DEVELOPMENT OF PREDICTIVE STRATEGIES TO IMPROVE PLANT-BASED PRODUCTION OF HIGH-VALUE PROTEINS IN CHLOROPLAST-TRANSFORMED TOBACCO

A Dissertation

Presented to the Faculty of the Graduate School

Of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jennifer Anne Schmidt

May 2020

DEVELOPMENT OF PREDICTIVE STRATEGIES TO IMPROVE PLANT-BASED PRODUCTION OF HIGH-VALUE PROTEINS IN CHLOROPLAST-TRANSFORMED TOBACCO

Jennifer Anne Schmidt, Ph.D.

Cornell University 2020

Global demand for high value proteins is rapidly expanding in critical areas of agriculture, biopharmaceuticals, and industrial processing. Current methods of producing these high-value proteins are cell cultures that have high capital costs, require specialized maintenance, and are relatively inflexible to changing market demands. Chloroplast-engineered high biomass crops may offer a cheap and versatile alternative; however, this technology is still largely limited to proof-of-concept experiments. Growth chamber trials demonstrated the remarkable metabolic flexibility of transplastomic *Nicotiana tabacum* plants to synthesize a recombinant cellulase to as much as 38% of TSP with no deleterious mutant phenotype. These engineered plants also maintained robust heterologous protein yields even when grown in the unpredictable conditions of open field cultivation, an important step towards viable marketability. In an effort to address some of the current limitations of transplastomic plants, additional growth analyses investigated ways that foreign proteins can impact the host plant, such as imposing greater resource demands, exhausting protein synthesis, altering gene expression, and enzymatically interfering with host plant metabolism. Finally, analyses of *in silico* and *in vivo* mRNA secondary structure identified a likely mechanism of function for the downstream box (DB) regulatory region to conserve the native folding conformation of the transgenic 5'UTR. Synthetically designed DBs confirmed this hypothesized mechanism for recombinant cellulase production in *N. tabacum* plastids with on-going work focusing on expanding applicability to other high-value proteins and host organisms.

BIOGRAPHICAL SKETCH

Jen Schmidt grew up in Walnutport, Pennsylvania. She graduated from Northampton High School in 2010 and received her Bachelor of Science degree from Cornell University in Science of Earth Systems engineering in 2014. She joined Dr. Beth Ahner's lab in the fall of 2014 where she began her work investigating and tailoring chloroplast-engineered plants for high-value protein production. Jen received a USDA NIFA pre-doctoral fellowship in December 2018 to support this work. During her PhD, Jen was a teaching assistant for the Department of Ecology and Evolutionary Biology as well as for the Department of Molecular Biology and Genetics. She also took part in several educational outreach programs, including Cornell's 4H career development program and the Graduate School's Kids' Day in Science. She was a mentor and tour guide for the Society of Women Engineers and Cornell College of Engineering. In April 2019, Jen was inducted into the Tau Beta Pi National Engineering Honors Society.

ACKNOWLEDGEMENTS

I would like to thank Drs. Beth Ahner, Maureen Hanson, and Klaas van Wijk for their guidance in my training and education. I would also like to thank Drs. Lubna Richter and Mohammad Yazdani for providing advice and technical expertise. Several members of the Hanson lab also contributed to my training and troubleshooting, particularly Drs. Myat Lin and Vishal Chaudhari. The friends and colleagues I made during my time at Cornell made my PhD particularly enjoyable, especially Gozde Gencer, Julia Miller, Nick Segerson, Annie Kruse, Elena Michel, and Jessica Maya.

I would not have been able to complete my PhD without the unwavering support of my family. In particular, I'd like to thank my mother, Barb, for taking on the roles of assistant, editor, personal chef, sommelier, illustrator, and therapist. My father, Greg, played a critical part in this endeavor as my car mechanic, financial advisor, and brainstorming expert. My brother, Tom, and sister-in-law, Katie, helped out as my moral support, comic relief, and tech support.

Funding for the work was partially provided by a USDA National Institute of Food and Agriculture Fellowships Grant Program (2018-67011-28027/project accession No. 1015534).

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH		iii
ACKNOWLEDGEMENTS		iv
TABLE OF CONTENTS		v
LIST OF FIGURES		vi
LIST OF TABLES		viii
CHAPTER 1	Chloroplast-engineering for production of recombinant proteins: molecular and physiological molecular strategies to reduce mutant phenotypes and improve foreign protein yields	1
CHAPTER 2	Field-grown tobacco plants maintain robust growth while accumulating large quantities of a bacterial cellulase in chloroplasts	15
CHAPTER 3	Mitigation of deleterious phenotypes in chloroplast-engineered plants accumulating high levels of foreign proteins	46
CHAPTER 4	Designing synthetic downstream boxes to predictably alter recombinant protein production in chloroplast-engineered tobacco	81
CHAPTER 5	Conclusions and future prospects	119
APPENDIX I	Tailoring synthetic regulators for enhanced accumulation of recombinant proteins in chloroplast-engineered *Chlamydomonas reinhardtii*	123
APPENDIX II	Summary of statistics reports	141

LIST OF FIGURES

Figure 2.1	Biomass accumulation and photosynthetic rate of field-grown tobacco	20
Figure 2.2	Cel6A and Rubisco protein quantification from leaf tissues	22
Figure 2.3	Comparison between trials of protein allocation between Cel6A, Rubisco, and all other endogenous soluble proteins	24
Figure 2.4	Hypothesized strategies for resource allocation and adaptation to foreign protein accumulation in transplastomic tobacco grown under different conditions	27
Figure 2.S1	Recorded insolation, temperature, and precipitation for both growth seasons	38
Figure 2.S2	TetC-Cel6A and NPTII-Cel6A accumulation normalized to leaf area	39
Figure 3.1	Growth assessment of TetC-*cel6*A tobacco plants from two to twelve weeks old	51
Figure 3.2	Effect of gibberellic acid on germination and seedling shoot elongation	54
Figure 3.3	Effect of ammonium nitrate and carbon dioxide on growth of transgenic tobacco	56
Figure 3.4	Alteration in recombinant cellulase and Rubisco abundance with varied ammonium nitrate and carbon dioxide levels	58
Figure 3.5	Variation in protein resource allocation dependent on ammonium nitrate and carbon dioxide content	60
Figure 3.6	Summary of codon bias for several recombinant proteins compared to tobacco RbcL	65
Figure 3.S1	Additional growth measurements of soil-grown transgenic tobacco	71
Figure 3.S2	Protein accumulation normalized for whole plant biomass	72
Figure 3.S3	Comparison of the effect of GA on average days until root emergence between genotypes	73
Figure 3.S4	Effect of enhanced CO_2 on dry weight, water storage, and whole plant Cel6A accumulation	74

Figure 4.1	Cartoon schematic of the transgenic mRNA	86
Figure 4.2	*In silico* structures of 5'end of the transgenic *cel6*A and *bglC* mRNA	87
Figure 4.3	Comparison of structure prediction for *in silico* and *in vivo* modeling of the chloroplast 23S rRNA using phylogenetic reference model	90
Figure 4.4	DMS reactivity and *in vivo* mRNA secondary structure for cel6A transgenic transcripts	92
Figure 4.5	*In silico* structure models of the synthetic DB-*cel6*A fusions	94
Figure 4.6	Growth assessment of chloroplast-transformed generation T1 plants	96
Figure 4.7	Cel6A protein and mRNA quantification in new chloroplast-transformed tobacco lines	98
Figure 4.S1	Southern blots for each of the homoplasmic transformed lines	111
Figure 4.S2	Germination and seed viability of transplastomic tobacco	112
Figure AI.1	Vector schematic of the algae chloroplast constructs	126
Figure AI.2	Predicted *in silico* mRNA secondary structure of the translation initiation region of the phytase constructs	127
Figure AI.3	Predicted *in silico* mRNA secondary structures for the transgenic IL2 transcripts	130

LIST OF TABLES

Table 2.S1	Fertilizer application to field-grown tobacco	35
Table 2.S2	Pesticide application to field-grown tobacco	36
Table 2.S3	Summary of cost analysis calculations and assumptions	37
Table 4.S1	Primers and linkers for *in vivo* RNA structure analysis	108
Table 4.S2	Synthetic DB plasmid construction primers	109
Table 4.S3	qPCR primers	110
Table AI.S1	Phytase transformation plasmid construction primers.	134
Table AI.S2	IL2 transformation plasmid construction primers	135
Table AII.1	Summary of field trial statistics	141
Table AII.2	Summary of age-dependent growth and Cel6A trial statistics	145
Table AII.3	Summary of statistics for effect of gibberellic acid on germination rate	147
Table AII.4	Summary of effect of Gibberellic acid application on stem length statistics	148
Table AII.5	Summary of CO_2 and ammonium nitrate chamber trial statistics	149
Table AII.6	Summary mRNA folding analysis and Cel6A accumulation statistics	153

CHAPTER 1

Chloroplast engineering for production of recombinant proteins: molecular and physiological strategies to reduce mutant phenotypes and improve foreign protein yields

Abstract

Well-established cultivation strategies coupled with more recent technical advancements in molecular biology make genetically engineered plants a promising alternative for high-value protein production to meet industrial, agricultural, and pharmaceutical demands. A growing body of work confirms that transplastomic plants are capable of producing large quantities of bioactive foreign proteins, as much as 40% of total soluble protein (TSP), without interfering with native plant metabolism. However, recombinant protein yields are project-specific and unpredictably vary from less than 1% to 70% of TSP. In an effort to improve the commercial applicability of this technology, one of the goals of the work presented herein is to reduce the uncertainty in establishing high yielding transplastomic tobacco lines by addressing two main challenges. The first is to understand the physiological demands of recombinant protein production on the host plant to increase heterologous protein yield without compromising plant growth. The second component focuses on fine-tuning gene regulation through components of mRNA secondary structure to reduce project-to-project variability in target protein synthesis.

Benefits and limitations of plastid engineering for large-scale protein production

Genetic engineering of the plastid genome was first accomplished in the freshwater algae *Chlamydomonas reinhardtii* in 1988 and soon after in *Nicotiana tabacum* [1, 2]. More recent advancements in molecular biology and tissue culture techniques have enabled the development of protocols for several other plant species, including potato, tomato, cotton, and canola [3-6].

Since then, plastid genetic engineering has been harnessed to better understand mechanisms of the plastid including photosynthesis, gene regulation, and plastid inheritance [7-9]. In particular, the considerable protein production capacity of higher plant chloroplasts make transplastomic plants a promising production method for high-value proteins [10].

Currently, high-value proteins are most often synthesized in bacterial or fungal cell culture systems with production costs averaging \$10 kg^{-1} protein [11]. However, these microbes, especially bacteria, are unable to perform many of the higher eukaryotic post-translational modifications (PTMs), limiting the usefulness of microbes for biopharmaceuticals [12]. Instead, engineered mammalian cells or milk from transgenic mammals are often used to synthesize these complex eukaryotic proteins. The risk of contamination in the mammalian cell cultures and the technical challenges of engineering mammals drive production costs ranging from \$100 - \$3000 kg^{-1} protein [13].

With well-established field cultivation practices, engineered biomass crops may be a cost-effective alternative to produce some of these high-value proteins. In fact, more than one hundred functional exogenous proteins have been produced from nuclear-engineered plants, including the Ebola treatment Zmapp™ [14, 15]. Additionally, a variety of monoclonal antibodies, vaccines, and cytokines have been successfully produced in transgenic plants as stable transformants, *in vitro* cell cultures, and via transient expression in higher plants (reviewed [16]). However, despite these successes, low foreign protein yields, often 1% of TSP or less, limit the usefulness of this technology. In comparison, recombinant protein yields from plastid-engineered GE plants often average 5-20% of TSP with exceptional transformants reaching 40-70% exogenous protein of TSP [17-25] . These high yields are often attributed to the lack of gene silencing in chloroplast, the high expression rates of native plastid genes, as well as the high

copy number of plastid genomes in leaf cells (reviewed in [10]). Furthermore, the near-complete maternal inheritance of plastids provides more stringent transgene containment than nuclear transgenes [9].

Although plastids are not capable of carrying out all mammalian PTMs, transplastomic plants can synthesize a variety of bioactive human medical proteins, including those to treat B cell lymphoma, anthrax, botulism, and herpes simplex virus (reviewed in [26]). Transplastomic plants have also successfully synthesized bacterial- and fungal-derived proteins for nutritional quality improvement of food and animal feed [27, 28], enhanced plant pest resistance [19, 29], as well as numerous enzymes with applications in industrial processes [20, 21, 23, 30].

Despite its applicability for a wide variety of proteins, commercialization of chloroplast engineering will require a shift in research focus from proof-of-concept to product development. Two such areas of focus are addressing the altered physiological demands of engineered plants and decreasing variability in foreign protein yields.

Physiological drivers of transgenic plant health and recombinant protein expression

Accumulation of a foreign protein represents an additional resource sink for the host plant, particularly for chloroplast-engineered plants accumulating on average 5-20% recombinant protein of total soluble protein (TSP). That many transplastomic plants can achieve such yields with little or no deleterious phenotypes demonstrates the considerable plasticity of tobacco plastids for reallocating resources to accommodate foreign protein synthesis [17, 18, 20, 21]. However, exceeding the limit of this buffering capacity often results in mutant phenotypes, as was shown in tobacco plants expressing PlyGBS to 70% of TSP and another line expressing CelB to 75% of TSP. In these cases, mutant phenotypes are likely triggered by excessive

depletion of plant protein resources, reducing native proteins beyond the plant's ability to maintain robust growth.

One method of resolving resource-limited growth of transplastomic plants is altering environmental conditions to optimize growing conditions. For instance, increasing light intensity and duration as well as increasing fertilization have been shown to improve transplastomic plant growth [17, 31]. Nitrogen is of particular interest for high-value protein production because it is a major component of proteins and is often a limiting resource in large-scale field cultivation. Increasing nitrogen input often leads to a general increase in amino acid synthesis, protein production, and biomass accumulation, which, in turn, can enhance foreign protein abundance [32].

In recent years, a lot of work has been done to understand the effect rising atmospheric CO_2 will have on plant growth and resource use. While this has yet to be applied to chloroplast-engineered plants, artificially elevating carbon dioxide (CO_2) may help alleviate mutant phenotypes. Plant response to elevated CO_2 depends strongly on several factors, including nutrient availability, water status, plant species, and abiotic growing conditions (i.e. insolation, temperature, etc.). However, well-fertilized C3 plants tend to exhibit significant increases in photosynthetic rate which lead to elevated carbohydrate production and biomass (reviewed in [32]). This enhancement of photosynthesis often triggers changes in gene expression, particularly the drastic down-regulation of many photosynthetic enzymes like Rubisco [32, 33]. Well-fertilized plants have also been shown to increase amino acid accumulation and up-regulate protein synthesis when grown in elevated CO_2 [34]. For a plant expressing a foreign protein, these metabolic changes could improve plant growth or enhance recombinant protein yields.

Understanding and tailoring transgene regulation to promote predictably high recombinant protein accumulation

While typical yields from transplastomic tobacco range from 5-20% of TSP, results are often unpredictable and yields of less than 1% of TSP are not uncommon [18, 35, 36]. One key to improving the marketability of plastid-engineered plants is developing a system to consistently yield large quantities of any recombinant protein of interest, necessitating a better understanding of plastid gene regulation as it pertains to the expression of foreign genes.

Early on, the 5' untranslated region (UTR) was identified as a major determinant of foreign gene expression for its role in regulating ribosome binding and translation initiation rates (reviewed in [37]). Native plastid 5'UTRs of highly expressed genes, like *rbcL, psbA,* and *atpB*, often yield high foreign protein accumulation when used in the transgenic cassette (reviewed in [38]). However, unintended recombination between these UTRs and their native copies can cause complications establishing correctly engineered lines [25]. Native UTRs can also trigger mutant phenotypes by competing with their native transcripts for essential RNA-binding factors. A notable example of such competition was reported in tobacco plants producing NPTII under regulation by the native tobacco *clpP* 5'UTR. The transgenic *clpP* 5'UTR interfered with processing of the native *clpP* mRNA, causing a severe stunted, chlorotic phenotype [39]. Non-native UTRs have the benefit of not interfering as directly with native gene elements; however, these UTRs risk lower protein yields if they are not as readily recognized by plastid translation machinery. Thus far, the T7G10 5'UTR has proven to be a very amenable choice, often yielding high levels of transgenic proteins [20-22, 24, 40].

A series of transcriptome thermodynamic modeling, *in vivo* mRNA structure analyses, and point mutation experiments have been used to better understand the mechanism of the

5'UTR's regulatory role [41-50]. Ribosomes contain a helicase activity only after they are fully assembled, so the procession of the pre-initiation complex is strongly dependent on the thermodynamic stability of the mRNA in the 5'UTR [37]. In many cases, 5'UTRs with thermodynamically weak structures result in greater protein accumulation [41, 44, 50]. Furthermore, 5'UTR structures that leave the SD and start codon single-stranded facilitate ribosome binding and translation initiation [41, 45].

Another important regulatory region in the transcript's 5'end is the downstream box (DB), located downstream of the start codon. The DB has also previously demonstrated a strong regulatory role in bacteria, where expression of dihydrofolate reductase increased by 4-fold when the bacteriophage T7G10 DB region was included in the expression cassette [51]. Initially, it was believed that this region functioned similarly to the Shine-Dalgarno (SD) sequence by interacting with a portion of the rRNA to aid in ribosome assembly [52]. However, that hypothesis was proven unlikely and, since then, the mechanism of the DB's function is still not well understood [53].

If the DB's primary role was regulating transcription, mRNA concentration would correlate with protein accumulation. However, several studies have shown that the identity of the DB and resulting protein concentrations are independent of transgenic mRNA concentration [20, 54]. In proteins where the initiator methionine is removed post-translationally, the first amino acid of the DB would act as the N-terminal protein stability determinant [7]. However, the DB's function extends beyond just the first amino acid because DBs with the same N-terminal amino acid have fostered widely different protein accumulation [20, 21]. Furthermore, silent mutations in the DB region still result in dramatic changes in protein accumulation, indicating that the DB's function is more dependent on the nucleotide sequence than the amino

acid sequence [54]. Thus, by process of elimination, it is commonly believed that the DB regulates translation rate or initiation.

Unlike the 5'UTR where a particular UTR may universally promote high expression for a variety of transgenes, the DB seems to be much more context dependent, with protein accumulation patterns depending strongly on the gene fused to the DB. For example, early work demonstrated that bacterial beta-glucosidase (BglC) expression was as much as ten times greater in constructs fused with the DB derived from the bacterial *nptII* gene than that of the *tetC* gene [21]. Meanwhile, these same DBs gave a reversed accumulation pattern when fused to a heterologous cellulase (Cel6A) with the *tetC* DB yielding significantly greater Cel6A protein than the *nptII* DB [20]. Furthermore, while total Cel6A yield was lower for both TetC-*cel6*A and NPTII-*cel6*A, the expression pattern between the two *cel6*A constructs was the same in plastid-engineered *C. reinhardtii,* suggesting that the mechanism of the DB's function is conserved between algae and higher plants [28]. With a doubling time of under ten hours, transplastomic algae would make a good platform for rapidly designing and testing DBs before scaling up to more labor intensive, but higher yielding, land plants.

One focus of this dissertation was to investigate a potential regulatory mechanism for the DB with the aim of tailoring this mechanism to produce large quantities of foreign proteins from plastid-engineered plants. In addition to industrial cellulases, preliminary work has also begun testing methods of designing DBs for different high-value proteins, including an agriculturally valuable phytase as well as a human immune system protein, interleukin-2, in the chloroplasts of *C. reinhardtii*. The molecular-scale analysis of DB regulatory control was coupled with a series of growth trials in *N. tabacum* to better understand resource reallocation in response to synthesis of an exogenous protein as a method of improving cultivation of these engineered plants.

References

1. Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM: Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. Science 1988, 240(4858):1534.
2. Svab Z, Hajdukiewicz P, Maliga P: Stable transformation of plastids in higher plants. PNAS 1990, 87(21):8526-8530.
3. Cheng L, Li H-P, Qu B, Huang T, Tu J-X, Fu T-D, Liao Y-C: Chloroplast transformation of rapeseed (*Brassica napus*) by particle bombardment of cotyledons. Plant Cell Rep 2010, 29(4):371-381.
4. Kumar S, Dhingra A, Daniell H: Stable transformation of the cotton plastid genome and maternal inheritance of transgenes. Plant Mol Biol 2004, 56(2):203-216.
5. Ruf S, Hermann M, Berger IJ, Carrer H, Bock R: Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. Nat Biotechnol 2001, 19(9):870.
6. Sidorov VA, Kasten D, Pang SZ, Hajdukiewicz PT, Staub JM, Nehra NS: Stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. The Plant Journal 1999, 19(2):209-216.
7. Apel W, Schulze WX, Bock R: Identification of protein stability determinants in chloroplasts. Plant J 2010, 63(4):636-650.
8. Hanson MR, Gray BN, Ahner BA: Chloroplast transformation for engineering of photosynthesis. J Exp Bot 2013, 64(3):731-742.
9. Svab Z, Maliga P: Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. PNAS 2007, 104(17):7003-7008.

10. Maliga P, Bock R: Plastid biotechnology: food, fuel, and medicine for the 21st century. Plant Physiol 2011, 155(4):1501-1510.

11. Liu G, Zhang J, Bao J: Cost evaluation of cellulase enzyme for industrial-scale cellulosic ethanol production based on rigorous Aspen Plus modeling. Bioprocess Biosyst Eng 2016, 39(1):133-140.

12. Yin J, Li G, Ren X, Herrler G: Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. J Biotechnol 2007, 127(3):335-347.

13. Doran PM: Foreign protein production in plant tissue cultures. Curr Opin Biotechnol 2000, 11(2):199-204.

14. Olinger GG, Pettitt J, Kim D, Working C, Bohorov O, Bratcher B, Hiatt E, Hume SD, Johnson AK, Morton J: Delayed treatment of Ebola virus infection with plant-derived monoclonal antibodies provides protection in rhesus macaques. PNAS 2012, 109(44):18030-18035.

15. Qiu X, Audet J, Wong G, Fernando L, Bello A, Pillet S, Alimonti JB, Kobinger GP: Sustained protection against Ebola virus infection following treatment of infected nonhuman primates with ZMAb. Scientific reports 2013, 3:3365.

16. Xu J, Towler M, Weathers P: Platforms for Plant-Based Protein Production. In: *Bioprocessing of Plant In Vitro Systems.* Edited by Pavlov A, Bley T. Cham: Springer; 2016: 1-40.

17. Bally J, Nadai M, Vitel M, Rolland A, Dumain R, Dubald M: Plant physiological adaptations to the massive foreign protein synthesis occurring in recombinant chloroplasts. Plant Physiol 2009, 150(3):1474-1481.

18. Castiglia D, Sannino L, Marcolongo L, Ionata E, Tamburino R, De Stradis A, Cobucci-Ponzano B, Moracci M, La Cara F, Scotti N: High-level expression of thermostable cellulolytic enzymes in tobacco transplastomic plants and their use in hydrolysis of an industrially pretreated *Arundo donax* L. biomass. Biotechnol Biofuels 2016, 9(1):154.

19. Dufourmantel N, Dubald M, Matringe M, Canard H, Garcon F, Job C, Kay E, Wisniewski JP, Ferullo JM, Pelissier B: Generation and characterization of soybean and marker-free tobacco plastid transformants over-expressing a bacterial 4-hydroxyphenylpyruvate dioxygenase which provides strong herbicide tolerance. Plant Biotechnol J 2007, 5(1):118-133.

20. Gray BN, Ahner BA, Hanson MR: High-level bacterial cellulase accumulation in chloroplast-transformed tobacco mediated by downstream box fusions. Biotechnol Bioeng 2009, 102(4):1045-1054.

21. Gray BN, Yang H, Ahner BA, Hanson MR: An efficient downstream box fusion allows high-level accumulation of active bacterial beta-glucosidase in tobacco chloroplasts. Plant Mol Biol 2011, 76(3):345-355.

22. Oey M, Lohse M, Kreikemeyer B, Bock R: Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. Plant J 2009, 57(3):436-445.

23. Petersen K, Bock R: High-level expression of a suite of thermostable cell wall-degrading enzymes from the chloroplast genome. Plant Mol Biol 2011, 76(3-5):311-321.

24. Tregoning JS, Nixon P, Kuroda H, Svab Z, Clare S, Bowe F, Fairweather N, Ytterberg J, van Wijk KJ, Dougan G: Expression of tetanus toxin fragment C in tobacco chloroplasts. Nucleic Acids Res 2003, 31(4):1174-1179.

25. Zhou F, Badillo-Corona JA, Karcher D, Gonzalez-Rabade N, Piepenburg K, Borchers AM, Maloney AP, Kavanagh TA, Gray JC, Bock R: High-level expression of human immunodeficiency virus antigens from the tobacco and tomato plastid genomes. Plant Biotechnol J 2008, 6(9):897-913.

26. Yusibov V, Kushnir N, Streatfield SJ: Antibody Production in Plants and Green Algae. Annu Rev Plant Biol 2016, 67:669-701.

27. Liu X, Zhang C, Wang X, Liu Q, Yuan D, Pan G, Sun SS, Tu J: Development of high-lysine rice via endosperm-specific expression of a foreign LYSINE RICH PROTEIN gene. BMC Plant Biol 2016, 16(1):147.

28. Richter LV, Yang H, Yazdani M, Hanson MR, Ahner BA: A downstream box fusion allows stable accumulation of a bacterial cellulase in *Chlamydomonas reinhardtii* chloroplasts. Biotechnol Biofuels 2018, 11(1):133.

29. Chakrabarti SK, Lutz KA, Lertwiriyawong B, Svab Z, Maliga P: Expression of the *cry9Aa2 B.t.* gene in tobacco chloroplasts confers resistance to potato tuber moth. Transgenic Res 2006, 15(4):481-488.

30. Kolotilin I, Kaldis A, Pereira EO, Laberge S, Menassa R: Optimization of transplastomic production of hemicellulases in tobacco: effects of expression cassette configuration and tobacco cultivar used as production platform on recombinant protein yields. Biotechnol Biofuels 2013, 6(1):1.

31. Stevens LH, Stoopen GM, Elbers IJ, Molthoff JW, Bakker HA, Lommen A, Bosch D, Jordi W: Effect of climate conditions and plant developmental stage on the stability of antibodies expressed in transgenic tobacco. Plant Physiol 2000, 124(1):173-182.

32. Stitt M, Krapp A: The interaction between elevated carbon dioxide and nitrogen nutrition: the physiological and molecular background. Plant, Cell Environ 1999, 22(6):583-621.

33. Moore B, Cheng SH, Rice J, Seemann J: Sucrose cycling, Rubisco expression, and prediction of photosynthetic acclimation to elevated atmospheric CO_2. Plant, Cell Environ 1998, 21(8):905-915.

34. Huang B, Xu Y: Cellular and Molecular Mechanisms for Elevated CO_2–Regulation of Plant Growth and Stress Adaptation. Crop Sci 2015, 55(4):1405-1424.

35. Hidalgo D, Abdoli-Nasab M, Jalali-Javaran M, Bru-Martínez R, Cusidó RM, Corchete P, Palazon J: Biotechnological production of recombinant tissue plasminogen activator protein (reteplase) from transplastomic tobacco cell cultures. Plant Physiol Biochem 2017, 118:130-137.

36. Lenzi P, Scotti N, Alagna F, Tornesello ML, Pompa A, Vitale A, De Stradis A, Monti L, Grillo S, Buonaguro FM: Translational fusion of chloroplast-expressed human papillomavirus type 16 L1 capsid protein enhances antigen accumulation in transplastomic tobacco. Transgenic Res 2008, 17(6):1091-1102.

37. Zoschke R, Bock R: Chloroplast translation: structural and functional organization, operational control, and regulation. The Plant Cell 2018, 30(4):745-770.

38. Maliga P: Plastid transformation in higher plants. Annu Rev Plant Biol 2004, 55:289-313.

39. Kuroda H, Maliga P: Overexpression of the *clpP* 5′-untranslated region in a chimeric context causes a mutant phenotype, suggesting competition for a *clpP*-specific RNA maturation factor in tobacco chloroplasts. Plant Physiol 2002, 129(4):1600-1606.

40. Kuroda H, Maliga P: Complementarity of the 16S rRNA penultimate stem with sequences downstream of the AUG destabilizes the plastid mRNAs. Nucleic Acids Res 2001, 29(4):970-975.

41. De Smit MH, Van Duin J: Control of prokaryotic translational initiation by mRNA secondary structure. In: *Prog Nucleic Acid Res Mol Biol.* vol. 38: Elsevier; 1990: 1-35.

42. De Smit MH, van Duin J: Control of translation by mRNA secondary structure in *Escherichia coli*: a quantitative analysis of literature data. J Mol Biol 1994, 244(2):144-150.

43. Ding Y, Kwok CK, Tang Y, Bevilacqua PC, Assmann SM: Genome-wide profiling of *in vivo* RNA structure at single-nucleotide resolution using structure-seq. Nat Protoc, 2015, 10(7):1050-1066.

44. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM: *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature 2014, 505(7485):696-700.

45. Gu W, Zhou T, Wilke CO: A Universal Trend of Reduced mRNA Stability near the Translation-Initiation Site in Prokaryotes and Eukaryotes. PLoS Comput Biol 2010, 6(2):e1000664.

46. Kwok CK, Ding Y, Tang Y, Assmann SM, Bevilacqua PC: Determination of *in vivo* RNA structure in low-abundance transcripts. Nat Commun 2013, 4.

47. Kwok CK, Tang Y, Assmann SM, Bevilacqua PC: The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. Trends Biochem Sci 2015, 40(4):221-232.

48. Seo SW, Yang J-S, Kim I, Yang J, Min BE, Kim S, Jung GY: Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. Metab Eng 2013, 15:67-74.

49. Shabalina SA, Ogurtsov AY, Spiridonov NA: A periodic pattern of mRNA secondary structure created by the genetic code. Nucleic Acids Res 2006, 34(8):2428-2437.

50. Yin J, Bao LC, Tian H, Gao XD, Yao WB: Quantitative relationship between the mRNA secondary structure of translational initiation region and the expression level of heterologous protein in *Escherichia coli*. J Ind Microbiol Biotechnol 2016, 43(1):97-102.

51. Sprengart ML, Fuchs E, Porter A: The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*. EMBO J 1996, 15(3):665.

52. Faxén M, Plumbridge J, lsaksson LA: Codon choice and potential complementarity between mRNA downstream of the initiation codon and bases 1471 –1480 in 16S ribosomal RNA affects expression of glnS. Nucleic Acids Res 1991, 19(19):5247-5251.

53. O'Connor M, Asai T, Squires CL, Dahlberg AE: Enhancement of translation by the downstream box does not involve base pairing of mRNA with the penultimate stem sequence of 16S rRNA. PNAS 1999, 96(16):8973-8978.

54. Kuroda H, Maliga P: Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. Plant Physiol 2001, 125(1):430-436.

CHAPTER 2

Field-grown tobacco plants maintain robust growth while accumulating large quantities of a bacterial cellulase in chloroplasts[1,2]

Abstract

High accumulation of heterologous proteins expressed from the plastid genome has sometimes been reported to result in compromised plant phenotypes. Comparisons of transplastomic plants to wild-type are typically made in environmentally-controlled chambers with relatively low light; little is known about the performance of such plants in field conditions. Here we report on two plastid-engineered tobacco lines expressing the bacterial cellulase Cel6A. Field-grown plants producing Cel6A at ~20% of total soluble protein (TSP) exhibit no loss in biomass or Rubisco content and only minor reductions in photosynthesis compared to wild-type. These experiments demonstrate that when grown in the field, tobacco possesses sufficient metabolic flexibility to accommodate high levels of recombinant protein by increasing total protein synthesis and accumulation and/or by reallocating unneeded endogenous proteins. Based on current tobacco cultivation practices and readily achievable recombinant protein yields, we estimate that particular proteins could be obtained from field-grown transgenic tobacco plants at costs three orders of magnitude less than current cell culture methods.

[1] Schmidt JA, McGrath JM, Hanson MR, Long SP, Ahner BA: Field-grown tobacco plants maintain robust growth while accumulating large quantities of a bacterial cellulase in chloroplasts. Nature Plants 2019, 5(7):715-721.

[2] I did all of the protein assays, statistical analyses, and co-authored the manuscript.

Introduction

Genetic engineering of plants is emerging as an innovative and sustainable means of meeting modern society's growing demand for valuable protein products. The market for recombinant proteins, expected to reach US$ 300 billion in the near future [1], is diverse and includes pharmaceuticals, nutritional improvement of food and feed products, as well as enzymes for consumer goods and various industrial processes. Such diversity necessitates versatile and scalable production platforms that can meet varying market demands. Currently most specialty protein synthesis is achieved through mammalian, insect, fungal, or bacterial cell cultures that are often expensive to maintain, prone to contamination, and inflexible to rapidly changing demand [2]. Field-grown transgenic crop plants may offer a safe and cost-effective alternative to augment or replace a significant fraction of this large and growing market [1, 3].

Early pioneers in this field focused on industrial proteins like GUS, lysozymes, trypsin, and amylases (reviewed in [4]). However, expression of proteins for use in biomedicine soon followed. For instance, in 2012 Protalix commercialized Elelyso™ to treat Type I Gaucher's disease, as a cheaper plant-produced alternative to the traditionally synthesized Cerezyme®[5, 6]. Currently, more than twenty plant-produced pharmaceutical proteins are undergoing clinical trials for the treatment of human diseases including Ebola, avian influenza, foot and mouth disease, and Newcastle disease, as well as for several veterinary vaccines [7, 8].

In plants, foreign protein expression can be mediated by the insertion of genes into the nuclear or the chloroplast genome or by transient expression through viral vectors [9]. Chloroplast engineering is largely untapped by plant biotechnology companies due to a relatively recent shift in research focus from proof-of-concept to commercial applicability [10, 11]. Commercial effort has been focused on nuclear transformation, which has proved far more

tractable than plastid transformation in most crops. However, plastid engineering is well established in tobacco, which provides an agronomically viable platform due to its large yield of leaf material. Further, plastids provide several unique advantages that may be key to commercialization. First, because of the nearly exclusive maternal inheritance of chloroplasts in most crop plants, chloroplast transformants have the ecological benefit that they are unlikely to disperse the transgene in pollen [12]. Second, introducing a gene into plastid DNA results in high copy number per cell, due to the multiplicity of plastomes (ranging from 2,000 to 20,000 copies per cell) [13, 14]. Finally, plastids intrinsically feature a high expression rate of many native genes and efficient homologous recombination machinery, allowing chloroplast-transformed plants to accumulate large amounts of recombinant protein [12, 15, 16].

However, while heterologous protein yields of up to 70% of total soluble protein (TSP) are possible [17], transformants accumulating massive amounts of foreign protein are often severely stunted. Growth restrictions of engineered plants may sometimes be caused by the low light environments of laboratory growth chambers. Under limiting light conditions, photosynthetic output is reduced and diversion of fixed carbon and other vital resources into recombinant protein synthesis will inevitably decrease growth. Alternatively, with higher light input, like that of the typical solar insolation of a field, light energy is often in excess of photosynthetic capacity [18], mitigating the burden of recombinant protein production. Further, for all but the most valuable transgenic proteins, growth chamber and greenhouse cultivation are too expensive for large-scale commercial production, so field cultivation offers the most economically sensible production scheme. The true test of industrial applicability of chloroplast-engineered plants is to analyze their growth and heterologous protein yield in the high-light, but less predictable environment of open-field cultivation. To our knowledge, this study is the first

to test the hypothesis that under field production conditions and in chambers delivering the equivalent of full sunlight, high levels of recombinant protein production in plastid-transformed plants can be achieved without compromising photosynthesis or harvestable biomass. An earlier study described biomass yield of dwarf tobacco transplastomic plants in the field, but no measurements of recombinant protein in the field-grown tissue were included, nor were the growth and photosynthetic rates of the transplastomic plants compared to wild-type [19].

The present study utilizes tobacco lines engineered to express an endoglucanase isolated from *Thermobifida fusca*, Cel6A, as a model protein to explore protein accumulation in plants. Cel6A is one member of a large group of industrially valuable thermotolerant cellulases used in the manufacturing of detergents and textiles, processing of food and animal feed, as well as in new markets such as cellulosic fuel and bioproduct production [20, 21]. Such cellulolytic enzymes can account for as much as 30% of the total production cost of biomass-derived green products according to life cycle assessments of cellulosic ethanol production [22]; thus, lower cost cellulases will improve the viability of cellulosic fuels and bioproducts relative to those derived from fossil fuels.

Herein we present results from studies that explore protein allocation in two different chloroplast-transformed lines when they are grown under a variety of conditions including two successive field seasons. The field trials represent a major step toward commercialization of chloroplast-transformed tobacco plants by demonstrating their profound resilience even while burdened with remarkable heterologous protein accumulation.

Results

Growth of chloroplast transformants in the field

The two chloroplast transformants used in this study were developed previously by our laboratories to show that the first fourteen codons near the 5'end of the open reading frame, also called the downstream box [23] (DB), can result in large gene-specific changes in exogenous protein accumulation [24]. Both transformants contain the *cel6A* gene for protein expression, one with the DB sequence native to the *tet*C gene (hereafter referred to as TetC-*cel6*A) and the other with the DB sequence from *npt*II (hereafter referred to as NPTII-*cel6*A). TetC-*cel6*A accumulates roughly ten times more Cel6A than NPTII-*cel6*A [24] which we hypothesized might limit the ability of TetC-*cel6*A plants to maintain wild-type (WT) biomass levels in the field.

In previous chamber trials experiments, Tetc-*cel6*A and NPTII-*cel6*A genotypes were reported to grow the same as WT tobacco [24]. In both field seasons, there was also no measurable difference in total above-ground biomass accumulation between the transformants and WT tobacco (Figure 2.1a, b). Final plant biomass was greater in Year 2 compared to Year 1, despite less total time in the field. This difference in yield is likely due to differences in weather conditions and/or fertilization procedures between the two growing seasons (Figure 2.S1, Table 2.S1). Due to regulatory limitations, plants were harvested just before flowering after 11 and 9 weeks in the field in Year 1 and Year 2, respectively; thus, seed yields were not measured in field-grown plants. However, no major fertility issues are observed in either TetC-*cel6*A or NPTII-*cel6*A tobacco lines when grown in chambers.

Photosynthetic efficiency was measured in the field at the time of harvest. During the Year 1 field season, measurements of photosynthesis defined by the maximum rate of

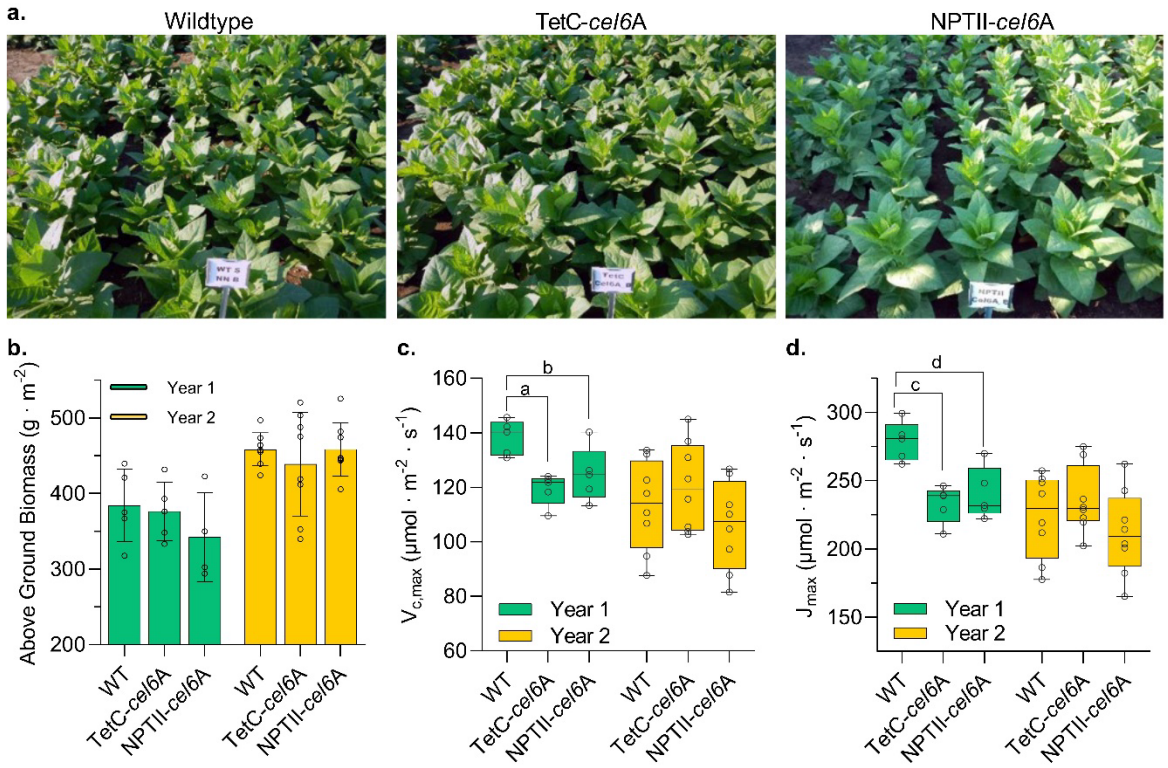


Figure 2.1: Biomass accumulation and photosynthetic rate of field-grown tobacco. **a.** Representative photos of field-grown tobacco. **b.** Above ground fresh weight of field-grown tobacco. n=5 plants for WT and TetC-*cel6*A in Field Year 1, n= 4 plants for NPTII-*cel6*A Field Year 1, and n= 8 plants for all genotypes in Field Year 2. Means and SEM are displayed along with individual data points for all biological replicates. **c.** $V_{c,max}$ for field-grown tobacco. Student's t-test p-values: a: $p< 0.002$, b: $p< 0.04$. **d.** J_{max} of field-grown tobacco. Student's t-test p-values: c: $p< 0.001$, d: $p< 0.008$. For both $V_{c,max}$ and J_{max} n= 5 plants for all genotypes in Field Year 1 and n= 8 plants for all genotypes in Field Year 2. Both $V_{c,max}$ and J_{max} results are shown as box-and-whisker plots with boxes displaying from the 25th to 75th percentiles, whiskers extending to the minimum and maximum values, and the center line indicates the median with all individual data points overlaid. Two-tailed Student's t-test p-values are provided in Table AII.1.

carboxylation, reflective of the activity of Rubisco ($V_{c,max}$), and the rate of electron transport through the photosystems (J_{max}), revealed a small difference between the transformants and WT (Figure 2.1c, d). Both transformed lines exhibited a 10-13% decrease in $V_{c,max}$ and a 13-16% decrease in J_{max} compared to WT. However, no difference from WT was measured in either transformed line in the Year 2 field season (Figure 2.1c, d). We also saw no differences between WT and transformed plants in these same parameters when plants were grown to the same stage in chambers (data not shown). Chloroplast transformants were phenotypically and, for the most part, photosynthetically indistinguishable from WT tobacco in both of the field cultivation trials.

Protein accumulation in transformed lines

As with earlier work with TetC-*cel6*A and NPTII-*cel6*A tobacco, the identity of the DB, consisting of the first forty-eight nucleotides after the start codon (Figure 2.2a), substantially influences recombinant cellulase expression. Immunoblots of leaf TSP reveal that field-grown plants are producing large amounts of Cel6A though, when reported as a fraction of TSP, less than plants grown in full sunlight chamber conditions (Figure 2.2b, c). TetC-*cel6*A plants grown in chambers accumulated as much as 40% Cel6A of TSP with an average abundance of 36% whereas NPTII-*cel6*A plants accumulated 6% on average. In both field seasons, TetC-*cel6*A plants accumulated an average of nearly 20% Cel6A of TSP, with no significant difference noted between seasons. In contrast, NPTII-Cel6A accumulation was comparable in Year 1 to the chamber-grown transformants, with an average accumulation of approximately 6% of TSP, but in Year 2, Cel6A levels were significantly lower with yields of 1% Cel6A of TSP (Figure 2.2c).

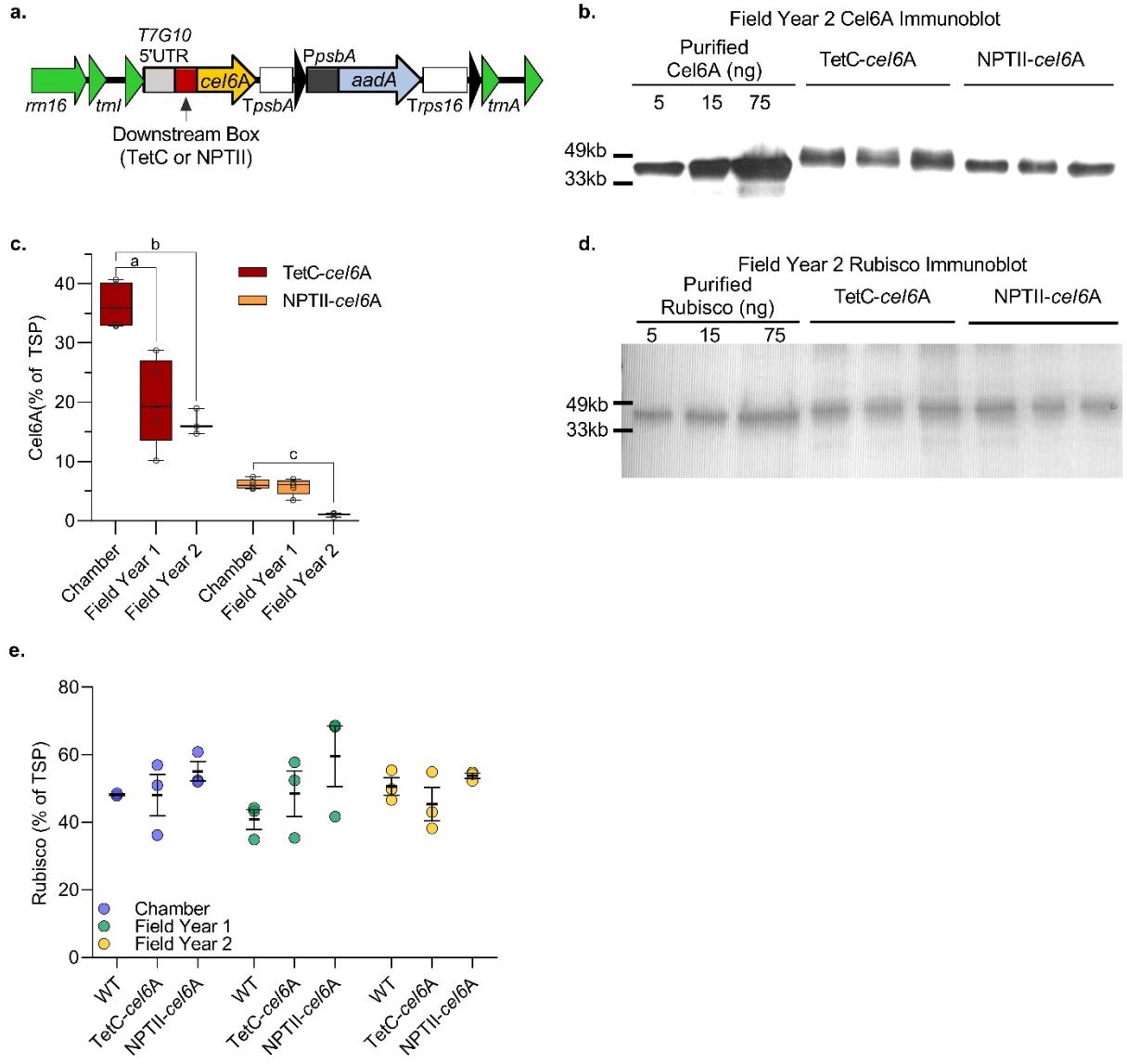


Figure 2.2: Cel6A and Rubisco protein quantification from leaf tissues. **a.** Vector schematic of transgenic constructs developed previously and published as Figure 1 in Gray *et al.* 2009. **b.** Representative Cel6A immunoblot. 200 ng TSP was loaded for each sample. **c.** TetC-Cel6A and NPTII-Cel6A accumulation presented as a percent of TSP. n= 5 plants for both genotypes in Chamber and Field Year 1 and n= 3 plants for both genotypes in Field Year 2. Results shown as box-and-whisker plots with boxes displaying from the 25th to 75th percentiles, whiskers extending to the minimum and maximum values, and the center line indicates the median with all individual data points overlaid. Student's t-test p-values a: $p< 0.003$, b: $p< 1.6x10^{-4}$, c: $p<2.6x10^{-5}$. **d.** Representative Rubisco immunoblot. 100 ng TSP was loaded for each sample. **e.** Rubisco accumulation as a percent of TSP. n= 3 plants for all genotypes and growth trials. Results shown as data points for all individual values along with means (bold bar) and standard error of the mean. Two-tailed Student's t-test p-values are provided in Table AII.1.

In separate immunoblots, Rubisco content was measured in the field-grown plants as well as in the chamber-grown plants. All plants maintained Rubisco at roughly 50% of TSP and no statistical differences were observed between plant varieties in a given growth scenario (Figure 2.2d, e). Lack of alteration of Rubisco content is consistent with the biomass and photosynthetic measurements described earlier.

Growth conditions appear to influence how plants accommodate the extra burden of producing a large amount of foreign protein. In the optimal chamber conditions, both transformed lines maintained the same level of total protein and Rubisco (per gram tissue and per meter squared of leaf area, Figure 2.3**, Chamber**), whereas what we have categorized as "other" proteins was reduced roughly in proportion to the amount of recombinant cellulase that accumulated in the tissue. In contrast, the tissue of the transformed plants grown in the Year 1 field season increased total TSP production to accommodate Cel6A accumulation, leaving the fraction of "other" proteins unchanged relative to WT tobacco (Figure 2.3**, Field Year 1**). The difference in total protein between the transformed plants and WT was not as great in Year 2 plants largely because protein levels were higher in Year 2 WT tobacco. In Year 2 TetC-*cel6*A plants, higher levels of recombinant protein (compared to NPTII-*cel6*A) resulted in a marginally higher TSP accumulation coupled with a slightly reduced "other" protein fraction (Figure 2.3**, Field Year 2**). The plants with lower Cel6A accumulation (the NPTII-*cel6*A plants) maintained a protein profile very similar to WT tobacco (Figure 2.3**, Field Year 2**). We note that TetC-Cel6A accumulation represented as $g \cdot m^{-2}$ leaf area (or per gram tissue) in the Year 1 field season is statistically indistinguishable from that of the chamber-grown plants and only slightly lower in the Year 2 field season (Figure 2.S2).

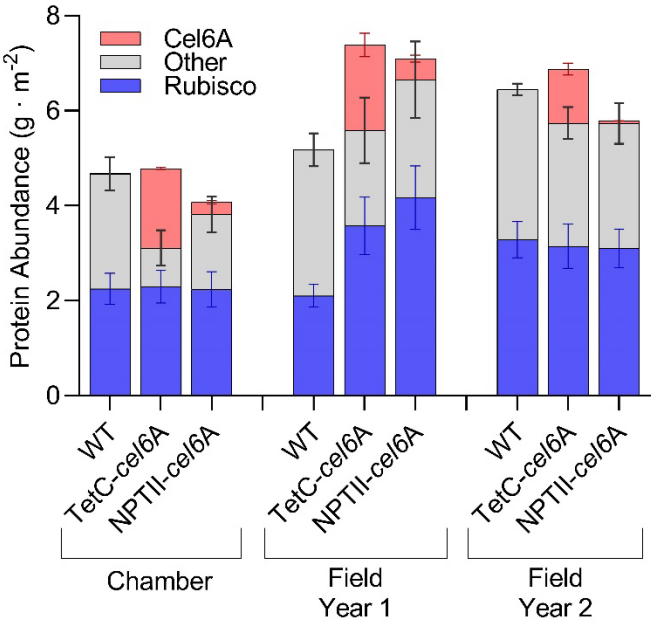


Figure 2.3: Comparison between trials of protein allocation between Cel6A, Rubisco, and all other endogenous soluble proteins. The full height of the stacked bars represents the total soluble protein accumulation normalized to leaf area. All results shown are the mean and standard error of the mean.

Discussion

Recombinant cellulase yields from full-sunlight chamber-grown TetC-*cel6*A were as high as 40% Cel6A of TSP with an average of 36%, placing them among the highest yielding plastid-engineered transformants reported [17, 25, 26]. NPTII-*cel6*A yields were, as expected, much more modest, supporting earlier work with these transformed lines [24]. Remarkably, Cel6A accumulation in our high-yielding line was not drastically reduced when grown in the field compared to the chamber.

Fundamental to our study was the question of the effect of the metabolic burden of producing large amounts of Cel6A while also growing in the more variable conditions of open field cultivation; therefore, we included the tobacco line, NPTII-*cel6*A, which was known to produce less Cel6A than line TetC-*cel6*A. We hypothesized that the lower Cel6A accumulation would afford NPTII-*cel6*A transformants a greater resilience to the unpredictable conditions of field cultivation. Remarkably, both TetC-*cel6*A and NPTII-*cel6*A plants were nearly indistinguishable from WT tobacco in the two consecutive field trials. Despite some minor reductions to photosynthetic rate, biomass accumulation was identical to WT for both transformed lines, indicating the Cel6A production did not have a meaningful impact on native metabolism. Furthermore, because both transformed lines exhibited the same degree of reduction in J_{max} and $V_{c,max}$ despite widely different levels of Cel6A accumulation, it is unlikely that the abundance of Cel6A is directly responsible for the decreases in photosynthesis. More likely, some other factor caused the observed reduction in photosynthesis; for example, the use of the native *psb*A untranslated regions (UTRs) in the transgene cassette might compete with the native 5'UTR in *psb*A transcripts for RNA-binding proteins, resulting in decreased *psb*A translation [27]. Competition for RNA binding proteins was observed when the *clp*P 5'UTR was

used to express the *neo* gene in chloroplast-transformed tobacco [28]. The robust maintenance of healthy growth in the *cel6*A lines differs from some other plastid-engineered tobacco lines that demonstrate chlorotic leaves and stunted growth even when grown in controlled-environment chambers [29]. One common cause of these poor phenotypes is often a reduction in some vital process, which can inhibit native metabolism depending on the severity of the reduction [17].

The response of plant resource allocation to constitutive foreign protein production is both a complex biological question as well as a critical component to determining the economic feasibility of these plants for high-value protein production. The chloroplast transformants of this study demonstrate distinct methods of adapting resource use to produce Cel6A, depending on their environmental conditions (Figure 2.4). One strategy, as exemplified by the chamber-grown cellulase transformants, is a reallocation of proteinaceous resources away from endogenous proteins to produce Cel6A (Figure 2.4**, Scenario 1**). Earlier studies with chloroplast-engineered transformants have measured significant reductions in endogenous proteins, including those related to carbon assimilation [30, 31]. The fact that these reductions did not result in a detectable negative phenotype suggests that these endogenous proteins may exist in excess in plants grown in carefully controlled environments. This appears to be the case with the Cel6A transformants grown in the chamber. Because there was no significant reduction in photosynthesis and no obvious difference in plant biomass, we can assume that no endogenous protein was reduced severely enough to impact native metabolism and that the majority of the reduction in endogenous proteins was from non-essential proteins. For instance, proteins and enzymes required for defense or stress tolerance would be dispensable in the controlled conditions of our chambers.

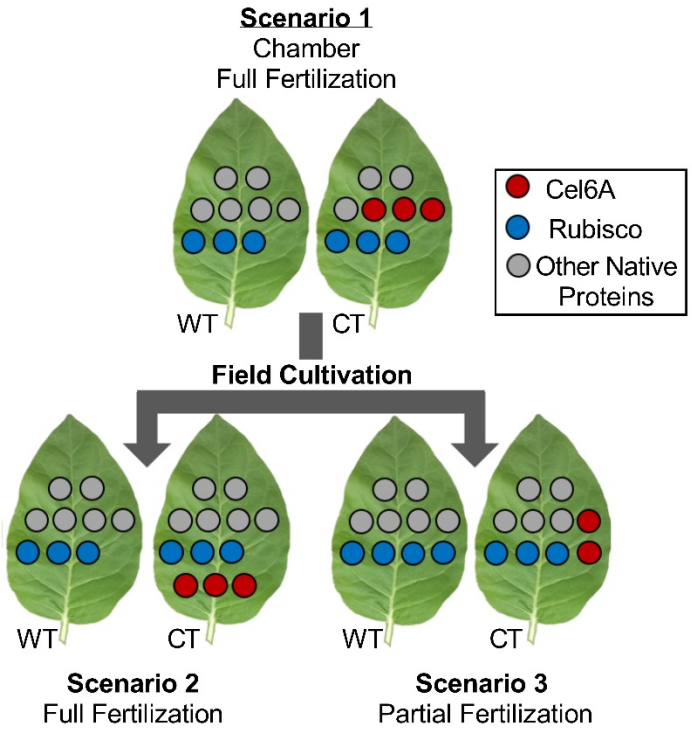


Figure 2.4: Hypothesized strategies for resource allocation and adaptation to foreign protein accumulation in transplastomic tobacco grown under different conditions. Wild-type (WT) and chloroplast-transformed (CT) tobacco abundance of Cel6A, Rubisco, and all other endogenous proteins are illustrated with red, blue, and gray dots, respectively. The total number of dots reflects total soluble leaf protein yield.

In contrast, the variability of environmental conditions during field cultivation likely necessitated a full or nearly full complement of endogenous proteins to maintain robust growth (Figure 2.4, **Scenarios 2** and **3**). In Year 1, transformed tobacco increased TSP synthesis to meet the demands of Cel6A synthesis while maintaining a full complement of native proteins (Figure 2.4, **Scenario 2),** whereas in Year 2 field transformants, reduced endogenous protein was accompanied by less Cel6A (Figure 2.4, **Scenario 3**). While this down-regulation of native proteins did not inhibit plant growth, enzyme product yield was reduced by 23% in TetC-*cel6*A and 86% in NPTII-*cel6*A from Field Year 2 compared to Field Year 1. Variation in plant response to field cultivation could be explained by differences in environmental conditions like insolation, precipitation, and fertilizer input (Figure 2.S1, Table 2.S1). The field was fertilized with potassium, phosphorous, and nitrogen prior to the Year 1 field season, but only nitrogen was replenished before the Year 2 season because residual potassium and phosphorous levels, as determined by soil tests, were deemed sufficient. Fertilizing a field on an as-needed basis is a standard, and economically practical farm management strategy, but the limitations exhibited by the Year 2 field transformants highlight the importance of carefully evaluating the unique nutrient requirements of chloroplast-engineered tobacco plants for commercial-scale production of a high value protein.

Even with the highly variable and unpredictable growth conditions, the TetC-*cel6*A transformants maintained robust growth and remarkable Cel6A yield. Thus, the TetC-*cel6*A plants present an ideal model to investigate the potential of plastid-engineered plants to be utilized as a production platform for high value proteins. Given standard tobacco harvest yields [32], TetC-Cel6A could be produced for US\$0.06 g^{-1} Cel6A (or US\$60 kg^{-1}), assuming a Cel6A yield of 20% of TSP (1.5 g Cel6A m^{-2}), as was the average for the Year 1 field plants (Table

2.S3). If, however, tobacco cultivation practices were optimized for maximizing protein rather than nicotine (e.g. by harvesting leaf tissue multiple times during a single growing season), harvest yields could be increased by more than 3.5-fold [33]. This increase in harvested biomass per acre would reduce Cel6A production costs to US$0.008 g^{-1} Cel6A (US$8 kg^{-1}), assuming protein yield is not severely reduced by multiple harvesting events. Currently, cellulases produced via traditional microbial bioreactors cost approximately US$10 kg^{-1} cellulase [34], making TetC-*cel6*A plants a viable alternative production method only after tobacco field cultivation has been optimized for high protein yield.

Downstream processing costs can account for as much as 95% of the final production cost to generate purified medical proteins from plant-based production systems [35] and we note that many high value, fully active medical proteins have been successfully synthesized in tobacco, like influenza antigens and antiretrovirals for HIV treatment [9, 36]. Thus, the final manufacturing cost of a pure protein produced from a plastid-engineered, field-grown tobacco plant is expected to be on the order of $0.16- $1.20 g^{-1}. For comparison, mammalian cell cultures utilizing Chinese hamster ovary cells have total production costs of $300-$3,000 g^{-1} recombinant protein while similar costs from transgenic goat's milk is $105 g^{-1} [3]. While it is generally unknown whether protein purification from plant chloroplasts will present particular challenges (or advantages)[37], the significantly lower production costs estimated here suggest there is great value in further exploration of field-grown chloroplast-engineered tobacco as a means to produce high value medical and industrial proteins.

Methods

Chamber experiments

Chloroplast-transformed tobacco cv. Samsun seeds expressing TetC-Cel6A and NPTII-Cel6A[24] (available upon request from corresponding author), were sown in a soil-less mix (LC1 Sunshine, Sun Gro Horticulture Inc, Vancouver, Canada) in 3.79 L pots. Two seeds were sown in each pot, and after germination, seedlings were thinned to 1 plant per pot. Plants were grown in two controlled-environment growth cabinets (PGC20, Conviron, Winnipeg, Canada) at 350 μmol m^{-2} s^{-1} PAR, relative humidity between 75 and 90%, 25/23 °C, and a 16/8 day-night cycle. Each week, plants were randomly reassigned to chambers and randomly rearranged within chambers. At 2 months old, the plants were harvested. Tissue was sampled from the youngest fully expanded leaf using a cork borer and immediately frozen in liquid nitrogen. Samples were shipped overnight on dry ice, and were stored at -80 °C.

Field experiments

The same seeds as described above were sown in a soil-less mix (LC1 Sunshine, Sun Gro Horticulture Inc, Vancouver, Canada) in polystyrene-foam floating-bed trays. Each cell of the tray had an opening at the bottom to allow movement of water. A 20-10-20 (N-P-K) fertilizer was applied to the hydroponic solution at a rate of 0.53 mg mL^{-1} (Table 2.S1**).** The trays were floated on water until plants were transplanted from the greenhouse to the field. The seedlings were topped to prevent shading and overcrowding at 3 and 4 weeks after germination.

Seedlings were transplanted to the field on June 23, 2014 for the Year 1 trial and July 8, 2015 for the Year 2 season. Prior to transplantation, the empty fields were fertilized with urea, phosphorous pentoxide, and potassium oxide in Year 1 and urea only in Year 2. Chemicals were

applied in the greenhouse and field to control fungus and pest growth (Table 2.S2). Plants were spaced in 15-inch grids in 8 plants by 8 plants plots in the Year 1 growing season and 6-by-6 plants in the Year 2 growing season. Supplemental irrigation was applied as needed during both field trials.

The outer rows were considered border rows and were not used for any measurements. Plants were harvested after 11 weeks of growth in Year 1 and 9 weeks in Year 2. Tissue was sampled from the youngest fully expanded leaf that received full sunlight using a cork borer and immediately frozen in liquid nitrogen. Samples were shipped overnight on dry ice and were stored at -80 °C. Aboveground portions of non-border plants were harvested and dried at 65°C in a drying oven before weighing.

Photosynthetic gas-exchange measurements

Photosynthetic gas exchange was measured using a portable gas exchange instrument (LI-6400-40 leaf chamber fluorometer; Li-Cor, Inc, Lincoln, Nebraska). Measurements were made on mature leaves in full sunlight. Temperature was controlled by setting cuvette block temperature of the instrument to 26°C. Humidity was manually controlled to achieve values between 65 to 80% relative humidity. Light levels were set to 2000 μmol m^{-2} s^{-1} photosynthetically active radiation. Carbon assimilation rate was measured at several CO_2 concentrations, starting at 400 μmol mol^{-1} and decreasing to 50 μmol mol^{-1}, then returned to 400 μmol mol^{-1} until Rubisco reactivated, then increasing to 1600 μmol mol^{-1}. $V_{c,max}$ and J_{max} were fit to the data [38], and values at 25 °C were calculated using temperature-response functions [39, 40].

Leaf protein extraction

Frozen wild-type, TetC-*cel6*A, and NPTII-*cel6*A, leaf samples were manually ground on ice using a micropestle for 10 minutes (n=5 plants for TetC-*cel6*A and NPTII-*cel6*A; n = 3 plants for wild type). A final volume of 500 µL of protein extraction buffer (20 mM Tris-HCl, 1% Triton X-100, 0.1% SDS, 1.5 mM PMSF, and 0.001% β-mercaptoethanol) was added to the ground tissue. The samples were vortexed thoroughly before being centrifuged for three minutes at 20,000 × *g* at 4 °C. The supernatant was extracted, aliquoted, and stored at -80 °C until use.

Total protein quantification

Total soluble protein from the tobacco leaf samples was quantified using a standard Bradford Assay. The BioRad Quick Start Bovine Serum Albumin Standard Set and 1X Dye Reagent (BioRad, Hercules, CA, USA) were used for all Bradford Assays. Tobacco samples were diluted with the extraction buffer described above to achieve a concentration within the linear range of the standard curve. Absorbance measurements were carried out on a BioTek Industries Synergy 4 plate reader (BioTek Industries, Winooski, VT, USA) at 595 nm.

Immunoblotting

Protein samples and standards were mixed in a 1:1 ratio with a 2X Laemmli Sample Buffer (BioRad, Hercules, CA, USA), electrophoresed through 12% polyacrylamide gels and then transferred to PVDF membranes (BioRad, Hercules, CA, USA). Purified Rubisco protein (Agrisera, Vännäs, Sweden) was used for standard curve quantitation. Purified Cel6A and the anti-Cel6A primary antibody [41, 42] were supplied by David Wilson (Cornell University, Ithaca, NY, USA). The anti-Rubisco antibody was generously donated by Martin Parry (University of Lancaster, UK) [43, 44]. The secondary antibody used for all blots was a

horseradish peroxidase-linked whole anti-rabbit IgG produced in donkey (GE Healthcare Life Sciences, Pittsburgh, PA, USA). All primary and secondary antibody treatments were diluted 1:20,000 in Antibody Signal Enhancer (Amresco, Solon, OH, USA). Samples from three plants comprised biological replicates that were analyzed for all Rubisco blots. Five biological replicates were analyzed for Cel6A in both genotypes in the Chamber and Field 2014, while three biological replicates were analyzed for Field 2015.

Membranes were developed using one of two methods. In the first method, membranes were treated with the ECL Western blotting reagents (Promega, Madison, WI, USA) for five minutes prior to exposure on CL-XPosure film (Thermoscientific, Waltham, MA, USA) and development in a film processor (Konica Minolta, Tokyo, Japan). The second method utilized a TMB stabilized substrate dye for horseradish peroxidase (Promega, Madison, WI, USA). Densitometry was performed using image analysis software (Integrated density in ImageJ [45] version 1.52a).

Dry weight analyses

Weight measurements were made on two separate samples of all genotypes before and after drying the tissue. Approximately 200 mg of fresh tobacco leaf tissue, consisting of 5 leaf punches of identical leaf area each, was dried in pre-weighed and pre-dried envelopes.

Statistical analysis

The chamber experiment was analyzed as a fully-randomized design with 5 replicate plants for each of the 3 lines, 15 plants in total. For gas exchange data, some curves could not be used to estimate J_{max} because stomatal closure prevented C_i from reaching high levels, so some genotypes had 4 replicates for J_{max} estimates.

The Year 1 field experiment was arranged in a fully-randomized design with 5 replicates for each of the 3 lines, 15 plots in total. A different field was used for the Year 2 season, and to control for spatial heterogeneity, the experiment was arranged in a randomized complete-block design with each of the 3 lines occurring once in each of 8 blocks, 24 plots total.

In chamber studies, observations were collected from individual plants, which were treated as experimental units. In field studies, plots were considered experimental units; if multiple observations were made within a plot, then values were averaged, and the mean of values for a single plot was used in further analyses. In no experiment were multiple observations made repeatedly through time.

For all analyses, means of values from genotypes (WT, TetC-*cel6*A, or NPTII-*cel6*A) within condition (Chamber, Field Year 1, or Field Year 2) were used. Differences between genotypes and conditions were evaluated using two-tailed, two-sample, Student's t-tests. Errors were considered normally, identically and independently distributed. Two-tailed Student's t-test p-values are provided in Table AII.1.

Supplementary Materials

Table 2.S1: Fertilizer application to field-grown tobacco

2014 Fertilizer application		
Date of application	**Fertilizer type**	**Application**
While in float trays	20-10-20 N-P-K	The solution in the float trays was maintained at an N concentration of 0.503 mg mL^{-1} by periodic testing and addition of a 20-10-20 fertilizer.
6/9/2014	P_2O_5	Applied and worked into soil at a rate of 11.2 g m^{-2}.
6/9/2014	K_2O	Applied and worked into soil at a rate of 22.4 g m^{-2}.
6/9/2014	Urea	Applied and worked into soil at a rate of 22.4 g m^{-2}.
6/19/2014		Transplanted to the field

2015 Fertilizer application		
Date of application	**Fertilizer type**	**Application**
While in float trays	20-10-20 N-P-K	The solution in the float trays was maintained at an N concentration of 0.503 mg $mL^{-1}$by periodic testing and addition of a 20-10-20 fertilizer.
6/30/2015	Urea	Applied and worked into soil at a rate of 22.4 g m^{-2}.
7/8/2015		Transplanted to the field

Table 2.S2: Pesticide application to field-grown tobacco

2014 Pesticide Applications		
Date of application	**Pesticide**	**Application**
6/4/2014	Terramaster 4EC	Float-tray solution was brought to a concentration of 0.078 mL L^{-1} Terramaster.
6/4/2014	Dithane	A 1.3 mL L^{-1} dithane solution was misted onto plant stalks.
6/8/2014	Dithane	A 1.3 mL L^{-1} dithane solution was misted onto plant stalks.
6/9/2014	Lorsban 15G	1.52 g m^{-2} Lorban was applied to the field and worked into the soil.
6/12/2014	Terramaster 4EC	Float-tray solution was brought to a concentration of 0.078 mL L^{-1} Terramaster.
6/15/2014	Dithane	A 1.3 mL L^{-1} dithane solution was misted onto plant stalks.
6/22/2014	Dithane	A 1.3 mL L^{-1} dithane solution was misted onto plant stalks.
6/23/2014	Platinum 75 SG	A 0.98 mL L^{-1} solution was used to water the bases of the plants.
7/15/2014	Dipel Pro DF	A 2.6 mL L^{-1} solution was sprayed on plant stalks.

2015 Pesticide Applications		
Date of application	**Pesticide**	**Application**
6/17/2015	Gnatrol	A 3.9 mL L^{-1} solution was sprayed onto the soil surface.
6/18/2015	Dithane	A 1.3 mL L^{-1} dithane solution was misted onto plant stalks.
6/22/2015	Terremaster 4EC	Float-tray solution was brought to a concentration of 0.078 mL L^{-1}Terramaster.
6/22/2015	Lorsban 15G	2.2 g m^{-2} Lorban was applied to the field and worked into the soil.
6/22/2015	Gnatrol	A 5.1 mL L^{-1}solution was sprayed onto the soil surface.
6/23/2015	Dithane	A 1.3 mL L^{-1} dithane solution was misted onto plant stalks.
6/30/2015	Lorsban 15G	2.2 g / m^{-2} Lorban was applied to the field and worked into the soil.
7/1/2015	Dithane	A 1.3 mL L^{-1} dithane solution was misted onto plant stalks.
7/10/2015	Platinum 75 SG	A 1.33 mL L^{-1} solution was used to water the bases of the plants.
7/21/2015	Dipel Pro DF	A 2.6 mL L^{-1} solution was sprayed on plant stalks.

Table 2.S3: Summary of cost analysis calculations and assumptions

Parameter	Value	
	Average Tobacco Yield	**High Tobacco Yield**
Tobacco biomass yield, dry weight (metric tons hectare^{-1})	2.7*	100†
Production cost of field-grown tobacco (US$ hectare^{-1})	6120[1]	4800[2]
Cel6A yield§ (g Cel6A g^{-1}dry tissue)	0.04	0.04
Cost of producing Cel6A (US$ g^{-1}Cel6A)	0.06	0.012

* fresh weight value from [3] converted to dry weight with approximate value of 10% dry matter

† Calculated from Albright USDA grant progress report on planting density

§ Average Cel6A yield for Field Year 1 TetC-*cel6*A tobacco

[1] Maksymowicz, B. and G. Palmer (1997). "Tobacco Management: Optimizing Profits." University of Kentucky Cooperative Extension(ARG 157).

[2] Scott, G. L. and J. F. Warren (2016). Tobacco production system, Google Patents.

[3] Foreman, L. (2006). Tobacco Production Costs and Returns in 2004, US Department of Agriculture, Economic Research Service.

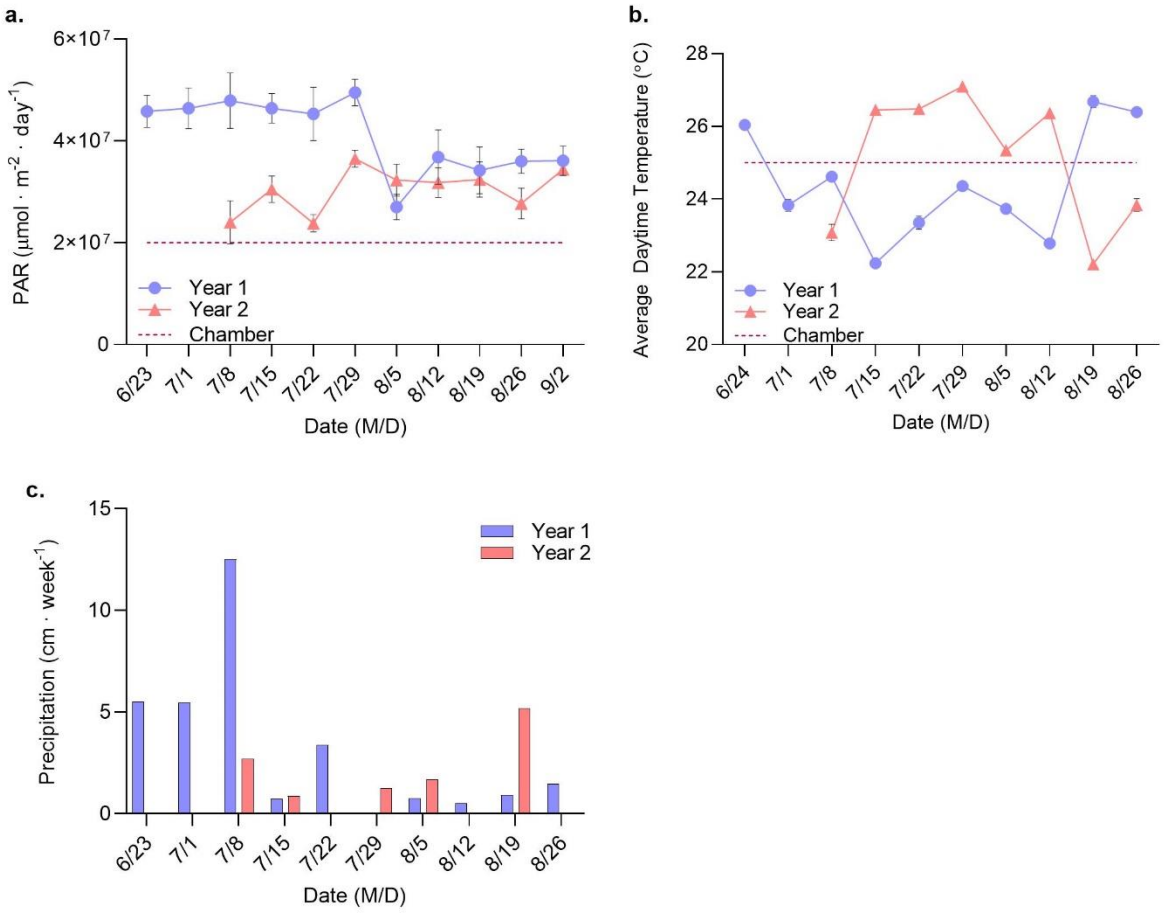


Figure 2.S1: Recorded insolation, temperature, and precipitation for both growth seasons. (a) Total insolation averaged for each week. n=7 independent samples for both field seasons. (b) Daily temperature averaged from 7:00am-7:00pm for each week. n=343 independent samples per week for Field Year 1 and n=511 independent samples per week for Field Year 2. (c) Total precipitation measured for each week. n=59 independent samples for both field seasons. Two-tailed Student's t-test p-values are provided in Table AII.1.

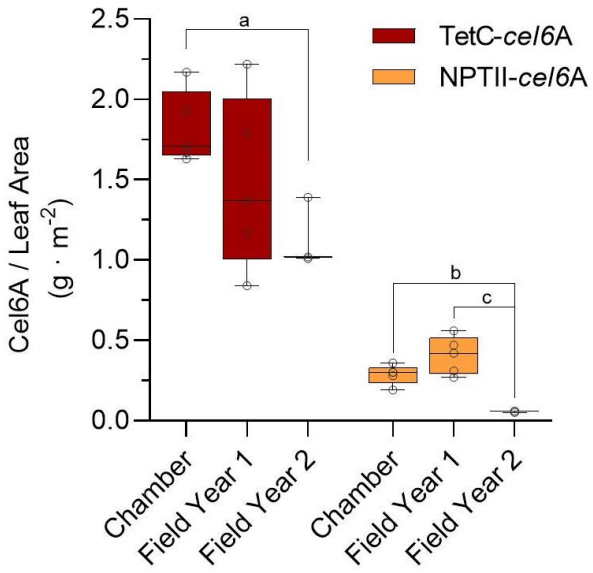


Figure 2.S2: TetC-Cel6A and NPTII-Cel6A accumulation normalized to leaf area. n= 5 plants for both genotypes in Chamber and Field Year 1 and n= 3 for both genotypes in Field Year 2. Results shown as box-and-whisker plots with boxes displaying from the 25th to 75 th percentiles and whiskers extending to the minimum and maximum values and the center line indicating the median with all individual data points are overlaid. Student's t-test p-values: a: $p<0.007$, b: $p<0.0009$, c: $p< 0.003$. Two-tailed Student's t-test p-values are provided in Table AII.1.

References

1. Palomares LA, Kuri-Breña F, Ramírez OT: Industrial recombinant protein production. In: *Biotechnology*. Edited by Doelle H, Rokem J, Berovic M, vol. 5. Abu Dhabi: EOLSS Publishers; 2002.
2. Twyman RM, Stoger E, Schillberg S, Christou P, Fischer R: Molecular farming in plants: host systems and expression technology. Trends Biotechnol 2003, 21(12):570-578.
3. Doran PM: Foreign protein production in plant tissue cultures. Curr Opin Biotechnol 2000, 11(2):199-204.
4. Xu J, Towler M, Weathers P: Platforms for Plant-Based Protein Production. In: *Bioprocessing of Plant In Vitro Systems.* Edited by Pavlov A, Bley T. Cham: Springer; 2016: 1-40.
5. Shaaltiel Y, Bartfeld D, Hashmueli S, Baum G, Brill-Almon E, Galili G, Dym O, Boldin-Adamsky SA, Silman I, Sussman JL: Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system. Plant Biotechnol J 2007, 5(5):579-590.
6. Zimran A, Brill-Almon E, Chertkoff R, Petakov M, Blanco-Favela F, Muñoz ET, Solorio-Meza SE, Amato D, Duran G, Giona F: Pivotal trial with plant cell–expressed recombinant glucocerebrosidase, taliglucerase alfa, a novel enzyme replacement therapy for Gaucher disease. Blood 2011, 118(22):5767-5773.
7. Yao J, Weng Y, Dickey A, Wang KY: Plants as factories for human pharmaceuticals: applications and challenges. Int J, Mol Sci 2015, 16(12):28549-28565.
8. Holtz BR, Berquist BR, Bennett LD, Kommineni VJ, Munigunti RK, White EL, Wilkerson DC, Wong KYI, Ly LH, Marcel S: Commercial-scale biotherapeutics

manufacturing facility for plant-made pharmaceuticals. Plant Biotechnol J 2015, 13(8):1180-1190.

9. Farsad A, Malekzadeh-Shafaroudi S, Zibaee S, Fotouhi F, Moshtaghi N: Transient Expression of HA1 Antigen of H5N1 Influenza Virus in Tobacco (*Nicotiana tabacum L.*) via Agro-infiltration. J Agr Sci Tech 2018, 19:439-451.

10. Fischer R, Vasilev N, Twyman RM, Schillberg S: High-value products from plants: the challenges of process optimization. Curr Opin Biotechnol 2015, 32:156-162.

11. Bock R: Engineering plastid genomes: Methods, tools, and applications in basic research and biotechnology. Annu Rev Plant Biol 2015, 66(1):211-241.

12. Svab Z, Maliga P: Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. PNAS 2007, 104(17):7003-7008.

13. Golczyk H, Greiner S, Wanner G, Weihe A, Bock R, Börner T, Herrmann RG: Chloroplast DNA in mature and senescing leaves: a reappraisal. The Plant Cell 2014, 26(3):847-854.

14. Shaver JM, Oldenburg DJ, Bendich AJ: Changes in chloroplast DNA during development in tobacco, *Medicago truncatula*, pea, and maize. Planta 2006, 224(1):72-82.

15. Leelavathi S, Reddy VS: Chloroplast expression of His-tagged GUS-fusions: a general strategy to overproduce and purify foreign proteins using transplastomic plants as bioreactors. Mol Breed 2003, 11(1):49-58.

16. Staub JM, Maliga P: Long regions of homologous DNA are incorporated into the tobacco plastid genome by transformation. The Plant Cell 1992, 4(1):39-45.

17. Oey M, Lohse M, Kreikemeyer B, Bock R: Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. Plant J 2009, 57(3):436-445.

18. Long S, Humphries S, Falkowski PG: Photoinhibition of photosynthesis in nature. Annu Rev Plant Biol 1994, 45(1):633-662.

19. Arlen PA, Falconer R, Cherukumilli S, Cole A, Cole AM, Oishi KK, Daniell H: Field production and functional evaluation of chloroplast-derived interferon-α2b. Plant Biotechnol J 2007, 5(4):511-525.

20. Bhat M: Cellulases and related enzymes in biotechnology. Biotechnol Adv 2000, 18(5):355-383.

21. Wilson DB: Studies of *Thermobifida fusca* plant cell wall degrading enzymes. Chem Rec 2004, 4(2):72-82.

22. Johnson E: Integrated enzyme production lowers the cost of cellulosic ethanol. Biofuels, Bioprod Biorefin 2016, 10(2):164-174.

23. Kuroda H, Maliga P: Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. Plant Physiol 2001, 125(1):430-436.

24. Gray BN, Ahner BA, Hanson MR: High-level bacterial cellulase accumulation in chloroplast-transformed tobacco mediated by downstream box fusions. Biotechnol Bioeng 2009, 102(4):1045-1054.

25. De Cosa B, Moar W, Lee S-B, Miller M, Daniell H: Overexpression of the *Bt cry2Aa2* operon in chloroplasts leads to formation of insecticidal crystals. Nat Biotechnol 2001, 19(1):71.

26. Maliga P: Plastid transformation in higher plants. Annu Rev Plant Biol 2004, 55:289-313.

27. Hirose T, Sugiura M: Cis-acting elements and trans-acting factors for accurate translation of chloroplast *psbA* mRNAs: development of an in vitro translation system from tobacco chloroplasts. EMBO J 1996, 15(7):1687-1695.

28. Kuroda H, Maliga P: Overexpression of the *clpP* 5′-untranslated region in a chimeric context causes a mutant phenotype, suggesting competition for a *clpP*-specific RNA maturation factor in tobacco chloroplasts. Plant Physiol 2002, 129(4):1600-1606.

29. Scotti N, Cardi T: Transgene-induced pleiotropic effects in transplastomic plants. Biotechnol Lett 2014, 36(2):229-239.

30. Bally J, Job C, Belghazi M, Job D: Metabolic adaptation in transplastomic plants massively accumulating recombinant proteins. PLoS One 2011, 6(9):e25289.

31. Bally J, Nadai M, Vitel M, Rolland A, Dumain R, Dubald M: Plant physiological adaptations to the massive foreign protein synthesis occurring in recombinant chloroplasts. Plant Physiol 2009, 150(3):1474-1481.

32. Foreman L: Tobacco production costs and returns in 2004: US Department of Agriculture, Economic Research Service; 2006.

33. Scott GL, Warren JF: Tobacco production system. In.: Google Patents; 2016.

34. Liu G, Zhang J, Bao J: Cost evaluation of cellulase enzyme for industrial-scale cellulosic ethanol production based on rigorous Aspen Plus modeling. Bioprocess Biosyst Eng 2016, 39(1):133-140.

35. Wilken LR, Nikolov ZL: Recovery and purification of plant-made recombinant proteins. Biotechnol Adv 2012, 30(2):419-433.

36. Rubio-Infante N, Govea-Alonso DO, Alpuche-Solís ÁG, García-Hernández AL, Soria-Guerra RE, Paz-Maldonado LMT, Ilhuicatzi-Alvarado D, Varona-Santos JT, Verdín-Terán L, Korban SS *et al*: A chloroplast-derived C4V3 polypeptide from the human immunodeficiency virus (HIV) is orally immunogenic in mice. Plant Mol Biol 2012, 78(4):337-349.

37. Stephan A, Hahn-Löbmann S, Rosche F, Buchholz M, Giritch A, Gleba Y: Simple purification of *Nicotiana benthamiana*-produced recombinant colicins: High-yield recovery of purified proteins with minimum alkaloid content supports the suitability of the host for manufacturing food additives. Int J, Mol Sci 2017, 19(1):95.

38. Farquhar Gv, von Caemmerer Sv, Berry J: A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. Planta 1980, 149(1):78-90.

39. Bernacchi C, Singsaas E, Pimentel C, Portis Jr A, Long S: Improved temperature response functions for models of Rubisco-limited photosynthesis. Plant, Cell Environ 2001, 24(2):253-259.

40. McMurtrie R, Wang YP: Mathematical models of the photosynthetic response of tree stands to rising CO_2 concentrations and temperatures. Plant, Cell Environ 1993, 16(1):1-13.

41. Caspi J, Barak Y, Haimovitz R, Gilary H, Irwin DC, Lamed R, Wilson DB, Bayer EA: *Thermobifida fusca* exoglucanase Cel6B is incompatible with the cellulosomal mode in contrast to endoglucanase Cel6A. Syst Synth Biol 2010, 4(3):193-201.

42. Caspi J, Irwin D, Lamed R, Shoham Y, Fierobe H-P, Wilson DB, Bayer EA: *Thermobifida fusca* family-6 cellulases as potential designer cellulosome components. Biocatal Biotransformation 2006, 24(1-2):3-12.

43. Lin MT, Hanson MR: Red algal Rubisco fails to accumulate in transplastomic tobacco expressing *Griffithsia monilis RbcL* and *RbcS* genes. Plant Direct 2018, 2(2):1-10.

44. Lin MT, Stone WD, Chaudhari V, Hanson MR: Enzyme kinetics of tobacco Rubisco expressed in *Escherichia coli* varies depending on the small subunit composition. Preprint at https://wwwbiorxivorg/content/101101/562223v1 2019:562223.

45. Schneider CA, Rasband WS, Eliceiri KW: NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012, 9(7):671.

CHAPTER 3

Mitigation of deleterious phenotypes in chloroplast-engineered plants accumulating high levels of foreign proteins[1,2]

Abstract

Medicine, agriculture, and industrial manufacturing all rely on high-quality proteins as major active components or process additives. Well-established crop cultivation techniques coupled with new advancements in genetic engineering may offer a cheaper and more versatile protein production platform than current methods. Chloroplast-engineered plants, like tobacco, have the potential to produce large quantities of high-value proteins, but often result in engineered plants with mutant phenotypes. For instance, we show that a *Nicotiana tabacum* line, TetC-*cel6*A, can produce an industrial cellulase at levels of up to 28% of total soluble protein (TSP) with a slight dwarf phenotype but no loss in biomass. In seedlings, the dwarf phenotype is recovered by exogenous application of gibberellic acid. We also demonstrate that, although these plants have a substantial degree of flexibility in protein metabolism, accumulating a foreign protein represents an added burden to the plants' metabolism that can make them more sensitive to limiting growth conditions such as low nitrogen. The biomass of nitrogen-limited TetC-*cel6*A plants was found to be as much as 40% lower than WT tobacco, although heterologous cellulase production was not greatly reduced compared to well-fertilized TetC-*cel6*A plants. Furthermore, cultivation at elevated carbon dioxide (1600 ppm CO_2) restored biomass accumulation in TetC-*cel6A* plants to that of WT, while also increasing total heterologous protein yield (mg Cel6A plant^{-1}) by 50-70%. Thus, we demonstrate that the alterations to protein expression triggered by growth at elevated CO_2 can help rebalance endogenous protein expression and/or increase foreign protein production in chloroplast-engineered tobacco.

[1] Full authors list: Schmidt, J., Richter, L., Condoluci, L., and Ahner, B.
[2] Submitted for review to *Biotechnology for biofuels*

Introduction

Proteins are used as enzyme additives in many industrial processes, as antibodies and medical peptides for pharmaceuticals, and as nutritional additives to food and animal feed (reviewed in [1]). These high-value proteins are mainly produced for such purposes using large scale cultures of bacterial, fungal, or mammalian cells. However, these systems are often expensive to establish, labor intensive to maintain, prone to contamination, and inflexible to changing market demands [2]. With well-established cultivation practices, a genetically engineered (GE) biomass crop, like tobacco, may represent a low-cost alternative ([3, 4] and Chapter 2).

While most GE plants used in commercial applications are generated by manipulation of the nuclear genome, transformation of the chloroplast genome for protein production has several key advantages. For instance, leaf cells contain a high copy number of chloroplast genomes (between 2,000 and 20,000 copies per leaf cell [5, 6]), the expression rates of many native chloroplast-encoded genes are high compared to those in the nucleus, and there are no gene silencing mechanisms in plastids (reviewed in [7]). As a result, chloroplast-engineered plants regularly achieve recombinant protein yields averaging 5-20% of total soluble protein (TSP) with exceptional plastid transformants reaching yields of 40-70% of TSP ([8-16] and Chapter 2).

While these high yields are promising for commercial applications, accumulating a non-native protein to that degree can have unforeseen impacts on the GE plant's metabolism. Several GE projects have yielded transplastomic plants with undesired mutant phenotypes like delayed growth, infertility, leaf chlorosis, and low biomass yield [9, 14-18]. These impacts are, however, highly variable and appear to depend on experiment-specific factors such as growth conditions, regulatory element selection and the particular recombinant protein.

Despite these challenges, some GE plants have been reported to accumulate foreign proteins to as much as 40% of TSP with little or no phenotypic changes. In fact, one group argued that some abundant native proteins, particularly Rubisco, serve as a dynamic N storage by accumulating to levels greater than is biologically required in wildtype (WT) tobacco; the authors posited that GE plants can reduce the abundance of these native proteins while synthesizing foreign proteins without interrupting native metabolism [8, 19]. In recent work, we showed that the response of plant protein metabolism to foreign protein synthesis is largely dependent on environmental factors. When grown in high-light growth chambers GE tobacco plants accumulating a recombinant cellulase to as much as 38% TSP, while phenotypically indistinguishable from WT plants, redistributed a substantial portion of protein resources away from some native proteins though not Rubisco (Chapter 2). However, when these same plants were grown in open fields, they increased total TSP production to effectively maintain constant levels of native proteins while producing exogenous cellulase at 20% of TSP on average (Chapter 2).

Deleterious mutant phenotypes can develop when foreign protein accumulation depletes native proteins beyond minimum levels required for endogenous metabolism. In these cases, mutant phenotypes may be alleviated by altering environmental conditions to reduce the need for specific endogenous proteins. For example, elevated CO_2 increases biomass while reducing photorespiration and the demand for photosynthetic machinery, particularly Rubisco, in C3 plants (reviewed in [20, 21]). We, therefore, hypothesize that in situations when synthesis of the recombinant protein strains the host plant's metabolism, elevated CO_2 may enable recovery of WT growth.

Mutant phenotypes can also be caused by the inclusion of features in the GE cassette that interfere with native gene expression. Regulatory elements native to the plastid genome of the host species or a closely related species are often used to mediate expression of transgenes because of their compatibility with the host's gene expression machinery [8, 15, 22-24]. However, the use of these DNA elements can result in potentially deleterious competition for transcriptional or translational factors with endogenous DNA elements within the engineered chloroplasts. For example, the use of the native tobacco *clpP* 5'untranslated region (UTR) for recombinant NPTII expression caused a pigment-deficient phenotype and a reduction in ClpP1 protease due to competition between the native and transgenic *clpP* 5'UTRs for a specific mRNA maturation factor [25]. Translational regulation by mRNA binding proteins is common in the chloroplast and not completely understood [26]; thus, constitutive expression of transgenic RNA sequences may have unforeseen consequences on endogenous protein accumulation. In addition, the specific codon usage in the transgenic cassette and insertion site choice in the plastid genome can also interfere with native protein expression [27, 28].

Finally, in some cases, the enzymatic activity of the recombinant protein may directly interfere with the host plant's metabolism. For example, Petersen and Bock suggested that the enzymatic activity of recombinant cell wall-degrading enzymes was responsible for the chlorotic leaves and stunted growth observed in their transformed tobacco [15]. Another group attributed alterations to auxin, gibberellic acid, and cytokinin levels to a chemical interaction of their exogenous β-glucosidase with hormone conjugates [29].

The chloroplast-engineered *N. tabacum* line used in this study, TetC-*cel6*A, expresses the endoglucanase Cel6A from the soil bacteria *Thermobifida fusca* [11]. Endoglucanases, along

with other cellulases, are used in cellulosic ethanol synthesis from cellulose feedstocks and are also needed for the production of some textiles, paper, detergent, beverages, and animal feeds (reviewed in [30]). Here, we measured Cel6A abundance in plants grown from germination to maturity to explore the underlying causes of transgenic phenotypic deviations from WT tobacco. We also manipulated atmospheric CO_2 and ammonium nitrate fertilizer application to alter protein resource allocation and study its impact on plant growth and Cel6A yield. Here, we aim to enhance the commercial applicability of GE plants for large-scale production of valuable proteins by using TetC-*cel6*A tobacco as a model to understand and mitigate mutant phenotypes caused by high recombinant protein synthesis.

Results

TetC-*cel6*A tobacco exhibited only a minor dwarf phenotype despite substantial accumulation of recombinant cellulase

The transplastomic and WT tobacco plants were grown in well-fertilized soil pots in growth chambers to document phenotypic differences and to measure recombinant cellulase accumulation as the plant tissue developed over time. TetC-*cel6*A tobacco plants were nearly indistinguishable from WT and produced the same amount of biomass (wet weight) and dry matter at the measured time points (Figure 3.1a, b, Figure 3.S1a). Leaf area was also not different between the two genotypes (Figure 3.S1b). The only altered phenotype in the transgenic plants under these growth conditions was a reduction in stem height of as much as 40% in TetC-*cel6*A plants compared to WT plants at 8 weeks and later (Figure 3.1c).

Cel6A yield as a percent of total soluble protein (TSP) increased over time in composite tissue samples from 9% Cel6A of TSP in 2-week-old seedlings to 28% in green leaves of 12-

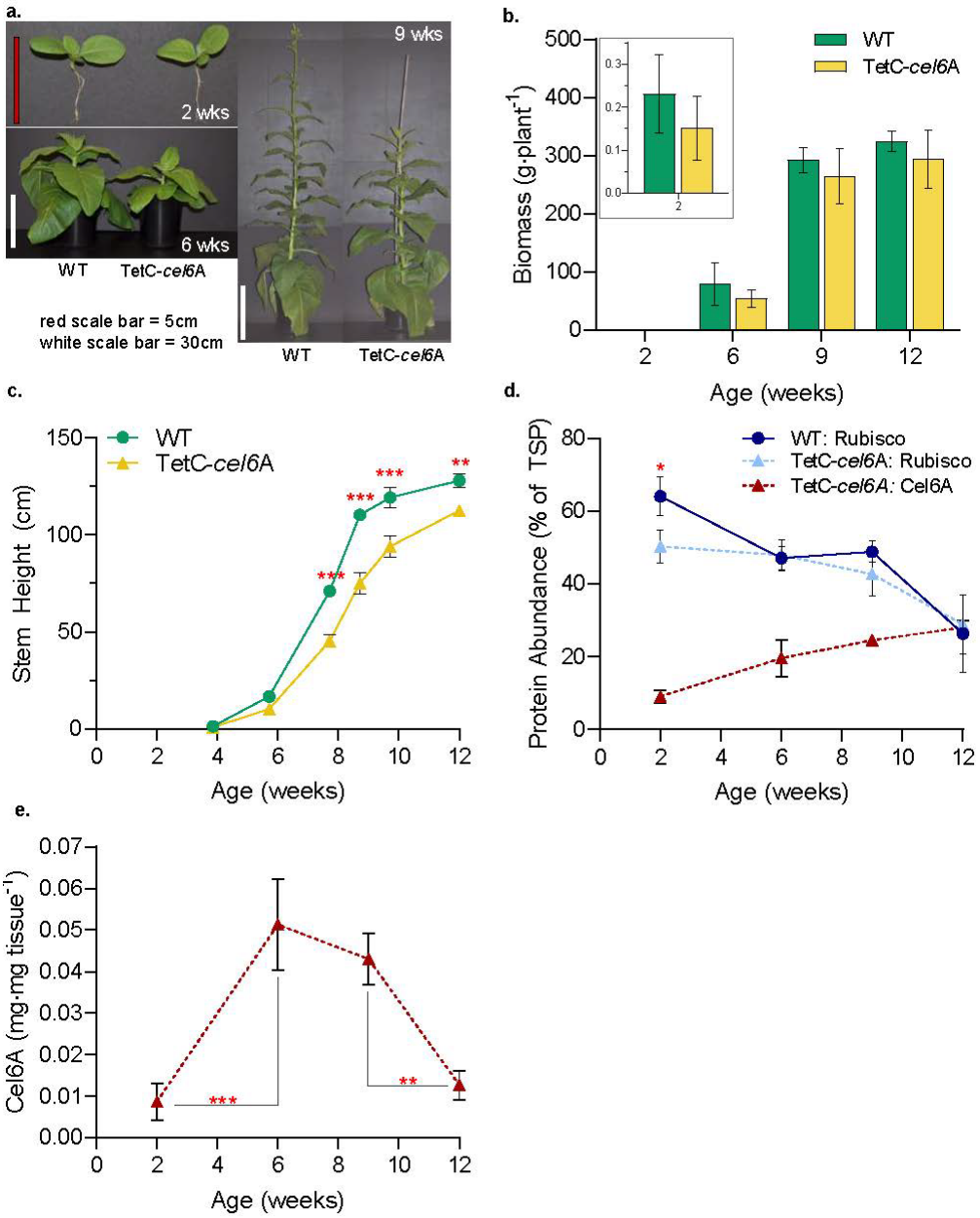


Figure 3.1: Growth assessment of TetC-*cel6*A tobacco plants from two to twelve weeks old. **a.** Representative photographs of WT and transgenic plants at 2, 6, and 9 weeks of age. **b.** Biomass measurements (n = 9 for 2 week harvest, n = 4 for 6 and 9 week harvest, and n = 3 for 12 week harvest). **c.** Measurements of stem elongation from soil to shoot apex (n = 4 for all data points except 12-week where n = 3). **d.** Cel6A and Rubisco leaf accumulation as a percent of TSP (n = 3). **e.** Cel6A yield normalized to mg Cel6A mg-1 dry leaf tissue (n = 3). Bar heights and data points correspond to the mean and error bars reflect the standard error of the means. Two-tailed t-test p-values for comparisons between genotypes and plant age are labeled in red asterisks as follows, $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***). A full summary of p-values is provided in Table AII.2.

week-old plants (Figure 3.1d), consistent with previous reports ([11] and Chapter 2). In contrast, Rubisco content decreased over time from approximately 50% of TSP in seedlings to 30% in 12-week-old plant tissue, with no difference noted between transgenic plants and WT except in 2-week-old TetC-*cel6*A seedlings which had significantly less Rubisco than WT (Figure 3.1d). The transgenic plants maintained similar total protein per plant as WT plants and exhibited the same pattern of reduced total protein due to senescence at 12 weeks (Figure 3.S2a). Cel6A concentration (normalized to tissue biomass) varied as a function of plant age, increasing by a factor of five from 2 to 6 weeks, holding steady for at least three weeks (6 to 9 weeks) and then decreasing at 12 weeks (Figure 3.1e); however, nine-week-old TetC-*cel6*A plants yielded the most Cel6A as a combined result of plant size and age-dependent protein levels (Figure 3.S2b). Consistent with previous work described in Chapter 2, chamber-grown TetC-*cel6*A plants reduced "other" proteins relative to WT, although this reduction was only statistically significant in 9-week-old plants because of replicate variability in protein measurements at other time points (Figure 3.S2c, d).

Exogenous gibberellic acid application alleviates the delayed germination and dwarf phenotype in TetC-*cel6*A tobacco

In addition to the dwarf stem phenotype noted above, we also previously observed a small delay in germination of the TetC-*cel6*A seeds. Combined, those two phenotypes led us to hypothesize that the presence of Cel6A in the GE plant tissue altered gibberellic acid (GA) metabolism because of GA's role in germination timing and internode elongation in higher plants [31, 32]. Without GA application, TetC-*cel6*A seeds took on average 6.3 (S.E. ± 0.1) days to germinate on agar plates, a full day longer than WT seeds (Figure 3.S3) but with the same profile in the

germination rate (Figure 3.2a). The addition of 1 μM GA to the plates eliminated the delay on average (5.5 days ± 0.14, Figure 3.S3), but the pattern of germination was not the same; 40% of theTetC-*cel6*A seeds germinated a day earlier than most WT seeds, but a subset of the transgenic seeds remained delayed (Figure 3.2a). WT seeds treated with and without 1 μM GA germinated in the same pattern as controls (Figure 3.2a) and with the same average time to germination (5.2 days [S.E. ±0.07] and 5.3 days [S.E. ±0.07] after seeding, respectively, Figure 3.S3).

The stem length of GA-treated and untreated seedlings was measured over time after germination. Untreated TetC-*cel6*A seedlings began to differentiate from WT at 11 days post-germination, with 15% shorter stems (Figure 3.2b, c). This difference in stem elongation increased with time, resulting in transgenic plants that were 20% shorter than WT at 17 days post-germination (p-value < 0.04; Figure 3.2b, c). In contrast, when provided with 1 μM GA, the stem length of both TetC-*cel6*A and WT seedlings was indistinguishable (Figure 3.2b, d) and more than three times longer than untreated shoots at the end of the experiment (Figure 3.2c, d). Thus, in addition to improving germination time, exogenous application of GA also abolished the dwarf stem phenotype observed in untreated TetC-*cel6*A plants.

Elevated CO_2 improves growth of TetC-*cel6*A plants when grown in nutrient-limiting conditions

We hypothesized that the burden of foreign protein synthesis may have a greater impact on plant phenotype under nutrient-limiting conditions. TetC-*cel6*A and WT tobacco were grown on an inert vermiculite media and fertilized with 1 mM, 4 mM, and 8 mM ammonium nitrate under ambient and elevated CO_2. At ambient CO_2, WT biomass increased significantly with each increase in ammonium nitrate as expected, whereas the biomass of TetC-*cel6*A tobacco

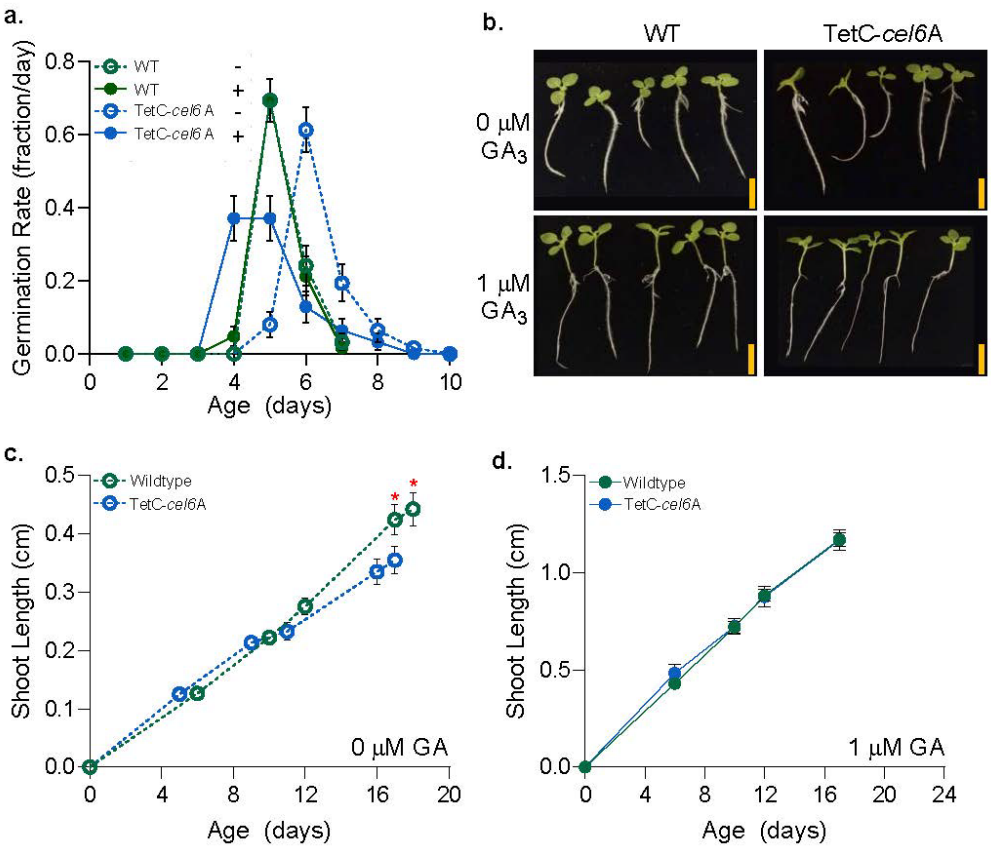


Figure 3.2: Effect of gibberellic acid on germination and seedling shoot elongation. **a.** Germination rate represented as the fraction of the total number of seeds that germinated per day (n = 62 independent plants per genotype and treatment). **b.** Representative photographs of seedlings approximately 2 weeks post-germination. **c.** Shoot elongation of untreated seedlings (n = 16 independent biological replicates). **d.** Shoot elongation of seedlings treated with 1 µM GA (n = 16 independent biological replicates). Data points represent the mean with error bars corresponding to the standard error of the means. Red asterisks (*) indicate two-tailed t-test *p*-values < 0.05. For simplicity not all significant comparisons are depicted on the graphs. A full summary of p-values is provided in Table AII.4.

increased from 1 mM to 4 mM but no further increase was observed at 8 mM (Figure 3.3a, b). Furthermore, TetC-*cel6*A tobacco plants were smaller relative to WT at all ammonium nitrate concentrations (Figure 3.3a, b), but in contrast to our expectations, biomass reductions were correlated with greater nitrogen (N) availability: transgenic plants receiving the most ammonium nitrate were the most reduced in biomass relative to WT (41% less at 8 mM; Figure 3.3a, b). This result suggests that there may be other stresses or limitations imposed by growth on the vermiculite medium contributing to the low biomass phenotype of TetC-*cel6*A.

The stunted phenotype observed in the transgenic plants was eliminated by elevating the atmospheric CO_2 concentration to 1600 ppm. At this higher level, the fresh weight per plant of the transgenic plants increased to match WT biomass levels at every ammonium nitrate concentration (Figure 3.3a, b) though the dry weight of the TetC-*cel6*A tobacco still lagged those of WT plants (Figure 3.S4a, b). This differential result reveals that some portion of the fresh weight biomass recovery caused by elevated CO_2 was due to an increase in tissue water content (Figure 3.S4 a, b). The biomass and dry weight of WT tobacco exhibited very little response to elevated CO_2; a statistically significant increase was only noted in WT plants fertilized with 4 mM ammonium nitrate (Figure 3.S4 a).

While increasing ammonium nitrate treatment increased total N accumulated per plant in both plant genotypes, at ambient CO_2, transgenic plants accumulated more N than WT at both higher N treatments (Figure 3.3c). This difference was abolished at the higher CO_2 level (Figure 3.3c). Consistent with this finding, C:N ratios were significantly lower in TetC-*cel6A* plants compared to WT at ambient CO_2 concentration, likely reflecting a combination of reduced carbon fixation and the comparatively higher levels of N assimilation (Figure 3.3d). At elevated

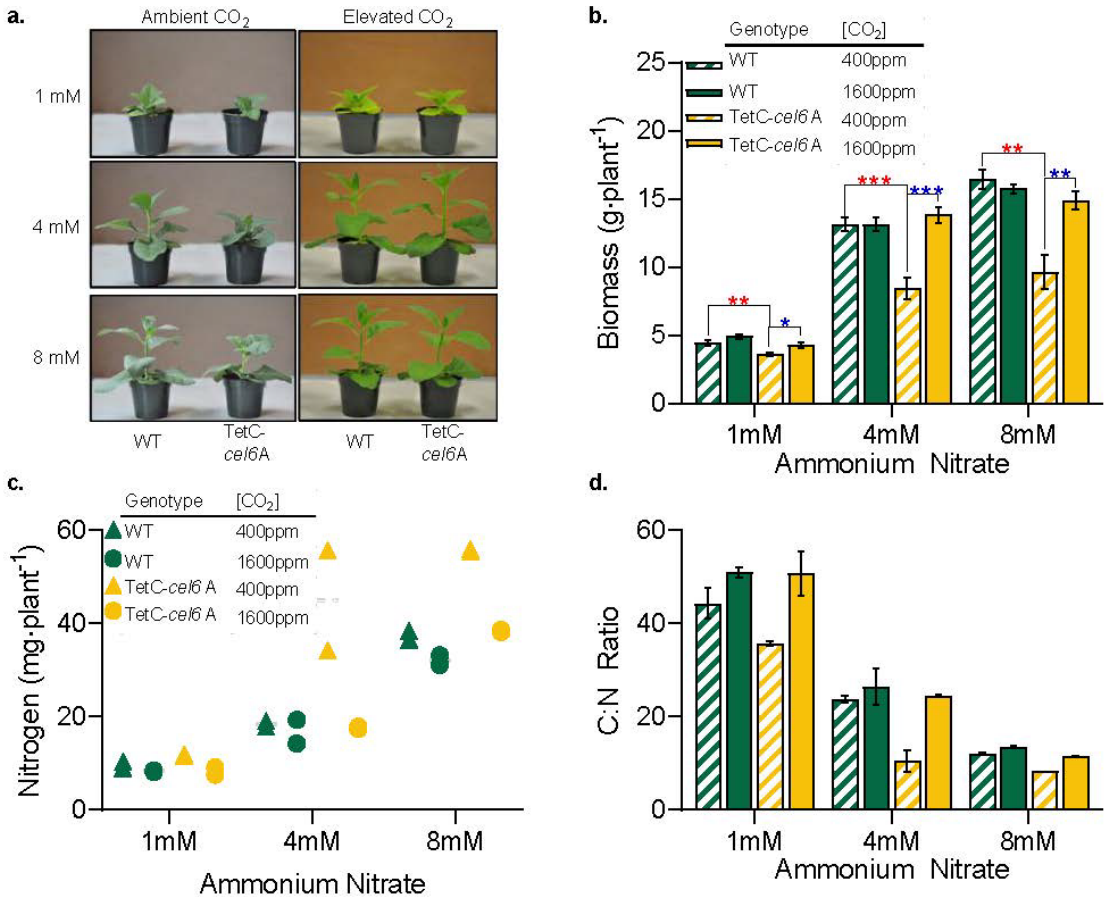


Figure 3.3: Effect of ammonium nitrate and carbon dioxide on growth of transgenic tobacco. **a.** Representative photos of six-week-old plants. **b.** Biomass accumulation (n = 6 independent biological replicates). **c.** Total leaf nitrogen content (n = 2 independent biological replicates). **d.** Leaf carbon to nitrogen ratio (n = 2 independent biological replicates). Bar heights and data points correspond to the mean and error bars reflect the standard error of the means. Two-tailed t-test p-values for comparisons between genotypes (red) and between CO_2 treatment (blue) were labeled with as follows, $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***). A full summary of p-values is provided in Table AII.5.

CO_2, C:N ratio in TetC-*cel6A* plants are identical to those of WT plants (Figure 3.3d), which we hypothesize is due to a recovery of native plant carbon metabolism in the transgenic plants.

Elevated CO_2 enables reallocation of protein resources, increasing recombinant cellulase production and bolstering accumulation of native proteins

Both nitrogen availability and CO_2 treatment had a significant impact on TetC-*cel6*A accumulation. Under ambient CO_2, the TetC-*cel6*A tobacco treated with both 1 mM and 8 mM ammonium nitrate yielded 10% Cel6A of TSP, while plants treated with 4 mM ammonium nitrate accumulated Cel6A to 20% of TSP (Figure 3.1d, Figure 3.4a). Increased CO_2 led to higher Cel6A yields at all ammonium nitrate levels. The highest concentration was measured in plants treated with 4 mM ammonium nitrate, averaging 25% Cel6A of TSP (Figure 3.4a), which exceeded the accumulation measured in similarly aged well-fertilized soil-grown plants at ambient CO_2 (Figure 3.1d). Furthermore, total leaf TSP (per plant) was similar between genotypes and generally unresponsive to CO_2 treatment (Figure 3.4b), except at 4 mM ammonium nitrate where elevated CO_2 led to 25% percent more whole plant protein. The slightly higher TSP in the 8 mM ammonium nitrate TetC-*cel6*A plants resulted in similar absolute levels of Cel6A ($mg \cdot plant^{-1}$) as those treated with 4 mM ammonium nitrate (Figure 3.S4c).

We also measured Rubisco abundance (% TSP) in all ammonium nitrate treatments with ambient and elevated CO_2. When grown in ambient CO_2, transformed tobacco and WT tissues maintained similar Rubisco levels at 40-50% TSP (Figure 3.4c). At elevated CO_2, plants of both genotypes reduced Rubisco abundance to 19-30% of TSP in WT and 17-20% in TetC-*cel6*A (Figure 3.4c).

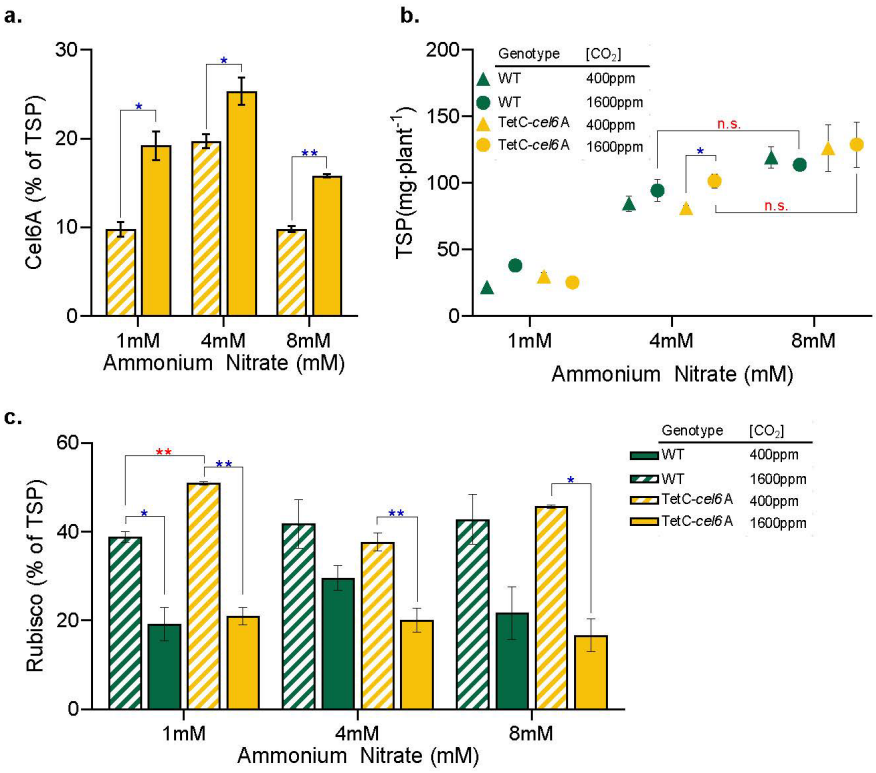


Figure 3.4: Alteration in recombinant cellulase and Rubisco abundance with varied ammonium nitrate and carbon dioxide levels. **a.** TSP accumulation. **b.** Cel6A accumulation measured as a percent of TSP. **c.** Rubisco accumulation measured as a percent of TSP. Bar heights and data points correspond to the mean and error bars reflect the standard error of the means (n = 3 independent biological replicates). Two-tailed t-test p-values for comparisons between genotypes (red) and between CO_2 treatment (blue) were labeled as follows, $p < 0.05$ (*) and $p < 0.01$ (**). A full summary of p-values is provided in Table AII.5.

Production of Cel6A without a corresponding increase in whole plant TSP indicates that some subset of endogenous leaf soluble proteins must be down-regulated to accommodate heterologous cellulase synthesis in TetC-*cel6*A tobacco (Figure 3.5, "Other"). Overall, at ambient CO_2 TetC-*cel6*A tobacco averaged 25% less "Other" proteins than WT tobacco. Meanwhile, at elevated CO_2, it is clear that reduced Rubisco levels leads to higher Cel6A accumulation, but transformants also increased their pool of "Other" proteins, suggesting a recovery of necessary native protein expression in the TetC-*cel6*A plants that enabled the return to WT growth.

Discussion

Successful use of plastid engineering for commercial production of proteins requires the generation of plants that have both high recombinant protein accumulation as well as reliable growth. However, chloroplast transformation sometimes unpredictably results in plants with mutant phenotypes like chlorotic leaves, stunted growth, and reduced fertility, even when grown in the well-managed conditions of growth chambers and greenhouses [9, 14-18]. These mutant phenotypes can be caused by either broadly overwhelming the host plant's protein metabolism or by more specific perturbations driven by particular recombinant proteins and/or transgenic regulatory elements.

We and others have shown that chloroplast-engineered tobacco has a sizable capacity to reallocate resources from endogenous proteins with no impact on plant growth; however, at some point, reduced levels of native proteins will compromise physiological processes. For example, Oey and colleagues argued that recombinant PlyGBS production to 70% of TSP exhausted native protein metabolism and severely stunted plant growth [14]. Recent work with TetC-*cel6*A tobacco, along with work from several other groups, have demonstrated that plastid engineered

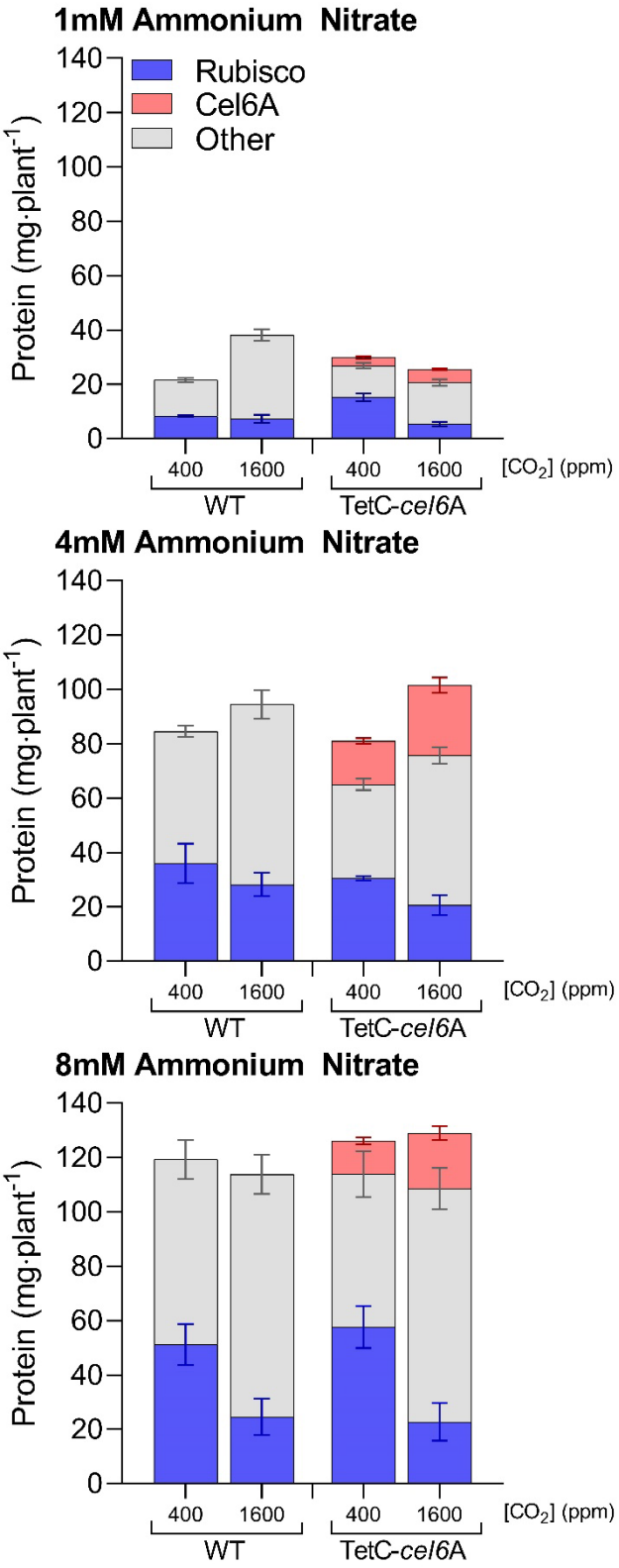


Figure 3.5: Variation in protein resource allocation dependent on ammonium nitrate and carbon dioxide content. Leaf TSP divided into Rubisco, Cel6A, and all "Other" endogenous soluble proteins for 1 mM (**a**), 4 mM (**b**), and 8 mM (**c**) ammonium nitrate.

plants can accumulate recombinant proteins up to 40% of TSP without deleterious phenotypes ([8, 9] and Chapter 2). This buffering capacity to accommodate foreign protein synthesis is highly dependent on growth conditions, as demonstrated by the reduced biomass relative to WT and lower Cel6A yields with lower N input for the TetC-*cel6*A tobacco.

Additionally, in this nitrogen limitation trial, we had hypothesized that Cel6A accumulation would be directly correlated with ammonium nitrate input. However, while total protein increased with increasing N, plants fertilized with 4 mM ammonium nitrate accumulated twice as much Cel6A, averaging 20% of TSP, as the plants in the 1 mM and 8 mM treatments. In contrast, one previous study found that N fertilization varying from 0 mM to 20 mM ammonium nitrate did not impact recombinant protein accumulation as a percent of TSP [8].

We propose that heterologous protein accumulation in response to changes in resource availability may be partially dependent on the foreign protein's resistance to degradation by the plastid's native protein degradation machinery. Degradation of heterologous protein appears to be highly variable and dependent on the properties of the foreign protein of interest [8, 18, 33, 34]. Nitrogen limitation leads to increased protein turnover and/or altered expression of specific proteases in plants to make amino acids available for new protein synthesis [35-37]. Cel6A's continued presence in 12-week-old plant tissue after the onset of sencenence suggests that it may be resistant to degradation and therefore more stable than native plant proteins. It is therefore possible that the high Cel6A yield noted at the intermediate N treatment results from a metabolic "sweet spot" where there is differential recycling of native proteins relative to Cel6A with less protein turnover at the highest N level and more tightly regulated protein synthesis at the lowest

N treatment. More experimentation would be required to confirm and then subsequently optimize nutrient application for enzyme production.

Similar to our N experiments, we hypothesized that increased CO_2 availability might increase foreign protein yield in our transgenic plants by increasing biomass and/or by altering protein allocation in the plant tissue [38]. Well-fertilized C3 plants grown in elevated CO_2 often exhibit increased carbon fixation and reduced total Rubisco abundance [20, 21, 39]. Both of these metabolic changes should enhance foreign protein production. Indeed, the increase in CO_2 concentration alleviated the stunted biomass phenotype in N-limited TetC-*cel6*A plants, and we observed an increase in Cel6A and "Other" native proteins concurrent with a significant decrease in Rubisco content. Thus, cultivating GE plants in high CO_2 may be a viable method of alleviating deleterious phenotypes caused by depletion of plant protein resources. While we quadrupled atmospheric CO_2 to 1600 ppm, gains in photosynthesis seem to taper off around 1000 ppm CO_2 at typical growing temperatures [40]. Further studies are needed to determine whether the synthesis of other proteins is similarly influenced and how much CO_2 enhancement is needed to optimize foreign protein accumulation.

In some cases, altered mutant phenotypes are caused not by a depletion of protein resources but by the enzymatic activity of the recombinant protein interfering with host plant metabolism. In fact, we propose that Cel6A's endoglucanase activity may be interfering with GA metabolism, leading to the minor mutant phenotypes observed in well-fertilized TetC-*cel6*A tobacco. The dwarf phenotype of our well-fertilized transgenic tobacco is likely due to a truncation of stem internode elongation rather than to reduced carbon metabolism because dry matter, biomass, and leaf area were all similar to WT. This truncation, coupled with delayed

germination, suggests that GA levels are diminished in our transgenic plants [31, 32]. Furthermore, increased water storage, reflected by elevated fresh weight to dry weight ratios, in our transgenic plants compared to WT is also indicative of decreased GA [41].

Since a portion of GA synthesis occurs in the chloroplast, it is possible that Cel6A is enzymatically interfering with GA intermediates, although a mechanism for this interaction is not clear. However, because of their ability to hydrolyze ether bonds of carbohydrates, recombinant cellulases have been reported numerous times to interfere with native host plant metabolism, particularly in the carbohydrate-rich chloroplast [42, 43]. As noted earlier, one study suggested that their recombinant beta-glucosidase activated hormone conjugates that increased plant biomass [29], while another study attributed low biomass and leaf chlorosis to the carbohydrate-binding affinity of their recombinant cellulases [15]. Alternatively, reductions in GA synthesis enzymes associated with the reallocation of protein resources to Cel6A could reduce synthesis of bioactive GA.

It is also possible for specific elements of the transgenic cassette to generate mutant phenotypes independent of protein synthesis by interfering with endogenous RNA-binding proteins. For example, the use of an endogenous gene's ribosome binding site (RBS) may alter native gene expression, as was suggested to explain the substantial decrease in Rubisco in plants expressing GFP and HPPD under control of the native *rbcL* RBS [8]. If ribosome binding at this RBS limits RbcL translation, then competition between the native and transgenic RBS could impede RbcL synthesis. Furthermore, the HPPD and GFP expression cassettes were both targeted for insertion adjacent to the native *rbcL,* which also may have impacted native expression of *rbcL.*

There are, however, other examples of high-accumulating transformants utilizing the *rbcL* leader sequence that do not exhibit reduced Rubisco [9]. This led us to consider whether downregulation of Rubisco in the GFP tobacco line may be due to codon use. If the codon use in the heterologous gene is very similar to RbcL, its expression may result in competition for charged tRNAs. Coding sequences utilizing many rare codons are often poorly expressed [44]; therefore, it is general practice to optimize codon use in heterologous genes to match the host system's native codon bias, as was done for GFP [8]. However, because the plastid is derived from bacteria, prokaryotic genes are often used directly with no modification.

To illustrate this point, we compared the codon use of several heterologous recombinant genes that were not codon optimized and the codon-optimized GFP to that of *N. tabacum's* native *rbcL* (Figure 3.6). The unoptimized genes include, our own TetC-*cel6*A [11] and NPTII-*bglC* [12], as well as another group's recombinant xylanase and beta-glucosidase, CelB [9]. The GE plants containing these transgenes all achieved foreign protein yields of at least 30% of TSP, except NPTII-BglC which accumulated to approximately 15% of TSP, and none of the engineered lines exhibited any deleterious mutant phenotypes.

Unsurprisingly, the codon use of codon-optimized GFP is far more similar to RbcL than the genes that were not optimized. Rubisco content, quantified by immunoblots or estimated from SDS-PAGE images included in the original papers, was most substantially reduced in the GFP-expressing tobacco, moderately reduced in HPPD-expressing plants, and not reduced at all in tobacco expressing TetC-Cel6A, NPTII-BglC, Xylanase, or CelB. HPPD is a unique case because tobacco possesses an endogenous copy of HPPD, which may alter plant response to foreign HPPD expression. Of the subset of recombinant proteins that are truly foreign to

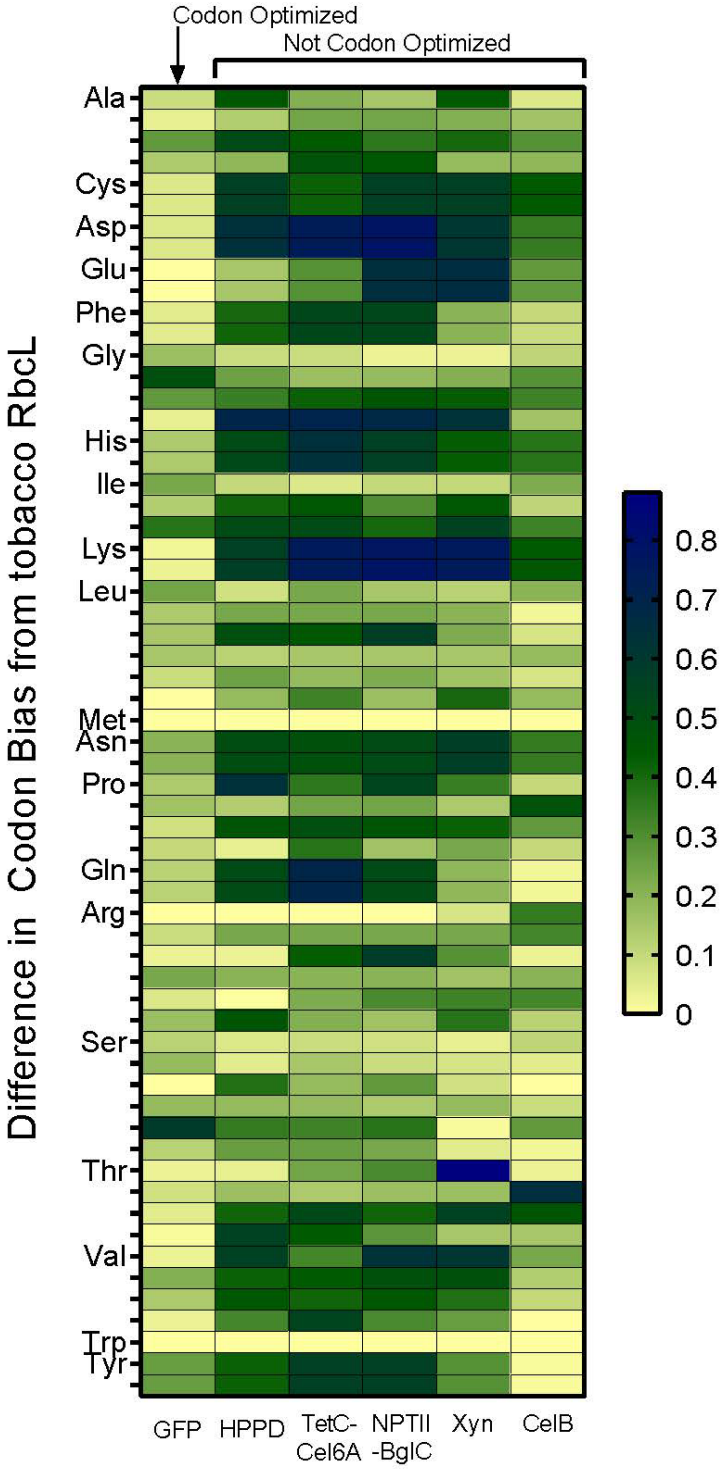


Figure 3.6: Summary of codon bias for several recombinant proteins compared to tobacco RbcL. Codon use (represented as proportions of each amino acid) of five different chloroplast-engineered recombinant genes was compared to that of *Nicotiana tabacum* RbcL to investigate the effect of codon optimization on Rubisco abundance in the transformed plants.

tobacco, only the codon-optimized GFP suffered notable reductions in Rubisco. Thus, we propose that depressed expression of endogenous proteins can be specific to some component of the expression cassette, like the 5'UTR or RBS, or be caused by a more general diversion of resources.

These experiments with TetC-*cel6*A tobacco add to a growing body of work that demonstrate the feasibility of plastid-engineered plants as a means of producing high-value proteins. However, some optimization is still needed to make GE plants a viable alternative to microbial and mammalian cell cultures. Commercializing GE plants for high-value protein production relies on generating transgenic plants that produce and accumulate recombinant protein at high levels with reliably healthy plant growth. A better understanding of plastid gene regulation would aid future work to balance protein resource allocation with heterologous protein synthesis.

Methods

Cultivation of soil-grown tobacco

For the soil cultivation experiment, seeds were sown directly into 1-gallon pots containing soil and 10 grams of 15-9-12 Osmocote fertilizer. Plants were watered as needed for the duration of the experiment with untreated tap water. The growth chamber was maintained on a 25°C/20°C day/night cycle with an 18-hour day length. Chamber solar radiation was 300 micromoles m^{-2} day^{-1} and relative humidity was set to 70%. Plant trays were rotated every third day throughout the chamber to ensure that all plants received on average the same growth conditions.

Sixteen plants were grown for each genotype, WT and TetC-*cel6*A, and four individuals were randomly selected for harvest at each of four time points (2, 6, 9, and 12 weeks of age).

Tissue Harvesting

At harvest, the aboveground biomass was weighed, and tissue was collected for protein and dry weight analysis. An equal size piece of tissue was collected from all green leaves of 6, 9, and 12-week-old plants while the entire plant above the roots was harvested from 2-week-old seedlings. All collected tissue was frozen in liquid nitrogen and stored at -80°C until further analyses.

Growth measurements

Stem length was measured from soil surface to apex with a tape measure. Measurements were taken six times throughout the duration of the experiment.

Gibberellic acid treatment

Seeds were sterilized in 100% ethanol for 1 minute and a 40% bleach, 0.5% SDS solution for 10 minutes before being rinsed four times in sterile water. Sterilized seeds were then placed evenly in two rows on a 15 cm diameter petri dish containing Murashige and Skoog media (3% sucrose) with either no GA_3 (bioactive GA) or 1 μM GA_3. The 1 μM GA_3 solution was prepared by dissolving powdered GA_3 (Sigma-Aldrich) in 100% ethanol and filter-sterilized through a 0.2 μm filter.

Plates were then placed in the dark at 4°C for three days before moving into a growth chamber with 16 hr day length at 21°C. Seeds were checked for root and cotyledon emergence every morning at the same time for nine days. Plates were also photographed for shoot and root length measurements for a total of 24 days. These measurements were carried out on the same 16 representative seedlings in ImageJ [45] (version 1.52a).

Cultivation of Nitrogen-limited and CO_2-enriched tobacco

WT and TetC-*cel6*A tobacco plants were grown in a Percival Scientific Inc. plant growth chamber (Iowa, USA) with 18-hour days, 22°C, 65% relative humidity, and illuminated with 280 micromoles m^{-2} day^{-1}. For the elevated CO_2 trial, CO_2 was supplemented with Pure Clean CO_2 (99.995% purity, Airgas, Pennsylvania, USA). Tobacco was seeded on soil and watered with fertilizer nutrient solution for 14 days before being transplanted into 4-inch diameter round pots with autoclaved vermiculite. Six plants were grown per ammonium nitrate treatment for both genotypes. Plants were watered every 3 days with 50 mL of premade liquid nutrient solution (1, 4, or 8 mM NH_4NO_3, 4mM $CaSO_4$, 30 µM $CaCl_2$, 250µM KH_2PO_4, 750 µM $MgSO_4$, 20 µM EDTA, 20 µM $FeSO_4$, 2 µM $ZnSO_4$, 0.5 µM $CuSO_4$, 2 µM $MnSO_4$, 0.5 µM Na_2MoO_4, and 42 µM H_3BO_3 [adapted from [39]). The plants were harvested for data analysis at 6 weeks old (after 4 weeks in the ammonium nitrate treatment).

Dry Weight Analyses

Weight measurements were made on two separate samples of all genotypes before and after drying the tissue. Approximately 200 mg of fresh tobacco leaf tissue was dried in pre-weighed and pre-dried envelopes.

Carbon and Nitrogen Analysis

Total carbon and nitrogen were analyzed by the Cornell Nutrient Analysis Laboratory (Ithaca, NY) using a combustion analysis of dry tissue (n=2).

Leaf Protein Extraction

Frozen WT and TetC-*cel6*A leaf samples were manually ground on ice using a micropestle for 10 minutes (n=3). A final volume of 500 µL (300 µL for 2-week old seedlings) of protein

extraction buffer (20 mM Tris-HCl, 1 mL Triton X-100, 0.1% SDS, 1.5 mM PMSF, and 0.001% β-mercaptoethanol) was added to the ground tissue. The samples were vortexed thoroughly before being centrifuged for three minutes at 20,000 × *g* at 4 °C. The supernatant was extracted, aliquoted, and stored at -80 °C until use.

Total Protein Quantification

Total soluble protein from the tobacco tissue samples was quantified using a standard Bradford Assay. The BioRad Quick Start Bovine Serum Albumin Standard Set and 1X Dye Reagent (BioRad, Hercules, CA, USA) were used for all Bradford Assays. Tobacco samples were diluted with the extraction buffer described above to achieve a concentration within the linear range of the standard curve. Absorbance measurements were carried out on a BioTek Industries Synergy 4 plate reader (BioTek Industries, Winooski, VT, USA) at 595 nm.

Immunoblotting

Protein samples and standards were mixed in a 1:1 ratio with a 2X Laemmli Sample Buffer (BioRad, Hercules, CA, USA), electrophoresed through 12% polyacrylamide gels and then transferred to PVDF membranes (BioRad, Hercules, CA, USA). Purified Rubisco protein (Agrisera, Vännäs, Sweden) was used for standard curve quantitation. Purified Cel6A and the anti-Cel6A primary antibody [34, 35] were supplied by the laboratory of the late professor David Wilson (Cornell University, Ithaca, NY, USA). The anti-Rubisco antibody was generously donated by Martin Parry (University of Lancaster, UK) [46, 47]. The secondary antibody used for all blots was a horseradish peroxidase-linked whole anti-rabbit IgG produced in donkey (GE Healthcare Life Sciences, Pittsburgh, PA, USA). All primary and secondary antibody treatments were diluted 1:20,000 in Antibody Signal Enhancer (Amresco, Solon, OH, USA).

Membranes were developed using a TMB stabilized substrate dye for horseradish peroxidase (Promega, Madison, WI, USA). Densitometry was performed using image analysis software (Integrated density in ImageJ [46]).

Codon Optimization

Transcript sequences were obtained for Beta-glucosidase CelB (NCBI Accession AF013169.2), endo-β-1,4-xylanase (NCBI Accession KJ466334) [9], HPPD [10](NCBI Accession DQ459069.1), GFP [8] (NCBI Accession EU870886.1), TetC-Cel6A [11], NPTII-BglC [12], and RbcL (NCBI Accession LT576836.1). Relative codon usage ratios were obtained using the Sequence Manipulation Suite: Codon Usage tool (http://www.bioinformatics.org/sms2/codon_usage.html). The absolute value of the difference of use proportion for each codon was compared between each test gene and *rbcL.*

Statistical analysis

Statistics were analyzed using the JMP (Version 14.3). Comparisons of the differences in means were analyzed with Student's t-test assuming unequal variances (alpha =0.05). Many of the significant p-values were included in figure captions or in the text. Errors were considered normally, identically and independently distributed. A full summary of Student's t-test p-values can be found in Tables AII.2-5.

Supplementary Materials

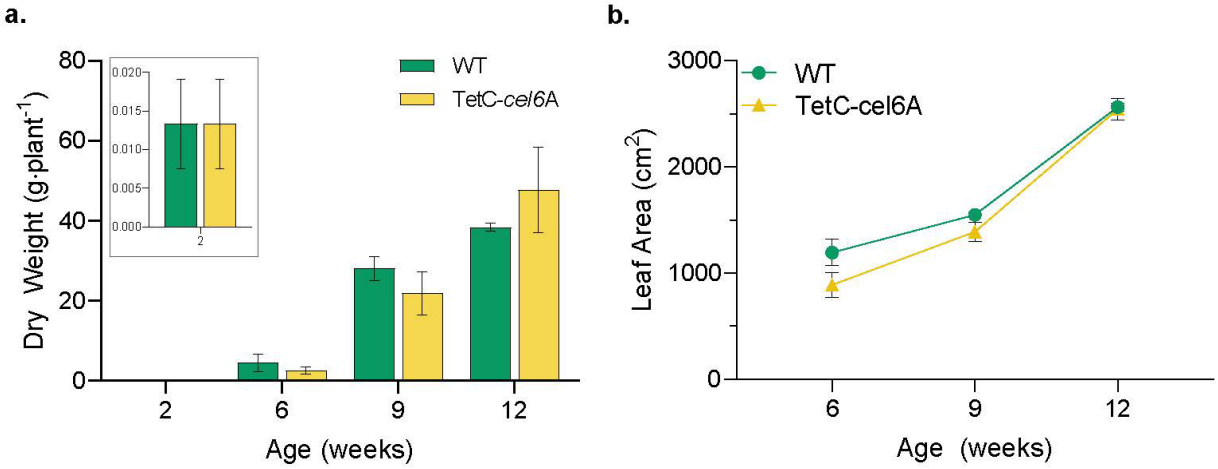


Figure 3.S1: Additional growth measurements of soil-grown transgenic tobacco. **a.** Dry weight and **b.** leaf area comparison between WT and TetC-*cel6*A tobacco by plant age. Bar heights and data points correspond to the mean and error bars reflect the standard error of the means. A full summary of p-values is provided in Table AII.2.

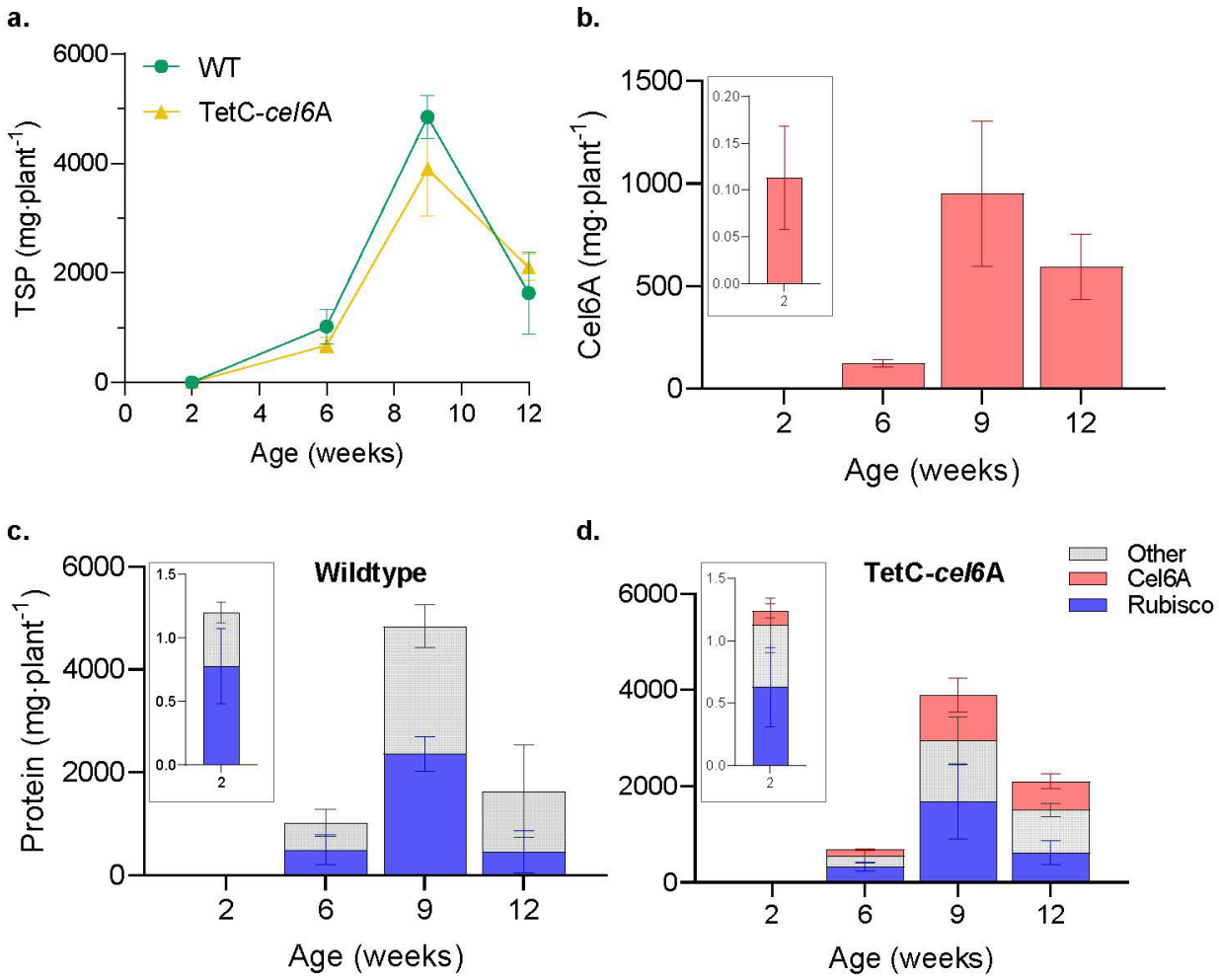


Figure 3.S2: Protein accumulation normalized for whole plant biomass. **a.** TSP and **b.** Cel6A from seedling to senescent plant. **c.** and **d.** Depict total protein allocated to Rubisco, Cel6A, and "Other" endogenous proteins. Bar heights and data points correspond to the mean and error bars reflect the standard error of the means (n =3). A full summary of p-values is provided in Table AII.2.

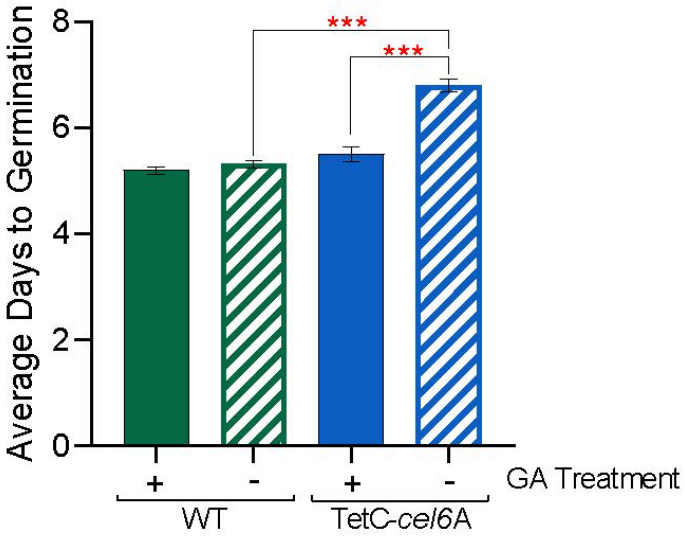


Figure 3.S3: Comparison of the effect of GA on average days until root emergence between genotypes. Bar heights correspond to the mean and error bars reflect the standard error of the means (n =60). Significant differences in two-tailed t-test *p*-values are labeled with red asterisks, $p < 0.001$ (***). A full summary of p-values is provided in Table AII.3.

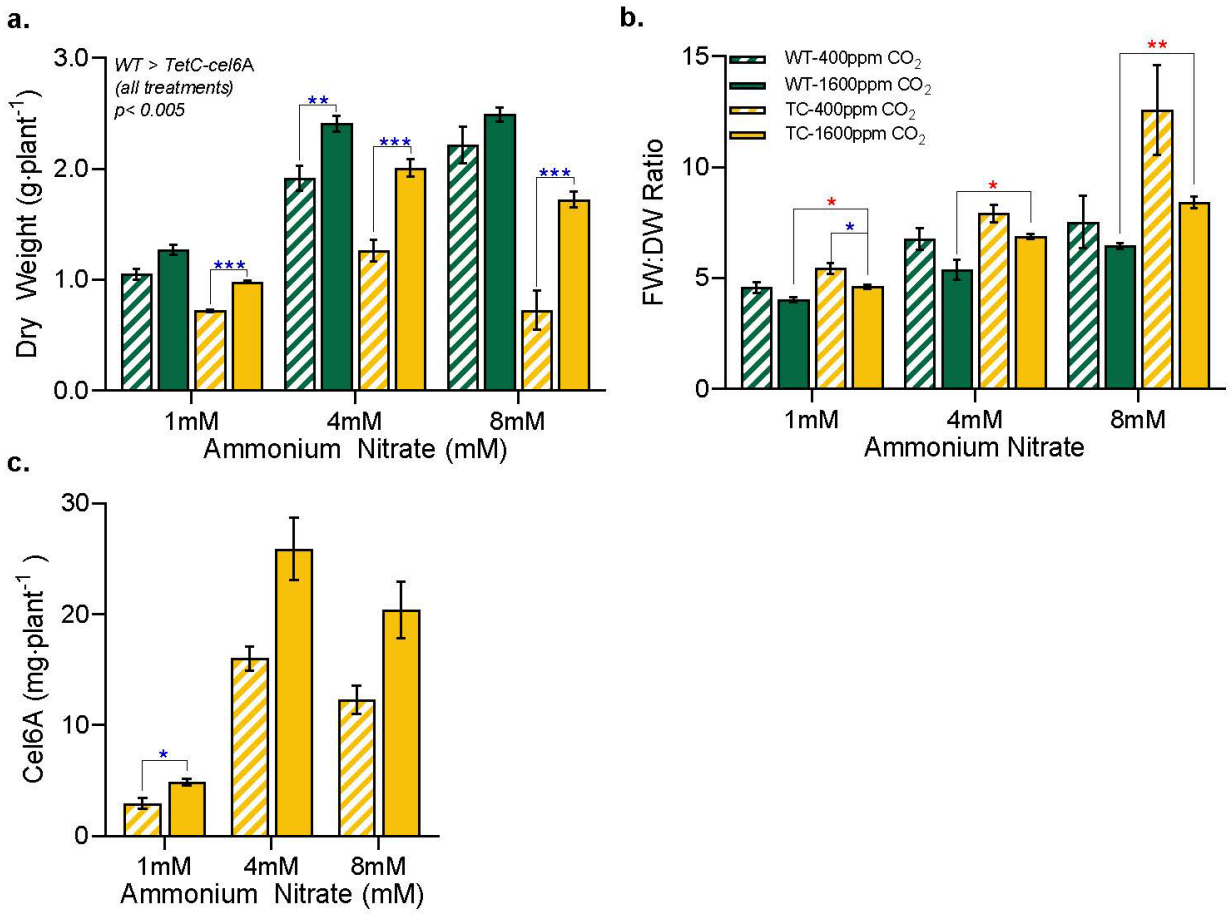


Figure 3.S4: Effect of enhanced CO_2 on dry weight, water storage, and whole plant Cel6A accumulation. **a.** Dry weight accumulation. **b.** Water storage comparison between genotypes and CO_2 treatment calculated as the ratio between fresh weight and dry weight. **c.** Cel6A yield normalized to biomass. Bar heights correspond to the mean and error bars reflect the standard error of the means (n =3). Two-tailed t-test *p*-values for comparisons between genotypes (red asterisks) and CO_2 treatment (blue asterisks) are labeled as follows, $p<0.05$ (*), $p<0.01$ (**), and $p< 0.001$ (***). A full summary of p-values is provided in Table AII.5.

References

1. Xu J, Towler M, Weathers P: Platforms for Plant-Based Protein Production. In: *Bioprocessing of Plant In Vitro Systems.* Edited by Pavlov A, Bley T. Cham: Springer; 2016: 1-40.
2. Twyman RM, Stoger E, Schillberg S, Christou P, Fischer R: Molecular farming in plants: host systems and expression technology. Trends Biotechnol 2003, 21(12):570-578.
3. Stoger E, Fischer R, Moloney M, Ma JK-C: Plant molecular pharming for the treatment of chronic and infectious diseases. Annu Rev Plant Biol 2014, 65:743-768.
4. Tusé D, Tu T, McDonald KA: Manufacturing economics of plant-made biologics: case studies in therapeutic and industrial enzymes. Biomed Res Int 2014, 2014.
5. Golczyk H, Greiner S, Wanner G, Weihe A, Bock R, Börner T, Herrmann RG: Chloroplast DNA in mature and senescing leaves: a reappraisal. The Plant Cell 2014, 26(3):847-854.
6. Shaver JM, Oldenburg DJ, Bendich AJ: Changes in chloroplast DNA during development in tobacco, *Medicago truncatula*, pea, and maize. Planta 2006, 224(1):72-82.
7. Maliga P, Bock R: Plastid biotechnology: food, fuel, and medicine for the 21st century. Plant Physiol 2011, 155(4):1501-1510.
8. Bally J, Nadai M, Vitel M, Rolland A, Dumain R, Dubald M: Plant physiological adaptations to the massive foreign protein synthesis occurring in recombinant chloroplasts. Plant Physiol 2009, 150(3):1474-1481.
9. Castiglia D, Sannino L, Marcolongo L, Ionata E, Tamburino R, De Stradis A, Cobucci-Ponzano B, Moracci M, La Cara F, Scotti N: High-level expression of thermostable

cellulolytic enzymes in tobacco transplastomic plants and their use in hydrolysis of an industrially pretreated *Arundo donax* L. biomass. Biotechnol Biofuels 2016, 9(1):154.

10. Dufourmantel N, Dubald M, Matringe M, Canard H, Garcon F, Job C, Kay E, Wisniewski JP, Ferullo JM, Pelissier B: Generation and characterization of soybean and marker-free tobacco plastid transformants over-expressing a bacterial 4-hydroxyphenylpyruvate dioxygenase which provides strong herbicide tolerance. Plant Biotechnol J 2007, 5(1):118-133.
11. Gray BN, Ahner BA, Hanson MR: High-level bacterial cellulase accumulation in chloroplast-transformed tobacco mediated by downstream box fusions. Biotechnol Bioeng 2009, 102(4):1045-1054.
12. Gray BN, Yang H, Ahner BA, Hanson MR: An efficient downstream box fusion allows high-level accumulation of active bacterial beta-glucosidase in tobacco chloroplasts. Plant Mol Biol 2011, 76(3):345-355.
13. Maliga P: Plastid transformation in higher plants. Annu Rev Plant Biol 2004, 55:289-313.
14. Oey M, Lohse M, Kreikemeyer B, Bock R: Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. Plant J 2009, 57(3):436-445.
15. Petersen K, Bock R: High-level expression of a suite of thermostable cell wall-degrading enzymes from the chloroplast genome. Plant Mol Biol 2011, 76(3-5):311-321.
16. Tregoning JS, Nixon P, Kuroda H, Svab Z, Clare S, Bowe F, Fairweather N, Ytterberg J, van Wijk KJ, Dougan G: Expression of tetanus toxin fragment C in tobacco chloroplasts. Nucleic Acids Res 2003, 31(4):1174-1179.

17. Chakrabarti SK, Lutz KA, Lertwiriyawong B, Svab Z, Maliga P: Expression of the *cry9Aa2 B.t.* gene in tobacco chloroplasts confers resistance to potato tuber moth. Transgenic Res 2006, 15(4):481-488.

18. Zhou F, Badillo-Corona JA, Karcher D, Gonzalez-Rabade N, Piepenburg K, Borchers AM, Maloney AP, Kavanagh TA, Gray JC, Bock R: High-level expression of human immunodeficiency virus antigens from the tobacco and tomato plastid genomes. Plant Biotechnol J 2008, 6(9):897-913.

19. Bally J, Job C, Belghazi M, Job D: Metabolic adaptation in transplastomic plants massively accumulating recombinant proteins. PLoS One 2011, 6(9):e25289.

20. Huang B, Xu Y: Cellular and Molecular Mechanisms for Elevated CO_2–Regulation of Plant Growth and Stress Adaptation. Crop Sci 2015, 55(4):1405-1424.

21. Stitt M, Krapp A: The interaction between elevated carbon dioxide and nitrogen nutrition: the physiological and molecular background. Plant, Cell Environ 1999, 22(6):583-621.

22. Kolotilin I, Kaldis A, Pereira EO, Laberge S, Menassa R: Optimization of transplastomic production of hemicellulases in tobacco: effects of expression cassette configuration and tobacco cultivar used as production platform on recombinant protein yields. Biotechnol Biofuels 2013, 6(1):1.

23. Lenzi P, Scotti N, Alagna F, Tornesello ML, Pompa A, Vitale A, De Stradis A, Monti L, Grillo S, Buonaguro FM: Translational fusion of chloroplast-expressed human papillomavirus type 16 L1 capsid protein enhances antigen accumulation in transplastomic tobacco. Transgenic Res 2008, 17(6):1091-1102.

24. Longoni P, Leelavathi S, Doria E, Reddy VS, Cella R: Production by Tobacco Transplastomic Plants of Recombinant Fungal and Bacterial Cell-Wall Degrading Enzymes to Be Used for Cellulosic Biomass Saccharification. Biomed Res Int 2015, 2015.

25. Kuroda H, Maliga P: Overexpression of the *clpP* 5′-untranslated region in a chimeric context causes a mutant phenotype, suggesting competition for a *clpP*-specific RNA maturation factor in tobacco chloroplasts. Plant Physiol 2002, 129(4):1600-1606.

26. Zoschke R, Bock R: Chloroplast translation: structural and functional organization, operational control, and regulation. The Plant Cell 2018, 30(4):745-770.

27. Bock R: Engineering plastid genomes: Methods, tools, and applications in basic research and biotechnology. Annu Rev Plant Biol 2015, 66(1):211-241.

28. Komar AA: The art of gene redesign and recombinant protein production: approaches and perspectives. In: *Protein Therapeutics.* Springer; 2016: 161-177.

29. Jin S, Kanagaraj A, Verma D, Lange T, Daniell H: Release of hormones from conjugates: chloroplast expression of β-glucosidase results in elevated phytohormone levels associated with significant increase in biomass and protection from aphids or whiteflies conferred by sucrose esters. Plant Physiol 2011, 155(1):222-235.

30. Bhat M: Cellulases and related enzymes in biotechnology. Biotechnol Adv 2000, 18(5):355-383.

31. Hedden P, Proebsting WM: Genetic analysis of gibberellin biosynthesis. Plant Physiol 1999, 119(2):365-370.

32. Piskurewicz U, Jikumaru Y, Kinoshita N, Nambara E, Kamiya Y, Lopez-Molina L: The gibberellic acid signaling repressor RGL2 inhibits *Arabidopsis* seed germination by

stimulating abscisic acid synthesis and ABI5 activity. The Plant Cell 2008, 20(10):2729-2745.

33. Birch-Machin I, Newell CA, Hibberd JM, Gray JC: Accumulation of rotavirus VP6 protein in chloroplasts of transplastomic tobacco is limited by protein stability. Plant Biotechnol J 2004, 2(3):261-270.

34. Oey M, Lohse M, Scharff LB, Kreikemeyer B, Bock R: Plastid production of protein antibiotics against pneumonia via a new strategy for high-level expression of antimicrobial proteins. PNAS 2009, 106(16):6579-6584.

35. Esquıvel MG, Ferreira RB, Teixeira AR: Protein degradation in C3 and C4 plants subjected to nutrient starvation. Particular reference to ribulose bisphosphate carboxylase/oxygenase and glycolate oxidase. Plant Sci 2000, 153(1):15-23.

36. Feller U, Anders I, Demirevska K: Degradation of rubisco and other chloroplast proteins under abiotic stress. Gen Appl Plant Physiol 2008, 34(1-2):5-18.

37. Kingston-Smith AH, Bollard AL, Minchin FR: Stress-induced changes in protease composition are determined by nitrogen supply in non-nodulating white clover. J Exp Bot 2005, 56(412):745-753.

38. Busch FA, Sage RF, Farquhar GD: Plants increase CO_2 uptake by assimilating nitrogen via the photorespiratory pathway. Nature Plants 2018, 4(1):46.

39. Geiger M, Haake V, Ludewig F, Sonnewald U, Stitt M: Influence of nitrate and ammonium nitrate supply on the response of photosynthesis, carbon and nitrogen metabolism, and growth to elevated carbon dioxide in tobacco. Plant, Cell Environ 1999, 22:1177-1199.

40. Kirschbaum M: The sensitivity of C3 photosynthesis to increasing CO_2 concentration: a theoretical analysis of its dependence on temperature and background CO_2 concentration. Plant, Cell Environ 1994, 17(6):747-754.

41. Biemelt S, Tschiersch H, Sonnewald U: Impact of altered gibberellin metabolism on biomass accumulation, lignin biosynthesis, and photosynthesis in transgenic tobacco plants. Plant Physiol 2004, 135(1):254-265.

42. Facchinelli F, Weber A: The Metabolite Transporters of the Plastid Envelope: An Update. Front Plant Sci 2011, 2(50).

43. Tomme P, Warren R, Gilkes N: Cellulose hydrolysis by bacteria and fungi. In: *Adv Microb Physiol.* vol. 37: Elsevier; 1995: 1-81.

44. Bulmer M: The selection-mutation-drift theory of synonymous codon usage. Genetics 1991, 129(3):897-907.

45. Schneider CA, Rasband WS, Eliceiri KW: NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012, 9(7):671.

46. Lin MT, Hanson MR: Red algal Rubisco fails to accumulate in transplastomic tobacco expressing *Griffithsia monilis RbcL* and *RbcS* genes. Plant Direct 2018, 2(2):1-10.

47. Lin MT, Stone WD, Chaudhari V, Hanson MR: Enzyme kinetics of tobacco Rubisco expressed in *Escherichia coli* varies depending on the small subunit composition. Preprint at https://wwwbiorxivorg/content/101101/562223v1 2019:562223.

CHAPTER 4

Designing synthetic downstream boxes to predictably alter recombinant protein production in chloroplast-engineered tobacco

Abstract

The rapid and efficient production of valuable proteins is in high demand for many areas of industrial manufacturing and biopharmaceutical synthesis. Genetically engineered plants may offer a flexible, cost-effective platform for producing these valuable proteins, but current understanding of gene expression limits the applicability of this technology. In higher plant chloroplasts, the downstream box (DB) regulatory region has the potential to significantly increase recombinant protein yields, but its role in influencing translation is poorly understood. Transcript folding patterns identified by *in silico* structure modeling and confirmed by *in vivo* DMS-treatment suggest that a DB that prevents the transgenic ORF from disrupting the native secondary structure of the construct's 5'UTR results in higher recombinant protein accumulation. Conversely, DBs that are not strong base-pairing matches for the downstream ORF accumulate less foreign protein. We confirmed the importance of this DB-gene dependent mRNA folding relationship using a series of six synthetic DBs. In addition to folding pattern, the thermodynamic stability of the DB's structures is inversely related to final heterologous protein accumulation, highlighting the importance of weak secondary structures around the translation initiation region. Thus, we have created and begun to test a method of synthetically designing DBs to enhance accumulation foreign protein synthesis from plastid engineered tobacco.

Introduction

High-value protein production from genetically engineered (GE) plants is a rapidly evolving technology with applications to improve nutritional quality of crop plants, produce biopharmaceuticals, and synthesize enzymes for industrial processing (see reviews [1, 2]). Genetic transformation of the nuclear genome accounts for the vast majority of commercialized GE crops, although plastid genome engineering is emerging as a promising alternative for synthesizing large quantities of valuable proteins [3]. Several unique properties of chloroplast gene expression contribute to high recombinant protein yields including the inherently high expression of native plastid genes, absence of gene silencing, and high copy number of plastid genomes in leaf cells (reviewed in [1]). Despite this potential, chloroplast engineering has, as of yet, remained confined to proof-of-concept laboratory experiments and requires additional work to become a marketable large-scale protein production platform.

One challenge to commercializing this technology is improving the consistency of recombinant protein yields. Currently, transplastomic plants average heterologous protein yields of 5-20% of TSP, but these values range anywhere from below 0.1% of total soluble protein (TSP) to as much as 70% of TSP (Chapter 2, [4-8]). Because foreign protein accumulation is dependent on many factors, it is often difficult to predict yield before committing to the lengthy and expensive process of establishing a transplastomic line. Thus, a better understanding of the regulatory mechanisms of chloroplast gene expression may reduce guesswork involved in transgenic engineering.

Because plastid gene expression is largely regulated at the level of post-transcription and translation, regulatory features of the mRNA, like the untranslated regions (UTRs) are critical determinants of protein abundance (reviewed in [9]). Much of this regulation is dependent on

transcript secondary structure, which itself is determined by base-pairing patterns in the ribonucleotides of the mRNA [10-19]. For instance, the 3'UTR is important for inhibiting exoribonuclease progression, often by forming a strong inverted repeat hairpin [20]. In *Nicotiana tabacum*, the *rbc*L and *psb*A 3'UTRs are commonly used in many of the highest yielding transplastomic constructs [4, 8, 21-26].

However, while the 3'UTR is important for mRNA stability, the 5'UTR plays a more dynamic role in regulating translation initiation. The secondary structure of the 5'UTR is critical to determining the accessibility and formation rate of interactions between the transcript and RNA-binding proteins (RBPs) for processes like intron excision, transcript stabilization, RNA editing, and ribosome complex assembly (reviewed in [27]). Both the local RNA structure and the thermodynamic stability (ΔG) of the Shine-Dalgarno (SD) sequence, if it is present, and the start codon are important for determining translation rates [10, 14]. In this regard, higher plant chloroplasts exhibit many prokaryotic regulatory features. For instance, as in bacteria, *in vivo* mRNA folding analyses have indicated that many highly expressed chloroplast transcripts have weak thermodynamic stability around the translation initiation region of the 5'UTR [10, 13, 19]. Conversely, strong secondary structures at the 5' end of the mRNA aid in inhibiting exoribonuclease degradation [28].

Another important regulatory region at the 5'end of the transcript is the downstream box (DB), named for its location just downstream of the start codon. The DB has previously demonstrated a strong regulatory role, where expression of the foreign protein increased by 4-fold when the bacteriophage T7G10 DB region was included in the expression cassette [29]. Initially, this region was hypothesized to function similarly to the SD sequence by interacting with a portion of the rRNA to aid in ribosome assembly, but enhancing the complementarity

between the DB of a plastid transgene and the 16S rRNA reduced foreign protein synthesis in plastids [29-32].

More recent work with DBs in chloroplast-engineered tobacco suggests that the regulatory function of the DB is dependent on the ORF downstream. In particular, the work presented in this manuscripts begins with an analysis of six transplastomic lines consisting of a combination of the DB regions from three genes, *tetC*, *npt*II, and *gfp*, fused to the full open reading frame (ORF) of either the endoglucanase, *cel6*A or the beta-glucosidase, *bglC* [24, 25]. Early work with these lines demonstrated that, while the GFP DB performed poorly for both foreign proteins, the TetC DB produced the most Cel6A and the NPTII DB resulted in the greatest BglC, indicating a context-dependent function of the DB.

In this manuscript, we first tested the role of mRNA secondary structure on DB function in these six lines by comparing *in silico* mRNA folding patterns with foreign protein accumulation. However, *in silico* mRNA structure modeling is limited by its inability to accurately capture the cellular environment where RNA-binding proteins and other molecular interactions can alter the structure of the transcript. Methods of analyzing *in vivo* structural patterns, such as chemical probing, can more accurately assess the folding pattern of RNA molecules [13]. One such chemical probe, dimethyl sulfate (DMS), methylates the Watson-Crick base pairing faces of adenine and cytosine residues in unstructured regions of RNA, like loops, bulges, and mismatches. These methylated bases halt reverse transcription and are detected by sequencing the 3'ends of cDNA created from treated RNA [12].

With these transplastomic plants, we were able to link transgenic mRNA *in vivo* secondary structural patterns and thermodynamic stability with resultant recombinant Cel6A

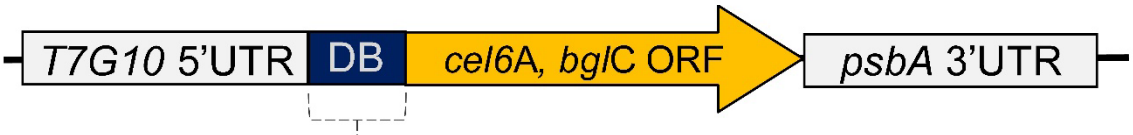
yield. We then designed a series of six synthetic DBs paired with the *cel6*A gene to confirm these patterns in mRNA structure and foreign protein yield.

Results

In silico mRNA secondary structure reveals folding patterns linked to recombinant protein accumulation

In silico mRNA secondary structural modeling of the six transplastomic tobacco lines predicted considerable variation in base-pairing interactions dependent on the DB-gene combination (Figure 4.1). When fused with *cel6*A, both the TetC and GFP DBs base paired with the beginning of the ORF while the NPTII DB exhibited greater compatibility for folding internally than with *cel6*A. In contrast, all three DBs base paired with the *bglC* ORF, although TetC and GFP did so further downstream in the gene (Figure 4.2a).

While there was notable variability in the specific structures of the different DB-gene combinations, the effect those interactions had on the folding of the 5'UTR may be more important. In its native conformation, the T7G10 5'UTR adopted an *in silico* predicted structure consisting of a short inverted repeat hairpin from bases 1-21 and a longer second hairpin with the SD sequence partially single-stranded in its loop (Figure 4.2a). The highest yielding transplastomic lines, TetC-*cel6*A and NPTII-*bglC*, had 5'UTR *in silico* structures that matched this folding pattern. However, the second hairpin in the 5'UTR was reduced in the lower yielding lines, NPTII-*cel6*A, TetC-*bglC*, and both GFP-gene constructs. In NPTII-*cel6*A, this reduction in the second hairpin sequestered the entire length of the SD sequence in base pairs at the bottom of another hairpin (Figure 4.2b).



DB	DB Sequence (5' - 3')
TetC	ATG *GCT AGC* AAA AAT CTG GAT TGT TGG GTC GAC AAT GAA GAA GAT ATA
NPTII	ATG *GCT AGC* ATT GAA CAA GAT GGA TTG CAC GCA GGT TCT CCG GCC GCT
GFP	ATG *GCT AGC* GGC AAG GGC GAG GAA CTG TTC ACT GGC GTG GTC CCA ATC
SynN1	ATG *GCT AGC* AAA AAT CTT GAT TGT TGG ATA CAG ATT TTT GAT AGC CAT
SynN2	ATG *GCT AGC* AAA AAT CTT GAT AAT TGG ATA CAG ATT AGA GAT AGC CAT
SynN3	ATG *GCT AGC* AAA TGT CTT GAT AAT TGG ATA AAC GAA AGA GAT AGC CAT
SynT1	ATG *GCT AGC* AAA AAT GAA GAC ATG TTG GGA TTG ACG TAT AAA GGA GAA
SynT2	ATG *GCT AGC* AAA AAT TTT GAC ATG TTG GCT TTG ACG GCT AAA AGA CAT
SynT3	ATG *GCT AGC* TGT TCC TGG TTG AAA GAC TTT AAC ATG AGA TAT GAT GAA

Figure 4.1: Cartoon schematic of the transgenic mRNA. Each transgenic transcript contains the T7G10 bacteriophage 5'UTR, *psbA* 3'UTR, a transgenic ORF (*cel6*A or *bglC*), and one of 8 DBs. The NheI restriction enzyme site at the start of each DB is in italics. Full vector schematic can be found in [25] and Chapter 2 Figure 2.2a.

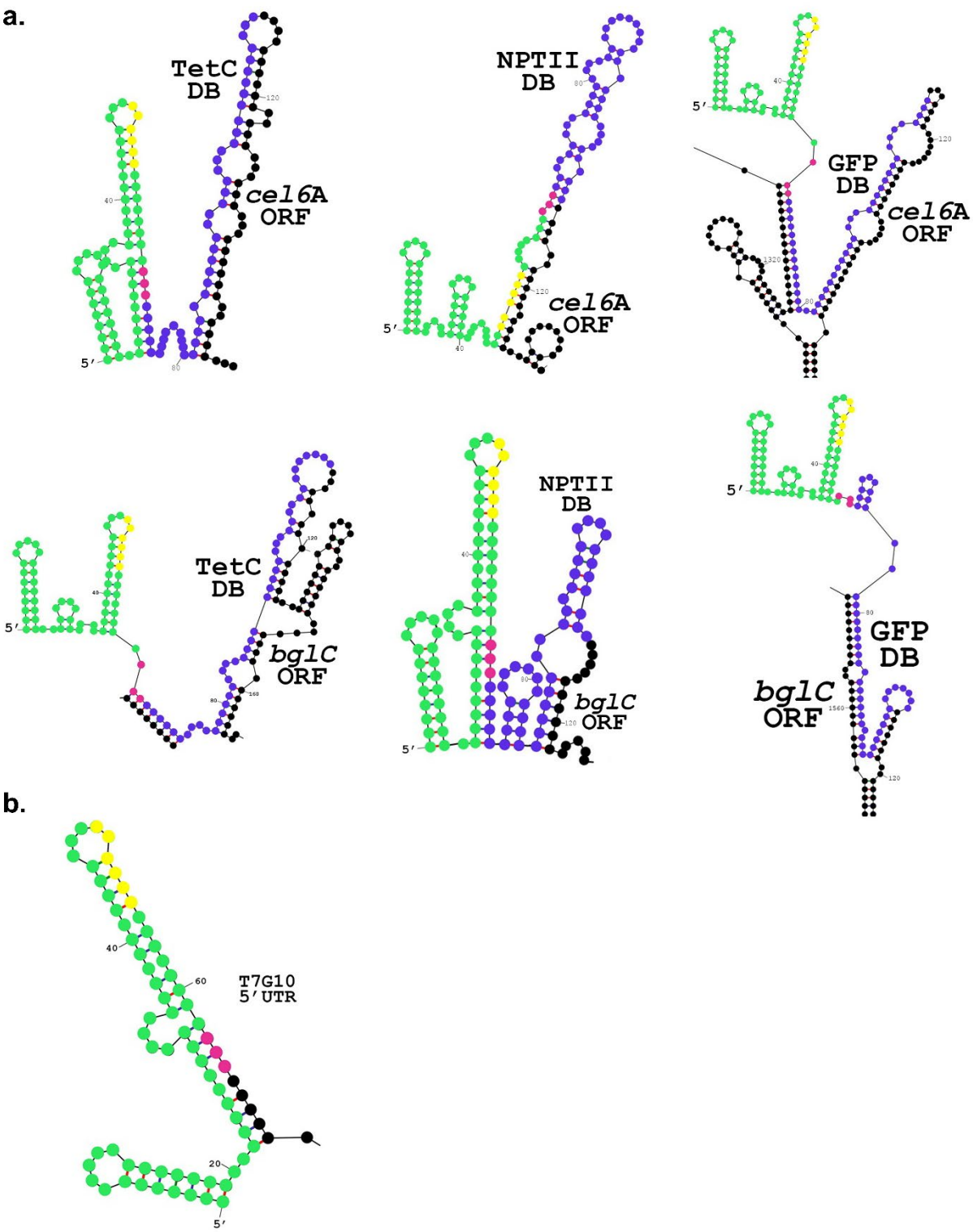


Figure 4.2: *In silico* structures of 5'end of the transgenic *cel6*A and *bglC* mRNA. Bases are colored by domain: T7G10 5'UTR (green), SD (yellow), start codon (pink), DB (blue), and ORF (black).

Our *in silico* modeling suggests that a DB that promotes higher foreign protein abundance prevents the ORF from interfering with the endogenous base-pairing patterns of the 5'UTR in the transgenic mRNA. However, *in silico* mRNA modeling only considers the thermodynamic properties of the mRNA base-pairing. It does not capture how the cellular matrix or RNA-binding proteins, including translating ribosomes, may alter the secondary structure of the transcript. For that reason, we conducted *in vivo* mRNA structure probing to identify base-pairing interactions in the mRNA within our transplastomic chloroplasts.

Chloroplast 23S rRNA folding demonstrates the use of DMS-treatment for *in vivo* RNA secondary structure prediction

Chemical probing methods of *in vivo* structure modeling rely on modifying ribonucleotides in a manner that is dependent on folding patterns. For examples, DMS methylates only unpaired bases and reverse transcription is terminated by these methylation points; thus, the 3'ends of the synthesized cDNA identify bases that are single-stranded. By sequencing and aligning these cDNA reads to the target mRNA strand, this method provides information about each residue's degree of base-pairing along the mRNA strand through its reactivity to DMS. Finally, this DMS reactivity is applied as a folding constraint to modeling software to improve the prediction of mRNA secondary structure. Also, to eliminate results generated from random reverse transcriptase drop-off or mRNA degradation, we analyzed water-treated negative control samples in parallel with our DMS-treated test samples [12].

Soil-grown *N. tabacum* seedlings were submersed in a treatment solution containing either 0.75% DMS or water as negative control. Total cellular RNA extracted from these plants was used as a template for cDNA synthesis with reverse transcription primers targeted to the transgenic *cel6*A ORF and the native plastid 23S rRNA [33].

To test that our DMS treatment was sufficient, we used the chloroplast 23S rRNA as a positive control because it has a well-established secondary structure derived from the evolutionary conservation of the tight structure-function association in rRNA [34, 35]. Also, its location within the chloroplast ensures that the DMS treatment fully penetrated the chloroplast double membrane.

Applying the DMS-derived constraints to the structure modeling program generated *in vivo* 23S rRNA structure predictions that more closely matched the reference structure than the unconstrained *in silico* models in many regions of the rRNA sequence (data not shown). For example, the *in vivo* DMS model improved prediction of the region spanning nucleotides 56-72 to perfectly match the reference model. This same region was poorly predicted using *in silico* modeling alone (Figure 4.3). Because this DMS-treatment technique constrains modeling predictions based on reactivity to DMS, any cellular condition that limits DMS accessibility to RNA molecules will be interpreted by the modeling software as base-pairing. For example, regions of the 23S rRNA that are buried within the ribosome complex are likely shielded from DMS treatment, and should, therefore, produce *in vivo* models with more base-pairing than the reference model. To illustrate this point we note that, although our *in vivo* model improves the prediction of nucleotides 198- 221 compared to the *in silico* model, the DMS-treated prediction contains six paired bases that are unparied in the reference model (Figure 4.3). Based on these results, we can use the information gathered from DMS treatment to improve the accuracy of the secondary structure predictions of our different DB-*cel6*A combinations.

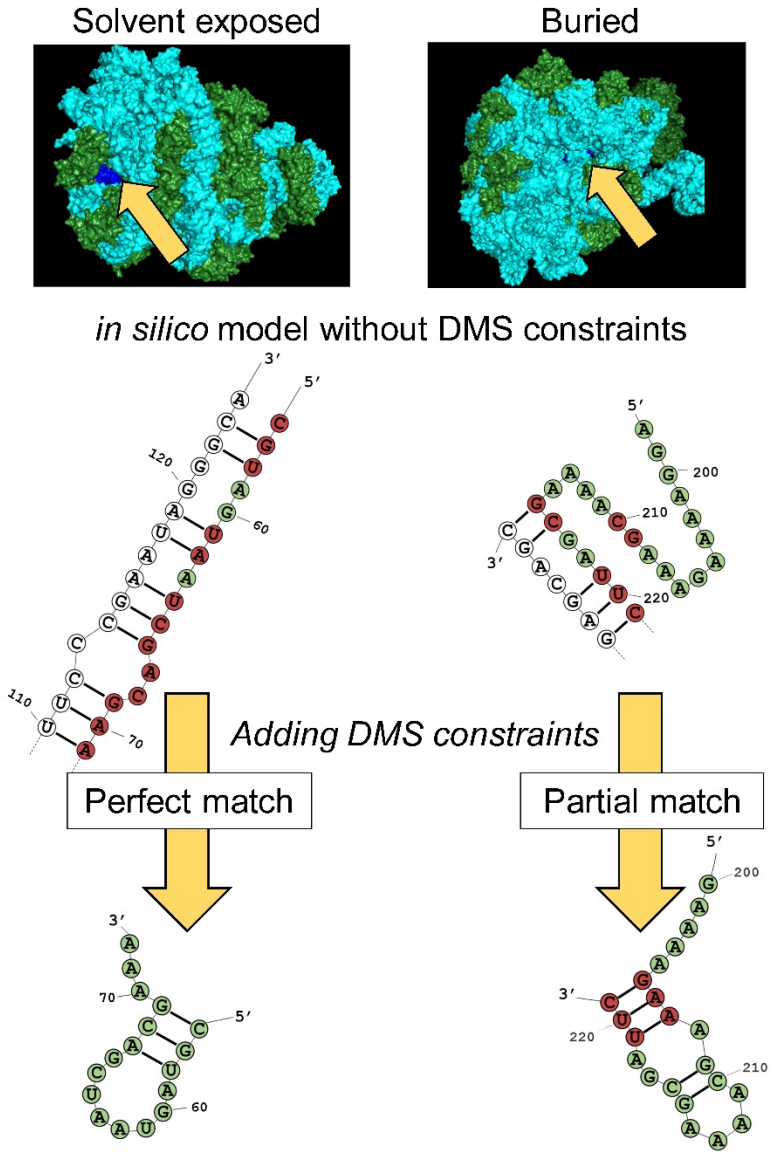


Figure 4.3: Comparison of structure prediction for *in silico* and *in vivo* modeling of the chloroplast 23S rRNA using phylogenetic reference model. Green bases match the reference model; dark red bases do not match the reference. Location of the region of the rRNA in the ribosome complex is colored dark blue in the images at the top and identified with the yellow arrow (the remainder of the rRNA is turquoise and the ribosomal proteins are green).

DMS-constrained mRNA structure modeling confirmed *in silico* folding patterns that correlated with recombinant protein accumulation

Mapping DMS reactivity along transgenic mRNA identified several regions of the transcript that were variably labeled by DMS (Figure 4.4a). We had anticipated that highest heterologous protein accumulation would be correlated with highest DMS reactivity resulting in a lack of base-pairing around the SD and start codon, as that is expected to facilitate ribosome binding. Instead, the highest DMS reactivity in this translation initiation region was observed in the transgenic mRNA that resulted in intermediate Cel6A abundance, NPTII-*cel6*A. Furthermore, both TetC-*cel6*A and NPTII-*cel6*A had the lowest DMS reactivity in the 5'UTR region with higher reactivity in the DB and ORF. In contrast, GFP-*cel6*A exhibited the highest DMS reactivity in the 5'UTR with very little reactivity measured in the ORF.

When we applied the DMS reactivity to constrain secondary structure predictions of our transgenic constructs, we found that the *in vivo* modeling results largely confirmed our *in silico*-based structure predictions despite some differences in the exact folding patterns. For instance, while *in silico* models predicted that the TetC DB paired with the 5'end of *cel6*A, *in vivo* models shift some of this pairing approximately 100 nucleotides downstream into the *cel6*A ORF. However, this slight change in base-pairing does not drastically alter TetC-*cel6*A transcript's predicted 5'UTR structure, which is still expected to mimic the native fold of the T7G10 5'UTR (Figure 4.4b). *In vivo* models were nearly identical to the *in silico* predictions for NPTII-*cel6*A mRNA, depicting a very similar internal hairpin (Figure 4.4c). Finally, GFP-*cel6*A transcripts exhibited the greatest variations between *in vivo* and *in silico* predictions. Where *in silico* models had the GFP DB pairing with a stretch of the 3'end of *cel6*A, *in vivo* models depict that

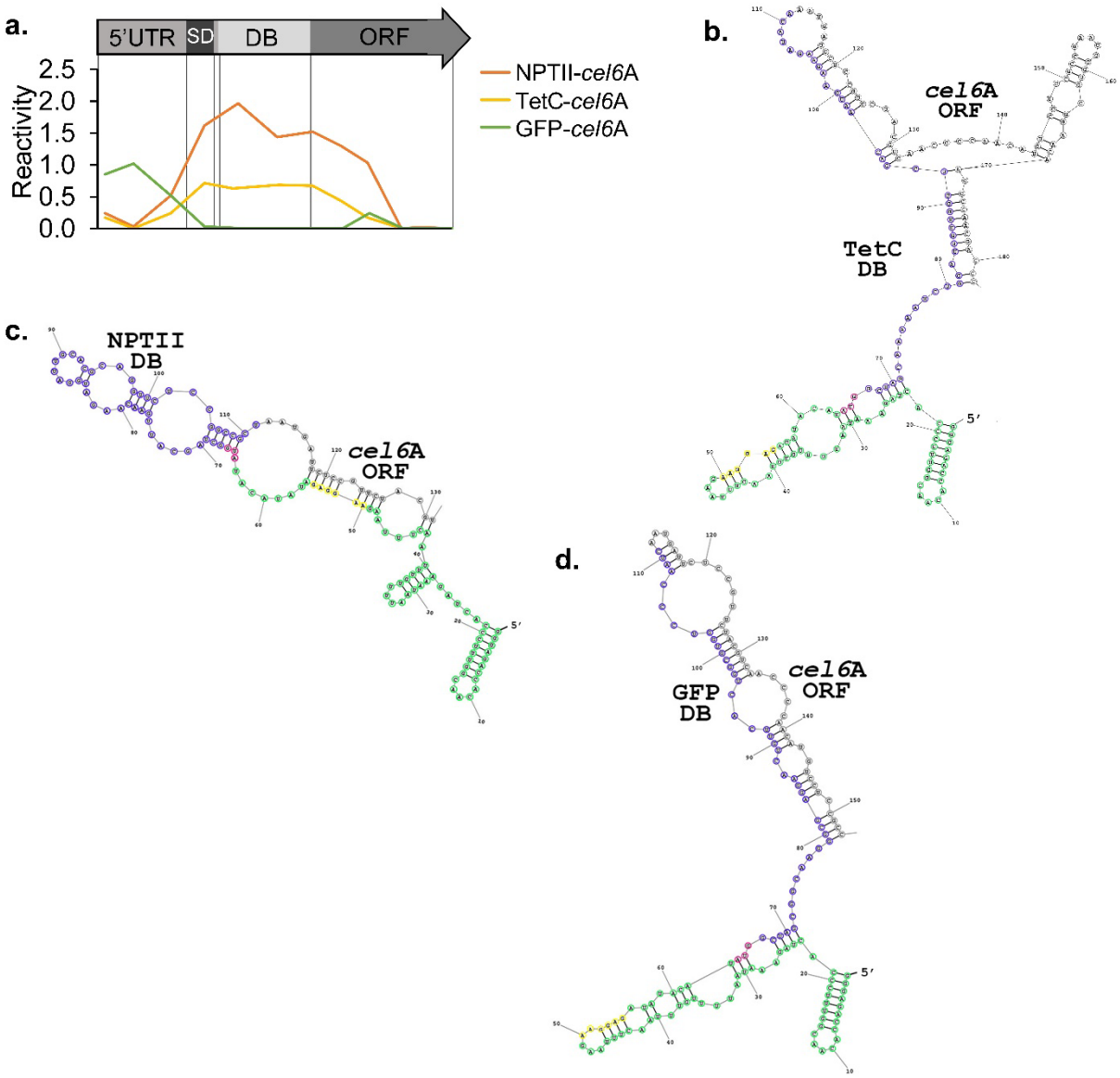


Figure 4.4: DMS reactivity and *in vivo* mRNA secondary structure for *cel6*A transgenic transcripts. **a.** DMS reactivity was averaged over 10nt windows along the first 250nt of the transcript for each DB-*cel6*A combination. **b.-d.** The most common *in vivo* secondary structures obtained for TetC-*cel6*A (b), NPTII-*cel6*A (c), and GFP-*cel6*A (d). Bases are colored by domain: T7G10 5'UTR (green), SD (yellow), start codon (pink), DB (blue), and ORF (black)

this DB pairs only with the 5'end of the ORF. Furthermore, this shift in base pairing alters the 5'UTR structure to more closely match the native fold for T7G10 (Figure 4.4d).

We also considered the impact of the mRNA thermodynamic stability on accumulation of our recombinant cellulase since higher thermodynamic stability predominantly correlates with lower production of the resulting protein. For instance, in our transgenic mRNA, GFP-*cel6*A produced the most thermodynamically stable transcripts and yielded the least amount of Cel6A (Table 4.1, Student's t-test p-value < 0.001). In contrast, TetC-*cel6*A was slightly more stable than NPTII-*cel6*A, suggesting that small variations in folding free energy may not be as impactful as the secondary structures. Furthermore, the GFP DB contain eleven rare plastid codons while TetC and NPTII contain four and three, respectively. This disparity could at least partially explain the exceptionally low GFP-Cel6A yields. Finally, due to the importance of the N-terminal amino acid on protein stability, we verified that all three DBs have the same amino acid, alanine, after the initiator methionine, due to the placement of a restriction enzyme site immediately after the start codon (Figure 4.1).

Designing synthetic DBs to test the role of DB secondary structure and folding free energy on Cel6A accumulation

We designed six new DBs paired with the *cel6*A ORF to test the impact of DB base-pairing structure and strength on recombinant Cel6A accumulation. Three of these new DBs were designed to fold internally like NPTII-*cel6*A (Figure 4.5a, SynN series) and three paired with the *cel6*A ORF similarly to TetC-*cel6*A (Figure 4.5b, SynT series). In both series, the DB labeled "1" (i.e. SynN1 and SynT1) was the most thermodynamically stable and stability progressively decreased with each consecutive DB. The nucleotide and amino acid sequence of the new synthetic DBs are less than 50% identical to their respective precursors TetC or NPTII DBs.

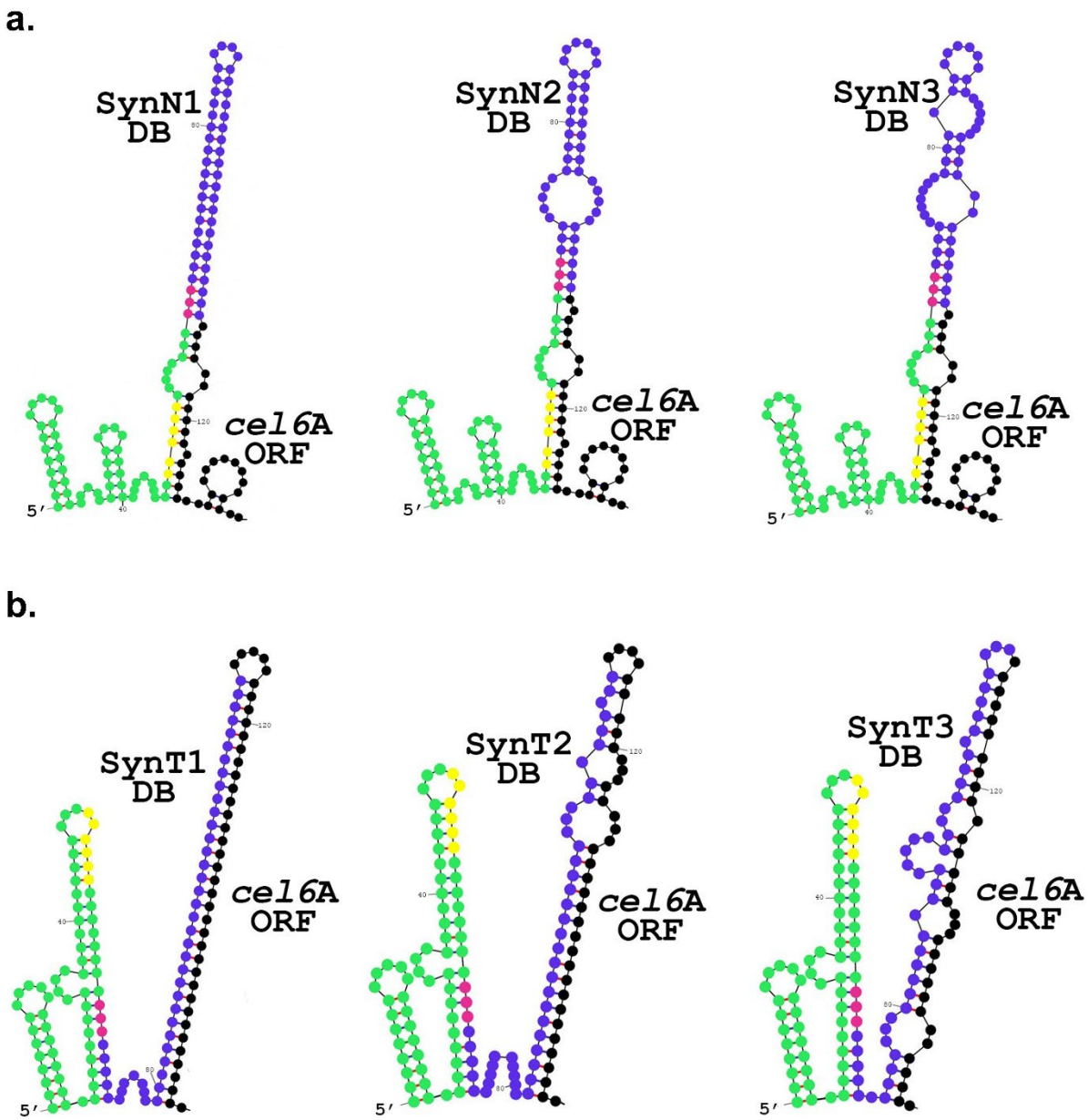


Figure 4.5: *in silico* structure models of the synthetic DB-*cel6*A fusions. **a.** SynN series mimics the secondary structure of NPTII-*cel6*A. **b.** SynT series mimics TetC-*cel6*A. For both series, the DB labeled "1" has the lowest free energy of folding and the DB labeled "3" has the highest.

Each synthetic DB also had the same number of rare codons, three, and the same N-terminal amino acid. Thus, this set of constructs enables us to test the individual and combined effects of thermodynamic stability and transcript folding patterns in synthetic DBs on heterologous protein accumulation.

Establishing and analyzing synthetic DB-*cel6*A mRNA and protein accumulation

Transformants for these six new DB-*cel6*A constructs as well as TetC-*cel6*A and NPTII-*cel6*A were generated via particle bombardment of *N. tabacum* (cv. Petit Havana) leaves and regenerated on sterile media containing spectinomycin. Homoplasmy was confirmed for the transformed shoots using Southern blots (Figure 4.S1). It took a minimum of four and a maximum of nine rounds of selection to obtain homoplasmic transformants, which is considerably more than most other reported tobacco plastid engineering projects. Furthermore, we were unable to isolate homoplasmic plants for SynT1-*cel6*A, SynT2-*cel6*A, or NPTII-*cel6*A even after at least ten consecutive rounds of selection on spectinomycin. We did obtain a homoplasmic SynN2-*cel6*A plant after nine rounds of selection, but laboratory constraints imposed by COVID-19 delayed analyses of this genotype.

We grew all available generation T1 plants in soil for nine weeks to identify any mutant phenotypes associated with transformation and Cel6A production. Biomass accumulation from the transgenic plants were all similar to WT tobacco except TetC-*cel6*A, which suffered reduced yields (Figure 4.6a, b). However, SynN1-*cel6*A, SynT3-*cel6*A, and TetC-*cel6*A plants all had significantly shorter stems than WT (Figure 4.6c). The transgenic tobacco lines also exhibited some delays in germination and reduction in seed viability compared to WT with TetC-*cel6*A being the most severely affected in both regards (Figure 4.S2).

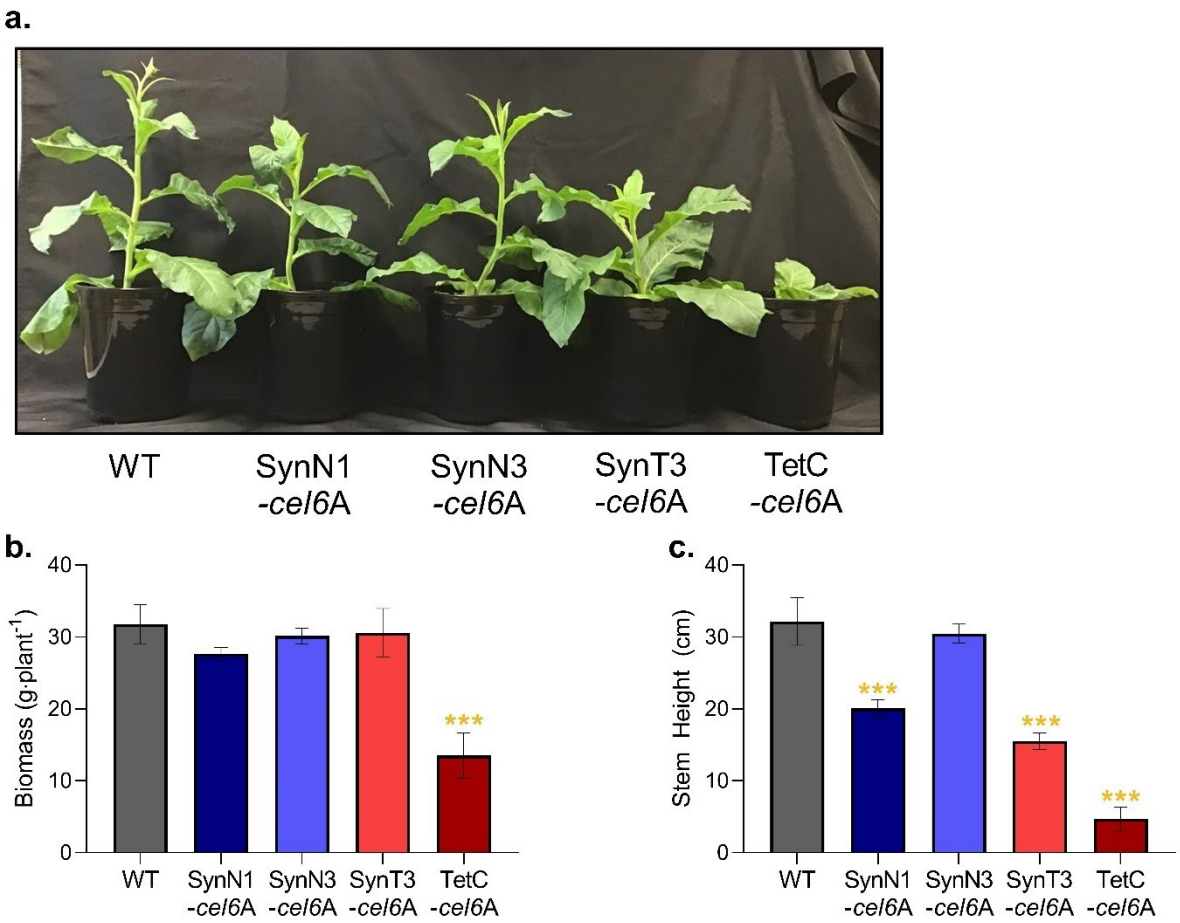


Figure 4.6: Growth assessment of chloroplast-transformed T1 plants. **a.** Pictures of representative plants of each genotype at harvest. **b.** Total aboveground biomass. **c.** Stem height. For biomass and stem height, n =4 for WT, SynN1-*cel6*A, and SynN3-*cel6*A while n=5 for TetC-*cel6*A and SynT3-*cel6*A. Bars indicate the mean of biological replicates and error bars depict the standard error of the means. Student's t-test *p* -values for comparisons between transgenic plants and WT tobacco are indicated as $p < 0.0005$ (***). Full statistical report can be found in Table AII.6.

Relative abundance of transgenic mRNA was compared between lines with qPCR normalized to transcripts for elongation factor 1-alpha. Transcript abundances for SynN1-*cel6*A, SynN3-*cel6*A, and TetC-*cel6*A were all very similar, but SynT3-*cel6*A accumulated two to three times more transgenic mRNA (Figure 4.7a).

In contrast to mRNA abundance, immunoblots revealed greater variations in Cel6A accumulation. TetC-*cel6*A and SynT3-*cel6*A accumulated the largest quantities of Cel6A, as predicted, with averages of 5% and 6.3% of TSP, respectively (Figure 4.7b, c). The DBs designed to fold like NPTII-*cel6*A accumulated much less Cel6A with SynN1-*cel6*A reaching barely detectable levels less than 0.001% Cel6A of TSP and SynN3-*cel6*A achieving averages of 0.1% Cel6A.

These new transformants also demonstrate the importance of the thermodynamic stability of secondary structures on Cel6A accumulation. For example, the SynN1 DB engaged in more base pairing than SynN3, making SynN1 DB a significantly more thermodynamically stable secondary structure than SynN3. Moreover, SynN1-*cel6*A yielded at least 500-times less Cel6A than SynN3-*cel6*A (Figure 4.7d). Similarly, while the difference in Cel6A accumulation between TetC-*cel6*A and SynT3-*cel6*A was statistically insignificant, TetC-*cel6*A yielded slightly less Cel6A protein and was predicted to be more thermodynamically stable (Figure 4.7e, Student's t-test of ΔG: p-value < 0.0001). By extrapolating the relationship between ΔG and Cel6A we estimated that SynN2-*cel6*A (to be tested at a later date) will yield intermediate Cel6A between SynN1-*cel6*A and SynN3-*cel6*A, while NPTII-*cel6*A would have achieved the highest Cel6A yields of that series. By the same logic, SynT2-*cel6*A and SynT1-*cel6*A would likely have yielded progressively less Cel6A than TetC-*cel6*A. Thus, we confirmed here that structural stability is inversely correlated with recombinant protein accumulation.

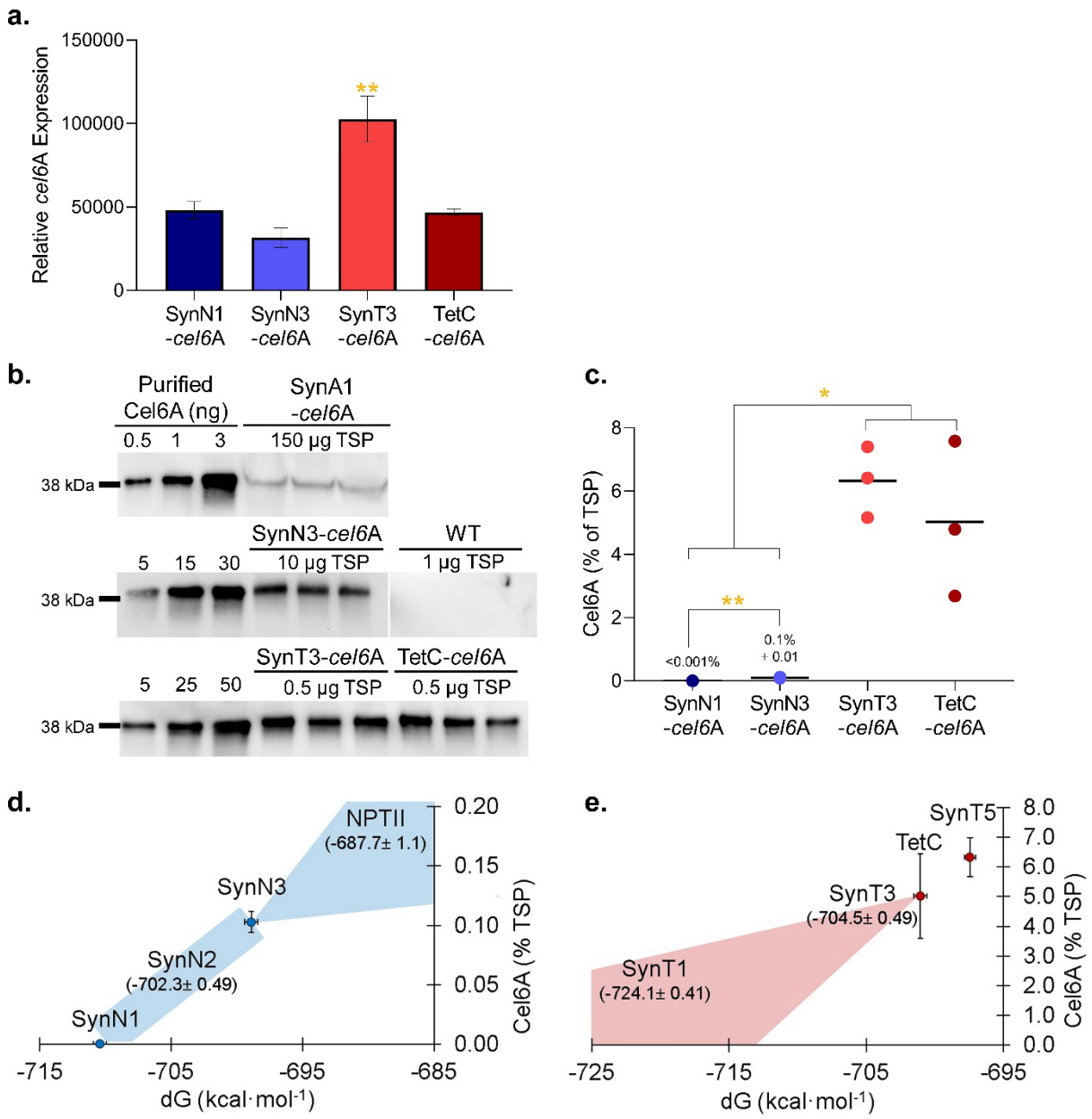


Figure 4.7: Cel6A protein and mRNA quantification in new chloroplast-transformed tobacco lines. **a.** Relative quantification of *cel6*A mRNA based on qPCR with internal reference gene. **b.** TSP quantification. Bars indicate the mean of biological replicates and error bars depict the standard error of the means. Student's t-test *p* -values for comparison between transgenic plants and WT tobacco are indicated as $p < 0.002$ (**). **c.** Cel6A immunoblots. **d.** Cel6A yield. **e.-f.** Comparison of Cel6A yield versus ΔG of folding for NPTII-like (e) and TetC-like (f) DBs with projected values for NPTII, SynN2, SynT1, and SynT2 DBs based on ΔG. Data points represent the values of each of the 3 biological replicates with the black indicating the mean of each set. Student's t-test *p* -values for comparison between transgenic plants are indicated as $p < 0.002$ (**) and $p < 0.05$ (*). Full statistical report can be found in Table AII.6.

Discussion

Here, we have postulated and tested the hypothesis that the mRNA secondary structure of the DB is a strong determinant of recombinant protein synthesis in engineered plastids. More specifically, our *in silico* and *in vivo* folding predictions suggested that the DB plays a role in buffering the base pairing interactions between the 5'UTR and ORF of chimeric transgene mRNA.

Transcriptome-wide *in silico* secondary structure analyses have reported that many endogenous transcripts fold in a modular fashion with each domain (i.e. 5'UTR, ORF, and 3'UTR) folding independently [18]. This modular folding, combined with the importance of 5'UTR structure on regulating translation, led us to postulate that the function of the 5'UTR strongly depends on its secondary structure in the context of the transgenic mRNA. However, transgenic mRNAs are often a patchwork of unrelated UTRs and ORFs chosen with little regard for how they may base pair together. Here, we propose that the DB may function to prevent the ORF from interfering with the native 5'UTR structure.

Using modern *in silico* structure software, we tested the effect of these DB folding patterns on resultant recombinant protein abundance by designing six synthetic DBs. These new DBs confirmed that promoting the native folding pattern of the 5'UTR yielded far more Cel6A than constructs that disrupted the native base pairing of the 5'UTR. The thermodynamic stability of the DB's secondary structures was also critical to exogenous cellulase yields. Our observation that stronger structures yielded less Cel6A than weaker structures is consistent with reports from native plastid transcriptome analyses as well as transgenic mRNA from engineered *E. coli* [10, 13, 19, 36].

A context-dependent mechanism for translation control also explains the lack of a consistently strong DB between different recombinant proteins. In addition to Gray et al.'s *cel6*A and *bglC* tobacco, other plastid engineering projects have demonstrated variability in recombinant protein yield with DB fusions [4, 5, 37, 38]. For instance, the DB from tobacco *rbcL* promoted the highest foreign protein production for the full-length NPTII protein as well as an HPV capsid protein [5, 37]. However, another report found that the DB from *atpB* produced more β-glucosidase than the *rbcL* DB, but their endoglucanase protein achieved its greatest yields when the DB was omitted entirely [4]. We would expect to see such variability if, as we propose, the function of the DB is not inherent to the DB itself but is dependent on its interaction with surrounding construct components.

Variations in physiology between host plant cultivars could also drive different recombinant protein expression levels. For instance, Conley et al reported protein-specific variations in foreign protein yields between *N. tabacum* cultivars when they expressed three recombinant proteins in separate nuclear transformations [39]. Here, we chose the cultivar Petit Havana as a host for chloroplast transformation because of its reduced generation time compared to Samsun. The metabolic differences between these cultivars that alter generation time could also explain the minor variation in Cel6A yield observed in TetC-*cel6*A transplastomic plants between the two cultivars. Regardless of yield differences, Petit Havana made an effective platform for testing the influence of the DB on foreign protein production.

One challenge we faced in engineering these Petit Havana lines was that we were unable to obtain homoplasmic transformants for three of the constructs, SynT1-*cel6*A, SynT2 -*cel6*A, and NPTII-*cel6*A. Because we used the same WT tissue for transformation, targeted all of the constructs to the same location in the plastid genome, and were able to successfully establish the

other four synthetic DB-gene constructs, the problem may be specific to the these three DBs. For instance, while we initially obtained 10-20 shoots from bombardment of most of the constructs, we only isolated six shoots for SynT1-*cel6*A. Two of these shoots were random mutations for spectinomycin resistance, while a third shoot was entirely chlorotic and died off after two rounds of selection. We were able to obtain homoplasmic transformants from bombardment with SynT2-*cel6*A and NPTII-*cel6*A; however, all subsequent sequencing of these shoots revealed highly mutated DB DNA sequences. A cursory search of the original DB nucleotide sequences of SynT1, SynT2, and NPTII revealed no significant similarities to the *N. tabacum* nuclear or plastid genomes. Likewise, an amino acid comparison between these DBs and the *N. tabacum* proteome failed to identify any obvious mechanisms for interference. However, given the complexity of gene expression in higher plants it is still possible that some unforeseen factor interfered with the host plant's metabolism too severely to obtain correct homoplasmic plants.

To date, a wide range of bioactive recombinant proteins have been produced in chloroplast engineered plants, including those from human, fungal, and bacterial origins [40] . Here, we tested the hypothesis that DBs could be used as mRNA structural buffers to maintain native 5'UTR function in transgene constructs to promote translation of a recombinant protein. Future work to test the utility of this mechanism with other recombinant proteins could help expand the applicability of transgenic plants as platforms for high-value protein production. If achieved, the ability to design a DB tailored to any given gene-of-interest could potentially reduce the time and resource intensive guess-and-check method of generating transplastomic plants.

Methods

RNA *in silico* structure modeling

In silico models were made using the mfold web server at 23°C [41] (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form2.3).

RNA *in vivo* labeling and structure modeling

For *in vivo* RNA structure probing, plants were grown in soil for 23 days. Seedlings were carefully removed from soil to maintain roots and washed in nuclease-free water to remove dirt and debris.

Procedure for *in vivo* RNA folding analysis was carried out as described in Ding *et al* 2015 [13] with modifications indicated below. The DMS reaction buffer consisted of 100 mM KCl, 40mM HEPES [pH 7.5], 0.5 mM $MgCl_2$, and either 0.75% DMS (treated samples) or the same volume of water (negative control samples). Three seedlings were added to this buffer in a 50 mL Falcon tube for 15 min with occasional swirling to mix. The reaction was terminated with the addition of 1.5 g DTT with continuous swirling for 2 min. The seedlings were then rinsed twice with nuclease-free water. The roots were removed, and the remainder of the seedlings were flash frozen in liquid nitrogen.

Total cellular RNA was extracted using RNeasy Plant Mini Kit (Qiagen) and the quality of the RNA was verified using a Bioanalyzer.

Gene-specific cDNA synthesis from the extracted RNA was carried out using SuperScript III first strand synthesis kit (Invitrogen) with primers specific to the *cel6*A, *bglC*, or the chloroplast 23S rRNA (Table 4.S1) [33].

Ligation of cDNA with the linker (Table 4.S1) was completed using Circligase (Lucigen, Madison, WI, USA) and reaction conditions from Ding *et al* 2015[12].

Preparation of the sequencing library was carried out as described in Ding *et* al 2015 [13] with Illumina TruSeq universal and indexed primer adaptors for multiplexed sequencing with the addition of the three-primer selective amplification system to select for target genes as described by Watters *et* al 2016 [42] (Table 4.S1).

Resulting PCR products were size-selected, cleaned, and sequenced with 150bp MiSeq by the Biotechnology Resource Center (Cornell University, Ithaca, NY, USA).

Sequencing results were analyzed by first using Adaptertrim in Linux to remove linker and RT primer sequences flanking the reads. Then, trimmed reads were aligned to target gene sequences using Bowtie2. Interpretation and constrained RNA structures were obtained using the Structurefold workflow in Galaxy. Folding was carried out at a simulated temperature of 37°C as is standard.

Construct Design

Synthetic downstream boxes were tested for predicted mRNA structure and folding free energy using mfold online software [41].

The original transformation plasmids, pTetCCel6AEC and pNPTIICel6AEC, from Gray *et al* (2009) were used to generate new transplastomic lines of TetC-*cel6*A and NPTII-*cel6*A in *N. tabacum* (cv. Petit Havana). The pTetCCel6AEC plasmid was also used as a template for all of the other constructs. The TetC-*cel6*A downstream box sequence was excised from the construct using the NheI restriction enzyme site at the 5'end of the DB sequence and the NotI restriction enzyme site located between the *cel6*A ORF and the *psb*A terminator. The remaining

downstream boxes were PCR amplified using a forward primer that annealed to the 5'end of the *cel6*A ORF with a 5'end tail containing each individual DB (Table 4.S2). The reverse primer, Cel6A-TpsbA-rev was the same for all constructs. The PCR products, containing DB-*cel6*A, were double-digested with NheI and NotI and ligated to the digested pTetCCel6AEC plasmid to create pSynA1Cel6A, pSynA2Cel6A, pSynA3Cel6A, pSynB1Cel6A, pSynB3Cel6A, and pSynB5Cel6A. These plasmids were transformed into chemically competent DH 5-alpha *E. coli*.

Tobacco transformation

Chloroplast transformation of *N. tabacum* was carried out as described by [43] with a few modifications detailed below. Each construct was bombarded on ten separate, sterile tobacco leaves with a final concentration of 1-2 ug plasmid DNA per shot. Selection was carried out with 0.5 mg mL-1 spectinomycin dichloride pentahydrate in RMOP media. Spectinomycin-resistant shoots were transferred to fresh RMOP + spectinomycin plates for as many rounds of selection as necessary to reach homoplasmy (as determined with Southern blots, see below).

Southern blotting

Southern blots were used to verify homoplasmy of transformation of the plastid genome. Tissue from shoots in selection was flash frozen on liquid nitrogen and ground into a fine powder. Genomic DNA was extracted using a 2% CTAB buffer (2% CTAB, 1.4M NaCl, 100mM Tris-HCl [pH 8.0], and 20mM EDTA [pH 8.0], filter-sterilized with 0.02µm filter), phase separated using 24:1 chloroform: iso-amyl alcohol, and precipitated with 100% isopropanol. Final DNA was quantified using a Nanodrop. 1-3ug of genomic DNA was then double digested with XhoI and HindII using NEB FastDigest. Digestion products were size-separated on a 0.8% agarose gel in TAE buffer. Digested DNA was then transferred onto a Hybond-N+ nylon transfer

membrane (Amersham) and probed with Anti-DIG-AP Fab Fragments (Roche). The membrane was then detected with CDP-Star (Roche) and ChemiDoc Imaging System (Biorad).

Homoplasmic plant growth

Homoplasmic T1 seeds were germinated on LM-111 soil in a growth chamber with an 18hr day length at 23°C. The flats were watered as needed with a nutrient solution containing 5% of 21-5-20 NPK. Seedlings were grown for 21 days before harvesting. To harvest, whole plant aboveground fresh weight was measured first before harvesting tissue from each true leaf for protein, RNA, and dry weight analyses. Tissue was flash frozen on liquid nitrogen and stored at -80°C until further analysis.

Leaf Protein Extraction and Quantification

Frozen leaf tissue samples were manually ground on dry ice using a micropestle (n=3). A final volume of 200 µL of protein extraction buffer (20 mM Tris-HCl, 1 mL Triton X-100, 0.1% SDS, 1.5 mM PMSF, and 0.001% β-mercaptoethanol) was added to the ground tissue. The samples were vortexed thoroughly before being centrifuged for three minutes at 20,000 × *g* at 4 °C. The supernatant was extracted, aliquoted, and stored at -80 °C until use.

Total soluble protein from the tobacco tissue samples was quantified using a standard Bradford Assay. The BioRad Quick Start Bovine Serum Albumin Standard Set and 1X Dye Reagent (BioRad, Hercules, CA, USA) were used for all Bradford Assays. Protein samples were diluted with water to achieve a concentration within the linear range of the standard curve. Absorbance measurements were carried out on a BioTek Industries Synergy 4 plate reader (BioTek Industries, Winooski, VT, USA) at 595 nm.

Immunoblotting

Protein samples and standards were mixed with a 4X Laemmli Sample Buffer (BioRad, Hercules, CA, USA), electrophoresed through 12% polyacrylamide gels and then transferred to PVDF membranes (BioRad, Hercules, CA, USA). Purified Cel6A and the rabbit anti-Cel6A primary antibody [34, 35] were supplied by the laboratory of the late professor David Wilson (Cornell University, Ithaca, NY, USA). The secondary antibody used for all blots was a horseradish peroxidase-linked whole anti-rabbit IgG produced in goat (Invitrogen, Waltham, MA). All primary and secondary antibody treatments were diluted 1:20,000 in Western blocker (Sigma,St. Louis, MO, USA).

Membranes were developed using ECL reagents (Promega) and visualized using ChemiDoc Imager (Biorad). Cel6A was quantified from western blots using ImageLab (Biorad).

RNA extraction, cDNA synthesis, and qPCR

RNA was extracted using RNeasy plant mini kit (Qiagen) from tissue of the same plants used for protein extraction. On-column DNA digestion was completed using RNase-free DNase I kit (Qiagen). Total RNA was eluted in nuclease-free water, quantified with Nanodrop, and used immediately for cDNA synthesis. The iScript cDNA Synthesis kit (Biorad, USA) was used for random priming cDNA synthesis of all samples (n =3 per genotype).

These cDNA reactions were diluted 1:4 with water and loaded into the SSoAdvanced Universal SYBR Green Supermix kit (Biorad, USA) with primers targeting the 5'end of the *cel6*A ORF (5end-*cel6*A_fwd/rev, Table 4.S3). Relative quantification was achieved with primers that targeted elongation factor 1 alpha (EF1A_fwd/rev, Table 4.S3).

Statistical analysis

Statistics were analyzed using the JMP (Version 14.3). Comparisons of the differences in means were analyzed with Student's t-test assuming unequal variances (alpha =0.05). Many of the significant *p*-values were included in figure captions or in the text. Errors were considered normally, identically and independently distributed. A full summary of Student's t-test *p*-values can be found in Tables AII.6.

Supplementary Materials

Table 4.S1: Primers and linkers for *in vivo* RNA structure analysis

Function	Name	Sequence (5'-3')
cDNA synthesis	cel6A_RT	GCTCTTCCGATCTTCGACCTGGCCGGTGATCT
	bglC_RT	GCTCTTCCGATCTCAGAAGGTGTCCCAGATGCT
	23SrRNA_RT	GCTCTTCCGATCTCCCGTTTTCACGGTTTAGGCT
Ligation linker	ssDNA_linker	/5Phos/NNNAGATCGGAAGAGCGTCGTGTAG/3SpC3/
PCR Selection	cel6A_Select	CTTTCCCTACACGACGCTCTTCCGATCTTCGACCTGGCCGGTGATCT*G*C*C*G*G
	bglC_Select	CTTTCCCTACACGACGCTCTTCCGATCTCAGAAGGTGTCCCAGATGCT*G*G*G*G
	23SrRNA_Select	CTTTCCCTACACGACGCTCTTCCGATCTCCCGTTTTCACGGTTTAGGCTG*C*T*C

[1] /5Phos/ indicates a monophosphate group on the 5'end
[2] /3SpC3/ indicates a 3 carbon spacer group on the 3'end
[3] Asterisks (*) indicate bases with a phosphorothioate backbone

Table 4.S2: Synthetic DB plasmid construction primers

Name	Sequence (5'-3')
Cel6A-TpsbA-rev	ATAGACTAGGCCAGGATCGCGGCCGCTCAGCTGGCGGCGCAGGT
SynN1_fwd	ATGGCTAGCAAAAATCTTGATTGTTGGATACAGATTTTTGCTAGCCAT AATGATTCTCCGTTCTAC
SynN2_fwd	ATGGCTAGCAAAAATCTTGATAATTGGATACAGATTAGAGCTAGCCAT AATGATTCTCCGTTCTAC
SynN3_fwd	ATGGCTAGCTGGTGTCTTGATAATTGGATAAACGAAAGAAAAAGCCAT AATGATTCTCCGTTCTAC
SynT1_fwd	ATGGCTAGCAAAAATGAAGACATGTTGGGATTGACGTATAAAGGAGAA AATGATTCTCCGTTCTAC
SynT3_fwd	ATGGCTAGCAAAAATTTTGACATGTTGGCTTTGACGGCTAAAAGACAT AATGATTCTCCGTTCTAC
SynT5_fwd	ATGGCTAGCTGTTCCTGGTTGAAAGACTTTAACATGAGATATGATGAA AATGATTCTCCGTTCTAC

Table 4.S3: qPCR primers

Name	**Sequence (5'-3')**
5end-cel6A_fwd	CCGATCCTGGTCGTGTA
5end-cel6A_rev	CGGTTCGACGATGATGTAG
EF1A_fwd	TGCCTTGTGGAAGTTTGAG
EF1A_rev	CCAGTGGTGGAGTCAATAATC

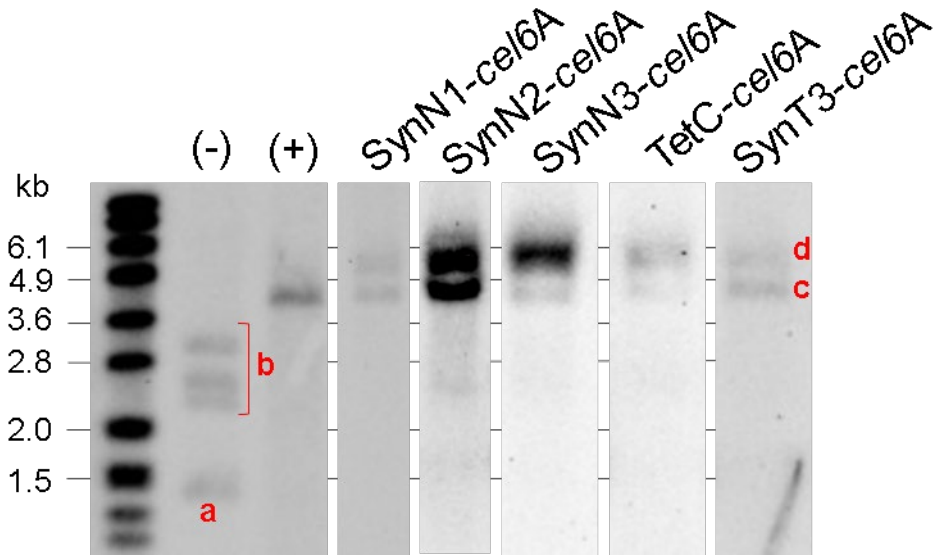


Figure 4.S1: Southern blots for each of the homoplasmic transformed lines. Untransformed WT Petit Havana was used for the negative control (-) in lane 2 while previously established TetC-*cel6*A tobacco (c.v. Samsun) was used as a positive control (+) in lane 3. The band identified as (a) is the expected untransformed band while the grouping of bands labeled (b) are untransformed DNA that was incompletely digested during Southern blot digestion. Band (c) is the expected transformed band and (d) is a product of unintended recombination described in [25]. Laboratory limitations imposed by COVID-19 quarantines delayed production of a complete southern blot of all homoplasmic lines, so this figure is a patchwork of each Southern blot where these lines were identified as homoplasmic.

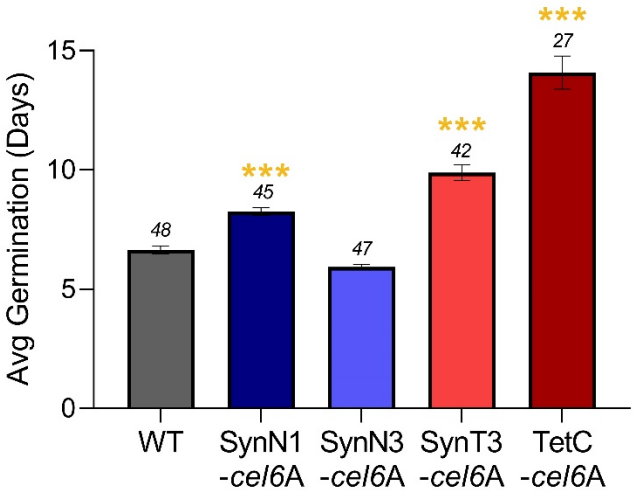


Figure 4.S2: Germination and seed viability of transplastomic tobacco. The number of seeds that germinated per genotype are given above the corresponding bar. Bars indicate the mean of biological replicates and error bars depict the standard error of the means. Student's t-test p - values for comparison between transgenic plants are indicated as $p < 0.002$ (**) and $p < 0.05$ (*). A full statistical report can be found in Table AII.6.

References

1. Maliga P, Bock R: Plastid biotechnology: food, fuel, and medicine for the 21st century. Plant Physiol 2011, 155(4):1501-1510.
2. Xu J, Towler M, Weathers P: Platforms for Plant-Based Protein Production. In: *Bioprocessing of Plant In Vitro Systems.* Edited by Pavlov A, Bley T. Cham: Springer; 2016: 1-40.
3. Leelavathi S, Reddy VS: Chloroplast expression of His-tagged GUS-fusions: a general strategy to overproduce and purify foreign proteins using transplastomic plants as bioreactors. Mol Breed 2003, 11(1):49-58.
4. Castiglia D, Sannino L, Marcolongo L, Ionata E, Tamburino R, De Stradis A, Cobucci-Ponzano B, Moracci M, La Cara F, Scotti N: High-level expression of thermostable cellulolytic enzymes in tobacco transplastomic plants and their use in hydrolysis of an industrially pretreated *Arundo donax* L. biomass. Biotechnol Biofuels 2016, 9(1):154.
5. Lenzi P, Scotti N, Alagna F, Tornesello ML, Pompa A, Vitale A, De Stradis A, Monti L, Grillo S, Buonaguro FM: Translational fusion of chloroplast-expressed human papillomavirus type 16 L1 capsid protein enhances antigen accumulation in transplastomic tobacco. Transgenic Res 2008, 17(6):1091-1102.
6. Oey M, Lohse M, Kreikemeyer B, Bock R: Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. Plant J 2009, 57(3):436-445.
7. Stevens LH, Stoopen GM, Elbers IJ, Molthoff JW, Bakker HA, Lommen A, Bosch D, Jordi W: Effect of climate conditions and plant developmental stage on the stability of antibodies expressed in transgenic tobacco. Plant Physiol 2000, 124(1):173-182.

8. Tregoning JS, Nixon P, Kuroda H, Svab Z, Clare S, Bowe F, Fairweather N, Ytterberg J, van Wijk KJ, Dougan G: Expression of tetanus toxin fragment C in tobacco chloroplasts. Nucleic Acids Res 2003, 31(4):1174-1179.

9. De Marchis F, Pompa A, Bellucci M: Plastid Proteostasis and Heterologous Protein Accumulation in Transplastomic Plants. Plant Physiol 2012, 160(2):571-581.

10. De Smit MH, Van Duin J: Control of prokaryotic translational initiation by mRNA secondary structure. In: *Prog Nucleic Acid Res Mol Biol.* vol. 38: Elsevier; 1990: 1-35.

11. De Smit MH, van Duin J: Control of translation by mRNA secondary structure in *Escherichia coli*: a quantitative analysis of literature data. J Mol Biol 1994, 244(2):144-150.

12. Ding Y, Kwok CK, Tang Y, Bevilacqua PC, Assmann SM: Genome-wide profiling of *in vivo* RNA structure at single-nucleotide resolution using structure-seq. Nat Protoc, 2015, 10(7):1050-1066.

13. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM: *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature 2014, 505(7485):696-700.

14. Gu W, Zhou T, Wilke CO: A Universal Trend of Reduced mRNA Stability near the Translation-Initiation Site in Prokaryotes and Eukaryotes. PLoS Comput Biol 2010, 6(2):e1000664.

15. Kwok CK, Ding Y, Tang Y, Assmann SM, Bevilacqua PC: Determination of *in vivo* RNA structure in low-abundance transcripts. Nat Commun 2013, 4.

16. Kwok CK, Tang Y, Assmann SM, Bevilacqua PC: The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. Trends Biochem Sci 2015, 40(4):221-232.

17. Seo SW, Yang J-S, Kim I, Yang J, Min BE, Kim S, Jung GY: Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. Metab Eng 2013, 15:67-74.

18. Shabalina SA, Ogurtsov AY, Spiridonov NA: A periodic pattern of mRNA secondary structure created by the genetic code. Nucleic Acids Res 2006, 34(8):2428-2437.

19. Yin J, Bao LC, Tian H, Gao XD, Yao WB: Quantitative relationship between the mRNA secondary structure of translational initiation region and the expression level of heterologous protein in *Escherichia coli*. J Ind Microbiol Biotechnol 2016, 43(1):97-102.

20. Stern DB, Gruissem W: Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell 1987, 51(6):1145-1157.

21. Bally J, Nadai M, Vitel M, Rolland A, Dumain R, Dubald M: Plant physiological adaptations to the massive foreign protein synthesis occurring in recombinant chloroplasts. Plant Physiol 2009, 150(3):1474-1481.

22. Chakrabarti SK, Lutz KA, Lertwiriyawong B, Svab Z, Maliga P: Expression of the *cry9Aa2 B.t.* gene in tobacco chloroplasts confers resistance to potato tuber moth. Transgenic Res 2006, 15(4):481-488.

23. Dufourmantel N, Dubald M, Matringe M, Canard H, Garcon F, Job C, Kay E, Wisniewski JP, Ferullo JM, Pelissier B: Generation and characterization of soybean and marker-free tobacco plastid transformants over-expressing a bacterial 4-

hydroxyphenylpyruvate dioxygenase which provides strong herbicide tolerance. Plant Biotechnol J 2007, 5(1):118-133.

24. Gray BN, Ahner BA, Hanson MR: High-level bacterial cellulase accumulation in chloroplast-transformed tobacco mediated by downstream box fusions. Biotechnol Bioeng 2009, 102(4):1045-1054.

25. Gray BN, Yang H, Ahner BA, Hanson MR: An efficient downstream box fusion allows high-level accumulation of active bacterial beta-glucosidase in tobacco chloroplasts. Plant Mol Biol 2011, 76(3):345-355.

26. Petersen K, Bock R: High-level expression of a suite of thermostable cell wall-degrading enzymes from the chloroplast genome. Plant Mol Biol 2011, 76(3-5):311-321.

27. Zoschke R, Bock R: Chloroplast translation: structural and functional organization, operational control, and regulation. The Plant Cell 2018, 30(4):745-770.

28. Miertens N, Remaut E, Fiers W: Increased stability of phage T7g10 mRNA is mediated by either a 5'-or a 3'-terminal stem-loop structure. Biol Chem Hoppe-Seyler 1996, 377(12):811-818.

29. Sprengart ML, Fuchs E, Porter A: The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*. EMBO J 1996, 15(3):665.

30. Faxén M, Plumbridge J, lsaksson LA: Codon choice and potential complementarity between mRNA downstream of the initiation codon and bases 1471 –1480 in 16S ribosomal RNA affects expression of *glnS*. Nucleic Acids Res 1991, 19(19):5247-5251.

31. Kuroda H, Maliga P: Complementarity of the 16S rRNA penultimate stem with sequences downstream of the AUG destabilizes the plastid mRNAs. Nucleic Acids Res 2001, 29(4):970-975.

32. O'Connor M, Asai T, Squires CL, Dahlberg AE: Enhancement of translation by the downstream box does not involve base pairing of mRNA with the penultimate stem sequence of 16S rRNA. PNAS 1999, 96(16):8973-8978.

33. Watters KE, Abbott TR, Lucks JB: Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. Nucleic Acids Res 2016, 44(2):e12-e12.

34. Harris EH, Boynton JE, Gillham NW: Chloroplast ribosomes and protein synthesis. Microbiol Mol Biol Rev 1994, 58(4):700-754.

35. Gutell RR, Lee JC, Cannone JJ: The accuracy of ribosomal RNA comparative structure models. Curr Opin Struct Biol 2002, 12(3):301-310.

36. Seo SW, Yang J, Jung GY: Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and translational efficiency in *Escherichia coli*. Biotechnol Bioeng 2009, 104(3):611-616.

37. Kuroda H, Maliga P: Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. Plant Physiol 2001, 125(1):430-436.

38. Ye GN, Hajdukiewicz PT, Broyles D, Rodriguez D, Xu CW, Nehra N, Staub JM: Plastid-expressed 5-enolpyruvylshikimate-3-phosphate synthase genes provide high level glyphosate tolerance in tobacco. The Plant Journal 2001, 25(3):261-270.

39. Conley AJ, Zhu H, Le LC, Jevnikar AM, Lee BH, Brandle JE, Menassa R: Recombinant protein production in a variety of *Nicotiana* hosts: a comparative analysis. Plant Biotechnol J 2011, 9(4):434-444.

40. Bock R: Engineering plastid genomes: Methods, tools, and applications in basic research and biotechnology. Annu Rev Plant Biol 2015, 66(1):211-241.

41. Zuker M: Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003, 31(13):3406-3415.

42. Watters KE, Angela MY, Strobel EJ, Settle AH, Lucks JB: Characterizing RNA structures *in vitro* and *in vivo* with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Methods 2016, 103:34-48.

43. Maliga P, Tungsuchat-Huang T: Plastid Transformation in *Nicotiana tabacum* and *Nicotiana sylvestris* by Biolistic DNA Delivery to Leaves. In: *Chloroplast Biotechnology: Methods and Protocols.* Edited by Maliga P, vol. 1132. New Yorks: Springer Science and Business Media; 2014: 147-164.

CHAPTER 5

Conclusions and future prospects

This dissertation describes a two-part application-driven investigation to understand and enhance recombinant protein expression in chloroplast-transformed plants. Here, we have analyzed the physiological impacts of foreign protein accumulation (Chapters 2 and 3) as well as identified translational control mechanisms underlying high target protein yield (Chapter 4). This work builds upon previous research that established six chloroplast-transformed *Nicotiana tabacum* lines producing recombinant *cel6*A or *bglC*.

As a plant model organism and well-established high biomass crop, *N. tabacum* is an effective host organism for many high-value proteins. Furthermore, the field trial described in Chapter 2 demonstrated that chloroplast-transformed tobacco can maintain healthy growth in open field cultivation while still yielding foreign protein up to 25% of TSP. However, current tobacco cultivation strategies are focused on nicotine yields rather than protein quantity. Preliminary work to develop alternative cultivation strategies has shown that increasing planting density and the frequency of harvests increase biomass yields, but further work is needed to assess the protein content of plant tissue harvested in this strategy [1].

Furthermore, tobacco may not be the best host plant for all high value protein applications. Instead, non-toxic and edible plants may be better alternatives for oral vaccines and nutritive additives. Engineering a food crop or non-toxic algae could substantially reduce downstream processing, purification, and storage costs if the target protein was used directly from the host organism [2]. However, current limitations in chloroplast engineering techniques and differences in protein production capacity necessitate the need for more work to fine-tune other host organisms for high-value protein production.

Over-expression of any transgenic protein is a nutrient sink and places an added burden on the host plant's metabolism that, in turn, increases the plant's demand for resources. For example, although TetC-*cel6*A tobacco had no deleterious phenotypes when well-fertilized, these plants developed a conditional mutant phenotype when cultivated with sub-optimal ammonium nitrate. Here, we focused on nitrogen input because it is an expensive fertilizer, often limiting in field cultivation, and a major component of proteins. However, a broader investigation of nutrient demand and utilization in engineered plants may identify other inorganic resources that are also critical for transplastomic plant growth and foreign protein yield.

Carbon fixation is another critical component of plant metabolism, and growing concerns of climate change in recent years have driven research into the effects of elevated CO_2 on agricultural plant production. Although many of the aspects of C3 plant response to elevated CO2 are controversial, we showed in Chapter 3 that enhanced atmospheric CO_2 is a possible method of alleviating mutant growth phenotypes in engineered plants. Our goal in growing TetC-*cel6*A tobacco at 1600ppm CO_2 was that reallocation of protein resources away from photosynthetic machinery would rebalance expression along other critical pathways. Indeed, in addition to enhancing biomass production, we observed a significant reduction in Rubisco coincident with an increase in other native proteins and a doubling of Cel6A accumulation.

However, physiological response to recombinant protein production is not the only factor determining the success of an engineered plant line. The final goal of this dissertation, described in Chapter 4, was to establish a potential mechanism of function for the downstream box (DB) regulatory region. This work built off of previous experiments that demonstrated an interaction between the DB and the ORF fused to it. *In silico* mRNA secondary structure models predicted that the highest yielding DB-gene pairs, TetC-*cel6*A and NPTII-*bglC*, maintained the same

5'UTR fold predicted for the T7G10 5'UTR in its native gene setting. As a proof-of-concept, we also showed that synthetic DBs with unique nucleotide and amino acid sequences, but similar *in silico* predicted secondary structures to TetC-*cel6*A and NPTII-*cel6*A, yielded Cel6A in a pattern consistent with our hypothesis (Chapter 4). Additionally, Appendix I outlines initial work testing the efficacy of this technique to design DBs for other high-value proteins. Thus far, we have established chloroplast-transformed *Chlamydomonas reinhardtii* lines with synthetic DBs tailored to enhance yields for a bacterial phytase as well as a human immune cytokine.

One caveat of this method is that it designs only for DBs that improve foreign protein yield as compared to randomly selecting a DB or omitting the DB entirely. Absolute recombinant protein accumulation is dependent on many other interconnected factors and will require more protein-specific optimization.

Additionally, at the molecular scale, this body of work focused on mRNA translation and did not consider potential post-translational of impacts the DB on protein folding and function. Factors inherent to the specific DB, like amino acid hydrophobicity or charge, may disrupt protein folding. Activity assays of the original GFP, TetC, and NPTII DBs confirmed that the recombinant Cel6A and BglC were active, although the new synthetic Cel6A DBs have not been tested. The engineered phytase lines currently under development are in the process of being analyzed for enzymatic activity.

Thus, this dissertation highlights the inter-connected complexity of higher plant protein metabolism. We have identified some components of plant physiology, resource use, and molecular-scale regulators that may help enhance foreign protein accumulation while reducing mutant phenotypes.

References

1. Scott GL, Warren JF: Tobacco production system. In., vol. US9462754B2. USA: Universal Leaf Tobacco Co Inc 2016.

2. Schillberg S, Twyman RM, Fischer R: Opportunities for recombinant antigen and antibody expression in transgenic plants—technology assessment. Vaccine 2005, 23(15):1764-1769.

APPENDIX I

Tailoring synthetic regulators for enhanced accumulation of recombinant proteins in chloroplast-engineered *Chlamydomonas reinhardtii*

Introduction

Chloroplast genetic engineering was first established in the freshwater microalgae *Chlamydomonas reinhardtii* in 1988 when researchers used plasmids containing *atp*B to restore photosynthesis in heterotrophic *atp*B knockout mutants [1]. Since then, chloroplast engineering has been used in reverse-genetic studies to better understand the plastome and plastid gene expression through site-directed mutations and gene knockouts (reviewed in [2]). More recently, though, research has shifted focus to applied uses for chloroplast engineering in the production of high-value proteins, like biopharmaceuticals, nutritional additives, and industrial enzymes (reviewed in [3]). *C. reinhardtii* presents a promising production system for high-value proteins because is amenable to phototrophic or heterotrophic growth in controlled batch cultures. *C. reinhardtii* is also generally regarded as safe, enabling the production of oral vaccines [4]. In fact, because *C. reinhardtii* is capable of folding complex proteins and forming disulfide bonds, it is a viable platform for many biopharmaceuticals [5]. To date, *C. reinhardtii* has been used successfully to produce more than fifty bioactive human medical proteins, including biopharmaceuticals to treat anthrax, foot and mouth disease, herpes, as well as a number of anti-cancer treatments (reviewed in [6]).

Although microalgae show promise, this technology is hindered by low heterologous protein yields. Chloroplast-engineered microalgae do not often exceed foreign protein yields of 5% of total soluble protein (TSP) with averages varying around 0.5-2% of TSP [6]. For all but

the most valuable recombinant proteins, such yields are likely too low for economical mass-production of these engineered lines, necessitating a focus on improving foreign protein yields.

In recent years, the 5'UTR has been the target of much of this work for its profound control of translation initiation and ribosome binding. 5'UTRs have been identified as targets of RNA-binding proteins that stabilize the transcripts and facilitate ribosome binding (reviewed in [7]). In addition, mRNA secondary structure, created by the base-pairing patterns of the ribonucleotides, is responsible for much of the regulatory function of the 5'UTR [8]. For instance, in many cases, a thermodynamically weak base-pairing pattern around the start codon and Shine-Dalgarno (SD) sequence enables easier ribosome binding and higher translation [9]. In particular, the native *rbcL*, *psbA*, *psbD*, and *atpA* promoter-5'UTRs combinations have all been used with modest success in chloroplast engineering projects [5].

The downstream box (DB), named for its location just downstream of the start codon has also been identified for its strong role in regulating protein accumulation in bacteria and plant chloroplast [10, 11]. In fact, many engineering projects have attempted to harness the power of the DB to drive higher accumulation of their target protein with varying success [10, 12-18]. For example, early work in *Nicotiana tabacum* demonstrated that a bacterial cellulase (Cel6A) fused with the DB derived from the *tetC* gene yielded significantly greater Cel6A protein than the DB from the *nptII* gene [12]. Furthermore, this expression pattern was conserved in transplastomic *C. reinhardtii,* suggesting that the mechanism of the DB's function for transgenic gene expression is conserved between algae and higher plants [17].

Recent *in vivo* mRNA secondary structure work with these *cel6*A tobacco lines have identified patterns between DB-induced mRNA folding and heterologous protein accumulation. Briefly, in the highest yielding tobacco line, TetC-*cel6*A, the T7G10 5'UTR most often had the

same folding pattern as that predicted for its native transcript. If the function of the 5'UTR is determined by its secondary structure, this suggests that a DB may promote high recombinant protein yields by preventing the transgenic ORF from interfering with the native base-pairing pattern of the 5'UTR.

Here, we further test this concept by designing a set of DBs for two different transgenic proteins of interest, a biopharmaceutical human immune system cytokine, Interluekin-2 (IL2), and a bacterial phytase for nutritional quality improvement of livestock feed.

Preliminary Results

Promoting phytase synthesis by adding a Shine-Darlgarno sequence and reducing base-pair interactions around the translation initiation region

The synthetic DBs described here were created to test different aspects of mRNA regulatory control. First, the goal behind designing the phytase DBs was to increase translation by improving the accessibility of the start codon for ribosome binding. *In silico* mRNA secondary structure models of the original phytase construct (Figure AI.1a &2a) predict that the start codon mostly is obscured in base pairing along a stem. Because the ribosome only has a helicase activity when it is fully assembled, the procession of the preinitiation complex is controlled by the stability and degree of folding around the start codon [19]. Thus, in many cases, weakly structured translation initiation regions more often result in higher accumulation of the encoded protein [9, 20, 21]. To address this feature, we designed two DBs, DB1 and DB2, tailored to the phytase ORF that are predicted to leave the start codon unpaired (Figure AI.2, c, e).

These two DBs achieve approximately the same goal but do so through different specific secondary structures. For instance, DB1 contains an inverted repeat that enables it to pair well

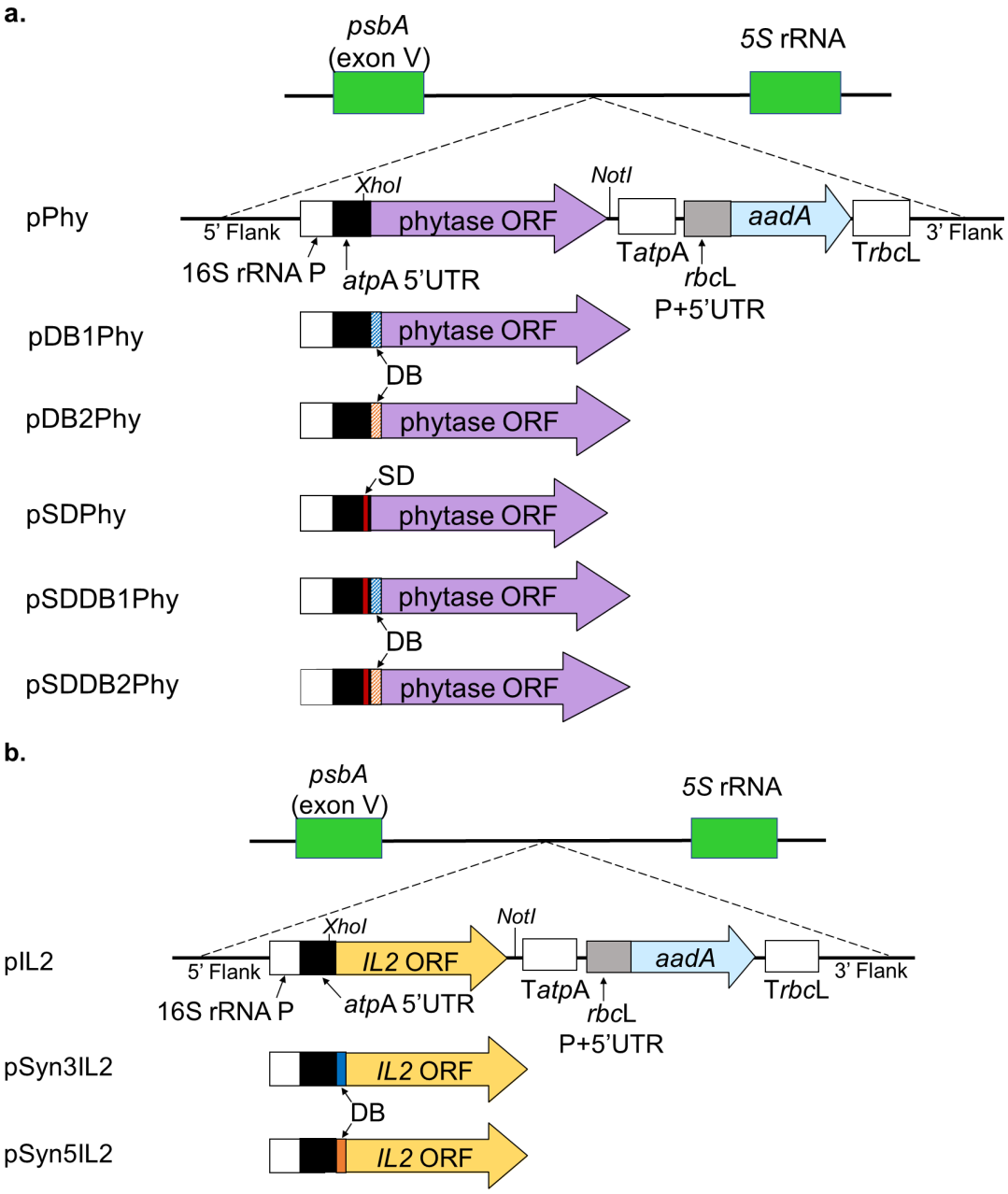


Figure AI.1: Vector schematic of the algae chloroplast constructs. **a.** the original phytase construct with variants including the addition of DBs (light blue or light orange) and/or SD (dark red). **b.** IL2 constructs without a DB or with the addition of Syn5 (dark blue) and Syn3 (orange) DBs. Both ORFs are under control of the 16S rRNA promoter as well as the *apt*A 5' and 3'UTRs. Spectinomycin resistance was used as the selectable marker via the *aadA* gene. All constructs were targeted for insertion between the *psbA* exon V and 5S rRNA.

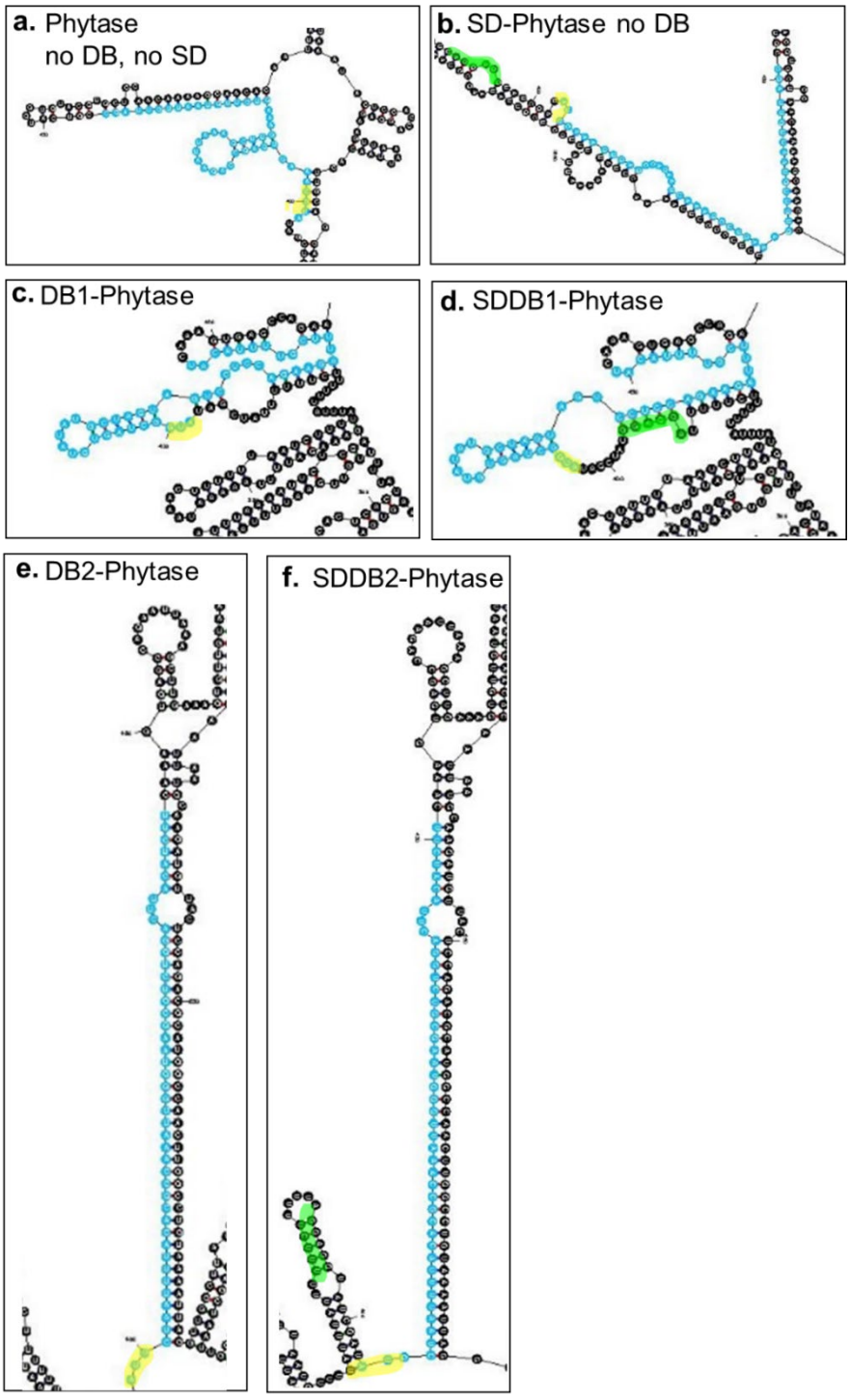


Figure AI.2: Predicted *in silico* mRNA secondary structure of the translation initiation region of the phytase constructs: **a.** phytase without an SD or DB, **b.** phytase with an SD, **c.** phytase with DB1 only, **d.** phytase with DB1 and an SD, **e.** phytase with DB2 only, and **f.** phytase with DB2 and SD.

with itself. Further downstream, DB1 also pairs with the 5'UTR. In contrast, DB2 exclusively forms base pairs with the phytase ORF, leaving the 5'UTR free to fold on its own. These differences in base-pairing interactions help test if translation initiation is aided simply by making the start codon more accessible or if other neighboring structural elements also influence translation regulation. For example, the start codon of the DB1-phytase mRNA is located in a bulge in the middle of a hairpin while the start codon in the DB2-phytase transcript is located in a single-stranded stretch between two hairpins. It is possible that the differences in these locations impacts ribosome assembly and early translation.

In addition to the DBs, we also tested whether an SD sequence would help increase translation of the exogenous phytase mRNA. Early work has yielded conflicting results over the importance of the SD sequence in plastid gene regulation in *C. reinhardtii* [22, 23]. Thus, we included plasmids containing the SD consensus sequence (AGGAGG) [24] 8nt upstream of the start codon to test the usefulness of this element in promoting translation in exogenous mRNA (Figure AI.2 b, d, &f).

Designing DBs to promote native 5'UTR folding patterns for transgenic IL2

The goal of designing the IL2 DBs focused less on the structure of the DB and, instead, sought to prevent the IL2 ORF from interacting with the *atp*A 5'UTR (Figure AI.1b). This goal was derived from observations in tobacco that the DB of the highest yielding *cel6*A line, containing the TetC DB, paired predominantly with the beginning of the *cel6*A ORF, allowing the T7G10 5'UTR to fold independently. Meanwhile, the lower yielding NPTII DB was folded internally with itself, allowing the 5'end of the *cel6*A ORF to base-pair with the 5'UTR. Furthermore, transcriptome analyses of mRNA *in vivo* secondary structure demonstrated that many native transcripts fold modularly with each domain (i.e. 5'UTR, ORF, and 3'UTR) base-pairing

internally [25]. This modular folding may be particularly important for prokaryotes and prokaryotic organelles like the plastid where translation occurs co-transcriptionally. In these scenarios, the 5'UTR emerges from the RNA polymerase first, enabling ribosomes to begin assembling around the SD and start codon. If the 5'UTR's function relied on its base-pairing with a region of the transcript very far downstream, translation would be disrupted until the full mRNA was synthesized and folded properly.

Thus, the synthetic DBs designed for IL2 test the hypothesis that a synthetic DB may function as a buffer to prevent the transgenic ORF from interrupting the mRNA folding of a non-native 5'UTR in chimeric heterologous constructs (Figure AI.3).

Conclusion and Future Outlook

In this report, we have begun to test varying methods of designing DBs tailored to increase expression of high-value proteins of different origins and functions. For instance, phytase is an important additive to improve the bioavailability of phosphorous in animal feed [26]. Meanwhile, the IL2 cytokine studied here is under testing for its use in cancer treatments [27].

Characterization of these transgenic lines is on-going. Planned experiments include immunoblots to quantify heterologous protein yields, qPCR to identify transgenic mRNA abundance, and growth curves to assess algal health. We are also optimizing a procedure to measure phytase enzymatic activity from crude cellular extracts as a means of minimizing processing costs.

Future work with this project may include refining these synthetic DBs even further. For instance, the IL2 DBs were created to be very stable to ensure that the desired structure was the most prominent fold for the transcripts. However, highly thermodynamically stable mRNA

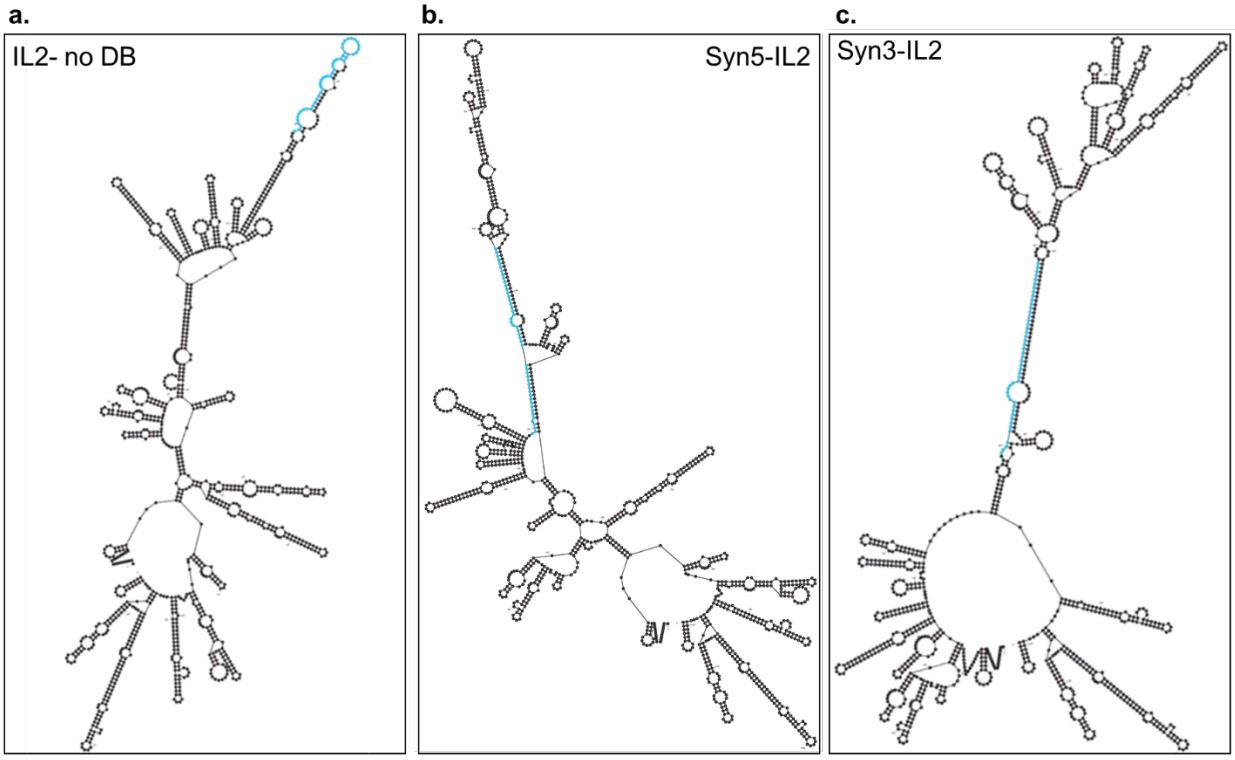


Figure AI.3: Predicted *in silico* mRNA secondary structures for the transgenic IL2 transcripts. **a.** IL2 without a DB. **b.** IL2 with the addition of the Syn5 DB. **c.** IL2 with the addition of the Syn3 DB. The first 48nt beginning with the start codon (i.e. the DB regions) are colored in blue.

secondary structure around the translation initiation region often reduces protein yields [28]. Thus, improving these DBs to reduce thermodynamic stability while maintaining the desired structure may increase IL2 yield.

Finally, this project was designed to be a proof-of-concept experiment testing our ability to predictably design synthetic DBs that improve foreign protein production. However, optimizing DBs alone may not be enough to make transgenic algae a cost-effective protein production platform. Additional work enhancing other aspects of plastid protein regulation like more species-specific codon optimization or the inclusion of adjuncts to improve protein stability may be necessary to further increase heterologous protein accumulation to economically viable levels [29]. It may also be worthwhile to transfer these constructs and principles to a higher yielding host organism, like the biomass crop *N. tabacum*. Foreign protein accumulation of 5-20% of TSP is common in plastid transformed tobacco which we expect would improve the feasibility of producing IL2 or phytase for large-scale applications [12, 13, 30, 31].

Methods

Construct design and plasmid construction

Transformation plasmids for the phytase-expressing algae were also constructed using the pCHR74 plasmid from Richter *et* al 2018 encoding TetC-*cel6*A as a template. The region containing the TetC DB and *cel6*A ORF was excised, using double digestion at the XhoI and NotI restriction enzyme sites (Figure AI.1a). The resulting plasmid contained the selectable marker and all of the necessary regulatory elements for transgene expression in the chloroplast of *C. reinhardtii*. The phytase gene used here was derived from the bacterial AppA2 enzyme (NCBI WP_094337278.1) with three amino acid substitutions to improve thermostability of the enzyme [32]. The nucleotide sequence of this enzyme was also codon optimized to match the *C.*

reinhardtii plastid bias. This phytase ORF was double-digested and ligated into the plasmid between the *atp*A 5'UTR and *atp*A 3'UTR, creating pPhy (Figure AI.1a). The SD sequence was added to pPhy using site-directed mutagenesis (SDM) primers SD_SDM_fwd/rev to create the plasmid pSDPhy. The DBs were inserted into pPhy using SDM insertion primers DB1_SDM_fwd/rev or DB2_SDM_fwd/rev to create pDB1Phy and pDB2Phy plasmids, respectively. These plasmids were then used as templates to add SD sequences using DB-specific forward primers SD-DB1_SDM_fwd and SD-DB2_SDM_fwd and the phy_SDM_rev reverse primer. These six new plasmids were all transformed into chemically competent DH5-alpha *E. coli* for storage and amplification. All primers used for these transformations can be found in Table AI.S1.

Transformation plasmids for the IL2-expressing algae were also constructed using the pCHR74 plasmid from Richter et al 2018 encoding TetC-*cel6*A as a template. As with the phytase plasmids, the plasmid region containing the TetC DB and *cel6*A ORF was excised via double digestion. We next codon-optimized and removed introns from the human IL2 ORF (Genbank AAA59140.1). The IL2 ORF was then PCR amplified using the IL2_insert_fwd and IL2_insert_rev primers and double-digested at the same restriction enzymes sites noted above. The digested IL2 gene and plasmid were ligated together, creating the pIL2. The DBs were inserted using SDM insertion primers Syn3-IL2_fwd/rev or Syn5-IL2_fwd/rev to create pSyn3IL2 and pSyn5IL2 plasmids, respectively (Figure AI.1b). These plasmids were also transformed into chemically competent DH5-alpha *E. coli*. All primers used for these transformations can be found in Table AI.S2.

C. reinhardtii transformation

The cell-wall mutant *C. reinhardtii* strain CR4349 was transformed via glass bead method. Cells in exponential growth phase were harvested and resuspended in 2mL TAP media (20 mM Tris, 17 mM Acetate, 0.68 mM K_2HPO_4, 7.26 mM KH_2PO_4, and 7.5 mM NH_4Cl) [33]. The transformation mixture consisting of 300 uL concentrated algae culture, 5ug plasmid DNA, and 300 mg of sterile 0.5 mm diameter glass beads was vortexed on high speed for 20 seconds. This mixture was then spread on TAP agar plates containing 0.15 mg · ml^{-1} spectinomycin. This process was repeated for a total of seven times for each construct. These plates were grown in low light for 24 hr before light intensity was increased. After approximately three weeks, seven colonies were selected from each construct and transferred to liquid TAP media containing 0.15 mg · ml^{-1} spectinomycin. All analyses were conducted after at least four rounds of selection on liquid TAP media containing 0.15 mg · ml^{-1} spectinomycin.

Supplementary Materials

Table AI.S1: Phytase transformation plasmid construction primers. Lowercase letters denote nucleotides added via site-directed mutagenesis

Plasmid Name	Primer Name	Sequence (5'-3')
pDB1Phy	DB1_SDM_fwd	tccagaagatttcgttttacatCAAAGTGAGCCAGAATTAAAG
	DB1_SDM_rev	ggactactagcatatacactagcCATATCGATAAAAAAGAAAAAATAAATAAAAG
pDB2Phy	DB2_SDM_fwd	taagcgtctggacttacatcttCAAAGTGAGCCAGAATTAAAG
	DB2_SDM_rev	cccaatttggcgtgtaaaattagCATATCGATAAAAAAGAAAAAATAAATAAAAG
pSDPhy	SD_SDM_fwd	aggaggTATCGATATGCAAAGTGAG
	SD_SDM_rev	AAAAAGAAAAAATAAATAAAAGATTAAAAAAG
pSDDB1Phy	SD-DB1_SDM_fwd	aggaggTATCGATATGGCTAGTGTATATG
	phy_SDM_rev	AAAAAGAAAAAATAAATAAAAGATTAAAAAAG
pSDDB2Phy	SD-DB2_SDM_fwd	aggaggTATCGATATGCTAATTTTACACGC
	phy_SDM_rev	AAAAAGAAAAAATAAATAAAAGATTAAAAAAG

Table AI.S2: IL2 transformation plasmid construction primers. Lowercase letters denote nucleotides added via site-directed mutagenesis

Plasmid Name	**Primer Name**	**Sequence (5'-3')**
pSyn5IL2	Syn5-IL2_fwd	accttcagttctgtggccttctGCACCAACTAGCAGCAGC
	Syn5-IL2_rev	ttcatcaaagcaaaaacttgcCATATCGATAAAAAAGAAAAAATAAATAAAAGATTAAAAAAGTTTATTTTTAAAATC
pSyn3IL2	Syn3-IL2_fwd	gctttgacaaaaggtaatccatctGCACCAACTAGCAGCAGC
	Syn3-IL2_rev	atcatcaccgtcctgacttgcCATATCGATAAAAAAGAAAAAATAAATAAAAGATTAAAAAAGTTTATTTTTAAAATC

References

1. Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM: Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. Science 1988, 240(4858):1534.
2. Purton S: Tools and techniques for chloroplast transformation of *Chlamydomonas*. In: *Transgenic microalgae as green cell factories.* Springer; 2007: 34-45.
3. Bock R: Engineering plastid genomes: Methods, tools, and applications in basic research and biotechnology. Annu Rev Plant Biol 2015, 66(1):211-241.
4. Dreesen IA, Charpin-El Hamri G, Fussenegger M: Heat-stable oral alga-based vaccine protects mice from *Staphylococcus aureus* infection. J Biotechnol 2010, 145(3):273-280.
5. Specht E, Miyake-Stoner S, Mayfield S: Micro-algae come of age as a platform for recombinant protein production. Biotechnol Lett 2010, 32(10):1373-1383.
6. Gong Y, Hu H, Gao Y, Xu X, Gao H: Microalgae as platforms for production of recombinant proteins and valuable compounds: progress and prospects. J Ind Microbiol Biotechnol 2011, 38(12):1879-1890.
7. Barkan A, Small I: Pentatricopeptide Repeat Proteins in Plants. Annu Rev Plant Biol 2014, 65(1):415-442.
8. Seo SW, Yang J-S, Kim I, Yang J, Min BE, Kim S, Jung GY: Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. Metab Eng 2013, 15:67-74.
9. De Smit MH, Van Duin J: Control of prokaryotic translational initiation by mRNA secondary structure. In: *Prog Nucleic Acid Res Mol Biol.* vol. 38: Elsevier; 1990: 1-35.

10. Kuroda H, Maliga P: Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. Plant Physiol 2001, 125(1):430-436.

11. Sprengart ML, Fuchs E, Porter A: The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*. EMBO J 1996, 15(3):665.

12. Gray BN, Ahner BA, Hanson MR: High-level bacterial cellulase accumulation in chloroplast-transformed tobacco mediated by downstream box fusions. Biotechnol Bioeng 2009, 102(4):1045-1054.

13. Gray BN, Yang H, Ahner BA, Hanson MR: An efficient downstream box fusion allows high-level accumulation of active bacterial beta-glucosidase in tobacco chloroplasts. Plant Mol Biol 2011, 76(3):345-355.

14. Herz S, Füßl M, Steiger S, Koop H-U: Development of novel types of plastid transformation vectors and evaluation of factors controlling expression. Transgenic Res 2005, 14(6):969-982.

15. Kolotilin I, Kaldis A, Pereira EO, Laberge S, Menassa R: Optimization of transplastomic production of hemicellulases in tobacco: effects of expression cassette configuration and tobacco cultivar used as production platform on recombinant protein yields. Biotechnol Biofuels 2013, 6(1):1.

16. Lenzi P, Scotti N, Alagna F, Tornesello ML, Pompa A, Vitale A, De Stradis A, Monti L, Grillo S, Buonaguro FM: Translational fusion of chloroplast-expressed human papillomavirus type 16 L1 capsid protein enhances antigen accumulation in transplastomic tobacco. Transgenic Res 2008, 17(6):1091-1102.

17. Richter LV, Yang H, Yazdani M, Hanson MR, Ahner BA: A downstream box fusion allows stable accumulation of a bacterial cellulase in *Chlamydomonas reinhardtii* chloroplasts. Biotechnol Biofuels 2018, 11(1):133.

18. Zhou F, Badillo-Corona JA, Karcher D, Gonzalez-Rabade N, Piepenburg K, Borchers AM, Maloney AP, Kavanagh TA, Gray JC, Bock R: High-level expression of human immunodeficiency virus antigens from the tobacco and tomato plastid genomes. Plant Biotechnol J 2008, 6(9):897-913.

19. Zoschke R, Bock R: Chloroplast translation: structural and functional organization, operational control, and regulation. The Plant Cell 2018, 30(4):745-770.

20. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM: *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature 2014, 505(7485):696-700.

21. Yin J, Bao LC, Tian H, Gao XD, Yao WB: Quantitative relationship between the mRNA secondary structure of translational initiation region and the expression level of heterologous protein in *Escherichia coli*. J Ind Microbiol Biotechnol 2016, 43(1):97-102.

22. Fargo D, Zhang M, Gillham N, Boynton J: Shine-Dalgarno-like sequences are not required for translation of chloroplast mRNAs in *Chlamydomonas reinhardtii* chloroplasts or in *Escherichia coli*. Molecular and General Genetics MGG 1998, 257(3):271-282.

23. Sakamoto W, Chen X, Kindle KL, Stern DB: Function of the *Chlamydomonas reinhardtii petD* 5'untranslated region in regulating the accumulation of subunit IV of the cytochrome b_6/f complex. Plant J 1994, 6(4):503-512.

24. Shine J, Dalgarno L: Determinant of cistron specificity in bacterial ribosomes. Nature 1975, 254(5495):34-38.

25. Shabalina SA, Ogurtsov AY, Spiridonov NA: A periodic pattern of mRNA secondary structure created by the genetic code. Nucleic Acids Res 2006, 34(8):2428-2437.

26. Lei XG, Porres JM: Phytase enzymology, applications, and biotechnology. Biotechnol Lett 2003, 25(21):1787-1794.

27. Mahmoudpour SH, Jankowski M, Valerio L, Becker C, Espinola-Klein C, Konstantinides S, Quitzau K, Barco S: Safety of low-dose subcutaneous recombinant interleukin-2: systematic review and meta-analysis of randomized controlled trials. Scientific reports 2019, 9(1):1-9.

28. Seo SW, Yang J, Jung GY: Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and translational efficiency in *Escherichia coli*. Biotechnol Bioeng 2009, 104(3):611-616.

29. Elghabi Z, Karcher D, Zhou F, Ruf S, Bock R: Optimization of the expression of the HIV fusion inhibitor cyanovirin-N from the tobacco plastid genome. Plant Biotechnol J 2011, 9(5):599-608.

30. Bally J, Nadai M, Vitel M, Rolland A, Dumain R, Dubald M: Plant physiological adaptations to the massive foreign protein synthesis occurring in recombinant chloroplasts. Plant Physiol 2009, 150(3):1474-1481.

31. Castiglia D, Sannino L, Marcolongo L, Ionata E, Tamburino R, De Stradis A, Cobucci-Ponzano B, Moracci M, La Cara F, Scotti N: High-level expression of thermostable cellulolytic enzymes in tobacco transplastomic plants and their use in hydrolysis of an industrially pretreated *Arundo donax* L. biomass. Biotechnol Biofuels 2016, 9(1):154.

32. Kim M-S, Weaver JD, Lei XG: Assembly of mutations for improving thermostability of *Escherichia coli* AppA2 phytase. Appl Microbiol Biotechnol 2008, 79(5):751.

33. Togasaki RK, Brunke K, Kitayama M, Griffith OM: Isolation of intact chloroplasts from *Chlamydomonas reinhardtii* with Beckman centrifugal elutriation system. In: *Progress in photosynthesis research.* Springer; 1987: 499-502.

APPENDIX II

Summary of statistics for all data analyses

Two-tailed Student's t-test p-values calculated assuming alpha = 0.05 and unequal variance unless otherwise noted.

Table AII.1: Summary of field trial statistics

Test Parameter	Comparison		Variance	p-value
	Condition	Genotype		
Biomass	Year 1	WT vs TetC-*cel6*A	equal	0.789
g plant^{-1}		WT vs NPTII-*cel6*A	equal	0.277
	Year 2	WT vs TetC-*cel6*A	unequal	0.462
		WT vs NPTII-*cel6*A	equal	0.993
	Year 1 vs Year 2	WT	unequal	0.022
		TetC-*cel6*A	equal	0.093
		NPTII-*cel6*A	equal	0.002
V_{cmax}	Chamber	WT vs TetC-*cel6*A	equal	0.253
µmol m^{-2} s^{-1}		WT vs NPTII-*cel6*A	unequal	0.206
	Year 1	WT vs TetC-*cel6*A	equal	0.001
		WT vs NPTII-*cel6*A	equal	0.034
	Year 2	WT vs TetC-*cel6*A	equal	0.393
		WT vs NPTII-*cel6*A	equal	0.385
	Chamber vs Year 1	WT	unequal	0.322
	Chamber vs Year 2	WT	unequal	0.742
	Year 1 vs Year 2	WT	unequal	0.003
	Chamber vs Year 1	TetC-*cel6*A	unequal	0.103
	Chamber vs Year 2	TetC-*cel6*A	equal	0.047
	Year 1 vs Year 2	TetC-*cel6*A	unequal	0.852
	Chamber vs Year 1	NPTII-*cel6*A	equal	0.007
	Chamber vs Year 2	NPTII-*cel6*A	equal	0.206
	Year 1 vs Year 2	NPTII-*cel6*A	equal	0.042
J_{max}	Chamber	WT vs TetC-*cel6*A	equal	0.555
µmol m^{-2} s^{-1}		WT vs NPTII-*cel6*A	equal	0.542
	Year 1	WT vs TetC-*cel6*A	equal	0.001
		WT vs NPTII-*cel6*A	equal	0.007
	Year 2	WT vs TetC-*cel6*A	equal	0.421

		WT vs NPTII-*cel6*A	equal	0.430
	Chamber vs Year 1	WT	unequal	0.266
	Chamber vs Year 2	WT	equal	0.318
	Year 1 vs Year 2	WT	unequal	0.001
	Chamber vs Year 1	TetC-*cel6*A	unequal	0.850
	Chamber vs Year 2	TetC-*cel6*A	equal	0.714
	Year 1 vs Year 2	TetC-*cel6*A	unequal	0.826
	Chamber vs Year 1	NPTII-*cel6*A	unequal	0.590
	Chamber vs Year 2	NPTII-*cel6*A	equal	0.507
	Year 1 vs Year 2	NPTII-*cel6*A	equal	0.092
TSP	Chamber	WT vs TetC-*cel6*A	equal	0.609
g m^{-2}		WT vs NPTII-*cel6*A	equal	0.905
	Year 1	WT vs TetC-*cel6*A	equal	0.018
		WT vs NPTII-*cel6*A	equal	0.040
	Year 2	WT vs TetC-*cel6*A	equal	0.452
		WT vs NPTII-*cel6*A	equal	0.527
	Chamber vs Year 1	WT	equal	0.580
	Chamber vs Year 2	WT	equal	0.094
	Year 1 vs Year 2	WT	equal	0.134
	Chamber vs Year 1	TetC-*cel6*A	equal	0.002
	Chamber vs Year 2	TetC-*cel6*A	equal	0.008
	Year 1 vs Year 2	TetC-*cel6*A	unequal	0.309
	Chamber vs Year 1	NPTII-*cel6*A	equal	0.021
	Chamber vs Year 2	NPTII-*cel6*A	equal	0.383
	Year 1 vs Year 2	NPTII-*cel6*A	equal	0.183
Rubisco	Chamber	WT vs TetC-*cel6*A	unequal	0.990
% of TSP		WT vs NPTII-*cel6*A	unequal	0.135
	Year 1	WT vs TetC-*cel6*A	unequal	0.379
		WT vs NPTII-*cel6*A	unequal	0.161
	Year 2	WT vs TetC-*cel6*A	equal	0.418
		WT vs NPTII-*cel6*A	unequal	0.342
	Chamber vs Year 1	WT	unequal	0.132
	Chamber vs Year 2	WT	unequal	0.439
	Year 1 vs Year 2	WT	equal	0.067

	Chamber vs Year 1	TetC-*cel6*A	equal	0.962
	Chamber vs Year 2	TetC-*cel6*A	equal	0.753
	Year 1 vs Year 2	TetC-*cel6*A	equal	0.728
	Chamber vs Year 1	NPTII-*cel6*A	unequal	0.674
	Chamber vs Year 2	NPTII-*cel6*A	unequal	0.700
	Year 1 vs Year 2	NPTII-*cel6*A	unequal	0.586
Rubisco	Chamber	WT vs TetC-*cel6*A	equal	0.936
g m^{-2}		WT vs NPTII-*cel6*A	equal	0.981
	Year 1	WT vs TetC-*cel6*A	unequal	0.121
		WT vs NPTII-*cel6*A	unequal	0.078
	Year 2	WT vs TetC-*cel6*A	equal	0.836
		WT vs NPTII-*cel6*A	equal	0.766
	Chamber vs Year 1	WT	equal	0.740
	Chamber vs Year 2	WT	equal	0.108
	Year 1 vs Year 2	WT	equal	0.058
	Chamber vs Year 1	TetC-*cel6*A	equal	0.138
	Chamber vs Year 2	TetC-*cel6*A	equal	0.210
	Year 1 vs Year 2	TetC-*cel6*A	equal	0.604
	Chamber vs Year 1	NPTII-*cel6*A	equal	0.065
	Chamber vs Year 2	NPTII-*cel6*A	equal	0.191
	Year 1 vs Year 2	NPTII-*cel6*A	equal	0.246
Cel6A	Chamber	TetC-*cel6*A vs NPTII-*cel6*A	unequal	2.72E-05
% of TSP	Year 1	TetC-*cel6*A vs NPTII-*cel6*A	unequal	0.011
	Year 2	TetC-*cel6*A vs NPTII-*cel6*A	unequal	0.006
	Chamber vs Year 1	TetC-*cel6*A	equal	0.002
	Chamber vs Year 2	TetC-*cel6*A	equal	0.0002
	Year 1 vs Year 2	TetC-*cel6*A	unequal	0.360
	Chamber vs Year 1	NPTII-*cel6*A	equal	0.555
	Chamber vs Year 2	NPTII-*cel6*A	unequal	2.53E-05
	Year 1 vs Year 2	NPTII-*cel6*A	unequal	0.001
Cel6A	Chamber	TetC-*cel6*A vs NPTII-*cel6*A	unequal	0.0001
g m^{-2}	Year 1	TetC-*cel6*A vs NPTII-*cel6*A	unequal	0.010
	Year 2	TetC-*cel6*A vs NPTII-*cel6*A	unequal	0.013

	Chamber vs Year 1	TetC-*cel6*A	unequal	0.250
	Chamber vs Year 2	TetC-*cel6*A	equal	0.006
	Year 1 vs Year 2	TetC-*cel6*A	unequal	0.265
	Chamber vs Year 1	NPTII-*cel6*A	equal	0.083
	Chamber vs Year 2	NPTII-*cel6*A	unequal	0.001
	Year 1 vs Year 2	NPTII-*cel6*A	unequal	0.003
Averaged weekly PAR µmol m^{-2} day^{-1}	Year 1 vs Year 2	N/A		
	week 1		equal	0.005
	week 2		equal	0.002
	week 3		unequal	0.006
	week 4		equal	0.001
	week 5		equal	0.216
	week 6		equal	0.428
	week 7		equal	0.772
	week 8		equal	0.053
	week 9		unequal	0.743
Averaged weekly daytime temperature °C	Year 1 vs Year 2	N/A		
	week 1		unequal	9.10E-03
	week 2		equal	4.80E-09
	week 3		unequal	5.00E-71
	week 4		equal	6.70E-46
	week 5		equal	3.80E-44
	week 6		equal	2.40E-31
	week 7		equal	4.80E-54
	week 8		unequal	1.10E-77
	week 9		unequal	1.40E-24
Total Rainfall (cm $week^{-1}$)	Total	N/A	equal	0.77

Table AII.2: Summary of age-dependent growth and Cel6A trial statistics

Parameter	Treatment	Comparison		p-value
Fresh Weight	2 weeks	WT	TetC-*cel6*A	0.0553
g plant^{-1}	6 weeks	WT	TetC-*cel6*A	0.2828
	9 weeks	WT	TetC-*cel6*A	0.428
	12 weeks	WT	TetC-*cel6*A	0.4155
	TetC-*cel6*A	2 weeks	6 weeks	0.0107
		6 weeks	9 weeks	<.0001
		9 weeks	12 weeks	0.4874
	WT	2 weeks	6 weeks	<.0001
		6 weeks	9 weeks	<.0001
		9 weeks	12 weeks	0.1737
Dry Weight	2 weeks	WT	TetC-*cel6*A	1
g plant^{-1}	6 weeks	WT	TetC-*cel6*A	0.2702
	9 weeks	WT	TetC-*cel6*A	0.1736
	12 weeks	WT	TetC-*cel6*A	0.2657
TSP	2 weeks	WT	TetC-*cel6*A	0.9318
mg plant^{-1}	6 weeks	WT	TetC-*cel6*A	0.4048
	9 weeks	WT	TetC-*cel6*A	0.3948
	12 weeks	WT	TetC-*cel6*A	0.6022
Rubisco	2 weeks	WT	TetC-*cel6*A	0.0285
% TSP	6 weeks	WT	TetC-*cel6*A	0.7854
	9 weeks	WT	TetC-*cel6*A	0.2143
	12 weeks	WT	TetC-*cel6*A	0.7694
	TetC-*cel6*A	2 weeks	6 weeks	0.9622
		6 weeks	9 weeks	0.7011
		9 weeks	12 weeks	0.0811
	WT	2 weeks	6 weeks	0.0458
		6 weeks	9 weeks	0.9856
		9 weeks	12 weeks	0.0109
Rubisco	2 weeks	WT	TetC-*cel6*A	0.575
mg plant^{-1}	6 weeks	WT	TetC-*cel6*A	0.4135
	9 weeks	WT	TetC-*cel6*A	0.2565
	12 weeks	WT	TetC-*cel6*A	0.5953
	TetC-*cel6*A	2 weeks	6 weeks	0.7725
		6 weeks	9 weeks	0.0146
		9 weeks	12 weeks	0.0509
	WT	2 weeks	6 weeks	0.2622
		6 weeks	9 weeks	0.0003

		9 weeks	12 weeks	0.0003
Cel6A	TetC-*cel6*A	2 weeks	6 weeks	0.0098
% TSP		6 weeks	9 weeks	0.2425
		9 weeks	12 weeks	0.4995
Cel6A	TetC-*cel6*A	2 weeks	6 weeks	0.8597
mg plant^{-1}		6 weeks	9 weeks	0.0037
		9 weeks	12 weeks	0.1924
Cel6A	TetC-*cel6*A	2 weeks	6 weeks	0.0003
mg mg^{-1}		6 weeks	9 weeks	0.494
		9 weeks	12 weeks	0.003
Stem length	27 days old	WT	TetC-*cel6*A	0.9196
cm	40 days old	WT	TetC-*cel6*A	0.1744
	54 days old	WT	TetC-*cel6*A	<.0001
	61 days old	WT	TetC-*cel6*A	<.0001
	68 days old	WT	TetC-*cel6*A	<.0001
	86 days old	WT	TetC-*cel6*A	0.0048

Table AII.3: Summary of statistics for effect of gibberellic acid on germination rate

GA Treatment	**Comparison**		**p-value**
Minus	WT	TetC-*cel6*A	<.0001
Plus	WT	TetC-*cel6*A	0.064
Genotype			
WT	Minus	Plus	0.240
TetC-*cel6*A	Minus	Plus	<.0001

Table AII.4: Summary of effect of Gibberellic acid application on stem length statistics

GA	Age (day)	Comparison		p-value
Minus	6	WT	TetC-*cel6*A	1
	10	WT	TetC-*cel6*A	1
	12	WT	TetC-*cel6*A	0.929
	17	WT	TetC-*cel6*A	0.016
	19	WT	TetC-*cel6*A	0.035
Plus	6	WT	TetC-*cel6*A	0.978
	10	WT	TetC-*cel6*A	1
	12	WT	TetC-*cel6*A	1
	17	WT	TetC-*cel6*A	1
Genotype	**Age (day)**			
WT	6	Minus	Plus	<.0001
	10	Minus	Plus	<.0001
	12	Minus	Plus	<.0001
	17	Minus	Plus	<.0001
TetC-*cel6*A	6	Minus	Plus	<.0001
	10	Minus	Plus	<.0001
	12	Minus	Plus	<.0001
	17	Minus	Plus	<.0001

Table AII.5: Summary of CO_2 and ammonium nitrate chamber trial statistics

Parameter	Treatment 1	Treatment 2	Comparison		p-value
Fresh Weight	TetC-*cel6*A	Ambient CO_2	8mM	1mM	0.0005
mg plant^{-1}			4mM	1mM	0.0033
			8mM	4mM	0.5945
		Elevated CO_2	8mM	1mM	<.0001
			4mM	1mM	<.0001
			8mM	4mM	0.3197
		1mM	Ambient CO_2	Elevated CO_2	0.0195
		4mM	Ambient CO_2	Elevated CO_2	0.0003
		8mM	Ambient CO_2	Elevated CO_2	0.0073
	WT	Ambient CO_2	8mM	1mM	<.0001
			4mM	1mM	<.0001
			8mM	4mM	0.0008
		Elevated CO_2	8mM	1mM	<.0001
			4mM	1mM	<.0001
			8mM	4mM	0.0003
	1mM	Ambient CO_2	WT	TetC-cel6A	0.0087
		Elevated CO_2	WT	TetC-cel6A	0.0207
	4mM	Ambient CO_2	WT	TetC-cel6A	0.0007
		Elevated CO_2	WT	TetC-cel6A	0.3918
	8mM	Ambient CO_2	WT	TetC-cel6A	0.0017
		Elevated CO_2	WT	TetC-cel6A	0.2757
Dry Weight	TetC-*cel6*A	Ambient CO_2	4mM	1mM	0.0111
g plant^{-1}			4mM	8mM	0.1077
			8mM	1mM	0.4766
		Elevated CO_2	4mM	1mM	<.0001
			8mM	1mM	<.0001
			4mM	8mM	0.0563
		1mM	Ambient CO_2	Elevated CO_2	0.0004
		4mM	Ambient CO_2	Elevated CO_2	<.0001
		8mM	Ambient CO_2	Elevated CO_2	<.0001
	WT	Ambient CO_2	8mM	1mM	<.0001
			4mM	1mM	<.0001
			8mM	4mM	0.0211
		Elevated CO_2	4mM	1mM	<.0001
			8mM	1mM	<.0001
			4mM	8mM	0.8766

		1mM	Ambient CO_2	Elevated CO_2	0.0014
		4mM	Ambient CO_2	Elevated CO_2	0.0015
		8mM	Ambient CO_2	Elevated CO_2	0.2017
	1mM	Ambient CO_2	WT	TetC-cel6A	0.0003
		Elevated CO_2	WT	TetC-cel6A	0.0003
	4mM	Ambient CO_2	WT	TetC-cel6A	<.0001
		Elevated CO_2	WT	TetC-cel6A	0.0038
	8mM	Ambient CO_2	WT	TetC-cel6A	<.0001
		Elevated CO_2	WT	TetC-cel6A	<.0001
TSP	TetC-*cel6*A	Ambient CO_2	4mM	1mM	0.0291
mg plant^{-1}			4mM	8mM	0.0478
			8mM	1mM	0.0014
		Elevated CO_2	4mM	1mM	0.0048
			8mM	1mM	0.0010
			4mM	8mM	0.2269
		1mM	Ambient CO_2	Elevated CO_2	0.2864
		4mM	Ambient CO_2	Elevated CO_2	0.0419
		8mM	Ambient CO_2	Elevated CO_2	0.9158
	WT	Ambient CO_2	8mM	1mM	<.0001
			4mM	1mM	0.0006
			8mM	4mM	0.0121
		Elevated CO_2	8mM	1mM	<.0001
			4mM	1mM	0.0005
			8mM	4mM	0.0747
		1mM	Ambient CO_2	Elevated CO_2	0.0035
		4mM	Ambient CO_2	Elevated CO_2	0.3947
		8mM	Ambient CO_2	Elevated CO_2	0.5610
	1mM	Ambient CO_2	WT	TetC-cel6A	0.0899
		Elevated CO_2	WT	TetC-cel6A	0.0096
	4mM	Ambient CO_2	WT	TetC-cel6A	0.6242
		Elevated CO_2	WT	TetC-cel6A	0.5162
	8mM	Ambient CO_2	WT	TetC-cel6A	0.7417
		Elevated CO_2	WT	TetC-cel6A	0.4671
Nitrogen	TetC-*cel6*A	Ambient CO_2	8mM	1mM	0.0309
mg plant^{-1}			4mM	1mM	0.064
			8mM	4mM	0.5212
		Elevated CO_2	8mM	1mM	<.0001
			8mM	4mM	0.0002
			4mM	1mM	0.0019
		1mM	Ambient CO_2	Elevated CO_2	0.1329

		4mM	Ambient CO_2	Elevated CO_2	0.2372
		8mM	Ambient CO_2	Elevated CO_2	0.0005
	WT	Ambient CO_2	8mM	1mM	0.0003
			8mM	4mM	0.0008
			4mM	1mM	0.0079
		Elevated CO_2	8mM	1mM	0.0036
			8mM	4mM	0.0129
			4mM	1mM	0.0648
		1mM	Ambient CO_2	Elevated CO_2	0.2763
		4mM	Ambient CO_2	Elevated CO_2	0.6192
		8mM	Ambient CO_2	Elevated CO_2	0.065
	1mM	Ambient CO_2	WT	TetC-cel6A	0.1913
		Elevated CO_2	WT	TetC-cel6A	0.9968
	4mM	Ambient CO_2	WT	TetC-cel6A	0.2436
		Elevated CO_2	WT	TetC-cel6A	0.8
	8mM	Ambient CO_2	WT	TetC-cel6A	0.023
		Elevated CO_2	WT	TetC-cel6A	0.6878
Rubisco	TetC-*cel6*A	Ambient CO_2	1mM	4mM	0.0006
% TSP			8mM	4mM	0.0086
			1mM	8mM	0.0532
		Elevated CO_2	1mM	8mM	0.5661
			4mM	8mM	0.6895
			1mM	4mM	0.9739
		1mM	Ambient CO_2	Elevated CO_2	0.0031
		4mM	Ambient CO_2	Elevated CO_2	0.0074
		8mM	Ambient CO_2	Elevated CO_2	0.0144
	WT	Ambient CO_2	8mM	1mM	0.8238
			4mM	1mM	0.8962
			8mM	4mM	0.9872
		Elevated CO_2	4mM	1mM	0.2850
			4mM	8mM	0.4543
			8mM	1mM	0.9152
		1mM	Ambient CO_2	Elevated CO_2	0.0253
		4mM	Ambient CO_2	Elevated CO_2	0.1416
		8mM	Ambient CO_2	Elevated CO_2	0.0617
	1mM	Ambient CO_2	WT	TetC-cel6A	0.0057
		Elevated CO_2	WT	TetC-cel6A	0.7005
	4mM	Ambient CO_2	WT	TetC-cel6A	0.5417
		Elevated CO_2	WT	TetC-cel6A	0.0703
	8mM	Ambient CO_2	WT	TetC-cel6A	0.6543

		Elevated CO_2	WT	TetC-cel6A	0.5224
Rubisco	TetC-*cel6*A	Ambient CO_2	8mM	1mM	0.0015
mg plant^{-1}			8mM	4mM	0.0135
			4mM	1mM	0.1229
		Elevated CO_2	8mM	1mM	0.0814
			4mM	1mM	0.1235
			8mM	4mM	0.9430
		1mM	Ambient CO_2	Elevated CO_2	0.0069
		4mM	Ambient CO_2	Elevated CO_2	0.1109
		8mM	Ambient CO_2	Elevated CO_2	0.0292
	WT	Ambient CO_2	8mM	1mM	0.0061
			4mM	1mM	0.0426
			8mM	4mM	0.2578
		Elevated CO_2	4mM	1mM	0.0448
			8mM	1mM	0.0881
			4mM	8mM	0.8539
		1mM	Ambient CO_2	Elevated CO_2	0.5428
		4mM	Ambient CO_2	Elevated CO_2	0.4223
		8mM	Ambient CO_2	Elevated CO_2	0.0601
	1mM	Ambient CO_2	WT	TetC-cel6A	0.0362
		Elevated CO_2	WT	TetC-cel6A	0.3494

Table AII.6: Summary mRNA folding analysis and Cel6A accumulation statistics

Parameter	Genotype 1	Genotype 2	p-Value
***in silico* dG of folding**	TetC-*cel6*A	TetC-*bgl*C	<.0001
	TetC-*cel6*A	SynT1-*cel6*A	<.0001
	TetC-*cel6*A	SynT2-*cel6*A	<.0001
kcal mol^{-1}	NPTII-*cel6*A	NPTII-*bglC*	<.0001
	NPTII-*cel6*A	SynN1-*cel6*A	<.0001
	NPTII-*cel6*A	SynN2-*cel6*A	<.0001
	NPTII-*cel6*A	TetC-*cel6*A	<.0001
	NPTII-*cel6*A	SynN3-*cel6*A	<.0001
	NPTII-*cel6*A	GFP-*cel6*A	<.0001
	GFP-*cel6*A	GFP-*bglC*	<.0001
	GFP-*cel6*A	TetC-*cel6*A	<.0001
	TetC-*bgl*C	GFP-*bglC*	<.0001
	TetC-*bgl*C	NPTII-*bglC*	0.8044
	NPTII-*bglC*	GFP-*bglC*	<.0001
	SynN1-*cel6*A	SynT1-*cel6*A	<.0001
	SynN2-*cel6*A	SynN1-*cel6*A	<.0001
	SynN2-*cel6*A	SynT2-*cel6*A	0.0048
	SynN3-*cel6*A	SynN1-*cel6*A	<.0001
	SynN3-*cel6*A	SynN2-*cel6*A	<.0001
	SynT2-*cel6*A	SynT1-*cel6*A	<.0001
	SynT3-*cel6*A	SynT1-*cel6*A	<.0001
	SynT3-*cel6*A	SynT2-*cel6*A	<.0001
	SynT3-*cel6*A	TetC-*cel6*A	<.0001
	SynT3-*cel6*A	SynN3-*cel6*A	0.055
***in vivo* dG of folding**	NPTII-*cel6*A	GFP-*cel6*A	<.0001
	NPTII-*cel6*A	TetC-*cel6*A	<.0001
kcal mol^{-1}	TetC-*cel6*A	GFP-*cel6*A	<.0001
Biomass	WT	TetC-*cel6*A	0.0002
g	SynT3-*cel6*A	TetC-*cel6*A	0.0002
	SynN3-*cel6*A	TetC-*cel6*A	0.0004
	SynN1-*cel6*A	TetC-*cel6*A	0.0017
	WT	SynN1-*cel6*A	0.3136
	SynT3-*cel6*A	SynN1-*cel6*A	0.4429
	SynN3-*cel6*A	SynN1-*cel6*A	0.5352
	WT	SynN3-*cel6*A	0.6902
	WT	SynT3-*cel6*A	0.7611
	SynT3-*cel6*A	SynN3-*cel6*A	0.9072
Stem Height	WT	TetC-*cel6*A	<.0001
cm	SynN3-*cel6*A	TetC-*cel6*A	<.0001
	WT	SynT3-*cel6*A	<.0001

	SynN1-*cel6*A	TetC-*cel6*A	<.0001
	SynN3-*cel6*A	SynT3-*cel6*A	<.0001
	WT	SynN1-*cel6*A	0.0003
	SynT3-*cel6*A	TetC-*cel6*A	0.0003
	SynN3-*cel6*A	SynN1-*cel6*A	0.0013
	SynN1-*cel6*A	SynT3-*cel6*A	0.0941
	WT	SynN3-*cel6*A	0.538
Germination	TetC-*cel6*A	SynN3-*cel6*A	<.0001
days	TetC-*cel6*A	WT	<.0001
	TetC-*cel6*A	SynN1-*cel6*A	<.0001
	TetC-*cel6*A	SynT3-*cel6*A	<.0001
	SynT3-*cel6*A	SynN3-*cel6*A	<.0001
	SynT3-*cel6*A	WT	<.0001
	SynN1-*cel6*A	SynN3-*cel6*A	<.0001
	SynT3-*cel6*A	SynN1-*cel6*A	<.0001
	SynN1-*cel6*A	WT	<.0001
	WT	SynN3-*cel6*A	0.0578
qPCR	SynT3-*cel6*A	SynN3-*cel6*A	0.0002
fold change	SynT3-*cel6*A	TetC-*cel6*A	0.0011
	SynT3-*cel6*A	SynN1-*cel6*A	0.0013
	SynN1-*cel6*A	SynN3-*cel6*A	0.1839
	TetC-*cel6*A	SynN3-*cel6*A	0.214
	SynN1-*cel6*A	TetC-*cel6*A	0.9192
Cel6A	SynT3-*cel6*A	SynN1-*cel6*A	0.0004
% TSP	SynT3-*cel6*A	SynN3-*cel6*A	0.0005
	TetC-*cel6*A	SynN1-*cel6*A	0.0019
	TetC-*cel6*A	SynN3-*cel6*A	0.0021
	SynT3-*cel6*A	TetC-*cel6*A	0.2702
	SynN3-*cel6*A	SynN1-*cel6*A	0.0003