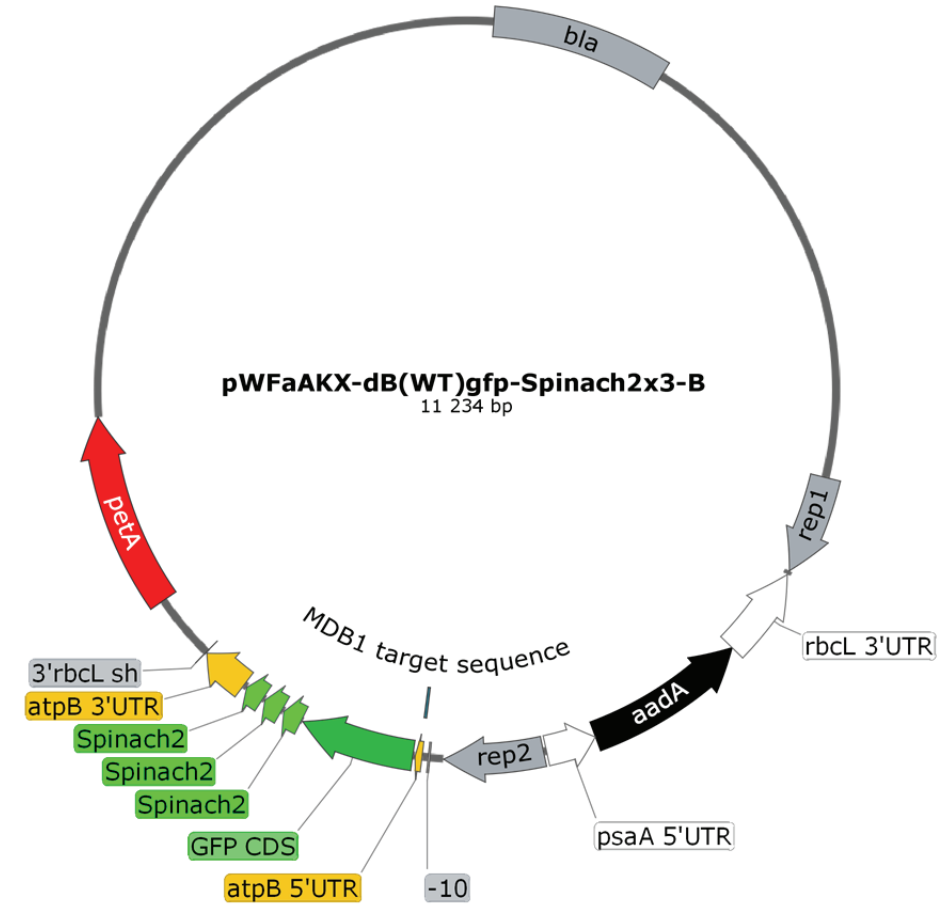
Table 3: Strains used in this study.




p5'atpBaAKX
9289 bp
(0) KpnI - TspMI - XmaI
(9285) Acc65I
SmaI (2)
BamHI (5)
XbaI (11)
SbfI (27)
BspQI - SapI (278)
PciI (394)
AhdI (1287)
NmeAIII (1435)
FspI (1509)
XmnI (1886)
PfoI (2320)
PpuMI (3228)
MfeI (3547)
EcoRI (3696)
MscI* (3758)
Eco53kI (4122)
SacI (4124)
SacII (4131)
BstXI (4132)
SpeI (4643)
BstBI (4955)
BssHII (5161)
BlpI (5577)
NcoI (5888)
NsiI (5987)
NruI (6131)
PaeR7I - PspXI - XhoI (6630)
EcoRV (7216)
BsrGI (7189)
SnaBI (7087)
BseRI (7049)
HpaI (7909)
BstZ17I (7825)
PmlI (7739)
BspDI - ClaI (8087)
atpB
atpB 5' UTR
rep1
5'psaA
aadA
3'rbcL
rep2
'orf1995
blaI
9000
1000
2000
3000
4000
5000
6000
7000
8000







Bibliographie

Anthonisen, I. L., Salvador, M. L. and Klein, U. (2001) Specific sequence elements in the 5' untranslated regions of rbcL and atpB gene mRNas stabilize transcripts in the chloroplast of Chlamydomonas reinhardtii. RNA (New York, N.Y.), 7, 1024-1033.

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T. and Nakamura, T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. PloS one, 8, e57286.

RÉSUMÉ

Depuis sa capture par un organisme eucaryote hétérotrophe ancestral et son évolution endosymbiotique de cyanobactérie à organelle, le chloroplaste a perdu de nombreux gènes (redondants ou transférés vers le génome nucléaire de la cellule hôte). Cependant, certains sont encore exprimés par le chloroplaste, notamment ceux impliqués dans la photosynthèse. Les sous-unités des complexes photosynthétiques sont ainsi codées en partie dans le noyau de la cellule et pour une autre partie dans le chloroplaste. Pour aboutir à la stœchiométrie nécessaire à l'assemblage et au bon fonctionnement de l'appareil photosynthétique, il est impératif que l'expression génétique des différents compartiments soit coordonnée. Pour ce faire, des facteurs spécifiques codés par le noyau contrôlent l'expression des gènes des organites : les OTAF (**O**rganellar **T**rans-**A**cting **F**actors). Chez l'algue unicellulaire *Chlamydomonas reinhardtii*, ceux-ci appartiennent à une grande famille de protéines à répétitions hélicoïdales ou protéines α-solénoïdes : les **protéines OPR** (**O**ctotrico**P**eptide **R**epeat). Ces protéines OPR sont constituées d'une succession de répétitions OPR dégénérées, chacune de ces répétitions de 38 résidus liant spécifiquement un nucléotide donné, grâce à certains acides aminés à des positions précises. La succession de répétitions permet ainsi la reconnaissance modulaire d'un ARNm spécifique. Le facteur OPR peut donc s'y fixer et permettre :

- Sa stabilisation, ou maturation (**facteur M**), le stroma du chloroplaste est riche en exonucléases, le facteur M bloque leur accès au messager. L'absence d'un facteur M provoque la perte totale de son ARNm cible.
- Sa traduction (**facteur T**), par un mécanisme qui reste mal connu.

Lors de ma thèse je me suis intéressée au rôle physiologique de deux facteurs OPR : **MTHI1** et **MDB1**. À l'aide de mutagenèse dirigée dans le chloroplaste de *C. reinhardtii* j'ai pu étudier combien ces facteurs sont essentiels à l'expression de gènes du chloroplaste. J'ai caractérisé un nouveau mutant de *MDB1* et confirmé le rôle crucial de cette protéine pour la stabilisation de l'ARNm chloroplastique ***atpB***, codant la sous-unité β de l'ATP synthase chloroplastique. J'ai aussi montré que la protéine MTHI1 était non seulement impliquée dans la stabilisation, mais également dans la traduction de l'ARNm d'***atpH***, et qu'elle jouait également un rôle actif dans la stabilisation et la traduction d'un autre transcrit ***atpI***, ces deux gènes codant les deux hémi-canaux à protons de l'ATP synthase. Plus largement, mes travaux s'inscrivent dans une étude approfondie du facteur MTHI1, qui suggère l'existence de facteurs auxiliaires spécifiques interagissant avec MTHI1 et renforçant son action sur *atpH* ou *atpI*. Globalement ces observations nous ont conduit à considérer un modèle peut être universel d'interaction d'OTAF en un complexe multimoléculaire assemblé autour de l'ARNm cible, chaque facteur renforçant la stabilité et l'affinité de l'interaction.

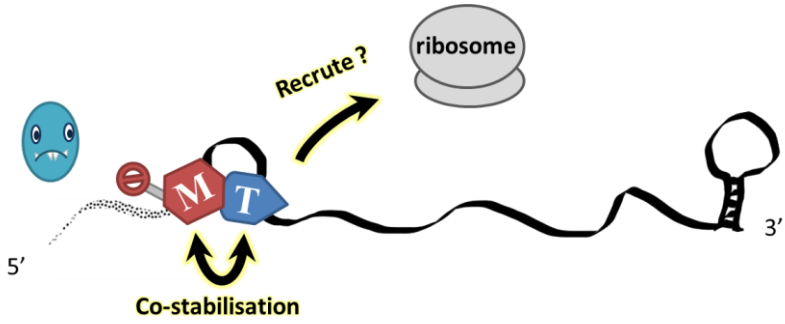


Figure 103: Modèle de l'interaction tripartite entre facteurs OTAF M et T sur leur ARNm cible, pourrait s'appliquer à MTHI1 et MDB1.

L'axe principal de ma thèse concernait l'étude des bases moléculaires des interactions protéine OPR/ARN, et notamment l'étude du « code OPR », qui permet de déduire à partir de la séquence d'une répétition OPR le nucléotide qu'elle reconnaitra préférentiellement. Une version préliminaire de ce code a été établie au laboratoire, par l'observation de la corrélation entre certains résidus à des positions précises du motif OPR et le nucléotide reconnu dans des cibles bien caractérisées moléculairement. Le résidu en position six semble être crucial pour l'interaction avec le nucléotide, il est le plus souvent polaire, permettant probablement la formation de liaisons hydrogènes avec la base, ou de petite taille, tolérant la présence de bases pyrimidiques volumineuses (A ou G). Mon objectif était de tester ce code théorique *in vivo*.

Position	Résidu										
3	X	X	X	X	R, K	X	X	X	X	X	X
4	X	P	X	X	X - P	X	X	X	X	X	X
5	X	Q	X - R, K	R, K	X	R	X - R	R, Q	X	R	R
6	E	G	D	D	D	Q	Q	A	H	S	N
Nucléotide reconnu	U	A	G	U	U	U	?	A	?	A	?

Table 8: « Code OPR » préliminaire : le nucléotide reconnu est indiqué en bas; le résidu en position 6 de la répétition OPR, capital pour la spécificité de reconnaissance, en grands caractères. X dénote un acide aminé quelconque, – des exclusions.

Les facteurs de stabilisation se fixent de façon stable sur leur ARN cible et génèrent une empreinte ARN, un petit ARN protégé par son interaction avec le facteur M de la dégradation par des exonucléases). Les facteurs M ne génèrent qu'une seule empreinte. De plus, leur délétion n'entraine des effets que sur un ARNm chloroplastique unique. Ces facteurs étant ainsi de toute évidence spécifiques et se liant solidement avec leur ARNm cible ont été choisis pour tester le code *in vivo*. La stratégie consistait à réaliser des mutations dans la séquence cible d'un ARNm pour perturber la fixation de son facteur M OPR et donc déstabiliser l'ARNm. Puis, de tenter de restaurer l'accumulation du messager muté en transformant dans les cellules une version de la protéine OPR mutée de manière complémentaire (en suivant le « code OPR » préliminaire (Table 8), modifiant spécifiquement les résidus 5 et 6 des répétitions OPR correspondant à la zone mutée dans la séquence ARN cible. Ainsi le code OPR préliminaire pourrait être confirmé ou infirmé pour différentes combinaisons.

Ces travaux, réalisés sur les transcrits ***atpB*** et ***atpH***, et leur facteur de stabilisation spécifique, **MDB1** et **MTHI1**, ont révélé une importante résilience de l'interaction des facteurs M et de leur transcrits chloroplastiques attitrés. Malgré les mutations dans les séquences cibles sur lesquelles se fixent ces facteurs M, peu d'effets sur la stabilisation du messager étaient observés (Figure 104 et Figure 105). Ce qui se traduit par la capacité de ces facteurs M à se lier à leur ARNm cible malgré des changement importants dans leur cible. Ces résultats sont d'autant plus surprenants que les facteurs M sont très spécifiques de leur cible.

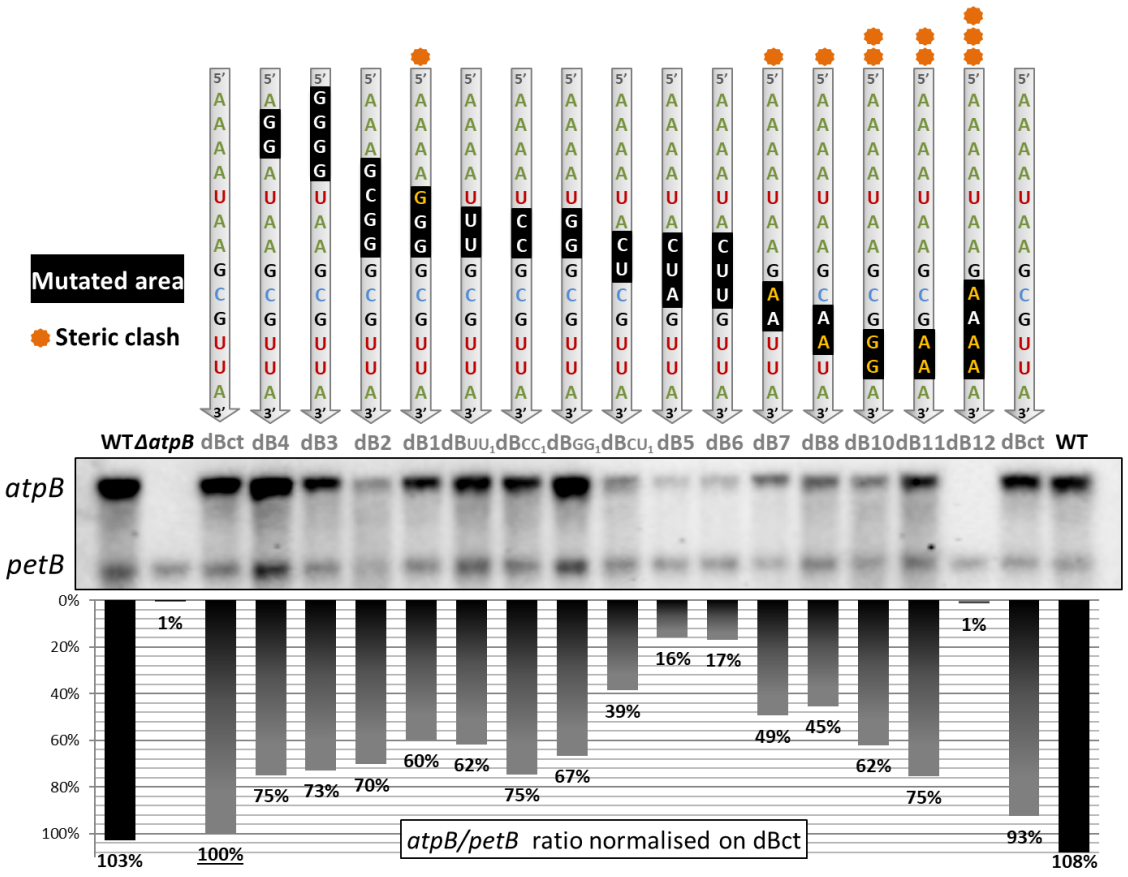


Figure 104: Gel d'ARN. Des mutations (surlignées en noir) réalisées dans la cible de MDB1 sur le 5' d'*atpB* ne conduisent qu'à une faible diminution de l'accumulation et donc de la stabilité du transcrit *atpB*. MDB1 parvient à s'y lier dans tous les cas, sauf pour la mutation dB12, très étendue. *petB* sert de contrôle de charge.

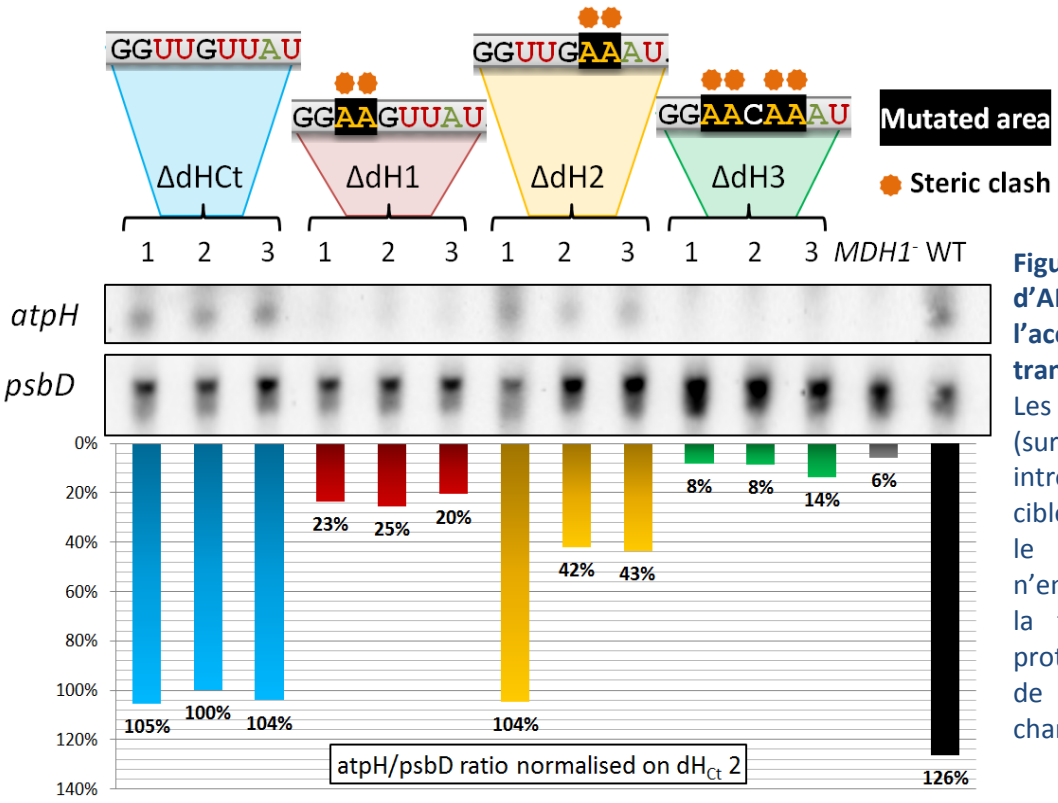
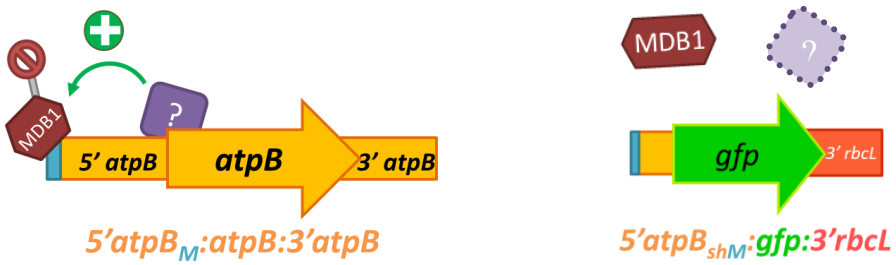


Figure 105: Gel d'ARN présentant l'accumulation du transcrit d'*atpH*. Les mutations (surlignées en noir) introduite dans la cible de MTHI1 sur le 5' d'*atpH*, n'empêchent pas la fixation de la protéine. *psbD* sert de contrôle de charge.

Seules des mutations très étendues parviennent à empêcher l'interaction facteur M/ARNm. Pour expliquer cette contradiction apparente et à la lumière de travaux antérieurs (Anthonisen *et al.*, 2001), nous avons émis l'hypothèse qu'*in vivo* d'autres facteurs auxiliaires contribuent à stabiliser le facteur M sur sa cible chloroplastique, renforçant encore l'hypothèse de complexes multimoléculaires suggérée plus haut. Pour tester cette hypothèse, j'ai réalisé des chimères ne comprenant qu'une petite portion de l'extrémité 5' non traduite du gène *atpB* (réduite à la séquence cible de MDB1), suivie d'une séquence exogène (codant pour une GFP), puis de l'extrémité 3' non traduite (NT) du gène chloroplastique *rbcL* (codant la grande sous-unité de la RuBisCO). Si un facteur secondaire interagit sur une autre région de l'ARNm *atpB* il ne devrait pas interagir avec cette séquence chimérique exogène.

En observant les niveaux d'accumulation de notre transcrit chimérique en présence ou absence du messager *atpB* endogène, nous avons constaté que la chimère souffrait considérablement de la compétition. Le transcrit *atpB* endogène devait manifestement accaparer la majeure partie des facteurs MDB1. Or ceci n'avait pas été observé précédemment pour d'autres chimères placées sous le contrôle du 5' NT *atpB*, où la répartition de MDB1 entre ARNm *atpB* et ARNm chimérique paraissait équivalente. En outre, ayant réalisé également des chimères similaires, avec un 5' NT d'*atpB* tronqué, mais porteuses du 3'NT d'*atpB*, nous n'avons pas observé d'amélioration de l'accumulation du transcrit en compétition, suggérant que le 3' NT n'est pas lié à cette meilleure propension à la stabilité. Au regard de ces résultats, nous pensons que des facteur(s) secondaire(s) interagissent avec la portion 3' du 5' NT d'*atpB*, absente dans les chimères au 5' NT court, et stabilisent MDB1 sur le transcrit. Ce complexe permettrait ainsi de former un système spécifique, reposant sur deux séquences d'*atpB*, mais résilient, des mutations modérées dans l'une des cibles pouvant être compensées en partie par la présence de l'autre facteur, pour contrôler l'expression de MDB1.



Construction initiale Construction chimérique

Par la suite, j'ai intégré des séquences cibles de MDB1 mutées dans nos chimères au 5'NT tronqué. Comme attendu, l'interaction entre MDB1 et l'ARN devient plus sensible aux mutations dans la séquence cible, que lorsque ces mêmes mutations sont introduites dans l'ARNm endogène (Figure 106). L'effet est d'autant plus marqué qu'ici nos chimères n'ont pas de compétiteurs, tous les facteurs MDB1 peuvent en théorie se consacrer à leur stabilisation, mais pour autant le niveau d'accumulation des transcrits chimériques est très faible. Cette expérience semble refléter l'affinité de MDB1 pour sa séquence cible. Elle nous a permis d'établir que la seconde portion de la séquence cible, côté 3', est reconnue plus spécifiquement par MDB1 que la portion en 5'.

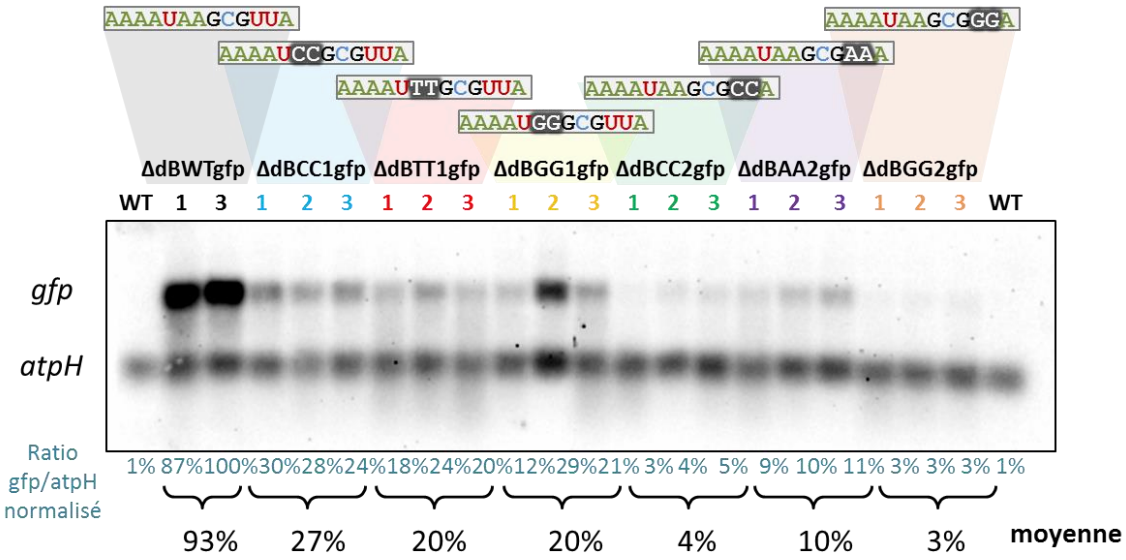


Figure 106: Gel d'ARN présentant l'accumulation du transcrit chimérique *gfp*. Les mutations (surlignées en noir) de la cible MDB1 entrainent cette fois une baisse plus importante du niveau de transcrit.

Grâce à ces résultats j'ai pu entrer dans la phase de validation du code. Nous avons dessiné des plasmides contenant la séquence de la protéine nucléaire MDB1 avec des mutations dans les répétitions OPR 6 et 7 ou 11 et 12, correspondant à la position des mutations dans la cible ARN, modifiant les résidus 5 et 6 en suivant le code préliminaire. J'ai transformé ces plasmides dans une souche déficiente en MDB1. J'ai pu récupérer des transformants exprimant bien le facteur **MDB1 muté**, puis les ai croisés avec les souches aux mutations dans les cibles chloroplastiques. Ceci nous a permis d'obtenir des descendants présentant le même niveau d'expression de la protéine MDB1 mutée. Cette confrontation entre facteur OPR et cible devrait nous permettre de valider ou d'infirmer le code de reconnaissance préliminaire. Ces travaux sur le code, non encore complètement achevés font partie d'un article en préparation.

Travaux complémentaires :

En plus de ces travaux sur les protéines OPR, j'ai également participé à des études de cRT-PCR sur *atpB* pour mieux comprendre sa dynamique de maturation et de dégradation, en lien avec MDB1 et son 3'UTR. Ces travaux aussi s'inscrivent dans un article.

J'ai aussi eu l'occasion de contribuer à la caractérisation de mutants de l'ATP-synthase et identifié un mutant de la sous-unité ATPG de l'ATP-synthase pour laquelle on ne possédait pas de mutants chez *Chlamydomonas*, et dont la caractérisation fera l'objet d'une publication.

ABSTRACT

During the post-endosymbiotic evolution of the chloroplast, its genome shrunk dramatically. Nowadays the various sub-units of the photosynthetic protein complexes are encoded partly in the nucleus, partly in the plastid genome. To achieve their correct assembly, the expression of the two genomes needs to be coordinated. Organellar Trans-Acting Factors (OTAF), encoded in the nucleus, are proteins which can bind specifically an organellar mRNA and control its expression. Among the several classes of OTAF factors the two that I studied were: the maturation and stabilisation (M) factors that stabilise and protect their cognate mRNA from exonucleases. And the translation (T) factors which are needed to initiate translation of a specific organellar mRNA. Among those OTAFs, the octotricopeptide repeat protein (OPR) family is abundant in *Chlamydomonas reinhardtii*. The OPR repeat is a degenerate motif of 38 amino-acids, folding into a tandem of antiparallel α-helices which can bind RNA. An individual OPR repeat is predicted to interact with one given nucleotide thanks to specificity-conferring residues at defined positions within the repeat. OPR proteins contain tracks of successive OPR motifs, each of them recognising one nucleotide, thus they can bind a specific RNA "target" sequence.

We aimed to study this specificity, called the "OPR code" starting with a draft code based on known OPR protein/mRNA pairs. A reliable code would enable the study of the many OPR factors with unknown targets. It could also be used to design and build proteins able to interact with specific RNAs. To confirm the draft "OPR code", I mutated *in vivo* the chloroplast targets of some OPR factors to disrupt the OPR/RNA interaction, and then tried to restore it by mutating the specificity-conferring residues in the corresponding repeats. Surprisingly, OPR/RNA interactions seem very resilient, challenging our view of how the specificity is established *in vivo*. Complementary functional studies that I performed on the OPR factors MDB1 and MTHI1 revealed how the control of chloroplast gene expression might rely on a network of OTAFs, with the M factor being the keystone of the expression system. Complexes of factors with moderate affinity for close RNA target sequences would cooperate to strengthen the interaction and anchor themselves on the mRNA. By cooperating, the overall affinity of the complex for the mRNA would be stronger, and the reliance on two or more RNA target sequences could allow compensating partly moderate mutations affecting one of them. Therefore, those putative systems would be both more specific and more resilient.