# Tight, Inducible Gene Expression Using Alternatively Spliced Exons from Plants

By

Tania Livania Gonzalez

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California at Berkeley

Committee in charge:

Professor Ming Chen Hammond, Chair
Professor Jamie Cate
Professor Russell Vance
Professor Brian Staskawicz

Fall 2015

Tight, Inducible Gene Expression Using Alternatively Spliced Exons from Plants



By Tania Livania Gonzalez

# ABSTRACT

Tight, Inducible Gene Expression Using Alternatively Spliced Exons from Plants

by

Tania Livania Gonzalez

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Ming Chen Hammond, Chair

Gene expression is regulated at multiple levels, from control of transcription to co-transcriptional processing steps to control of translation and decay. This multi-layered approach allows for sophisticated gene expression programs to be implemented. Previously, our lab engineered HyP5SM, a plant-derived alternatively spliced cassette exon which can be inserted directly and tracelessly into a variety of open reading frames to inducibly regulate protein expression in dicot plants. HyP5SM takes advantage of a gene regulation strategy common in nature – alternative splicing coupled to nonsense-mediated decay – to produce "default off, inducible on" gene expression. Here, we explored the benefits of combining two levels of regulation, a conditional promoter and an NMD-targeted cassette exon, and showed that we could obtain tight, inducible regulation of toxic genes. We demonstrate that addition of HyP5SM can regulate even the hypersensitive response (HR) phenotype, a defensive programmed cell death response initiated by disease-resistant plants upon detection of specific pathogen effector proteins. Effector-triggered immune pathways are important for disease-specific resistance to plant pathogens. We combine the dexamethasone inducible promoter and the HyP5SM cassette exon to regulate pathogen effector proteins. The inducible promoter alone results in leaky effector protein and HR, but HyP5SM renders the leaky transcript non-productive, thus eliminating leaky protein and the resulting auto-immune responses. Furthermore, plants inducibly recover both effector protein expression and the HR phenotype. We have tested this with Bs2/AvrBs2- and RPP1/ATR1Δ51-dependent hypersensitive response pathways in *Nicotiana benthamiana* and *Nicotiana tabacum,* respectively. We also show that *Arabidopsis thaliana* plants transgenic for these resistance/effector gene pairs are viable, healthy, and can complete their full life cycle (seed-to-seed) without inducing leaky HR phenotype from the pathogen effector transgene, unless the HR immune phenotype is specifically induced with dexamethasone. We also present progress identifying additional factors involved in the alternative splicing regulation of HyP5SM. The HyP5SM cassette can be generally applied to regulate other genes in dicot plants, and may be utilized with conditional, constitutive, or native promoters. Finally, we present progress toward engineering new variants of HyP5SM for gene regulation in mammalian cells.

This dissertation is dedicated to my family, my teachers, and my fiancé.

# ACKNOWLEDGEMENTS

This work was possible because of many people who have supported both me and my research. I am very thankful for everyone who took the time to train me in techniques and methods, give me feedback on my writing, and listen to my research problems and offer advice.

First, a huge thanks to my adviser, Professor Ming Chen Hammond. I joined her lab during its second year, when the group was small enough for everyone to present at group meetings. Ming has been a very involved and supportive principal investigator and thesis chair. She taught me many of the lab skills I now know. I had not worked with RNA before joining her lab, and now I am an expert who gives advice to other labs. Ming helped me troubleshoot my experiments when they did not work and shared my excitement over new data when they did. I appreciate Ming's attention to detail when editing my manuscripts and helping interpret data. I'm so glad she chose me as a student. Even when she is most busy, Ming always does her best to make time for us. Ming made sure that I presented at conferences and local symposia regularly, that I had the connections I needed for my projects, that I learned a variety of technical skills so I would be a strong job candidate, and that I was always funded. Ming works very hard for everyone in the lab; she can be sure that it has not escaped our notice.

I also thank my peers in earlier P5SM projects. Scott Hickey, Qian Qin, Malathy Sridhar, Geoffrey Liou, Pooja Vijayendra, and Alexander Westermann pioneered the work to show that *At*P5SM could function outside its native *TFIIIA* gene context and engineered the published hybrid P5SM versions using *A. thaliana* and *O. sativa* sequence. Their work made my work easier. Malathy kindly infiltrated *Nicotiana benthamiana* leaves with my rotation project constructs after I ran out of time, and her leaf scans were used to make **Figure 5.1.1**.

For **Chapters 2 and 3**, I thank Professor Brian Staskawicz and his lab for discussing gene regulation issues that affect plant immunology. I thank Douglas Dahlbeck, Ana Raquel Figueiredo, Allison Schwartz, and Adam Steinbrenner from Brian Staskawicz's lab for providing constructs, *Nicotiana tabacum* plants, 3680 *Arabidopsis thaliana* seeds, and technical advice. I thank Douglas Dahlbeck for discussions and Sandra Goritschnig and my co-authors for feedback on the *Nucleic Acids Research* manuscript. I thank Yeon Lee and Malik Francis from Don Rio's lab for training and help purifying recombinant MaltoseBindingProtein-*Os*L5-6xHis and Exotoxin-*Os*L5-6xHis fusion proteins from *E.coli* in sufficient quantities for generating antigen to make an antibody to *Os*L5, and then for purifying said antibody from serum. Malik especially walked me through the process from start to finish and showed me a lot of techniques to optimize antigen and antibody purifications. Both Yeon and Malik were very kind to provide their time. I also thank former undergraduate Geoffrey Liou for cloning pTA7001 *ATR1Δ51-FLAG*. I thank Chun-Ting "Ting" Kuo for attempting qPCR experiments to quantify the amount of leaky *avrBs2* transcript from pTA7001 *avrBs2-HA*. For the manuscript that has been expanded to **Chapters 2 and 3**, I designed and performed experiments, analyzed data, prepared figures. I co-wrote the manuscript with Ming Chen Hammond. Ming also supervised the study, designed experiments, helped

analyze data. Yan Liang and Dominique Loqué generated the multi-gene pTKan vectors used to make transgenic *A. thaliana* plants. Brian Staskawicz provided guidance and advice on effector-R protein pairs and the hypersensitive response. Bao Nguyen and I tested the multi-gene pTKan vectors and generated transgenic *A. thaliana* plants. Bao joined the lab just after I had floral dipped wild-type *A. thaliana* plants and she assisted me with selection of the first transgenic generation ($T_1$) and every subsequent generation up to $T_3$. Ming and I had opened 1 position for the Undergraduate Research Apprenticeship Program (URAP), and I chose Bao out of 40 applicants. Her enthusiasm and sharp intellect was clear from the very first time I met her. Bao Nguyen was a dedicated, curious, and hard-working undergraduate who made such an obvious contribution to the project that she was added as a co-author during the paper's revision process. I know she will go on to do great things.

For **Chapter 4**: Rotation student James Nuñez helped clone genes from *A. thaliana* when we started the splice factor project. Undergraduate Pooja Vijayendra performed the leaf scans to test RSZ32, RSZ33, and SRp34b that produced the data in **Figure 4.3**. Pooja also cloned out the RSZ32 coding sequence from *A. thaliana* cDNA. Amgen Scholar Megan Dowdle tested the RSZ32 coding sequence in *N. benthamiana* several times thinking that she was incorrectly normalizing the *Agrobacterium* solutions, until we accepted that the RSZ32 coding sequence was suppressing *DsRed2* fluorescence. When my undergraduate Bao Nguyen and I took over, we found the same effect with RSZ33 coding sequence. Megan also used site-specific mutagenesis to clone the RSZ32 mutants (**Figure 4.5**) which Bao and I tested in *N. benthamiana*. Bao and I also cloned the C-terminal tagged constructs and found that the location of the epitope tag affected protein function. Since we had only tested three SR proteins, I proceeded to clone out the genomic and coding sequences of other SR proteins known to interact with RSZ33 to see if they had an effect on *At*P5SM. During this time, graduate student Rebekah Kitto joined the lab and assisted me with leaf fluorescence experiments to test RSZp22 genomic and coding sequences. Rebekah kindly dedicated time to collect the raw data for the RSZp22 quantifications while I was busy writing my dissertation and analyzing earlier data. I also thank post-doc Rachel Senturia and rotation student Robert Nichols for attempts to purify recombinant *Os*L5 and RSZ32 protein from *E. coli*. They helped lay the groundwork for important future experiments.

For the mammalian HyP5SM project in **Chapter 5**, I thank Amgen Scholar Ryan Muller for testing *EGFP-HyP5SM* in HEK293T cells, resulting in the RT-PCR in **Figure 5.2.1**. I thank post-doc Yeon Lee for training both Ryan and me in BSL-2 work with HEK293T cells, including training for cell maintenance, transfections with lipofectamine, visualizing cells with the Haemocytometer and the inverted microscope, and isolating RNA and protein with TRIzol reagent. I had never worked with mammalian cell culture before and Yeon Lee was very kind to provide training, supervision, and detailed advice. Thank you, Prof. Don Rio, for being so welcoming and allowing us to use your BSL-2 lab space. I also thank the entire Splicing Supergroup, including Yeon Lee, Don Rio, and Liana Lareau, for providing feedback on the mammalian project.

I also thank Professor John Coates' lab, including John himself, and his post-docs Hans Carlson and Misha Mehta for teaching me to work with anaerobic *Geobacter sulfurreducens*. Their training allowed me to make a meaningful contribution to the

manuscript titled, “GEMM-I riboswitches from *Geobacter* sense the bacterial second messenger cyclic AMP-GMP”, published in the *Proceedings of the National Academy of Sciences of the United States* (2015) in volume 112, pages 5383-5388.

I thank my excellent thesis committee, who met with me and gave me an outside perspective on my projects. They provided troubleshooting advice when experiments were not working. They also told me when I had too many projects and needed to focus.

I thank my fellow graduate students. I thank my graduate program cohort, the entering class of 2010 in Molecular and Cell Biology; I am thankful for our qualifying exam study sessions, data clubs, and social events. My lab members are great and have been a huge support network. I am thankful for the collaborative and friendly atmosphere in the Hammond lab. Thank you to everyone who ever picked up my plates from the incubator over the weekend; you helped me have life outside of lab. Thank you to everyone who brought snacks for the group; you made long days in lab easier. Thank you to our post-docs and undergraduates. I thank my senior (now graduated) graduate students, Dr. Scott Hickey and Dr. Colleen Kellenberger, for being role model scientists, for being so approachable and generous with advice and protocols, and for generally being great people. Thank you, Colleen, most especially for playing matchmaker and encouraging my fiancé to ask me out on that first date. You did good. Thank you to the current students for making lab a wonderful atmosphere: Cindy Wang for your inspirational cross-disciplinary talent; Zach Hallberg for your intense hard-working self, your cheerful attitude, all your high-fives, and for always caring about people; Yichi Su, for your concise summaries during group meeting, cheerful attitude, and for either driving us or lending us your car when we went to MCB retreats; to all the current second years, Todd Wright, Johnny Truong, Andrew Dippel, and Rebekah Kitto for doubling the size of the lab and suddenly making us a much bigger group – I am glad you all joined our lab! Especially Rebekah. Thanks, Rebekah, for becoming my successor to the P5SM projects! P5SM in plants was the founding project of the Hammond lab. I am excited to see where you will take splicing regulation next.

I also thank the students in the UC Berkeley chapter of the Society for the Advancement of Chicanos/Latinos and Native Americans in Science (SACNAS) for mentorship, friendship, and inspiration. Thank you, Patty Garcia Guzman, Galo Garcia, Maria Mouchess, Jose Estrada, Adrienne Greene, Akemi Kunibe, Vanessa de la Rosa, Lissette Andres, and Brandon Gaytán for guidance and mentorship. Thank you, Brian Castellano, Brandon Brooks, and the current students for keeping SACNAS alive. Also, a huge thanks to our faculty mentor, Professor David Weisblat; the MCB Graduate Affairs Office (especially Berta Parra); and Carisa Orwig for institutional support.

My family played a central role in my amazing support network. I thank my parents, who filled our home with field guides and dictionaries, who took me to the park and the library, and who instilled in me the love of science and learning. My mother walked us to the public library so often that the librarians learned our names. Her discussions and counter-hypotheses during television news segments taught me to think critically about correlation and causation. My father’s curiosity, talent for remembering facts, love of biology, and extensive science vocabulary in Spanish taught me to explain science in different ways. I went into high school anticipating that I would

love biology, because he loved it. My siblings and their varied interests also taught me about different fields of science; they are creative geniuses. I especially thank my sister, Alin, who completed a Master's degree while I was at UC Berkeley, and who shared her field biology and graduate school experiences with me.

I also owe a debt of gratitude to my teachers, who challenged me and fought for me and helped me navigate academia. California passed Proposition 187 when I was in elementary school and the negative effect it would have had on my poor immigrant neighborhood was terrifying to my 7-year-old self. The anti-immigrant rhetoric and the state's voting results made me feel so unwanted and scared, but my teachers were vocally opposed to the proposition and firmly told us that they would never abandon us or our school. That stage in my life had a lasting impact on my views on education, teaching me that it was not something I could ever take for granted. I thank my Cesar Chavez Elementary School teachers, Winnie Porter and Norman Mattox, who are my friends to this very day. I still have the mini-dictionary I received during the "I'm Going To College!" field trip the school organized with San Francisco State University. I also thank my Aptos Middle School teachers, especially Eileen Coulter and Matt Chapman who nominated me for the Gifted and Talented Education program. I thank Mrs. James who pushed me to apply for UC Berkeley's summer Academic Talent Development Program. I almost did not apply because it seemed so expensive, but my parents saw the course catalog in my opened backpack and told me to go ahead; that summer, I went from the least to most computer-savvy student in my middle school homeroom. I also thank the teachers and staff at Lowell High School, all of them, for being so invested in our education. You made my transition to college easy.

At my undergraduate school, UC Davis, I thank the Biology Undergraduate Scholars Program (BUSP), the NSF-funded Collaborative Learning at the Interface of Mathematics and Biology (CLIMB) program, and the McNair Scholars Program for putting me on a path toward the PhD. I have a confession: when the freshman orientation survey asked us about the highest level of education we wanted to achieve, I only said PhD because I was vain and knew that I didn't want an MD. I did not *actually* think I was going to graduate school. I just liked the idea of research. My family was poor and I thought getting a PhD would require huge loans; it seemed unaffordable and unrealistic. I thank UC Davis for calling my bluff and showing me that graduate school was a real option. I thank my undergraduate principal investigators and mentors: Professors Anne Britt, Richard Grosberg, Sebastian Schreiber, and Judy Callis; Dr. Carole Hom, Dr. Gina Holland; and Mr. Rolf Unterleitner. You gave me opportunities, inspired me to work hard, advised me, wrote letters of recommendation for me, and supported me throughout my undergraduate career. I am eternally grateful.

I thank my fiancé, Mike, whom I met at a Molecular and Cell Biology holiday party during my second year of graduate school, and who brightens my life. Mike made the ups and downs of graduate school easier. I love him and I like him, so much.

Thank you to everyone who has helped me and whom I may have omitted!

# TABLE OF CONTENTS

## LIST OF ABBREVIATIONS

| | |
|---|---|
| 3'ss, 5'ss | 3' splice site, 5' splice site |
| 3'UTR | 3' untranslated region, the mRNA transcript after the stop codon |
| 5'UTR | 5' untranslated region, the mRNA transcript before the start codon |
| 35S | Cauliflower Mosaic Virus 35S promoter, a constitutive plant promoter |
| 6xUAS | 6x Upstream Activating Sequence; DNA element bound by Gal4; here used as an abbreviation to describe the Dex/GVG-inducible promoter |
| *At*L5 | L5 ribosomal protein from *Arabidopsis thaliana* |
| *At*P5SM | P5SM of *Arabidopsis thaliana*, published in Hammond et al. 2009 |
| ATR1 | Pathogen effector from oomycete *Hyaloperonospora arabidopsidis* |
| ATR1Δ51 | N-terminal truncated ATR1, removed exit signal for stability in plant cells |
| avrBs2 | Pathogen effector from bacterium *Xanthomonas euvesicatoria* |
| Bs2 | Resistance gene from pepper plants resistant, recognizes AvrBs2 |
| Dex | Dexamethasone, glucocorticoid that induces expression from 6xUAS/GVG |
| EGFP | Enhanced Green Fluorescent Protein, a modified form of GFP |
| EJC | Exon-Junction Complex |
| Emoy2 | Describes allele of ATR1, recognized by the RPP1-WsB allele |
| GFP | Green Fluorescent Protein |
| GVG | Chimeric transcription factor Gal4$_{\text{bindin domain}}$-VP16$_{\text{activation domain}}$-Glucocorticoid Receptor$_{\text{binding domain}}$, localized to the nucleus upon Dex binding, where it promotes gene expression by binding 6xUAS |
| HyP5SM | Engineered hybrid of *At*P5SM and *Os*P5SM, responsive to *Os*L5 |
| L5 | L5 ribosomal protein, which binds P5SM and promotes P5SM exclusion |
| NMD | nonsense-mediated decay, a pathway that degrades mRNA with PTCs |
| *Os*L5 | L5 ribosomal protein from *Oryza sativa* (rice), promotes *Os*P5SM and HyP5SM exon skipping |
| *Os*P5SM | P5SM of *Oryza sativa* (rice), published in Hickey et al. 2012 |
| P5SM | Plant 5S rRNA Mimic, conserved structured RNA in plant *TFIIIA* genes. The term "P5SM" is also used as a general term encompassing various versions of P5SM-containing cassette exons (including flanking introns). |
| pBinAR | Binary vector used to constitutively express transgenes in plants. |

| | |
|---|---|
| pTA7001 | Binary vector used to express Dex-inducible transgenes in plants. |
| PTC | Premature Termination Codon, an early stop codon |
| R protein | Resistance protein, a receptor that specifically recognizes a pathogen effector and initiates immune responses that result in disease resistance |
| RPP1 | Resistance gene from *Arabidopsis thaliana,* recognizes ATR1 |
| RSZ32 | Splice factor, SR protein |
| RSZ33 | Splice factor, SR protein |
| RUST | Regulated Unproductive Splicing and Translation |
| SNP | Single nucleotide polymorphism, a point mutation |
| SR | Serine/Arginine-rich domain characterizing the SR splice factor family |
| SRp34b | Splice factor, SR protein |
| SP-I | Splice product I, in which the entire P5SM cassette is excluded, thus resulting in a “scarless” coding sequence that may be translated to protein. Co-expression with L5 promotes this splice product. |
| SP-II | Splice product II, in which the introns of the P5SM cassette are excluded, but the P5SM “suicide” exon itself remains. Because the P5SM exon has a premature termination codon, this splice product is targeted by nonsense-medicated decay and thus non-productive. |
| UPF1 | Up-Frameshift Protein 1, part of the NMD complex |
| WsB | Describes allele of RPP1, recognizes the ATR1-Emoy2 allele. In this manuscript, no other alleles of RPP1 or ATR1 are used. |

# CHAPTER 1:
# An Introduction to Alternative Splicing and Gene Regulation

## 1.1: Splicing in Various Systems (Plants, Animals, Yeast, Even Prokaryotes!)

### Discovery of Splicing

In 1961, Francis Crick and his collaborators published a paper presenting evidence for their hypothesis that the genetic code for proteins was composed of triplicate bases that coded for individual amino acids (1). By 1965, the three-base codons had been tied to amino acids and the "RNA code" could be used to translate from RNA sequence to protein sequence (2). However, in the 1970s, increasing information about gene and protein sequences led to the observations that sometimes the protein-coding regions of genes and genomes ("exons") were disrupted by stretches of nucleotides that were not translated into the final protein product (3). These non-coding sequences have been called "intervening sequences," "intragenic sequences," and finally "introns." In 1980, an experiment by the Abelson group removed an intron sequence in a yeast transfer RNA suppressor and reported that the artificially spliced sequence still resulted in functional protein when expressed in yeast (4). These intervening sequences were found not just in protein-coding sequences, but also in ribosomal RNA (rRNA) precursors in eukaryotes like the protozoa *Tetrahymena thermophile* (5). Furthermore, Thomas Cech's group found that some of these introns could self-splice out of their precursor mRNA (pre-mRNA) *in vitro*, leading to the first evidence that RNA could be catalytic (6). Today, self-splicing introns have been found even in the 16S rRNA of the giant sulfur bacteria *Thiomargarita namibiensis,* as well as in Archaea and eukaryotic organelles (7, 8).

### Origin of Introns and the Spliceosome

There are three classes of introns. Group I self-splicing introns and Group II self-splicing introns have been found in both prokaryotes and eukaryotes and catalyze their own splicing *in vitro* without protein factors, although some need proteins to splice efficiently (5, 7, 9, 10). The third group of introns are typically called "spliceosomal introns" and these are only found in eukaryotes, because only eukaryotes have the necessary protein complex to mediate their splicing (11). This spliceosome complex is composed of uridylate-rich small nuclear ribonucleoprotein particles (U snRNPs, pronounced "you snurps"), non-U snRNPs, and various associated co-factors (12). The spliceosome finds its RNA target using RNA-RNA interactions and facilitates the two successive *trans*-esterification reactions required for RNA splicing (12).

There are two competing models for the origin of introns, “intron-early” and “intron-late” (8). The intron-early model postulates that introns interrupted exons in the very early stages of life, pre-eukaryotes, that these interruptions shaped gene evolution, and that introns were lost in prokaryotes through active selection of condensed genomes. The intron-late model postulates that intron appeared in eukaryotes and are continually being added to protein-coding sequences *de novo*. With some exceptions, introns sequences are not strictly conserved (13). However, there are introns positions that are highly conserved not just within-kingdoms, but also between. A large-scale genome comparison found that 14% of animal introns and 13% of fungal introns matched plant intron position, supporting the intron-early model (14). An analysis of 16S rRNA from a population of giant sulfur bacteria found 131 introns in 4 independent insertion sites, of which 1 site was identical and 3 were near intron insertion sites in eukaryotic nuclear rRNA and mitochondrial rRNA (7). The researchers noted that giant sulfur bacteria contain large energy reserves, possibly making replication and transcription less costly for them, which also appears to support an intron-early model (7). However, another analysis of introns in 16S rRNAs found variable occupation of these insertion sites and suggests that regions of the rRNA sequence might be hot spots for intron insertion, providing an alternative hypothesis that supports the intron-late model (7, 15).

**Recognition of Introns and Exons, and Comparisons Between Organisms**

There are three key components to introns recognition: a 5’ splice site, a branch point site, and a 3’ splice site. In *Saccharomyces cerevisiae*, splicing requirements are very strict: few genes have introns, the intron length does not vary much, introns are usually located at the far 5’ end of genes, the sequence UACUA**A**C strictly denotes the location of the branch point site (bold, underlined), and the splice site consensus sequences are strict (16). Vertebrates and plant introns are not so dependent on a strict consensus branch point site, although they do have traits that characterize introns (**Table 1.1**). For example, plant introns are characterized by AU-rich sequences, and dicots have a particularly strong preference for U bases (17, 18). Typically, the AU content is 15-20% higher in introns than neighboring exons (18). This is generally true for all flowering plants, although dicot plants have a greater shift in AU composition from introns to exons (74.3/55.0%), as compared to monocots (58.7/42.7%); this sharp transition in base preferences is not found for vertebrates (18, 19).

Additionally, the size of introns and exons varies a lot among kingdoms and even species. *S. cerevisiae* genes rarely have introns (3%), whereas over 40% of *S. pombe* genes contain introns (**Table 1.2**). Mammals tend to have large very large introns (often >3 kb) and exons an order of magnitude shorter (~200 nt), whereas *S. cerevisiae* is shows just the size relationship. Comparatively, plant introns and exons are more similar in size to each other (**Table 1.2**). The relative size differences are directly tied to the molecular mechanisms of recognition for spliced sequences, which are not the same among organisms (12). In yeast, almost all information required for proper splicing of the intron is located in the intronic sequence itself, as opposed to the surrounding exon sequence (20). Only the immediate exon sequence (~10 bp) is important for splice site definition, since yeast have strong consensus sequence (21, 22). By contrast, mammalian cells discriminate exons, not introns, and consequently have maximum size

limitations for their exons. An early study that looked at 1600 primate exons not located at the ends of genes found that only 3.5% were longer than 300 nt, and fewer than 1% exceeded 400 nt (23). This “exon-definition” mechanism of splicing is thought to occur by protein-protein interactions between the spliceosome at the 3’ splice site and protein splice factors at the 5’ splice site of the exon (23). Exons artificially made shorter than 50 nt tend to be skipped in mammalian cells, possibly due to steric hindrance of the exon-defining complexes (24). Plants have both intron-defining and exon-defining features. As described previously, there is a clear difference in nucleotide composition between plant exons and introns. Additionally, plants have a wider range of splice factors than mammalian cells (25).

**TABLE 1.1:** *Characteristics of introns that splice well in various systems.*

| Organism | 5’ Splice Site | Branch site | 3’ Splice Site | Ref |
|---|---|---|---|---|
| ***Homo sapiens*** (humans) | $A_{63}G_{79}/G_{100}T_{100}$ | yUn**A**y, weak consensus. Frequency of Y increases -30 from intron, reaching 75-90% at positions -15 to -5. Y-rich tracts thought to help branch site selection. | $C_{77}A_{100}G_{100}$/g | (12, 18, 26) |
| ***Saccharomyces cerevisiae*** (budding yeast) | AAG/GTAUGU | UACUA**A**C (Strict!) Third A is branch point. | $AY_{98}A_{100}G_{100}$/ | (12, 27, 28) |
| ***Arabidopsis thaliana*** (dicot plant) | $A_{62}G_{79}/G_{100}T_{99.4}$ | AU-rich and U-rich tracts near 3’ splice site, but no strict consensus. | $C_{66}A_{100}G_{100}$/ | (12, 17, 18) |
| ***Zea mays*** (monocot plant) | $AG/G_{100}T_{99}$ | AU-rich tracts near 3’ splice site. | Intron ends in AG always. | (12, 19) |

*Y = A or G; lowercase is used when poor conservation of bases is specifically noted. Subscript numbers denote percent occurrence when available. Slashes indicate different sides of the exon/intron or intron/exon junction.

## Alternative Splicing and Serine/Arginine-Rich Proteins

Alternative splicing describes a mechanism of gene regulation characterized by pre-mRNA processed to more than one final mRNA product, not including the unspliced pre-mRNA. Alternative splicing is an important regulated method to increase protein diversity (29). Alternative splicing to non-productive products is also a well-known method of controlling transcript levels (see **Section 1.2**). In *S. cerevisiae*, few genes have introns and those that do tend to be constitutively spliced (**Table 1.2**). Alternative splicing is not common in *S. cerevisiae*, but a few rare examples exist, such as the *PTC7* gene which can splice to two different proteins (30). In contrast, alternative splicing is a very common in plants and animals, with approximately 61% of intron-containing plant genes regulated by alternative splicing and over 95% of human genes regulated by alternative splicing (31, 32). Alternative splicing includes variable splice

site choices, including alternative 5' splice sites, alternative 3' splice sites, intron retention, mutually exclusive exons , and cassette exons (sometimes retained, sometimes skipped). Alternative splicing events are represented differently in various organisms (**Table 1.3**).

**TABLE 1.2:** *Comparison of average splicing characteristics among organisms.*

| Organism | Gene size (kb) | Genes containing introns | Introns per gene | Exon size (nt) | Intron size (nt) | Intron-containing genes that alt. splice | Ref |
|---|---|---|---|---|---|---|---|
| ***Saccharomyces cerevisiae*** | 1.3 | 3% | 1 | 1274 | 270 | Three known: *PTC7, YML034W, MTR2* | (20, 30, 33, 34) |
| ***Saccharomyces pombe*** | 1.4 | 43-45% | 1-2 | 140 | 40-93 | - | (20, 35) |
| ***Arabidopsis thaliana*** (dicot plant) | 2.0 | >98% | Usually >1 | 217 | 145-167 | >61% | (31, 36) |
| ***Oryza sativa*** (rice, monocot plant) | 2.7 | - | - | 201-254 | 356-413 | >30% (early estimate) | (37–39) |
| ***Homo sapien*** (human) | 28 | >92% | 8.4-9.37 | 171-200 | 3560-5500 | >95% | (20, 32, 38, 40) |
| ***Drosophila*** | 1.1 | - | 3.98 | 141 | 564 | 19% | (36, 37, 40) |
| ***Caenorhabditis elegans*** | 1.6 | - | - | 100 | 467 | - | (37) |

*Some values were not available or found, indicated by dash line. Introns per gene refers to typical or average number of introns per intron-containing gene, not per all genes in the genome.

**TABLE 1.3:** *Types of alternative splicing among organisms.*

| Organism | Intron retention | Cassette exons | Single Exon Skipped | Multiple Exons Skipped | Alternative Donor Site (Alt. 5' SS) | Alternative Acceptor Site (Alt. 3' SS) | Ref |
|---|---|---|---|---|---|---|---|
| ***Arabidopsis thaliana*** (dicot plant) | 41-56% | 3-8% | - | - | 10.2-18% | 21.9-38% | (37, 41) |
| ***Oryza sativa*** (rice, monocot) | 33-53.5% | 11-13.8% | - | - | 11.3-22% | 15.1-34% | (37, 41) |
| ***Homo sapien*** (human) | 5% | 69% | 58% | 11% | 26% | | (42) |

*Percentages of total splicing events.

Alternative splicing provides another level of regulation for organisms to respond to environmental cues such as temperature, heat, light, and drought stress (43–46). Weak splice sites may be especially responsive to their environment. For example, an *A. thaliana* temperature-sensitive mutant with normal flower development at 16°C, but aberrant flower development at 28°C, was found to be due to a single base change from A to U to that weakened the 5' splice site of an intron in *APETALA3 (AP3)* from the consensus AG/GU exon/intron junction to mutant UG/GU (47). The mutant gene produced two protein products, one functional and one inactive, neither protein affected by temperature. The mutant plant phenotype was found to be due to differences in functional protein levels due to changes in splice ratios that were further shifted at higher temperatures (47).

Splice site selection uses RNA-RNA interactions with spliceosome components, such as U1 snRNP binding to the consensus 5' splice site (12). However, in organisms with weaker consensus splice sites and higher levels of alternative splicing, U1 snRNP is not always sufficient for 5' splice site selection (48–50). Eukaryotes have families of serine/arginine-rich (SR) proteins, some of which act as splice factors that promote recruitment of the spliceosome to pre-mRNA (51, 52). In HeLa cell extract with a β-globin splicing model, human SR protein ASF/SF2 increasingly promoted selection of the 5' splice site nearest the 3' splice site in a dosage-dependent manner (49). In HeLa cell extract, human SR proteins, SRp30 and SRp40, promoted splicing intron skipping and an intron+exon skipped splice product of the SV40 early pre-mRNA splicing model, again in a dosage-dependent manner (50). SR proteins have been found to bind exonic or intronic splicing enhancer (ESE or ISE) sequences, promoting recruitment of spliceosome components such as U1 snRNP or the later U4/U6.U5 tri-snRNP complex (50, 53). Some SR proteins are linked to non-splicing processes. For example, human SR protein SC35 promotes transcriptional elongation in a sequence-dependent manner (54). SC35 also promotes intron exclusion of target pre-mRNAs *in vitro* with HeLa cell extract (55). This dual function is not a surprise, since splicing is known to happen co-transcriptionally (56, 57). I discuss SR proteins more in **Chapters 4** and **5**.

## 1.2: Pairing of Alternative Splicing and Nonsense-Mediated Decay

In this section, I discuss an RNA surveillance pathway that is often coupled to alternative splicing. This nonsense-mediated decay mechanism is the basis for the "suicide" quality of the alternatively spliced exon I will discuss in a later section.

### Nonsense-Mediated Decay

Not all alternatively spliced transcripts result in alternative protein products. When eukaryotes produce mRNAs with premature termination codons (PTCs), those transcripts are selectively degraded by an RNA surveillance pathway called nonsense-mediated decay (NMD). NMD is thought to be a quality-control mechanism to prevent the accumulation of truncated proteins which might have deleterious dominant-negative or gain-of-function activity (58, 59). NMD also acts as a normal gene regulation mechanism that controls the abundance of transcripts (60). NMD happens in all studied

eukaryotes, but there does not appear to be a universal NMD pathway (58, 61). The core components of the NMD complex itself (UPF1, UFP2, UPF3A/B) appears to be highly conserved in all eukaryotes, but cofactors vary, and there are differences in how the NMD complex is recruited to PTCs and differences in the final method of RNA degradation (58). Because NMD is best studied in mammalian cells and not yet well-understood in plants, this section focuses briefly on what is known in mammalian NMD pathways and discusses known differences in other organisms.

In order for NMD to function, it first requires the ability to distinguish between the final stop codon and PTCs. NMD target recognition is a topic currently in flux. In mammals, it was long well-established that NMD distinguishes PTCs relative to the location of an Exon-Junction Complex (EJC), a protein complex that is deposited 20-24 nt upstream of exon-exon junctions on spliced mRNA and remains bound to the mRNA when it exists the nucleus (62–65). In mammalian cells, PTCs must be 50-55 nt upstream of an EJC in order to efficiently promote EJC-enhanced NMD (66). Transcripts without introns are immune to NMD in mammalian cells (66, 67). Assembly of the NMD complex requires Up-Frameshift Protein 1 (UPF1), a highly conserved 5'→3' ATP-dependent helicase protein in eukaryotes (60). UPF1 interacts with the EJC, cap-binding proteins, and elongation release factors of the ribosome, (64, 68). Ribosome stalling at PTCs allows the NMD complex to form and UPF1 to inhibit ribosome release, whereas ribosomes near the end of the mRNA interact instead with poly(A) binding proteins which inhibit NMD (69, 70). NMD requires translation so that a stalled ribosome is present and NMD-associated factors such as EJC are not displaced (58). NMD is inhibited by translation inhibitors such as cycloheximide, anisomycin, and puromycin (71).

Until very recently, NMD was thought to *only* affect transcripts during the pioneer round of translation in mammalian cells (58, 68, 72, 73). NMD was specifically thought to target only transcript bound by cap-binding protein heterodimer CBP80-CBP20, characteristic of newly synthesized mRNA, and not mature eIF4E-capped mRNA (71, 74). The model behind this was that eIF4E-capped mRNA had already lost the EJC and other required NMD factors due to displacement by passing ribosomes (71, 75). However, HeLa cell experiments published in 2013 show that eIF4E-capped mRNAs with long 3'UTRs are also targeted by NMD (68, 76). There are now thought to be both EJC-independent and EJC-enhanced NMD pathways in mammalian cells (77).

In *S. cerevisiae*, a species where only 3% of genes have introns, NMD is not efficient and mRNAs may be targeted by NMD even after the pioneer round of translation; in fact, multiple rounds of translation (> 200 rounds) are required to noticeably reduce the quantity of PTC-bearing transcripts in yeast (78). Most components of the EJC are absent in *S. cerevisiae* and *S. pombe*, so PTCs recognition does not require EJCs (79). Instead, *S. cerevisiae* cells distinguish transcripts with PTCs by recognizing abnormally long 3' untranslated regions (3'UTRs), characterized by long distances between the stop codon and the poly(A) tail (80, 81). In *S. pombe,* where < 5% of genes contain introns, splicing of introns does promote NMD, but no EJC is required and the length of the 3'UTR does not have an effect (79).

In plants, NMD is not as well researched; the exact rules governing NMD target selection and the essential components for NMD in plants are not yet established (61). However, there is evidence that both retained introns and long 3'UTRs promote NMD, most likely by overlapping NMD pathways resembling both yeast and mammalian NMD (61, 82). Plants contain an EJC-like complex like mammalian cells, but tethering the EJC proteins to the 3'UTR does not have the same NMD-promoting effect seen for mRNAs in mammalian cells, so the function is not exactly the same (82, 83). Overexpression of plant EJC proteins promotes expression of intron-containing genes in *A. thaliana*, suggesting a role in gene regulation perhaps similar to the translation-enhancing role of EJCs for intronless genes in mammals (84, 85). Tethering UPF1 the 5'UTR of an intronless gene also promotes translation in mammalian cells, whereas UFP1 promotes NMD instead if at the 3'UTR (85, 86). In plants, UPF1 results in a strong improvement in NMD efficiency whether tethered to either the 5' or 3' UTRs (82). This difference might be because intronless genes containing PTCs are not immune to NMD in plants, unlike in mammalian cells (87, 88). In *N. benthamiana* plants, increasing length of 3'UTRs in intronless genes is proportional to more efficient NMD (82). As for NMD targeting intron-containing genes, it was recently shown that plants follow a "50 np rule" similar to the mammalian rule (66, 83, 89). Expression of constructs with introns located at different distances downstream from a PTC showed that introns located > 50 nt from the PTC were able to efficiently trigger NMD in *N. benthamiana* (83).

Finally, transcripts can acquire PTCs in various ways. PTCs can be due to frameshifts or point mutations that result in PTCs, coding sequences interrupted due to transposons, programmed DNA rearrangements, and by failure to excise introns or alternative exons in the final mRNA (58). The coupling of alternative splicing and NMD is common in organisms with alternative splicing and is thought to be an evolved gene regulation mechanism (58).

**Regulated Unproductive Splicing and Translation (RUST)**

Coupling of alternatively splicing and NMD is called Regulated Unproductive Splicing and Translation (RUST) (90, 91). Alternatively spliced cassette exons that contain PTCs have been called "suicide exons" and "poison exons" because their retention in the final splice product generally makes the mRNA a target for NMD. Defects in NMD are known to have severe effects in eukaryotes, especially eukaryotes whose genomes are heavily regulated by splicing. UPF1 knockouts are lethal for mammals, *Drosophila*, zebrafish, and plants (92–95), but not for *Caenorhabditis elegans* or *S. cerevisiae* albeit they do show growth and metabolic defects (58, 60). In HeLa cells, NMD can be reduced by RNAi-mediated silencing of UPF1 for short-term experiments (< 1 week), but over time the reduced levels of UPF1 cause genome instability, activation of DNA damage response pathways, and cell cycle arrest (96). In mice, a conditional knockout of UFP2 in the hematopoietic stem cell compartment (inside bone marrow) led to permanent and irreparable loss of hematopoietic stem cells and mice died within 10 days (97). These protein knockdowns all cause a wide range of effects; potentially some phenotypes could be due to functions of the NMD core components that are not strictly related to PTCs (58). More specifically, many human diseases have been associated with small nucleotide polymorphisms (SNPs) that create PTCs (98). For example, 80% of mutations in the cadherin-1 gene, associated with

hereditary diffuse gastric cancer, were SNPs that created PTCs (98). Rare PTCs causally associated with diseases are the most common targets for gene therapy (99).

In *A. thaliana*, approximately 18% of alternative splicing events lead to non-productive mRNA that is targeted by NMD (100). Reduction of NMD is possible by maintaining sickly T-DNA-interrupted UPF1 mutants in hemizygous lines, but homozygous lines are not viable (92). Various T-DNA insertions to disrupt other components of NMD in plants also usually result in conditionally viable plants with defects in growth or fertility, but interestingly they have constitutively over-active immune pathways (92). Later RNA-seq experiments showed that *A. thaliana* plants use NMD to regulate turnover of immune receptors with Toll/Interleukin-1 Receptor (TIR) domains (101). RUST is also common in the regulation of circadian clock genes in *A. thaliana* (102).

In eukaryotes with extensive alternative splicing, RUST is a common and important method of gene regulation. As the previous sections show, changes to the location of a PTC in transcripts can have big effects on gene regulation. In laboratories, however, splicing-based gene regulation is not common as a biotechnological tool. In the following section, I will discuss previous splicing-based methods of artificial gene regulation and introduce an alternatively spliced “suicide” exon from plants.

## 1.3: Engineered Splicing Regulation and P5SM

### Intron-Based Regulation

The addition of introns to genes of interest allows researchers to reach higher gene expression levels than what is possible without introns. If plants, addition of introns can increase transcription levels 5-100 fold, presumably due to the coupling of transcription and splicing (103, 104). A similar 10-100 fold improvement in mRNA levels was seen from multiple intron-containing versus intronless transgenes in microinjected fertilized mouse eggs, although in this case the effect was not reproduced in cultured cells (105). Introns are also beneficial for vectors used for gene therapy in human patients; transgene expression is a limiting factor in gene therapy, and the addition of synthetic introns to *trans*-splicing adeno-associated viral vectors improves transduction efficiency by raising the barrier for mRNA accumulation in 293 cells and muscle Dys-3-positive cells (106).

On the other hand, addition of introns can solve leaky expression issues when amplifying plasmids in *E. coli* for use in mammalian cells. A chicken β-actin promoter driving *Cre* recombinase construct for use in HEK293T cells was found to result in unwanted and toxic leaky expression in *E. coli* cells (used to amplify the vector), but the addition of an intron in the *Cre* gene eliminated leaky expression in *E. coli* and solved the toxicity problem without disrupting expression in HEK293T cells (107).

## Chemically-Regulated Splicing with Aptamers

An aptamer is a single-stranded RNA element that becomes structured upon specific and high-affinity binding a small molecule ligand (108). Riboswitches contain aptamers and also regulatory elements which alter gene expression upon ligand-binding (108). The naturally-occurring thiamine pyrophosphate (TPP) riboswitch regulates alternative splicing of introns in the 5' ends of three *Neurospora crassa* (filamentous fungus) genes (109). Dong-Suk Kim *et al*. used the theophylline aptamer to make an inducible-off method of *in vitro* gene regulation for HeLa nuclear extract (110). They achieved this by cloning the theophylline aptamer in an intron between the pyrimidine (Y)-rich tract that marks the intron branch site and the 3' splice site. Theophylline inhibited intron-skipping in a dose-dependent manner, with up to 75% splicing inhibition was seen in a 2 hour assay with 2 mM theophylline (110).

Julia Weigand and Beatrix Suess used a similar idea to develop an aptamer-based and chemically-suppressible method of *in vivo* gene regulation for *S. cerevisiae* (111). They achieved this by cloning the tetracycline aptamer near the 5' exon/intron junction of a synthetic intron. Upon addition of tetracycline, the aptamer becomes structured, obscuring the 5' splice site and leading to intron retention, which disrupts gene expression (111). Weigand and Suess were able to modulate the decrease in expression of their GFP reporter from 2.5- to 4.3-fold by moving the aptamer closer to the 5' splice site, and improved GFP reduction 6.2-fold by altering the aptamer structure until the consensus sequence formed part of the aptamer's stem (111).

The advantage of these aptamer-based methods is that genes can inducibly and transiently be turned off with a chemical inducer. This is unlike conditional promoters, which have generally been designed for gene expression to be inducibly turned on.

## Plant 5S rRNA Mimic (P5SM), a Cassette Suicide Exon That Regulates TFIIIA

Cassette exons are alternatively spliced exons which are either entirely retained or skipped in the final mRNA splice product. Cassette exons are called "suicide" exons or "poison" exons when they contain a premature-termination codon (PTC) and are regulated by nonsense-mediated decay (see **Section 1.2**).

Plant *Transcription Factor IIIA* (*TFIIIA)* genes contain a conserved cassette exon which contains a highly conserved structured RNA with sequence and structure similarity to 5S ribosomal RNA (112, 113). This mimic, coined Plant 5S rRNA Mimic (P5SM), is found in the *TFIIIA* genes of all plants, but not in green algae or non-plants (113). This is interesting because TFIIIA protein is thought to be a limiting transcription factor for the recruitment of RNA Polymerase III to 5S rDNA genes in *Xenopus* oocytes (114). If TFIIIA is also limiting in plants, it makes sense that low 5S rRNA levels activate a feedback loop to produce more TFIIIA protein. The mechanism connecting this feedback loop is L5 ribosomal protein, the binding partner for 5S rRNA which forms a ribonucleoprotein component of the large subunit of plant ribosomes (112, 115). L5 ribosomal protein preferentially binds 5S rRNA, but will bind P5SM when 5S rRNA levels are low (112). L5 binding to P5SM promotes exon skipping of the P5SM-containg suicide exon in *TFIIIA* pre-mRNA, thereby promoting functional *TFIIIA* splice product which results in increased TFIIIA protein levels (112). This regulation strategy of TFIIIA

protein may be a way of dynamically balancing stoichiometric ratios of paired downstream molecular partners. Thus, in plants, P5SM in *TFIIIA* is part of a feedback look to regulate levels of 5S rRNA and unbound L5 ribosomal protein (112).

While animals do not appear to regulate their 5S rRNA or TFIIIA protein levels in the same RUST-mediated way as plants, they do regulate other components of the ribosome using a similarly dynamic suicide exon strategy. *C. elegans* ribosomal proteins L3, L7a, L10a, and L12 – as well as human ribosomal proteins L12 and L3 – are auto-regulated in negative feedback loops that utilizes nonsense-mediated decay and alternative splicing (116, 117). Like the highly conserved P5SM element in plants, alternatively spliced intron 3 in human ribosomal protein L3 gene shows unusually high degrees of conservation with other animals, with some segments preserving 50% and up to 90% sequence identity between *C. elegans, Drosophila melanogaster*, and humans (116).

Animal TFIIIA protein activity is mediated by binding of TFIIIA protein with its 5S rRNA product, a binding event that also occurs in plants (118, 119). In plants, low levels of 5S rRNA promote indirectly resulting in a shift of TFIIIA alternative splicing to make more functional protein. In *Xenopus* oocytes, 5S rRNA has a more direct effect: high levels inhibit TFIIIA protein function by competing with its 5S rDNA transcription target, whereas low levels of 5S rRNA free TFIIIA (118). Supporting this competition model, mutant TFIIIA proteins with reduced affinity for 5S rRNA result in elevated transcriptional activity at 5S rDNA genes, resulting in higher levels of 5S rRNA (118).

### Engineered Hybrid P5SM

Previous member of the Hammond lab showed that the *A. thaliana* P5SM cassette could be moved into other gene contexts and still regulate alternative splicing in an *At*L5 ribosomal protein-dependent manner (120). Furthermore, they engineered a hybrid sequence between *O. sativa* (rice) P5SM and *A. thaliana* P5SM for orthogonal gene expression in dicot plants (120). This Hybrid P5SM (HyP5SM) responds to monocot *Os*P5SM protein, but not dicot *At*L5, and is able to regulate transgenes in dicot *N. benthamiana* with almost 100-fold induction upon *Os*L5 co-expression (120). My colleagues showed that HyP5SM could regulate reporter genes and non-toxic genes orthogonally (120). In later chapters, as needed, I will discuss the details of the rational design that led to HyP5SM.

### Addressing Problems That Require Splicing-Based Regulation

It is no surprise that constitutive and conditional promoters have been developed as tools for artificial gene regulation before splicing-based regulation. *E. coli* and yeast, organisms used to research many basic pathways, do not use splicing except for rare examples. Conceptually, it is also more straight-forward to take a promoter and clone a gene of interest downstream. Chemically inducible promoter are already available for *E.coli*, yeast, plants, and mammalian systems (121–124). Researchers are accustomed to using promoters to regulate genes of interest. Why, then would one want to develop splicing-based gene regulation?

For this dissertation work, I present work testing the limits of the HyP5SM suicide exon as a biotechnological tool for gene regulation by attempting (and succeeding) to regulate the expression of pathogen effector proteins which trigger hypersensitive immune responses in plants (**Chapters 2 and 3**). My work shows that alternative splicing offers a new level of gene regulation which is readily compatible with existing promoter-based regulation. A researcher does not need to choose one or the other, but may instead decide to combine the two methods for improved control of gene regulation. **Chapters 2 and 3** demonstrate this with immune-triggering genes that cause programmed cell death in plants. The phenotype that these genes trigger cannot be regulated with inducible promoters alone due to leaky protein expression that accumulates past the threshold required for immune detection.

In **Chapter 4**, I present progress toward finding the putative SR protein(s) in plants which we believe promote P5SM exon definition in plants. In **Chapter 5**, I present the results of early attempts to re-engineer a hybrid P5SM suicide exon so that it is targeted by mammalian SR proteins in HEK293T cells.

# CHAPTER 2:

# Evaluating HyP5SM to Regulate the Hypersensitive Response in Plants

*Portions of this work were published in the following scientific journal:

Gonzalez, T. L.; Liang Y.; Nguyen B. N.; Staskawicz B. J.; Loqué, D.; Hammond, M. C. (2015) *Nucleic Acids Research*, **43**, 7152-61.


## Abstract

Disease-specific resistance is a highly desirable agricultural trait. Plants achieve this with effector-triggered immunity (ETI), a pathway activated when plant disease resistance (R) proteins recognize the presence of pathogen effector proteins delivered into host cells. The ETI response generally encompasses a defensive "hypersensitive response" (HR) that involves programmed cell death at the site of pathogen recognition. While many R protein and effector protein pairs are known to trigger HR, many components of the ETI signaling pathway remain elusive. Effector genes regulated by inducible promoters cause background HR due to leaky protein expression, preventing the generation of relevant transgenic plant lines. By employing the HyP5SM suicide exon, we have developed a strategy to tightly regulate effector proteins such that HR is chemically inducible and non-leaky. This alternative splicing-based gene regulation system was shown to successfully control Bs2/AvrBs2-dependent and RPP1/ATR1Δ51-dependent HR in *Nicotiana benthamiana* and *Nicotiana tabacum*, respectively.


## 2.1: Introduction to Plant Disease Resistance and Effector-Triggered Immunity

Plant disease and health are constant concerns for the agricultural industry and for worldwide food security. As climate change alters the agricultural landscape and creates more favorable conditions for plant pathogens, diseases that have been confined to sub-regions of the world may gain broader geographic footholds (125, 126). Understanding how plants successfully combat pathogens is crucial to maintaining healthy agriculture. Plants recognize pathogen associated molecular patterns (PAMPs) such as bacterial flagellin and fungal chitin, and initiate general PAMP-triggered immunity (PTI) responses (127, 128). However, successful pathogens circumvent this

primary innate immune system of plants by delivering effector proteins into plant cells to target and suppress the immune signaling pathway, thus making the plant susceptible to infection (128).

In the arms race between plants and pathogens, plants have evolved receptor proteins that distinguish strain- or race-specific pathogen effectors (128–130). Effector-triggered immunity (ETI) is a secondary plant innate immune pathway that is activated when plants possess the appropriate receptor, or "disease resistance protein" (R protein), to recognize a specific pathogen effector protein. The ETI response is associated with a hypersensitive response (HR) that involves localized cell death and generation of reactive oxygen species, thereby limiting growth of the pathogen. Although many R protein and effector protein pairs are known to trigger HR and disease resistance, the signaling events that lead to ETI and how much these pathways differ from PTI remains unclear; a lot more is known about how the pathways overlap (131–133).

In order to distinguish specific effector-triggered responses from PAMP-triggered responses, it would be beneficial to generate transgenic plants with an R gene that can inducibly express the corresponding pathogen effector protein to initiate the immune response. In addition, experiments with these types of transgenic plants would enable monitoring of transcriptional and biochemical differences at specific and short time points. However, part of the challenge facing ETI research is the limitation imposed by the HR phenotype. Bacterial, oomycete, and fungal pathogens generally produce effector proteins in their own cells and then deliver them into the host plant (128, 134–136). Thus, very low thresholds (137, 138) of pathogen effector protein are required to trigger HR, which makes generation of transgenic plants for mutant screens and transcriptome analysis difficult (139).

Herein, we describe employing a suicide exon in order to tightly regulate the expression of effector proteins, such that the HR is chemically inducible but background HR is fully suppressed (**Figure 2.1, 2.2**). Previously, we have shown that insertion of the HyP5SM splicing cassette (**Figure 2.3a**) into any gene of interest results in retention of a suicide exon by default, which triggers nonsense-mediated decay (NMD) of the non-productive spliced product (112, 120). Alternatively, co-expression with *Os*L5 protein results in skipping of the suicide exon (120). In this study, we establish that HyP5SM circumvents the problem of leaky transcription from a Dex-inducible promoter by effectively blocking protein expression. Whereas background levels of the effector AvrBs2 from the bacterial spot disease pathogen *Xanthomonas euvesicatoria* (140), previously called *Xanthomonas campestris* pv. *vesicatoria* (141), triggers visible HR in plants harboring the R gene *Bs2* (142), the immune response is suppressed by insertion of the HyP5SM splicing cassette into the *avrBs2* gene. Furthermore, we show that skipping of the suicide exon can be triggered by Dex induction of *Os*L5 expression, which leads to recovery of the HR phenotype by chemical activation of both the promoter and alternative splicing. We demonstrate that this dual regulation strategy is generalizable by showing that similar results are obtained for ATR1, an effector from the downy mildew pathogen *Hyaloperonospora arabidopsidis* (previously *Peronospora parasitica*), which triggers HR in tobacco plants harboring the R gene *RPP1* (143–145). In **Chapter 3**, we show that stably transformed plants harboring the resistance gene

and the dual regulated effector are viable, healthy, can be propagated through multiple generations, and can inducibly initiate the HR phenotype.

## 2.2: Results

### HyP5SM is functional in bacterial effector AvrBs2 and eliminates detectable leaky protein expression that leads to background hypersensitive response

AvrBs2 is a type III secreted effector protein highly conserved and required for virulence in strains of the bacterial spot disease pathogen *X. euvesicatoria* (140). The *Bs2* disease resistance gene from *Capsicum chacoense* (pepper) is functional in *Solanum lycopersicum* (tomato) and increases the yield of marketable tomatoes by 2.5-fold in field tests (142, 146). Unfortunately for studying this effector-R gene pair, we have found that leaky transcription from the Dex-inducible 6xUAS promoter (123) results in background expression of AvrBs2 protein, which activates HR in *N. benthamiana* plants co-transformed with Bs2 (**Figure 2.1a**). Accordingly, even without Dex treatment, *avrBs2* transcripts are detectable by RT-PCR (**Figure 2.3b**) and HA-tagged AvrBs2 protein is observed by western blot (**Figure 2.4a**). Thus, HR has prevented the generation of stable transgenic plants that inducibly express AvrBs2 in the presence of Bs2 (unpublished, D. Dahlbeck and B. J. Staskawicz).

In order to address this problem, we inserted the HyP5SM splicing cassette into the *avrBs2* coding sequence (**Figure 2.5**). It was previously shown that the splicing cassette minimally requires insertion after an AG dinucleotide which forms part of the extended 5' splice site sequence (120). Cassette insertion between codons encoding Glu and Val (E/V; GAG/GTA) as well as Glu and Pro (E/P; GAG/CCA) had been well tolerated in the context of *Enhanced Green Fluorescent Protein (EGFP),* which reveals that the extended 3' splice site could deviate substantially from the consensus sequence (120). However, in the context of the *avrBs2* coding sequence, splicing cassette function appears more sensitive to the extended 3' splice site context. Specifically, placing the splicing cassette between codons encoding E308 and V309 resulted in maintenance of splicing fidelity and regulation (**Figure 2.3c, 2.5**), whereas mis-splicing was observed for insertion between codons encoding E123 and P124 (**Figure 2.6**). Thus, we continued with the E/V insertion site for *avrBs2-HyP5SM-HA* and all text references relate to this construct unless otherwise stated.

RT-PCR analysis of *avrBs2-HyP5SM-HA* show that in the absence of the specific splicing factor *Os*L5, which is ribosomal protein L5 from the monocot *Oryza sativa* (rice) that binds the P5SM RNA element (120), the suicide exon is retained in the spliced product (SP-II) (**Figure 2.3c**). As expected, endogenous L5 protein does not result in observable background activation of the splicing cassette, as the HyP5SM cassette was engineered to be unresponsive to L5 from dicots such as *N. benthamiana* and *A. thaliana* (120). In contrast, constitutive co-expression of *Os*L5 promotes skipping of the suicide exon and leads to a shorter spliced product (SP-I, **Figure 2.3**). These results were confirmed by sequence analysis of the spliced products and are observed regardless of whether the gene construct is expressed via a constitutive (35S, a cauliflower mosaic virus promoter) or Dex-inducible (6xUAS) promoter. We did not

observe any of the splicing isoforms previously associated with aberrant splicing of the cassette in weak sequence contexts (120), however this analysis does not rigorously exclude all possible mis-splicing events.

Notably, the HyP5SM splicing cassette does not reduce leaky transcription from the Dex-inducible promoter, as transcripts are clearly observed by RT-PCR in the absence of Dex (**Figure 2.3c**). However, HyP5SM adds a second layer of regulation such that induction of alternative splicing is required for protein expression. Thus, we have nominally constructed a two-input regulatory system in which the chemical inducer Dex and the splicing factor *Os*L5 are required for full activation of gene expression (**Figure 2.1b, 2.2**). However, the observation that HR can be triggered with only one input (**Figure 2.4b, 2.7**) suggests that both transcription and exon skipping are leaky to some extent. The sensitivity of HR signaling is such that cell death is visible even when no protein is detected by a standard western blot (**Figure 2.4a, lane 8**). Additional experiments confirm that some exon skipping occurs upon Dex induction even without *Os*L5, presumably because the transcription is activated in excess of the levels of the endogenous splice factor that promotes exon retention (**Figure 2.8**). More importantly, the absence of both Dex and *Os*L5 eliminates HR (**Figure 2.4b**). Only mild chlorosis is observed on that leaf half, similar to results obtained for leaf infiltration with *Agrobacterium* harboring no AvrBs2 effector (**Figure 2.9**).

**Tight regulation of the HR phenotype by dual chemical induction of transcription and exon skipping**

Although omission of the splicing factor effectively suppresses protein expression from the HyP5SM-containing *avrBs2* construct, we considered that transgenic plant experiments would instead require an inducible copy of *Os*L5. However, as the inducible promoter employed would likely be leaky, it was an open question whether background expression of *Os*L5 would trigger sufficient AvrBs2 protein expression to cause HR. To test this, we performed the subsequent experiments with *Os*L5 under the control of 6xUAS. In this case, the two-input regulatory system is constructed such that the same chemical inducer Dex serves as both the input for activation of transcription and exon skipping (**Figure 2.2d**).

Promisingly, background expression of *Os*L5 in the absence of Dex does not trigger HR when the two-input *avrBs2-HyP5SM-HA* construct is employed, in contrast to the *avrBs2-HA* construct, which does not contain the splicing cassette (**Figure 2.10a**). Chemical induction of *avrBs2-HyP5SM-HA* and *OsL5* still produces the robust HR phenotype, although the onset of visible HR is delayed (< 16 hours) relative to the *avrBs2-HA* construct. Presumably this delay is due to leaky build-up of effector protein in the latter case, before the *N. benthamiana* leaves are Dex-induced (typically 16-18 hpi). Consistent with this interpretation, a time course analysis shows no detectable AvrBs2 protein expression until 16 h after Dex induction for the two-input system, whereas protein expression was detected as soon as 4 h after Dex induction for the one-input system (**Figure 2.11a, b**). Induction of high levels of AvrBs2 protein occurs at the same time as expression of *Os*L5 protein in the two-input system (**Figure 2.11c**).

Correspondingly, RT-PCR and western blot analyses support that chemical induction of *Os*L5 promotes exon skipping to make SP-I the major splice product, which

leads to high AvrBs2 expression (**Figure 2.10b, c**). Interestingly, in the absence of Dex induction, it appears that leaky expression of *Os*L5 can generate a small amount of AvrBs2 protein from the HyP5SM-containing construct, but the signal is much lower than observed for leaky expression of the *avrBs2-HA* construct (**Figure 2.10c**). We showed that this effect is probably due to leaky expression of *Os*L5, because no AvrBs2 protein is observed when co-expressed with empty pTA7001 vector, which carries another copy of GVG transcription factor (**Figure 2.11d**). Nevertheless, the results suggested that this combined promoter- and splicing-based regulation will be sufficiently tight to generate viable transgenic plants to study immune responses.

### HyP5SM is also functional in oomycete effector ATR1Δ51

Pathogen effectors are recognized by different classes of resistance proteins. The most characterized classes are CC-NB-LRR and TIR-NB-LRR, with a conserved nucleotide binding (NB) site and a leucine rich repeat (LRR) domain for auto-inhibition (147) and effector specificity (148). The N-terminal domain is known to be an important determinant for the downstream signaling pathway, with NDR1 being required for CC-NB-LRR protein signaling and the EDS1/PAD4/SAG1 complex being required for TIR-NB-LRR protein signaling (148–151).

AvrBs2 is recognized by Bs2, which has an uncharacterized N-terminal domain, but is homologous to the *Rx* gene (CC-NB-LRR) from potato (142, 152, 153). On the other hand, the ATR1 effector from the oomycete *H. arabidopsidis*, agent of downy mildew in *A. thaliana*, is recognized by the *A. thaliana* RPP1 resistance protein, which is in the TIR-NB-LRR class. For efficient expression in plants, we utilized a construct, ATR1Δ51, that has been truncated to remove the N-terminal eukaryotic secretion sequence and predicted translocation region (145). Co-expression of ATR1Δ51 and RPP1 produces weak HR in *N. benthamiana* and strong HR in *N. tabacum* (**Figure 2.9**), so we utilized the latter for our HR experiments.

In order to demonstrate that the HyP5SM splicing cassette can be applied to regulate a different effector protein in both *Nicotiana* species, we inserted the HyP5SM splicing cassette into the *ATR1Δ51* coding sequence. We made two versions that placed the splicing cassette between Glu and Ala codons (E/A sites) after codons encoding E128 or E168 (**Figure 2.5**). Both constructs maintained splicing fidelity and regulation (**Figure 2.12a, 2.13**). We continued with the E168/A169 insertion site for *ATR1Δ51-HyP5SM-FLAG* and all text references relate to this construct unless otherwise stated. Furthermore, regulation of protein expression by Dex induction and co-expression of *Os*L5 was observed by western blot against anti-FLAG (**Figure 2.12b, c**). In this case, no protein was detectable in the absence of Dex, showing that the one-input system is less leaky in this case than for AvrBs2 in *N. benthamiana*.

However, background HR in the absence of Dex treatment of *N. tabacum* is still observed for the one-input system upon co-expression of RPP1 and inducible ATR1Δ51 (**Figure 2.12d**). Onset of visible HR due to leaky expression is delayed (6-9 dpi) relative to constitutive or induced expression of ATR1Δ51 (2.5-4 dpi). As expected, the dual regulatory system using *6xUAS::ATR1Δ51-HyP5SM-FLAG* and *6xUAS::OsL5* eliminates the problem with background HR. Only mild chlorosis similar to that observed

for empty vector, which expresses the GVG transcription factor, is observed up to 13 dpi (**Figure 2.12d**, **2.13**). Furthermore, induction of HR with 30 µM Dex occurred on a similar timescale and strength for constitutive, one-input, and two-input systems (**Figure 2.13**).

Taken together, these data show that our system can be used to tightly regulate both highly leaky and less leaky effector proteins. Importantly, background ATR1Δ51 levels still exceed the threshold for triggering HR, which appears to be below the detection limit for western blots. In addition, we show that HyP5SM functions robustly in *N. tabacum*.

## 2.3: Discussion

To our knowledge, this study is the first time that regulation of transcription and exon skipping has been combined to make an artificial two-input regulatory system for use in plants. We show that the dual system is more tightly regulated (i.e. has lower background) than the promoter-only system, as evidenced by suppression of leaky HR (**Figure 2.1a**), an immune response sensitive to lower levels of protein expression than a western blot. Importantly, our study demonstrates the general principle that layering regulatory elements reduces the leakiness of gene expression, even when each regulatory element remains leaky. Background transcription from the 6xUAS promoter presumably occurs because some GVG transcription factor enters the nucleus even without binding dexamethasone (Dex). Background skipping of the HyP5SM suicide exon could result from limiting levels of the exon-defining splice factor that is proposed to bind HyP5SM RNA (112, 120) and/or from leaky expression of Dex-inducible *Os*L5. The dual regulatory system reduces leaky gene expression by requiring both transcription and exon skipping for protein production (**Figure 2.1b**).

By the same principle, maximum gene expression becomes dependent on both regulatory elements. Since full activation of exon skipping is dependent on the *Os*L5 protein, which in this case is also expressed using a Dex-inducible promoter, there is a delay in full induction of protein expression as compared to the leaky construct (**Figure 2.11**). In part, this effect is due to the leaky construct having a head-start in protein accumulation before Dex induction, relative to the non-leaky construct. Leaves are typically treated with Dex 16-18 h after infiltration and the observed delay is less than 16 h (**Figures 2.7, 2.11**), consistent with this conclusion. In addition, a threshold of *Os*L5 protein levels in the nucleus must be reached before splicing switches to the exon-skipped product as the major splice product, which likely reflects the equilibrium constants for competitive binding to HyP5SM in the cell. Since alternative splicing requires *Os*L5 binding to each precursor mRNA and the levels of *Os*L5 are probably limiting when activation of *avrBs2* transcription begins, this step may limit or slow build-up of AvrBs2 protein levels. This point is illustrated by comparing western blots when *Os*L5 is driven by the constitutive 35S promoter versus the Dex-inducible 6xUAS promoter. When co-expressed with *35S::OsL5*, AvrBs2 protein levels appear comparable 8 h after Dex induction with or without HyP5SM regulation (**Figure 2.4a**), whereas AvrBs2 protein from the two-input system is not yet seen at that time when using *6xUAS::OsL5* (**Figure 2.11**). Possible solutions to this issue include performing

Dex induction twice with optimized timing, or a using different inducible promoter to control *Os*L5 while the gene of interest is still under the Dex inducible system.

Although protein accumulation is more gradual with HyP5SM, chemical induction of the HR phenotype appears to be only slightly delayed (12 hours) for Bs2/AvrBs2-dependent HR (**Figure 2.7**) and not noticeably delayed for RPP1/ATR1Δ51-dependent HR (**Figure 2.13**). These results suggest that protein accumulation quickly surpasses the threshold required to induce HR. Indeed, we have shown that without *Os*L5 protein, Dex induction still promotes sufficient protein expression from *avrBs2-HyP5SM* to trigger HR (**Figures 2.4b, 2.7, 2.11**). The option to omit *Os*L5 simplifies the HyP5SM regulation system; researchers may clone the HyP5SM cassette into their gene of interest and utilize pre-existing plasmids with desired promoters, without needing to clone in an additional gene or co-infiltrate with an additional plasmid. However, we recommend that the stability and expression level of the protein of interest be considered before omitting *Os*L5. For example, we found that AvrBs2 protein accumulated to high levels and was easily visualized on a western blot, whereas ATR1Δ51 protein levels were consistently lower. Because of this, stable plant lines were generated using *Os*L5 with the RPP1/ATR1Δ51 system (see **Chapter 3**).

Because omission or inclusion of *Os*L5 affects maximum proteins levels of the HyP5SM-regulated gene (**Figure 2.11**), HyP5SM regulation is tunable to specific needs. For plant biology studies, it is often desirable to match endogenous levels of gene expression. Chemically inducible promoters can drastically overshoot native expression levels and are difficult to tune. Overexpression can lead to artifacts like loss of biologically-relevant regulation (154). For example, overexpression of Dex-inducible *Pseudomonas syringae* AvrPphB effector protein in transgenic *A. thaliana* resulted in acylation-independent HR in an otherwise acylation-dependent HR pathway (154). In contrast, the dual regulation system leads to moderate induction and even levels of protein expression that lasts at least 72 h after Dex treatment (**Figure 2.11**). In the future, we are interested in applying this dual regulation to achieve inducible, native levels of transgene expression in plants.

Another key advantage of our system is that it enables two independent promoters to be “stacked” in parallel to regulate a single gene, or possibly multiple genes. While here we showed the stacking of two identical promoters, this need not be the case. For example, plant biotechnology applications may involve targeting proteins to specific tissues or cellular compartments, or limiting protein expression to certain developmental stages. This may be achieved by expressing the HyP5SM-regulated gene of interest with a tissue- or stage-specific promoter along with an inducible *Os*L5 gene. Another successful approach to promoter stacking has been described, in which tissue-specific and chemically inducible expression of a gene of interest was effected by stacking two transcription factors/activators. Ethanol induction of GUS expression (*alcA::GUS*) is effected by the AlcR transcription factor under the control of a UAS promoter (*UAS::alcR*), which in turn is recognized by the mGAL4-VP16 transactivator, whose expression was localized to the endosperm in an *Arabidopsis* line from the Haseloff enhancer trap library (155). However, this latter approach is limited in scope to combining chemical induction and tissue specificity, whereas our strategy is generally applicable to combining any type of promoters. In fact, a third layer of regulation may be

achieved using our system by changing the promoter driving GVG from the constitutive 35S promoter to make a similar tissue-specific Dex-inducible system.

In conclusion, we have established that the strategy of combining promoter and suicide exon elements leads to tight regulation of pathogen effector proteins and prevents background triggering of even very sensitive phenotypes such as HR. We also showed for the first time that the HyP5SM splicing cassette functions in *N. tabacum* as well as *N. benthamiana*, and we expect that it will function in other dicot plants as well. A similar suicide exon may be engineered for monocot plants in a similar manner as has been described for the current HyP5SM (120) used here. Notably, unlike other inducible expression systems for plants, both the HyP5SM and *Os*L5 sequences are entirely derived from plants (120). If combined with native plant promoters, a fully plant-derived dual regulation system may be constructed. Finally, the dual regulation system we describe is generally applicable to genes that initiate severe or undesired phenotypes upon leaky expression, and provides a straightforward and promising way to generate many previously unattainable transgenic plants.

## 2.4: Figures

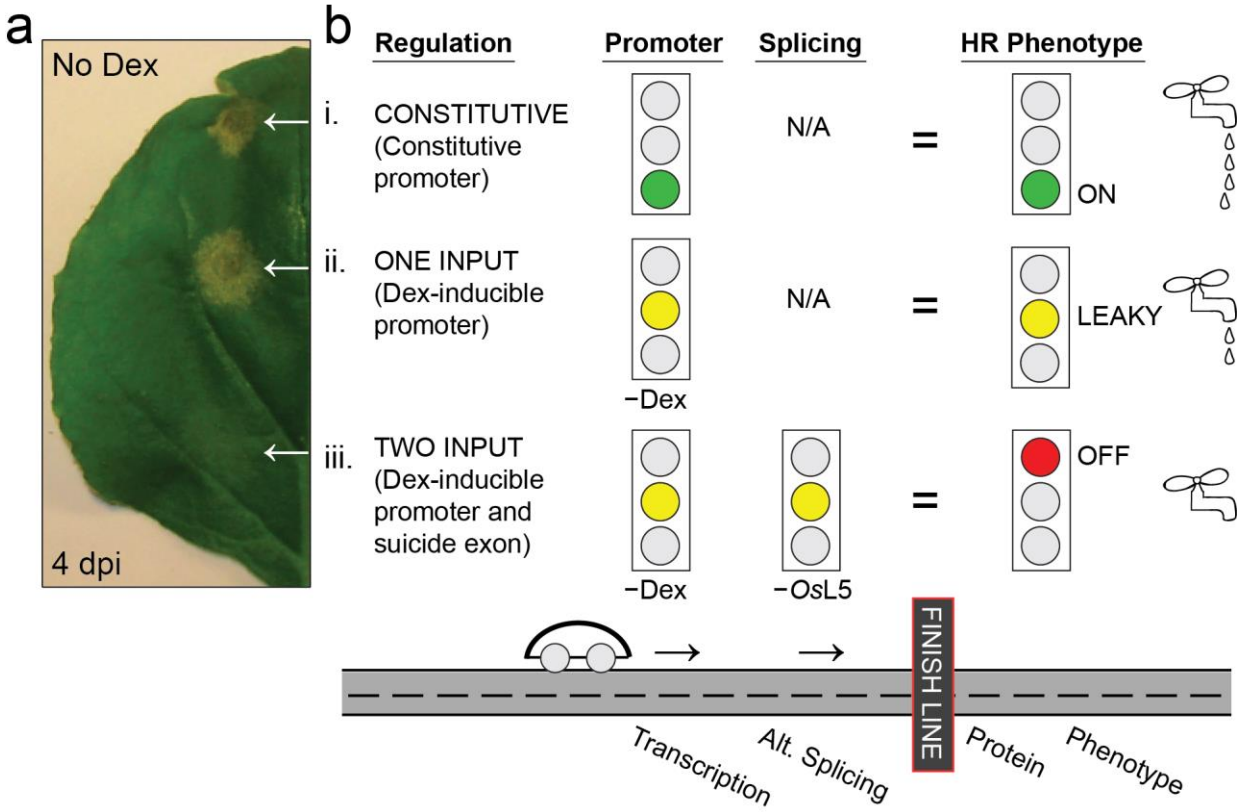


**FIGURE 2.1. Dual transcription and splicing regulation eliminates leaky hypersensitive response.** (a) *N. benthamiana* leaf with three spots transiently co-transformed with *Agrobacterium tumefaciens* to introduce *SuperPromoter*::*Bs2-HA* resistance gene and either i. constitutive (35S) *avrBs2-HA*, ii. Dex-inducible (*6xUAS*) *avrBs2-HA*, or iii. Dex-inducible *avrBs2-HyP5SM-HA.* The leaf was not treated with Dex and *OsL5* is not co-expressed. (b) Scheme representing how a two input system that combines promoter and suicide exon elements to control transcription and alternative splicing, respectively, results in tighter regulation of phenotype compared to a one input system (promoter alone).

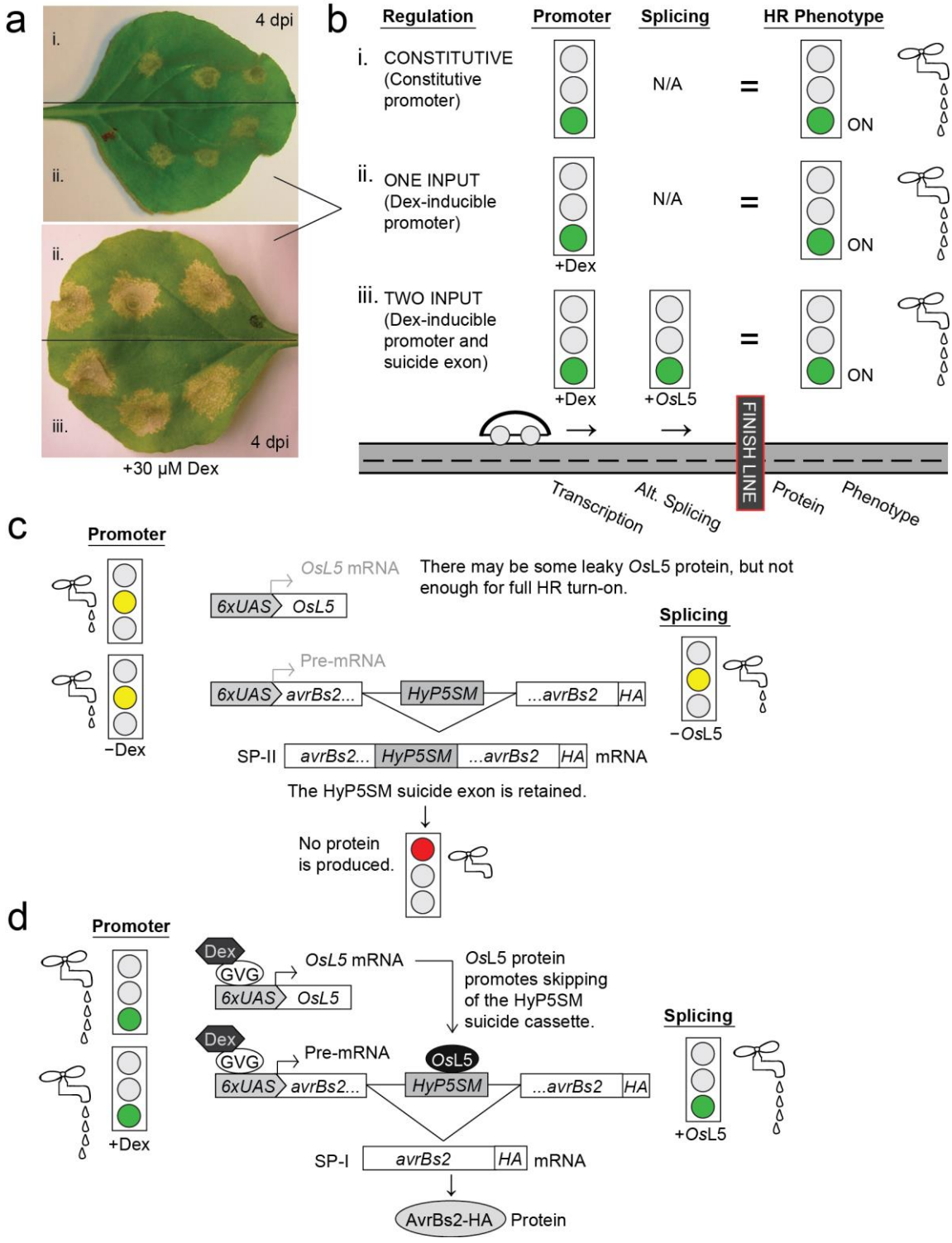


**FIGURE 2.2. Schematics of gene regulation with Dex induction.** (a) *N. benthamiana* leaves spot infiltrations to compare i. pMD1 *avrBs2-HA*, ii. pTA7001 *avrBs2-HA*, and iii. pTA7001 *avrBs2-HyP5SM-HA*. All spots were co-infiltrated with p1776 *Bs2-HA*. The bottom leaf was also co-infiltrated with pTA7001 *OsL5-6xHis* on both sides. Leaves sprayed with Dex. (b) Model of induced expression. (c), (d) Models showing how the two input regulation system (c) prevents leaky AvrBs2-HA protein and (d) can inducible promote AvrBs2-HA expression.

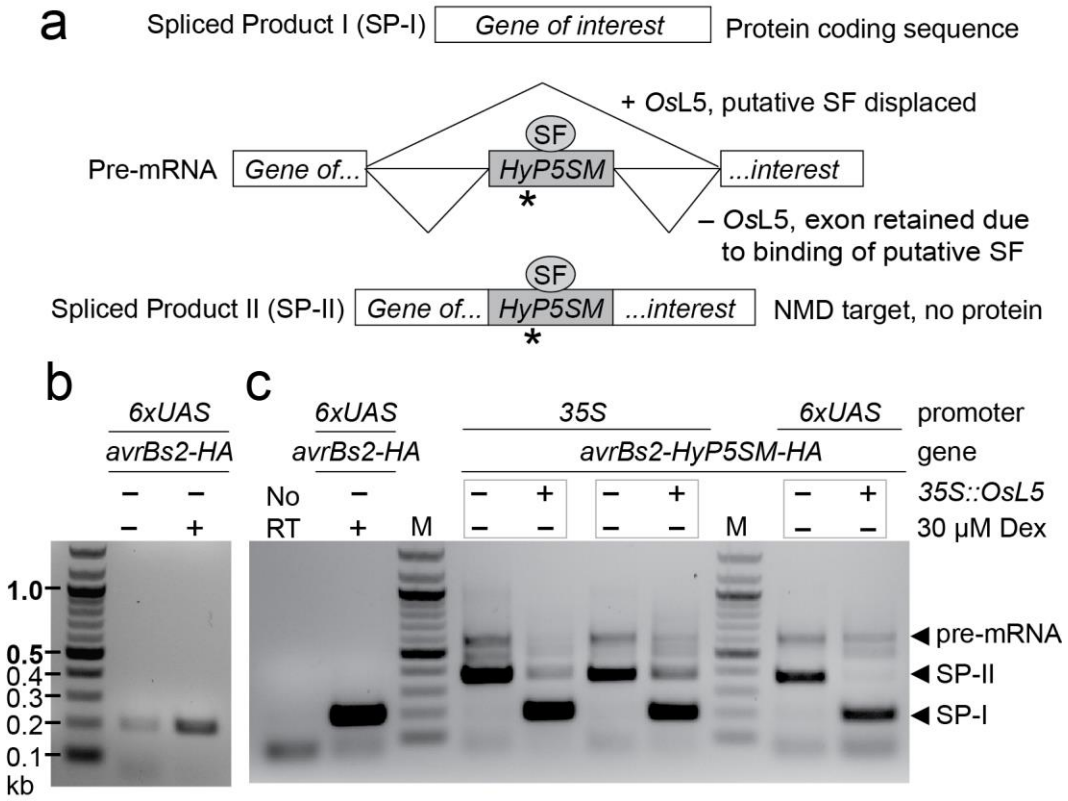


**FIGURE 2.3. HyP5SM functions in bacterial effector gene *avrBs2-HA*, but does not eliminate leaky transcription.** (a) Model of HyP5SM splicing regulation by *Os*L5 displacement of an endogenous splice factor (SF) that promotes exon retention by default. A premature termination codon (asterisk) is present in the HyP5SM suicide exon. (b) RT-PCR detection of *avrBs2-HA* transcripts from transiently transformed *N. benthamiana* leaves with and without Dex induction of transcription at 16 hours post-infiltration (hpi), with tissue collected at 2 dpi. (c) RT-PCR detection of spliced products for *avrBs2-HyP5SM-HA* with or without *Os*L5 induction of exon skipping, from tissue collected 2 dpi. Results are shown for the construct expressed either using 35S or 6xUAS promoters without Dex induction. After 35 cycles of PCR amplification, the 35S samples appear to be saturated, so levels should not be compared to the 6xUAS samples. No RT is shown as a negative control and *avrBs2-HA* from a different experiment is shown as a positive control for SP-I. Grey boxes indicate samples taken from different infiltrated halves of the same leaf.

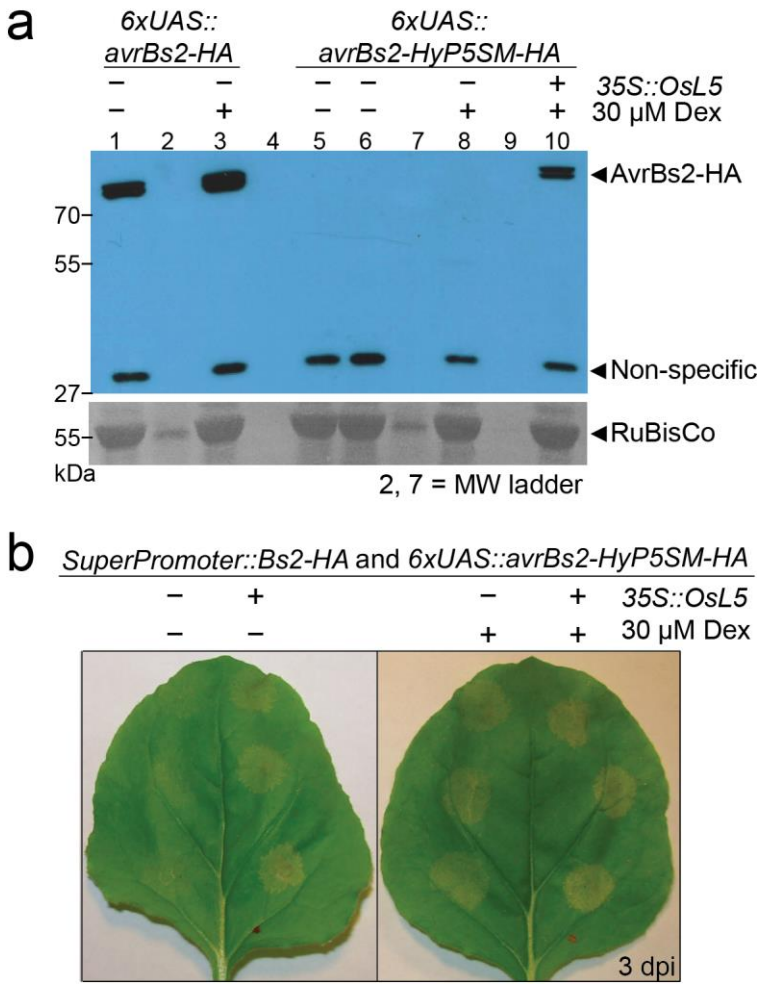


**FIGURE 2.4. The two-input regulatory system overcomes both leaky transcription and exon skipping to tightly control protein expression and HR phenotype.** (a) Anti-HA western blot to detect protein expression from *avrBs2-HA* or *avrBs2-HyP5SM-HA* constructs induced or mock induced at 18 hpi, with tissue collected at 26 hpi. Shown at the bottom is Ponceau S staining of the nitrocellulose membrane to visualize the large subunit of RuBisCo as a loading control. (b) Induction of the HR phenotype from the effector construct *avrBs2-HyP5SM-HA* requires co-transformation with *Bs2-HA* and either co-transformation with *OsL5*, treatment with 30 µM Dex, or both. Neither chlorosis nor HR is visible in the absence of both inputs. Lanes 4 and 9 are empty.

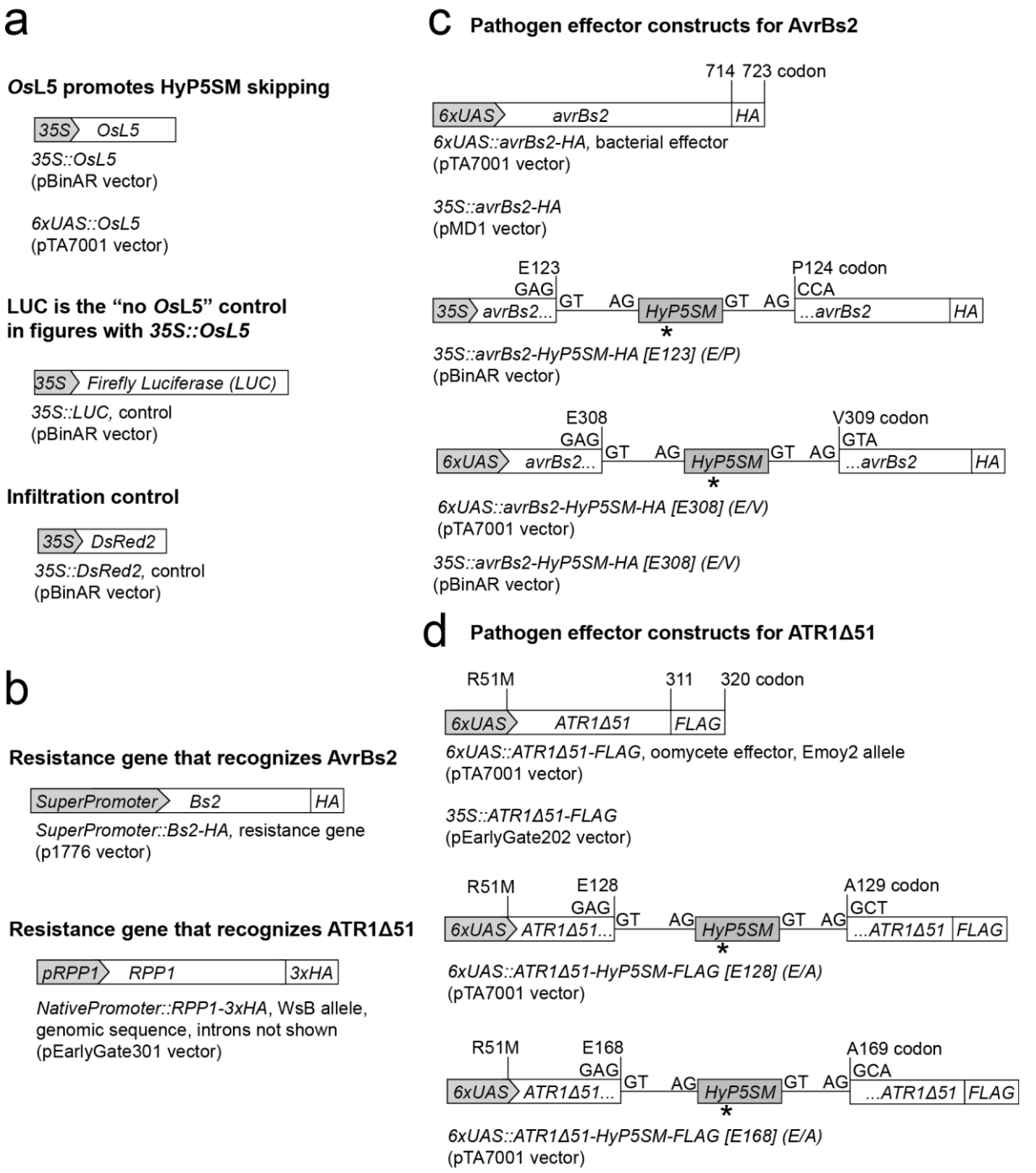


**FIGURE 2.5. Gene constructs used for transient transformations of *Nicotiana* plants in this study.** (a) Splicing effector and control constructs. (b) Resistance gene constructs. (c) Constructs for the bacterial effector AvrBs2 experiments. (d) Constructs for the eukaryotic pathogen effector ATR1Δ51 experiments. The N-terminal 51 amino acid truncation removes an eukaryotic exit signal.(156) The numbers above the constructs refer to amino acid codon positions. The dexamethasone inducible pTA7001 vector also includes *35S::GVG,* a transcription factor that binds the *6xUAS* promoter.(123) The p1776 vector has a strong constitutive chimeric octopine and manopine synthase promoter later renamed the "SuperPromoter" (157, 158).

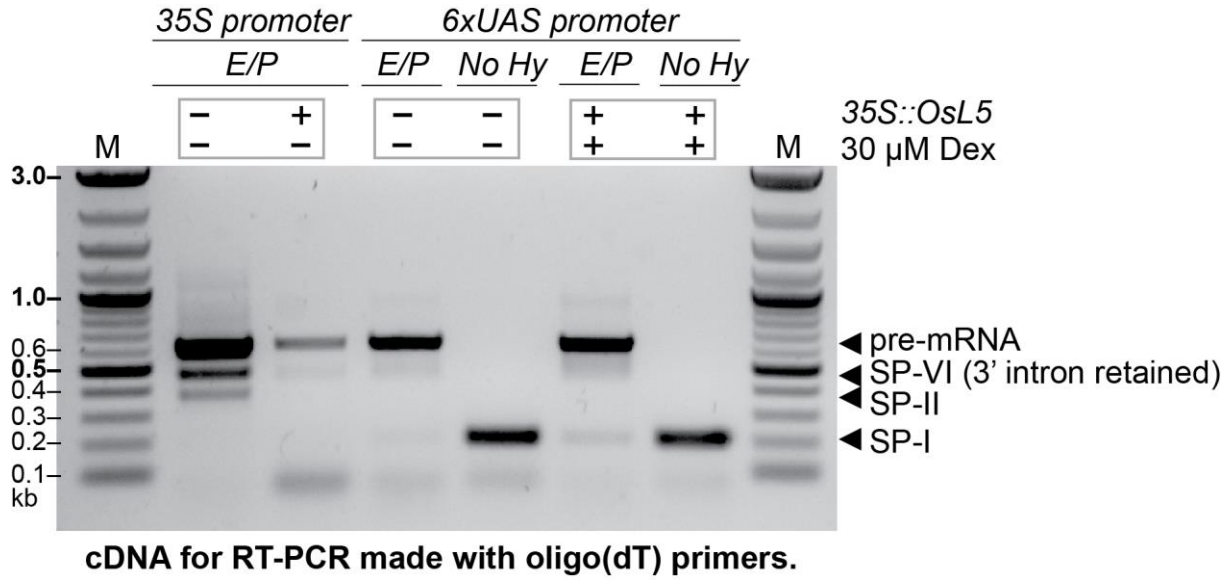


**FIGURE 2.6. The E/P insertion site for *avrBs2-HyP5SM-HA* does not splice efficiently.** RT-PCR for *avrBs2-HyP5SM-HA* (E/P insertion site) and *avrBs2-HA* ("No Hy") samples from *N. benthamiana*. Co-expression of *35S::OsL5* was expected to produce the SP-I splice product (HyP5SM skipped), but *avrBs2-HyP5SM-HA* (E/P) exhibits poor HyP5SM exon skipping. Major intron-retained splice products were sequenced and identified as pre-mRNA and SP-VI (SP-II + 3' intron). Gray boxes indicate that samples come from different halves of the same leaf. M = 2-log DNA ladder.

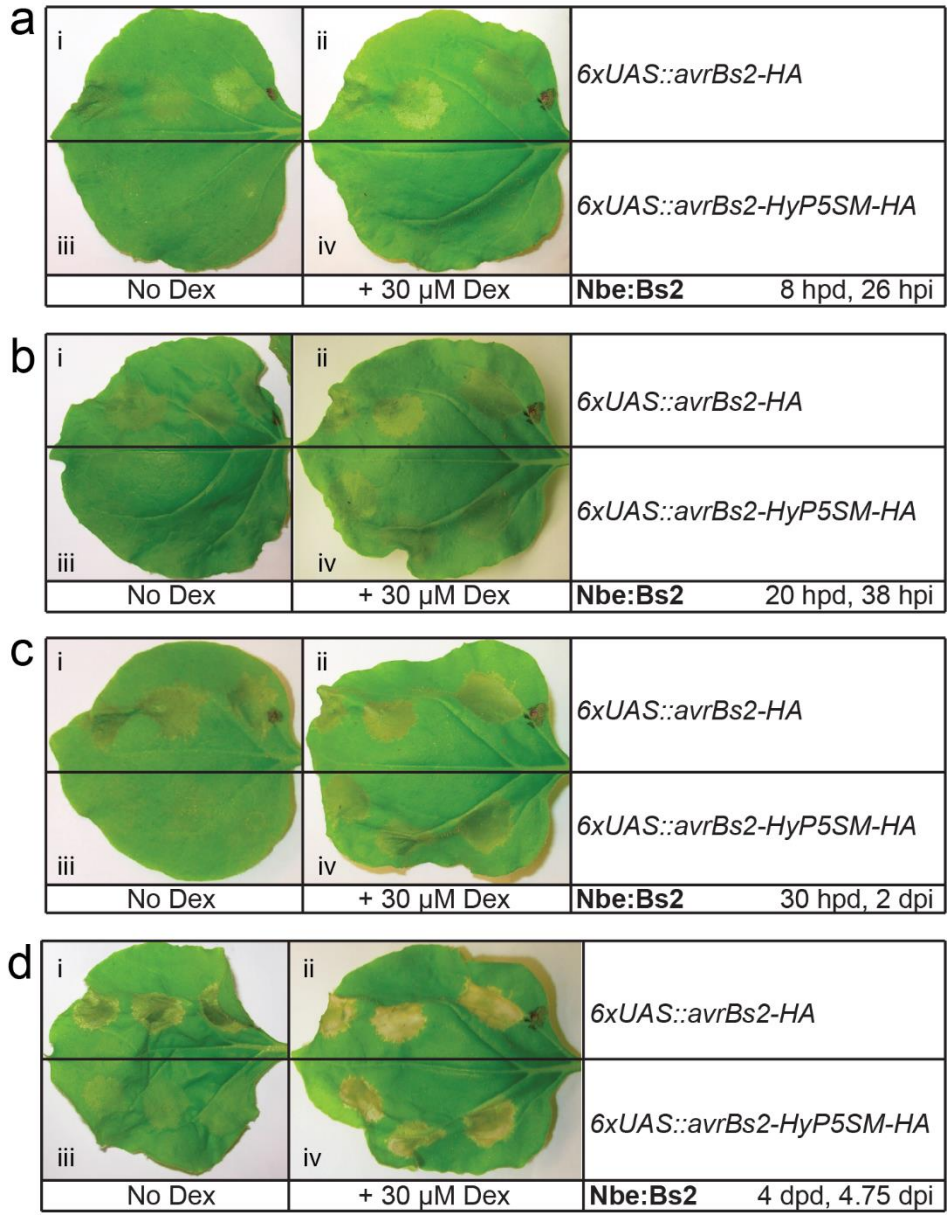


**FIGURE 2.7. *N. benthamiana* transgenic for Bs2 shows HR development from *6xUAS::avrBs2-HyP5SM-HA* over time, without *Os*L5.** Leaves of *N. benthamiana* stably expressing Bs2 were transiently infiltrated with either pTA7001 *avrBs2-HA* or pTA7001 *avrBs2-HyP5SM-HA*. They were either Dex induced or not induced at 18 hpi. Photos were taken (a) 8 h, (b) 20 h, (c) 30 h, and (d) 4 d later. With HyP5SM regulation, Dex-induced HR appears delayed 12 hours (a, b).

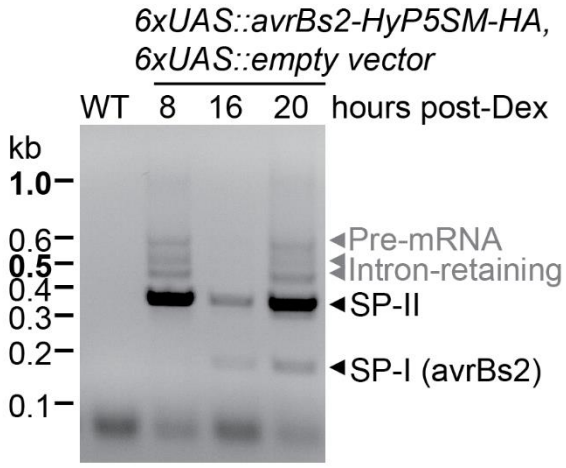


**FIGURE 2.8. RT-PCR showing SP-I from *avrBs2-HyP5SM-HA* visible in the absence of *Os*L5.** RT-PCR from *N. benthamiana* samples co-infiltrated with pTA7001 *avrBs2-HyP5SM-HA* and pTA7001 empty vector, in a parallel experiment accompanying the RT-PCR in Figure 4b. Samples were collected, processed, and run on an agarose gel alongside samples for Figure 4b. This image is cropped from the full gel.

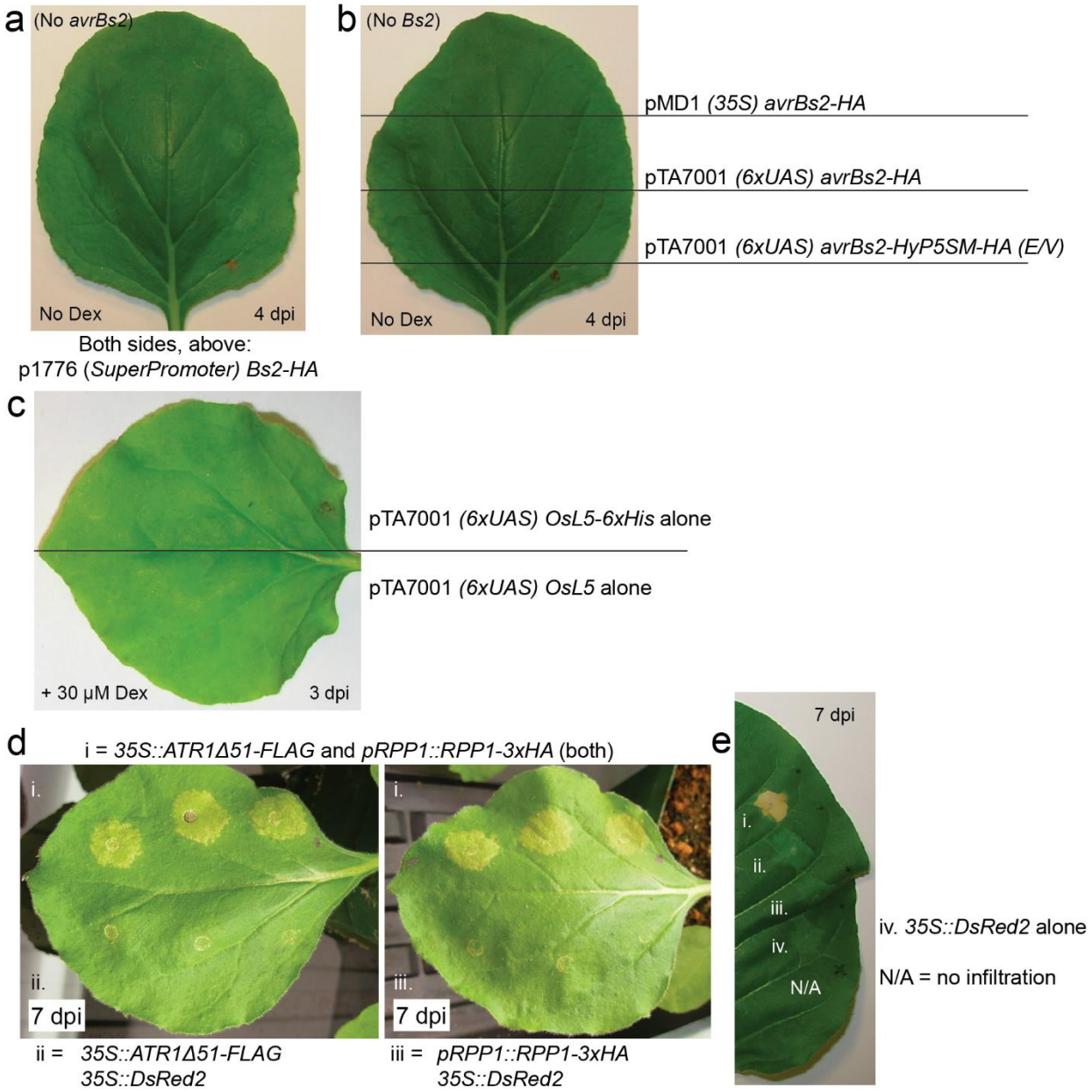


**FIGURE 2.9. Resistance gene and effector gene controls for the hypersensitive response.** *N. benthamiana* leaves were spot infiltrated or spot co-infiltrated with the indicated constructs using a needle-less syringe, with three spots on each leaf half. Total *Agrobacterium* (a-c) $OD_{600}$ = 0.75 or (d-e) $OD_{600}$ = 0.9. (a) *Bs2-HA* alone sometimes produces visible minor chlorosis (pale green color), but not the hypersensitive response. (b) *avrBs2* constructs do not initiate a hypersensitive response in the absence of the resistance gene *Bs2*. (c) The Dex-induced pTA7001 *OsL5* or *OsL5-6xHis* vectors alone do not result in a hypersensitive response. The red dot seen on all leaves near the stem was used to mark infiltrated leaves during the experiment. (d, e) Control infiltrations for ATR1 and RPP1. DsRed2 is added to keep $OD_{600}$ consistent. RPP1/ATR1Δ51-dependent HR is much stronger in (e) *N. tabacum* than (d) *N. benthamiana*.

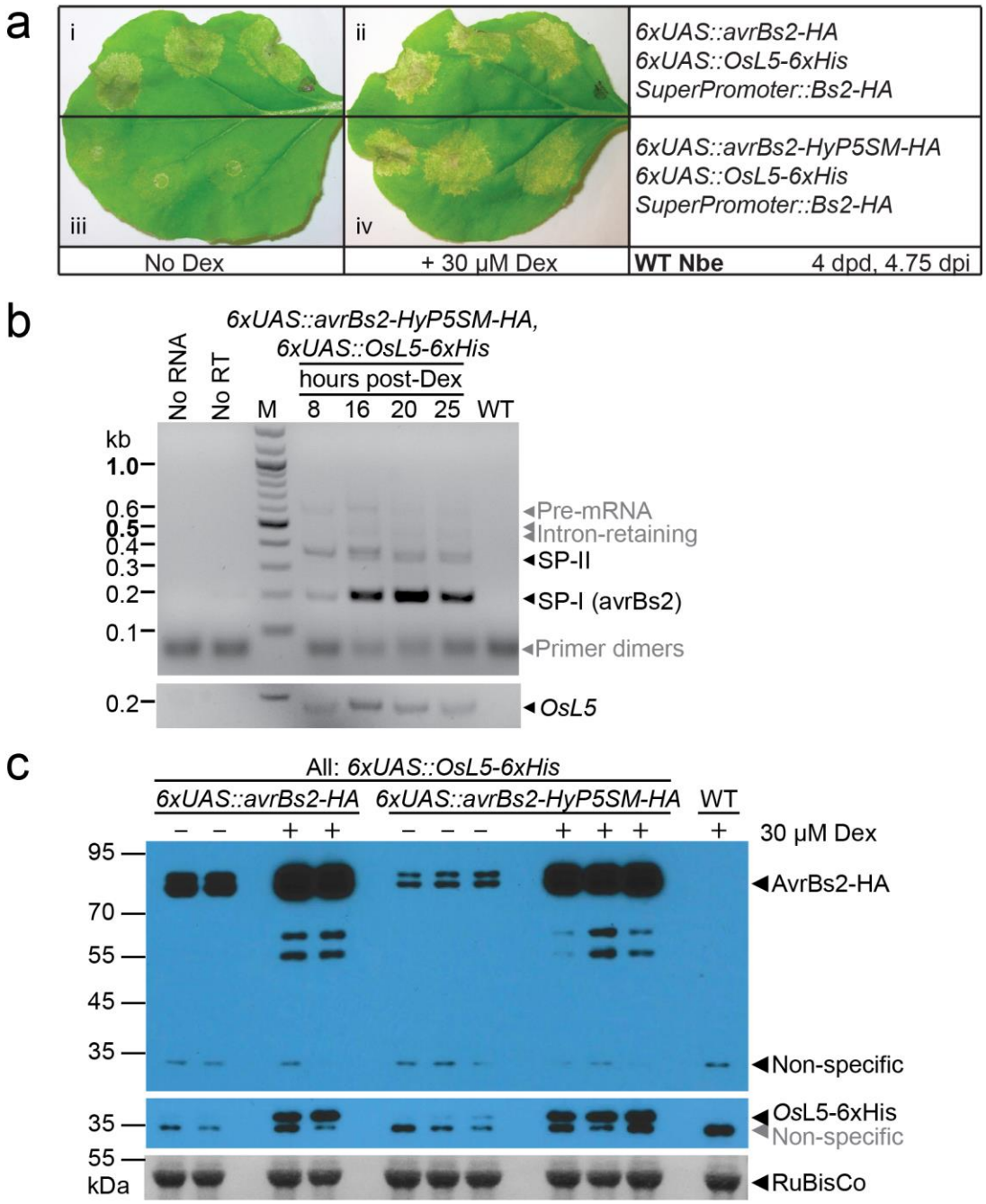


**FIGURE 2.10. Comparison of single and dual regulation by chemical induction.** (a) Representative wild-type *N. benthamiana* leaves were spot infiltrated to compare the extent of HR phenotype from (i, ii) pTA7001 *avrBs2-HA* (only regulated by the Dex-inducible promoter) or (iii, iv) pTA7001 *avrBs2-HyP5SM-HA* (also includes splicing regulation), with or without 30 µM Dex treatment. Leaf spots were also co-infiltrated with pTA7001 *OsL5-6xHis* and p1776 *Bs2-HA*. (b) Top: RT-PCR detection of *avrBs2-HyP5SM-HA* transcripts from transiently transformed *N. benthamiana* leaves at different time points post induction with Dex. Controls shown are no RNA or no RT enzyme in cDNA synthesis reaction, and RT-PCR from wild-type tissue. Bottom: RT-PCR detection of *OsL5* transcripts. (c) Top: Anti-HA western blot to detect protein expression from *avrBs2-HA* or *avrBs2-HyP5SM-HA* constructs induced or mock induced with Dex. Leaf tissue were also co-infiltrated with pTA7001 *OsL5-6xHis*, except for the wild-type control. Middle: Anti-*Os*L5 western blot. Bottom: Ponceau S stain of the large subunit of RuBisCo as loading control.

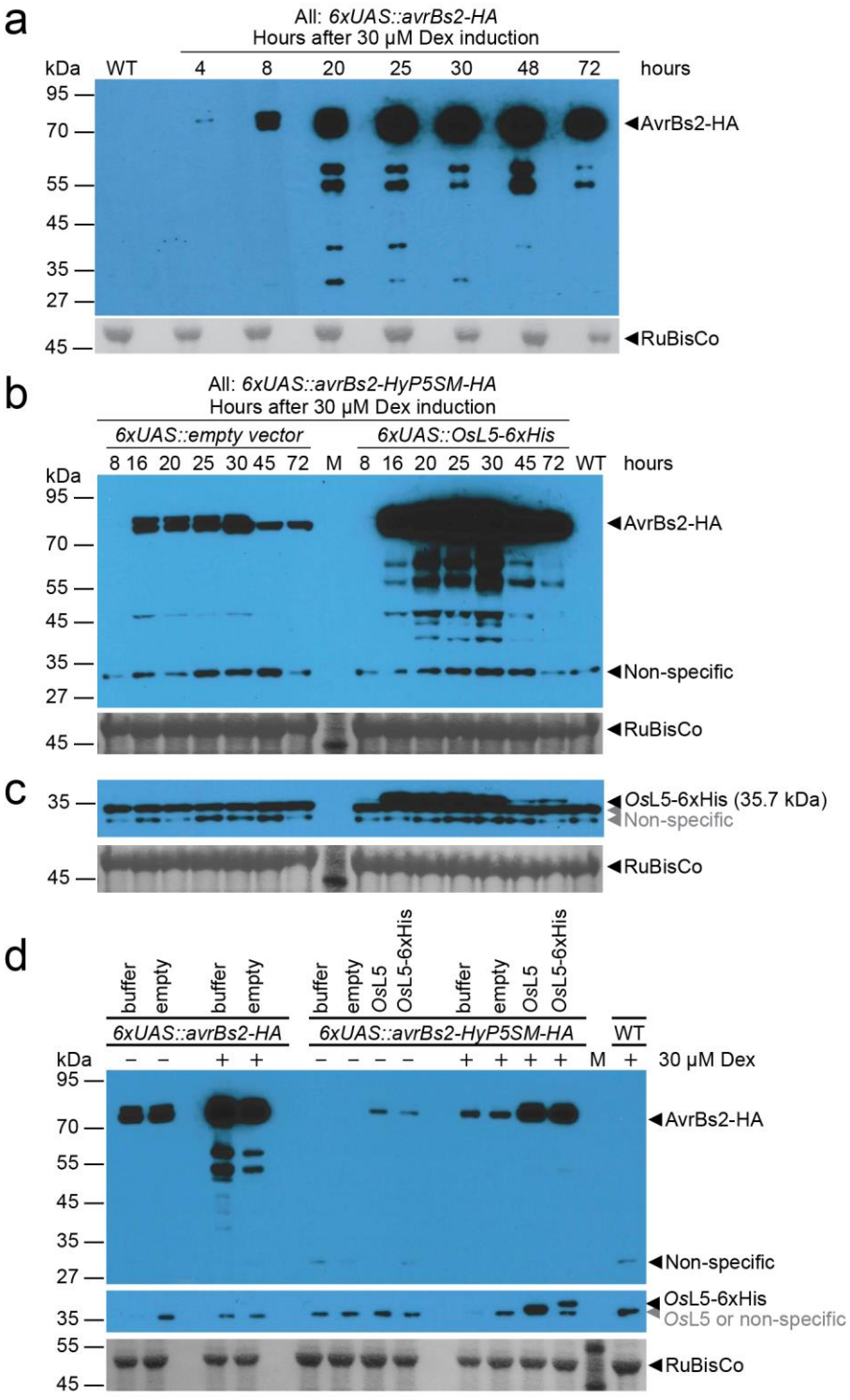


**FIGURE 2.11. Western blot time course of AvrBs2-HA and *Os*L5-6xHis protein expression from dual and single regulated constructs.** Western blot from *N. benthamiana* tissue transiently transformed, then induced with 30 µM Dex at 17 hpi. Tissue was collected at (a, b, c) the indicated times or (d) 24 h after treatment. (a, d) 5 µL or (b, c) 15 µL of crude proteins were run on 4-12% NuPAGE Bis-

Tris gels with 1X MOPS buffer, then transferred to nitrocellulose. (a, b) Anti-HA blot shows AvrBs2-HA protein development from induced (a) pTA7001 *avrBs2-HA* alone or (b) pTA7001 *avrBs2-HyP5SM-HA* co-infiltrated either with pTA7001 empty vector or pTA7001 *OsL5-6xHis*. (c) Anti-*Os*L5-6xHis blot shows development of *Os*L5-6xHis protein (35.7 kDa) at slightly higher molecular weight than a non-specific band which may potentially be endogenous *Nb*L5 protein (estimated 34.6 kDa). (d) Anti-HA western blot (top) and anti-OsL5-6xHis western blot (middle) comparing the extent of protein expression from one-input and two-input regulation systems. *N. benthamiana* leaves were co-infiltrated in 1:1 mixes with the indicated pTA7001 *avrBs2* constructs, plus either buffer, pTA7001 empty vector, pTA7001 *OsL5*, or pTA7001 *OsL5-6xHis*. The "buffer" control displays the leakiness of pTA7001 without the confounding variable of an additional copy of the GVG transcription factor in pTA7001, while also maintaining equal final $OD_{600}$ of all *avrBs2* constructs. Lanes 3, 6, 11 are empty. Ponceau S stain of RuBisCo large subunit is shown as a loading control below each western blot. M = NEB #P7711 marker. WT = wild-type *N. benthamiana*.

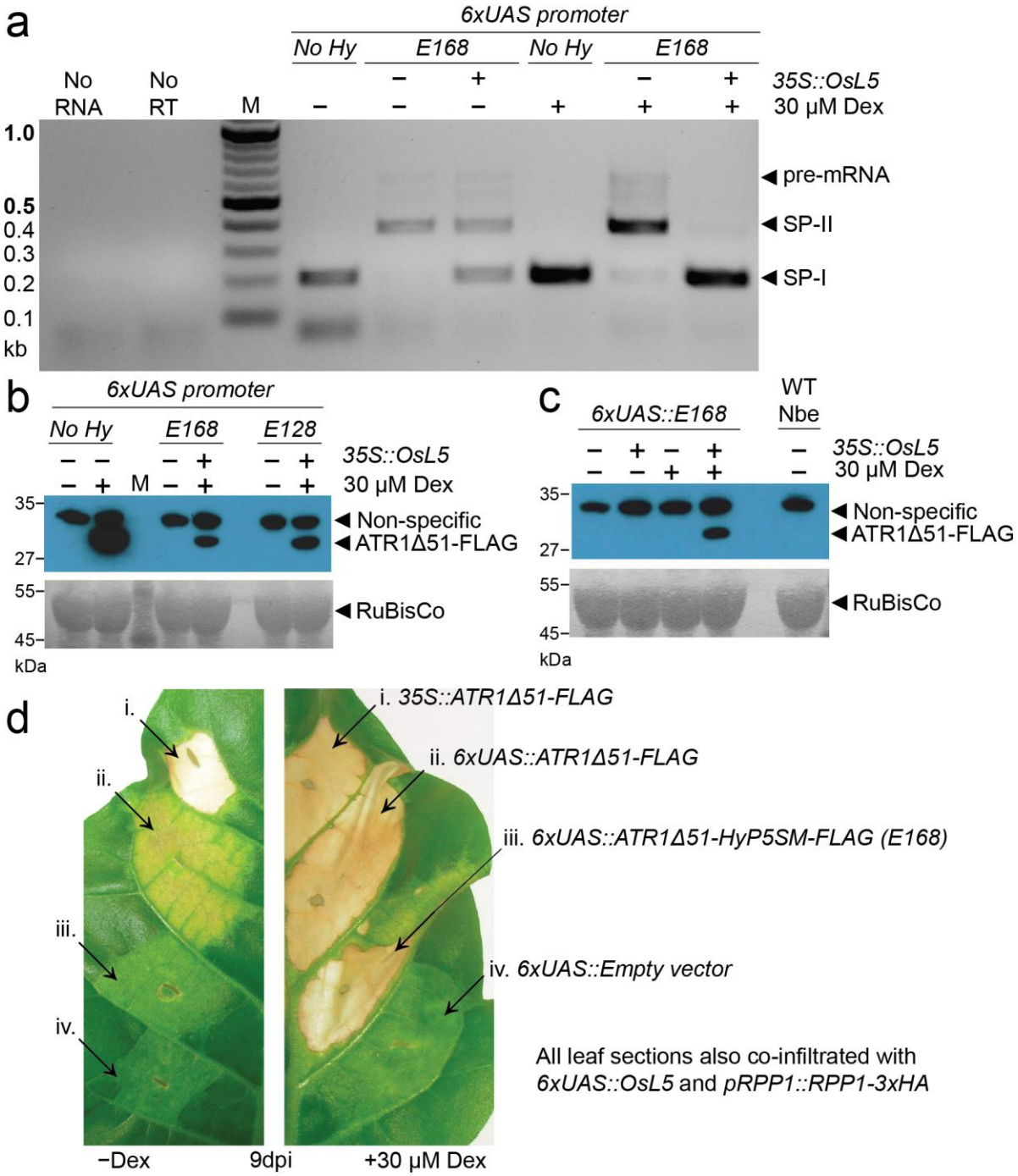


**FIGURE 2.12. HyP5SM tightly regulates the hypersensitive response induced by the oomycete effector gene *ATR1Δ51-FLAG* in *N. tabacum.*** (a) RT-PCR detection of transcripts from *ATR1Δ51-FLAG* ("No Hy") or *ATR1Δ51-HyP5SM-FLAG* ("E168") transiently transformed in *N. benthamiana* leaves. The constructs were expressed under the control of 6xUAS, with and without Dex induction, and with and without OsL5 co-expression. (b, c) Anti-FLAG western blot using *N. benthamiana* tissue from the same experiment in (a). The RuBisCo large subunit visualized by Ponceau S staining of the nitrocellulose membrane serves as a loading control. (d) Comparison of the HR phenotype in *Nicotiana tabacum* for constitutive (i), Dex-inducible promoter (ii), and Dex-inducible promoter and splicing factor (iii) regulated *ATR1Δ51-FLAG*. Empty vector (iv) serves as a negative control. Left panel shows the extent of leaky HR without Dex induction and right panel shows induction of HR with 30 µM Dex.

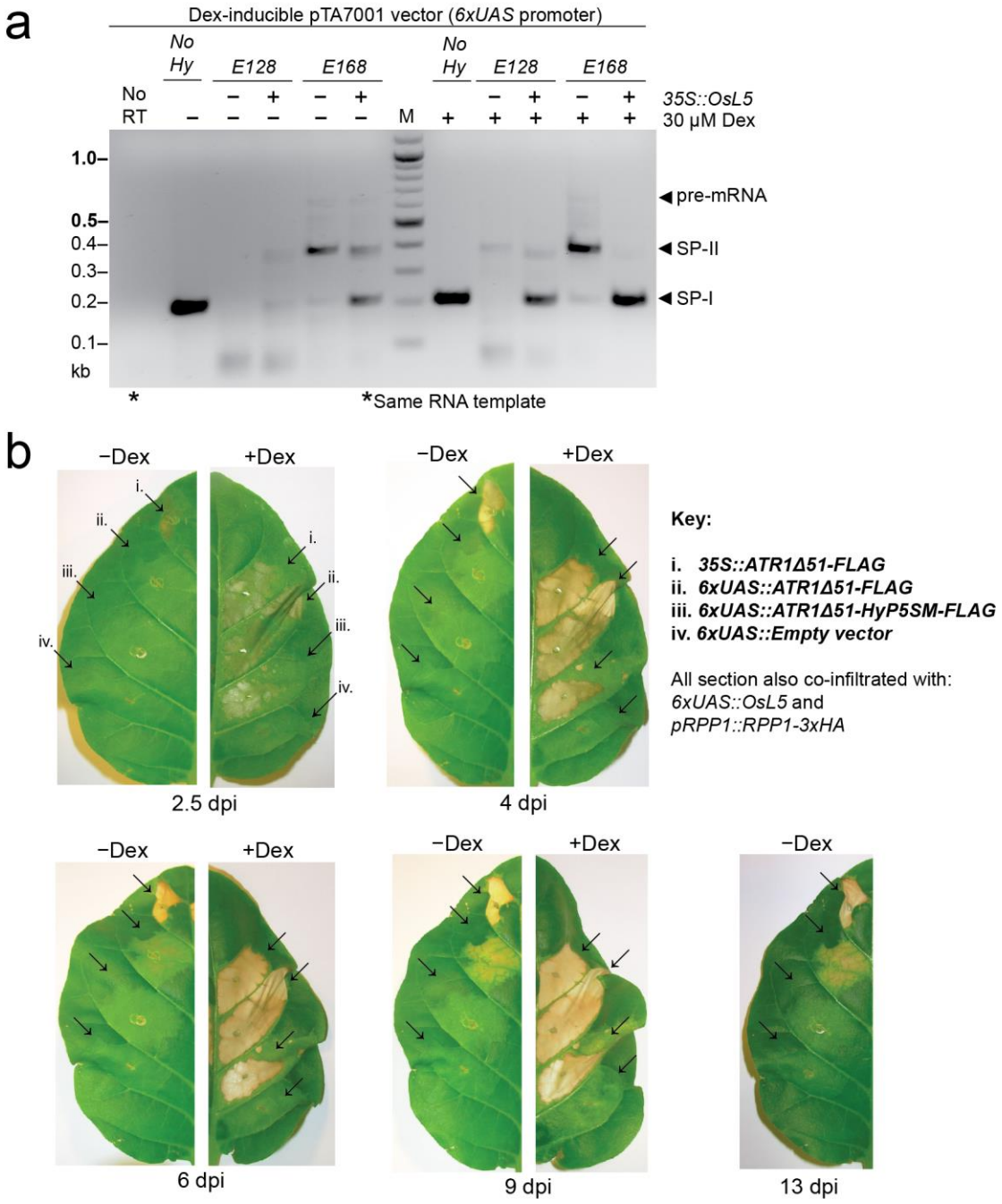


**FIGURE 2.13. RT-PCR analysis of *ATR1Δ51-HyP5SM-FLAG* constructs and HR time course.** (a) RT-PCR on 1.5% agarose gel of *6xUAS::ATR1Δ51-HyP5SM-FLAG* (with HyP5SM inserted after E128 or E168 codons) and *6xUAS::ATR1Δ51-FLAG* ("No Hy"). "No RT" is no reverse transcriptase control for genomic DNA contamination. RNA was extracted from *N. benthamiana* tissue from the same experiment described in Figure 5a. (b) Time course showing development of the hypersensitive response phenotype in transiently transformed *N. tabacum* leaves. The images follow two leaf halves through time. Because Dex is cell permeable, the two leaf halves are from different leaves. The leaky hypersensitive response from *6xUAS::ATR1Δ51-FLAG* is delayed, but apparent around 6 dpi.

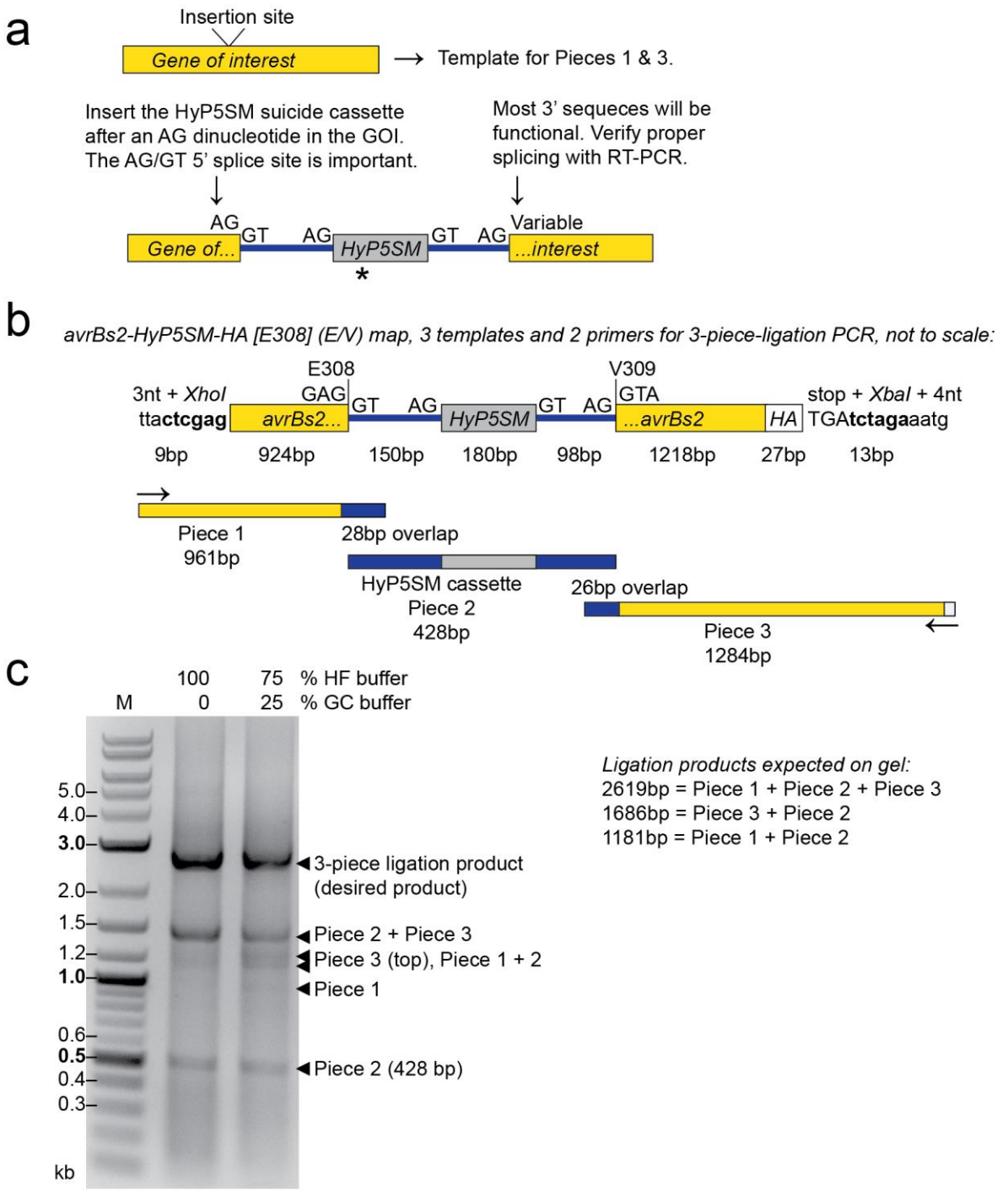


**FIGURE 2.14. Insertion of the HyP5SM cassette into *avrBs2-HA*.** (a) Chose an insertion site after an AG dinucleotide (120). The 5' intron provides the GT dinucleotide to complete the minimal AG/GT 5' splice site. (b) HyP5SM is introduced into the gene of interest using 3-piece-ligation PCR. The three template pieces and final product are shown. Introns are in blue. Exons are in yellow and gray. The 3xHA tag is white. (c) 1% agarose gel showing the 3-piece-ligation products. "HF buffer" and "GC buffer" refer to the buffers for Phusion HF DNA Polymerase (New England Biolabs). HyP5SM contains a premature termination codon (asterisk).

## 2.5: Materials and Methods

### Oligonucleotides and DNA Constructs

Sequences for all synthetic primers used for performing PCR and RT-PCR experiments are described in **Table A1**. The HyP5SM splicing cassette sequence was inserted into constructs by overlap extension PCR as discussed in Supplemental Methods (**Figure 2.14**) and a previous publication (120). For *avrBs2* constructs, silent mutations to the proximal codons were introduced to conform to the extended splice site consensus, and are indicated in **Table A1**. Effector constructs were cloned into the pBinAR vector ($Kan^R$) (159) for constitutive expression using the CaMV 35S promoter and into the pTA7001 vector ($Kan^R$) (123) for Dex-inducible expression using the 6xUAS promoter. The *avrBs2-HA* gene was cloned from pMD1 *avrBs2-HA* (160) and *ATR1Δ51-FLAG* (Emoy2 allele) was cloned out of pENTR/D-TOPO *ATR1Δ51-FLAG* (145). The Staskawicz lab provided the pTA7001 empty vector (originally from Aoyama and Chua) (123), the pEG202 *35S::ATR1Δ51-FLAG* vector (145), and the R gene vectors, p1776 *Bs2-HA* and pEG301 *pRPP1*::*RPP1-3xHA* (WsB allele) (145). The *Bs2-HA* gene was driven by a constitutive chimeric octopine and manopine synthase promoter ("SuperPromoter") (158, 161) in the p1776 vector ($Kan^R$) (157). The SuperPromoter results in higher expression than the 35S promoter (161). The HyP5SM cassette, pBinAR *OsL5*, and control pBinAR *firefly luciferase (LUC)* are from the original HyP5SM paper (120). *OsL5* and *OsL5-6xHis* were amplified from pBinAR *OsL5* and inserted into pTA7001 by restriction digest cloning.

### *Agrobacterium*-mediated Transient Leaf Transformations

*Agrobacterium tumefaciens* strain GV2260 ($Rif^R$, $Carb^R$) was used for all plasmids in the AvrBs2 experiments except for p1776 *Bs2-HA*. *A. tumefaciens* strain GV3101 ($Rif^R$, $Gent^R$) was used for p1776 *Bs2-HA*, the multi-gene pTKan vectors, and all plasmids in the ATR1Δ51 experiments. *Nicotiana benthamiana* plants were grown 3 weeks in a greenhouse (Oxford Tract Greenhouse Facility), then transferred to an open growth cart under constant light and ambient temperatures (24-27°C) for approximately 1 week before experiments. *Nicotiana tabacum* var Turk plants were grown in a growth chamber under 16/8 hour day/night cycles, then transferred to the open growth cart 2 days before experiments.

*Agrobacterium* was grown in LB Miller broth with appropriate antibiotics (all 50 µg/mL) at 28°C. *Agrobacterium* was grown in 5 mL starter cultures overnight, then added to 30 mL selection media and grown for another 20-24 h. All liquid cultures were shaken at 225 rpm. Cells were collected with 10 min 4700 rpm. Cells were washed (10 mM MES, 10 mM $MgCl_2$, pH 5.6), then collected again. Cells were resuspended with induction buffer (wash buffer + 150 µM acetosyringone) and incubated for 2-3 h. Cell density was normalized with induction buffer to $OD_{600}$ = 0.5 or 0.75. Normalized Agro solutions were mixed 1:1 or 1:1:1, for a final $OD_{600}$ = 0.25 of each construct. R gene was included in HR experiments and omitted for RT-PCR and western blot experiments to avoid HR. Plant leaves were infiltrated with *Agrobacterium* solutions using a 1 mL needle-less syringe.

**Dexamethasone Induction and Tissue Collection**

Dexamethasone (Sigma, CAS 50-02-2) was dissolved in ethanol to make a 30 mM stock solution. *Nicotiana* plants were induced with 30 µM dexamethasone (+0.1% ethanol, in Agro wash buffer) at 15-20 hpi, kept in the dark with restricted air circulation for 2-4 h, then returned to regular growth conditions with constant light. For *N. benthamiana,* injecting with a needle-less syringe or thoroughly spraying the leaves were both effective induction methods in our hands, but spraying was preferred because it caused less tissue damage. For *N. tabacum*, only induction by needle-less syringe was tested.

At 8-72 hours post-dexamethasone induction (hpd), 2-4 leaf discs were collected using the end of a 1 mL pipette tip (approximately 15-30 mg tissue), snap frozen in liquid nitrogen, and stored at -80°C until ready for processing. RNA or crude proteins were extracted from tissue samples pulverized with stainless steel beads and TissueLyser II (Retsch) for 25 Hz for 15-30 seconds in pre-cooled adapters.

**RT-PCR**

RNA was extracted from pulverized tissue using GeneMATRIX Universal RNA Purification Kit (CHIMERx) with RL buffer. RNA was treated with RQ1 DNase (Promega), then 400-1000 ng of RNA was used for cDNA synthesis using iScript Select cDNA Synthesis kit (BioRad) or SuperScript III Reverse Transcriptase (Thermo Fisher Scientific). Splicing was assessed using PCR with Taq polymerase (New England Biolabs) and primers designed ~100 bp outside the splice sites such that the size of amplified SP-I would be approximately 200 bp. Primers are described in **Table A1**.

**Protein Extraction, Western Blots, and Ponceau S Stains**

Total protein was extracted from pulverized tissue using 200-300 µL of Laemmli (162) Buffer (0.24 M Tris-Cl pH 6.8, 6% SDS, 30% glycerol, 16% β-mercaptoethanol, 0.006% bromophenol blue, 10 M urea) as previously described (145). Extract was run on denaturing 10% or 4-12% Bis-Tris NuPAGE gels with MOPS buffer (Invitrogen), then transferred to nitrocellulose membrane using a tank transfer at 300 mA for 1.5 h (for AvrBs2-HA) or 1 h (for ATR1Δ51-FLAG). Membranes were stained for total protein with Ponceau S (0.5% Ponceau S, 1% acetic acid), and the large subunit of RuBisCo (approx. 55 kDa) is shown below western blots as a loading control. Membranes were blocked overnight with 5% milk in TBST buffer (20 mM Tris-Cl, 0.5 M NaCl, 0.05% Tween-20, pH 7.5). AvrBs2-HA protein was detected using 1:1000 of rat Anti-HA HRP-conjugated antibody (Roche, 3F10). ATR1Δ51-FLAG was detected using 1:2000 of mouse Anti-FLAG M2 HRP-conjugated antibody (Sigma, A8592). Antibodies for *Os*L5-6xHis were made by expressing and purifying exotoxinA-*Os*L5-6xHis, then submitting this protein to a commercial vendor (Josman, LLC) for generation of *Os*L5-6xHis-specific antibodies in rabbits. *Os*L5 protein was detected using rabbit anti-*Os*L5-6xHis 1:40,000 and goat anti-rabbit HRP-conjugated 1:5000 (BioRad, 170-6515). SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher Scientific) and film paper developed in a dark room were used to visualize the location of the HRP enzyme.

## 2.6: Additional Methods: Cloning the HyP5SM Cassette Into a Gene of Interest

*Step 1: Choose an insertion site in the sequence of the gene of interest*.

Three sequence elements are essential for proper splicing: the 5' splice site, the branch site, and the 3' splice site (163). The branch sites are internal components of introns and thus not subject to change when choosing a site for HyP5SM cassette insertion into the gene of interest. The splice sites, on the other hand, are formed by sequence at the junctions of exons and introns. The HyP5SM cassette is comprised of an intron-exon-intron sequence. The gene of interest will form exons on either end of the HyP5SM cassette, thus generating two *de novo* exon/intron and intron/exon junctions; for simplicity, the exon sequences of the new junctions are described here as the 5' splice site and 3' splice site, respectively, since the intron sequence does not change. Because splicing-based gene regulation is not yet a common biotechnological tool, choosing an insertion site may seem daunting to new users. However, a key feature of the HyP5SM cassette is that the sequence requirements are very minimal. This has simplified implementing HyP5SM in new contexts. This section highlights essential requirements for the 5' splice site and discusses non-essential, but helpful tips for the 3' splice site.

The most conserved splice site is the 5' splice site (12). The HyP5SM cassette must be cloned after an **AG** dinucleotide in the gene of interest in order to form a functional 5' splice site (120). The following silent mutations may be used if cloning the cassette in frame. Fortunately, dicot plants do not have strong codon preferences for these amino acids (164).

| | |
|---|---|
| Glutamate (E): | GAA → GAG |
| Glutamine (Q): | CAA → CAG |
| Lysine (K): | AAA → AAG |

In our hands, we have found that HyP5SM is functional when out-of-frame as well (unpublished). If silent mutations are unwanted, the HyP5SM cassette may be cloned after an **AG** dinucleotide in any frame. The HyP5SM exon has premature termination codons in all three reading frames, so it can function as a suicide exon in any frame.

The 3' splice site is variable. The 3' splice site closest to the consensus sequence is **GU**, but many other sites will work as well. The engineered HyP5SM cassette contains introns from the endogenous *A. thaliana* P5SM cassette, which has a 3' splice site of **AGA** (112). In Hickey *et al.* 2012 Figure 2, mutational analysis of the endogenous 3' splice site for *At*P5SM showed proper splicing even with a proline codon **CCA** which resembles the consensus sequence very poorly (120). However, when **CCA** was used as the 3' splice site for *avrBs2-HyP5SM* (E/P insertion site), it resulted in improper splicing (**Figure 2.6**), and so the use of **CCN** as a 3' splice site is discouraged for future user. Based on the consensus sequence for 3' splice sites in dicot plants (12), the following 3 nucleotide pattern is expected to function well: **(G|A)(T|A|G)(A|T|G|C)**. The nucleotides are listed from most-to-least common for each position, albeit the consensus for each consecutive position becomes weaker. This suggested working

consensus includes both the ideal **GTN**, as well as the endogenous *At*P5SM **AGA** 3' splice sites.

For best results, clone 2-3 versions of the gene of interest with HyP5SM in different insertion sites and check the splicing of each construct with RT-PCR. Gel extract and sequence the splice products to verify that the HyP5SM cassette is splicing as expected. Choose the best insertion site empirically.

Example with *avrBs2:*

***E/V (GAA/GTG → GAG/GTA)*** – Silent mutations introduce the essential AG dinucleotide for the 5' splice site and make the 3' splice site closer to the HyP5SM endogenous 3' splice site (AGA). The second silent mutation may not have been necessary because this insertion site is close to the minimal consensus 3' splice site for plants (GTN) (12).

*Step 2: Make the three templates for 3-piece-ligation PCR (also called extension PCR).*

Make three separate dsDNA templates with a high fidelity polymerase.

Piece 1:

[NNN][XhoI site][ATG---*avrBs2 first part*---GAG][GT--5'intron overhang--

Template: *avrBs2-HA*

Primers: The *avrBs2* forward cloning primer (TLG45). The reverse primer (TLG44) anneals to *avrBs2* at the chosen insertion site and adds an overhang. This same sequence overhang can be used for cloning HyP5SM into other genes.

Piece 2:

[GT---5'intron---AG]**[HyP5SM exon]**[GT---3'intron---AG]

Template: Any gene with the full HyP5SM cassette already cloned into it. In this case, we used *EGFP-HyP5SM* (120).

Primers: Primers DNA37 (forward) and DNA38 (reverse).

Piece 3:

---3'intron overhang---AG][GTA---*avrBs2 second part*---][HA][XbaI][NNNN]

Template: *avrBs2-HA*

Primers: The forward primer (TLG43) anneals to *avrBs2* at the insertion site and adds an overhang. The *avrBs2* reverse primer (TLG46) adds an XbaI site.

Gel extract each product from an agarose gel. Calculate the molar concentration of the PCR products.

*Step 3: Perform 3-piece-ligation PCR with a high fidelity polymerase.*

Templates: The three PCR products in a 1:1:1 molar ratio. For *avrBs2-HyP5SM-HA (E/V)* cloning, the final concentration of each template was approximately 0.15 nM.

Primers: The *avrBs2-HA* forward and reverse cloning primers (TLG45, TLG46). The overhangs in the templates also act as primers.

*Step 4:* The desired full-length product was gel purified, inserted into a cloning vector (Invitrogen pCR2.1 TOPO), and sequenced to confirm that the junctions were correct and that there were no frame shifts. The sequence perfect product was then cloned into a binary vector for plant expression.

See **Figure 2.14** for an illustrated diagram and the 3-piece-ligation PCR products that generated *avrBs2-HyP5SM-HA* (E/V).

Note: For large genes (>3 kb) that are difficult to PCR, Gibson assembly may be a useful alternative cloning method.

# CHAPTER 3:

# Generating Viable Transgenic Plants with Inducible Hypersensitive Response

*Portions of this work were published in the following scientific journal:

Gonzalez, T. L.; Liang Y.; Nguyen B. N.; Staskawicz B. J.; Loqué, D.; Hammond, M. C. (2015) *Nucleic Acids Research*, **43**, 7152-61.

## Abstract

In **Chapter 2**, the addition of the HyP5SM suicide exon was shown to more stringently regulate effector-triggered hypersensitive responses in transiently transformed *Nicotiana* plants. Without HyP5SM, the Dex-inducible promoter resulted in too much leaky background expression of AvrBs2 and ATR1Δ51 effector proteins. The resulting hypersensitive response phenotype was undesirably constitutive. With HyP5SM, inducible regulation was recovered. We predicted that HyP5SM would allow us to generate transgenic plants with inducible hypersensitive response. In **Chapter 3**, we show the generation of viable and healthy transgenic *Arabidopsis thaliana* plants that inducibly initiate RPP1/ATR1Δ51-triggered HR. Beyond enabling studies on the ETI pathway, our regulatory strategy is generally applicable to reduce or eliminate undesired background expression of transgenes.

## 3.1: Contemporary Methods to Make Transgenic Plants With Inducible HR

As demonstrated in **Chapter 2**, combining the dexamethasone (Dex) inducible promoter and the HyP5SM "suicide" cassette can stringently control the expression of pathogen effectors *avrBs2* and *ATR1Δ51* such that no visible hypersensitive response (HR) appears in leaves transiently co-expressing the corresponding resistance gene. Moreover, effector protein expression and the HR phenotype can be inducibly recovered. Without HyP5SM, leaky effector protein activates the resistance gene and plant tissue undergoes programmed cell death, a circumstance that has prevented the generation of disease resistant plants transgenic for various effectors. These plants are

desirable for experiments that cannot be completed with transiently transformed tissue: for example, forward genetic screens to find important components of HR pathways, analysis of early-stage proteomic/transcriptional changes that occur when a specific effector is expressed (as opposed to whole pathogen inoculation, which triggers many more pathways), and analysis of these same proteomic/transcriptional changes under different environmental conditions or at different developmental stages.

A highly desired quality of an R/effector-transgenic plant is the ability to inducible immune responses without the need of pathogen infection assays, *Agrobacterium*-mediated transient transformations, viral vector-mediated transient transformations, or other gene-delivery methods that activate PAMP-triggered immunity. *Ve1/Ave1*-transgenic *Arabidopsis thaliana* is an example of a viable R/effector-transgenic plant that cannot initiate immune responses. *Ve1* is a tomato gene which recognizes *Verticillium dahliae* fungal effector *Ave1*, triggers an HR phenotype in tomato and *N. tabacum*, and confers resistance against *Verticillium* wilt (165). *Ve1* is unable to trigger a hypersensitive response in *A. thaliana* plants even though other *Ve1*-mediated disease resistance responses to *V. dahliae* do occur (166). However, the *Ve1/Ave1*-expressing *A. thaliana* plants were not able to mount a defense against a *ΔAve1* mutant of *V. dahlia*, suggesting that the plants require a pathogen-provided version of the Ave1 effector. This, then, suggests that the *Ave1* transgene in the plant is not activating the immune system and may explain why the *Ve1/Ave1 A. thaliana* plants lack an autoimmune hypersensitive response. Ve1 resistance protein is membrane-bound, contains extracellular leucine-rich repeats (eLRRs) and has no clear motifs for intracellular signaling (166). It is possible that the lack of HR may be because Ve1/Ave1 recognition occurs outside the cell. Although this plant line is transgenic for an R/effector gene pair, it still requires inoculation with a pathogen to initiate immune responses, and thus it is not a candidate to study transcriptional changes from a single immune trigger. In our literature search, we were careful to note that not all transgenic plants with inducible effector proteins were able to mount an inducible HR phenotype.

To our knowledge, the dexamethasone (Dex)-inducible plasmids developed by Aoyama and Chua, pTA7001 and pTA7002 (123), are the major method that has been able to generate continuing transgenic lines expressing both R/effector gene pairs required to induce HR. McNellis et al. (1998) were able to generate transgenic *A. thaliana* plants to study the RPS2/AvrRpt2-dependent ETI pathway by regulating the effector AvrRpt2 with a Dex-inducible promoter (123, 167). The *RPM1* resistance gene has also successfully tolerated inducible regulation of its recognized effectors, including Dex-inducible *avrB* or *avrRpm1*, or estradiol-inducible *avrRpm1* (137, 168). *A. thaliana* expressing both the *RPS5* resistance gene and Dex-inducible *avrPphB* is also viable (154). Unfortunately, promoter regulation is not enough to generate transgenic plants for all R/effector pairs. In our experience, healthy and viable *N. benthamiana* plants expressing Bs2/AvrBs2 or *A. thaliana* plants expressing RPP1/ATR1 have not been able to be generated. Anecdotally, additional examples have been encountered by other researchers, but these negative results are difficult to publish. Beyond applications in immunology research, leaky protein expression from inducible promoters is a common limitation in transgenic plant generation (169).

Another strategy to germinate transgenic plants when R/effector co-expression otherwise leads to lethality is to take advantage of the HR phenotype's temperature sensitivity. In tomato, this is accomplished by crossing resistant plants harboring only the R gene with susceptible plants harboring only the effector gene, and then germinating the F1 seeds at elevated temperatures (32-33°C) to suppress HR (139, 170). Lowering the temperature then induces HR, but this method is not suitable for all R/effector pairs (139, 170, 171), sometimes alters rather than suppresses the HR phenotype (171), and may result in pleiotropic effects such as changes in hormone levels (172). Also, it has only been used for very young plants and cannot generate continuing transgenic lines (139, 170). Without transgenic plants viable to adulthood, the only way to design a mutant screen would be to continually cross the two parent lines, mutagenize the dormant seeds, then exploiting the temperature sensitivity of HR. However, this is not practical due to the low yield and labor involved in plant crossings; an effective selection screen requires an abundance of seeds.

We propose that employing the HyP5SM cassette in combination with a conditional promoter is an improved method for making R/effector-transgenic plants. In **Chapter 2**, this dual regulation strategy successfully controlled HR in transiently transformed leaf tissue by reducing background effector protein levels. However, when plants were co-expressed with Dex-inducible *OsL5-6xHis* and dual regulated effector, low-level leaky protein expression was still observed by western blot (**Figure 2.10c, 2.11d**). One concern for generating stable transgenic plants is that this low-level leakiness, although much less than leaky expression from the Dex-inducible promoter alone, could still cause lethality if the effector protein is able to accumulate to HR-triggering levels. In this chapter, we demonstrate that we can generate viable and healthy transgenic *A. thaliana* plants that can be chemically induced to initiate the hypersensitive response. Thus, HyP5SM regulation is shown to be a reliable plant biotechnological tool to regulate difficult-to-control genes in even stable transgenic plants.

## 3.2: Results

### Generation and testing of multi-gene constructs

In order to demonstrate that our system enables generation of viable transgenic plants, we constructed a multi-gene pTKan vector containing Dex-inducible *OsL5-6xHis*, Dex-inducible *ATR1Δ51-HyP5SM-FLAG*, *35S::GVG* (the Dex-responsive transcription factor from pTA7001), and *pNOS::DsRed2* reporter gene in the T-DNA insertion cassette (C59, **Figure 3.1**). We elected to focus on the RPP1/ATR1-dependent HR pathway because the *RPP1* gene comes from an *A. thaliana* plant resistant to downy mildew, and so a transgenic plant will allow us to study a biologically-relevant pathway. Transient transformation experiments with the C59 pTKan vector in *N. tabacum* showed that it can promoted HR after Dex induction (**Figure 3.2a**), although the phenotype was weaker than the HR seen from co-infiltrating pTA7001 *ATR1Δ51-HyP5SM-FLAG*, pTA7001 *OsL5*, and pEG301 *RPP1-3xHA* vectors (**Figure 2.12d**). The multi-gene pTKan vector contains only one copy of the GVG transcription factor, whereas previous experiments included multiple copies of GVG due to inclusion of separate pTA7001

vectors. Thus, the weaker HR phenotype may be due to titration of the GVG transcription factor or insufficient induction of *Os*L5 protein. Moreover, due to the large T-DNA size of the C59 pTKan vector, full-length T-DNA transfer efficiency might be lower than for the three co-infiltrated vectors. We tested these hypotheses by co-infiltrating pTKan C59 with either pTA7001 empty vector or pTA7001 *OsL5* since GVG transcription factor and *Os*L5 are located in the beginning and middle of the T-DNA of C59 respectively. Adding extra copies of either GVG transcription factor or *OsL5* promoted stronger HR, suggesting that they are limiting factors (**Figure 3.2b**).

**HyP5SM enables generation of stable *Arabidopsis* plant lines that inducibly activate the hypersensitive response**

To transform plants with C59, we performed floral dips of *A. thaliana* line 3860 containing *RPP1-3xHA* (WsB allele) acquired from B. J. Staskawicz (173) and of *A. thaliana* Col-0 ecotype, which does not respond to the Emoy2 allele of *ATR1* used here because it does not harbor the *RPP1* WsB allele. Transgenic seedlings were grown on selection plates and transplanted to soil after 10 days. Plants looked healthy and displayed no obvious growth defects compared to wild-type (**Figures 3.3, 3.4**). Furthermore, Dex induction of three independent second-generation lines resulted in an RPP1/ATR1-dependent hypersensitive response characterized by initial chlorosis, browning, tissue collapse, and eventual drying of induced leaves (**Figures 3.3, 3.4**). Injection of individual leaves with 15 µM Dex or spraying the whole plant with 30 µM Dex both resulted in HR (**Figures 3.3, 3.4**). Leaf tissue reproducibly underwent induced HR. We did not observe HR in floral tissue and shoot apical meristem tissue in any of the independent transgenic lines (not shown).

In order to facilitate research on bacterial spot disease resistance in tomato and pepper plants, we designed similar multi-gene pTKan vectors for the Bs2/AvrBs2 pair. As before, all vectors contain Dex-inducible *OsL5-6xHis*, *pNOS::DsRed2*, and *35S::GVG* (**Figure 3.1**). Two vectors were constructed with the addition of *35S::Bs2-FLAG* resistance gene, pTKan C92 contains Dex-inducible *avrBs2-HA* and pTKan C93 contains Dex-inducible *avrBs2-HyP5SM-HA* (**Figure 3.1**). In experiments paralleling Figure 4a and S4, the C92 vector lacking HyP5SM-regulation showed constitutive HR phenotype in *N. benthamiana* despite absence of Dex, whereas dual regulated *avrBs2-HyP5SM-HA* showed both clean background and Dex-induced development of HR (**Figure 3.5**). Unlike with pTKan C59, the HR phenotype from the multi-gene plasmid did not appear to be weaker than seen previously. The induced HR from C92 and C93 appeared to be of similar strength as well, but no quantification was attempted. Two control vectors were constructed: pTKan C226 lacks *Bs2*, and pTKan C228 contains *EGFP-HyP5SM* in place of *avrBs2* (**Figure 3.1**). As expected, neither vector promotes HR on its own, but complementation by co-expressing either *avrBs2* or *Bs2* as needed results in visible Bs2/AvrBs2-dependent HR (**Figure 3.5**). Due to time constraints, transgenic *N. benthamiana* was not attempted; however, these multi-gene vectors may be used in the future to make transgenic *Solanaceae* plants to study agriculturally relevant disease resistance.

Meanwhile, we were able to transform *A. thaliana* Col-0 with C93, C226, and C228. Like the RPP1/ATR1 plants, these plants were viable and healthy as well. Dex

induction resulted in a Bs2/AvrBs2-dependent phenotype that was distinct from HR, observed as gradual darkening of the leaves to a purple color (**Figure 3.6**). The *X. euvesicatoria* pathogen (source of *avrBs2*) does not infect *A. thaliana* and the *Bs2* gene is not known to function in *A. thaliana*, so we are unclear if this is a true or biologically relevant immune response. Nevertheless, these results together demonstrate that the dual regulation system works well in stable transgenic lines.

## 3.3: Discussion

Impressively, our two-input system enables the generation of viable, healthy, transgenic plants harboring both effector and R genes. We have made independent stable transgenic *A. thaliana* lines carrying dual regulated *ATR1Δ51-HyP5SM* in the presence of *RPP1* and Dex-inducible *OsL5*. We have shown that the HR phenotype is suppressed by default, yet can be chemically induced (**Figures 3.3, 3.4**). These transgenic plants will allow us to study *RPP1* disease resistance under different environmental conditions, at different stages of plant development, in different tissues, and at short time scales. Significantly, these plants are viable during their whole lifetime and produce fertile progeny. Continuing transgenic lines across multiple generations makes possible a forward genetics screen. Mutagenized homozygous seedlings may be Dex-induced and screened for absence of the expected HR phenotype, revealing potential dominant mutations disrupting the hypersensitive response pathway. Although the pTKan C59 vector resulted in a weaker HR when transiently expressed in *N. tabacum* plants, the HR phenotype was reliably reproducible in stable transgenic *A. thaliana* plant leaves sprayed with Dex.

We hypothesize that this tissue-specific absence of HR in floral or shoot apical meristem tissue could be due to differential expression of a promoter, perhaps the Dex-inducible promoter or the native *RPP1* promoter which drives *RPP1* in our transgenic plants. No expression profile was found for *RPP1* in literature or available databases, and articles with Dex-inducible plants have not focused on floral or shoot apical meristem tissue. The *35S* promoter driving Dex-responsive GVG may also be responsible; although *35S* is considered a constitutive promoter, it is reported to have variable strength in different tissue across species, and it has been observed previously that transgenic *A. thaliana* lines expressing 35S-driven genes in leaves may not always express them in flower tissue (174, 175). Because HyP5SM/*Os*L5 splicing regulation is engineered from sequences important for ribosome formation, they are unlikely to be the issue (112, 120). Furthermore, a key benefit of the HyP5SM regulation system is that the promoters may be exchanged for projects where the *35S* or Dex-inducible promoters are undesired.

In closing, the hypersensitive response of effector-triggered immunity is an excellent phenotype to demonstrate the strong down-regulation of transgene basal expression that we can achieve with the HyP5SM cassette. The minimal sequence requirements and broad applicability of the HyP5SM cassette to any gene means that transgenic plant research will no longer be limited to studying the signaling pathways of R genes that can tolerate leaky effector protein expression. Anecdotally, leaky expression is a problem not only for genes that trigger immune responses, but also for

genes that trigger other amplified signal transduction pathways such as hormone signaling pathways, and for cell-wall disrupting enzymes important for sugar extraction and biofuels research. The HyP5SM cassette represents a new biotechnological tool that can be used in conjunction with various promoters, making a versatile gene regulation system.

## 3.4: Figures

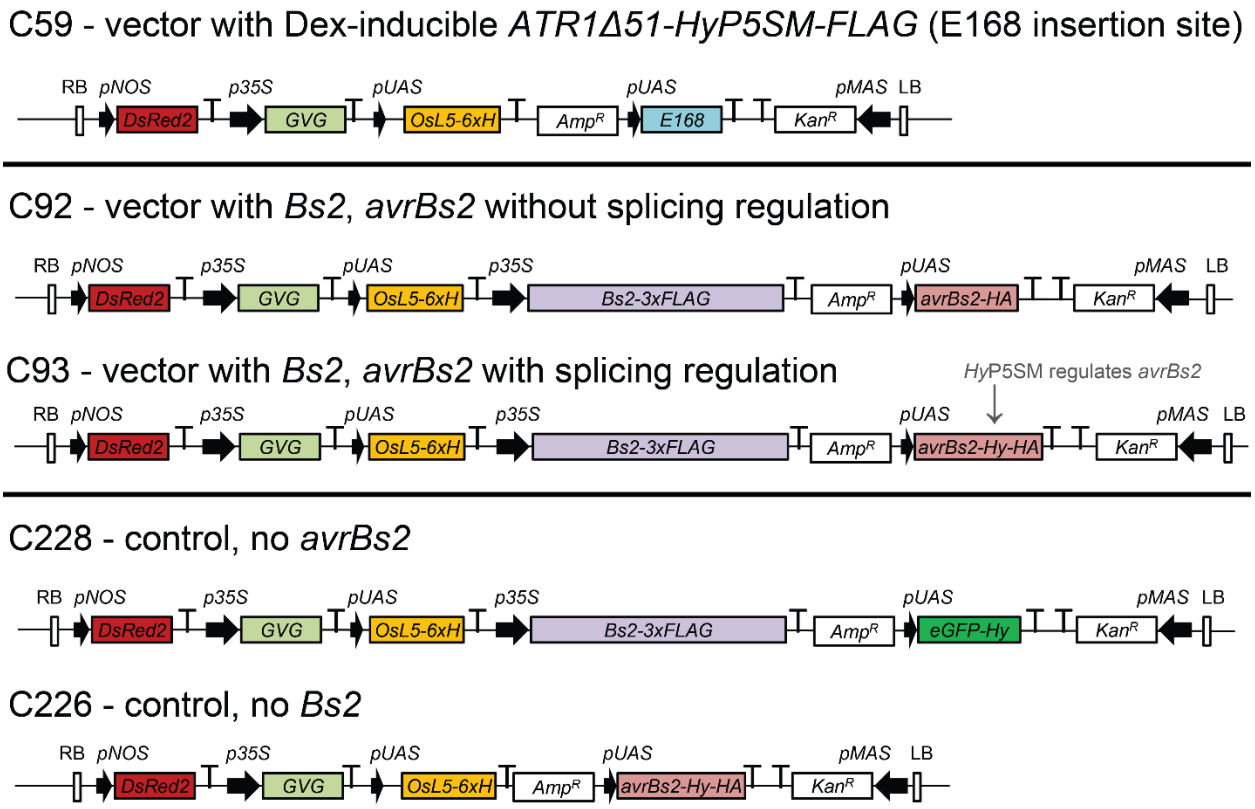


**FIGURE 3.1. The multi-gene pTKan constructs used for transgenic *Arabidopsis*.** See supplemental methods for cloning description. Abbreviations include pNOS = nopaline synthase promoter, p35S = Cauliflower Mosaic Virus 35S promoter, pUAS = p6xUAS promoter (induced by GVG transcription factor in presence of dexamethasone), 6xH = 6xHis tag, E168 = *ATR1Δ51-HyP5SM-FLAG* (with the splicing cassette inserted after codon E168), pMAS = manopine synthase promoter, Hy = HyP5SM cassette.

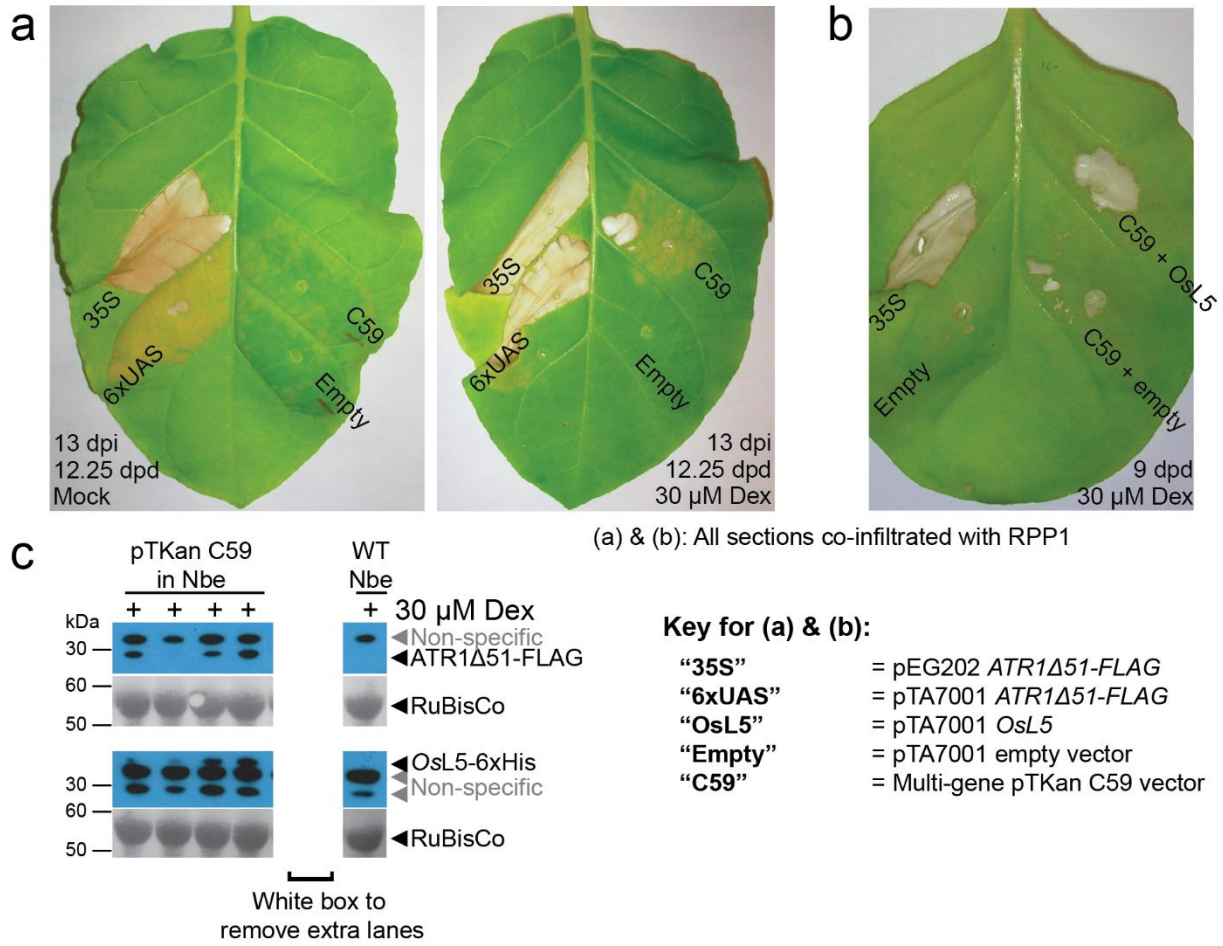


**FIGURE 3.2. Transient expression of the multi-gene plasmid pTKan C59 shows hypersensitive response and protein.** (a) Infiltrations to test "C59", the multi-gene construct pTKan C59 (which lacks RPP1, but contains Dex-inducible *ATR1Δ51-HyP5SM-FLAG* and *OsL5-6xHis*). All sections were co-infiltrated with pEG301 *RPP1-3xHA* in a 1:1 mix. The mock-induced leaf shows no hypersensitive response from pTKan C59. With Dex, the hypersensitive response is induced. (b) To investigate why the total hypersensitive response seems to be weaker from pTKan C59, *N. tabacum* was co-infiltrated with 1:1:1 mixes of RPP1, the indicated ATR1 or empty construct, and either pTA7001 *OsL5* or additional pTA7001 empty vector (which, although empty after the Dex-inducible promoter, still contains *35S::GVG* transcription factor). (c) *N. benthamiana* was transiently transformed with pTKan C59 and Dex-induced to promote protein expression. Western blots show expression of ATR1Δ51-FLAG (top: anti-FLAG) and *Os*L5-6xHis protein (bottom: anti-*Os*L5-6xHis) from four biological replicates. Below each western blot is a Ponceau S stain of RuBisCo.

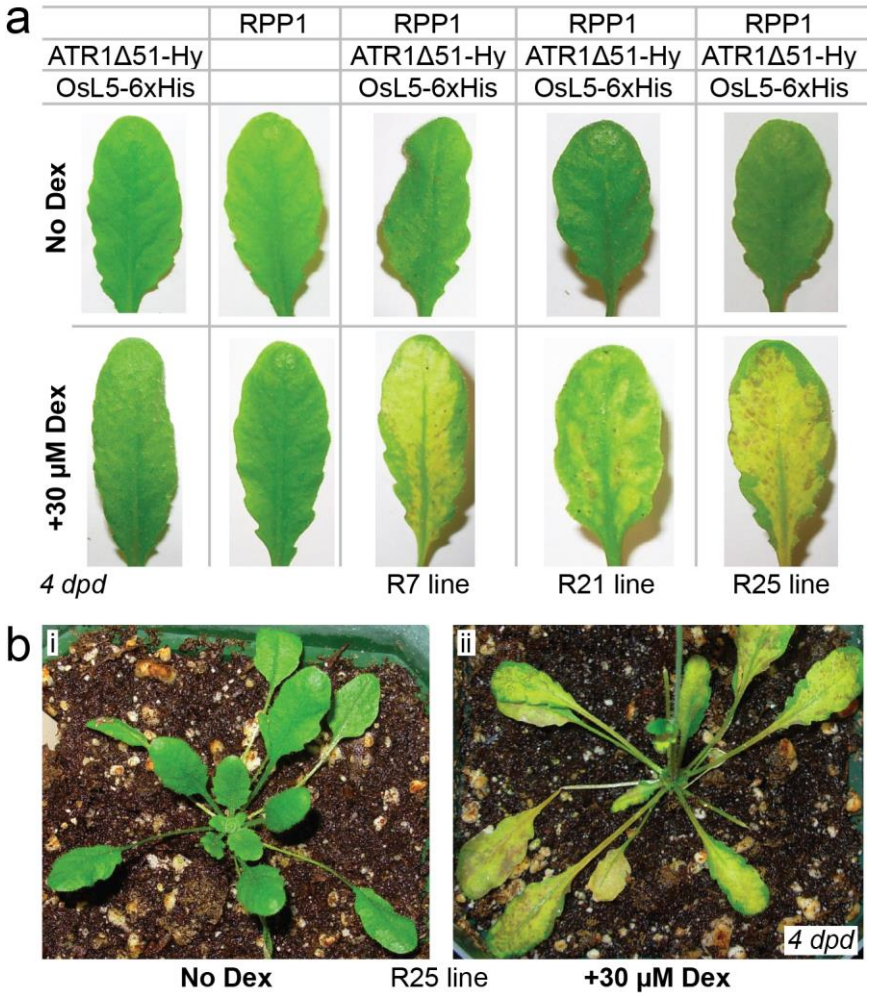


**FIGURE 3.3. Transgenic *Arabidopsis thaliana* plant lines that contain both the resistance gene and the dual regulated effector gene are viable, and the hypersensitive response is chemically inducible.** (a) Representative leaves from plant lines carrying the effector (ATR1□51-HyP5SM) and *Os*L5, the R gene (RPP1), or all three genes, without and with Dex induction. (b) Rosettes of line 25 plants without and with Dex induction.

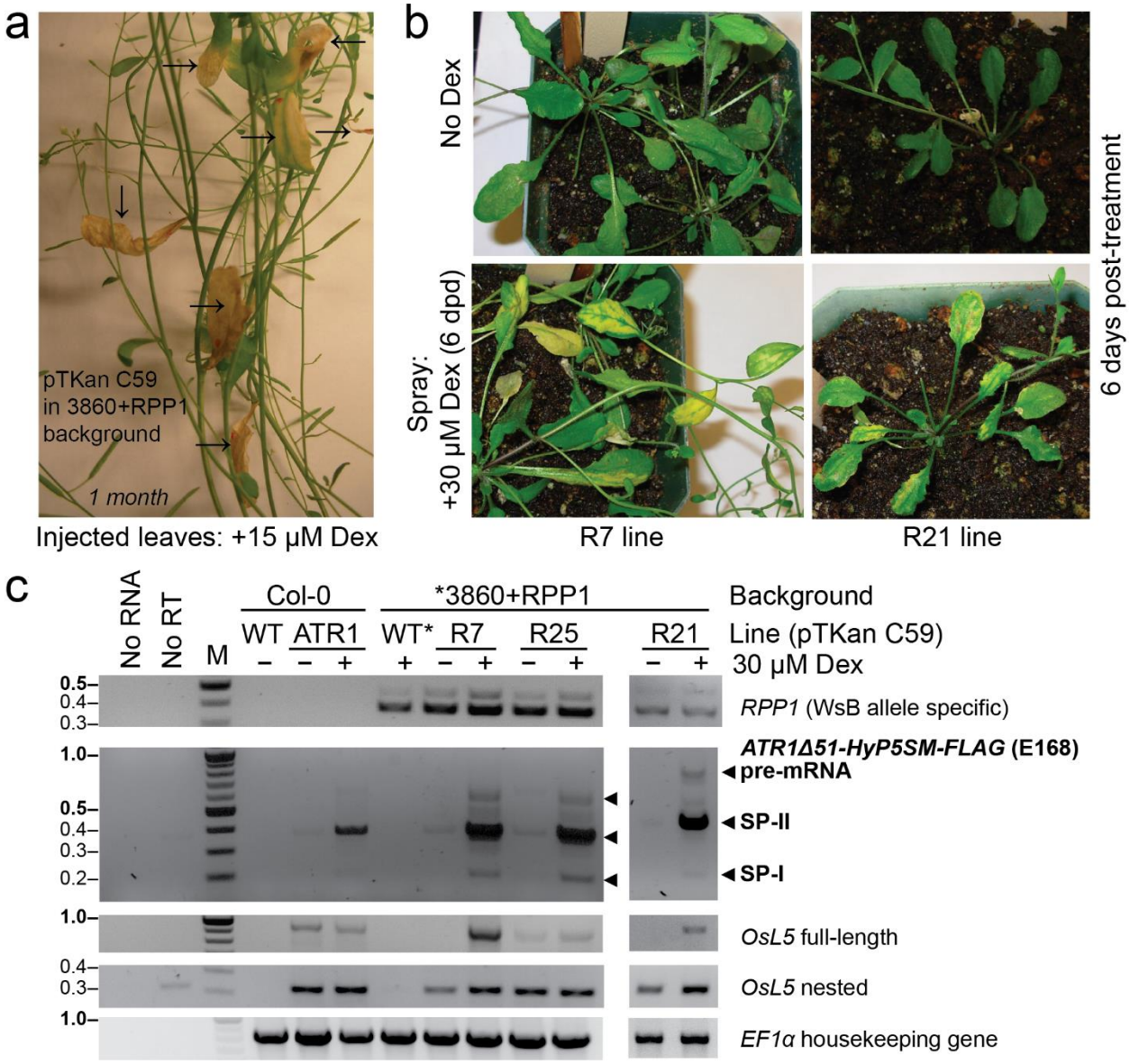


**FIGURE 3.4. Induced HR in transgenic *A. thaliana* plants expressing RPP1 and Dex-inducible ATR1Δ51 protein.** Plants are as described in Figure 6. (a) Injecting individual leaves of a bolting and early flowering plant with 15 µM Dex (see arrows) and growing in humid conditions on long days resulted in accelerated senescence of the induced leaves relative to the rest of the plant, but slower cell death relative to transient transformations (weeks instead of days). (b) Spraying late rosette-stage plants with 30 µM Dex and moving the plants to open growth carts on constant light resulted in faster cell death. These additional images of lines R7 (multiple rosettes in one pot) and R21 (single rosettes) accompany Figure 6b. (c) RT-PCR analysis of different plant lines. Col-0 WT contains no transgenes, Col-0 ATR1 contains pTKan C59 but no RPP1, WT* is 3860+RPP1 background line for R7, R25, and R21, which contain pTKan C59. The higher band observed for *RPP1* is an intron-retaining spliced product. Pre-mRNA, SP-I, and SP-II are observed for *ATR-HyP5SM-FLAG*. A nested PCR was performed on the PCR product from full-length *OsL5*, resulting in stronger signal.

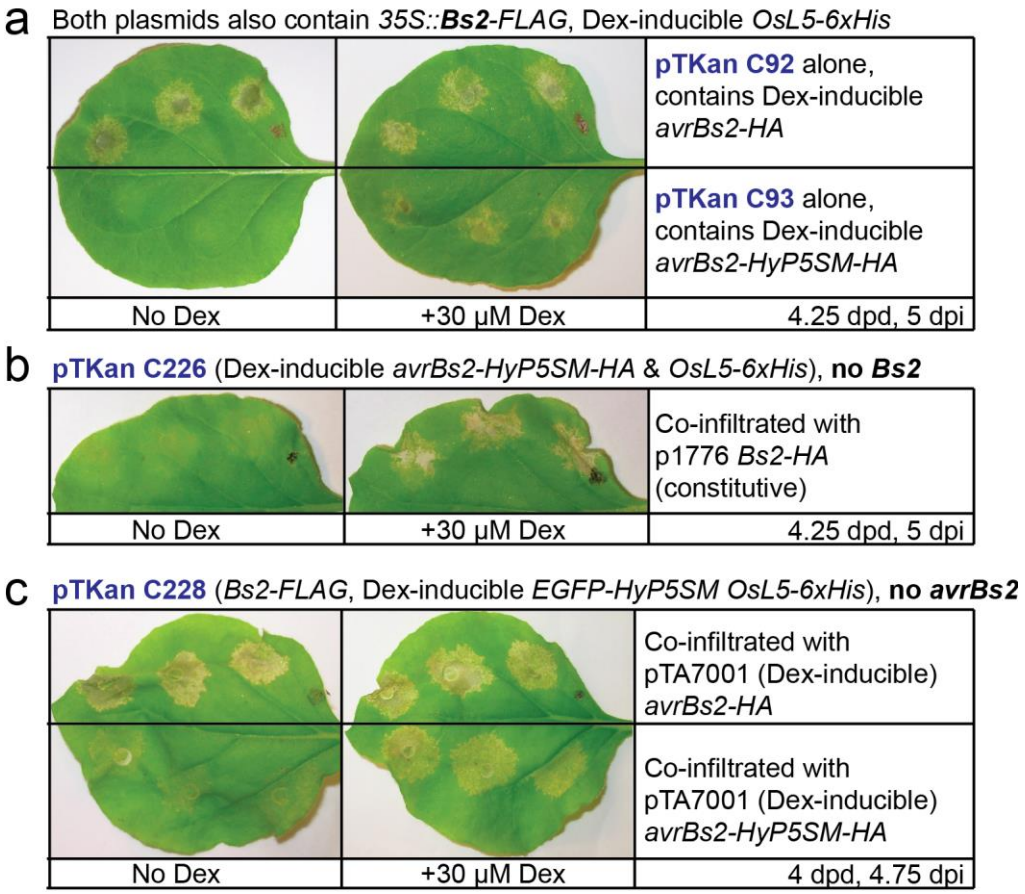


**FIGURE 3.5. Functional testing of multi-gene pTKan *Bs2/avrBs2* plasmids.** The multi-gene pTKan plasmids were tested for their ability to promote Bs2/avrBs2-dependent HR using either Dex-induction and/or complementation. (a) Plasmids pTKan C92 and C93 contain genes for some version of *Bs2* and *avrBs2*. Only pTKan C93 (with splicing-regulated *avrBs2-HyP5SM-HA*) shows OFF/ON regulation of the hypersensitive response. (b) pTKan C226 is like pTKan C93, but without *Bs2*. Complementation shows that its copy of *avrBs2* is functional. (c) Similarly, pTKan C228 is like pTKan C93, except that it contains *EGFP-HyP5SM* in place of *avrBs2-HyP5SM-HA*. Complementation shows that its copy of *Bs2* is functional. The pTKan C93, C226, and C228 plasmids were used to make transgenic *A. thaliana* plants. Note: All pTKan plasmids here also contain *pNOS::DsRed2* and *35S::GVG* transcription factor gene (for Dex-induction of the *6xUAS* promoter).

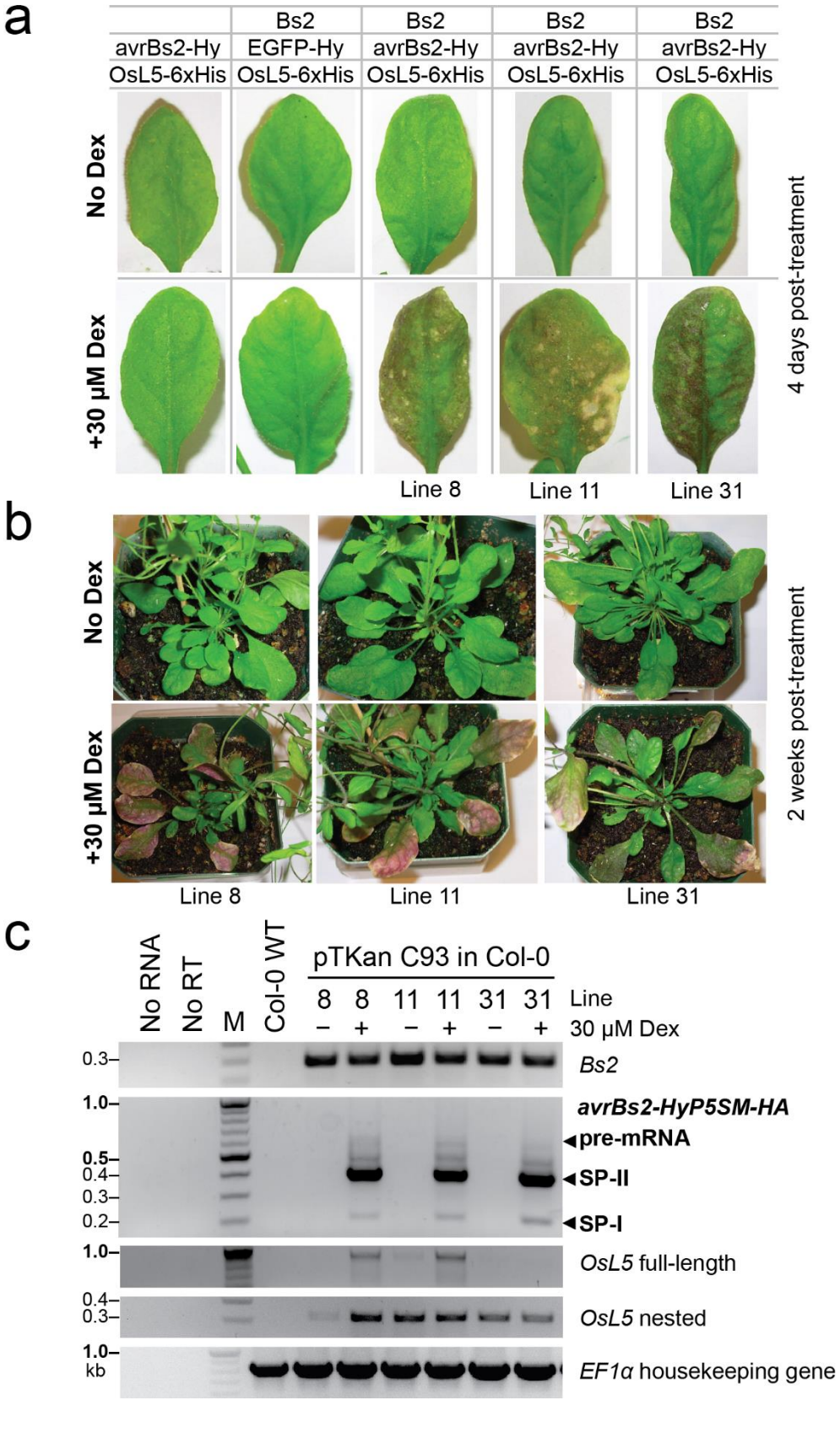
a
Bs2
avrBs2-Hy
OsL5-6xHis
Bs2
EGFP-Hy
OsL5-6xHis
Bs2
avrBs2-Hy
OsL5-6xHis
Bs2
avrBs2-Hy
OsL5-6xHis
No Dex
+30 µM Dex
4 days post-treatment
Line 8
Line 11
Line 31
b
No Dex
+30 µM Dex
2 weeks post-treatment
Line 8
Line 11
Line 31
c
No RNA
No RT
M
Col-0 WT
pTKan C93 in Col-0
8 8 11 11 31 31
Line
− + − + − +
30 µM Dex
0.3
Bs2
1.0
avrBs2-HyP5SM-HA
pre-mRNA
0.5
0.4
SP-II
0.3
0.2
SP-I
1.0
OsL5 full-length
0.4
0.3
OsL5 nested
1.0
kb
EF1α housekeeping gene

**FIGURE 3.6. Transgenic *A. thaliana* plants show a Bs2/AvrBs2-dependent phenotype.** (a) Transgenic plants expressing some combination of *6xUAS::avrBs2-HyP5SM-HA* or *6xUAS::EGFP-HyP5SM* (negative control), *6xUAS::OsL5-6xHis,* and *p35S::Bs2-3xFLAG*. Three second generation independent lines are shown for the plants containing both the *avrBs2* and the *Bs2* genes (made with pTKan C93 in Col-0). Dex induction stimulates Bs2/AvrBs2-dependent leaf darkening to a purple color. (b) Rosette images showing the progression of the phenotype, 2 weeks after treatment. (c) RT-PCR analysis of different plant lines. Col-0 WT contains no transgenes and is the background line for R8, R11, and R31, which contain pTKan C93 (Bs2, AvrBs2-HyP5SM-HA, OsL5). RNA was extracted from leaf tissue collected from the same plants before and 1 day after 30 μM Dex spray. Pre-mRNA, SP-I, and SP-II are observed for *AvrBs2-HyP5SM-HA*. A nested PCR was performed on the PCR product from full-length *OsL5*, resulting in stronger signal.

## 3.5: Materials and Methods

### Cloning the Multi-Gene pTKan Plasmids for Faster Generation of Transgenic *Arabidopsis thaliana* Plants

*Step 1: Construction of the pTKan expression vector*.

The p35S promoter was amplified by PCR from pRT100 (176) and cloned into pCRblunt (Invitrogen) to generate *pBlunt-p35S*. *pTKan-p35S-GW* binary vector was generated by digesting *p35S* promoter from *pBlunt-p35S* with XhoI/AvrII, and then inserted between XhoI/AvrII restriction sites of *pTKan-GW-R1R2* vector (177). The *pNOS-DsRed2-tNOS* sequence was synthesized and fused into ApaI site of *pTKan-GW-R1R2* vector by in-fusion cloning (In-Fusion® HD, Clontech) to generate *pTKan-pNOS-DsRed2-tNOS-p35S-GW-R1R2* which was subsequently digested with AvrII restriction enzyme to fuse with *GVG-tRbcsE9* and *p6xUAS* fragments. The in-fusion product is the expression vector *pTKan-pNOS-DsRed2-tNOS-p35S-GVG-tRbcsE9-p6xUAS-GW-R1R2*.

*Step 2: Construction of the entry vectors*

To generate Gateway entry clones, attB or attBr-flanked PCR products were cloned into pDONR 221 vectors with corresponding attP or attPr sites (MultiSite Gateway®Pro, Invitrogen) by BP recombination reaction. In this way, *OsL5-6xHis* was cloned into pDONR221 P1-P4 vector; *eGFP-HyP5SM, ATR1Δ51-HyP5SM-FLAG (E168 insertion site)*, and *avrBs2-HyP5SM-HA* were cloned into pDONR221 P3-P2 vector, respectively; *tG7-p6xUAS* was cloned into pDONR221 P4r-P3r vector. pDORN221-*tG7-AmpR-pUAS* R4-R3 and pDONR221-*tG7- p35S-Bs2-3xFlag-tNOS-AmpR-pUAS* R4-R3 constructs were generated by in-fusion cloning (In-Fusion® HD, Clontech). Briefly, AmpR alone or *AmpR* together with *p35S-Bs2-3xFlag-tNOS* was inserted in between *tG7* and *p6xUAS* at HindIII site of the pDONR221 *tG7-p6xUAS* R4-R3 vector in single in-fusion cloning reactions. The PCR template of *p35S-Bs2-3xFlag-tNOS* was provided by pMD1 Bs2-3xFLAG from the Staskawicz lab. The insertion of AmpR helps to stabilize the existence of two *p6xUAS* promoters in the final expression constructs.

*Step 3: Construction of the multi-gene plasmids*

In a MultiSite Gateway reaction (MultiSite Gateway®Pro, Invitrogen), genes or fragments of interest in pDONR221 P1-P4 vector, pDONR221 P4r-P3r vector and pDONR221 P3-P2 vector were connected and incorporated into GW R1-R2 expression vector in a sequential manner. MultiSite Gateway reaction was carried out with the expression vector *pTKan-pNOS-DsRed2-tNOS-p35S-GVG-tRbcsE9-p6xUAS-GW-R1R2* and various entry vectors from Step 1 and Step 2. Entry vector combinations for the final expression constructs were listed in the following table. LR Clonase™ II Plus enzyme mix and equal molar of each component construct were used in the MultiSite Gateway reaction according to the manufacturer's manual.

The pTKan vector selectable markers for *E. coli* are spectinomycin resistance (SpecR), and if applicable, ampicillin resistance (AmpR). In our hands, we found that *Agrobacterium* could utilize the pMAS-driven KanR gene for plant selection, but SpecR may be used for *Agrobacterium* selection as well.

**TABLE 3.5.1:** *Pieces used to generate the multi-gene pTKan vectors.*

| Final construct | DNA element 1 | DNA element 2 | DNA element 3 |
|---|---|---|---|
| **C59** | pDONR221-OsL5-6xHis L1L4 | pDONR221-tG7-AmpR-pUAS R4R3 | pDONR221-ATR1Δ51-HyP5SM-FLAG L3L2 |
| **C92** | pDONR221-OsL5-6xHis L1L4 | pDONR221-tG7-p35S-Bs2-3xFlag-tNOS-AmpR-pUAS R4R3 | pDONR221-avrBs2-HA L3L2 |
| **C93** | pDONR221-OsL5-6xHis L1L4 | pDONR221-tG7-p35S-Bs2-3xFlag-tNOS-AmpR-pUAS R4R3 | pDONR221-avrBs2-HyP5SM-HA L3L2 |
| **C226** | pDONR221-OsL5-6xHis L1L4 | pDONR221-tG7-AmpR-pUAS R4R3 | pDONR221-avrBs2-HyP5SM-HA L3L2 |
| **C228** | pDONR221-OsL5-6xHis L1L4 | pDONR221-tG7-p35S-Bs2-3xFlag-tNOS-AmpR-pUAS R4R3 | pDONR221-eGFP-HyP5SM L3L2 |

## Generation of Transgenic *Arabidopsis thaliana* Plants

To facilitate the introduction of multiple genes into *A. thaliana*, pTKan binary vectors were designed (**Table A2**, **Figure 3.1**) and constructed as described in Supplemental Methods. *A. thaliana* Col-0 ecotype plants were grown in a growth chamber on long day conditions (16/8 h day/night cycles) and transformed using the *Agrobacterium*-mediated floral dip method (178). Briefly, *Agrobacterium tumefaciens* strain GV3101 containing the binary pTKan vectors were cultured in LB Lennox (low salt) media for 16-24 h at 28°C, 225 rpm, with rifampicin (50 µg/mL), gentamycin (50 µg/mL), and kanamycin (50 µg/mL) selection. Cells were pelleted at 4700 rpm for 15 minutes and resuspended in floral dip buffer (5% sucrose, 10 mM $MgCl_2$, 10 mM MES buffer pH 5.6, 0.018% Silwet L-77). The flowering stems of young *A. thaliana* plants

were dipped in the *Agrobacterium* solution for 30-60 seconds. Plants were left to recover in the dark for 16-24 h, grown on long days to maturity, then dried and the seeds were collected by sifting.

To select for transgenic plants, the seeds of floral-dipped plants were ethanol-sterilized and plated on phytoagar media in a sterile biosafety cabinet. Phytoagar media was prepared as follows: 2.16 g/L MS, 20 g/L sucrose, 0.35 g/L MES, water added, pH to 5.7-5.8, and then 7.5 g/L phytoagar added. Media was autoclaved for 30 minutes, filter-sterilized kanamycin (50 µg/mL) and cefotaxime (200 µg/mL) were added once media had cooled, then plates were poured and allowed to set inside the sterile cabinet. Phytoagar plates with seeds were sealed with breathable microfilm, wrapped in foil, and stored at 4°C for 2-4 days. Plates were unwrapped and transferred to a growth chamber for germination under long days. Transgenic plants are viable with kanamycin treatment and were transferred to soil at 10 days, then rosette leaves were scanned for expression of the DsRed2 reporter gene using a Typhoon laser image scanner (ex/em: 532/580 nm). DsRed2 fluorescence was used to estimate expression of the other transgenes, thus reducing the initial labor required to analyze and compare independent lines. Lines expressing higher levels of DsRed2 fluorescence were carried forward. Subsequently, expression of other transgenes was verified by RT-PCR.

*A. thaliana* 3860 ecotype plants homozygous for *pRPP1::RPP1-WsB-3xHA::tRPP1* (Basta resistant, kanamycin sensitive) were acquired from the Staskawicz group (37), and transformed with pTKan C59 (**Figure 3.1**) by floral dip. The resulting transgenic plants are Basta and Kan resistant, and carry copies of the R gene *RPP1*, the effector gene *ATR⊿51-HyP5SM-Flag* (Emoy2 allele, inducible), *OsL5-6xHis* (inducible), the transcription factor *GVG*, and *DsRed2*. As a control, homozygous Col-0, which does not recognize the Emoy2 allele of *ATR1*, was transformed with the pTKan C59 vector. Data shown for plants carry both effector and R genes are for T2 lines (either heterozygous or homozygous). Control plants carrying just the effector gene or the R gene are homozygous lines.

**Dex-Induction of Transgenic Plants to Induce HR**

Transgenic *Arabidopsis thaliana* plants were induced as in **Chapter 2** by needle-less syringe with 15 µM dexamethasone or by spraying with 30 µM dexamethasone.

**RT-PCR**

Generally as in **Chapter 2**. The only different in tissue collection is that whole rosette leaves were collected from transgenic *A. thaliana* plants instead of leaf discs as for *N. benthamiana*. The cDNA was synthesized with Super Script III Reverse Transcriptase (Thermo Fisher Scientific) and oligo(dT) primers. Primers are described in **Table A1**.

**Protein Extraction, Western Blots, and Ponceau S Stains**

Same as in **Chapter 2**.

# CHAPTER 4:

# Understanding the Mechanism of P5SM Regulation


## Abstract

In **Chapter 2**, Dex-induced expression from pTA7001 *avrBs2-HyP5SM-HA* in the absence of *OsL5* co-expression showed increasing amounts of AvrBs2-HA protein. This result provided further evidence for a possible exon defining factor for P5SM cassettes. In this chapter, we present progress toward identifying this putative splice factor. The broader impacts of this goal are two-fold. First, identifying what causes P5SM to be recognized as an exon will allow us to further engineer P5SM-based biotechnological tools in plant and possibly non-plant systems. Second, native P5SM is a conserved structured RNA involved in the production of 5S ribosomal RNA, and thus better understanding P5SM regulation will give us insight into ribosome biosynthesis regulation in plants.


## 4.1: Introduction

### The Plant 5S rRNA Mimic (P5SM) Splicing Model

We are interested in understand the alternative splicing mechanism of Plant 5S rRNA Mimic (P5SM), a highly conserved RNA structure in plants that is part of a suicide cassette exon in *Transcription Factor IIIA* (*TFIIIA*) (112). The P5SM cassette has proven to be very modular, meaning that the whole sequence can be moved from one gene context to another (with very minor sequence requirements) and still retain its function. We are able to shift the splice product ratios using a known RNA-binding protein factor, L5 ribosomal protein. L5 protein binds 5S rRNA preferentially, but will bind P5SM when 5S rRNA levels are low, thus altering *TFIIIA* splicing to promote TFIIIA protein expression, which in turn promotes transcription of 5S rRNA (112). Thus, endogenous P5SM splicing is part of a feedback loop to maintain adequate stoichiometric ratios of L5 ribosomal protein and 5S rRNA in plants. The modular property and known mechanistic details of P5SM have together allowed us to adapt and reengineered P5SM as a tool for transgene regulation in dicot plants (120, 179).

Verifying or modifying the splicing regulation model for P5SM will give us a better understanding of how plants control stoichiometric ratios of 5S rRNA and L5 ribosomal protein. This RNA-protein complex is a structural component required for biogenesis of the large 60S ribosome subunit (180, 181). Differential regulation of endogenous P5SM in *TFIIIA* genes by one or multiple splice factors could be tied to regulation of protein synthesis in different tissue types or under different environmental stresses. Because

the L2 loop region of dicot *Arabidopsis thaliana* P5SM and monocot *Oryza sativa* (rice) P5SM is different, expanding our knowledge of P5SM-protein interactions may give rare insight into how ribosome biosynthesis is differentially regulated among plants. For biotechnological purposes, better understanding the endogenous P5SM mechanism may help us engineer new P5SM-based cassettes for orthogonal gene regulation in different organisms.

Previous experimental work provided circumstantial evidence for a splicing regulation model wherein exon inclusion or skipping correlates to P5SM binding to an exon-defining splice factor or L5, respectively (112, 120). The current model of P5SM splicing regulation proposes that a putative protein factor binds the L2 hairpin loop of the P5SM RNA structure and promotes exon definition and exon inclusion (112). In Hickey et al. 2012, this model was the basis for the engineering of the much-improved Hybrid P5SM (HyP5SM) cassette (120). In an initial attempt to find a P5SM/L5 pair for orthogonal gene regulation in dicot plants, the rice *Os*P5SM structured element was given *Arabidopsis thaliana*-derived exon flanks (the same sequences flanking the original *At*P5SM structured element) as well as *At*P5SM cassette introns. In transiently transformed dicot *Nicotiana benthamiana* leaves, *EGFP-OsP5SM* responded to *Os*L5 protein by promoting exon skipping to favor the functional splice product, SP-I, but exon retention was so low that SP-I was the major product even in the absence of *Os*L5 (120). Thus, *Os*P5SM was not an adequate “suicide” exon for gene regulation in dicot plants. From previous mutations of the *At*P5SM structure, the L2 hairpin loop of *At*P5SM was known to be important for exon retention, but not required for *At*L5 binding (112). With a simple 4 nucleotide addition by site-specific mutagenesis, the *Os*P5SM L2 hairpin loop was changed to the purine-rich *At*P5SM sequence, and the resulting hybrid structured element became HyP5SM. In the absence of *Os*L5, the major splice product of HyP5SM is the non-productive exon-retaining SP-II mRNA, a target for nonsense-mediated decay which gives the cassette its “suicide” cassette nickname (120). We have predicted the existence of an endogenous splice factor that promotes P5SM exon retention by binding to the L2 loop since 2009 (112), and all experimental evidence thus far supports this model (112, 120, 179).

In my publication, Gonzalez et al. 2015 (see **Chapter 2**), I observed that excessive transcription levels of a HyP5SM-regulated gene unexpectedly promoted exon skipping (179). When working with the *avrBs2-HyP5SM* gene jointly regulated by the dexamethasone (Dex)-inducible promoter and the HyP5SM cassette exon, I found that inducing the promoter alone with Dex was able to promote accumulation of modest levels of AvrBs2 protein even in the absence of *Os*L5 (**Figure 2.11**). We hypothesize that the high Dex-induced transcription levels raised *avrBs2-HyP5SM* transcript to high enough concentrations that the putative protein factor that promotes HyP5SM exon inclusion was titrated away.

Herein, to further validate the regulatory model for P5SM, we begin experiments to identify the putative splice factor that promotes exon inclusion. We show the results of a reporter screens which identify RSZ32, RSZ33, RSZp22, SCL33, and RSp31 as proteins that promote reduced expression of *At*P5SM-regulated genes. We show RT-PCR results with RSZ32 genomic and coding sequences, and a mutant. Finally, we

show progress toward a direct test of RNA-protein interactions and discuss future directions.

## 4.2: Results and Discussion

### Selecting Serine/Arginine-Rich Proteins in Plants

Alternative splicing is known to be regulated by various RNA-binding proteins such as serine/arginine-rich (SR) proteins which promote splice site selection by recruiting the spliceosome to weaker splice sites, as well as heterogeneous nuclear ribonucleoparticle (hnRNP) proteins which are known to repress splicing (20, 182–184). For any given gene, we can identify the most abundant splice products through RNA-seq data or by sequencing bands on an RT-PCR gel (see **Chapter 2** for example), but identifying what specific splice factor interactions are important for splice site selection is more challenging. Previously, bioinformatics analysis of RNA-seq databases and *in vitro* SELEX experiments have been used to find patterns in the RNA consensus binding sequences for specific splice factors, mostly from mammalian splice factors (185). Motif searches of the *A. thaliana* genome find 395 genes predicted to encode splicing-associated proteins, including 109 potential splice factors, but very few have been characterized or specifically tied to alternative splicing (183, 186). Thus, the search was focused on the super family of serine/arginine-rich (SR) proteins, characterized by an RNA Recognition Motif (RRM) domain and an arginine/serine-rich (RS) domain (52). *A. thaliana* has 19 SR proteins (52).

We hypothesized that an exon-defining splice factor was required for binding to the purine-rich L2 loop, which may serve as an exonic splicing enhancer (ESE) sequence (187, 188). A targeted approach was taken to analyze the effect of SR proteins that have previously been implicated in alternative splicing in plants (112, 120). We selected candidate genes: RSZ32, RSZ33, SRp34b, SRp34/SR1, SR30, RSZp22, SCL33, and SR45 (52, 183, 189). See **Table 4.1** for accession information. We utilized sequences from the model plant *A. thaliana* (ecotype Colombia-0 when applicable) because it has the most comprehensive genomics, transcriptomics, and proteomics databases for any plant to date (190). Additionally, the sequence requirements, structural information, and molecular regulation of *At*P5SM is currently the best understood (112).

The RS2Z splice factor family was especially interesting because it is a plant-specific SR protein family with a unique domain architecture (**Figure 4.1**) and no predicted homology to mammalian splice factors (189). The two zinc knuckle (ZnK: $CX_2CX_4HX_4C$) domains that characterize the RS2Z family have been shown in other proteins to be important for protein-protein and possible RNA-protein interactions (191). Zinc knuckles contains zinc-chelating residues that are stabilized in the presence of zinc, allowing turns in the protein's structure (192). Notably, TFIIIA protein contains several zinc finger ($C_2H_2$) motifs required for folding and binding to 5S rRNA (119). The L2 loop of P5SM corresponds to the region of 5S rRNA that is bound by TFIIIA, but not L5 ribosomal protein (193). However, TFIIIA protein does not bind the P5SM element, at least not in *A. thaliana* (112).

**TABLE 4.1:** *Plant SR proteins used in this chapter.*

| Sub-family | Name | TAIR Accession | UniProtKB Entry | Protein length (AA) | Ref |
|---|---|---|---|---|---|
| RS2Z | atRSZ32 | At3g53500 | Q9FYB7 | 284 | (189) |
| RS2Z | atRSZ33 | At2g37340 | Q8VYA5 | 290 | (189, 194–197) |
| SF2/ASF | atSRp34b | At4g02430 | F4JHI7 | 278 | (189) |
| SF2/ASF | atSRp34 /SR1 | At1g02840 | O22315 | 303 | (189, 198, 199) |
| SF2/ASF | atSR30 | At1g09140 | Q9XFR5 | 268 | (189, 199, 200) |
| SF2/ASF | atSRp34a | At3g49430 | A2RVS6 | 300 | (201) |
| 9G8 | atRSZp22 | At4g31580 | O81126 | 200 | (189, 202, 203) |
| RS | atRSp31 | At3g61860 | P92964 | 264 | (189) |
| SCL | atSCL33 | At1g55310 | Q9SEU4 | 287 | (189) |
| SR45 | atSR45 | At1g16610 | Q9SEE9 | 414 | (200, 204) |

*Shorthand: at = *Arabidopsis thaliana*

The sequence of the two RS2Z family members, RSZ32 and RSZ33 (**Figure 4.2**), suggests some structural similarity to human 9G8, but they are not close homologues (189). RSZ33 protein auto-regulates its own splicing (194), and presumably this is true for the highly similar RSZ32 as well. The Barta group has found that constitutive overexpression of RSZ33 is toxic in transgenic *A. thaliana* plants; the genome sequence results in sickly transgenic plants and the protein coding sequence (RSZ33 CDS isoform 1) results in sterile plants (194). Since the endogenous *At*P5SM cassette acts as a “suicide” exon in *TFIIIA* that results in splice products targeted by nonsense-mediated decay (112), overexpression of a splice factor that promotes exon retention of the *At*P5SM suicide exon is expected to reduce functional TFIIIA protein levels. In turn, low TFIIIA levels are predicted to result in reduced recruitment of RNA Polymerase III (Pol III) to 5S rDNA genes, and chronically low levels of 5S rRNA. Overexpression of such a splice factor, then, would be predicted to disrupt ribosome biosynthesis.

### Leaf Fluorescence Assay with *EGFP-AtP5SM*

To quickly screen and easily view phenotypic changes in leaves, we elected to use the *EGFP-AtP5SM* reporter gene and the *DsRed2* infiltration control used previously (120). Splicing-regulated green fluorescence from the reporter was normalized by constitutively expressed red fluorescence. Constructs incorporating the *At*P5SM cassette exhibit some basal reporter fluorescence, presumably because it cross-interacts with endogenous L5 protein. Thus, we can screen for splice factors that favor exon retention by observing reduced reporter fluorescence. Although this is an indirect test that cannot verify protein-RNA interactions, the fluorescence screens are advantageous as a first test because they are quantifiable *in vivo*, are faster and less laborious than protein purifications, and allow us to use full-length proteins which have reportedly been difficult to purify (189).

## The RS2Z SR Family Proteins, RSZ32 and RSZ33, Are Initial Hits

We used *Agrobacterium*-mediated infiltrations of *N. benthamiana* leaf halves to compare the effect of the SR proteins against *firefly luciferase* (LUC), a gene which does not affect *At*P5SM splicing. Each leaf had two different samples (with either LUC or an SR protein) on either leaf half to reduce differences caused by variable infiltration success among leaves, making the data paired. Differences in green/red fluorescence ratios between samples were analyzed using Student's two-tailed, paired *t*-test. We tested N-terminal 3x-Myc tagged genomic sequences of RSZ32, RSZ33, and SRp34b. In case an increase in exon retention would not be noticeably higher than background, we also competed each SR protein co-expressed against *At*L5 to see if the SR proteins would be able to reduce the effect of *At*L5.

Leaf scans showed that RSZ32 and RSZ33 genomic sequences both significantly reduced EGFP reporter fluorescence with or without co-expression of *At*L5, but SRp34b had no significant effect (**Figure 4.3**). This suggests that the reduction in reporter fluorescence is due to specific functions of RSZ32 and RSZ33, and not trivially due to general pleiotropic effects from over-expression of serine/arginine-rich proteins. RSZ33 genomic sequence was able to reduce reporter fluorescence induced by +*At*L5 to levels comparable with background expression (+LUC) for *EGFP-AtP5SM* (**Figure 4.3d**). RSZ32 also had a strong and significant effect ($p < 0.01$), but the suppression of the EGFP reporter was not as strong as with RSZ33 (**Figure 4.3**).

## RSZ32 and RSZ33 Coding Sequences Affect the DsRed2 Infiltration Control

In preparation for protein purifications to test RNA-protein interactions *in vitro*, we cloned the RSZ32 coding sequence (CDS) from *A. thaliana* cDNA. We repeated the leaf infiltration experiment to test the coding sequence *in vivo*. RSZ32 CDS has a visibly stronger suppression effect on fluorescence from the *EGFP-AtP5SM* reporter, but we were not able to reliably quantify this suppression effect because suppression of fluorescence from our *35S::DsRed2* normalization control was also observed (**Figure 4.4a**). Since we are using an intronless coding sequence of *DsRed2*, the reduced red fluorescence from our "constitutive" control is likely due to upstream molecular interactions and not a direct DsRed2-RSZ32 interaction.

One possible explanation for reduced fluorescence of our control is that over-expression of RSZ32 CDS resulted in critically low levels of TFIIIA protein, an already limiting transcription factor for the recruitment of RNA Polymerase III to 5S rDNA genes (114). If the pathway to 5S rRNA transcription is disrupted enough, ribosome biosynthesis will be affected. The reduced *DsRed2* expression we observe might be emblematic of a total reduction in protein expression. We attempted to visualize differences in 5S rRNA levels by loading total RNA onto an agarose gel, but the 5S rRNA bands were too weak for quantification (not shown). In the future, these hypotheses can be tested by RT-PCR to look for changes in *TFIIIA* splicing and western blots to look for alterations of TFIIIA protein as well as various housekeeping proteins. However, with the current data, we cannot rule out that the reduced fluorescence is not due to pleiotropic effects unrelated to TFIIIA protein levels. We do not think that *TFIIIA* pre-mRNA is the only endogenous target of RSZ32.

Meanwhile, to verify that RSZ32 CDS can have a reduction effect on other genes, we repeated the leaf scan experiment using reporter pBinAR *TFIIIA-M8-EGFP* (112), a mutant construct with *At*P5SM in its native context in genomic TFIIIA. The M8 mutant has a truncation in the L2 loop that constitutively disrupts suicide exon definition, resulting in constitutive splicing to a productive mRNA product and thus constitutive green fluorescence from the TFIIIA-EGFP fusion protein. RSZ32 CDS reduced fluorescence from the *TFIIIA-M8-EGFP* reporter, and suppression of the DsRed2 control was observed here as well (**Figure 4.4b**). As expected, we saw a similar result with the *EGFP-OsP5SM* reporter (**Figure 4.4c**). RSZ33 CDS had a similar effect on reporters (**Figure 4.4d**).

**RT-PCR Shows That RSZ32 Promotes *At*P5SM Exon Retention**

We could not use the leaf scans to infer how RSZ32 CDS affected reporter splice ratios because RSZ32 CDS had a pleiotropic effects on overall protein expression. Our initial attempts to express recombinant RSZ32 in *E. coli* for *in vitro* RNA-protein binding assays also proved problematic (not shown). We decided to go ahead and examine the effects of RSZ32 genomic, wild-type CDS, and mutant CDS constructs on reporter splicing by RT-PCR. We cloned RSZ32 CDS mutations in conserved residues to look for changes in *AtP5SM* splicing. We made three domain mutants, each with two mutations to conserved residues. The ΔRRM alanine mutant (Y46A/F48A) disrupts conserved aromatic residues of the Rnp1 sub-domain which have been found to be important for RNA-protein interactions in RSZp22 (205). The ΔZn1 (C101S/C104S) and ΔZn2 (C123S/C126S) mutants disrupt the two zinc knuckle domains in RSZ32 by eliminating the thiol side chains from the first two cysteines in the $CX_2CX_4HX_4C$ zinc knuckle motif, without drastically altering the polarity or size of the side chains (**Figure 4.5a**). Genomic RSZ32, wild-type RSZ32 CDS, and mutant RSZ32 CDS constructs were co-expressed in *N. benthamiana* with reporter *EGFP-AtP5SM*.

Genomic RSZ32 shows a modest reduction in SP-I which supports the hypothesis that RSZ32 is preventing skipping of the suicide exon, but SP-II accumulation is not clearly noticeable (**Figure 4.5b**, lanes 4-7). This may be due to the fact that SP-II mRNA is subject to nonsense-mediated decay, so modest increases in SP-II may not be visible. Addition of wild-type RSZ32 CDS shows a clear shift in the reporter splice ratio from SP-I to SP-II, compared to the LUC control which has no effect on splicing (**Figure 4.5b**, lanes 8-9). This supports the hypothesis that RSZ32 CDS has an effect on endogenous *TFIIIA* splicing and that this may at least partially explain its pleiotropic effects on overall protein expression (**Figure 4.4a**). The ΔRRM mutant samples were ambiguous, with ΔRRM showing an unexpected shift from SP-I to SP-II as compared to LUC (**Figure 4.5b**, lanes 10-11), but also a shift to SP-II as compared to wild-type CDS (**Figure 4.5b**, lanes 13-14).

RT-PCR results are not yet available for the zinc knuckle mutants. However, infiltration of the coding sequence and mutants in *N. benthamiana* leaves showed a mild chlorosis (yellowing) phenotype resulting from the RSZ32 CDS, ΔRRM, and ΔZn1 constructs, as compared to the LUC control (not shown). This phenotype supports the hypothesis that general protein expression is being affected by RSZ32. Interestingly, in the same experiment, the observed chlorosis effect was much weaker for ΔZn2 for all

biological replicates (not shown). This, coupled with the RT-PCR results for ΔRRM, suggests that the ΔRRM and ΔZn1 mutants may retain some RSZ32 function, whereas ΔZn2 may be a less functional mutant. If so, the second zinc knuckle domain may be more important than the first zinc knuckle domain for RSZ32 interaction with downstream players. This hypothesis should be examined in the future; there is not enough evidence at this time to draw strong conclusions.

**Screening of Known RSZ33 Interacting Proteins**

Meanwhile, because wild-type RSZ32 promoted *At*P5SM exon retention (**Figure 4.5b**), we decided to examine known SR protein binding partners to see if these splice factors would have similar effects on protein expression. RSZ32 protein is not well-characterized, but the Barta group has performed several experiments to characterize the protein-protein interactions of RSZ33. C-terminal truncated RSZ33 lacking the SP-dipeptide repeat domain was shown to interact with *A. thaliana* splice factors SRp34/SR1, RSZp21, RSZp22, SCL28, SCL30, SCL33, and SC35 in a yeast-2-hybrid screens and *in vitro* binding assays (189). SCL33 was specifically interesting because its transcriptional activity overlaps with RSZ33's expression and the proteins bind each other (189). RSZp21, RSZp22 are notable because they are members of a family characterized by a single zinc knuckle motif, similar to human 9G8 SR protein (203).

Representative family members were co-infiltrated into *N. benthamiana* with *EGFP-AtP5SM, DsRed2,* as well as +*At*L5 for increased signal (**Figure 4.6**). As expected, the RSZp22 and SCL33 suppressed fluorescence from *EGFP-AtP5SM* (**Figure 4.6a**). However, SCL33 had a weaker effect, and suppression of the reporter was only significantly observed with the coding sequence (**Figure 4.6df**). Neither the genomic nor the tested coding sequences had an effect on DsRed2, allowing us to interpret the effects on reporter expression as non-pleiotropic (**Figure 4.6**).

The negative controls were meant to be known RSZ33 non-interacting proteins, SRp34b and RSp31. Unexpectedly, RSp31 showed a significant ($p < 0.01$) albeit modest suppression of EGFP expression. Previous researchers did not find RSp31 to interact with RSZ33 either through yeast-2-hybrid screens (with truncated RSZ33ΔSP) or co-immunoprecipitations from *E.coli* samples expressing full-length RSZ33 (189). In the experiments shown here, we utilized full-length splice factors expressed *in planta*. RSp31 may interact with RSZ32/RSZ33 in these conditions, or alternatively RSp31 may suppress EGFP expression indirectly by affecting expression of other splice factors. We did not observe a suppression of the DsRed2 normalization control with RSp31 (not quantified, but shown in **Figure 4.6e**), which suggests that the reduction of reporter fluorescence is at least somewhat specific and not due to general protein suppression.

**Epitope Tag Location Affects the Function of RSZ32 and RSZ33**

At this point, we wanted to proceed with protein purification of the SR proteins expressed in *N. benthamiana* tissue. With purified protein, we would be able to directly test RNA-protein binding between *At*P5SM and the splice factors *in vitro* through various methods, including gel-shift assays or DRaCALA assays or affinity chromatography. In order to ensure that we recovered full-length protein in our purifications, we re-cloned the SR protein constructs to remove the N-terminal 3xMyc

tag and instead have C-terminal FLAG and 3xHA tags. Unfortunately, upon testing the new tagged splice factors, we discovered that the C-terminal tag disrupted the previously observed effects on *At*P5SM-regulated gene expression (**Figure 4.7**). Whereas 3xMyc-RSZ32 genomic sequence reliably reduces fluorescence from *EGFP-AtP5SM* (**Figure 4.3, 4.7**), RSZ32-3xHA has no significant effect (**Figure 4.7b**). No issue was found with the construct's identity after re-sequencing, and its protein expression appears to be very robust as visualized by an anti-HA western blot (**Figure 4.7c**). The same reduced function was seen with the C-terminal FLAG tag (not shown), suggesting that it is the location and not the sequence composition of the tag that is important.

Notably, the Barta group was able to see nuclear localization with RSZ33-GFP fusions, so a C-terminal tag does not affect all functions of the splice factors (189, 196). We predict that the C-terminal tags disrupt protein-RNA, but not protein-protein interactions. A recent study reports that one of the two kinases known to phosphorylate serines in RS domains, CLK1, can efficiently target serines in both RS and SP dinucleotides (206). All serine/arginine-rich proteins have some SP dinucleotides in their RS domain, in addition to the RS dinucleotides for which the domain and proteins are named (206). Both RSZ32 and RSZ33 contain particularly long SP-repeat domains (189) (**Figure 4.1**). In human SRSF1 protein, phosphorylation of SP dinucleotides was found to promote different conformational changes and cooperative binding to the target RNA, possibly by affecting accessibility to the RRM domain (206). The RRM domain of RSZp22 can be mutated to disrupt RNA recognition without disrupting protein nucleotide localization (205), and we think this is true for RSZ33 as well. We speculate that adding a C-terminal epitope tag may either be preventing accessibility to CLK1 or disrupting C-terminal conformations important for RNA-binding activity.

Our data taken together, we have identified candidate serine/arginine-rich proteins that regulate *At*P5SM, directly or indirectly, and have predicted other hits untested here due to time limitations. We have begun characterization of the RSZ32 protein and coding sequence (**Figure 4.3, 4.5**), previously only identified by bioinformatics and as part of a larger screen for effects of abscisic acid treatment on expression SR genes (46, 189). We discovered an effect on overall protein expression due to RSZ32 CDS and RSZ33 CDS over-expression which was not seen for other SR protein coding sequences (**Figure 4.4, 4.6**). We also presented evidence suggesting that the C-terminal domain is important for RNA target interaction (**Figure 4.7**).

## 4.3: Future Directions and Proposed Experiments

### Is the Interaction Between the SR Proteins and *At*P5SM Direct?

The future direction of this project must now focus on getting direct evidence for RNA-protein interactions between *At*P5SM and the identified SR proteins. The fluorescence leaf scans with reporter *EGFP-AtP5SM* and normalization control *DsRed2* provided useful information to screen for expression changes *in planta*, but we cannot conclude that RSZ33 and related proteins are binding *At*P5SM RNA. SR proteins are known to auto-regulate themselves and other SR proteins (197), so the results of over-expression *in planta* could be due to SR protein-protein interactions rather than

interactions with the *At*P5SM element. In order to directly assay for RNA-protein interactions, I propose the following two approaches:

First, we could attempt an affinity purification using biotinylated *in vitro* transcribed *At*P5SM sequence bound to streptavidin magnetic beads. RNA-binding proteins from plants have been found with this method before (207). Wild-type *A. thaliana* total protein extract would be RNase-treated and incubated with the beads at room temperature (207), then gently washed several times. *At*P5SM-bound protein could be eluted by washing the column with a solution containing recombinant *At*L5 protein expressed in *E. coli* as previously described (112). Based on the P5SM model, *At*P5SM binds the putative splice factor at the L2 loop, but this interaction is disrupted at high levels of *At*L5 ribosomal protein (112). We are taking advantage of this disruption to recover non-L5 proteins bound at *At*P5SM. To control for unrelated interactions, we can repeat the experiment using the M8 mutant of *At*P5SM (112), containing a truncated L2 loop that results in constitutive exon skipping, presumably due to loss of the putative exon-defining protein. The resulting elutions would be compared by mass spectrometry to find proteins that bind wild-type *At*P5SM and not the M8 mutant. The advantage of such a screen is that it is unbiased and could yield proteins we have not yet considered. The challenges of such a screen are that the exon-defining protein may be too low abundant to be detected by mass spectrometry. Splicing-regulating proteins such as the SR proteins are not highly expressed and previous lab members who attempted such an unbiased screen could not detect specific interactions (unpublished, personal communication with Kathrin Leppek). Potentially, incubating the *At*P5SM-bound beads with *A. thaliana* nuclear extract instead of total protein could yield better results; this has worked for mammalian splice factors (208). Protocols for nuclear extract preparation from plants are available (209).

A second method that is more compatible with low abundant proteins may be to utilize Differential Radial Capillary Action of Ligand Assay (DRaCALA) to detect RNA-protein interactions between *in vitro* transcribed *At*P5SM and either cell extract from *N. benthamiana* infiltrated to over-express different SR proteins or purified SR proteins. A disadvantage of the DRaCALA method is that it is not unbiased (210); we would need to select beforehand which proteins will be part of the screen. A strong advantage, however, is that this method can be used with different concentrations of *At*P5SM RNA and candidate protein, allowing us to measure binding affinity and other kinetics of the RNA-protein interactions (210). As an initial test, serial dilutions of wild-type *N. benthamiana* total protein extract can be blotted onto nitrocellulose membrane to find a dilution point where DRaCALA no longer shows an interaction between radiolabeled *At*P5SM and endogenous proteins such as *Nb*L5. At this dilution point, total protein extract from *N. benthamiana* leaf tissue transiently expressing *A. thaliana* 3xMyc-SR proteins can be tested for an interaction with *At*P5SM and the M8 mutant negative control. This procedure allows for higher-throughput testing of candidate *A. thaliana* proteins without the need for protein purification. Suggested candidates include: all 19 SR proteins, *At*L5 as a positive control, and LUC as a negative control. If background signal from endogenous *N. benthamiana* proteins is too high, the SR proteins may be enriched using the anti-3xMyc antibody. Importantly, recombinant protein from *E.coli* may not yield functional protein for RNA-protein binding assays. Post-translational

phosphorylation of the RS domain have been shown in other SR proteins to be required for binding target RNA sequences (211).

### Are the Pleiotropic Effects from RSZ32 and RSZ33 Due to Reduced TFIIIA Levels?

In addition to finding direct RNA-protein interactions, it is also worth investigating the reason for reduced DsRed2 expression with RSZ32 CDS and RSZ33 CDS. The intronless, plant codon-optimized *DsRed2* gene is driven by the constitutive 35S Cauliflower Mosaic Virus promoter in those experiments, yet the RS2Z family splice factors affect its expression. We speculate that what we are seeing is a loss of total protein biosynthesis, but this has not yet been proven. Western blotting for TFIIIA protein, P5SM-regulated EGFP protein, and various housekeeping genes may prove a reduction of protein. The loading control should be a protein with low turnover whose levels will not be quickly affected by the expected disruption of ribosomal biosynthesis. This experiment may require a time course.

### Do RSZ32 and RSZ33 Have Distinct Roles?

Finally, although RSZ32 and RSZ33 are highly conserved (75% identity) and have thus far produced similar results, they may not function in the same way *in planta*. Some consideration should be given to expression profile differences between RSZ32 and RSZ33. To our knowledge, RSZ32 expression has not yet been characterized, but some expression information for RSZ33 is available (189).

## 4.4: Figures

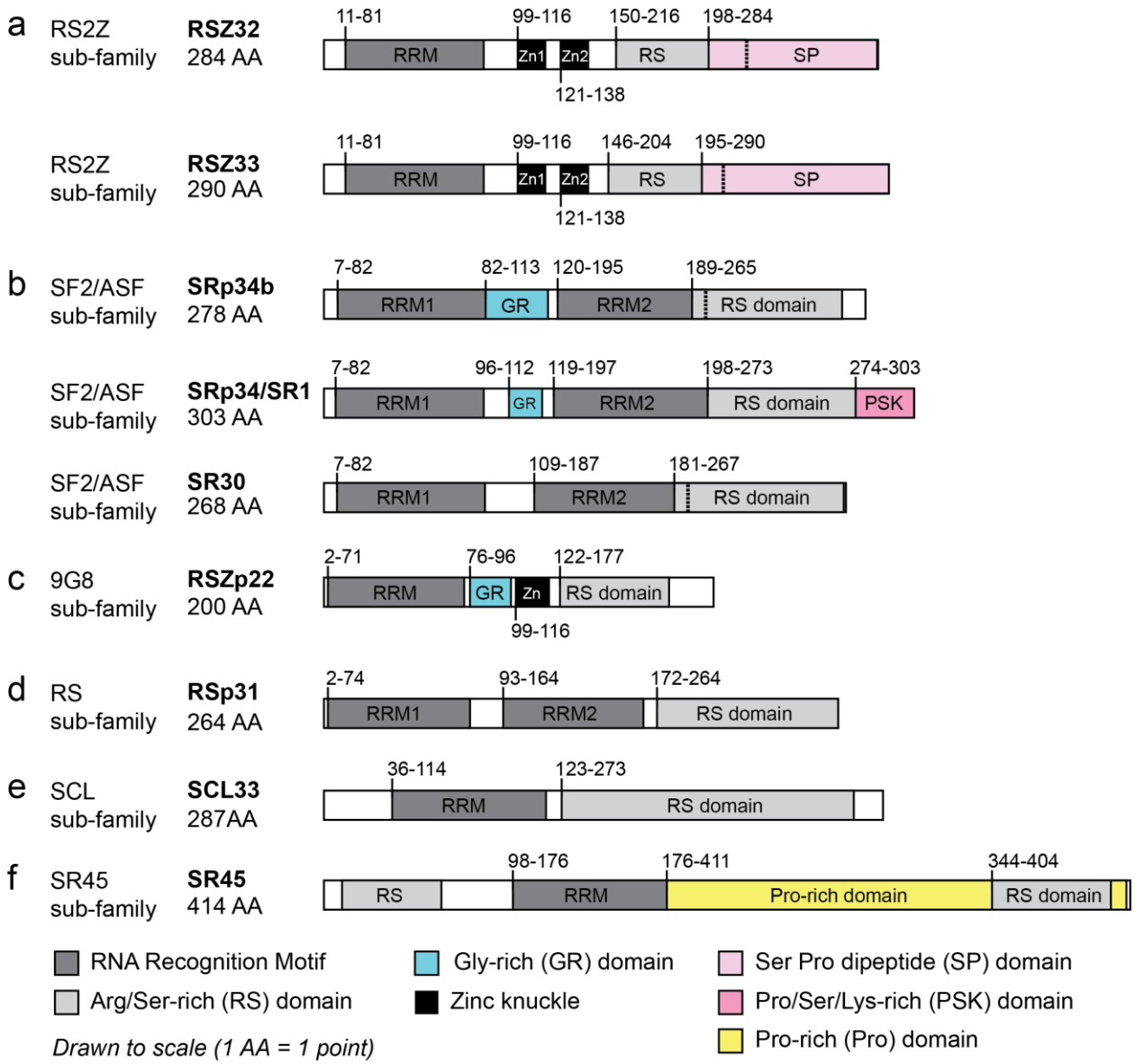


**FIGURE 4.1. Domain architecture of splice factor proteins used in this study.** Serine/arginine-rich (SR) proteins were cloned out of *Arabidopsis thaliana* Col-0 genomic DNA or cDNA. (a) RS2Z family characterized by two zinc knuckle domains, no mammalian homologs; (b) SF2/ASF family, characterized by a true RRM followed by a second pseudo-RRM domain; (c) 9G8 family characterized by a single zinc knuckle domain; (d) RS family; (e) SCL family; (f) SR45 family.

```
Sequence 1: RSZ32        284 aa
Sequence 2: RSZ33        290 aa


RSZ32     MPRYDDRYGNTRLYVGRLSSRTRTRDLERLFSRYGRVRDVDMKRDYAFVEFSDPRDADDA    60
RSZ33     MPRYDDRYGNTRLYVGRLSSRTRTRDLERLFSRYGRVRDVDMKRDYAFVEFGDPRDADDA    60
          ************************************************.********

RSZ32     RYYLDGRDFDGSRITVEASRGAPRGSRDNGSRGPPPGSGRCFNCGVDGHWARDCTAGDWK    120
RSZ33     RHYLDGRDFDGSRITVEFSRGAPRGSRDFDSRGPPPGAGRCFNCGVDGHWARDCTAGDWK    120
          *:***************  **********  .*******:**********************

RSZ32     NKCYRCGERGHIERNCKNSPSPKKARQGGSYSRSPVKSRSPRRRRSPSRSRSYSRGRSYS    160
RSZ33     NKCYRCGERGHIERNCKNSP--KKLRRSGSYSRSPVRSRSPRRRRSPSRS--LSRSRSYS    176
          ********************  ** *:.********:*************   **.****

RSZ32     RSRSPVRR-EKSVEDRSRSPKAMERSVSPKGRDQSLSPDR----KVIDASP--------K    227
RSZ33     RSRSPVRRRERSVEERSRSPKRMDDSLSPRARDRSPVLDDEGSPKIIDGSPPPSPKLQKE    236
          ******** *:***:****** *: *:**:.**:*   *     *:**.**        :

RSZ32     RGSDYDGSPKENGNGRNSASPIVGGGESPVGLNGQDRSPIDDEAELSRPSPKGSESP       284
RSZ33     VGSDRDGGSPQDNGRNSVVSPVVGAGGDSS---KEDRSPVDDDYEPNRTSPRGSESP       290
           *** **.. :::. .. .**:**.* ..     :****:**: * .*.**:*****
```

**FIGURE 4.2. Clustal W multiple sequence alignment between RSZ32 and RSZ33.** The proteins are 71.1% identical. The alignment below the sequences marks amino acids as identical (asterisk), similar (colon), somewhat similar (period) or not similar at all (blank space, yellow highlighting). The RRM domain is indicated with light blue text; zinc knuckle domains, red text, essential residues underlined; RS domain, purple text.

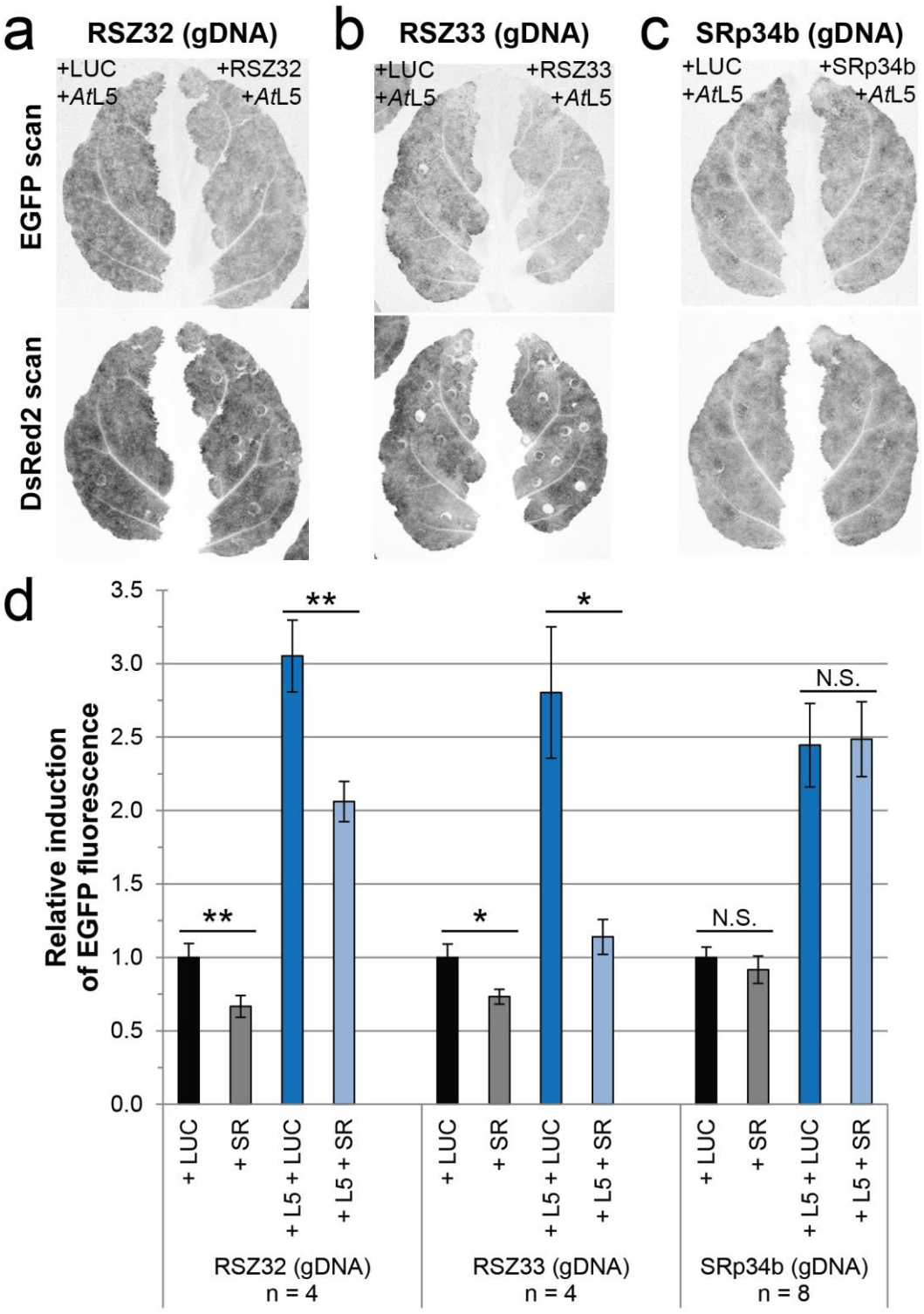


**FIGURE 4.3. Small initial screen finds RSZ32 and RSZ33.** *N. benthamiana* leaves were transiently co-transformed with 1:1:1 or 1:1:1:1 mixes of *Agrobacterium* carrying pBinAR (35S promoter) *EGFP-AtP5SM*, *DsRed2,* either the indicated SR protein gene (genomic sequence) or control *firefly luciferase* (LUC), and *AtL5* where indicated. Unless otherwise stated, all serine/arginine-rich splice factors in this chapter have an N-terminal 3xMyc-tag. (a-c) Representative leaf scans at 3 dpi for EGFP or DsRed2 fluorescence, for *+AtL5* leaves. A similar experiment was conducted for leaves without *AtL5* co-

expression, scans not shown. (d) Quantification of leaf scans by Pooja Vijayendra, reanalyzed for significance and re-graphed by Tania Gonzalez. EGFP fluorescence from the *EGFP-AtP5SM* reporter was normalized using DsRed2 to account for differences in *Agrobacterium* infiltration efficiency between leaves, and EGFP/DsRed2 ratios for each batch of leaves were then normalized again to the "+LUC" control to compare effects of co-expression with *At*L5, SR protein, and *At*L5 + SR protein. Error bars are standard error of the mean. Significance analyzed with Student's two-tailed, paired *t*-test. Significance markers: * $p < 0.05$; ** $p < 0.01$; N.S. = not significant, $p > 0.05$.

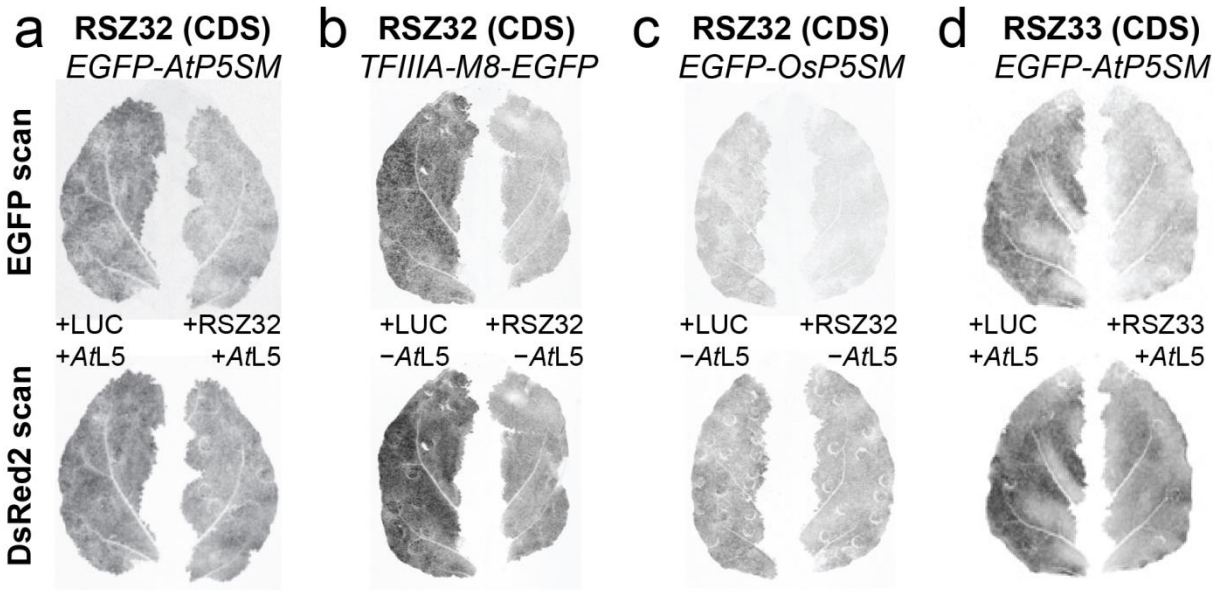


**FIGURE 4.4. RSZ32 and RSZ33 coding sequences reduce DsRed2 expression.** Representative *N. benthamiana* leaf scans 3 dpi showing that DsRed2 fluorescence is reduced by the RS2Z family members. The *DsRed2* gene is driven by the constitutive 35S promoter and is typically used as a normalization control for fluorescence from the EGFP reporters. (a-c) *3xMyc-RSZ32(CDS)*, (d) *3xMyc-RSZ33(CDS)*. The reporter genes are (a, d) *EGFP-AtP5SM;* (b) *TFIIIA-M8-EGFP*, an *At*P5SM mutant from Hammond et al. 2009 which has a truncation in the L2 loop and constitutively skips the P5SM M8 mutant exon; (c) *EGFP-OsP5SM*, which has a naturally truncated L2 loop. (b, c) Leaf scans by Megan Dowdle.

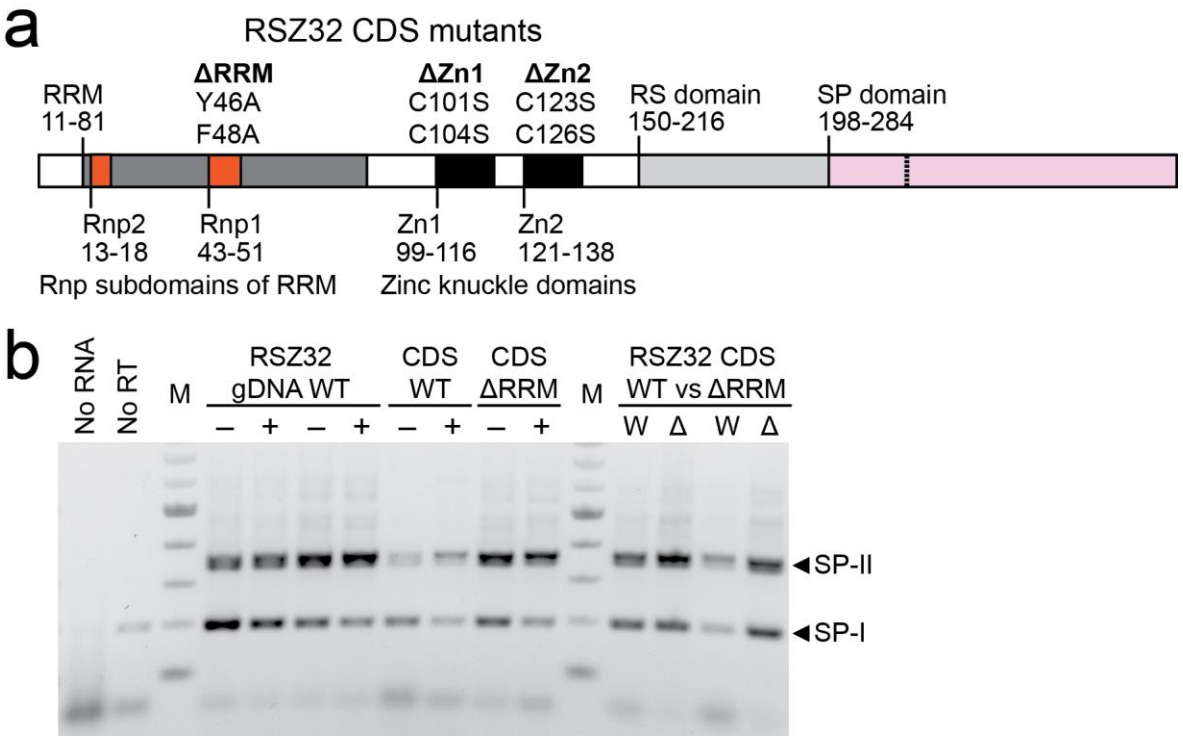


**FIGURE 4.5. RSZ32 coding sequence mutations.** (a) Domains lengths are drawn to scale relative to each other. RSZ32 coding sequence (CDS) is 284 amino acids in length. The Rpn1 subdomain of RRM was mutated Y46A/F48A to disrupt the aromatic rings which are an integral part of binding to nucleic acids (205). The first zinc knuckle domain (Zn1) was mutated C101S/C104S. The second zinc knuckle domain (Zn2) was mutated similarly C123S/C126S. These cysteine to serine mutations maintain the polar characteristic of the residues at those locations, but target the thiol side chains that help characterize the zinc knuckle domain, mutating the $CX_2CX_4HX_4C$ motif to a $SX_2SX_4HX_4C$ motif. These mutations can disrupt disulfide bonds potentially necessary for protein-protein or protein-RNA interactions (191). Mutations were cloned by Megan Dowdle using site-specific mutagenesis. (b) RT-PCR with cDNA synthesized using iScript RT (BioRad) with oligo(dT) primers. RT-PCR samples from same leaves are next to each other (-/+ splice factor) for paired comparisons. For some samples, there are two biological replicates. Primers amplify the splice products of *EGFP-AtP5SM*. All RSZ32 constructs have an N-terminal 3xMyc tag. No *At*L5 was co-expressed for the samples used for RT-PCR. M = 2-log DNA ladder.

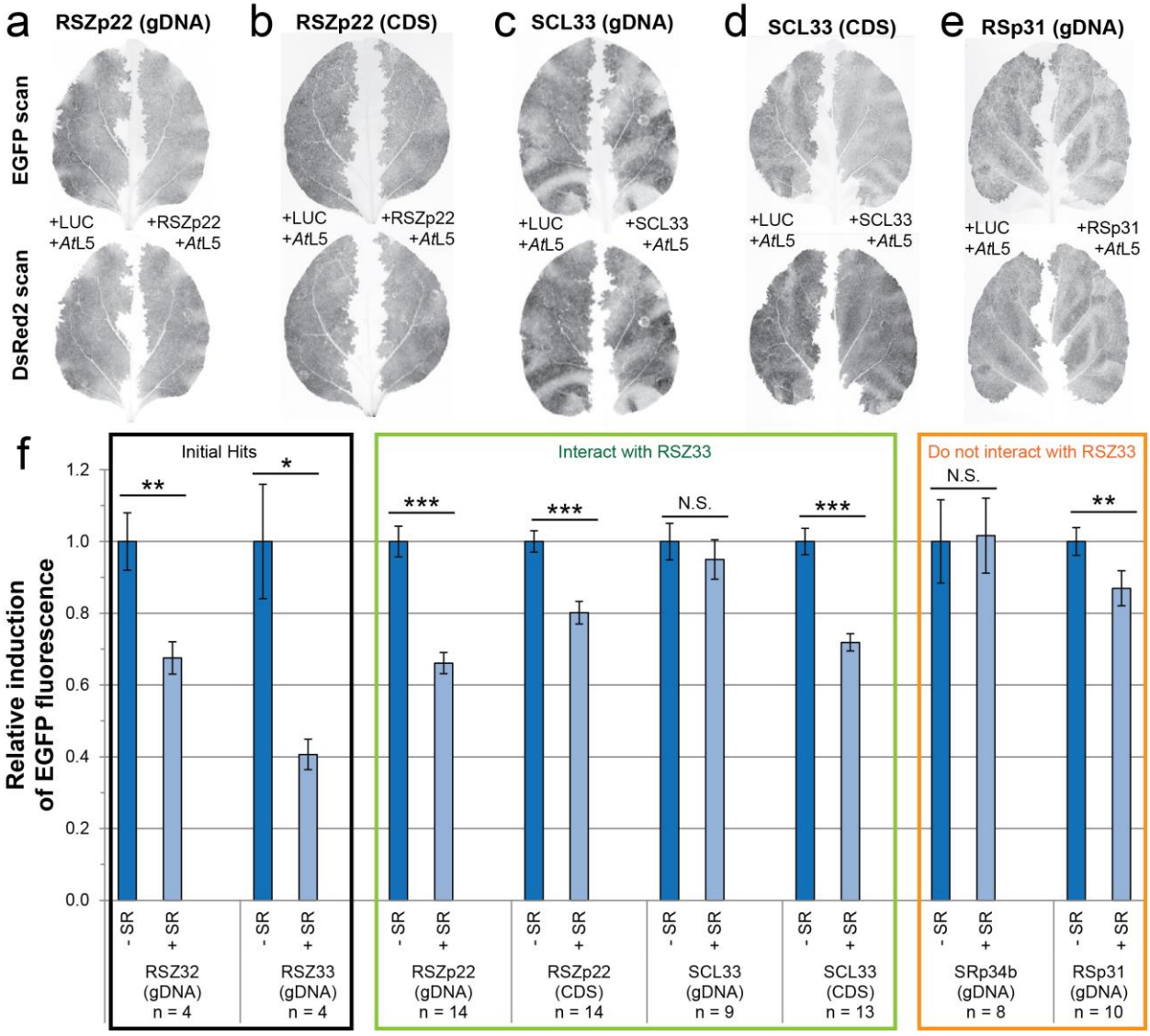


**FIGURE 4.6. Different SR proteins tested for effect on *EGFP-AtP5SM* (all +*At*L5).** (a-e) Representative *N. benthamiana* leaves co-infiltrated with pBinAR *EGFP-AtP5SM* (reporter gene), *DsRed2* (fluorescence control), *AtL5*, and either *firefly luciferase* (LUC) or the indicated N-terminal 3xMyc-tagged SR gene. Leaves were scanned 3 dpi. (a) RSZp22 genomic sequence, (b) RSZp22 coding sequence, (c) SCL33 genomic sequence, (d) SCL33 coding sequence, (e) RSp31 genomic sequence. Both RSZp22 and SCL33 have previously been shown to interact with RSZ33, whereas RSp31 is not expected to interact with RSZ33 (189). (f) EGFP/DsRed2 fluorescence ratio for each leaf normalized to the +LUC +*At*L5 sample ("-SR"). Raw data for RSZ32, RSZ33, SRp34b is the same as used for **Figure 4.3**, but normalized to the +LUC +*At*L5 sample ("-SR") instead of +LUC –*At*L5 samples (omitted here) in order to allow comparison with other splice factors. Error bars are standard error of the mean. Significance analyzed with Student's two-tailed, paired *t*-test. Significance markers: * $p < 0.05$; ** $p < 0.01$; *** $p<0.001$; N.S. = not significant, $p > 0.05$.

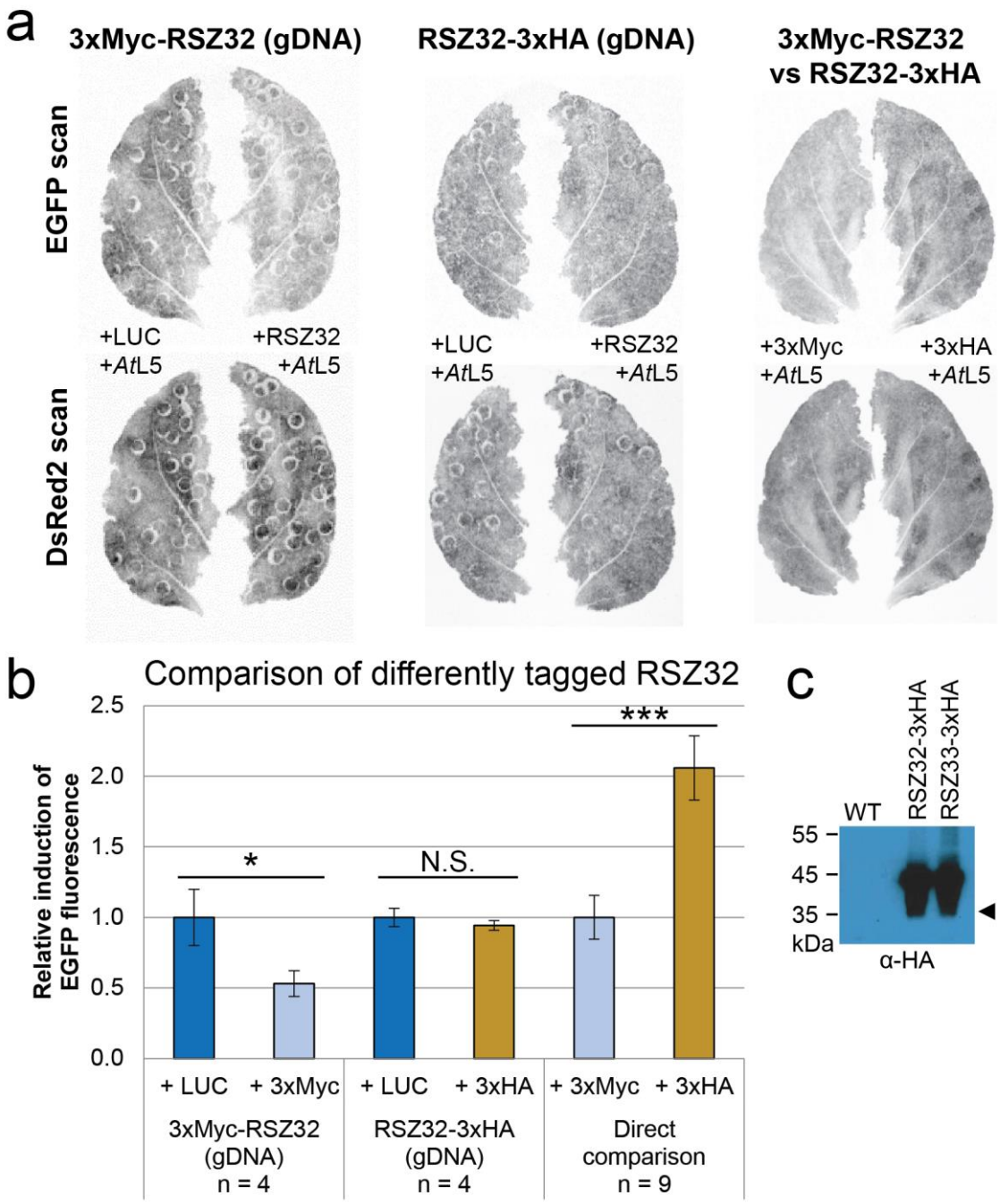


**FIGURE 4.7. The position of the epitope tag affects function.** (a) *N. benthamiana* leaves co-infiltrated as before (with +*At*L5), comparing: LUC versus an RSZ32 genomic sequence tagged with an N-terminal 3xMyc tag or a C-terminal 3xHA tag, or a direct comparison between the two tagged constructs. Scanned 3 dpi. (b) Quantified EGFP fluorescence scaled to DsRed2 expression from the same leaf portion, then normalized to +LUC or, in the direct comparison, to the 3xMyc-tagged construct. Error bars and significance markers as before. LUC versus 3xMyc-RSZ32 data here is an experimental replicate of data in Figure 4.2, not repeated data. (c) Anti-HA western blot of LUC ("WT"), RSZ32-3xHA, and RSZ33-3xHA samples shows protein expression. Significance analyzed with Student's two-tailed, paired *t*-test. Significance markers as before. Leaves were infiltrated by Bao Nguyen and Tania Gonzalez.

## 4.5: Materials and Methods

### Cloning of Genomic and Coding Sequences from *Arabidopsis thaliana*

Wild-type *Arabidopsis thaliana* Col-0 ecotype plants were grown to rosette stage and 1-3 leaves of tissue were collected. Tissue was flash frozen in liquid nitrogen and pulverized using steel beads with a TissueLyser II, then returned to liquid nitrogen for further processing. Genomic DNA was extracted using DNeasy Plant Mini Kit (QIAGEN) and used directly in PCR. RNA was extracted using GeneMATRIX Universal RNA Purification Kit with RL buffer (CHIMERx), treated with RQ1 DNase (Promega), and used for cDNA synthesis using oligo(dT) primers and SuperScript III (BioRad). Genomic sequences were PCR amplified between the ATG start codon and the necessary stop codon, as defined by the genomic sequences or coding sequences annotated by The Arabidopsis Information Resource (TAIR) which correspond to the protein sequence of splice factors used in relevant publications cited in results. Most SR proteins had multiple isoforms, so identifying the coding sequence that translates to the full-length protein used in cited literature was important; the UniProt database was useful for this. Amplified bands corresponding to the expected gene size were gel-extracted using the QIAquick Gel Extraction Kit (QIAGEN), then cloned into pCR2.1 TOPO vector using the TOPO TA Cloning Kit (Invitrogen) and verified by sequencing.

The pBinAR *DsRed2* binary vector (112, 159) was modified by double digesting with KpnI-HF/BamHI-HF restriction enzymes (New England Biolabs) to cut the vector between the constitutive 35S cauliflower mosaic promoter and the *DsRed2* gene, preserving both elements. A yeast-optimized *3xMyc* tag was ligated into the KpnI/BamHI site to make pBinAR *35S-KpnI-3xMyc-BamHI-DsRed2-SalI*, also called pBinAR *3xMyc-DsRed2*. This vector was utilized as a backbone for the splice factors unless otherwise noted, resulting in proteins with an N-terminal 3xMyc tag. The pBinAR *3xMyc-DsRed2* vector was digested with BamHI-HF/SalI-HI restriction enzymes and treated with Calk Intestinal Alkaline Phosphatase (New England Biolabs). The splice factors were PCR amplified with primers that added BamHI and SalI sites, digested with BamHI-HF/SalI-HF, and purified with the QIAquick PCR Purification Kit (QIAGEN) to remove enzymes. For splice factors that contained an internal SalI site, the insert was flanked by BamHI and SpeI sites instead, because SpeI and SalI overhangs are compatible. The digested inserts were ligated into the pBinAR vector using T4 ligase (New England Biolabs). Ligation products were used to transform competent *E. coli* cells, either TOP10 cells (ThermoFisher Scientific) or NEB10-beta cells (New England Biolabs), and positive transformations were selected using LB agar plates with 50 µg/mL kanamycin. Positive colonies were sequenced again to confirm successful ligation products.

The SR proteins with C-terminal 3xHA and FLAG tags were cloned using pBinAR vectors with 3xHA or FLAG tags cloned out of *Bs2-3xHA* and *Bs2-FLAG* constructs from Chapters 2 and 3, also by PCR amplification and restriction digest cloning. The integrity of the tags was verified by sequencing.

## Cloning of RSZ32 CDS Mutants

Site-specific mutations as described in the results section were achieved using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) and PCR primers designed as specified in the manual. The QuikChange PCR reaction product was digested with DpnI to eliminate the *E.coli*-amplified template vector, leaving only the un-methylated PCR product. One Shot TOP10 Chemically Competent *E.coli* cells (ThermoFisher Scientific) were transformed with the DpnI-digested product, and sequenced through the mutagenesis target to confirm a successful mutation.

## *Agrobacterium*-mediated Leaf Infiltrations of *Nicotiana benthamiana*

*Nicotiana benthamiana* plants were grown 3 weeks in a greenhouse (Oxford Tract Greenhouse Facility), then transferred to an open growth cart under constant light and ambient temperatures (24-27°C under the lights) for approximately 1 week before experiments. *Agrobacterium tumefaciens* strain GV2260 was electroporated with the pBinAR binary vectors containing the desired genes of interest, independently. GV2260 with pBinAR vector was selected on LB agar plates with kanamycin (50 µg/mL), carbenicillin (50 µg/mL), and rifampicin (50 µg/mL). Single colonies were re-streaked onto a fresh selection plate or LB broth culture for further experiments. GV2260 culture was pelleted and washed with 0.5X PBS buffer, then normalized to $OD_{600}$ = 0.8 as previously described (120). Normalized *Agrobacterium* solutions were mixed 1:1:1 or 1:1:1:1 as needed and co-infiltrated into the back of leaves with a 1 mL blunt syringe. Approximately 2 leaves were infiltrated per plant. Leaves were blotted gently with a clean kimwipe after infiltration to remove excess *Agrobacterium* solution and reduce potential cross-contamination. Plants were left in the growth carts on unwatered plant trays at room temperature with the lights off overnight, then lights were turned on again and plants were left at constant light. Plants were watered as needed, 1-3 days post-infiltration (dpi). Tissue was collected 2 dpi for RNA and 3 dpi for protein.

## Leaf Fluorescence Scans and Quantification EGFP Induction

*N. benthamiana* leaves were scanned for fluorescence 3 days post-infiltration (dpi) using a Typhoon 9410 Variable Mode Imager or Typhoon TRIO+ (Amersham Biosciences) with the fluorescence setting under normal sensitivity, PMT = 300V, “pressed sample” mode enabled to flatten the leaves, with excitation/emission filters for EGFP (488/520 nm) or DsRed2 (532/580 nm). The leaves were placed with the infiltrated side (the back of the leaves) facing the glass. Raw data from the .gel images was collected using manufacturer-provided software, ImageQuant as follows: 4 equal boxes were located on DsRed2-expressing areas on each leaf half for every infiltrated leaf, avoiding leaf veins and syringe-damaged areas. The DsRed2 boxes were copied over to the EGFP scan so that each box would be in the same position. “Volume” and “area” information was exported for boxes from each scan, where here volume accounts for the quantifying area and pixel strength inside the area. In Microsoft Excel, total DsRed2 was divided by average DsRed2 for each box. The lowest 15% of DsRed2/avgDsRed2 was inspected using conditional formatting on Excel. Leaves with >3 “low” DsRed2 boxes were discarded, and leaves with uneven “low” DsRed2 levels between leaf halves were discarded. The remaining leaves had EGFP volume divided by DsRed2 volume to get an EGFP/DsRed2 ratio for each box. The EGFP/DsRed2 ratio

was averaged for each set of 4 boxes to get the average value of that leaf half. In cases where leaves were normalized to LUC, the average EGFP/DsRed2 for LUC-infiltrated leaf halves was used to normalize each individual leaf half. The data was normalized this way to avoid artificially setting the standard deviation of LUC to zero, as would occur if each leaf were normalized to its own LUC value.

To analyze whether data sets were significantly different, LUC-normalized EGFP/DsRed2 values for LUC-expressing or SR protein-expressing leaf halves were compared using Student's paired *t*-test assuming a two-tailed distribution. The resulting p-values were marked significant (*) if $p < 0.05$, more significant (**) if $p < 0.01$, and most significant (***) if $p < 0.001$. Error bars on graphs are standard error of the mean (standard deviation divided by the square root of the sample size n) to account for different sample sizes.

Claims that SR proteins did not affect DsRed2 are based on the observation that the lowest 15% of DsRed2 expression among leaf halves was fairly evenly distributed between -/+ SR samples (not shown, but seen during quantification). Variability between separate leaves, but not between leaf halves, is a normal and expected quality of the raw data. It is why we infiltrate leaf halves and normalize reporter green fluorescence by DsRed2 fluorescence. When attempting to quantify results with RSZ32 CDS and RSZ33 CDS, the lowest DsRed2 expression samples were clearly over-represented in the +SR leaf halves (not shown).

### RNA Extraction and RT-PCR for *N. benthamiana* Experiments

Leaf tissue was processed for RT-PCR as described in **Chapter 2**. Briefly, RNA was extracted from tissue collected 2 dpi using the GeneMATRIX Universal RNA Purification Kit with RL buffer (CHIMERx). RNA was DNase-treated to remove residual genomic DNA using RQ1 DNase (Promega), then normalized by concentration and equal amounts of RNA were used to synthesize cDNA using iScript cDNA Synthesis Kit and the provided oligo(dT) primers (BioRad). The cDNA synthesis reaction was run for twice the time recommended by the kit (1.5 hours instead of 30-60 minutes) since this has resulted in increased cDNA synthesis in the past. The iScript reverse transcriptase enzyme has RNaseH activity, so it degrades an RNA template as soon as it is used to make a cDNA molecule, making it a useful reverse transcriptase for semi-quantitative RT-PCR. Amplifying the region outside of the *At*P5SM insertion site shows the *EGFP-AtP5SM* splice products. RT-PCR products were run on 1-2% agarose gels.

### Protein Extractions and Western blots

See **Chapter 2** methods. Briefly: crude protein was extracted, equal volumes were run on a denaturing polyacrylamide gel, protein was transferred to nitrocellulose membrane. The membrane blocked with TBS + 5% powdered milk + 0.05% Tween-20 overnight at 4°C. The next day, the membrane was equilibrated to room temperature for 2 hours, blotted with anti-HA HRP-conjugated (3F10) used at 1:1000 dilution in TBS + 0.05% Tween-20 for 1 hour, then washed 4 times with TBST buffer. Chemiluminescence from the horseradish peroxidase enzyme was assessed using SuperSignal West Pico Chemiluminescent Substrate (ThermoFisher Scientific) and film paper developed in a dark room.

# CHAPTER 5:

# Further Engineering of the HyP5SM system


## Abstract

This chapter is a compilation of different projects to advance the utility of P5SM-based splicing regulation. In **Section 5.1**, we show that AtP5SM is not functional immediately after two codons (the start codon and a glutamate codon to supply the required AG dinucleotide) in a vector without a 5' UTR. In **Section 5.2**, we show that the outer splice sites of the *EGFP-HyP5SM* cassette are recognized in HEK293T cells, resulting in almost all exon-skipped splice product. We modify the L2 loop with known human splice factor binding sequences in attempts to recover exon recognition. The goal is to engineer a suicide exon for gene regulation in mammalian cells. We did not succeed here, but suggestions for future designs are provided based on the RT-PCR results.


## 5.1: Far Upstream Sequence Insertions of P5SM

### Challenges For Implementation of Splicing-Based Regulation

Cloning and expression vectors have multi-cloning sites for inserting a gene of interest, and no location is expected to be better than another. For restriction digestion and ligation-based cloning, the researcher only typically needs to verify that the restriction enzymes selected will not cut their gene of interest internally. For recombination-based cloning, internal cut sites are not generally a concern. Consequently, if a researcher wants to clone 100 genes into an inducible vector, the genes can likely be cloned into the same multi-cloning site. This simple insertion site decision is a key biotechnological benefit of promoter-based regulation.

In contrast, the cloning methods for splicing-based regulation have thus far required the researcher to select an insertion site in the target gene. This is true not only for P5SM cassettes, but also for methods that artificially alter splicing by cloning an aptamer element next to splice sites (110, 111). This internal sequence insertion site is an additional choice that researchers must make. The sequencing requirement of P5SM are minimal: it must be cloned after an AG dinucleotide, and the 3' splice site exon region should not be the worst match to the consensus plant 3' splice site (CC following the intron), but it also is not required to be a good match (12, 120, 179), see **Chapter 2** for examples. However, despite the forgiving quality of P5SM likely due to strong internal splicing enhancers, the fact remains that not all insertions of P5SM into a gene of interest are guaranteed to function. This is a major drawback of splicing-based

regulation. This uncertainly may cause researchers to hesitate about using P5SM cassettes for their own projects.

To address these concerns, we chose to investigate the possibility of a very far 5' insertion of P5SM. Essentially, we wanted to see if we could turn P5SM into an N-terminal tag that could be cloned into any currently-existing binary vector for plant expression. This approach would allow for an almost "traceless" N-terminal tag, a P5SM cassette which acts as a suicide exon unless P5SM skipping is induced by co-expressing L5 ribosomal protein. The resulting protein product of interest would either not be tagged, or would have a maximum of two amino acids. We used *Arabidopsis thaliana* P5SM (*At*P5SM) because it is the model previously used to determine the P5SM insertion site requirements (120). We expect results to be applicable to Hybrid P5SM (HyP5SM) as well since both cassettes contain the same *A. thaliana* 5' and 3' introns (120).

## Results

We designed an *Enhanced Green Fluorescence Protein* (*EGFP*) reporter constructs with the minimum sequence required for *At*P5SM insertion (**Figure 5.1.1a**). Conceptually, we need a methionine codon (start codon) followed by a glutamate, lysine, or glutamine codon (GAG, AAG, or CAG, respectively) supplying the required AG dinucleotide 5' splice site sequence before the intron. The 3' splice site sequence can be variable. However, to eliminate another choice that researchers would have to make, we considered a "MER" tag. The initial ME codons (AUGGAG) are expected to have a positive effect on translation initiation. The G base in the +4 position (where the A of the AUG methionine codon is +1) makes the methionine a stronger start codon because it conforms to Kozak consensus sequence for plants (212). The arginine codon was chosen because the *At*P5SM cassette in its endogenous *Transcription Factor IIIA* (*TFIIIA*) gene is flanked by GAG/AGA (E/R) codons (112).

Fortuitously, the second amino acid of EGFP is naturally a valine, and the GTN valine codons match the consensus 3' sequence after introns in plants (12). This allowed us to try a "ME" tag as a first attempt, which would produce EGFP protein with a single glutamate addition after its start codon (**Figure 5.1.1b**). We also tested a "MEM" tag, where the *At*P5SM cassette is flanked by GAG/AUG (E/M) codons preceding the *EGFP* reporter. The "MEM" tag allows us to test for potential translation reinitiation after the *At*P5SM cassette.

*N. benthamiana* leaves were transiently co-infiltrated with the *EGFP* reporter, *DsRed2* (infiltration control), and either *firefly luciferase* (LUC, does not affect splicing) or *AtL5* (promotes *At*P5SM exon skipping). The "ME" and "MEM" tagged *EGFP* constructs were tested along with the original *EGFP-AtP5SM* construct which lacks an N-terminal tag. Fluorescence scans show induced protein expression only for the original construct (**Figure 5.1.2**). Neither the "ME" nor the "MEM" tagged *EGFP* reporters showed a difference between the LUC and *At*L5 leaf halves for any biological replicate (n = 4, **Figure 5.1.2**).

### Discussion and Future Directions

In retrospect, this negative result is not surprising. The constructs were cloned into the regular KpnI/SalI restriction sites of the pBinAR vector used previously (112). Transcription is driven by a constitutive 35S promoter very close to the KpnI site. It is likely that the ME and MEM tags failed because there was simply not enough upstream sequence to be recognized as a first exon in front of the *At*P5SM cassette. This was a design flaw on our part, but the solution is clear: we need more upstream sequence.

These constructs had another flaw which we did not consider upon initial design. Recombinant proteins with a glutamate as the second amino acid have been observed to have reduced stability and a shorter half-lives in cells due to what is called the "N-end rule pathway" (213–215). The ubiquitin-proteosome pathway targets proteins for degradation at different rates, and one variable that affects protein turnover is the second amino acid of a peptide (214). Glutamate is considered a secondary destabilizing amino acid and it has been shown to result in increased proteasome-targeting of proteins when used as an N-terminal tag in *A. thaliana* (215). We did not have a *ME-EGFP* control without *At*P5SM cassette to see if the loss of green fluorescence could be explained by this E-tag effect. However, the N-end rule pathway is expected to reduce green fluorescence, but not eliminate it all together (215). We think the mis-splicing hypothesis is a more likely reason for loss of functional EGFP.

In the future, we recommend testing P5SM cassettes cloned into 5'UTR sequences. A literature search did not yield obvious candidates for model 5'UTRs in plants, so this would require screening various 5'UTRs for one that spliced correctly and expressed well in leaf tissue for transient transformations, at least. If an N-terminal peptide tag is desired instead of a 5'UTR, care should be taken to choose a sufficiently long sequence for proper exon recognition. The smallest exon known in plants is the potato invertase mini-exon 2 (9 nt), but exons this short are very rare (216). Since P5SM originally comes from the *TFIIIA* gene, its 5'UTR may be a fitting first candidate.

## 5.2: Re-Engineering HyP5SM for Splicing Regulation in Mammalian Cells

### Introduction

After the success of using Hybrid Plant 5S rRNA Mimic (HyP5SM) exon to regulate severe plant immune phenotypes (**Chapters 2 and 3**), we wondered if such a strategy would be applicable in mammalian cells. To our knowledge, there are no engineered HyP5SM-like exons for gene regulation in mammalian cells. However, just as in plants, mammalian cells also naturally couple alternative splicing and nonsense mediated decay to regulate gene expression using endogenous "suicide exons" or "poison exons" (90). This strategy is called Regulated Unproductive Splicing and Translation (RUST). Although many transcripts are known to have splice products that are non-productive due to nonsense-mediated decay (90, 91, 116), the obstacles to

implementing any such suicide exons as tools for genetically-encoded artificial or orthogonal gene regulation are not trivial.

First, the alternatively spliced sequence must be modular, i.e. able to be moved as one piece into new sequence contexts and still maintain its function. For this property, individual introns or cassette exons (with their flanking introns included) are ideal candidates. The success of *At*P5SM and HyP5SM owes a lot to its minimal sequence requirements; it can be cloned after an AG dinucleotide in a gene of interest and will almost surely work (120, 179). We have only found two exceptions where this did not work: once, when the 3' sequence context was as far away as possible from the consensus 3' splice site (179) albeit even this poor context has worked for other genes (120), and then in the ME-tag case described in **Section 5.1** where we hypothesize there was not enough upstream sequence for a first exon before the P5SM 5' intron. Except for these two extreme cases, *At*P5SM and HyP5SM have proven to be very modular cassette exons.

Second, the splice sites must be recognized strongly enough in the new context that only the inserted sequence is subject to splicing. If inserting a foreign sequence results in splicing upstream or downstream such that the gene of interest is internally truncated, that foreign sequence is an unreliable and inconvenient method of gene regulation. Sequences that splice using alternative 5' or 3' splice sites may be more context-dependent and are not likely to be modular enough.

Third, the default splice ratio should be shifted far toward the non-productive (NMD-targeted) splice product. The more the non-productive splice product is favored, the more stringent the regulation of the gene of interest will be. The modifications from *Os*P5SM to HyP5SM show that a spliced element may be rationally designed to strongly improve preference for the non-productive splicing route, even if the starting cassette exon favors the productive splice product (120).

Finally, there should be an easy and reliable researcher-manipulatable way of switching splice product ratios from the non-productive to the productive splice product, such that gene expression can be intentionally induced. This last point, in particular, is difficult to achieve. The splicing fields is replete with transcriptional profiling papers describing changes in splicing due to gradual and not naturally binary environmental factors such as drought, heat stress, starvation, and so forth (45, 217, 218). This makes sense from a biological systems viewpoint; endogenously, RUST-regulated genes are likely responding to dynamic and small changes inside cells, and subtle expression changes from many sources can lead to significant phenotype alterations. P5SM is also endogenously part of a feedback loop that responds to gradual increases in unbound L5 ribosomal protein (112). For lab-controlled gene regulation purposes, we can co-express or induce L5 ribosomal protein to high levels to make a pseudo-binary P5SM response. With *At*P5SM, the induced gene expression is up to 7-fold when constitutive *At*L5 is co-expression (120). With HyP5SM, induced expression is almost 100-fold when constitutive *Os*L5 is co-expressed (120). The order of magnitude difference in induction levels between *At*P5SM- and HyP5SM-regulated genes in *Nicotiana benthamiana* is because HyP5SM was rationally designed to be a more binary splicing system.

## Results and Discussion

*EGFP-HyP5SM splices to SP-I in HEK293T cells*

Since HyP5SM in plants has all the desired characteristics of an alternatively spliced sequence suitable for stringent gene regulation, we wanted to see if it would be feasible to use it as a starting point for an engineered splicing cassette for mammalian cells. To determine if the plant introns would be properly recognized by mammalian cells, pcDNA3.1(+) *EGFP-HyP5SM* was transfected into HEK293T cells using lipofectamine. Mammalian and plant introns have slightly different consensus sequences for the splice sites and especially for the branch site, but cross-kingdom splicing fidelity has been reported previously (12). RNA was extracted and used for RT-PCR to look at *EGFP-HyP5SM* splice products. The agarose gel showed a single specific band at the expected size for exon-skipped SP-I, and sequencing confirmed that the plant splice sites were correctly identified and used by HEK293T cells (**Figure 5.2.1**). This result demonstrates that the productive splice product can be recovered from HEK293T cells expressing *EGFP-HyP5SM*. We proceeded to investigate what might be affecting HyP5SM splicing in HEK293T cells.

*Human L5 ribosomal protein does not promote HyP5SM exon skipping in plants*

HyP5SM is correctly spliced out, but does not act as a suicide exon in mammalian cells. Although SP-I is undesirable as the major splice product, previous engineering from *At*P5SM to *Os*P5SM to HyP5SM has shown that SP-I levels may be reduced. In *N. benthamiana* leaves, *At*P5SM shows moderate levels of SP-I by RT-PCR, even without co-expression of *At*L5 protein (120). This background level of SP-I is hypothesized to be due to endogenous *N. benthamiana* L5 ribosomal protein interacting with *At*P5SM, since both sequences come from dicot plants. To overcome this issue, a hybrid sequence was engineered using portions of the *At*P5SM cassette and the more distantly related *Oryza sativa* (rice) P5SM. The resulting HyP5SM responds only to monocot *Os*L5 and shows reduced background SP-I in *N. benthamiana,* supporting the hypothesis that endogenous L5 ribosomal protein was an issue with *At*P5SM (120). Since HyP5SM appears to splice completely to SP-I in HEK293T, we hypothesize that this is due to loss of an exon-defining splice product present in plants, but not in human cells (**Figure 5.2.1**). An alternative hypothesis is that human L5 ribosomal protein interacts with HyP5SM and actively promotes SP-I. In order to rule out this possibility, we tested to see if *Hs*L5 would promote expression from *EGFP-HyP5SM* in our usual plant system. We graciously obtained the coding sequence for *Hs*L5 from Kathleen Collin's group, sequence verified it, and cloned it into the binary expression vector, pBinAR, for *Agrobacterium*-mediated infiltration and constitutive expression in plants. *N. benthamiana* leaves were co-infiltrated with pBinAR *EGFP-HyP5SM* (splicing reporter), pBinAR *DsRed2* (infiltration control), and either pBinAR *firefly luciferase* (*LUC*, does not affect splicing) or pBinAR *Os*L5 (promotes SP-I) or pBinAR *Hs*L5. Plants were left under constant light for 3 days post-infiltration (dpi) to accumulate maximum protein levels as per the usual protocol (112). Leaves were scanned for EGFP fluorescence (ex/em: 488/520 nm) and DsRed2 fluorescence (ex/em: 532/580 nm). Leaf halves with *Hs*L5

showed fluorescence similar to that seen in the leaf halves with *LUC* (**Figure 5.2.2**). Human L5 ribosomal protein does not appear to actively promote HyP5SM exon skipping or gene expression from HyP5SM-regulated genes. We proceeded with the hypothesis that HEK293T lacks the necessary splice factor for exon definition.

*Co-expression of RSZ32 does not promote HyP5SM exon retention in HEK293T*

The members of the plant-specific RS2Z sub-family of serine/arginine-rich (SR) proteins, RSZ32 and RSZ33, are candidates for the putative splice factor that promotes *At*P5SM and HyP5SM exon retention in dicot plants (**Chapter 4**). We hypothesized that they might be the missing factors in HEK293T cells. The inner splice sites of the HyP5SM cassette, the 5'intron/exon and exon/3'intron junctions, might not be recognized strongly enough in HEK293T cells for HyP5SM exon definition. Exogenous SR proteins that promote exon definition in their native organisms have previously been found to promote splicing of weak introns in unicellular organisms lacking alternative splicing (20). With this in mind, pcDNA3.1(+) *RSZ32 coding sequence (CDS)* was co-transfected with pcDNA3.1(+) *EGFP-HyP5SM* in HEK293T cells. After a day, transfected cells were examined under a microscope for EGFP expression and then collected for RNA extraction and RT-PCR analysis. Cells expressed EGFP fluorescence and no difference was observed (**Figure 5.2.3c**). After extracting RNA and verifying splicing by RT-PCR, a faint band at the expected size for SP-II was observed. This was unexpected since a previous pilot experiment had not shown any SP-II (**Figure 5.2.1**). We attempted to extract the band from the gel for sequence verification, but it was too faint to recover.

The amplified bands in the RT-PCR did not show a clear difference in splice ratios or the predicted SP-II product with RSZ32 as compared to empty vector (**Figure 5.2.3ab**), supporting the observations with the fluorescence microscope. This result can be explained various ways: RSZ32 contains a domain architecture not found in mammalian splice factors and the protein may not be stable outside of plant cells. The Barta lab which discovered RSZ32 and RSZ33 has reported difficulties expressing full-length recombinant RSZ33 in *E.coli* and yeast (189), and presumably stability could be an issue for RSZ32 in HEK293T cells as well. RSZ32 and RSZ33 share 71% identity and likely have related functions and interactions. RSZ33 is known to interact with other plant-specific splice factors such as the SCL sub-family of SR proteins (189), and none of these proteins are available in HEK293T cells either. We decided not to attempt to reconstitute the RSZ32 or RSZ33 protein complexes in human cells. Exogenous expression of foreign splice factors would surely have many pleiotropic effects, so we sought a different solution to recover HyP5SM exon definition.

*Selection of SR protein-binding sequences for rational design of mammalian HyP5SM*

Based on both our model of HyP5SM exon retention (**Figure 5.2.1**) and the diverse sequence and length of the L2 loop in P5SM structures between plants (112), it is worth considering that the factor(s) that promote P5SM exon retention may vary between monocots and dicots and other plants. The P2 stem structure below the L2 loop has nucleotides conserved with >97% identity among plants (112). The P2 stem structure is required for interaction of L5 ribosomal protein with both P5SM RNA and

eukaryotic 5S rRNA, but the L2 loop is not required for L5 ribosomal protein binding (112, 193). Monocot *Os*P5SM lacks the purine-rich sequence required for exon retention of dicot *At*P5SM, yet it still appears to function similarly as a suicide exon. If the splice factors responsible for exon definition of the otherwise highly conserved P5SM structure may vary and bind diverse L2 loops, then we propose that altering the L2 loop to contain motifs known to bind mammalian splice factors may be a way to recover SP-II as the primary splice product in HEK293T cells. The rational design of HyP5SM is precedent suggesting that this strategy may work.

We began with targeted bioinformatics and database searches to compile information about individual mammalian splice factors. We narrowed our list of target mammalian SR proteins to those highly associated with alternative splicing and splice site selection, and discarded SR proteins best known for other regulatory functions such as transcription regulation. We searched the scientific literature for RNA motifs empirically found to bind these mammalian splice factors through such techniques as Systematic Evolution of Ligands by Exponential Enrichment (SELEX) and UV-crosslinking experiments (**Table 5.2.1**). We proceeded to search the literature for examples of motifs that were tested *in vitro* using splicing-competent cell extracts and that were found to promote splice factor recruitment, intron excision, and other changes in splicing. Motifs verified *in vivo* were highly desired, but not found. Notably, previous researchers have observed that the consensus sequence itself is sometimes not the best sequence to recruit SR proteins (55). This may be due to an averaging effect of the consensus that masks important nucleotide pairings present in various individual high binding affinity sequences. For consensus sequences listed in **Table 5.2.1**, we looked for the best “hits” in the original source articles and selected those sequences when available. In cases where top hits were very similar to each other, we chose a top hit and then another hit that was less similar. We did not know which loop modification, if any, might work best in our P5SM structure context, so we wanted some sequence diversity in our initial experiments. **Table 5.2.2** shows the selected motifs.

**TABLE 5.2.1:** *Binding Motifs for Selected Mammalian Splice Factors.*

| SR protein | Binding | Method | Ref |
|---|---|---|---|
| **ASF/SF2** | RGAAGAAC | SELEX | (219) |
| | AGGACRRAGC | SELEX | (219) |
| | SRSASGA | Functional SELEX | (220) |
| | UGRWG | CLIP | (221) |
| **SC35** | AGSAGAGUA | SELEX | (219) |
| | GUUCGAGUA | SELEX | (219) |
| | GRYYcSYR | Functional SELEX | (55) |
| | UGUUCSAGWU | SELEX | (222) |
| | GWUWCCUGCUA | SELEX | (222) |
| | GGGUAUGCUG | SELEX | (222) |
| | GAGCAGUAGKS | SELEX | (222) |
| | AGGAGAU | SELEX | (222) |
| | UGCNGYY | Functional SELEX | (208) |
| **9G8** | UCAACA | UV cross-linking | (223) |
| | ACGAGAGAY | SELEX | (222) |
| | GGACGACGAG | Functional SELEX | (208) |
| **SRp55/B52** | USCGKM | Functional SELEX | (220) |
| | UCAACCAGGCGAC | SELEX | (224) |
| **SRp40** | GAGCAGUCGGCUC | SELEX | (211) |
| | ACDGS | Functional SELEX | (220) |

**N**: any; **R**: purine (A/G); **Y**: pyrimidine (T/C); **D**: A, G, or U; **K**: U or G; **M**: A or C; **S**: G or C; **W**: A or U

*Table modified from Long and Caceres review (185).

*Selected sequences inserted into the L2 loop of OsP5SM*

Fortuitously, the L2 loop of *Os*P5SM was found to closely match one of the SELEX consensus sequences for ASF/SF2 protein: RGAAGAAC, where R = A or G (219). A sequence matching the consensus, AGA**A**GAAC, requires adding only one nucleotide (bolded and underlined) to the *Os*P5SM L2 loop. This observation implicates the ASF/SF2 sub-family of plant SR proteins as potential regulators of *Os*P5SM splicing in its endogenous organism, *Oryza sativa* (rice). We decided to add pcDNA3.1(+) *EGFP-OsP5SM* to our screen. The additional +1 nt insertion was also generated to match the RGAAGAAC consensus and called Hy-ASFv1; another mutant to match the SRSASGA consensus was called Hy-ASFv2 (**Table 5.2.1, Figure 5.2.4**). The selected sequences were cloned by performing site-specific mutagenesis PCR using *Os*P5SM as a template (same backbone as *Hy*P5SM).

To further target ASF/SF2, we followed the example of Take and Manley and used three copies of octamer AGAAGAAC ("A3" motif), which together have been shown to function as a splicing enhancer (219). This octamer binding site for ASF/SF2 was found using SELEX, then tested for function *in vitro* by placing 80 nt downstream of the second intron of α-globin (219). This intron is normally poorly spliced out due to weak splice sites, by the A3 motif enhanced splicing *in vitro* in an ASF/SF2-dependent manner. The A3 insertion adds +60 nt to the L2 loop of *Os*P5SM, now called HyP5SM-A3. This insertion is the largest of the L2 loop modifications, so we were especially concerned that it might affect the integrity of the P5SM RNA structure. Using the Mfold server to computationally predict RNA structure of the entire HyP5SM-A3 exon sequence, the highest scoring structures all predicted that the P2 stem which holds the L2 loop will still form (**Figure 5.2.5**). Significantly, the AA dinucleotide bulge in the P2 loop is preserved in the top five predicted structure models (**Figure 5.2.5**). An A(A|U) dinucleotide bulge is highly conserved across P5SM structures in plants (present as AA >90%, sometimes AU). The 5S rRNA structure also has a highly conserved dinucleotide bulge in the equivalent Helix III stem, present as CA in *A. thaliana* and as AA in *Xenopus laevis* (112, 225). This small bulge is essential for L5-5S rRNA and L5-P5SM binding and is thought to be the initial contact point between L5 ribosomal protein and 5S rRNA, and presumably this is true for P5SM as well (112, 225, 226). For inducible gene expression purposes in the future, it is desirably to maintain the ability for HyP5SM-A3 to interact with L5 ribosomal protein. Since the RNA structure prediction program estimates that the essential dinucleotide bulge and other important structural components of P5SM will remain intact, the three octamers of A3 and spacer regions (+60 nt total) were cloned into the L2 loop of *Os*P5SM using extension PCR.

In order to target SC35 protein, the Krainer group's D2 motif was cloned into the *Os*P5SM L2 loop (+20 nt) to generate HyP5SM-D2 (**Figure 5.2.4**). The D2 motif is a specific exonic splicing enhancer (ESE) sequence, one of the highest scoring sequences in the functional SELEX experiment that produced the GRYYcSYR octamer consensus for SC35 protein binding (55). It is notable because it was tested in two different contexts and found to promote exon skipping *in vitro* for both, making it more likely that it may function for HyP5SM-D2 as well. Although not as large as the A3 sequence, the D2 motif expands the L2 loop by a considerable 20 nt. Mfold analysis to predict RNA structure for HyP5SM-D2 estimates that the P2 stem will form properly in this HyP5SM variant as well (not shown).

Finally, to target the 9G8 protein, a zinc knuckle-containing SR protein like RSZp22 in plants (shown to alter HyP5SM-regulated splicing in **Chapter 4**), we utilized the 6-18 and 6-24 motifs identified through SELEX by Schaal and Maniatis (208). These motifs add 18 nt and 19 nt, respectively, and Mfold again predicts that neither will affect the integrity of the P2 stem (not shown). Both motifs were tested functionally by cloning them downstream of the model splicing reporter *Drosophila dsx*, a substrate with weak splice sites that requires an enhancer to splice. The 6-18 sequence required 9G8 protein to splice, whereas both 9G8 and SC35 could promote splicing with the 6-24 sequence (208).

**TABLE 5.2.2: Modified *Os*P5SM L2 loop sequences.**

| Name | Motif | L2 loop | Target Splice Factor(s) | Δnt | Ref |
|---|---|---|---|---|---|
| HyP5SM [top]<br>*Os*P5SM1 [bottom] | Mutated to resemble *At*P5SM loop | …UCCCAaagaGAGAACU…<br>…UCCCA----GAGAACU… | ? | +4 | (120) |
| *Os*P5SM1 | - | …UCCCAGAGAACU… | Unsure, but similar to ASF/SF2 motif | +0 | (120) |
| Hy-ASFv1 | RGAAGAAC | …UCCC**AGAaGAAC**U… | **ASF/SF2** (SELEX) | +1 | (219) |
| Hy-ASFv2 | SRSASGA | …UCC**gAGAcGA**ACU… | **ASF/SF2** (SELEX) | +1 | (220) |
| HyP5SM-A3 | A3 motif (AGAAGAAC octamer and spacers) | …UCCCAGA au **ucu cgac agaagaac ucu cgac agaagaac ucu cgac agaagaac ucgagucuaga** GAACU… | **ASF/SF2** (Functional) | +60 | (219) |
| HyP5SM-D2 | D2 motif (GRYYcSYR consensus) | …**UCCCA ggggacauacucggccgcag GAGAAC**U… | **SC35** (Functional) | +20 | (55) |
| Hy-6-24 | 6-24 motif | …**UCCCA uuugcggucuccggccucc--- C**U… | **SC35 and 9G8** (Functional) | +19 | (208) |
| Hy-6-18 | 6-18 motif | …U**CCgAcgccauggacgac GAG**AACU… | **9G8** (Functional) | +18 | (208) |

*Loop sequence lengths are 8 to 60 nt. Red text is change from *Os*P5SM. Bold and underlined text is region that matches the motif or previously tested sequence.

*Splicing from new HyP5SM variants shows some exon retention*

HEK293T cells were transfected with constitutive expression vector pcDNA3.1(+) carrying the indicated HyP5SM variants. After incubating overnight, cells were collected for RNA extract, DNase treatment, cDNA synthesis (with oligo dT primers) and RT-PCR. An initial RT-PCR with primers annealing about 100 bp outside the splice sites (such that SP-I is 200 bp) showed no noticeable SP-II amplification (**Figure 5.2.6a**). However, repeating the RT-PCR with SP-II specific primers showed SP-II product for most of the HyP5SM variants (**Figure 5.2.6bc**). With primers specific for SP-II, we were able to visualize some SP-II product even from the original *EGFP-HyP5SM*. The best amplification of SP-II came from *EGFP-HyP5SM-A3*, the one variant with three copies of a consensus sequence. The bands from *EGFP-HyP5SM-D2* are also very good, as expected from a top result from a functional SELEX screen. The 6-24 bands appear weaker. Overall, these are good results that indicate that the splice ratios may be altered using SR protein binding sequences inserted into the L2 loop. Although none of the tested samples indicates that we successfully engineered a suicide exon — as would be indicated by a major splice product of SP-II – the results are not completely

negative. This is promising for a first attempt and suggests that HyP5SM may yet be able to be engineered for use in mammalian cells. Notably, the RT-PCR using primers to amplify all splice products shows a specific band around 500 bp which could correspond to splice products retaining the 5' intron and exon (478 bp for HyP5SM) or the exon and 3' intron (530 bp for HyP5SM) (**Figure 5.2.6c**). The presence of this band may indicate that one of the inner splice site junctions between the introns and the HyP5SM exon variants is not efficiently targeted in HEK293T cells. These intron-retaining splice products are visible in plant samples as well (see **Chapter 2**), but they are never stronger than SP-I or SP-II in properly splicing samples.

*Intron strength analysis*

The *A. thaliana*-derived introns in *EGFP-HyP5SM* were scanned for predicted human branch sites using the SVM-BPfinder tool for mammalian U2 branch point prediction, set to *Homo sapien*. The server is hosted by Universitat Pompeu Febra, Barcelona at:

< http://regulatorygenomics.upf.edu/Software/SVM_BP/>

Both the 3' and 5' introns had at least three segments matching the human branch site consensus sequence >75% (see bolded black text in **Figure 5.2.7**). The sequence of the introns was also scanned manually for pyrimidine-rich (Y-rich) segments which are a characteristic of mammalian intron branch sites, differing slightly from the plant preference for AU-rich segments in introns (12). The entire 5' intron (150 bp) contains 57 U (38%) and 21 C (24%), together 52% Y-rich. The entire 3' intron (98 bp) is composed of 48 U (48.98%) and 16 C (16.33%), together 65.31% Y-rich. Additionally, the 3' intron had longer clusters of Y-rich segments, reaching 79% in some cases. The 5' intron had shorter Y-rich segments spaced around the intron. By this analysis, the 3' intron is predicted to be a stronger intron for mammalian cells. The 5' intron is relatively lacking in pyrimidine content and lengthy Y-rich segments. Based on this information, we predict that the potential intron-retaining product present in the RT-PCRs (**Figure 5.2.6**) is a splice product retaining both the 5' intron and the HyP5SM exon variants. This splice product is expected to be non-functional because the coding sequence would still be interrupted by the 3' intron, resulting in premature termination codons and mRNA targeted by nonsense-mediated decay.

**Future directions**

Based on the RT-PCR with primers specific for SP-II, it appears that the A3 motif worked best from all the HyP5SM variants (**Figure 5.2.6b**). When designing the first batch of variants, we were concerned about adding large sequences to the L2 loop. We did not want to disrupt the structured P5SM RNA element. However, it appears from Mfold predictions and the RT-PCR that this concern was not warranted. In the future, we recommend that the next batch of HyP5SM variants for testing in mammalian cells have duplicate or triplicate consensus sequences separated by spacers as done for the A3 motif (**Figure 5.2.5**). The longer length of the L2 loop ("loop" used here casually for lack of empirical structure information) may be beneficial specifically because it has more consensus binding sequence copies or, in the case of HyP5SM-D2 which only had one consensus sequence, the longer (>8 nt) consensus binding sequences may improve affinity.

In the future, splicing to the SP-II product may be improved by modifying or replacing the *A. thaliana* 5' intron which is a component of all HyP5SM variant cassettes tested here, including the original HyP5SM and *Os*P5SM cassettes. The plant 5' intron is enriched for uracils, but not for cytosines, reducing its total pyrimidine content. This can be remedied by utilizing a 5' mammalian intron instead.

Finally, it is unclear how much the lack of apparent SP-II in RT-PCR gels is due to production rates (no enough splicing to SP-II) or degradation rates (nonsense-mediated decay). Plants are predicted to have less efficient nonsense-mediated decay targeting than mammalian cells and intron-retaining splice products are more common in RT-PCRs with plant samples (197). In order to better visualize the transcribed splice ratios and assess which HyP5SM variant may be resulting in more SP-II, we propose repeating the transfection experiments and RT-PCR under reduced nonsense-mediated decay conditions. This can be done in HeLa cells transfected to express shRNAs that target Upf1 protein, an essential protein in the nonsense-mediated decay complex. This Upf1 knockdown will make it easier to screen for constructs that splice to SP-II in the future. The goal is to design a construct that acts as a suicide exon by default, resulting in low background expression of the gene of interest, but which can be inducibly skipped to recover protein expression. In plants, this is achieved with the HyP5SM suicide exon and induced skipping of HyP5SM by co-expression with *Os*L5. In mammalian cells, HyP5SM has some promising initial results, but requires additional engineering to become a mammalian suicide exon. If this is achieved, we predict that expression can be recovered with *Os*L5 due to the strong binding affinity between the *Os*P5SM RNA backbone structure and *Os*L5 ribosomal protein. Based on previous binding research between L5 ribosomal protein and 5S rRNA (193), *Os*L5 is not expected to require additional co-factors in mammalian cells.

## 5.3: Figures

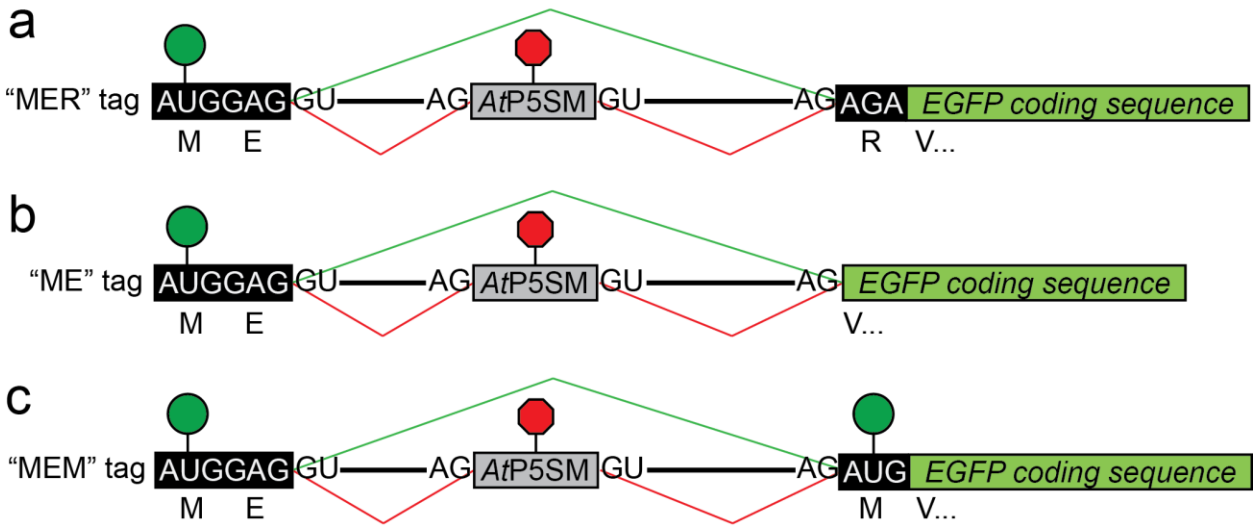


**FIGURE 5.1.1. Constructs to test far upstream insertion of *At*P5SM cassette.** These constructs were designed to test the possibility of making *At*P5SM act as an N-terminal tag. Green circles are start codons. Red octagons indicate the premature stop codon in the *At*P5SM exon. Splicing products decisions are noted by red lines (for SP-II) and green lines (for SP-I). (a) MER tag would be used for other genes, (b) ME tag for EGFP, (c) MEM tag. Tania Gonzalez cloned constructs for testing.

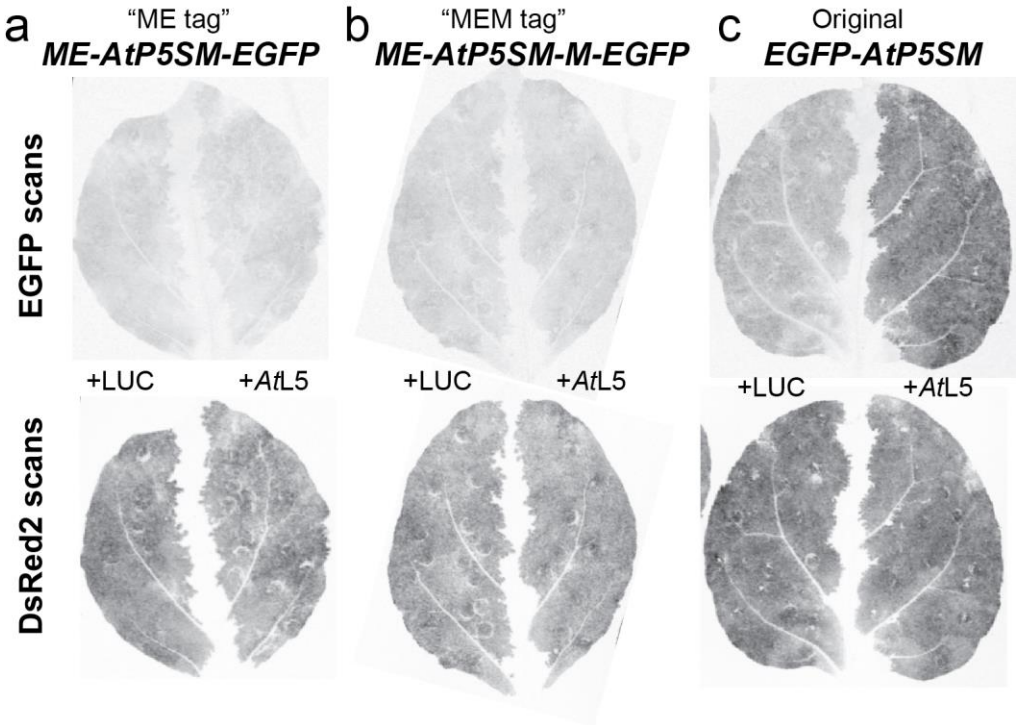


**FIGURE 5.1.2. Leaf scans show no EGFP expression from far upstream constructs.** *N. benthamiana* leaves scanned 3 dpi for EGFP and DsRed2 fluorescence. (a) *ME-EGFP-AtP5SM*, (b) *MEM-EGFP-AtP5SM*, (c) Original *EGFP-AtP5SM* as control. Malathy Sridhar performed experiment.

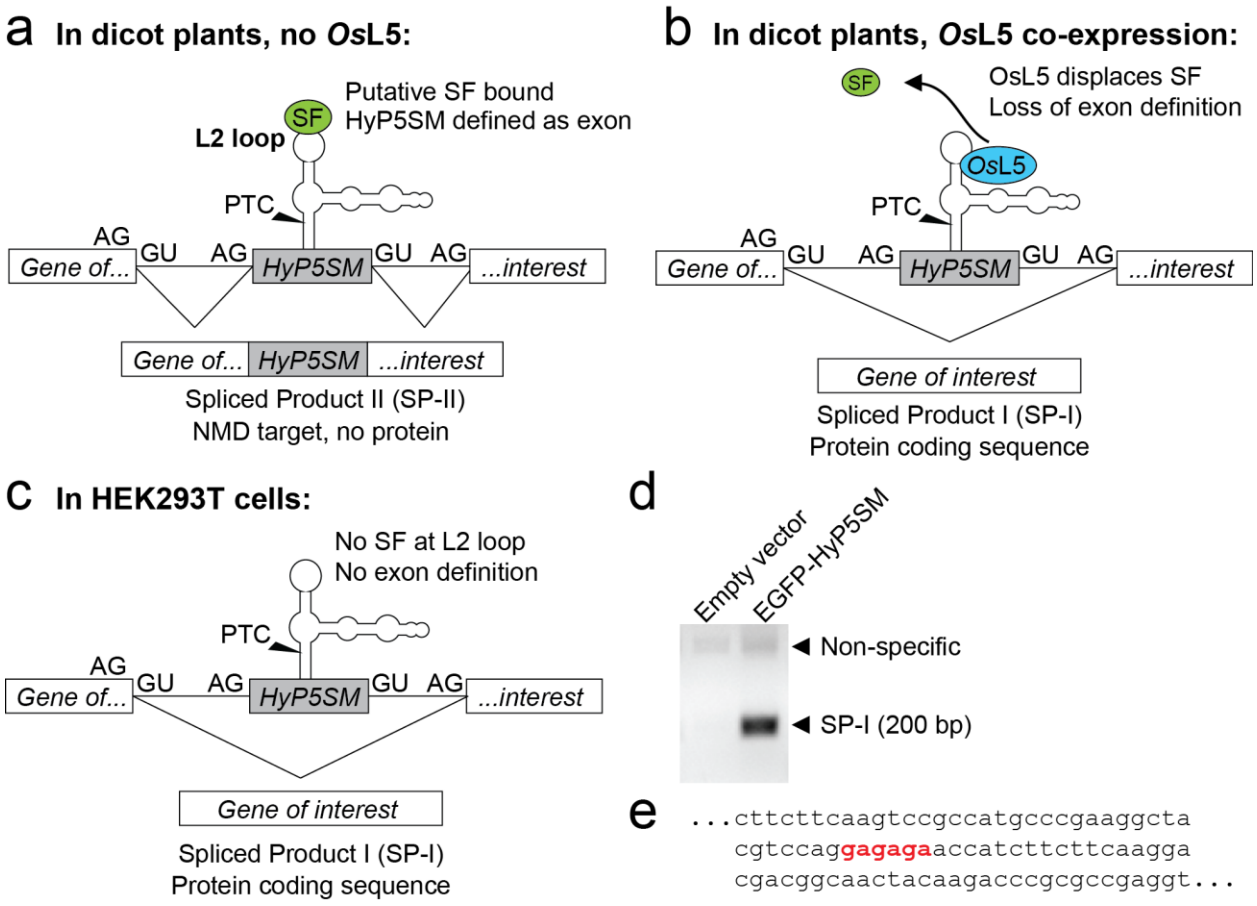


**FIGURE 5.2.1. HyP5SM exon definition and alternative splicing in plants and mammalian cells.** (a-c) Models for HyP5SM regulation and expected major splice products. (a) In dicot plants such as *N. benthamiana* and *A. thaliana*, SP-II is the major default splice product. (b) Upon co-expression of *Os*L5, the splicing shifts to favor SP-I, presumably through a displacement model. (c, d) In HEK293T cells, SP-I is the major splice product, presumably due to absence of the putative splice factor (SF) in plants. This model also explains why induction of strong transcription from HyP5SM-regulated genes in plants results in SP-I in the absence of *Os*L5 (see **Chapters 2**). (d) RT-PCR to see the splice products from HEK293T cells transfected with pcDNA3.1(+) *EGFP-HyP5SM* or empty vector. No *Os*L5 is present. Only one splice product was detected. Gel extraction and sequencing confirms that it is SP-I. (e) Sequence around the E/R codons (GAG/AGA) in *EGFP* where the HyP5SM cassette was inserted. The SP-I product joins the GAG and AGA codons together again, resulting in recovery of the *EGFP* coding sequence.

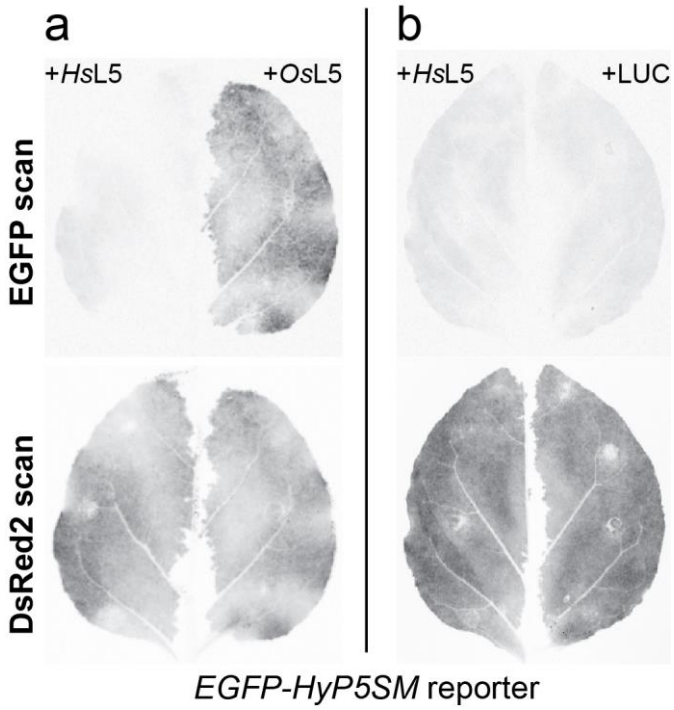


**FIGURE 5.2.2. Human L5 ribosomal protein does not promote expression from *EGFP-HyP5SM* in plants.** *N. benthamiana* leaves transiently transformed by *Agrobacterium*-mediated infiltration with pBinAR *EGFP-HyP5SM*, *DsRed2*, and either *firefly luciferase* (LUC) or a coding sequence of L5 ribosomal protein from rice (*Os*L5) or humans (*Hs*L5). (a) Comparison of *Hs*L5 versus positive control *Os*L5 on different leaf halves. (b) Comparison of *Hs*L5 versus negative control LUC on different leaf halves. Leaves were scanned together at 3 dpi for EGFP (488/520) and DsRed2 (532/580) fluorescence.

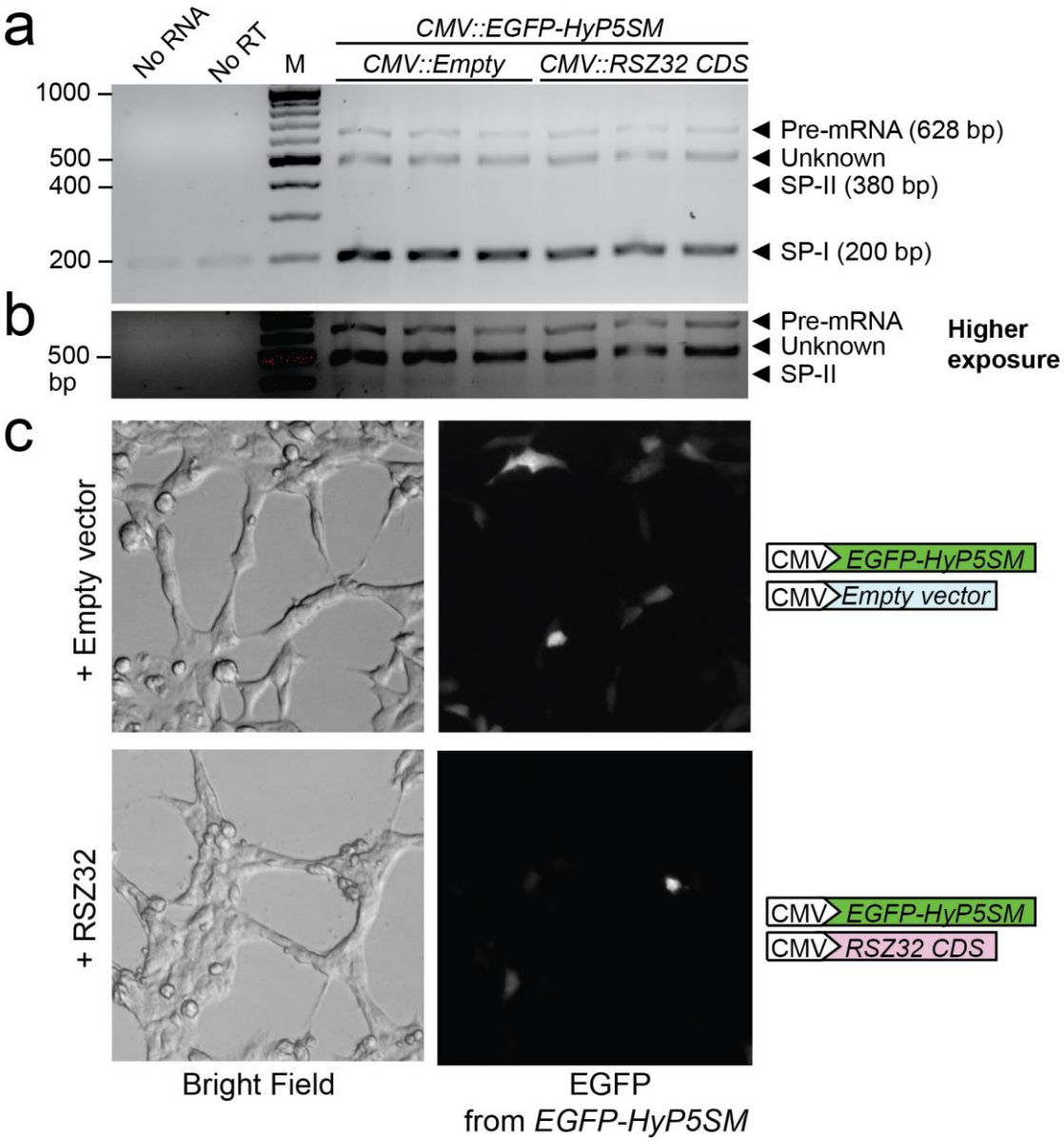


**FIGURE 5.2.3. RSZ32 has no effect on *EGFP-HyP5SM* in HEK293T, compared to empty vector.** (a) RT-PCR for *EGFP-HyP5SM* splice products. HEK293T cells were co-transfected with pcDNA3.1(+) *EGFP-HyP5SM* and either pcDNA3.1(+) *RSZ32 CDS* or pcDNA3.1(+) empty vector. (b) Same RT-PCR and higher exposure shows faint SP-II band indicating that there is some amount of exon retention in HEK293T cells. (c) Microscope image at 20X magnification for bright field and EGFP fluorescence. HEK293T cells were co-transfected as in (a).

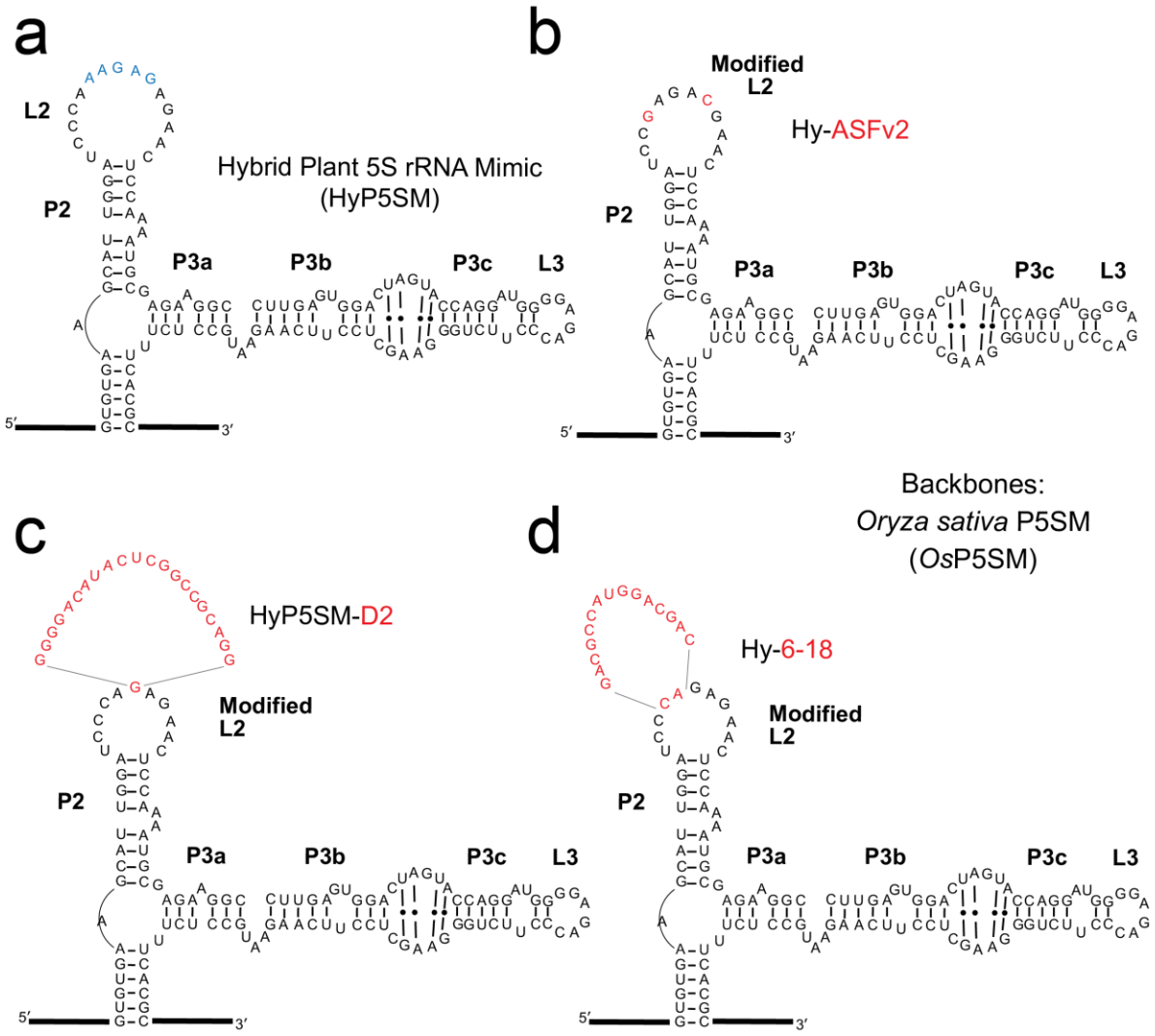


**FIGURE 5.2.4. L2 Loop Modifications of Various Sizes.** *Os*P5SM is used as a structural backbone for all modifications shown here. (a) The original Hybrid P5SM (HyP5SM) is a modification of the *Os*P5SM structure (+4 nt in blue). (b) The L2 loop of *Os*P5SM is already similar to some various consensus binding sequences for ASF/SF2 protein, and point mutations (red text) were sufficient to conform to the consensus. (c) HyP5SM-D2 has a considerably larger insertion (red text). The D2 motif sequence is predicted to bind SC35 protein. (d) Insertion of another functional sequence. The 6-18 motif is predicted to bind 9G8 protein, a zinc knuckle-containing SR protein.

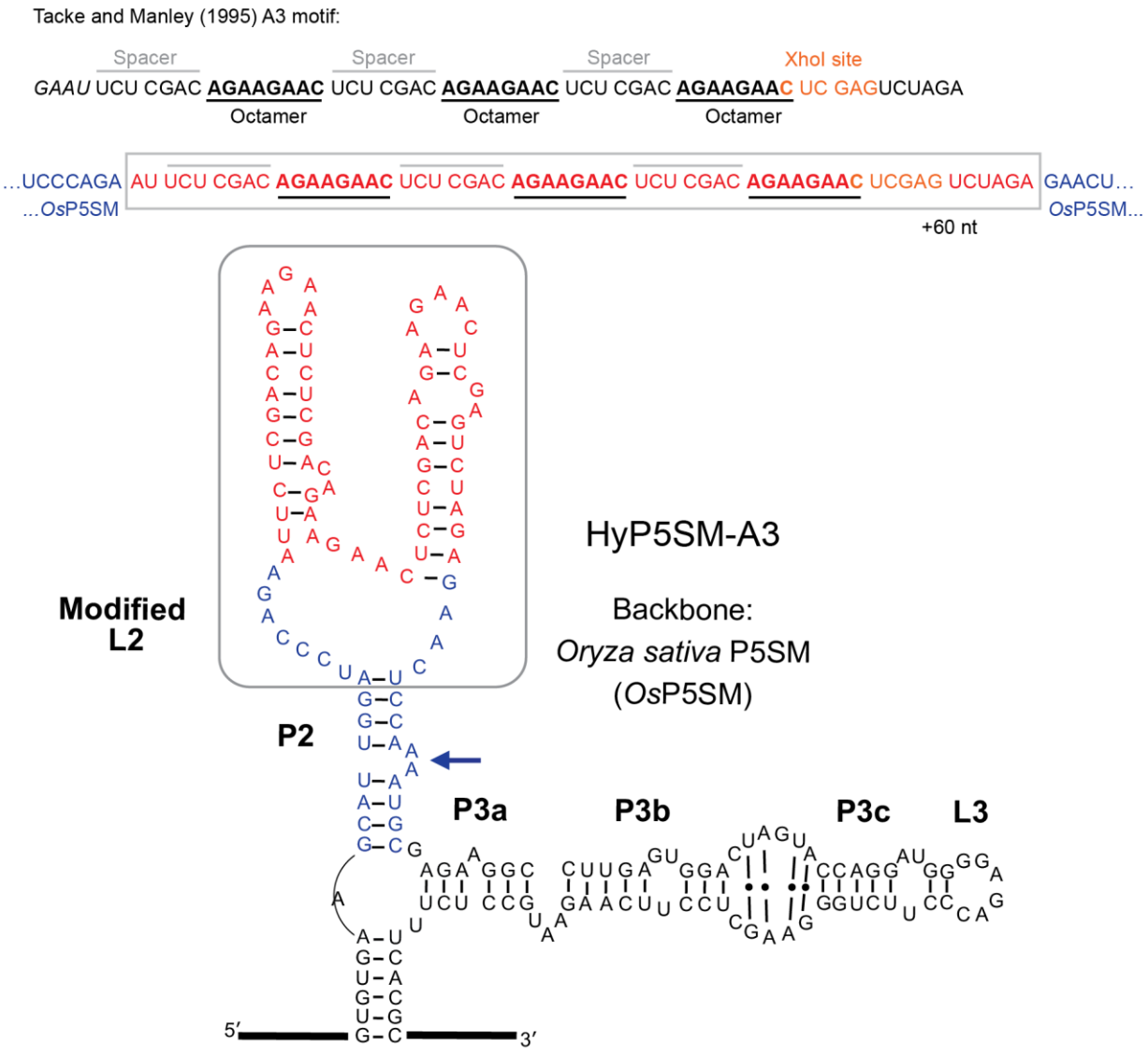


**FIGURE 5.2.5. Predicted HyP5SM-A3 structure preserves P2 stem and essential dinucleotide bulge.** The longest insertion into the L2 loop is the +60nt addition of Tacke and Manley's A3 sequence shown previously to function as a splicing enhancer targeted by AFS/SF2 protein (219). The sequence was strategically inserted into a location of the L2 loop that minimized the size of the insert (red text). The entire HyP5SM-A3 exon (including sequence flanking the structured region) was input into Mfold to analyze predicted structures. The top predicted structures all preserve the base of the original L2 loop and the P2 stem (blue text), including the critical AA dinucleotide bulge (blue arrow). Shown here is the P2 stem and modified L2 loop in the top predicted structure. The *Os*P5SM backbone structure is supported by in-line proving analysis in previous research and the figure is a modified version based on that structure (112).

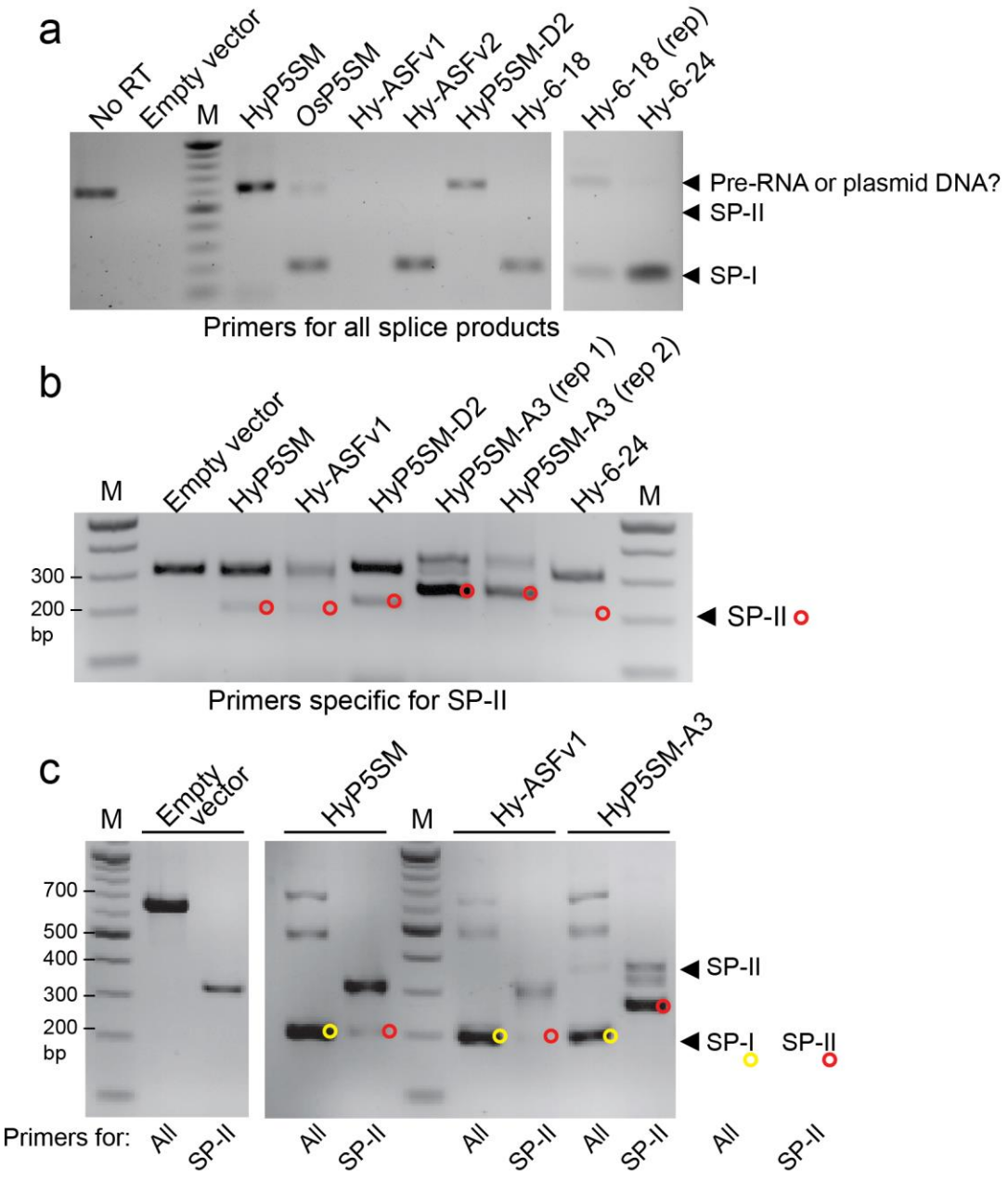


**FIGURE 5.2.6. RT-PCR shows improved exon retention.** HEK293T cells were transfected with pcDNA3.1(+) vector carrying the indicated *EGFP-HyP5SM-variant*. (a) RT-PCR using primers outside the splice sites primarily finds SP-I. Some samples appear weaker than others. This might be due to degradation issues in the samples, so bands should not be compared quantitatively. (b) RT-PCR with different biological replicates. Using SP-II specific primers successfully amplifies exon-retaining SP-II (red circle on right of band indicates expected size). For some samples, such as HyP5SM-A3 and HyP5SM-D2, the SP-II band is very apparent. For some genes, there are biological replicates (indicated with "rep"). (c) More samples, with both primer sets. Yellow circle on right side of bands indicates expected SP-I size. SP-II specific RT-PCR samples are repeated from (b). Note non-specific bands from empty vector.


**FIGURE 5.2.7. *Arabidopsis* intron branch site strength analysis.** The DNA amplified from the RT-PCR primer set for all splice products from *EGFP-HyP5SM (EGFP* in green text), 628 bp for pre-mRNA. Full sequence lines account for 100 bp. The HyP5SM 5' intron and 3' intron are 98 bp and 150 bp, respectively (black text). The HyP5SM exon (pink text) is 180 bp. Predicted human branch sites in introns are indicated in black bold text. Regions of enriched for pyrimidines (a characteristic of branch sites) are labeled as Y-rich, with % pyrimidine content and segment length noted. The 3' intron has more Y-rich segments than the 5' intron, and they are less interrupted.

## 5.4: Materials and Methods

### Site-Specific Mutagenesis for Modified HyP5SMs for Mammalian Expression

Primer pairs were designed according to the guidelines of the QuikChange manual (Invitrogen). PCR reactions of 60 µL volume were set up using *PfuTurbo* DNA Polymerase AD with 1X of the supplied stock 10X Cloned *Pfu* Reaction Buffer AD, plus the QuikSolution from the QuikChange Lightning Kit (Invitrogen). Reactions were set up in duplicate with 0% or 4.2% DMSO because the P5SM element is highly structured and was expected to cause problems. Primers were added in 0.3 µM final concentration. The reactions contained 100 ng of pcDNA3.1(+) *Os*P5SM as template. Reactions were found to work best with an extension time of 1 min/kb, rounded up, and a lower temperature PCR program: 95°C 2m initial melting, 18 cycles of (95°C 20s, 55°C 30s, 68°C 7m), 68°C 15m final extension, 4°C hold.

The segment with longer insertions (>4 bp) were cloned using extension PCR, with two primers adding the overhang and amplifying away from the L2 loop, then joined together in a 2-piece-ligation PCR. See **Chapter 2.6** for detailed advice for joining large pieces together by PCR.

### Cloning Vectors for Expression in HEK293T

The pcDNA3.1(+) vector used here drives gene transcription using the human cytomegalovirus immediate-early gene (CMV) promoter. *EGFP-HyP5SM* variants and *RSZ32 CDS* were cloned into the vector's multi-cloning site using restriction digestion and ligation. To ensure strong transcription, a Kozak consensus sequence for mammals was cloned upstream of the start codon for each construct. Ligation products were used to transform chemically competent TOP10 *E.coli*, selected on LB agar with 50 µg/mL carbenicillin. Positive colonies were cultured, miniprepped (with QIAGEN kit), and sent for sequencing to verify identity. Minipreps were eluted in sterile water for transfection into HEK293T cells.

### HEK293T Transfections

HEK293T cells were cultured in a UC Berkeley BSL-2 facility under sterile conditions. They were grown with Dulbecco's Modified Eagle Medium (DMEM) high glucose, supplemented with 10% fetal bovine serum (HyClone Fetal Bovine Serum from Thermo Fisher Scientific) and pyruvate (Gibco, CAT # 11995-065). The incubator was set to 37°C and with 5% $CO_2$ to maintain physiological pH. The cell media was pre-warmed to 37°C for 30 min before it was added to cells, to avoid temperature shock. Cells were maintained by replacing growth media and diluting concentration as needed. Every 2-3 days once ~90% confluency was reached: media was removed, cells were washed with 1X PBS buffer (, cells were treated with trypsin solution to detach (incubated in 37°C for 4-5 minutes), the trypsin solution was removed, and fresh growth media was added. Cells were split 1:6 and immediately plated onto clean petri dishes after contact with fresh media (because they begin to attach again).

A day before transfections, cells were split again 1:5. On the day of transfection, cells had reached ~65-70% confluency. Cell viability was accessed using a small aliquot

of freshly trypsinized and resuspended cells (100 µL) stained with Trypan blue dye (100 µL) and counted under a Haemocytometer. A minimum of 2 quadrants were counted, where 1 live cell (not blue stained) per quadrant represents 1 x $10^4$ cells/mL in the stock. Since cells were diluted 1:2 with Trypan blue, here 2 live cells (not blue stained) represents 1 x $10^4$ cells/mL in the original undiluted stock of resuspended cells. Based on this calculated live cell concentration, 2.13 x $10^6$ cells were added per 60 mm petri dish, or 4.5 x $10^5$ cells were added per well in a 12-well plate.

DNA for transfection was purified with QIAGEN's miniprep spin column kit, eluted with sterile water. Plasmid concentrations were normalized to facilitate transfections.

Cells were transfected or co-transfected with maximum 1 µg vector DNA using OPTI-MEM (REF # 31985-062, Gibco by Life Technologies) and Lipofectamine 2000 (REF # 11668-027, Invitrogen) following the suggested protocol.

**Sequencing Splice Products from HEK293T**

Splice products were sequence verified by running RT-PCR product on an agarose gel and extracting the desired band for purification by QIAquick Gel Extraction Kit (QIAGEN). The purified oligo was inserted into pCR2.1-TOPO vector following the TOPA TA Cloning Manual (Thermo Fisher Scientific) and vectors in positive *E.coli* colonies were sequenced following standard practices.

**RNA Extraction and RT-PCR**

Prior to RNA isolation, media was removed and cells were washed with 1X PBS buffer. A volume of 1 mL TRIzol Reagent (REF # 15596-026, Ambion by Life Technologies) was added per well in 12-well plates. RNA was isolated using TRIzol, following the reagent protocol supplied by Ambion. RNA was DNase-treated with RQ1 (Promega) and cDNA was synthesized using oligo(dT) primers and AMV Reverse Transcriptase (New England Biolabs). RT-PCR was performed using primers binding *EGFP* at sites ~10 bp outside the *EGFP-HyP5SM* splice sites (for SP-I size of 200 bp) or with a primer set made specific for SP-II by amplifying over the HyP5SM exon junctions with *EGFP*. RT-PCR was run with *Taq* DNA Polymerase with ThermoPol buffer (New England Biolabs).

**Microscopy**

HEK293T were visualized on an inverted microscope (AxioObserver) in the UC Berkeley Molecular Imaging Center. The multi-D acquire setting was set for DAPI and GFP, and the field of focus was previewed using for BrightField (BF) to select areas with cells. Exposure settings: GFP: 0.020 sec exposure; BF: 20 msec exposure.

**RNA Structure Prediction**

RNA sequence was input into the Mfold web server hosted by the RNA Institute (227). The temperature was set to 37°C and other settings were left as default. Mfold URL: <http://unafold.rna.albany.edu/?q=mfold>

## REFERENCES


1. Crick, F.H., Barnett, L., Brenner, S. and Watts-Tobin, R.J. (1961) General nature of the genetic code for proteins. *Nature*, **192**, 1227–32.

2. Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F. and O'Neal, C. (1965) RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proc. Natl. Acad. Sci. U. S. A.*, **53**, 1161–8.

3. Abelson, J. (1979) RNA Processing and the Intervening Sequence Problem. *Annu. Rev. Biochem.*, **48**, 1035–1069.

4. Wallace, R., Johnson, P., Tanaka, S., Schold, M., Itakura, K. and Abelson, J. (1980) Directed deletion of a yeast transfer RNA intervening sequence. *Science*, **209**, 1396–1400.

5. Zaug, A.J. and Cech, T.R. (1980) In vitro splicing of the ribosomal RNA precursor in nuclei of Tetrahymena. *Cell*, **19**, 331–8.

6. Kruger, K., Grabowski, P., Zaug, A., Sands, J., Gottschling, D. and Cech, T. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147–157.

7. Salman, V., Amann, R., Shub, D.A. and Schulz-Vogt, H.N. (2012) Multiple self-splicing introns in the 16S rRNA genes of giant sulfur bacteria. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 4203–8.

8. Koonin, E. V (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol. Direct*, **1**, 22.

9. Cech, T. (1990) Self-splicing of group I introns. *Annu. Rev. Biochem.*, **59**, 543–568.

10. Lambowitz, A. and Belfort, M. (1993) Introns as mobile genetic elements. *Annu. Rev. Biochem.*, **62**, 587–622.

11. Roy, S.W. and Gilbert, W. (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat. Rev. Genet.*, **7**, 211–21.

12. Simpson, G.G. and Filipowicz, W. (1996) Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spliceosomal machinery. *Plant Mol. Biol.*, **32**, 1–41.

13. Nurtdinov, R.N. (2003) Low conservation of alternative splicing patterns in the human and mouse genomes. *Hum. Mol. Genet.*, **12**, 1313–1320.

14. Fedorov, A., Merican, A.F. and Gilbert, W. (2002) Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 16128–33.

15. Jackson, S., Cannone, J., Lee, J., Gutell, R. and Woodson, S. (2002) Distribution of rRNA introns in the three-dimensional structure of the ribosome. *J. Mol. Biol.*, **323**, 35–52.

16. Klinz, F.-J. and Gallwitz, D. (1985) Size and position of intervening sequences are critical for the splicing efficiency of pre-mRNA in the yeast Saccharomyces cerevisiae. *Nucleic Acids Res.*, **13**, 3791–3804.

17. Baynton, C., Potthoff, S., McCullough, A. and Schuler, M. (1996) U-rich tracts enhance 3' splice site recognition in plant nuclei. *Plant J.*, **10**, 703–711.

18. Wiebauer, K., Herrero, J.J. and Filipowicz, W. (1988) Nuclear pre-mRNA processing in plants: distinct modes of 3'-splice-site selection in plants and animals. *Mol. Cell. Biol.*, **8**, 2042–2051.

19. Luehrsen, K.R. and Walbot, V. (1994) Intron creation and polyadenylation in maize are directed by AU-rich RNA. *Genes Dev.*, **8**, 1117–1130.

20. Ram, O. and Ast, G. (2007) SR proteins: a foot on the exon before the transition from intron to exon definition. *Trends Genet.*, **23**, 5–7.

21. Meyer, M., Plass, M., Pérez-Valle, J., Eyras, E. and Vilardell, J. (2011) Deciphering 3′ss Selection in the Yeast Genome Reveals an RNA Thermosensor that Mediates Alternative Splicing. *Mol. Cell*, **43**, 1033–1039.

22. Gahura, O., Hammann, C., Valentová, A., Půta, F. and Folk, P. (2011) Secondary structure is required for 3' splice site recognition in yeast. *Nucleic Acids Res.*, **39**, 9759–67.

23. Berget, S.M. (1995) Exon Recognition in Vertebrate Splicing. *J. Biol. Chem.*, **270**, 2411–2414.

24. Dominski, Z. and Kole, R. (1991) Selection of splice sites in pre-mRNAs with short internal exons. *Mol. Cell. Biol.*, **11**, 6075–6083.

25. Lopato, S., Mayeda, A., Krainer, A.R. and Barta, A. (1996) Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 3074–9.

26. Gao, K., Masuda, A., Matsuura, T. and Ohno, K. (2008) Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.*, **36**, 2257–67.

27. Domdey, H., Apostol, B., Lin, R.-J., Newman, A., Brody, E. and Abelson, J. (1984) Lariat structures are in vivo intermediates in yeast pre-mRNA splicing. *Cell*, **39**, 611–621.

28. Spingola, M., Grate, L., Haussler, D. and Ares, M.. J. (1999) Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *RNA*, **5**, 221–234.

29. Graveley, B.R. (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.*, **17**, 100–107.

30. Juneau, K., Nislow, C. and Davis, R.W. (2009) Alternative splicing of PTC7 in Saccharomyces cerevisiae determines protein localization. *Genetics*, **183**, 185–94.

31. Syed, N.H., Kalyna, M., Marquez, Y., Barta, A. and Brown, J.W.S. (2012) Alternative splicing in plants--coming of age. *Trends Plant Sci.*, **17**, 616–23.

32. Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–5.

33. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. (1996) Life with 6000 genes. *Science*, **274**, 546, 563–7.

34. Davis, C.A., Grate, L., Spingola, M. and Ares, M. (2000) Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. *Nucleic Acids Res.*, **28**, 1700–6.

35. Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., et al. (2002) The genome sequence of Schizosaccharomyces pombe. *Nature*, **415**, 871–80.

36. Mao, R., Raj Kumar, P.K., Guo, C., Zhang, Y. and Liang, C. (2014) Comparative analyses between retained introns and constitutively spliced introns in Arabidopsis thaliana using random forest and support vector machine. *PLoS One*, **9**, e104049.

37. Barbazuk, W.B., Fu, Y. and McGinnis, K.M. (2008) Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome Res.*, **18**, 1381–92.

38. Yu, J., Hu, S., Wang, J., Wong, G.K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science*, **296**, 79–92.

39. Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S.M. and Buell, C.R. (2006) Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics*, **7**, 327.

40. Shepard, S., McCreary, M. and Fedorov, A. (2009) The peculiarities of large intron splicing in animals. *PLoS One*, **4**, e7853.

41. Wang, B.-B. and Brendel, V. (2006) Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 7175–80.

42. Gupta, S., Zink, D., Korn, B., Vingron, M. and Haas, S.A. (2004) Genome wide identification and classification of alternative splicing based on EST data. *Bioinformatics*, **20**, 2579–2585.

43. Aprile, A., Mastrangelo, A.M., De Leonardis, A.M., Galiba, G., Roncaglia, E., Ferrari, F., De Bellis, L., Turchi, L., Giuliano, G. and Cattivelli, L. (2009) Transcriptional profiling in response to terminal drought stress reveals differential responses along the wheat genome. *BMC Genomics*, **10**, 279.

44. Bozsó, Z., Maunoury, N., Szatmari, A., Mergaert, P., Ott, P.G., Zsíros, L.R., Szabó, E., Kondorosi, E. and Klement, Z. (2009) Transcriptome analysis of a bacterially induced basal and hypersensitive response of Medicago truncatula. *Plant Mol. Biol.*, **70**, 627–46.

45. Staiger, D. and Brown, J.W.S. (2013) Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell*, **25**, 3640–56.

46. Cruz, T.M.D., Carvalho, R.F., Richardson, D.N. and Duque, P. (2014) Abscisic Acid (ABA) Regulation of Arabidopsis SR Protein Gene Expression. *Int. J. Mol. Sci.*, **15**, 17541–64.

47. Sablowski, R.W.M. and Meyerowitz, E.M. (1998) Temperature-Sensitive Splicing in the Floral Homeotic Mutant apetala3-1. *Plant Cell*, **10**, 1453–1464.

48. Roca, X., Krainer, A.R. and Eperon, I.C. (2013) Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes Dev.*, **27**, 129–44.

49. Krainer, A., Conway, G. and Kozak, D. (1990) The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell*, **62**, 35–42.

50. Zahler, a M. and Roth, M.B. (1995) Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. *Proc. Natl. Acad. Sci. U. S. A.*, **92**, 2642–6.

51. Fu, X. (1995) The superfamily of arginine/serine-rich splicing factors. *RNA*, **1**, 663–680.

52. Barta, A., Kalyna, M. and Lorković, Z.J. (2008) Plant SR Proteins and Their Functions. *Curr. Opin. Microbiol. Immunol.*, **326**, 83–102.

53. Roscigno, R.F. and Garcia-Blanco, M.A. (1995) SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA*, **1**, 692–706.

54. Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S. and Fu, X.-D. (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.*, **15**, 819–26.

55. Liu, H.X., Chew, S.L., Cartegni, L., Zhang, M.Q. and Krainer, A.R. (2000) Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol. Cell. Biol.*, **20**, 1063–71.

56. Cramer, P., Cáceres, J.F., Cazalla, D., Kadener, S., Muro, a F., Baralle, F.E. and Kornblihtt, a R. (1999) Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol. Cell*, **4**, 251–8.

57. Bentley, D.L. (2014) Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.*, **15**, 163–75.

58. Chang, Y.-F., Imam, J.S. and Wilkinson, M.F. (2007) The Nonsense-Mediated Decay RNA Surveillance Pathway. *Annu. Rev. Biochem.*, **76**, 51–74.

59. Culbertson, M.R. (1999) RNA surveillance. Unforeseen consequences for gene expression, inherited genetic disorders and cancer. *Trends Genet.*, **15**, 74–80.

60. Culbertson, M.R. and Leeds, P.F. (2003) Looking at mRNA decay pathways through the window of molecular evolution. *Curr. Opin. Genet. Dev.*, **13**, 207–214.

61. Kerényi, Z., Mérai, Z., Hiripi, L., Benkovics, A., Gyula, P., Lacomme, C., Barta, E., Nagy, F. and Silhavy, D. (2008) Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay. *EMBO J.*, **27**, 1585–95.

62. Tange, T.Ø., Nott, A. and Moore, M.J. (2004) The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.*, **16**, 279–84.

63. Shibuya, T., Tange, T.Ø., Sonenberg, N. and Moore, M.J. (2004) eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat. Struct. Mol. Biol.*, **11**, 346–51.

64. Chamieh, H., Ballut, L., Bonneau, F. and Le Hir, H. (2008) NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat. Struct. Mol. Biol.*, **15**, 85–93.

65. Le Hir, H., Izaurralde, E., Maquat, L.E. and Moore, M.J. (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.*, **19**, 6860–9.

66. Nagy, E. and Maquat, L.E. (1998) A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.*, **23**, 198–199.

67. Brocke, K.S., Neu-Yilik, G., Gehring, N.H., Hentze, M.W. and Kulozik, A.E. (2002) The human intronless melanocortin 4-receptor gene is NMD insensitive. *Hum. Mol. Genet.*, **11**, 331–5.

68. Rufener, S.C. and Mühlemann, O. (2013) eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. *Nat. Struct. Mol. Biol.*, **20**, 710–717.

69. Silva, A.L., Ribeiro, P., Inácio, A., Liebhaber, S.A. and Romão, L. (2008) Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. *RNA*, **14**, 563–76.

70. Ivanov, P. V, Gehring, N.H., Kunz, J.B., Hentze, M.W. and Kulozik, A.E. (2008) Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J.*, **27**, 736–47.

71. Maquat, L.E. (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.*, **5**, 89–99.

72. Brogna, S. and Wen, J. (2009) Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Mol Biol*, **16**, 107–113.

73. Ishigaki, Y., Li, X., Serin, G. and Maquat, L.E. (2001) Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell*, **106**, 607–17.

74. Matsuda, D., Hosoda, N., Kim, Y.K. and Maquat, L.E. (2007) Failsafe nonsense-mediated mRNA decay does not detectably target eIF4E-bound mRNA. *Nat. Struct. Mol. Biol.*, **14**, 974–979.

75. Lejeune, F., Ishigaki, Y., Li, X. and Maquat, L.E. (2002) The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J.*, **21**, 3536–45.

76. Durand, S. and Lykke-Andersen, J. (2013) Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. *Nat. Struct. Mol. Biol.*, **20**, 702–709.

77. Metze, S., Herzog, V.A., Ruepp, M.-D. and Mühlemann, O. (2013) Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. *RNA*, **19**, 1432–48.

78. KEELING, K.M. (2004) Leaky termination at premature stop codons antagonizes nonsense-mediated mRNA decay in S. cerevisiae. *RNA*, **10**, 691–703.

79. Wen, J. and Brogna, S. (2010) Splicing-dependent NMD does not require the EJC in Schizosaccharomyces pombe. *EMBO J.*, **29**, 1537–1551.

80. Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S. and Jacobson, A. (2004) A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature*, **432**, 112–8.

81. DENISE MUHLRAD and ROY PARKER (1999) Aberrant mRNAs with extended 3′ UTRs are substrates for rapid degradation by mRNA surveillance. *RNA*, **5**, 1299–1307.

82. Kertész, S., Kerényi, Z., Mérai, Z., Bartos, I., Pálfy, T., Barta, E. and Silhavy, D. (2006) Both introns and long 3′-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. *Nucleic Acids Res.*, **34**, 6147–6157.

83. Nyikó, T., Kerényi, F., Szabadkai, L., Benkovics, A.H., Major, P., Sonkoly, B., Mérai, Z., Barta, E., Niemiec, E., Kufel, J., et al. (2013) Plant nonsense-mediated mRNA decay is controlled by different autoregulatory circuits and can be induced by an EJC-like complex. *Nucleic Acids Res.*, **41**, 6715–28.

84. Mufarrege, E.F., Gonzalez, D.H. and Curi, G.C. (2011) Functional interconnections of Arabidopsis exon junction complex proteins and genes at multiple steps of gene expression. *J. Exp. Bot.*, **62**, 5025–36.

85. Nott, A., Le Hir, H. and Moore, M.J. (2004) Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes Dev.*, **18**, 210–22.

86. Lykke-Andersen, J., Shu, M.D. and Steitz, J.A. (2000) Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell*, **103**, 1121–31.

87. Voelker, T.A., Moreno, J. and Chrispeels, M.J. (1990) Expression analysis of a pseudogene in transgenic tobacco: a frameshift mutation prevents mRNA accumulation. *Plant Cell*, **2**, 255–61.

88. Hoof, A. and Green, P.J. (1996) Premature nonsense codons decrease the stability of phytohemagglutinin mRNA in a position-dependent manner. *Plant J.*, **10**, 415–424.

89. Hori, K. and Watanabe, Y. (2007) Context analysis of termination codons in mRNA that are recognized by plant NMD. *Plant Cell Physiol.*, **48**, 1072–8.

90. Lewis, B.P., Green, R.E. and Brenner, S.E. (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 189–92.

91. Lejeune, F. and Maquat, L.E. (2005) Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr. Opin. Cell Biol.*, **17**, 309–15.

92. Riehs-Kearnan, N., Gloggnitzer, J., Dekrout, B., Jonak, C. and Riha, K. (2012) Aberrant growth and lethality of Arabidopsis deficient in nonsense-mediated RNA decay factors is caused by autoimmune-like response. *Nucleic Acids Res.*, **40**, 5615–24.

93. Yoine, M., Nishii, T. and Nakamura, K. (2006) Arabidopsis UPF1 RNA Helicase for Nonsense-mediated mRNA Decay is Involved in Seed Size Control and is Essential for Growth. *Plant Cell Physiol.*, **47**, 572–80.

94. Rehwinkel, J., Letunic, I., Raes, J., Bork, P. and Izaurralde, E. (2005) Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. *RNA*, **11**, 1530–1544.

95. Wittkopp, N., Huntzinger, E., Weiler, C., Saulière, J., Schmidt, S., Sonawane, M. and Izaurralde, E. (2009) Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival. *Mol. Cell. Biol.*, **29**, 3517–28.

96. Azzalin, C.M. and Lingner, J. (2006) The human RNA surveillance factor UPF1 is required for S phase progression and genome stability. *Curr. Biol.*, **16**, 433–9.

97. Weischenfeldt, J., Damgaard, I., Bryder, D., Theilgaard-Mönch, K., Thoren, L.A., Nielsen, F.C., Jacobsen, S.E.W., Nerlov, C. and Porse, B.T. (2008) NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev.*, **22**, 1381–96.

98. Bhuvanagiri, M., Schlitter, A.M., Hentze, M.W. and Kulozik, A.E. (2010) NMD: RNA biology meets human genetic medicine. *Biochem. J.*, **430**, 365–77.

99. Keeling, K.M. and Bedwell, D.M. (2011) Suppression of nonsense mutations as a therapeutic approach to treat genetic diseases. *Wiley Interdiscip. Rev. RNA*, **2**, 837–52.

100. Brown, J.W.S., Simpson, C.G., Marquez, Y., Gadd, G.M., Barta, A. and Kalyna, M. (2015) Lost in Translation: Pitfalls in Deciphering Plant Alternative Splicing Transcripts. *Plant Cell*, **27**, 2083–7.

101. Gloggnitzer, J., Akimcheva, S., Srinivasan, A., Kusenda, B., Riehs, N., Stampfl, H., Bautor, J., Dekrout, B., Jonak, C., Jiménez-Gómez, J.M., et al. (2014) Nonsense-mediated mRNA decay modulates immune receptor levels to regulate plant antibacterial defense. *Cell Host Microbe*, **16**, 376–90.

102. Filichkin, S. a and Mockler, T.C. (2012) Unproductive alternative splicing and nonsense mRNAs: a widespread phenomenon among plant circadian clock genes. *Biol. Direct*, **7**, 20.

103. Callis, J., Fromm, M. and Walbot, V. (1987) Introns increase gene expression in cultured maize cells. *Genes Dev.*, **1**, 1183–200.

104. Rose, A.B. and Beliakoff, J.A. (2000) Intron-mediated enhancement of gene expression independent of unique intron sequences and splicing. *Plant Physiol.*, **122**, 535–42.

105. Brinster, R.L., Allen, J.M., Behringer, R.R., Gelinas, R.E. and Palmiter, R.D. (1988) Introns increase transcriptional efficiency in transgenic mice. *Proc. Natl. Acad. Sci. U. S. A.*, **85**, 836–40.

106. Lai, Y., Yue, Y., Liu, M. and Duan, D. (2006) Synthetic intron improves transduction efficiency of trans-splicing adeno-associated viral vectors. *Hum. Gene Ther.*, **17**, 1036–42.

107. Mähönen, A.J., Airenne, K.J., Lind, M.M., Lesch, H.P. and Ylä-Herttuala, S. (2004) Optimized self-excising Cre-expression cassette for mammalian cells. *Biochem. Biophys. Res. Commun.*, **320**, 366–71.

108. Weigand, J.E. and Suess, B. (2009) Aptamers and riboswitches: perspectives in biotechnology. *Appl. Microbiol. Biotechnol.*, **85**, 229–36.

109. Cheah, M.T., Wachter, A., Sudarsan, N. and Breaker, R.R. (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, **447**, 497–500.

110. Kim, D.-S., Gusti, V., Pillai, S.G. and Gaur, R.K. (2005) An artificial riboswitch for controlling pre-mRNA splicing. *RNA*, **11**, 1667–77.

111. Weigand, J.E. and Suess, B. (2007) Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. *Nucleic Acids Res.*, **35**, 4179–85.

112. Hammond, M.C., Wachter, A. and Breaker, R.R. (2009) A plant 5S ribosomal RNA mimic regulates alternative splicing of transcription factor IIIA pre-mRNAs. *Nat. Struct. Mol. Biol.*, **16**, 541–549.

113. Fu, Y., Bannach, O., Chen, H., Teune, J.-H., Schmitz, A., Steger, G., Xiong, L. and Barbazuk, W.B. (2009) Alternative splicing of anciently exonized 5S rRNA regulates plant transcription factor TFIIIA. *Genome Res.*, **19**, 913–21.

114. Brown, D.D. and Schlissel, M.S. (1985) A positive transcription factor controls the differential expression of two 5S RNA genes. *Cell*, **42**, 759–67.

115. Scripture, J.B. and Huber, P.W. (1995) Analysis of the binding of Xenopus ribosomal protein L5 to oocyte 5 S rRNA. The major determinants of recognition are located in helix III-loop C. *J. Biol. Chem.*, **270**, 27358–27365.

116. Cuccurese, M., Russo, G., Russo, A. and Pietropaolo, C. (2005) Alternative splicing and nonsense-mediated mRNA decay regulate mammalian ribosomal gene expression. *Nucleic Acids Res.*, **33**, 5965–77.

117. Mitrovich, Q.M. (2000) Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in C. elegans. *Genes Dev.*, **14**, 2173–2184.

118. Pittman, R.H., Andrews, M.T. and Setzer, D.R. (1999) A Feedback Loop Coupling 5 S rRNA Synthesis to Accumulation of a Ribosomal Protein. *J. Biol. Chem.*, **274**, 33198–33201.

119. Mathieu, O., Yukawa, Y., Prieto, J.-L.L., Vaillant, I., Sugiura, M. and Tourmente, S. (2003) Identification and characterization of transcription factor IIIA and ribosomal protein L5 from Arabidopsis thaliana. *Nucleic Acids Res.*, **31**, 2424–2433.

120. Hickey, S.F., Sridhar, M., Westermann, A.J., Qin, Q., Vijayendra, P., Liou, G. and Hammond, M.C. (2012) Transgene regulation in plants by alternative splicing of a suicide exon. *Nucleic Acids Res.*, **40**, 4701–4710.

121. Kosiba, B.E. and Schleif, R. (1982) Arabinose-inducible promoter from Escherichia coli. *J. Mol. Biol.*, **156**, 53–66.

122. Picard, D., Schena, M. and Yamamoto, K.R. (1990) An inducible expression vector for both fission and budding yeast. *Gene*, **86**, 257–261.

123. Aoyama, T. and Chua, N.-H. (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J.*, **11**, 605–612.

124. Gossen, M. and Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci.*, **89**, 5547–5551.

125. Chakraborty, S. and Newton, A.C. (2011) Climate change, plant diseases and food security: an overview. *Plant Pathol.*, **60**, 2–14.

126. Shaw, M.W. and Osborne, T.M. (2011) Geographic distribution of plant pathogens in response to climate change. *Plant Pathol.*, **60**, 31–43.

127. Zhang, J. and Zhou, J.-M. (2010) Plant immunity triggered by microbial molecular signatures. *Mol. Plant*, **3**, 783–793.

128. Chisholm, S.T., Coaker, G., Day, B. and Staskawicz, B.J. (2006) Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response. *Cell*, **124**, 803–14.

129. Shao, Z.-Q., Zhang, Y.-M., Hang, Y.-Y., Xue, J.-Y., Zhou, G.-C., Wu, P., Wu, X.-Y., Wu, X.-Z., Wang, Q., Wang, B., et al. (2014) Long-Term Evolution of Nucleotide-Binding Site-Leucine-Rich Repeat Genes: Understanding Gained from and beyond the Legume Family. *Plant Physiol.*, **166**, 217–34.

130. Botella, M.A., Parker, J.E., Frost, L.N., Bittner-Eddy, P.D., Beynon, J.L., Daniels, M.J., Holub, E.B. and Jones, J.D. (1998) Three genes of the Arabidopsis RPP1 complex resistance locus recognize distinct Peronospora parasitica avirulence determinants. *Plant Cell*, **10**, 1847–60.

131. Dodds, P.N. and Rathjen, J.P. (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.*, **11**, 539–548.

132. Tsuda, K., Sato, M., Stoddard, T., Glazebrook, J. and Katagiri, F. (2009) Network properties of robust immunity in plants. *PLoS Genet.*, **5**, e1000772.

133. Boccara, M., Sarazin, A., Thiébeauld, O., Jay, F., Voinnet, O., Navarro, L. and Colot, V. (2014) The Arabidopsis miR472-RDR6 Silencing Pathway Modulates PAMP- and Effector-Triggered Immunity through the Post-transcriptional Control of Disease Resistance Genes. *PLoS Pathog.*, **10**, e1003883.

134. Bozkurt, T.O., Schornack, S., Banfield, M.J. and Kamoun, S. (2012) Oomycetes, effectors, and all that jazz. *Curr. Opin. Plant Biol.*, **15**, 483–492.

135. Büttner, D. and Bonas, U. (2002) Getting across—bacterial type III effector proteins on their way to the plant cell. *EMBO J.*, **21**, 5313–5322.

136. Stergiopoulos, I. and Wit, P.J.G.M. de (2009) Fungal Effector Proteins. *Annu. Rev. Phytopathol.*, **47**, 233–263.

137. Tornero, P., Chao, R.A., Luthin, W.N., Goff, S.A. and Dangl, J.L. (2002) Large-scale structure-function analysis of the Arabidopsis RPM1 disease resistance protein. *Plant Cell*, **14**, 435–450.

138. Bendahmane, A. (1999) The Rx Gene from Potato Controls Separate Virus Resistance and Cell Death Responses. *Plant Cell*, **11**, 781–792.

139. De Jong, C.F., Takken, F.L., Cai, X., de Wit, P.J. and Joosten, M.H. (2002) Attenuation of Cf-mediated defense responses at elevated temperatures correlates with a decrease in elicitor-binding sites. *Mol. Plant. Microbe. Interact.*, **15**, 1040–1049.

140. Kearney, B. and Staskawicz, B.J. (1990) Widespread distribution and fitness contribution of Xanthomonas campestris avirulence gene avrBs2. *Nature*, **346**, 385–386.

141. Jones, J.B., Lacy, G.H., Bouzar, H., Stall, R.E. and Schaad, N.W. (2004) Reclassification of the xanthomonads associated with bacterial spot disease of tomato and pepper. *Syst. Appl. Microbiol.*, **27**, 755–762.

142. Tai, T.H., Dahlbeck, D., Clark, E.T., Gajiwala, P., Pasion, R., Whalen, M.C., Stall, R.E. and Staskawicz, B.J. (1999) Expression of the Bs2 pepper gene confers resistance to bacterial spot disease in tomato. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 14153–14158.

143. Chou, S., Krasileva, K. V, Holton, J.M., Steinbrenner, A.D., Alber, T. and Staskawicz, B.J. (2011) Hyaloperonospora arabidopsidis ATR1 effector is a repeat protein with distributed recognition surfaces. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 13323–13328.

144. Botella, M., Parker, J., Frost, L., Bittner-Eddy, P., Beynon, J., Daniels, M., Holub, E. and Jones, J. (1998) Three genes of the Arabidopsis RPP1 complex resistance locus recognize distinct Peronospora parasitica avirulence determinants. *Plant Cell*, **10**, 1847–1860.

145. Krasileva, K., Dahlbeck, D. and Staskawicz, B. (2010) Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. *Plant Cell Online*, **22**, 2444–2458.

146. Horvath, D.M., Stall, R.E., Jones, J.B., Pauly, M.H., Vallad, G.E., Dahlbeck, D., Staskawicz, B.J. and Scott, J.W. (2012) Transgenic Resistance Confers Effective Field Level Control of Bacterial Spot Disease in Tomato. *PLoS One*, **7**, e42036.

147. Rairdan, G.J. and Moffett, P. (2006) Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation. *Plant Cell*, **18**, 2082–2093.

148. Eitas, T.K. and Dangl, J.L. (2010) NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Curr. Opin. Plant Biol.*, **13**, 472–477.

149. Rustérucci, C., Aviv, D.H., Holt, B.F., Dangl, J.L. and Parker, J.E. (2001) The disease resistance signaling components EDS1 and PAD4 are essential regulators of the cell death pathway controlled by LSD1 in Arabidopsis. *Plant Cell*, **13**, 2211–2224.

150. Parker, J., Holub, E. and Frost, L. (1996) Characterization of eds1, a mutation in Arabidopsis suppressing resistance to Peronospora parasitica specified by several different RPP genes. *Plant Cell*, **8**, 2033–2046.

151. Aarts, N., Metz, M., Holub, E., Staskawicz, B.J., Daniels, M.J. and Parker, J.E. (1998) Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R gene-mediated signaling pathways in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 10306–10311.

152. Leister, R.T., Dahlbeck, D., Day, B., Li, Y., Chesnokova, O. and Staskawicz, B.J. (2005) Molecular Genetic Evidence for the Role of SGT1 in the Intramolecular Complementation of Bs2 Protein Activity in Nicotiana benthamiana. *Plant Cell*, **17**, 1268–1278.

153. Moffett, P., Farnham, G., Peart, J. and Baulcombe, D.C. (2002) Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death. *EMBO J.*, **21**, 4511–4519.

154. Dowen, R.H., Engel, J.L., Shao, F., Ecker, J.R. and Dixon, J.E. (2009) A family of bacterial cysteine protease type III effectors utilizes acylation-dependent and -independent strategies to localize to plasma membranes. *J. Biol. Chem.*, **284**, 15867–15879.

155. Sakvarelidze, L., Tao, Z., Bush, M., Roberts, G.R., Leader, D.J., Doonan, J.H. and Rawsthorne, S. (2007) Coupling the GAL4 UAS system with alcR for versatile cell type-specific chemically inducible gene expression in Arabidopsis. *Plant Biotechnol. J.*, **5**, 465–476.

156. Krasileva, K. V, Dahlbeck, D. and Staskawicz, B.J. (2010) Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. *Plant Cell*, **22**, 2444–2458.

157. Day, B., Dahlbeck, D., Huang, J., Chisholm, S.T., Li, D. and Staskawicz, B.J. (2005) Molecular Basis for the RIN4 Negative Regulation of RPS2 Disease Resistance. *Plant Cell*, **17**, 1292–1305.

158. Lee, L.-Y., Kononov, M.E., Bassuner, B., Frame, B.R., Wang, K. and Gelvin, S.B. (2007) Novel plant transformation vectors containing the superpromoter. *Plant Physiol.*, **145**, 1294–300.

159. Höfgen, R. and Willmitzer, L. (1990) Biochemical and genetic analysis of different patatin isoforms expressed in various organs of potato (Solanum tuberosum). *Plant Sci.*, **66**, 221–230.

160. Mudgett, M., Chesnokova, O., Dahlbeck, D., Clark, E., Rossier, O., Bonas, U. and Staskawicz, B. (2000) Molecular signals required for type III secretion and translocation of the Xanthomonas campestris AvrBs2 protein to pepper plants. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 13324–13329.

161. Ni, M., Cui, D., Einstein, J., Narasimhulu, S., Vergara, C.E. and Gelvin, S.B. (1995) Strength and tissue specificity of chimeric promoters derived from the octopine and mannopine synthase genes. *Plant J.*, **7**, 661–676.

162. Laemmli, U.K. (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, **227**, 680–685.

163. Brown, J.W.S. and Simpson, C.G. (1998) Splice Site Selection in Plant Pre-mRNA Splicing. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **49**, 77–95.

164. Murray, E.E., Lotzer, J. and Eberle, M. (1989) Codon usage in plant genes. *Nucleic Acids Res.*, **17**, 477–498.

165. De Jonge, R., Peter van Esse, H., Maruthachalam, K., Bolton, M.D., Santhanam, P., Saber, M.K.K., Zhang, Z., Usami, T., Lievens, B., Subbarao, K. V, et al. (2012) Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 5110–5.

166. Zhang, Z., van Esse, H.P., van Damme, M., Fradin, E.F., Liu, C.-M. and Thomma, B.P.H.J. (2013) Ve1-mediated resistance against Verticillium does not involve a hypersensitive response in Arabidopsis. *Mol. Plant Pathol.*, **14**, 719–27.

167. McNellis, T., Mudgett, M., Li, K., Aoyama, T., Horvath, D., Chua, N. and Staskawicz, B. (1998) Glucocorticoid-inducible expression of a bacterial avirulence gene in transgenic Arabidopsis induces hypersensitive cell death. *Plant J.*, **14**, 247–257.

168. Nimchuk, Z., Marois, E., Kjemtrup, S., Leister, R.T., Katagiri, F. and Dangl, J.L. (2000) Eukaryotic fatty acylation drives plasma membrane targeting and enhances function of several type III effector proteins from Pseudomonas syringae. *Cell*, **101**, 353–363.

169. Padidam, M. (2003) Chemically regulated gene expression in plants. *Curr. Opin. Plant Biol.*, **6**, 169–177.

170. Hong, W., Xu, Y.-P.P., Zheng, Z., Cao, J.-S.S. and Cai, X.-Z.Z. (2007) Comparative transcript profiling by cDNA-AFLP reveals similar patterns of Avr4/Cf-4- and Avr9/Cf-9-dependent defence gene expression. *Mol. Plant Pathol.*, **8**, 515–527.

171. Romero, A.M., Kousik, C.S. and Ritchie, D.F. (2002) Temperature Sensitivity of the Hypersensitive Response of Bell Pepper to Xanthomonas axonopodis pv. vesicatoria. *Phytopathology*, **92**, 197–203.

172. Jovel, J., Walker, M. and Sanfaçon, H. (2011) Salicylic acid-dependent restriction of Tomato ringspot virus spread in tobacco is accompanied by a hypersensitive response, local RNA silencing, and moderate systemic resistance. *Mol. Plant. Microbe. Interact.*, **24**, 706–718.

173. Rehmany, A.P., Gordon, A., Rose, L.E., Allen, R.L., Armstrong, M.R., Whisson, S.C., Kamoun, S., Tyler, B.M., Birch, P.R.J. and Beynon, J.L. (2005) Differential Recognition of Highly Divergent Downy Mildew Avirulence Gene Alleles by RPP1 Resistance Genes from Two Arabidopsis Lines. **17**, 1839–1850.

174. Wilkinson, J.E., Twell, D. and Lindsey, K. (1997) Activities of CaMV 35S and nos promoters in pollen: implications for field release of transgenic plants. *J. Exp. Bot.*, **48**, 265–275.

175. De Mesa, M.C., Santiago-Doménech, N., Pliego-Alfaro, F., Quesada, M.A. and Mercado, J.A. (2004) The CaMV 35S promoter is highly active on floral organs and pollen of transgenic strawberry plants. *Plant Cell Rep.*, **23**, 32–8.

176. Töpfer, R., Matzeit, V., Gronenborn, B., Schell, J. and Steinbiss, H.H. (1987) A set of plant expression vectors for transcriptional and translational fusions. *Nucleic Acids Res.*, **15**, 5890.

177. Eudes, A., George, A., Mukerjee, P., Kim, J.S., Pollet, B., Benke, P.I., Yang, F., Mitra, P., Sun, L., Cetinkol, O.P., et al. (2012) Biosynthesis and incorporation of side-chain-truncated lignin monomers to reduce lignin polymerization and enhance saccharification. *Plant Biotechnol. J.*, **10**, 609–620.

178. Bent, A. (2006) Arabidopsis thaliana floral dip transformation method. *Methods Mol. Biol.*, **343**, 87–103.

179. Gonzalez, T.L., Liang, Y., Nguyen, B.N., Staskawicz, B.J., Loque, D. and Hammond, M.C. (2015) Tight regulation of plant immune responses by combining promoter and suicide exon elements. *Nucleic Acids Res.*, **43**, 7152–61.

180. Lin, E., Lin, S.W. and Lin, A. (2001) The participation of 5S rRNA in the co-translational formation of a eukaryotic 5S ribonucleoprotein complex. *Nucleic Acids Res.*, **29**, 2510–6.

181. Leidig, C., Thoms, M., Holdermann, I., Bradatsch, B., Berninghausen, O., Bange, G., Sinning, I., Hurt, E. and Beckmann, R. (2014) 60S ribosome biogenesis requires rotation of the 5S ribonucleoprotein particle. *Nat. Commun.*, **5**, 3491.

182. Nilsen, T.W. and Graveley, B.R. (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature*, **463**, 457–63.

183. Meyer, K., Koester, T. and Staiger, D. (2015) Pre-mRNA Splicing in Plants: In Vivo Functions of RNA-Binding Proteins Implicated in the Splicing Process. *Biomolecules*, **5**, 1717–1740.

184. Cáceres, J.F., Stamm, S., Helfman, D.M. and Krainer, a R. (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science*, **265**, 1706–9.

185. Long, J.C. and Caceres, J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.*, **417**, 15.

186. Wang, B.-B. and Brendel, V. (2004) The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. *Genome Biol.*, **5**, R102.

187. Shen, H., Kan, J.L.C. and Green, M.R. (2004) Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol. Cell*, **13**, 367–76.

188. Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA*, **6**, 1197–211.

189. Lopato, S., Forstner, C., Kalyna, M., Hilscher, J., Langhammer, U., Indrapichate, K., Lorković, Z.J. and Barta, A. (2002) Network of Interactions of a Novel Plant-specific Arg/Ser-rich Protein, atRSZ33, with atSC35-like Splicing Factors. *J. Biol. Chem.*, **277**, 39989–39998.

190. Koornneef, M. and Meinke, D. (2010) The development of Arabidopsis as a model plant. *Plant J.*, **61**, 909–21.

191. Vo, L.T., Minet, M., Schmitter, J.M., Lacroute, F. and Wyers, F. (2001) Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. *Mol. Cell. Biol.*, **21**, 8346–56.

192. Krishna, S.S. (2003) Structural classification of zinc fingers: SURVEY AND SUMMARY. *Nucleic Acids Res.*, **31**, 532–550.

193. DiNitto, J.P. and Huber, P.W. (2003) Mutual induced fit binding of Xenopus ribosomal protein L5 to 5S rRNA. *J. Mol. Biol.*, **330**, 979–992.

194. Kalyna, M., Lopato, S. and Barta, A. (2003) Ectopic Expression of atRSZ33 Reveals Its Function in Splicing and Causes Pleiotropic Changes in Development. *Mol. Biol. Cell*, **14**, 3565–3577.

195. Kalyna, M., Lopato, S., Voronin, V. and Barta, A. (2006) Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res.*, **34**, 4395–4405.

196. Lorković, Z.J., Hilscher, J. and Barta, A. (2008) Co-localisation studies of Arabidopsis SR splicing factors reveal different types of speckles in plant cell nuclei. *Exp. Cell Res.*, **314**, 3175–3186.

197. Göhring, J., Jacak, J. and Barta, A. (2014) Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in Arabidopsis. *Plant Cell*, **26**, 754–64.

198. Lazar, G., Schaal, T., Maniatis, T. and Goodman, H.M. (1995) Identification of a plant serine-arginine-rich protein similar to the mammalian splicing factor SF2/ASF. *Proc. Natl. Acad. Sci.*, **92**, 7672–7676.

199. Lopato, S., Kalyna, M., Dorner, S., Kobayashi, R., Krainer, A.R. and Barta, A. (1999) atSRp30, one of two SF2/ASF-like proteins from Arabidopsis thaliana, regulates splicing of specific plant genes. *Genes Dev.*, **13**, 987–1001.

200. Mori, T., Yoshimura, K., Nosaka, R., Sakuyama, H., Koike, Y., Tanabe, N., Maruta, T., Tamoi, M. and Shigeoka, S. (2012) Subcellular and subnuclear distribution of high-light responsive serine/arginine-rich proteins, atSR45a and atSR30, in Arabidopsis thaliana. *Biosci. Biotechnol. Biochem.*, **76**, 2075–81.

201. Lorković, Z.J. and Barta, A. (2002) Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant Arabidopsis thaliana. *Nucleic Acids Res.*, **30**, 623–35.

202. Huang, X.-Y., Niu, J., Sun, M.-X., Zhu, J., Gao, J.-F., Yang, J., Zhou, Q. and Yang, Z.-N. (2013) CYCLIN-DEPENDENT KINASE G1 Is Associated with the Spliceosome to Regulate CALLOSE SYNTHASE5 Splicing and Pollen Wall Formation in Arabidopsis. *Plant Cell*, **25**, 637–48.

203. Lopato, S., Gattoni, R., Fabini, G., Stevenin, J. and Barta, A. (1999) A novel family of plant splicing factors with a Zn knuckle motif: examination of RNA binding and splicing activities. *Plant Mol. Biol.*, **39**, 761–73.

204. Golovkin, M. and Reddy, A.S. (1999) An SC35-like protein and a novel serine/arginine-rich protein interact with Arabidopsis U1-70K protein. *J. Biol. Chem.*, **274**, 36428–38.

205. Rausin, G., Tillemans, V., Stankovic, N., Hanikenne, M. and Motte, P. (2010) Dynamic nucleocytoplasmic shuttling of an Arabidopsis SR splicing factor: role of the RNA-binding domains. *Plant Physiol.*, **153**, 273–284.

206. Keshwani, M.M., Aubol, B.E., Fattet, L., Ma, C.-T., Qiu, J., Jennings, P.A., Fu, X.-D. and Adams, J.A. (2015) Conserved proline-directed phosphorylation regulates SR protein conformation and splicing function. *Biochem. J.*, **466**, 311–22.

207. Combier, J.P., de Billy, F., Gamas, P., Niebel, A. and Rivas, S. (2008) Trans-regulation of the expression of the transcription factor MtHAP2-1 by a uORF controls root nodule development. *Genes Dev.*, **22**, 1549–59.

208. Schaal, T.D. and Maniatis, T. (1999) Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell. Biol.*, **19**, 1705–19.

209. Fiil, B.K., Qiu, J.-L., Petersen, K., Petersen, M. and Mundy, J. (2008) Coimmunoprecipitation (co-IP) of Nuclear Proteins and Chromatin Immunoprecipitation (ChIP) from Arabidopsis. *CSH Protoc.*, **2008**, pdb.prot5049.

210. Donaldson, G.P. and Lee, V.T. (2011) A rapid assay for affinity and kinetics of DNA-protein interactions. *FASEB J*, **25**, 564.2–.

211. Tacke, R., Chen, Y. and Manley, J.L. (1997) Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 1148–53.

212. Fütterer, J. and Hohn, T. (1996) Translation in plants--rules and exceptions. *Plant Mol. Biol.*, **32**, 159–189.

213. Bivona, L., Zou, Z., Stutzman, N. and Sun, P.D. (2010) Influence of the second amino acid on recombinant protein expression. *Protein Expr. Purif.*, **74**, 248–56.

214. Graciet, E. and Wellmer, F. (2010) The plant N-end rule pathway: structure and functions. *Trends Plant Sci.*, **15**, 447–53.

215. A., B. and P., S. (2002) A test of fusion protein stability in the plant Arabidopsis thaliana reveals degradation signals from ACC synthase and from the plant N-end rule pathway. *Plant Cell Rep.*, **21**, 174–179.

216. Simpson, C.G., Thow, G., Clark, G.P., Jennings, S.N., Watters, J.A. and Brown, J.W.S. (2002) Mutational analysis of a plant branchpoint and polypyrimidine tract required for constitutive splicing of a mini-exon. *RNA*, **8**, 47–56.

217. Kakumanu, A., Ambavaram, M.M.R., Klumas, C., Krishnan, A., Batlang, U., Myers, E., Grene, R. and Pereira, A. (2012) Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. *Plant Physiol.*, **160**, 846–67.

218. Filichkin, S. a, Priest, H.D., Givan, S. a, Shen, R., Bryant, D.W., Fox, S.E., Wong, W.-K. and Mockler, T.C. (2010) Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Genome Res.*, **20**, 45–58.

219. Tacke, R. and Manley, J.L. (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.*, **14**, 3540–51.

220. Liu, H.X., Zhang, M. and Krainer, A.R. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.*, **12**, 1998–2012.

221. Sanford, J.R., Coutinho, P., Hackett, J.A., Wang, X., Ranahan, W. and Caceres, J.F. (2008) Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS One*, **3**, e3369.

222. Cavaloc, Y., Bourgeois, C.F., Kister, L. and Stévenin, J. (1999) The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA*, **5**, 468–483.

223. Lynch, K.W. and Maniatis, T. (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. *Genes Dev.*, **10**, 2089–2101.

224. Shi, H., Hoffman, B.E. and Lis, J.T. (1997) A specific RNA hairpin loop structure binds the RNA recognition motifs of the Drosophila SR protein B52. *Mol. Cell. Biol.*, **17**, 2649–57.

225. Scripture, J.B. and Huber, P.W. (1995) Analysis of the Binding of Xenopus Ribosomal Protein L5 to Oocyte 5 S rRNA. **270**, 27358–27365.

226. DiNitto, J. (2003) Mutual Induced Fit Binding of Xenopus Ribosomal Protein L5 to 5S rRNA. *J. Mol. Biol.*, **330**, 979–992.

227. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

## Appendix

**TABLE A1:** *Primers used for cloning and analysis in Chapters 2 and 3.*

Sequence Legend:

**lowercase bold** = restriction enzyme (RE) site

lowercase & not bold = random nucleotides for efficient RE digest or overlapping sequence for infusion cloning

lowercase with gray highlight = att site for BP recombination reaction

UPPERCASE = primer sequence

UPPERCASE UNDERLINE = point mutation

***UPPERCASE BOLD ITALIC*** = overlap with HyP5SM

F = forward primer; R = reverse primer

| NAME | | EXPERIMENT; 5'→3' SEQUENCE | | SITE |
|---|---|---|---|---|
| *>>> RT-PCR analysis of splice products. Primers are to amplify all splice products unless otherwise specified.* | | | | |
| | | ***avrBs2-HyP5SM-HA* (E/V)** | | |
| TLG36 | | CGGAAAACTCGCTGGCGTCCA | F | |
| TLG58 | | TGCGAATCACCAACGGCATTTCAC | R | |
| | | ***avrBs2-HyP5SM-HA* (E/V)**, SP-I specific | | |
| TLG83 | | ACGTCGAGGTAAGTTCCGATGG | F | |
| TLG84 | | AGCATCTGCTCCACACCG | R | |
| | | ***avrBs2-HyP5SM-HA* (E/P)** | | |
| TLG26 | | AGACCCTGCAAGGCAAG | F | |
| TLG57 | | TGCGACCTTGTTGTGCATCGATCA | R | |
| | | ***ATR1Δ51-HyP5SM-FLAG* (E128)** | | |
| GL35 | | GAACGGGATGATTTGATTGGCGAG | F | |
| TLG81 | | ACTGCTTCCTCCAATCGGTGC | R | |
| | | ***ATR1Δ51-HyP5SM-FLAG* (E168)** | | |
| TLG82 | | CACTAGGCTAGTAACAACCTATTCGG | F | |
| GL36 | | GGGTGCGAAAAAGTCAACATCGTG | R | |
| *>>> Other RT-PCR primers for A. thaliana samples* | | | | |
| | | ***RPP1* (WsB allele specific, amplifies coding sequence and intron-retaining product)** | | |
| TLG155 | | GCTCTACATGAGAGACTGCAAGG | F | |
| TLG156 | | GCGTTTCCAAAAGAGGGAAGC | R | |
| | | ***Bs2* resistance gene** | | |
| TLG191 | | GATTGTCGGGATGGGAGGCA | F | |
| TLG192 | | ACGCCATCCCACACTTCACA | R | |
| | | ***EF1alpha* housekeeping gene from *A. thaliana*** | | |
| MS62 | | GCTCTATGGAAGTTCGAGACC | F | |
| MS63 | | GTGTGGCAATCGAGAACTGG | R | |
| | | ***OsL5* full-length** | | |
| TLG47 | | atta**ctcgag**ATGGGAGGGTTTGTCAAGACCC | F | XhoI |

| | | | | |
|---|---|---|---|---|
| TLG52 | | tatgct**actagt**TCACTCATCATCCTCTTCCTCGTC | R | SpeI |
| | | ***OsL5* internal product, for nested PCR** | | |
| TLG86 | | CTCACGCTCCGTGGTTTGGACCAGG | F | |
| TLG87 | | CTCAGGTTCCTCCTCCGCCATAGACC | R | |
| *>>> Other RT-PCR primers for N. benthamiana samples* | | | | |
| | | ***OsL5* internal product optimized for specificity** | | |
| TLG88 | | ACCGTGTCTTTGGTGCCCTCAAG | F | |
| TLG87 | | CTCAGGTTCCTCCTCCGCCATAGACC | R | |
| *>>> Extension primers to insert HyP5SM cassette into avrBs2-HA at E/V codon site (GAA/GTC → GAG/GTA). E=E308.* | | | | |
| TLG43 | | ***ATGGTTTTCACTCTTTTGGTGTGTAG***GTAAGTTCCGATGG CGTGCCGGTGTT | F | |
| TLG44 | | ***CTCATGACAAGAGGATGCATAAATCTAC***CTCGAC GTCCAGCTCCAGATTGCGGTA | R | |
| *>>> Extension primers to insert HyP5SM cassette into avrBs2-HA at E/P codon site (GAG/CCG→GAG/CCA). E=E123.* | | | | |
| TLG32 | | ***CACTCTTTTGGTGTGTAG***CCAGTGTATCTGGATACCGCC | F | |
| TLG33 | | ***GACAAGAGGATGCATAAATCTAC***CTCCAGTGTGCCGGCA GCAA | R | |
| TLG34 | | GCTGCCGGCACACTGGAG***GTAGATTTATGCATCCTCTTGT C*** | F | |
| TLG35 | | CGGTATCCAGATACACTGG***CTACACACCAAAAGAGTGAA AACC*** | R | |
| *>>> Extension primers to insert HyP5SM cassette into ATR1Δ51-FLAG at E128 codon site (E/A; GAA/GCT→GAG/GCT)* | | | | |
| GL31 | | ***CACTCTTTTGGTGTGTAG***GCTCTCGCCACTAGGCTAG | F | |
| GL30 | | ***GACAAGAGGATGCATAAATCTAC***CTCATCATAGGTATCAT GGAGTGG | R | |
| *>>> Extension primers to insert HyP5SM cassette into ATR1Δ51-FLAG at E168 codon site (E/A; GAA/GCA→GAG/GCA)* | | | | |
| GL33 | | ***CACTCTTTTGGTGTGTAG***GCAGTGGCATCACTATGGAA | F | |
| GL32B | | ***GACAAGAGGATGCATAAATCTAC***CTCCTCCAATCGGTGC G | R | |
| *>>> Primers to amplify the full-length HyP5SM cassette, including introns, for use in 3-piece-ligation PCR, previously published* (120) | | | | |
| DNA37 | | GTAGATTTATGCATCCTCTTGTCATGAG | F | |
| DNA38 | | CTACACACCAAAAGAGTGAAAACCAT | R | |
| *>>> Cloning into pTA7001 (into the XhoI/SpeI site), XbaI ends are compatible with SpeI ends* | | | | |
| | | ***avrBs2-HA; avrBs2-HyP5SM-HA* (E/P and E/V)** | | |
| TLG45 | | tta**ctcgag**ATGCGTATCGGTCCTCTGCAACCTTC | F | XhoI |
| TLG46 | | catt**tctaga**CTACGCATAGTCAGGAACATCGTATGGGTAATCC | R | XbaI |
| | | ***ATR1Δ51-FLAG; ATR1Δ51-HyP5SM-FLAG* (E128 and E168)** | | |
| GL28 | | tta**ctcgag**ATGGCGCAGACAGCTC | F | XhoI |
| GL29 | | catt**tctaga**TTAGCCTTTGTCGTCATCG | R | XbaI |
| | | ***OsL5*** | | |
| TLG47 | | atta**ctcgag**ATGGGAGGGTTTGTCAAGACCC | F | XhoI |
| TLG52 | | tatgct**actagt**TCACTCATCATCCTCTTCCTCGTC | R | SpeI |
| | | ***OsL5-6xHis*** | | |
| TLG47 | | atta**ctcgag**ATGGGAGGGTTTGTCAAGACCC | F | XhoI |
| TLG80 | | cagtg**actagt**TCAGTGGTGATGATGGTGATGA | R | SpeI |
| *>>> Cloning into pBinAR (into the KpnI/SalI site)* | | | | |
| | | ***avrBs2-HyP5SM-HA* (E/P and E/V)** | | |
| TLG53 | | tta**ggtacc**ATGCGTATCGGTCCTCTGCAACCTTC | F | KpnI |
| TLG54 | | catt**gtcgac**CTACGCATAGTCAGGAACATCGTATGGGTA | R | SalI |

| | | **Primers to make *OsL5-6xHis* from *OsL5* template** | | |
|---|---|---|---|---|
| TLG51 | | atta**ggtacc**ATGGGAGGGTTTGTCAAGACCC | F | KpnI |
| TLG65 | | ctt**gtcgac**TCAGTGGTGATGATGGTGATGACCGGATCCCTCA TCATCCTCTTCCTCGTC | R | SalI |
| *>>> Sequencing primers* | | | | |
| | | **pTA7001** | | |
| TLG39 | | AACGACGGCCAGTGAATTCTCGAAG | F | |
| | | **pBinAR** | | |
| BinFor | | CTATCCTTCGCAAGACCCTTCCTCTATA | F | |
| BinRev | | AATATCATGCGATCATAGGCGTCT | R | |

**TABLE A2:** *Primers used to generate multi-gene pTKan plasmids in Chapter 3.* Color coding is as in Table A1.

| **NAME** | **PLASMID PORTION; OLIGO IN 5'→3' SEQUENCE** | | | **SITE** |
|---|---|---|---|---|
| | | **CaMV 35S promoter** | | |
| YL1 | | CATGGAGTCAAAGATTCAAAT | F | |
| YL2 | | CCCG**cctagg**AGTCCCCCGTGTTCTCTCCAA | R | AvrII |
| | | **pNOS-DsRed2-tNOS** | | |
| YL3 | | ggggatcctctagagGATACATGAGAATTAAGGGAG | F | |
| YL4 | | tccgcggacgtccc**gggccc**GAGCTTGCATGCCGGTCG | R | ApaI |
| | | **GVG-tRbcsE9** | | |
| YL5 | | acgggggactcctag**ctcgagcctagg**ATGAAGCTACTGTCTTCTAT CG | F | XhoI AvrII |
| YL6 | | gagggctgtcacccgAGTGTTTTACTCCTCATATTAACTTCGG | R | |
| | | **p6xUAS promoter** | | |
| YL7 | | gaggagtaaaacactCGGGTGACAGCCCTC | F | |
| YL8 | | acaggctagc**cctagg**CAGCGTGTCCTCTCCAAATG | R | AvrII |
| | | ***OsL5-6xHis*** | | |
| YL9 | | acaagtttgtacaaaaaagcaggcttcATGGGAGGGTTTGTCAAGAC | F | attB1 |
| YL10 | | acaactttgtatagaaaagttgggtgTCAGTGGTGATGATGGTGATG | R | attB4 |
| | | ***ATR1Δ51-HyP5SM-FLAG (E168)*** | | |
| YL11 | | acaactttgtataataaagttggcATGGCGCAGACAGCTC | F | attB3 |
| YL12 | | accactttgtacaagaaagctgggtcTTAGCCTTTGTCGTCATCGTC | R | attB2 |
| | | ***AmpR* for insertion alone** | | |
| YL13 | | cccgggtaccaagctAATTCTTGAAGACGAAAGGG | F | |
| YL14 | | tagactcgag**aagctt**TACCAATGCTTAATCAGTGA | R | HindIII |
| | | ***p35S-Bs2-3xFlag-tNOS* for insertion with *AmpR*** | | |
| YL15 | | cccgggtaccaagctCATGGAGTCAAAGATTCAAATAG | F | |
| YL16 | | ccctttcgtcttcaagaattGATCTAGTAACATAGATGACACC | R | |
| | | ***AmpR* for insertion with *p35S-Bs2-3xFlag-tNOS*** | | |
| YL17 | | ctatgttactagatcAATTCTTGAAGACGAAAGGG | F | |
| YL14 | | tagactcgag**aagctt**TACCAATGCTTAATCAGTGA | R | HindIII |
| | | ***avrBs2-HyP5SM-HA*** | | |
| YL18 | | acaactttgtataataaagttggcATGCGTATCGGTCCTCTGC | F | attB3 |
| YL19 | | accactttgtacaagaaagctgggtaCTACGCATAGTCAGGAACATCG | R | attB2 |
| | | ***eGFP-HyP5SM*** | | |
| YL20 | | acaactttgtataataaagttggcATGTCTAGAGTGAGCAAGGGC | F | attB3 |
| YL21 | | accactttgtacaagaaagctgggtaCTTGT ACAGCT CGTCC ATGCC | R | attB2 |