Development of RNA-based Genetic Control Elements for Predictable Tuning of Protein Expression in Yeast

Thesis by

Andrew Harris Babiskin

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2011

(defended November 29, 2010)

Acknowledgements

First, I would like to express my appreciation to my research advisor, Christina Smolke. Her enthusiasm for her students, her hands-on approach, and her accessibility were major factors in my development as an independent researcher. Her guidance has been immeasurable in the development and the progress of my research projects.

Over the years, I developed close relationships with several colleagues in lab and in the community at the California Institute of Technology. My fellow Smolke lab members have provided a great atmosphere for the facilitation of the sharing of information (and vectors) and for social interaction. Maung Win, Joe Liang, Leo d'Espaux and I spent numerous enjoyable hours together -- whether lunch at Chandler or coffee break at Red Door -- discussing research, politics, pop culture, and whatever else came to mind. I would like to thank Drew Kennedy for spending many half-price burger and chicken sandwich dinners with me during my time at Stanford University. Outside of lab, I developed close friendships with Armin Sorooshian (my roommate of 4 years), Marc Woodka, Ubaldo Córdova-Figueroa, Edgardo García-Berríos, and Muang Win. We enjoyed many visits to all-you-can-eat buffets, our sometimes successful seasons of intramural basketball, our unsuccessful attempt at intramural football, and assembling the greatest Gradiators team of all time. I am lucky to have great colleagues who are also such great friends. Outside of Caltech, I would like to thank Justin Aefsky, Mark Gaylord, and Peter Schmidt. My many trips to San Diego to visit them always gave me a nice, quick weekend away from graduate school. I express my appreciation to my friends back on the east coast who I always had a great time with whenever I was visiting.

For their constant support and love, I would like to thank my parents, Carol and Robert Babiskin, and my sisters, Rachel and Jenny. They have made me everything I am today and my success can be attributed to their involvement. To my deceased grandfather, Julius Babiskin, I would like to acknowledge his doctorate in physics as my source of inspiration for achieving my own doctorate. I would also like to thank my grandparents, Melvin and Lee Rosen, for their constant love and support. Being so far away was tough, but I enjoyed visiting home and going out to lunches with them. Even though they may not understand what I am actually doing, I was always glad to provide them with additional bragging material.

Last but not least, I would like to thank my wife, Rosario Babiskin, whom I love deeply. She has been with me for almost my entire graduate career. Because of her, I had a family away from home with whom I could spend numerous holidays with and eat ridiculous amounts of amazing food. We share together the cutest dog in the world, Cleo. When I moved away to Stanford, it was tough on the both of us being apart and not knowing when we would live together again, but she stayed strong. Though that period was difficult, we now have memories that we will cherish forever, including our wedding. Having her back living with me after our marriage has been amazing and her support and understanding has been invaluable in these final months as I finished my doctoral research and wrote my thesis.

Development of RNA-based Genetic Control Elements for Predictable Tuning of Protein Expression in Yeast

Andrew Harris Babiskin

B.S., University of Maryland

M.S., California Institute of Technology

Ph.D., California Institute of Technology

Abstract

The proper functioning of many biological processes and synthetic genetic networks depends on the precise tuning of expression levels of key protein components. With growing interests in eukaryotic hosts and the increasing complexity of networks in synthetic biology, there is a need for the expansion of the genetic toolbox, particularly for the bioprocessing and biosynthesis applications in the yeast *Saccharomyces cerevisiae*. The available control elements in yeast generally focus on the regulation of transcription through alternative promoter systems. Synthetic RNA-based control elements placed in the untranslated regions (UTRs) of transcripts have the ability to regulate the posttranscriptional mechanisms of translation initiation and transcript stability. Such posttranscriptional elements have the added advantage of being coupled to any promoter for enhanced control strategies.

Two types of posttranscriptional elements were examined in this thesis. The first type is a class of RNA hairpins baring AGNN tetraloops that are cleaved by the *S. cerevisiae* RNase III enzyme Rnt1p. By locating these hairpins in the 3' UTR of a

transcript, the endonucleolytic cleavage due to Rnt1p activity resulted in the rapid degradation of the transcript. We developed two libraries of RNA hairpins based on the randomization of critical regions in Rnt1p substrates that affect the enzyme's ability to associate and/or cleave the hairpin. The modulation of the strength of binding and cleavage by Rnt1p resulted in changes in the steady-state transcript levels and thus protein levels. Through integration of an aptamer into the stem of an Rnt1p hairpin, we were able to develop a riboswitch based upon the direction inhibition of Rnt1p cleavage through association of the ligand in the sites of cleavage. The second type of posttranscriptional elements examined is the placement of internal ribosome entry sites (IRESes) in the 5' UTR that initiate translation independent of the 5' cap through direct interaction with the ribosomal machinery. We propose that the activity of small sequential IRESes can be tuned through varying the complementarity with the 18S ribosomal RNA (rRNA) to advance the creation of yeast multicistronic vectors. The application of Rnt1p hairpins and IRESes provide a key tool in synthetic biology for the construction of complex genetic networks in yeast where the predictable tuning of gene expression is necessitated.

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

1.1. Synthetic biology and metabolic engineering in *Saccharomyces cerevisiae*

Synthetic biology is an emerging field that joins biology and engineering to design and build new biological systems exhibiting desired functions, such as the biosynthesis of drugs and biofuels in microorganisms and genetic therapies that can target diseased cells in humans[1-4]. Synthetic biologists have focused on the development and application of genetic tools and engineering principles to design and implement synthetic gene networks and the rewiring or reprogramming of endogenous cellular networks[5]. Developed genetic regulatory tools function in the cellular environment to control transcriptional, posttranscriptional, and posttranslational processes. Precise levels of gene expression are critical for the proper functioning of genetic networks[6-8]. As complexity increases with the size of engineered networks, there is a growing need for control elements that allow for the fine-tuning of the levels of protein components in the network[9]. There is a particular need for the development of genetic regulatory tools that function in eukaryotes, as the majority of devices to date have been built in prokaryotes.

Metabolic engineering is defined as the redirection of cellular metabolism for the production of valuable chemicals and the removal of harmful or toxic compounds from the environment[10]. Research in this area often involves the implementation of gene expression tools to precisely control enzyme levels and thus regulate flux through natural or heterologous pathways[11-13]. While synthetic chemistry has traditionally been the main method used to synthesize chemicals for a wide variety of industries, many chemicals,

particularly ones with multiple chiral carbon centers, have proven extremely difficult to synthesize through these traditional methods[14]. Metabolic engineering addresses these challenges by utilizing enzymes to perform chemical conversions, which generally exhibit stereospecificity, thereby resulting in the efficient production of chiral products. Enzymatic reactions performed inside cells offer several advantages over *in vitro* based systems in that cells can be used to generate and replenish the desired enzymes and necessary cofactors from inexpensive starting materials and provide appropriate precursor chemicals[15]. However, the redirection and construction of cellular metabolic networks is not as straightforward as cloning the genes that encode the appropriate enzymes into the cell. Cellular productivity can be negatively impacted by metabolic burden associated with enzyme overexpression[16-17], the accumulation of cytotoxic intermediates[12, 18-19], and the redirection of cellular resources from central metabolism[20-22]. The tuning of enzymes levels has been found to be crucial for optimizing metabolic flux to alleviate these detrimental issues and achieve the desired function, namely increased product yield[12, 16, 23-24].

There are many examples of plants and other higher-level organisms that naturally produce chemicals that are of interest to various industries[25-27]. In particular, many plant species produce compounds with diverse pharmacological activities that are of interest as drug molecules[25]. These compounds have been traditionally extracted from their natural hosts. However, higher-order eukaryotic cells have very long doubling times and, due to differentiation, not every cell necessarily produces the product of interest. For example, natural products of interest in plants have been found to amass at low quantities and extraction procedures can be difficult due to the production of other chemically

similar compounds and the use of toxic solvents[28-29]. In addition, there are additional costs associated with the land and resources (including manpower) required to grow plants. It is desirable to transfer the ability to make these chemicals into organisms that grow more rapidly on inexpensive energy sources to lower the cost of these compounds. The construction of a biosynthetic network begins with the selection of the appropriate organism that naturally produces required intermediates or demonstrates similar chemistries[30]. A common tactic is to reconstitute the system in common host organisms used in industrial fermentation applications: a bacterium, *Escherichia coli*; and a eukaryotic microorganism, *Saccharomyces cerevisiae*.

Although *E. coli* is robust and fast growing, there are limitations in its ability to effectively express enzymes from eukaryotic host organisms. Many of these problems arise from differences in the protein expression pathways between bacteria and eukaryotic organisms. For instance, posttranslational processes such as glycosylation and the localization of enzymes to intracellular membranes are present in eukaryotes but not in prokaryotes. Therefore, *S. cerevisiae* can overcome these deficiencies present in *E. coli*, while having advantages over other eukaryotic cell lines due to its small fully-sequenced genome, fast doubling time, ability to grow in inexpensive chemically defined medium, and ease of scale-up to fermentation vessels similar to those used to grow *E. coli*[31-32]. Additional advantages of *S. cerevisiae* are associated with the accumulated knowledge of the organism's genetics, physiology, and biochemistry, its classification as GRAS (generally regarded as safe) by the U.S. Food and Drug Administration (FDA), and its tolerance at low pH levels and high concentrations of sugar and ethanol[33].

Enzyme levels are typically regulated by controlling the copy number of heterologous genes, transcription efficiency, translation efficiency, transcript abundance, and protein abundance[33]. In *S. cerevisiae*, very few genetic tools exist to control transcript levels and the translation of transcripts. The majority of genetic tools developed to date have focused on the incorporation of different endogenous promoter systems or the re-engineering of promoters to modulate the transcriptional output or the response to factors of transcriptional activation[34-38]. Posttranscriptional elements have the advantage of being coupled to any promoter of choice, providing for enhanced control strategies. Internal ribosome entry sites (IRESes) and AU-rich elements (AREs) have demonstrated the ability to modulate gene expression in yeast, while more recently, antisense- and ribozyme-based riboswitches have shown the ability to enhance or repress gene expression due to presence of a small molecule effector[39-43].

In the following sections of the Introduction, a detailed explanation of eukaryotic posttranscriptional mechanisms will be provided. Numerous RNA-based elements that regulate or bypass these mechanisms will be described. Finally, two specific regulatory elements, Rnt1p hairpin substrates and IRESes, will be described, including their function in *S. cerevisiaie*.

1.2. Common pathways of transcriptional decay and translation in yeast

The cellular processes of transcription, translation, and transcript turnover are common across all eukaryotes and prokaryotes. However, the eukaryotic gene expression pathway is more complex and contains intermediate steps between transcription and translation that provide further mechanisms of control such as splicing, transcript editing,

and nuclear export. In eukaryotes, a mature transcript is formed through a series of coupled processing events (Figure 1.1). Initially, a pre-messenger RNA (pre-mRNA) is transcribed from a gene by RNA polymerase II. The first processing step in the production of a translationally-competent transcript is the addition of a 5' cap, which contains a methylated guanine nucleotide[44]. The purpose of the cap is to protect the transcript from degradation by 5' to 3' exonucleases and to stimulate the initiation of translation[45]. The next processing step is the removal of introns, intervening noncoding sequences found within the coding region, to form the mature transcript through a process called splicing[44]. The final step before nuclear export to the cytoplasm is a 3' end modification in which a poly(A) tail is added. The transcript is then exported to the cytoplasm where it undergoes cytoplasmic decay or translation to produce protein molecules. Cells control the level of proteins by regulating each one of these steps, from the chromatin remodeling necessary for transcription of many genes to posttranslational protein stability.

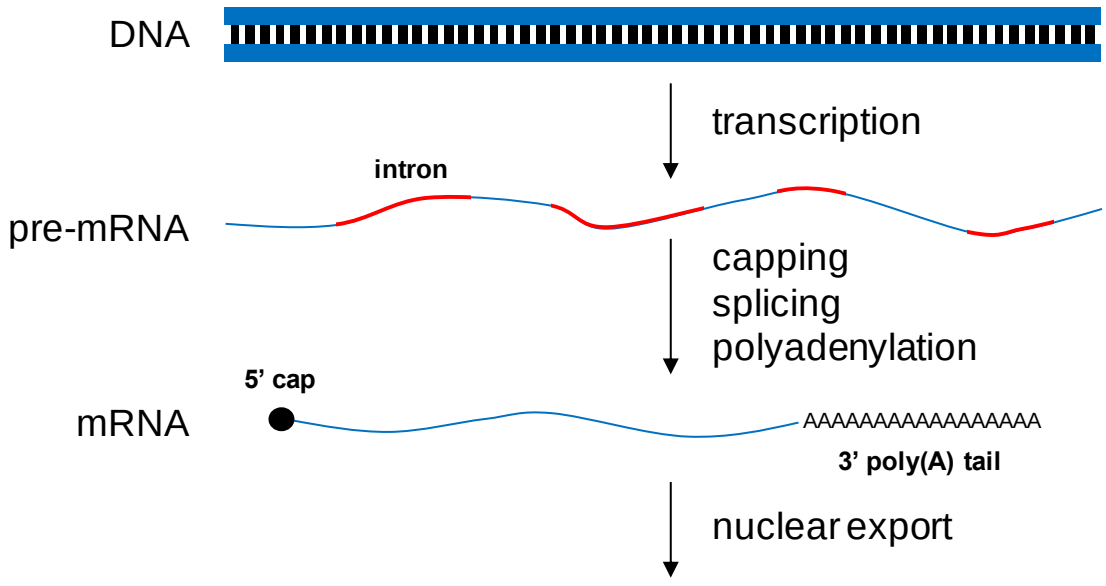


Figure 1.1. Maturation of eukaryotic transcripts after transcription.

1.2.1. Deadenylation-dependent decapping pathway of transcript degradation

There are several mechanisms by which transcripts are degraded, including the deadenylation-dependent decapping pathway, the deadenylation-independent decapping pathway, and the endoribonucleolytic cleavage pathway. In *S. cerevisiae*, the most common degradation pathway is deadenylation-dependent decapping.

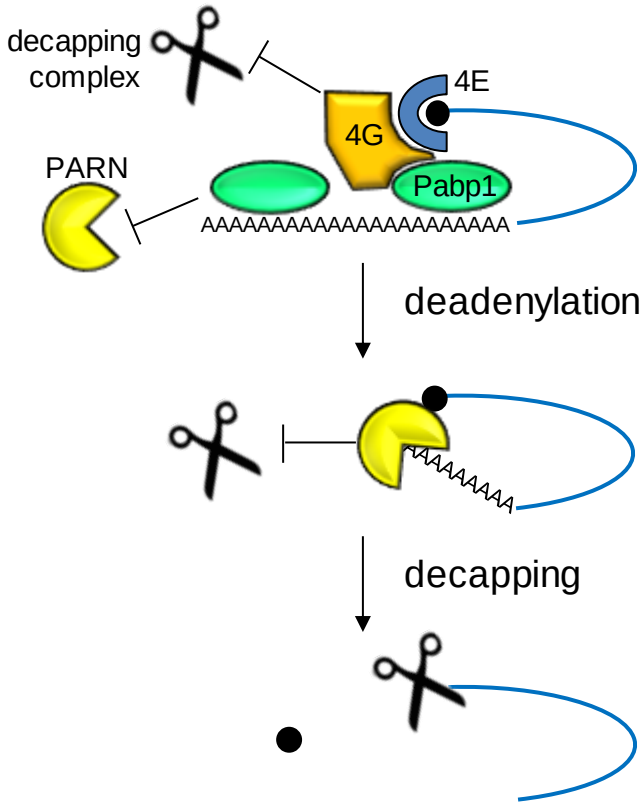


Figure 1.2. The deadenylation-dependent decapping pathway of transcript degradation in eukaryotes. 4E denotes eIF4E and 4G denotes eIF4G. Adapted from Wilusz et al. (2001)[46].

Transcripts are present in a circular conformation due to the interaction of the cap-binding protein, eukaryotic transcription factor (eIF)4E, on the 5' cap and the poly(A) binding protein (Pab1p) on the 3' poly(A) tail mediated through binding to eIF4G, a scaffolding protein (Figure 1.2). This circularization promotes translation and prevents the activity of decapping and deadenylation enzymes[46]. The interaction of Pab1p

with the poly(A) tail inhibits deadenylation[47]. When the poly(A) ribonuclease (PARN) binds to the 5' cap, it disrupts the cap's interaction with eIF4E causing a displacement of Pab1p, which allows deadenylation to occur[46, 48]. Deadenylation is the rate-limiting step in transcript decay[49]. While there are several different deadenylases that could be functioning[50], the predominant form in yeast is Ccr4p[51]. Once deadenylation is completed, PARN no longer stays associated with the cap and the decapping complex of Dcp1 and Dcp2 cleaves off the cap allowing a 5' to 3' exonuclease (Xrn1p) to rapidly degrade the rest of the transcript[46-47, 52]. 3' to 5' exonucleolytic activity does occur after deadenylation, but it tends to be slower than the activity of Xrn1p[49, 53].

1.2.2. Cap-dependent translation initiation

Translation initiation begins when the 40S small ribosomal subunit associates with two eIFs, eIF2 and eIF3, and the initiator methionine tRNA to form the 43S preinitiation complex (Figure 1.3)[54]. eIF2 must also be bound by guanosine 5'-triphosphate (GTP) in order for it to associate with the 40S ribosome. On the transcript, the 5' cap is bound by a cap-binding protein complex, eIF4F, which consists of three subunits: eIF4A, an RNA helicase; eIF4E, the actual cap-binding protein; and eIF4G, a scaffolding protein[55-56]. The transcript is initially in a closed, circular form due to eIF4G's interactions with both eIF4E at the 5' cap and Pab1p at the 3' poly(A) tail. The 43S complex binds eIF4F to form the 48S complex and scanning of the transcript begins for the initiating AUG start codon[57]. The scanning by the complex for AUG is caused by more initiation factors powered by ATP. Once the start codon is located, eIF1 and eIF1A bind to stabilize the binding of the 48S complex to the transcript[56]. eIF5 then stimulates

the hydrolysis of GTP on eIF2 followed by the association of the large 60S ribosomal subunit with the 48S complex to form the complete 80S ribosome[44, 56]. The complex is then in a form where translation can initiate and proceeds to the elongation stage.

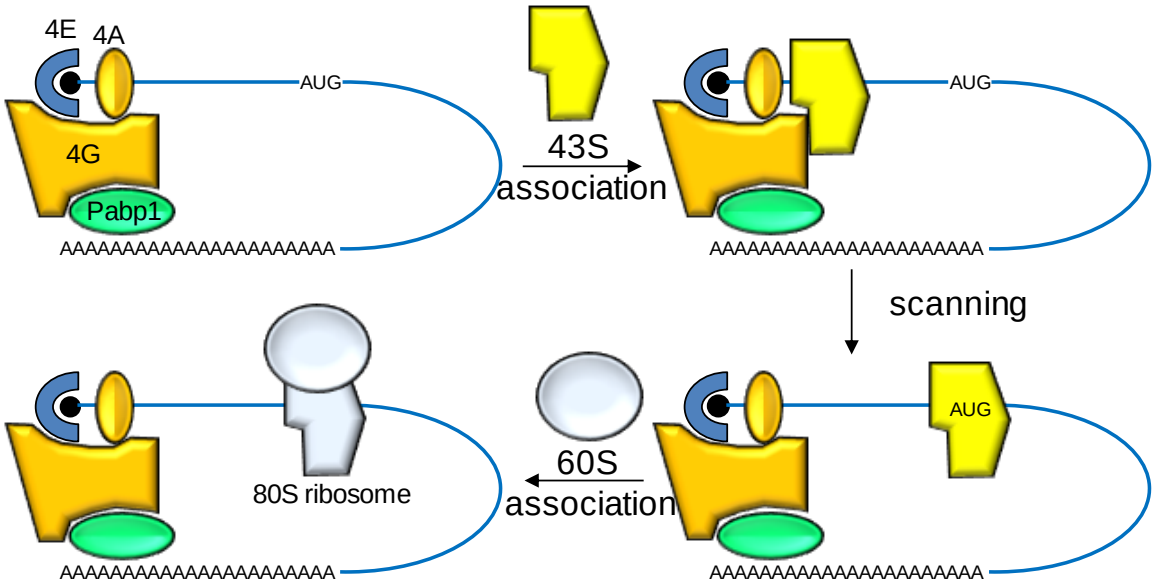


Figure 1.3. The mechanism of eukaryotic cap-dependent translation. 4A denotes eIF4A; 4E denotes eIF4E; and 4G denotes eIF4G. Adapted from Klann and Dever (2004)[58].

1.3. Posttranscriptional regulation through transcript stability and translation

1.3.1. Control of transcript decay

As discussed in Section 1.2.1, the deadenylation-dependent decapping pathway is the primary mechanism by which eukaryotic transcripts are degraded. Briefly, transcripts contain a 3' poly(A) tail whose interactions with proteins inhibit decapping. Once the tail is removed, decapping proceeds and the transcript is degraded 5' to 3' by an exonuclease. Eukaryotes have evolved additional non-coding elements in their genetic untranslated regions (UTRs) and sometimes in the actual coding regions by which this pathway can be

attenuated or bypassed[52]. These elements include deadenylation-independent decapping elements, transcript stability elements, and elements conveying endonucleolytic activity.

In the deadenylation-independent decapping pathway, transcript degradation proceeds without the removal of the 3' poly(A) tail. This pathway proceeds through the recruitment of elements that enhance decapping. The *RPS28B* transcript in *S. cerevisiae* contains a stem-loop structure within its 3' UTR and encodes for a protein, Rps28B, that binds directly to that stem-loop[59]. The Rps28B protein product recruits a decapping-enhancing protein, Edc3, and Edc3, in turn, recruits several other factors that lead to the decapping of the transcript. The *EDC1* transcript in *S. cerevisiae* encodes for a decapping-enhancing protein, Edc1, though it is unknown if Edc1 plays a role in *EDC1* degradation[52]. *EDC1* contains a stretch of uridine nucleotides that interacts with the poly(A) tail inhibiting deadenylation[60]. The decapping of the transcript is caused by several protein factors including those associated with deadenylation.

Transcript stability elements are located at multiple positions on the transcript, but primarily in the 3' UTR[52]. The largest class of elements that has been examined is the ARE in the 3' UTR. AREs are identified by a consensus AUUUA pentamer, but its activity is dependent on the context and number of those pentamers[52]. AREs can destabilize the transcript through the interaction of the sequence itself or ARE-binding proteins with the transcript decay protein complex[61-62]. AREs can also stabilize transcripts, where proposed mechanisms are based on competition with destabilizing factors or inhibition of deadenylation-dependent decapping decay, such as through strengthening the interaction between PABP and the poly(A) tail[52]. In addition, the PUF family of proteins binds to UG-rich sequences found in the 3' UTR of transcripts.

Destabilization by PUF binding is due to recruitment of deadenylases[63]. As another example, the proteins αCP1 and αCP2 are responsible for the stabilization of several genes through interactions with pyrimidine-rich elements in the 3' UTR. It is believed that the observed transcript stabilization is due to interactions with PABP that protect the poly(A) tail from deadenylases[64].

Endoribonucleolytic decay can also be described as deadenylation-independent and decapping-independent decay. Internal cleavage of the transcript results in the generation of two RNA fragments with unprotected ends (Figure 1.4)[52]. The 3' fragment is susceptible to exonucleolytic decay by Xrn1p in the 5' to 3' direction, while the 5' fragment is degraded in the same manner once the cap is removed or by 3' to 5' exonucleases. In eukaryotes containing the RNA interference (RNAi) pathway, gene expression is modulated through directed endonucleolytic cleavage, referred to as "Slicer" activity, of the target transcript[65]. Cleavage is mediated through components of the RNA-induced silencing complex (RISC), which contains the RNase III enzyme variant Dicer[66-68]. MicroRNAs (miRNAs) or small interfering RNAs (siRNAs) are loaded onto RISC and direct the complex to the transcript through perfect or nearly perfect (some mismatches allowed) base-pairing between the transcript and the miRNA/siRNA[69-71]. There are numerous endoribonucleases that regulate expression levels in eukaryotes, although for many of the enzymes, such as PRM1, IRE1, and RNase MRP, *cis*-acting consensus binding regions have not yet been determined[72-74]. As an alternative to the endonucleolytic cleavage caused by *trans* factors, transcript degradation in a diverse group of eukaryotes can also be mediated through *cis* self-cleaving catalytic RNA structures called ribozymes[75]. The RNase III variant Drosha and the *S. cerevisiae*-specific

RNase III variant Rnt1p has been shown to cleave transcripts containing stem loop structures[76-77]. Rnt1p specifically processes transcripts containing hairpins with AGNN tetraloops[78] and is explained in further detail in Section 1.3.3.

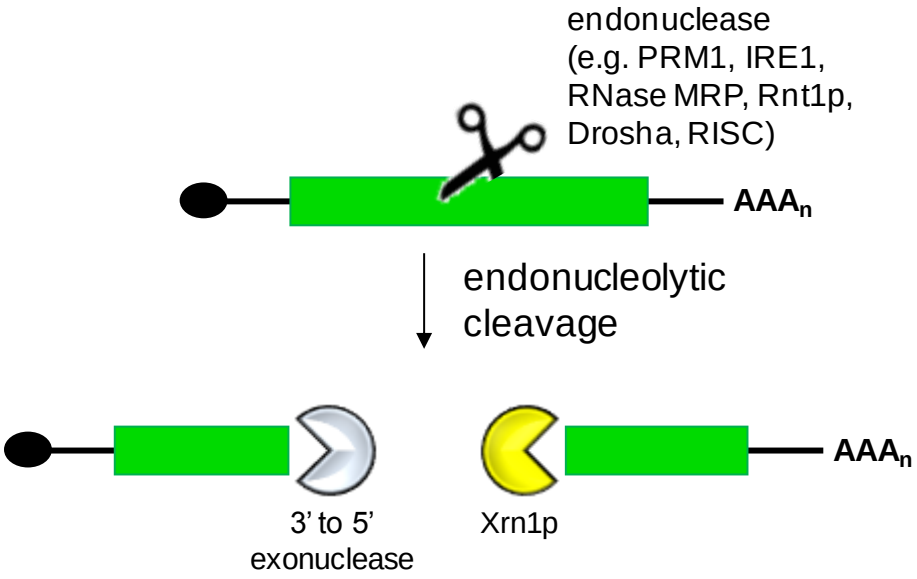


Figure 1.4. The processing of eukaryotic transcripts following endonucleolytic cleavage is independent of the 5' cap and the 3' poly(A) tail. Scissors denote the endonuclease. Adapted from Garneau et al. (2007)[52].

1.3.2. Control of the initiation of translation

As discussed in Section 1.2.2, the initiation of translation in eukaryotes is mediated through protein interactions at the 5' cap between eIFs and the small ribosomal subunit. The preinitiation complex scans the transcripts for the AUG start codon, where the large ribosomal subunit binds and translation begins. The majority of translational control is due to interference with the normal processes of the ribosome and the eIFs (Figure 1.5)[79]. Cap-independent translation by IRESes is described in its own section (1.3.4). The majority of these elements are located in the 5' UTR.

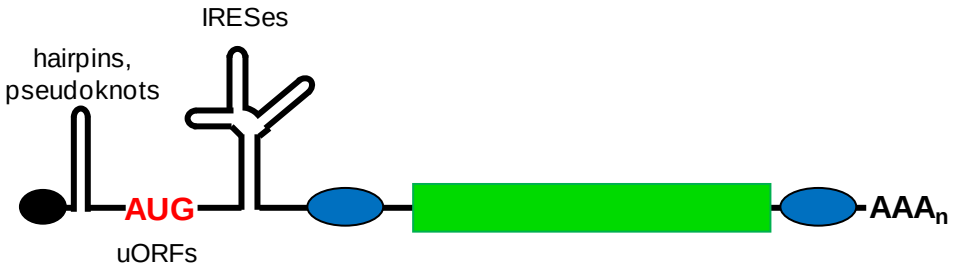


Figure 1.5. Genetic elements that affect the initation of translation in eukaryotes. The blue ovals in the 5' and 3' UTRs represent binding sites for protein factors that typically inhibit translation. Adapted from Gebauer and Hentze (2004)[79].

Translational repression for the ferritin transcript is mediated through a stem-loop structure called an iron-responsive element (IRE) located 40 nucleotides from the cap in the 5' UTR[80-81]. Iron regulatory proteins (IRPs) bind the IRE blocking the recruitment of the preinitiation complex to the 5' cap due to steric hinderance[82]. Steric hinderance is also observed with secondary and tertiary structures in the 5' UTR, such as RNA hairpins and pseudoknots[79]. Transcripts containing a U-rich sequence known as a cytoplasmic polyadenylational element (CPE) in the 3' UTR interact with the CPE-binding protein (CPEB)[83]. CPEB represses translation by associating with another protein, Maskin, that contains an eIF4E-binding domain, which inhibits eIF4E's interaction with eIF4G[84]. Translational repression can also interfere with the ribosome after it has been bound to the cap. The sex-lethal protein (Sxl) binds to U-rich sites on the *msl-2* transcript located in both the 5' and 3' UTR[85]. The binding of Sxl interferes with ribosomal scanning. Another cap-independent method interferes with the association of the large ribosomal submit in *LOX3* transcripts[86]. Here, two proteins, hnRNP K and hnRNP E, bind a differentiation-control element (DICE), which is a repeated CU-rich element, in the 3' UTR and block formation of the 80S ribosome. In the previous section, we discussed endonucleolytic cleavage mediated by the interaction of siRNA and miRNA to RISC.

The miRNA silencing pathway is also known to repress translation through direct or indirect interference with eIFs[65, 87].

1.3.3. RNA processing by the RNase III enzyme Rnt1p

The RNase III family is a class of enzymes that cleaves double-stranded RNA (dsRNA)[88]. Dicer is an RNase III enzyme in humans and other eukaryotes that cleaves dsRNA into 21–23 nt fragments referred to as siRNAs that go on to induce gene silencing though the RNAi pathway[89]. Drosha, another eukaryotic RNase III involved in the RNAi pathway, is involved in the processing of miRNA from long dsRNA transcripts referred to as primary (pri-)miRNA[90]. Rnt1p was discovered in *S. cerevisiae* due to similarities to the *E. coli* RNase III[91] and has been shown to cleave cellular ribosomal RNA (rRNA) precursors, small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), and messenger RNA (mRNA)[91-94]. This protein is localized to the nucleus[95] and contains two domains: an RNase III domain and a dsRNA-binding domain (dsRBD)[96].

The RNA hairpin substrates of Rnt1p contain a consensus AGNN tetraloop with a cleavage site 14–16 base-pairs (bp) from the tetraloop[78]. The AGNN tetraloop forms a predetermined fold that is recognized by the dsRBD[88, 97]. The dsRNA region of Rnt1p substrates has an effect on the binding affinity and cleavage rate with this enzyme. The base-pairs immediately below the tetraloop can impact Rnt1p binding, while sequences near the cleavage site influence the cleavage rate[98]. These observations led to the definition of three regions on Rnt1p substrates: the initial binding and positioning box (IBPB) which consists of the tetraloop; the binding stability box (BSB) which is the base-paired region immediately adjacent to the tetraloop; and the cleavage efficiency box

(CEB) which is the region containing and surrounding the cleavage site[98]. An unique feature of Rnt1p is that it uses the tetraloop as its primary binding site, whereas for other RNase III enzymes it is the RNA helix[99].

Rnt1p is involved in the natural regulation of several genes in *S. cerevisiae*. Previously, transcripts had been discovered that undergo endonucleolytic cleavage but by factors other than Rnt1p[100-101]. The *RPS22B* and *RPL18A* transcripts contain intronic Rnt1p substrates that deplete unspliced transcripts as well as reducing levels of the mature transcripts[94]. The *MIG2* transcript contains an Rnt1p substrate in its coding region that increases the transcript's sensitivity to glucose-dependent degradation[77]. Several transcripts involved in iron uptake or iron mobilization contain Rnt1p substrates in the coding region that help avoid cytotoxicity to the cellular iron starvation reponse[102]. The diversity of structure and sequences in natural Rnt1p substrates, as well as the identification of critical regions, support that a set of engineered Rnt1p hairpins can be generated with differential activity.

1.3.4. Translation initiation mediated through internal ribosome entry sites

IRESes were initially discovered during the analysis of the 5' UTRs of picornaviral transcripts where it was determined that the transcripts lacked a 5' cap and translation continued in the absence of the cap-binding protein, eIF4F[103]. IRESes are critical elements for the translation of the genome of several viruses, including the Hepatitis A virus (HAV)[104], the Hepatitis C virus[105], the foot-and-mouth-disease virus (FMDV)[106], and the human immunodeficiency virus (HIV)[107]. Viral IRESes contain a diverse range of secondary and tertiary structures that mimic components of the

ribosomal machinery or mimic the interaction of such components with the 5' cap and other protein factors (Figure 1.6)[108-113]. IRESes illustrate the scaffolding power of RNA structures and the creative mechanisms by which viruses have evolved to essentially hijack the host-based expression machinery.

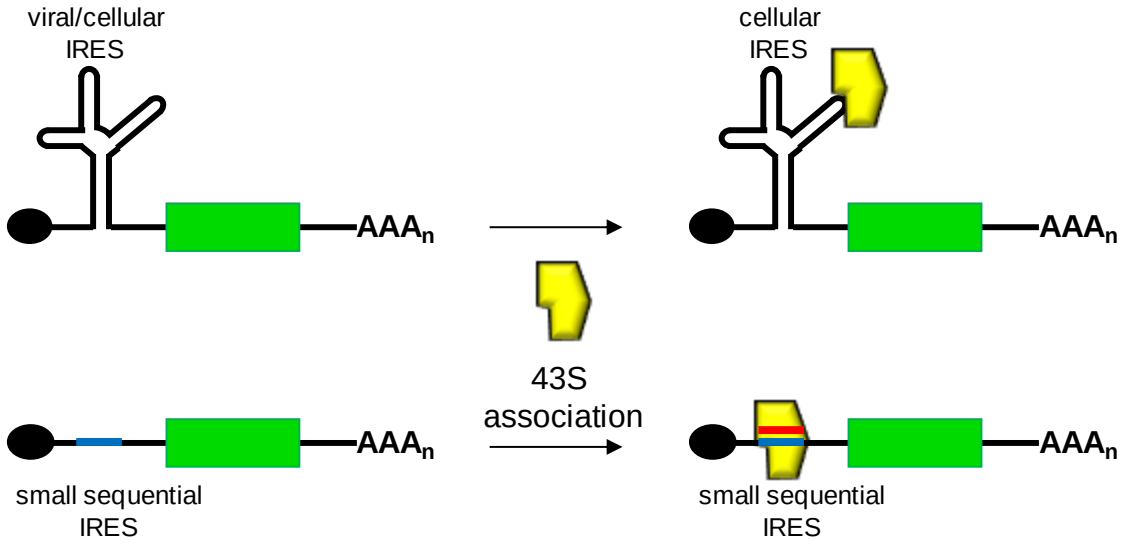


Figure 1.6. Simplified schematic of the interactions with structural and sequential IRESes with the translational machinery. Red line denotes the sequences on the 18S rRNA complementary to the IRES sequence (blue line).

Cellular IRESes were first discovered when researchers observed that immunoglobulin heavy chain binding protein (BiP) continued being expressed after cap-dependent translation had been shut down due to poliovirus infection[109]. Many cellular IRESes characterized thus far contain a Y-shaped stem-loop structure upstream of the initiation codon; however, the activity of these IRESes may not necessarily depend on that secondary structure[114-115]. In a mouse cell line, deletional studies of the structured 5' UTR of the Gtx protein, which demonstrated IRES activity, identified a 9 nucleotide (nt) module that retained the ability to internally initiate translation[114]. When multiple modules of the 9-nt module were placed in tandem, a synergistic effect was observed as

overall IRES activity increased. This 9-nt segment was determined to be completely complementary to a segment of the 18S ribosomal RNA, a critical component of the ribosomal machinery (Figure 1.6)[116]. In *S. cerevisiae*, two IRES-containing 5' UTRs for the *YAP1* and *p150* genes were also found to contain several regions of complementarity to 18S rRNA[39]. These studies have demonstrated that cellular IRESes contain regions that directly base-pair to the 18S rRNA. This mechanism of translation initiation by cellular IRESes in eukaryotes suggest that their function is analogous to Shine-Dalgarno sequences in prokaryotes, which initiate translation through direct base-pairing with the prokaryotic analogue of the 18S rRNA, the 16S rRNA[117]. Based on this observation, a short segment of nucleotides in the intercistronic region (IR) of a yeast and mammalian dicistronic vector were randomized and screened for IRES activity by expression of the second cistron[118-119]. The resultant IRESes demonstrated complementarity to the 18S rRNA.

Prokaryotic genes are typically expressed from operons, where multiple coding regions are located on one transcript under the control of a single promoter. Each coding region contains a Shine-Dalgarno sequence upstream of its start codon in order to initiate translation of each gene. Viruses are also known to produce multicistronic transcripts or genomes. For example, the entire positive-strand genome of HCV is contained on a single piece of RNA[120]. The entire genome is translated through an IRES at the 5' end. The resultant polyprotein is then processed by a series of proteases and peptidases to create each individual protein product. HIV translation is similar to HCV except that its IRES can also cause translation initation at multiple start codons resulting in different protein products[121]. In a manner akin to prokaryotic operons, multicistronic transcripts

can be generated through the introduction of an IRES element before each gene. Retroviral mulitcistronic vectors had been developed in mammalian systems where multiple viral IRESes were incorporated[122-123]. Recently, a dicistronic reporter construct had been characterized in *S. cerevisiae* where the *YAP1* and *p150* IRES were placed in the IR to alter the ratio of expression between two genes[124]. Since only the *p150* IRES worked in this system, the work highlights the need for additional IRESes to be discovered or engineered in yeast to increase the ability to tune gene expression through this method.

1.4. Interrelationship among the thesis projects

Chapter I provides an overview of the field of synthetic biology and metabolic engineering and gives an in-depth examination of the cellular processes of transcript translation initiation and decay and the RNA elements that control these processes. Chapter II describes the development of a library of RNA hairpins that regulate posttranscriptional decay to attenuate gene expression due to the endonucleolytic processing of the hairpins by the *S. cerevisiae* RNase III Rnt1p. The library is based on the randomization of nucleotides associated with controlling the cleavage rate by the enzyme. Chapter III describes a second library of Rnt1p-cleaved hairpins based on the randomization of nucleotides associated with the binding of Rnt1p to the hairpin. In addition, the two library elements are integrated combinatorially to extend the accessible levels of gene expression. Chapter IV describes the integration of the small molecule-responsive aptamers into Rnt1p substrates to achieve ligand-controlled cleavage. The engineered riboswitches function through direct inhibition of Rnt1p activity by ligand

binding in proximity of the cleavage sites and switching dynamics are altered through incorporation of additional aptamers and Rnt1p-based modules, as well as the construction of multiple switch devices in tandem. Chapter V describes a strategy to select for a library of small sequential IRES elements with various strengths to initiate translation at physiological conditions. These elements will aid in the development of yeast 'operons' or multicistronic vectors where relative gene expression levels can be controlled. These research projects collectively demonstrate the capacity of utilizing RNA-based control elements to predictably tune gene expression levels in *S. cerevisiae*.

References

1. Chen, Y.Y., Jensen, M.C. & Smolke, C.D. Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc Natl Acad Sci U S A* **107**, 8531-8536 (2010).
2. Culler, S.J., Hoff, K.G. & Smolke, C.D. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science* **330**, 1251-1255 (2010).
3. Lee, S.K., Chou, H., Ham, T.S., Lee, T.S. & Keasling, J.D. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr Opin Biotechnol* **19**, 556-563 (2008).
4. Lee, S.Y., Kim, H.U., Park, J.H., Park, J.M. & Kim, T.Y. Metabolic engineering of microorganisms: general strategies and drug production. *Drug Discov Today* **14**, 78-88 (2009).

5. Andrianantoandro, E., Basu, S., Karig, D.K. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol* **2**, 2006 0028 (2006).

6. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338 (2000).

7. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339-342 (2000).

8. Basu, S., Mehreja, R., Thiberge, S., Chen, M.T. & Weiss, R. Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A* **101**, 6355-6360 (2004).

9. Khalil, A.S. & Collins, J.J. Synthetic biology: applications come of age. *Nat Rev Genet* **11**, 367-379 (2010).

10. Smolke, C., Martin, V. & Keasling, J. in Protein expression technologies: current status and future trends. (ed. F. Baneyx) (Horizon Bioscience, Wymondham, Norfolk; 2004).

11. Pitera, D.J., Paddon, C.J., Newman, J.D. & Keasling, J.D. Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. *Metab Eng* **9**, 193-207 (2007).

12. Pfleger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* **24**, 1027-1032 (2006).

13. Peralta-Yahya, P.P. & Keasling, J.D. Advanced biofuel production in microbes. *Biotechnol J* **5**, 147-162 (2010).

14. Hughes, E.H. & Shanks, J.V. Metabolic engineering of plants for alkaloid production. *Metab Eng* **4**, 41-48 (2002).

15. Smolke, C.D., Martin, V.J. & Keasling, J.D. in Protein Expression Technologies: Current Status and Future Trends. (ed. F. Baneyx) (Horizon Scientific Press, U.K.; 2004).

16. Jin, Y.S., Ni, H., Laplaza, J.M. & Jeffries, T.W. Optimal growth and ethanol production from xylose by recombinant Saccharomyces cerevisiae require moderate D-xylulokinase activity. *Appl Environ Microbiol* **69**, 495-503 (2003).

17. Jones, K.L., Kim, S.W. & Keasling, J.D. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab Eng* **2**, 328-338 (2000).

18. Zhu, M.M., Lawman, P.D. & Cameron, D.C. Improving 1,3-propanediol production from glycerol in a metabolically engineered Escherichia coli by reducing accumulation of sn-glycerol-3-phosphate. *Biotechnol Prog* **18**, 694-699 (2002).

19. Zhu, M.M., Skraly, F.A. & Cameron, D.C. Accumulation of methylglyoxal in anaerobically grown Escherichia coli and its detoxification by expression of the Pseudomonas putida glyoxalase I gene. *Metab Eng* **3**, 218-225 (2001).

20. Alper, H., Jin, Y.S., Moxley, J.F. & Stephanopoulos, G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. *Metab Eng* **7**, 155-164 (2005).

21. Alper, H., Miyaoku, K. & Stephanopoulos, G. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* **23**, 612-616 (2005).

22. Paradise, E.M., Kirby, J., Chan, R. & Keasling, J.D. Redirection of flux through the FPP branch-point in Saccharomyces cerevisiae by down-regulating squalene synthase. *Biotechnol Bioeng* **100**, 371-378 (2008).

23. Dueber, J.E. et al. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* **27**, 753-759 (2009).

24. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* **102**, 12678-12683 (2005).

25. Morant, M., Bak, S., Moller, B.L. & Werck-Reichhart, D. Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Curr Opin Biotechnol* **14**, 151-162 (2003).

26. Veen, M. & Lang, C. Production of lipid compounds in the yeast Saccharomyces cerevisiae. *Appl Microbiol Biotechnol* **63**, 635-646 (2004).

27. Hughes, E.H., Hong, S.B., Gibson, S.I., Shanks, J.V. & San, K.Y. Metabolic engineering of the indole pathway in Catharanthus roseus hairy roots and increased accumulation of tryptamine and serpentine. *Metab Eng* **6**, 268-276 (2004).

28. Chemler, J.A. & Koffas, M.A. Metabolic engineering for plant natural product biosynthesis in microbes. *Curr Opin Biotechnol* **19**, 597-605 (2008).

29. Silva, G.L., Lee, I. & Kinghorn, A.D. in Natural Products Isolation. (ed. R.J.P. Cannell) (Humana Press, Totowa, New Jersey; 1998).

30. Alper, H. & Stephanopoulos, G. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat Rev Microbiol* **7**, 715-723 (2009).

31. Cereghino, G.P. & Cregg, J.M. Applications of yeast in biotechnology: protein production and genetic analysis. *Curr Opin Biotechnol* **10**, 422-427 (1999).

32. Gerngross, T.U. Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nat Biotechnol* **22**, 1409-1414 (2004).

33. Nevoigt, E. Progress in metabolic engineering of Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **72**, 379-412 (2008).

34. Hawkins, K.M. & Smolke, C.D. The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae. *J Biol Chem* **281**, 13485-13492 (2006).

35. Nevoigt, E. et al. Engineering promoter regulation. *Biotechnol Bioeng* **96**, 550-558 (2007).

36. Nevoigt, E. et al. Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. *Appl Environ Microbiol* **72**, 5266-5273 (2006).

37. Mumberg, D., Muller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156**, 119-122 (1995).

38. Jeppsson, M., Johansson, B., Jensen, P.R., Hahn-Hagerdal, B. & Gorwa-Grauslund, M.F. The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant Saccharomyces cerevisiae strains. *Yeast* **20**, 1263-1272 (2003).

39. Zhou, W., Edelman, G.M. & Mauro, V.P. Transcript leader regions of two Saccharomyces cerevisiae mRNAs contain internal ribosome entry sites that function in living cells. *Proc Natl Acad Sci U S A* **98**, 1531-1536 (2001).

40. Vasudevan, S. & Peltz, S.W. Regulated ARE-mediated mRNA decay in Saccharomyces cerevisiae. *Mol Cell* **7**, 1191-1200 (2001).

41. Lautz, T., Stahl, U. & Lang, C. The human c-fos and TNFalpha AU-rich elements show different effects on mRNA abundance and protein expression depending on the reporter in the yeast Pichia pastoris. *Yeast* **27**, 1-9 (2010).

42. Bayer, T.S. & Smolke, C.D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* **23**, 337-343 (2005).

43. Win, M.N. & Smolke, C.D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A* **104**, 14283-14288 (2007).

44. Alberts, B. Molecular biology of the cell, Edn. 4th. (Garland Science, New York; 2002).

45. Day, D.A. & Tuite, M.F. Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. *J Endocrinol* **157**, 361-371 (1998).

46. Wilusz, C.J., Wormington, M. & Peltz, S.W. The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol* **2**, 237-246 (2001).

47. Tucker, M. & Parker, R. Mechanisms and control of mRNA decapping in Saccharomyces cerevisiae. *Annu Rev Biochem* **69**, 571-595 (2000).

48. Dehlin, E., Wormington, M., Korner, C.G. & Wahle, E. Cap-dependent deadenylation of mRNA. *Embo J* **19**, 1079-1086 (2000).

49. Jacobson, A. & Peltz, S.W. Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu Rev Biochem* **65**, 693-739 (1996).

50. Mitchell, P. & Tollervey, D. mRNA stability in eukaryotes. *Curr Opin Genet Dev* **10**, 193-198 (2000).

51. Tucker, M. et al. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. *Cell* **104**, 377-386 (2001).

52. Garneau, N.L., Wilusz, J. & Wilusz, C.J. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* **8**, 113-126 (2007).

53. Caponigro, G. & Parker, R. Mechanisms and control of mRNA turnover in Saccharomyces cerevisiae. *Microbiol Rev* **60**, 233-249 (1996).

54. Huang, H. & Donahue, T.F. in mRNA Metabolism and Post-transcriptional Gene Regulation. (eds. J.B. Harford & D.R. Morris) (Wiley-Liss, New York; 1997).

55. Richter, J.D. & Sonenberg, N. Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature* **433**, 477-480 (2005).

56. Weaver, R. Molecular biology, Edn. 2nd. (McGraw-Hill, New York; 2005).

57. Dever, T.E. Gene-specific regulation by general translation factors. *Cell* **108**, 545-556 (2002).

58. Klann, E. & Dever, T.E. Biochemical mechanisms for translational regulation in synaptic plasticity. *Nat Rev Neurosci* **5**, 931-942 (2004).

59. Badis, G., Saveanu, C., Fromont-Racine, M. & Jacquier, A. Targeted mRNA degradation by deadenylation-independent decapping. *Mol Cell* **15**, 5-15 (2004).

60. Muhlrad, D. & Parker, R. The yeast EDC1 mRNA undergoes deadenylation-independent decapping stimulated by Not2p, Not4p, and Not5p. *EMBO J* **24**, 1033-1045 (2005).

61. Anderson, J.R. et al. Sequence-specific RNA binding mediated by the RNase PH domain of components of the exosome. *RNA* **12**, 1810-1816 (2006).

62. Chen, C.Y. et al. AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* **107**, 451-464 (2001).

63. Goldstrohm, A.C., Hook, B.A., Seay, D.J. & Wickens, M. PUF proteins bind Pop2p to regulate messenger RNAs. *Nat Struct Mol Biol* **13**, 533-539 (2006).

64. Wang, Z., Day, N., Trifillis, P. & Kiledjian, M. An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA in vitro. *Mol Cell Biol* **19**, 4552-4560 (1999).

65. Valencia-Sanchez, M.A., Liu, J., Hannon, G.J. & Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev* **20**, 515-524 (2006).

66. Liu, J. et al. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437-1441 (2004).

67. Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M.C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* **18**, 1655-1666 (2004).

68. Song, J.J., Smith, S.K., Hannon, G.J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434-1437 (2004).

69. Guo, H.S., Xie, Q., Fei, J.F. & Chua, N.H. MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for arabidopsis lateral root development. *Plant Cell* **17**, 1376-1386 (2005).

70. Mallory, A.C. et al. MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region. *EMBO J* **23**, 3356-3364 (2004).

71. Yekta, S., Shih, I.H. & Bartel, D.P. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594-596 (2004).

72. Gill, T., Cai, T., Aulds, J., Wierzbicki, S. & Schmitt, M.E. RNase MRP cleaves the CLB2 mRNA to promote cell cycle progression: novel method of mRNA degradation. *Mol Cell Biol* **24**, 945-953 (2004).

73. Hollien, J. & Weissman, J.S. Decay of endoplasmic reticulum-localized mRNAs during the unfolded protein response. *Science* **313**, 104-107 (2006).

74. Yang, F. & Schoenberg, D.R. Endonuclease-mediated mRNA decay involves the selective targeting of PMR1 to polyribosome-bound substrate mRNA. *Mol Cell* **14**, 435-445 (2004).

75. Serganov, A. & Patel, D.J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet* **8**, 776-790 (2007).

76. Gregory, R.I., Chendrimada, T.P. & Shiekhattar, R. MicroRNA biogenesis: isolation and characterization of the microprocessor complex. *Methods Mol Biol* **342**, 33-47 (2006).

77. Ge, D., Lamontagne, B. & Elela, S.A. RNase III-mediated silencing of a glucose-dependent repressor in yeast. *Curr Biol* **15**, 140-145 (2005).

78. Chanfreau, G., Buckle, M. & Jacquier, A. Recognition of a conserved class of RNA tetraloops by Saccharomyces cerevisiae RNase III. *Proc Natl Acad Sci U S A* **97**, 3142-3147 (2000).

79. Gebauer, F. & Hentze, M.W. Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol* **5**, 827-835 (2004).

80. Gray, N.K. & Hentze, M.W. Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. *EMBO J* **13**, 3882-3891 (1994).

81. Muckenthaler, M., Gray, N.K. & Hentze, M.W. IRP-1 binding to ferritin mRNA prevents the recruitment of the small ribosomal subunit by the cap-binding complex eIF4F. *Mol Cell* **2**, 383-388 (1998).

82. Stripecke, R., Oliveira, C.C., McCarthy, J.E. & Hentze, M.W. Proteins binding to 5' untranslated region sites: a general mechanism for translational regulation of mRNAs in human and yeast cells. *Mol Cell Biol* **14**, 5898-5909 (1994).

83. Mendez, R. & Richter, J.D. Translational control by CPEB: a means to the end. *Nat Rev Mol Cell Biol* **2**, 521-529 (2001).

84. Stebbins-Boaz, B., Cao, Q., de Moor, C.H., Mendez, R. & Richter, J.D. Maskin is a CPEB-associated factor that transiently interacts with elF-4E. *Mol Cell* **4**, 1017-1027 (1999).

85. Gebauer, F., Grskovic, M. & Hentze, M.W. Drosophila sex-lethal inhibits the stable association of the 40S ribosomal subunit with msl-2 mRNA. *Mol Cell* **11**, 1397-1404 (2003).

86. Ostareck, D.H., Ostareck-Lederer, A., Shatsky, I.N. & Hentze, M.W. Lipoxygenase mRNA silencing in erythroid differentiation: The 3'UTR regulatory complex controls 60S ribosomal subunit joining. *Cell* **104**, 281-290 (2001).

87. Humphreys, D.T., Westman, B.J., Martin, D.I. & Preiss, T. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci U S A* **102**, 16961-16966 (2005).

88. Wu, H. et al. A novel family of RNA tetraloop structure forms the recognition site for Saccharomyces cerevisiae RNase III. *Embo J* **20**, 7240-7249 (2001).

89. Lingel, A. & Izaurralde, E. RNAi: finding the elusive endonuclease. *Rna* **10**, 1675-1679 (2004).

90. Lee, Y., Jeon, K., Lee, J.T., Kim, S. & Kim, V.N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**, 4663-4670 (2002).

91. Elela, S.A., Igel, H. & Ares, M., Jr. RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell* **85**, 115-124 (1996).

92. Chanfreau, G., Elela, S.A., Ares, M., Jr. & Guthrie, C. Alternative 3'-end processing of U5 snRNA by RNase III. *Genes Dev* **11**, 2741-2751 (1997).

93. Chanfreau, G., Rotondo, G., Legrain, P. & Jacquier, A. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *Embo J* **17**, 3726-3737 (1998).

94. Danin-Kreiselman, M., Lee, C.Y. & Chanfreau, G. RNAse III-mediated degradation of unspliced pre-mRNAs and lariat introns. *Mol Cell* **11**, 1279-1289 (2003).

95. Catala, M., Lamontagne, B., Larose, S., Ghazal, G. & Elela, S.A. Cell cycle-dependent nuclear localization of yeast RNase III is required for efficient cell division. *Mol Biol Cell* **15**, 3015-3030 (2004).

96. Filippov, V., Solovyev, V., Filippova, M. & Gill, S.S. A novel type of RNase III family proteins in eukaryotes. *Gene* **245**, 213-221 (2000).

97. Wu, H., Henras, A., Chanfreau, G. & Feigon, J. Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc Natl Acad Sci U S A* **101**, 8307-8312 (2004).

98. Lamontagne, B. et al. Sequence dependence of substrate recognition and cleavage by yeast RNase III. *J Mol Biol* **327**, 985-1000 (2003).

99. Lebars, I., Lamontagne, B., Yoshizawa, S., Aboul-Elela, S. & Fourmy, D. Solution structure of conserved AGNN tetraloops: insights into Rnt1p RNA processing. *EMBO J* **20**, 7250-7258 (2001).

100. Presutti, C., Ciafre, S.A. & Bozzoni, I. The ribosomal protein L2 in S. cerevisiae controls the level of accumulation of its own mRNA. *EMBO J* **10**, 2215-2221 (1991).

101. Vreken, P. & Raue, H.A. The rate-limiting step in yeast PGK1 mRNA degradation is an endonucleolytic cleavage in the 3'-terminal part of the coding region. *Mol Cell Biol* **12**, 2986-2996 (1992).

102. Lee, A., Henras, A.K. & Chanfreau, G. Multiple RNA surveillance pathways limit aberrant expression of iron uptake mRNAs and prevent iron toxicity in S. cerevisiae. *Mol Cell* **19**, 39-51 (2005).

103. Nomoto, A., Kitamura, N., Golini, F. & Wimmer, E. The 5'-terminal structures of poliovirion RNA and poliovirus mRNA differ only in the genome-linked protein VPg. *Proc Natl Acad Sci U S A* **74**, 5345-5349 (1977).

104. Glass, M.J. & Summers, D.F. A cis-acting element within the hepatitis A virus 5'-non-coding region required for in vitro translation. *Virus Res* **26**, 15-31 (1992).

105. Tsukiyama-Kohara, K., Iizuka, N., Kohara, M. & Nomoto, A. Internal ribosome entry site within hepatitis C virus RNA. *J Virol* **66**, 1476-1483 (1992).

106. Kuhn, R., Luz, N. & Beck, E. Functional analysis of the internal translation initiation site of foot-and-mouth disease virus. *J Virol* **64**, 4625-4631 (1990).

107. Buck, C.B. et al. The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. *J Virol* **75**, 181-191 (2001).

108. Kieft, J.S. Viral IRES RNA structures and ribosome interactions. *Trends Biochem Sci* **33**, 274-283 (2008).

109. Hellen, C.U. & Sarnow, P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev* **15**, 1593-1612 (2001).

110. Kaminski, A. & Jackson, R.J. The polypyrimidine tract binding protein (PTB) requirement for internal initiation of translation of cardiovirus RNAs is conditional rather than absolute. *Rna* **4**, 626-638 (1998).

111. Pilipenko, E.V. et al. A cell cycle-dependent protein serves as a template-specific translation initiation factor. *Genes Dev* **14**, 2028-2045 (2000).

112. Pestova, T.V., Shatsky, I.N., Fletcher, S.P., Jackson, R.J. & Hellen, C.U. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev* **12**, 67-83 (1998).

113. Martinez-Salas, E., Ramos, R., Lafuente, E. & Lopez de Quinto, S. Functional interactions in internal translation initiation directed by viral and cellular IRES elements. *J Gen Virol* **82**, 973-984 (2001).

114. Chappell, S.A., Edelman, G.M. & Mauro, V.P. A 9-nt segment of a cellular mRNA can function as an internal ribosome entry site (IRES) and when present in linked multiple copies greatly enhances IRES activity. *Proc Natl Acad Sci U S A* **97**, 1536-1541 (2000).

115. Yang, Q. & Sarnow, P. Location of the internal ribosome entry site in the 5' non-coding region of the immunoglobulin heavy-chain binding protein (BiP) mRNA: evidence for specific RNA-protein interactions. *Nucleic Acids Res* **25**, 2800-2807 (1997).

116. Hu, M.C., Tranque, P., Edelman, G.M. & Mauro, V.P. rRNA-complementarity in the 5' untranslated region of mRNA specifying the Gtx homeodomain protein: evidence that base- pairing to 18S rRNA affects translational efficiency. *Proc Natl Acad Sci U S A* **96**, 1339-1344 (1999).

117. Shine, J. & Dalgarno, L. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A* **71**, 1342-1346 (1974).

118. Owens, G.C., Chappell, S.A., Mauro, V.P. & Edelman, G.M. Identification of two short internal ribosome entry sites selected from libraries of random oligonucleotides. *Proc Natl Acad Sci U S A* **98**, 1471-1476 (2001).

119. Zhou, W., Edelman, G.M. & Mauro, V.P. Isolation and identification of short nucleotide sequences that affect translation initiation in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **100**, 4457-4462 (2003).

120. Moradpour, D., Penin, F. & Rice, C.M. Replication of hepatitis C virus. *Nat Rev Microbiol* **5**, 453-463 (2007).

121. Ricci, E.P., Rifo, R.S., Herbreteau, C.H., Decimo, D. & Ohlmann, T. Lentiviral RNAs can use different mechanisms for translation initiation. *Biochem Soc T* **36**, 690-693 (2008).

122. De Felipe, P. & Izquierdo, M. Tricistronic and tetracistronic retroviral vectors for gene transfer. *Hum Gene Ther* **11**, 1921-1931 (2000).

123. de Felipe, P. & Izquierdo, M. Construction and characterization of pentacistronic retrovirus vectors. *J Gen Virol* **84**, 1281-1285 (2003).

124. Edwards, S.R. & Wandless, T.J. Dicistronic regulation of fluorescent proteins in the budding yeast Saccharomyces cerevisiae. *Yeast* **27**, 229-236 (2010).

Chapter II: A synthetic library of RNA control modules for predictable tuning of gene expression in yeast

Abstract

Advances in synthetic biology have resulted in the development of genetic tools that support the design of complex biological systems encoding desired functions. The majority of efforts have focused on the development of regulatory tools in bacteria, whereas fewer tools exist for the tuning of expression levels in eukaryotic organisms. Here, we describe a novel class of RNA-based control modules that provide predictable tuning of expression levels in the yeast *Saccharomyces cerevisiae*. A library of synthetic control modules that act through posttranscriptional RNase cleavage mechanisms was generated through an *in vivo* screen, where structural engineering methods were applied to enhance the insulation and modularity of the resulting components. The library of sixteen synthetic RNase substrates exhibit a wide range of gene regulatory activities (spanning 8% and 85% at the protein level). This new class of control elements can be combined with any promoter to support titration of regulatory strategies encoded in transcriptional regulators and thus more sophisticated control schemes. We applied these synthetic controllers to the systematic titration of flux through the ergosterol biosynthesis pathway, where feedback regulation was observed to maintain production of ergosterol and thus cellular growth rates. This work provides insight into endogenous control strategies and highlights the utility of this control module library for manipulating and probing biological systems.

2.1. Introduction

Synthetic biology is advancing capabilities for engineering biological systems exhibiting desired functions. The proper functioning of synthetic genetic circuits often relies on precise control and tuning of the expression levels of key protein components. For example, the proper functioning of synthetic gene networks exhibiting complex dynamic behaviors has been shown to depend on the appropriate matching of levels of protein components in the engineered networks[1-3]. The tuning of protein levels to obtain functioning circuits has been commonly achieved by screening randomized gene expression control elements for those sequences that provide the desired regulatory strength[3-5]. As another example, the optimization of engineered metabolic networks has been shown to depend on the precise control of enzyme levels and activities[4, 6-7]. The tuning of enzyme levels is critical for reducing metabolic burden due to enzyme overexpression[8-9], decreasing accumulation of toxic intermediates by balancing pathway flux[4, 10], and redirecting cellular resources from native pathways without negatively impacting the health and viability of the engineered host by knocking out required enzymes[11-12]. As such, the development of well-characterized gene expression control modules that can be used to predictably tune the levels of proteins are key to the design of robust genetic systems.

While many gene regulatory tools have been developed for use in *Escherichia coli*[13-16] fewer such tools exist for the precise tuning of expression levels in the budding yeast, *Saccharomyces cerevisiae*. However, *S. cerevisiae* is a relevant organism in industrial processes, including biosynthesis and biomanufacturing strategies[17-22], such that as more complex genetic networks are engineered into yeast it becomes critical to

have tools that allow for the facile programming of gene expression levels. The existing methods for tuning gene expression levels in *S. cerevisiae* rely on transcriptional control mechanisms in the form of inducible and constitutive promoter systems. Many inducible promoters do not provide tunable control systems due to their on/off switch-like behavior, where the amount of inducer molecule controls the likelihood that a given cell is repressed or fully expressing the desired protein[23]. While engineered variants have been constructed that offer more tunable responses[24-25], these systems can exhibit other undesirable properties, due to nonspecific or pleotropic effects associated with the inducing molecule or limitations associated with costs in using the inducing molecule in large-scale processes. As an alternative strategy, a promoter library was recently developed based on mutating the constitutive TEF1 promoter[26]. The resulting library of promoter parts comprised 11 promoter variants that spanned expression levels from 8% to 120%, providing a useful tool for controlling expression levels in yeast. However, control modules based on transcriptional mechanisms require the use of a particular promoter, which may be limiting to certain applications. For example, the use of a specific or native promoter may be desired to retain cellular control mechanisms associated with the given promoter. RNA-based control modules based on posttranscriptional mechanisms may offer an advantage by allowing these control elements to be coupled to any promoter of choice, providing for enhanced control strategies and finer resolution tuning of expression levels.

Endoribonucleases play key roles in RNA processing across diverse cellular systems[27]. In eukaryotic cells, endoribonuclease cleavage in the untranslated regions (UTRs) or coding regions of a transcript can result in rapid degradation of that transcript

by exoribonucleases. The RNase III family is a class of enzymes that cleaves double-stranded RNA (dsRNA)[28]. The *S. cerevisiae* RNase III enzyme, Rnt1p, recognizes RNA hairpins that contain a consensus AGNN tetraloop and cleaves its substrates 14 nucleotides (nt) upstream and 16 nt downstream of the tetraloop[29]. Rnt1p harbors an RNase III domain and a dsRNA-binding domain (dsRBD)[28], where the AGNN tetraloop of an Rnt1p substrate forms a predetermined fold that is recognized by the dsRBD[30]. Rnt1p is localized to the nucleus, where it has been shown to cleave cellular ribosomal RNA (rRNA) precursors, small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), and messenger RNAs (mRNA)[31-34]. However, despite extensive characterization of this RNA processing enzyme, neither natural nor synthetic Rnt1p substrates have been used to control gene expression levels in yeast.

We have demonstrated that Rnt1p substrates can be utilized as effective posttranscriptional gene control modules when placed in the 3' UTR of a target transcript. We utilized this Rnt1p regulatory construct with a cell-based screening strategy to develop a library of synthetic Rnt1p substrates that exhibit a wide range of gene regulatory activities (spanning 8% and 85%) to tune Rnt1p processing efficiency. *In vivo* and *in vitro* assays demonstrate that the library of control elements modulate transcript and protein levels through variations of the Rnt1p processing efficiency. The library of Rnt1p elements was applied to predictably modulate flux through an endogenous ergosterol biosynthesis network through the direct integration of the synthetic components with an endogenous gene target, highlighting the broader utility of these synthetic control modules. The described Rnt1p substrate library provides a new set of

control modules that can be used to predictably tune gene expression in yeast with any desired promoter.

2.2. Results

2.2.1. Implementing Rnt1p hairpins as RNA-based gene regulatory components

Rnt1p is an RNase III enzyme that cleaves consensus hairpin structures in *S. cerevisiae*. In order for a hairpin to be effectively recognized and cleaved by Rnt1p it must have the following consensus elements: an AGNN tetraloop and four base-pairs immediately below the tetraloop (Figure 2.1A). An Rnt1p substrate can be divided into three critical regions: the initial binding and position box (IBPB), comprising the tetraloop; the binding stability box (BSB), comprising the base-paired region immediately adjacent to the tetraloop; and the cleavage efficiency box (CEB), comprising the region containing and surrounding the cleavage site[29]. The CEB has no reported sequence or structural requirements. Rnt1p will initially position itself and bind to the tetraloop and cleave the hairpin at two locations within the CEB: between the 14th and 15th nts upstream of the tetraloop and the 16th and 17th nts downstream of the tetraloop. Most naturally-occurring Rnt1p hairpins have been identified in noncoding RNAs (ncRNAs), where Rnt1p plays a critical role in ncRNA processing[31-33]. Synthetic *trans*-acting RNA guide strands were recently utilized to direct Rnt1p processing of a target ncRNA[35]. Rnt1p hairpins have also been identified within the coding region of at least one endogenous yeast gene, *MIG2*, where Rnt1p was shown to play a role in controlling expression levels of that gene[36]. However, the ability of Rnt1p hairpins to function as

genetic control modules in regulating the expression of heterologous genes has not been previously examined.

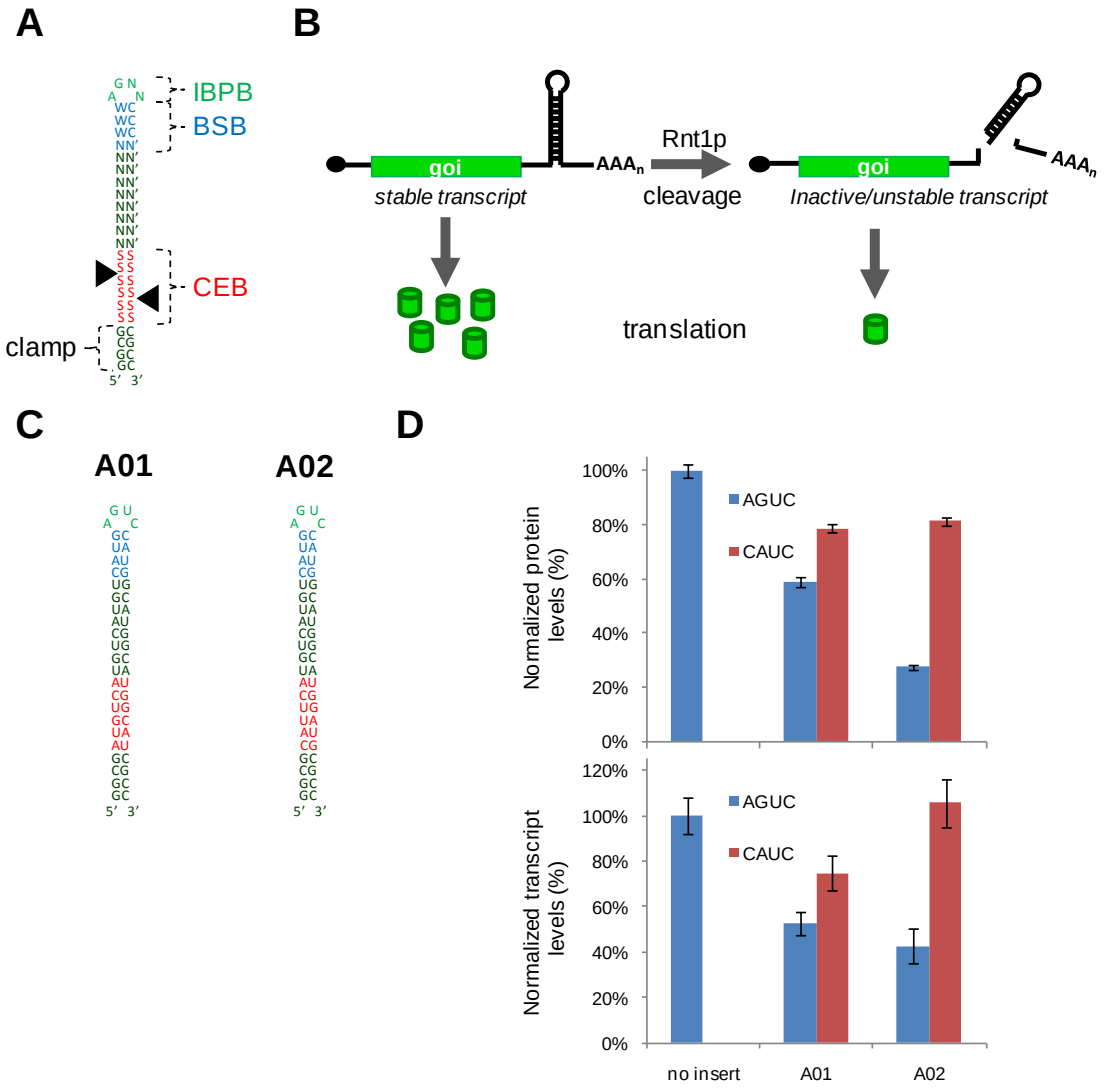


Figure 2.1. Genetic control elements based on Rnt1p hairpins. (A) Consensus elements of an Rnt1p hairpin. Color scheme is as follows: cleavage efficiency box (CEB), red; binding stability box (BSB), blue; initial binding and positioning box (IBPB), green. Black triangles represent location of cleavage sites. The clamp region is a synthetic sequence that acts to insulate and maintain the structure of the control element. (B) Schematic illustrating the mechanism by which Rnt1p hairpins act as gene control elements when placed in the 3' UTR of a gene of interest (goi). Barrels represent protein molecules. (C) Sequences and structures of Rnt1p hairpin controls. (D) The transcript and protein levels associated with Rnt1p hairpins and their corresponding mutated tetraloop (CAUC) controls support that the observed gene regulatory activity is due to Rnt1p processing. Normalized protein expression levels are determined by measuring the

median GFP levels from a cell population harboring the appropriate construct through flow cytometry analysis and values are reported relative to that from an identical construct lacking a hairpin module (no insert). Reported values and their error are calculated from the mean and standard deviation from the three identically-grown samples, respectively. Transcript levels are determined by measuring transcript levels of *yEGFP3* and a house-keeping gene, *ACT1*, through qRT-PCR and normalizing the *yEGFP3* levels with their corresponding *ACT1* levels. Normalized transcript levels are reported relative to that from an identical construct lacking a hairpin module. Reported values and their error are calculated from the mean and standard deviation from three identically-prepared qRT-PCR reactions, respectively.

We designed a system that utilizes Rnt1p-mediated hairpin cleavage to regulate gene expression in yeast through the modular insertion of Rnt1p hairpins in UTRs of a gene (Figure 2.1B). Specifically, we inserted Rnt1p hairpins as gene control elements within the 3' UTR of a transcript to direct cleavage to that region, thereby inactivating the transcript and resulting in rapid transcript degradation. While directing cleavage to the 5' UTR of a transcript would be expected to similarly inactivate the transcript, insertion of secondary structures in the 5' UTR of eukaryotic transcripts has been shown to result in nonspecific translational inhibition due to affects of structural elements on ribosomal scanning[37], such that resulting gene regulatory effects would likely not be specific to the desired cleavage mechanism. We designed and built a low-copy Rnt1p characterization plasmid (pCS321) to quantify the gene regulatory properties of Rnt1p substrates in yeast. Unique restriction sites for inserting Rnt1p hairpins were located 2 nts downstream of the stop codon of a gene encoding a yeast enhanced green fluorescent protein (*yEGFP3*)[38].

We first examined the ability of Rnt1p hairpins to function as gene control elements when placed downstream of a heterologous reporter gene in yeast. We adapted two Rnt1p hairpins with different CEBs, A01 and A02, that had been previously characterized through *in vitro* assays (R31-27 and R31D, respectively, in [39]) (Figure 2.1C). The hairpins were modified by placing a G-C rich base-paired region, or clamp,

below the 18 nt stem of the Rnt1p substrates to ensure structural stability of the hairpins when placed within the context of our Rnt1p characterization construct and to provide a proper stem length for effective cleavage *in vivo*. The hairpins were inserted into the characterization plasmid, and regulatory efficiencies were determined by monitoring cellular fluorescence by flow cytometry and transcript levels by quantitative real-time PCR (qRT-PCR) (Table 2.1). Negative controls for Rnt1p hairpins were constructed by mutating the tetraloop sequence to CAUC to impede Rnt1p activity while maintaining the secondary structure of the hairpins. The fluorescence and transcript data for A01 and A02 show that nucleotide modifications in the CEB result in the attenuation of *in vivo* gene expression, and the mutated tetraloop controls support that the observed regulatory effects are due to Rnt1p processing (Figure 2.1D). Flow cytometry histograms of the control hairpins demonstrate that regulatory activity causes a population shift with reduced median levels (Supplementary Figure 2.1A).

Table 2.1. *In vivo* characterization data for the Rnt1p cleavage library. All normalized protein and transcript levels are determined as described in Figure 2.1D.

Substrate	Normalized protein levels (%)	Normalized transcript levels (%)
C01	84% ± 6%	68% ± 3%
C02	80% ± 3%	71% ± 7%
C03	55% ± 1%	60% ± 9%
C04	20% ± 1%	31% ± 4%
C05	55% ± 2%	51% ± 5%
C06	33% ± 2%	55% ± 5%
C07	41% ± 1%	67% ± 13%
C08	11% ± 0%	12% ± 2%
C09	25% ± 1%	28% ± 4%
C10	46% ± 1%	66% ± 8%
C11	11% ± 0%	56% ± 10%
C12	81% ± 6%	75% ± 12%
C13	8% ± 0%	12% ± 1%
C14	85% ± 3%	83% ± 6%
A01	59% ± 2%	53% ± 5%
A02	28% ± 1%	43% ± 8%
no hairpin	100% ± 3%	100% ± 8%

2.2.2. Design and selection of an Rnt1p cleavage library to achieve tunable gene regulatory control

Gene regulatory elements that allow the precise and predictable tuning of expression levels are important tools in synthetic biology for the control of gene circuits. The generation of well-characterized libraries of gene control elements that exhibit varying regulatory properties has resulted in useful tools in bacteria[13-15]. Similar strategies have been applied to develop libraries of transcriptional control elements, specifically constitutive promoters, in yeast[26, 40]. However, certain applications will require circuit designs where either native or inducible promoter systems will be desired, such that the ability to integrate posttranscriptional control elements that act downstream of desired promoter systems will be required. Libraries of tuned posttranscriptional stability control elements have not been developed to date in *S. cerevisiae*.

Based on the different gene regulatory activities observed from A01 and A02, we examined whether a larger library of synthetic Rnt1p hairpins could be engineered to develop a set of tuned posttranscriptional control elements. We developed an Rnt1p library based on randomizing the CEB (12 nt) to generate Rnt1p hairpins that exhibit different gene regulatory activities due to altered enzyme processing rates and identified synthetic Rnt1p substrates through a cell-based fluorescence screen (Figure 2.2A). The designed library has a diversity of ~1.7 x 10^7 different hairpin sequences. Due to the flexibility of the structural and sequence requirements for the CEB and the ability of each of the library members to bind Rnt1p through the maintained tetraloop structure, we anticipated that a large percentage of the library members would exhibit some cleavage activity. The goal of the functional screen of the library was to identify a set of Rnt1p

hairpins that will provide a range of different regulatory activities and be useful as modular gene control elements.

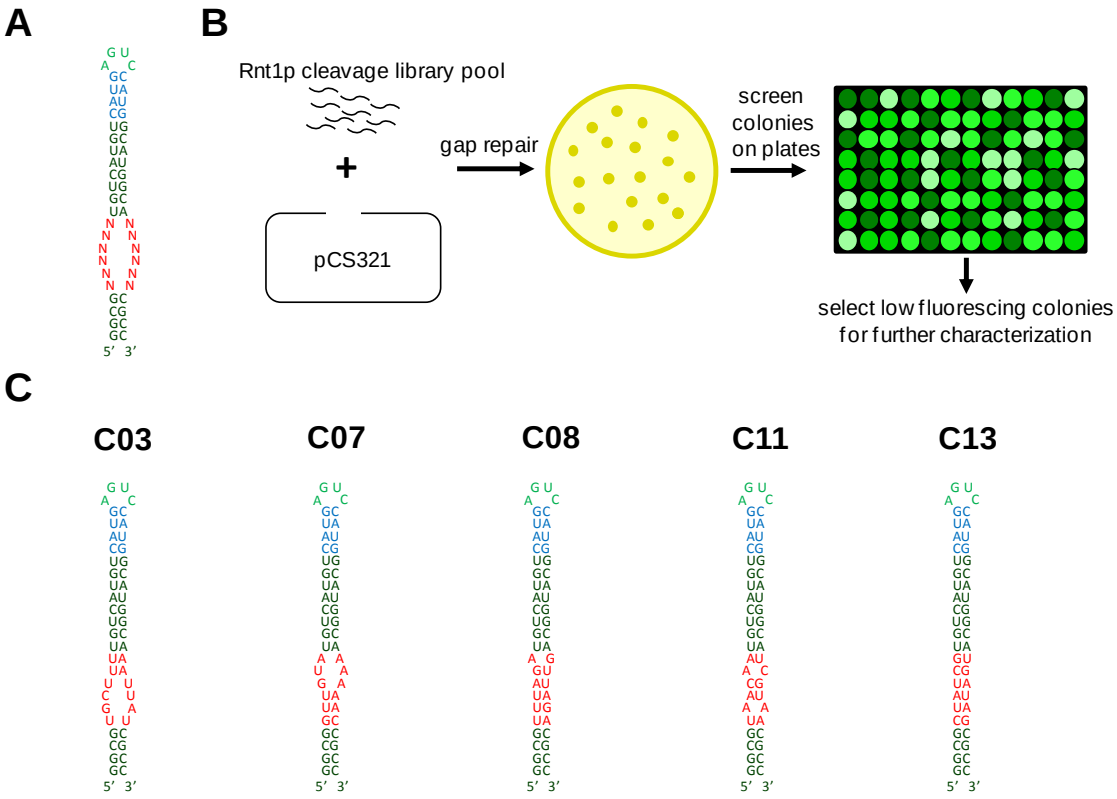


Figure 2.2. Design and *in vivo* screening of an Rnt1p cleavage library. (A) Sequence and structure of Rnt1p hairpin library containing the 12 randomized nucleotides in the CEB. (B) An *in vivo*, fluorescence-based screen of Rnt1p hairpin activity. The library pool is cloned through gap-repair into yeast, and clones are screened on a plate reader for sequences resulting in low fluorescence. (C) Sequences and structures of select library members highlight the diversity of the selected library sequences. The color scheme for hairpin sequences is described in Figure 2.1A.

The cleavage library was transformed into yeast through a gap-repair strategy and individual colonies were initially characterized for gene regulatory activity through assaying cellular fluorescence on a plate reader (Figure 2.2B). A total of 318 colonies were characterized and from this initial screening, constructs from 41 low-expressing colonies were sequenced (Supplementary Table 2.1). The sequences of the selected Rnt1p hairpins were analyzed by RNAstructure (http://rna.chem.rochester.edu/

RNAstructure.html) to determine the predicted secondary structure of the hairpins. No consensus secondary structure was identified from the 41 isolates due to the diversity of associated structures. The CEBs of the recovered hairpins are either completely base-paired or contain one or two bulges of different size and location (Figure 2.2C). To ensure modularity of the synthetic Rnt1p hairpins to other genetic constructs, we removed library candidates that were 'structurally weak'. Structurally-weak hairpins were identified based on two properties: 1) predicted ability of the hairpin sequence to fold into multiple secondary structure conformations with similar free energies and 2) interactions with flanking sequences. In total, 16 Rnt1p cleavage library substrates were identified as synthetic control modules (Table 2.1, Supplementary Figure 2.2).

2.2.3. A synthetic Rnt1p hairpin library exhibits a range of gene regulatory activities **in vivo**

The range of regulatory activities spanned by the cleavage library was measured at the protein expression and transcript levels. Flow cytometry analysis of the synthetic Rnt1p hairpins indicated that the selected set of hairpins spanned a large gene regulatory range – from 7.9% (C13) to 84.7% (C14) (Table 2.1, Figure 2.3A). The regulatory activities of the selected hairpins are fairly evenly distributed across this range allowing for precise tuning of expression levels based on insertion of different synthetic Rnt1p hairpins. Flow cytometry histograms of the library hairpins confirm that regulatory activity causes a population shift with reduced median GFP levels (Supplementary Figure 2.1B). The negative controls demonstrated that the majority of knockdown observed from each hairpin is due to Rnt1p processing (Figure 2.3B). The controls also indicate that the

hairpin structures can have slight effects on gene expression (compared to the construct with no hairpin insertion set at 100%), likely due to some effects of the inserted structures on normal translation or degradation processes.

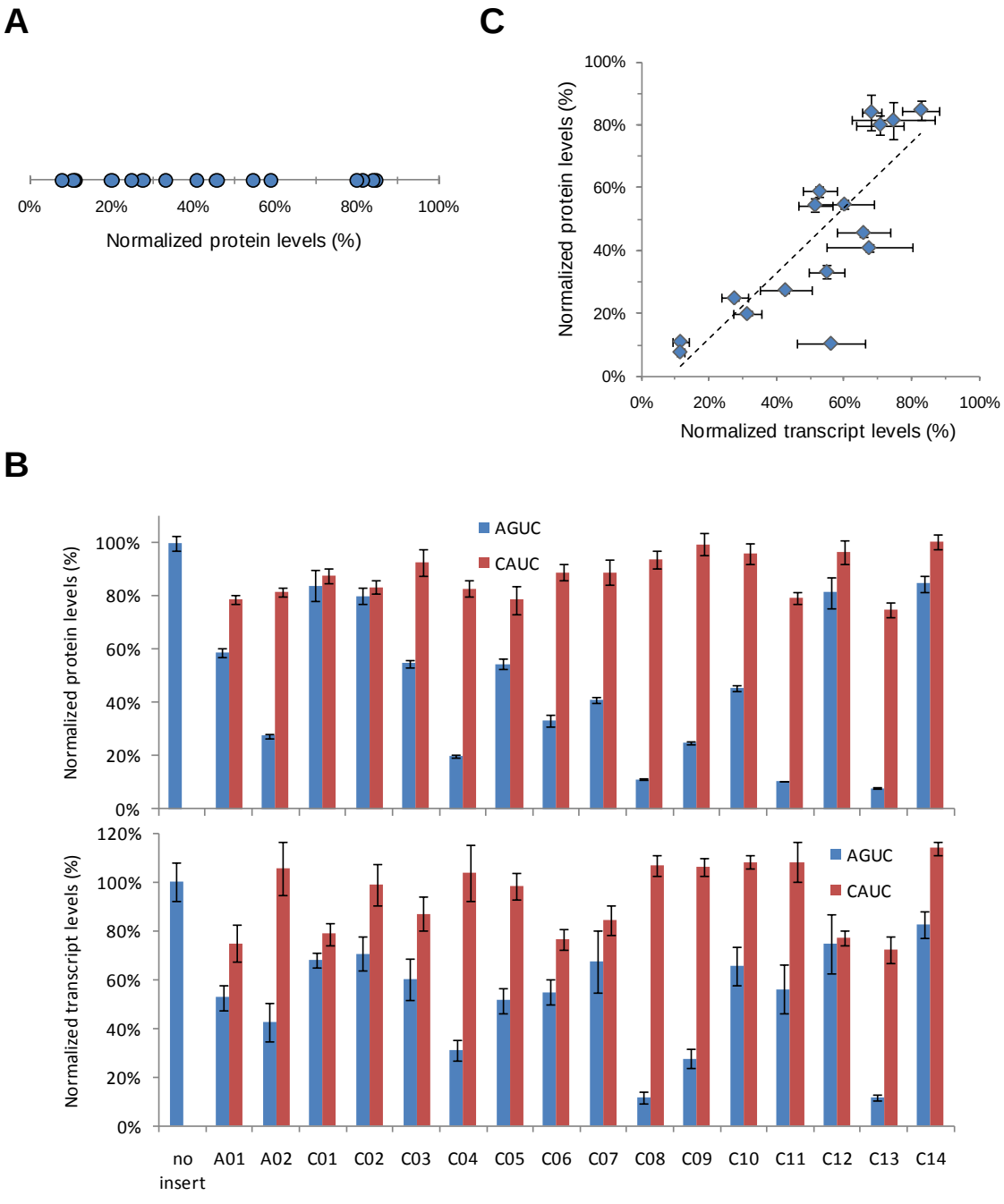


Figure 2.3. *In vivo* characterization of the selected Rnt1p cleavage library. (A) The gene regulatory range of the Rnt1p library spans a broad range of protein expression levels. (B) The transcript and protein levels associated with all Rnt1p library members and their corresponding mutated tetraloop (CAUC) controls supports that the observed gene regulatory activity is due to Rnt1p processing. (C) Correlation analysis of protein and transcript levels from the Rnt1p hairpin library members supports a strong correlation between the two measures of gene regulatory activity. All normalized protein and transcript levels and their error are determined as described in Figure 2.1D.

The activity of the synthetic Rnt1p hairpins was further confirmed by monitoring steady-state transcript levels in cells harboring the Rnt1p constructs (Table 2.1, Figure 2.3B). Rnt1p hairpins generally resulted in reduced transcript levels compared to a construct harboring no Rnt1p hairpins and to a construct harboring a mutated tetraloop. In addition, a plot of normalized *yEGFP3* expression levels versus normalized *yEGFP3* transcript levels indicates that there is a strong positive correlation ($r = 0.817$) between the two measures (Figure 2.3C). Specifically, with decreasing transcript levels a similar decrease in protein levels was generally observed, as further supported from a Spearman's rank correlation coefficient (ρ) value of 0.818. Unintended effects of the hairpins on translation and transcript stability caused by interference of the structures on the machinery controlling those processes may contribute to deviations from linearity. However, observed deviations are not entirely due to structural variability between the library members, as hairpins with similar secondary structure (i.e., A01, A02, C13) do not demonstrate an exact linear relationship between transcript and protein levels.

2.2.4. Rnt1p library hairpins maintain regulatory activity in a different genetic context

The utility of any genetic control element requires that the control module retain its activity under different genetic contexts. The modular function of the synthetic Rnt1p

hairpins, as measured by maintenance of gene regulatory activity, may be impacted by differences in 3' UTR, promoter, and gene sequences. We cloned the synthetic Rnt1p hairpins into a second construct harboring a different promoter (TEF1), terminator (CYC1), and gene (*ymCherry*), and measured the regulatory activities of each hairpin through flow cytometry and qRT-PCR assays (Supplementary Table 2.1). The data indicate a strong positive correlation ($r = 0.897$) and a strong preservation of rank order ($\rho = 0.882$) between the *ymCherry* protein and transcript levels, confirming that gene regulatory activity by the Rnt1p modules is due to the reduction of steady-state transcript levels (Figure 2.4A). Flow cytometry histograms confirm that regulatory activity associated with the Rnt1p hairpins in this second construct causes a population shift with reduced mean fluorescence levels (Supplementary Figure 2.3). The functional modularity of the hairpins was determined by performing a correlation analysis between expression data for the hairpins in the *yEGFP3* and *ymCherry* constructs (Figure 2.4B), which demonstrated a strong positive correlation ($r = 0.856$) between the two data sets. A lack of functional modularity was observed for one hairpin (C06), which did not maintain its gene regulatory activity in the second construct (33% yEGFP3 vs. 89% ymCherry) (Figure 2.4B; red point). The data suggests that the flanking sequences in the *ymCherry* construct may be disruptive to the structural integrity of C06, thereby affecting cleavage efficiency in the CEB, although RNA folding software does not predict alternative hairpin structures. The majority of data indicate that *ymCherry* expression tended to be slightly higher than that of *yEGFP3*, likely due to the differences in transcriptional strength between the TEF1 and GAL1 promoters, where TEF1 resulted in a greater absolute number of transcripts per cell (data not shown). The data indicate that there was

a strong preservation of rank order (ρ = 0.848) between the two different genetic contexts, supporting the functional modularity of the Rnt1p hairpin library.

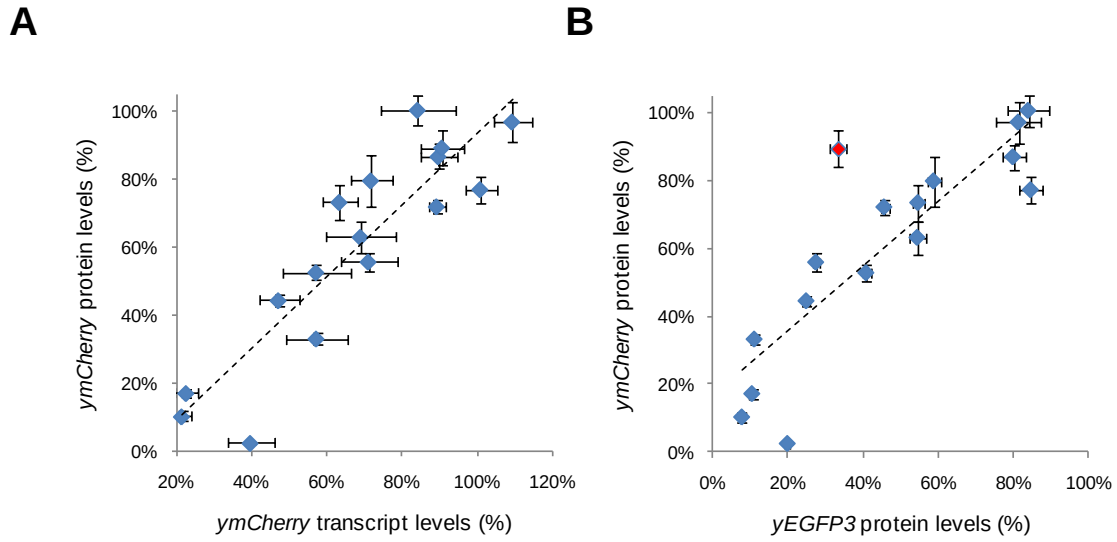


Figure 2.4. Demonstration of functional modularity of the hairpin library in the context of a different genetic construct. (A) Correlation analysis of ymCherry protein and transcript levels from the Rnt1p hairpin library members supports a strong correlation between the two measures of gene regulatory activity. Normalized protein and transcript levels and their error are determined as described in Figure 2.1D with the mean ymCherry fluorescence used for the protein level measurement. (B) Correlation analysis of ymCherry and yEGFP3 protein levels from the Rnt1p hairpin library members demonstrates a strong correlation between gene regulatory activities in different genetic contexts and preservation of library rank-order. Red data point, C06.

2.2.5. In vitro *characterization demonstrates that Rnt1p library members achieve differential activity through alterations in Rnt1p cleavage rates*

We hypothesized that the variation in transcript processing and subsequent protein expression levels exhibited by the Rnt1p hairpin library is due to alterations of Rnt1p cleavage rates through alterations of the CEB sequence and/or structure. We analyzed the reaction through a Michaelis-Menten model, with the substrate (S) being the hairpin

transcript, the enzyme (E) being Rnt1p, and the product (P) being the cleaved pieces of the transcript. Under these conditions, the following reaction occurs:

$$E + S \underset{k_{-1}}{\overset{k_1}{\leftrightharpoons}} ES \xrightarrow{k_2} E + P$$

The rate of product formation (V) is modeled as:

$$V = \frac{V_{max} * [S]}{K_M + [S]} = \frac{k_2 * [E]_0 * [S]}{K_M + [S]}$$

The maximum rate of product formation (V_{max}) is the product of the total enzyme concentration ($[E]_0$) and k_2. Alterations in the cleavage efficiency will have an effect on the value of k_2 and thus V_{max}.

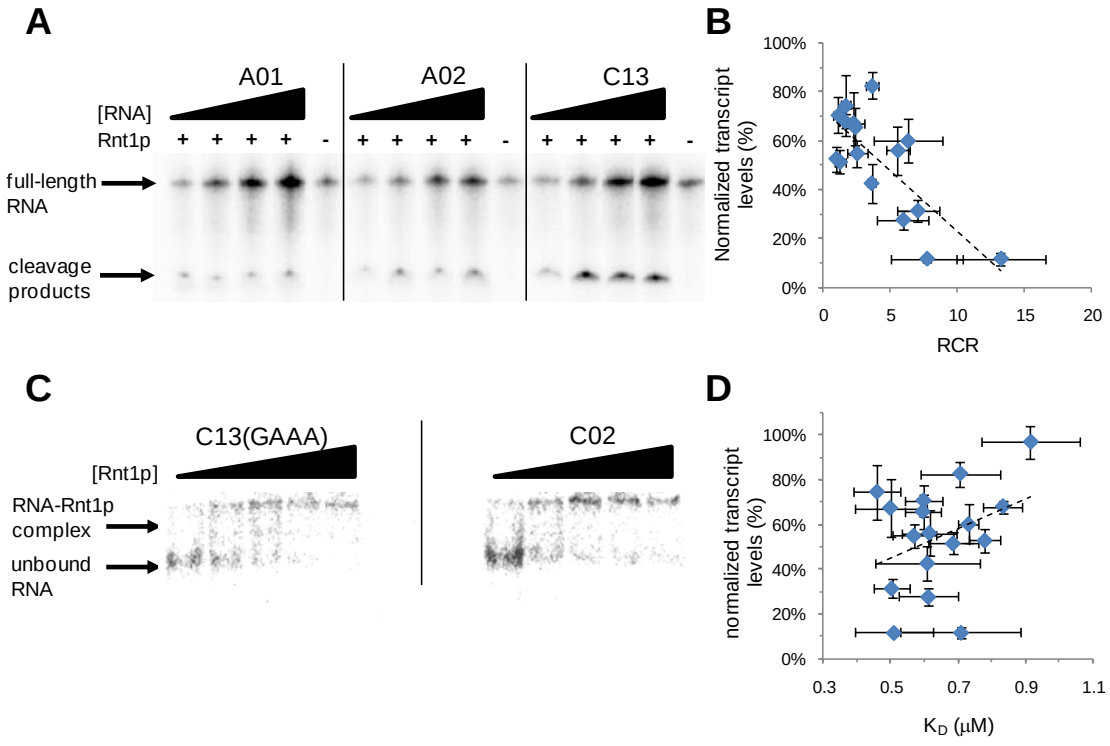


Figure 2.5. *In vitro* characterization of the Rnt1p library supports the tuning of gene regulatory activity through modulation of cleavage rates. (A) Representative cleavage reaction assays and analyses by denaturing polyacrylamide gel electrophoresis on hairpins A01, A02, and C13. The top band corresponds to full-length RNA; the bottom

band corresponds to the three cleavage products expected from Rnt1p processing. Due to added sequences flanking the Rnt1p hairpin for insulation, the three cleavage products differ in size by 1 nt and cannot be resolved into individual bands under the assay conditions. RNA is added to the following final concentrations in each reaction (left to right; in μM): 0.2, 0.35, 0.5, 0.6–0.8. Reactions lacking Rnt1p are with 0.2 μM RNA. (B) Correlation analysis of relative cleavage rate (RCR) and normalized *yEGFP3* transcript levels supports a strong correlation between cleavage rate and gene regulatory activity. Reported RCR values are determined from a Michaelis-Menten model parameter fit using Prism 5 (GraphPad) and standard error was calculated from the software. (C) Representative mobility shift assays and analyses by nondenaturing polyacrylamide gel electrophoresis on the mutated tetraloop (C13-GAAA) and C02. The top band corresponds to RNA-Rnt1p complexes; the bottom band corresponds to unbound RNA. Rnt1p is added to the following final concentrations in each reaction (left to right; in μM): 0, 0.42, 0.83, 1.25, 1.66. (D) Correlation analysis of binding affinity (K_D) and normalized *yEGFP3* transcript levels indicates a very weak correlation between binding affinity and gene regulatory activity. Reported K_D values are determined from a modified Scatchard model parameter fit using Prism 5 (GraphPad) and standard error was calculated from the software.

We performed *in vitro* RNA cleavage reactions with purified Rnt1p to determine relative values of k_2 for each synthetic Rnt1p hairpin. Reactions were run with varying concentrations of *in vitro* synthesized radiolabeled RNA encoding an Rnt1p hairpin flanked by A-rich sequences (see Materials and Methods) and a constant concentration of purified Rnt1p. Reaction products were separated by denaturing polyacrylamide gel electrophoresis and quantified through phosphorimaging analysis (Figure 2.5A). The resulting data were fit to the Michaelis-Menten model to calculate a relative cleavage rate (RCR), which is directly proportional to V_{max}. The RCR value for A01 is set to 1 and the rest of the reported values normalized to A01. The RCR values for each synthetic Rnt1p hairpin were determined through this analysis method (Table 2.2). There is a direct relationship ($r = –0.763$) between the measured RCR and gene regulatory activity for the synthetic Rnt1p hairpins (Figure 2.5B). Specifically, increasing Rnt1p's ability to cleave a substrate results in lowered transcript levels and thus lower protein expression levels. Notably, the transcript levels saturate at high RCRs, indicating that increasing the

cleavage rates above a certain threshold results in limiting decreases in transcript levels. The trend saturates at approximately 10% transcript levels, suggesting that increasing *in vitro* cleavage rates beyond an RCR value of approximately 8 will not result in an increase in the amount of transcript being processed *in vivo* and that we have approached the maximum amount of knockdown that can be achieved with a single substrate in this system. In initial control tests, we found that the mutant tetraloop (CAUC) was cleaved *in vitro* under excessive protein concentrations (i.e., nine times greater than that used in the cleavage assay). As such, another mutant tetraloop (GAAA) that exhibited no cleavage under excessive protein concentrations *in vitro* was used as a control for the cleavage assays.

Table 2.2. *In vitro* characterization data for Rnt1p cleavage library.

Substrate	RCR	K_D (μM)
C01	1.66 ± 0.43	0.83 ± 0.06
C02	1.17 ± 0.41	0.60 ± 0.05
C03	6.33 ± 2.51	0.73 ± 0.03
C04	7.06 ± 1.56	0.50 ± 0.05
C05	1.28 ± 0.36	0.68 ± 0.07
C06	2.55 ± 0.73	0.57 ± 0.06
C07	2.27 ± 0.81	0.50 ± 0.11
C08	13.25 ± 3.29	0.71 ± 0.18
C09	5.98 ± 1.92	0.61 ± 0.09
C10	2.42 ± 0.25	0.60 ± 0.06
C11	5.58 ± 0.83	0.61 ± 0.08
C12	1.71 ± 0.37	0.46 ± 0.07
C13	7.75 ± 2.64	0.51 ± 0.12
C14	3.66 ± 0.44	0.71 ± 0.12
A01	1.00 ± 0.12	0.78 ± 0.05
A02	3.62 ± 0.32	0.61 ± 0.16
C13 (GAAA)	0*	0.91 ± 0.15

*Immeasurable due to lack of product formation

While changes in the CEB are anticipated to result in changes to the Rnt1p processing efficiencies, it is also possible that the introduced sequence alterations may result in changes to the binding affinities between the hairpins and Rnt1p. To examine

whether the synthetic Rnt1p hairpins exhibit any changes in binding affinity to Rnt1p, we performed *in vitro* binding assays with purified Rnt1p. Binding reactions were run with 20 nM of *in vitro* synthesized radiolabeled RNA encoding an Rnt1p hairpin and varying concentrations of purified Rnt1p in the absence of magnesium. As magnesium and other divalent metal ions are essential to Rnt1p function[41], these reaction conditions allow for Rnt1p to bind to the substrates without subsequent cleavage. Bound products were separated by nondenaturing polyacrylamide gel electrophoresis and quantified through phosphorimaging analysis (Figure 2.5C). We analyzed the reaction through a modified Scatchard equation in which the fraction of unbound RNA (R) to total RNA (R_0) is plotted against the enzyme (E) concentration. The equation is as follows:

$$Z = \frac{R}{R_o} = \frac{K_D}{K_D + [E]}$$

The dissociation constant, K_D, for each synthetic Rnt1p hairpin was determined through this analysis method (Table 2.2). The data indicate that there is no correlation between K_D and *in vivo* gene regulatory activity (Figure 2.5D). The values of K_D for the synthetic Rnt1p hairpins span a narrow range and are weakly correlated with transcript knockdown ($r = 0.351$). A lack of correlation was anticipated due to expectations of mutations in the CEB primarily affecting cleavage. For a series of four hairpins with approximately 70% transcript levels, the reported K_D cover the range of the entire library, suggesting that nucleotide modifications in the CEB can have an effect on protein binding. The mutant tetraloop control binds with a similar K_D as the library hairpins, albeit the binding is weaker than the library. While the mutant tetraloop does not impact binding greatly, it severely impacts the ability of Rnt1p to bind in a conformation that allows cleavage, which has been previously reported[39].

2.2.6. Control of endogenous ERG9 expression by 3' UTR replacement with Rnt1p library members

In many metabolic engineering applications, there is a balance that must be maintained between diverting cellular metabolites to the production of desired compounds and the conversion of those metabolites to molecules required for cell growth and viability. In these cases, completely knocking out endogenous genes to remove the drain caused by native cellular pathways is not an option and genetic tools that allow for precise titration of enzyme expression are desired such that flux through the endogenous pathway can be minimized to that required to maintain cell viability. Farnesyl pyrophosphate (FPP) is one such cellular metabolite that is an important precursor both to industrially-relevant molecules and to molecules required for cell viability in yeast[19, 42-43]. Squalene synthase, encoded by the *ERG9* gene, is responsible for catalyzing the conversion of two molecules of FPP to squalene, the first precursor in the ergosterol biosynthetic pathway in *S. cerevisiae*[44] (Figure 2.6A). In a series of 14 catalytic steps, squalene is converted to ergosterol, the analogue of cholesterol in mammalian cells. Ergosterol is an essential component of yeast cells due to its effect on structural stability of the cell membrane[22]. Therefore, controlled reduction of ergosterol levels will allow metabolic flux to be diverted from sterol synthesis to value-added products from FPP.

We examined the ability of our posttranscriptional genetic control modules to modulate flux through the ergosterol biosynthetic pathway in a predictable manner by incorporating several members of the Rnt1p library into the 3' UTR of the *ERG9* gene. *ERG9* is in close proximity to *CTF8* on the reverse strand of the chromosome (~50 nts

between stop codons), and there is a lack of information on the transcription terminators

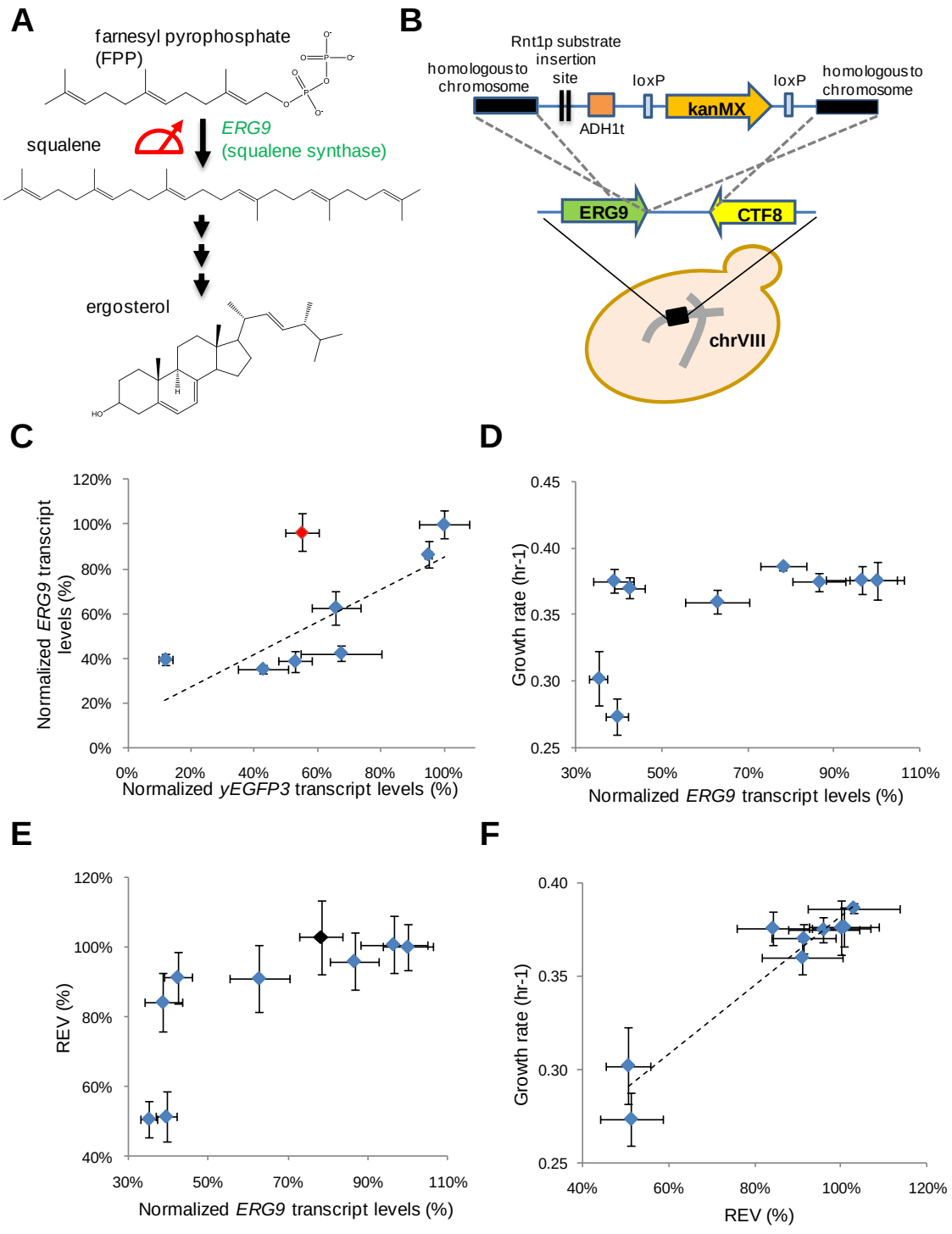


Figure 2.6. Synthetic Rnt1p hairpins enable posttranscriptional control over endogenous *ERG9* expression levels. (A) Simplified schematic of ergosterol biosynthesis from FPP showing key components for this work. Squalene is converted to ergosterol through 14

enzymatic steps. The dial highlights that *ERG9* levels are tuned with the synthetic Rnt1p control modules. (B) Schematic of the construct and strategy utilized for introducing the synthetic Rnt1p control modules into the 3' UTR of the endogenous *ERG9* gene. The construct is designed to replace the native *ERG9* 3' UTR with a synthetic 3' UTR harboring an Rnt1p hairpin through homologous recombination between the integration cassette and chrVIII. The illustrated strategy maintains the native feedback regulation acting through transcriptional mechanisms to control *ERG9* levels. (C) Correlation analysis of *yEGFP3* transcript levels and *ERG9* transcript levels indicates that the synthetic Rnt1p hairpins maintain their gene regulatory activity in a different genetic context. Normalized *ERG9* transcript levels and their error are determined as described in Figure 2.1D. Red data point, C06. (D) Correlation analysis of cellular growth rate and *ERG9* transcript levels indicates that the titration of *ERG9* levels results in two distinct phenotypic regimes – 'fast-growing' and 'slow-growing'. Growth rates are determined by measuring the OD_{600} during a time course and fitting the data to an exponential growth curve using Prism 5 and standard error was calculated from the software. Black data point, wild-type yeast strain. (E) Correlation analysis of relative ergosterol values (REVs) and *ERG9* transcript levels indicates that ergosterol levels remain relatively consistent across varying *ERG9* levels above a certain threshold value (approximately 40% normalized transcript levels). REVs are determined by extracting unsaponified sterols and measuring the absorbance of signature peaks associated with ergosterol in the UV spectrum. Reported REV values and their error are calculated from the mean and standard deviation from the three identical aliquots from sterol extractions, respectively (F) Correlation analysis of cellular growth rate and REV indicates that the two phenotypic measures of *ERG9* levels are strongly correlated.

of both genes. Thus, we designed a construct to integrate the entire 3' UTR and ADH1 terminator from the library plasmid (pCS321) immediately following the *ERG9* stop codon and before the intervening nucleotides between *ERG9* and *CTF8* (Figure 2.6B). We built an Rnt1p control module integration plasmid based on the library plasmid, where *yEGFP3* was replaced by *ERG9* and the marker *loxP-KanMX-loxP*[45] was inserted downstream of the ADH1 terminator to provide resistance against G418. PCR-amplification from the integration plasmid results in DNA cassettes that can be directly integrated into the desired location of the yeast genome. Six members of the synthetic Rnt1p library (A01, A02, C06, C07, C08, C10) along with a mutant tetraloop control (C13, GAAA tetraloop) were integrated into the 3' UTR of *ERG9* to cover the regulatory range of the library.

The regulatory activity of the seven examined Rnt1p hairpins for *ERG9* was initially assessed by measuring *ERG9* transcript levels (Table 2.3). The *ERG9* transcript levels were directly compared to the transcript levels measured from the *yEGFP3* construct. The Rnt1p hairpins resulted in reduced transcript levels compared to the integrant harboring no Rnt1p hairpins and to the construct harboring a mutated tetraloop. The transcript levels of the no hairpin integrant were higher than the transcript levels from the wild-type yeast strain, indicating that the 3' UTR replacement results in increased *ERG9* transcript levels, likely due to altered transcript stability. In addition, a plot of the normalized *yEGFP3* transcript levels versus normalized *ERG9* transcript levels for each synthetic Rnt1p hairpin reveals a strong positive correlation ($r = 0.844$) between the two measures when the hairpin C06 is excluded from the analysis, supporting the ability of the synthetic Rnt1p hairpins to act as predictable genetic control modules (Figure 2.6C). The modularity of the hairpin set (excluding C06) was further supported by a ρ-value of 0.771, indicating a preservation of rank order. C06 was previously determined to not maintain function in the context of the *ymCherry* construct (Figure 2.4B; red point). The difference in regulatory activity observed from the C06 hairpin in the context of the endogenous *ERG9* gene (Figure 2.6C; red point) further suggests that C06 may not be as well-insulated from different genetic contexts as the other tested hairpins, although RNA folding software does not predict alternative structures. In addition, the *ERG9* levels flatten and do not drop below ~40%, indicating that the natural feedback regulation associated with the *ERG9* promoter may act to maintain levels at this minimum value.

Table 2.3. Gene regulatory and phenotypic measures of the impact of Rnt1p hairpins on *ERG9* expression.

Substrate	Normalized *ERG9* transcript levels (%)	Growth rate (hr^{-1})	REV (%)
wild-type	78% ± 5%	0.386 ± 0.003	103% ± 11%
no insert	100% ± 6%	0.376 ± 0.014	100% ± 7%
C13 (GAAA)	87% ± 6%	0.375 ± 0.007	96% ± 8%
C06	96% ± 8%	0.376 ± 0.010	101% ± 8%
C07	42% ± 4%	0.370 ± 0.008	91% ± 7%
C08	40% ± 3%	0.273 ± 0.014	51% ± 7%
C10	63% ± 7%	0.360 ± 0.009	91% ± 9%
A01	39% ± 5%	0.375 ± 0.009	84% ± 8%
A02	35% ± 2%	0.302 ± 0.021	51% ± 5%

While sterol synthesis is vital to cell growth, knockouts of enzymes in the downstream ergosterol biosynthetic pathway are viable due to the retained ability to incorporate intermediate sterols into the cellular membrane[46]. However, knockouts of enzymes in the early part of the pathway, including *ERG9* and enzymes leading up to the production of FPP, are lethal. As such, we examined the effect of decreased *ERG9* expression on the cell growth rate. The OD_{600} of yeast strains harboring the different Rnt1p hairpins in the 3' UTR of *ERG9* was measured during the exponential growth phase. The growth rate, k, was calculated by fitting the OD_{600} data to an exponential growth curve (Table 2.3). A plot of the growth rate versus *ERG9* transcript levels for each strain reveals that above a certain threshold level of *ERG9* production, differences in growth rates in this 'fast-growing' regime are negligible (Figure 2.6D) This data suggest that decreasing the amount of *ERG9* in this regime does not significantly impact the flux through the ergosterol biosynthetic pathway due to overproduction of *ERG9* in the wild-type strain or feedback control of *ERG9*[47]. Below a certain threshold level a 'slow-growing' regime is observed, characterized by a substantial drop-off in the cell growth

rate. Interestingly, the two slow-growing strains harbor the Rnt1p hairpins exhibiting the strongest gene silencing activities, but demonstrate similar *ERG9* levels to two of the fast-growing strains. It is possible that feedback regulation is acting to increase *ERG9* expression to the desired setpoint in the slow-growing strains, but the perturbations introduced in these strains result in other impacts on the pathway that inhibit the endogenous control systems from restoring cellular growth to wild-type rates.

In order to better understand the cellular processes linking *ERG9* production and cell growth rate, we measured the amount of ergosterol, the end product of the *ERG9* biosynthetic pathway. Each culture was inoculated with the same amount of cells and allowed to grow for 8 hr before the cultures were saponified. The UV spectrum of unsaponified sterols was determined and used to calculate a relative ergosterol value (REV), which is normalized against the control containing no Rnt1p substrate (set to 100%) (Table 2.3). A plot of REV versus *ERG9* transcript levels reveals a similar relationship as observed between growth rate and *ERG9* levels, with the exception that the fast-growing strains exhibit a slight positive correlation between REV and *ERG9* transcript levels (Figure 2.6E). A plot of growth rate versus REV illustrates the two regimes (slow-growing and fast-growing) and highlights the strong positive correlation between these two phenotypic measures ($r = 0.953$) (Figure 2.6F). There is little difference in the amount of ergosterol per cell (determined as the ratio of REV and OD_{600}), indicating that alteration of flux through the *ERG9* pathway results in changes in the time required for a cell to produce sufficient ergosterol to duplicate and not decreased levels of ergosterol molecules in the cell membrane.

2.3. Discussion

We have developed a novel class of genetic control modules in *S. cerevisiae* based on Rnt1p cleavage. Although an Rnt1p substrate has been shown to play a role in regulating the expression of the endogenous *MIG2* gene[36], our work describes the first synthetic gene regulatory system based on engineered Rnt1p hairpins. A library of synthetic Rnt1p hairpins that span a wide range of gene regulatory activities was generated to act as posttranscriptional control modules by placing these elements in the 3' UTR of a target gene. To ensure the modularity of the synthetic Rnt1p substrates, two design strategies were implemented. First, a 'clamp' region was added to the base of each hairpin. Second, only sequences that formed single predicted hairpin structures at the lowest free energies were included within the Rnt1p control module set. These properties minimize any potential of flanking sequences to disrupt desired folding of the control modules, thereby reducing the likelihood of varying function within different genetic contexts. We observed a significant improvement in the correlation between transcript and protein levels for hairpins that exhibited both properties (Supplementary Figure 2.4), suggesting that undesired interactions between hairpin and flanking sequences can affect translation. The functional modularity of the resulting hairpin library, as measured by maintenance of regulatory activity and rank order, was demonstrated under three different genetic contexts (Figure 2.4B, Figure 2.6C), where in each case one hairpin did not retain its expected activity, likely due to improper folding. Our studies indicate that for any given genetic system, one of the library members (less than 10%) may not exhibit regulatory activity, where the coverage of the reported library will allow a researcher to select hairpins that span the desired regulatory range.

Previous *in vitro* studies identified three regions of the hairpin substrates critical to Rnt1p cleavage activity: the CEB, BSB, and IBPB[29]. The BSB and IBPB are regions of the hairpin that affect the overall binding of the protein, such that modifications of the nucleotides in these regions can inhibit Rnt1p binding and subsequent cleavage. The CEB is a region of the hairpin that affects the processing of the stem by Rnt1p, such that nucleotide modifications in this region are expected to specifically modulate the cleavage rate. We developed an Rnt1p library based on randomization of the CEB region and screened this library *in vivo* for substrates with altered processing efficiencies. While earlier *in vitro* studies retained base-pairing within the CEB of modified Rnt1p substrates[29], our library screen demonstrated that the CEB has substantial structural flexibility in maintaining function *in vivo*, as the majority of library members retained processability. Cleavage assays with purified Rnt1p support that variations in the CEB within the set of synthetic Rnt1p substrates alter the processing efficiency in a manner that is directly correlated with the observed gene regulatory activity, whereas binding assays indicate no relationship between binding affinity and processing efficiency (or gene regulatory activity). The data support that modifications in the CEB directly influence the ability of Rnt1p to cleave the hairpin, whereas interactions between the nucleotides in the CEB and RBDs of Rnt1p likely lead to the small variation in observed K_D values[29].

The set of synthetic Rnt1p substrates developed in this work represents the first engineered library of transcript stability control modules in the yeast *S. cerevisiae*. While other posttranscriptional regulatory elements, such as internal ribosome entry sites (IRESes) and AU-rich elements (AREs), have been applied to regulate heterologous gene

expression in yeast, such genetic elements have exhibited substantial variability in activity and have not been engineered as synthetic libraries of control modules exhibiting a wide range of activities[48-50]. In addition, a library of short synthetic internal ribosome entry sites (IRESes) that act through translation initiation was previously developed for yeast[51]. However, these short IRESes result in substantially reduced expression levels compared with cap-dependent translation mechanisms, such that the resulting library spans a much narrower range of regulatory activities than exhibited by the synthetic Rnt1p library. The most similar control module library is one that was developed based on mutating a constitutive promoter (TEF1) in yeast, which spans a similar range of gene regulatory activities as the described Rnt1p hairpin library (TEF: 8–120%; Rnt1p: 2–100%)[26]. However, the two libraries exhibit different coverage of these ranges, where the Rnt1p library provides greater coverage of expression levels between 2–60% and the TEF promoter library provides greater coverage between 60–100%.

A unique advantage of control modules based on posttranscriptional processes is that such elements can be readily used in combination with one another and with other genetic control modules, such as promoter elements and other transcriptional regulators, to achieve more finely-tuned and expanded regulatory schemes. As one example, inducible promoters are commonly used to turn on and set the expression levels of genes by controlling the concentration of the inducing molecule exogenously added to the system. However, such transcriptional control modules on their own are limited to applying identical regulatory activities to multiple gene targets within a given system. The combination of an inducible promoter and our engineered Rnt1p substrates will allow for the relative gene regulatory activities at a given inducer concentration to be

modulated (based on the Rnt1p substrate), thus enabling ligand-mediated control over multiple genes with different expression levels.

In addition to expanding the utility of inducible promoters in the context of multi-gene circuits, the ability to combine posttranscriptional control modules with any promoter element has key advantages in the context of endogenous networks. Endogenous networks often have critical control strategies in place, such as feedback regulation, that commonly operate at the level of transcriptional processes. The combination of the synthetic Rnt1p substrates with endogenous genetic targets allows specific engineered control strategies to be added to a system, while retaining native regulatory schemes that may play an important role in the overall system operation. Therefore, these posttranscriptional control modules provide a useful toolset for predictably modulating specific components in complex biological systems and can be further used to probe and study native regulatory networks. One consideration in the implementation of these genetic modules is that their gene regulatory activities may be affected by variation in the ratio of cellular levels of Rnt1p to transcript levels. While absolute activities of the synthetic Rnt1p hairpins are expected to vary with substantial changes in Rnt1p levels, the rank order of the hairpin activities is expected to be maintained.

To demonstrate the utility of these posttranscriptional control modules, we implemented a synthetic control strategy directed to modulating a key enzyme component of the endogenous ergosterol synthesis network by combining the Rnt1p control modules with the *ERG9* genetic target. Previous work had replaced the endogenous *ERG9* promoter with a MET3 repressible promoter, and demonstrated a

sharp decrease in ergosterol levels with full transcriptional repression[12]. In contrast, our engineered control strategy was anticipated to allow the system to retain previously identified transcriptional feedback control around the *ERG9* gene[47], while allowing for titration of *ERG9* transcript levels. Generally, transcript levels between *ERG9* and *yEGFP3* with a given hairpin correlated strongly. However, relative *ERG9* levels did not fall below ~40% regardless of the Rnt1p hairpin strength, indicating an endogenous feedback mechanism that maintains *ERG9* expression levels at that threshold value. Interestingly, the data indicate a 'buffer' region in the endogenous control strategy around *ERG9* levels, where wild-type levels are set substantially higher than that threshold value. The synthetic Rnt1p hairpin set allowed for systematic titration of *ERG9* expression levels and the identification of two regimes for the system. Strains expressing over ~40% *ERG9* transcript levels exhibited high ergosterol levels and growth rates. Strains harboring two synthetic Rnt1p hairpins resulting in the lowest expression levels exhibited a significant reduction in the amount of ergosterol produced and growth rate. Interestingly, these 'slow-growing' strains have similar levels of *ERG9* to two strains in the 'fast-growing' regime. One possible explanation for these observations is that the diminished cellular sterol levels result in positive regulation of the *ERG9* promoter below a certain threshold value, maintaining expression at a minimum level. However, dialing down *ERG9* levels below a critical value can affect the cell through unknown mechanisms that do not permit restoration of ergosterol levels or cell growth rate through the endogenous control system[47]. Therefore, this work supports the unique ability of the synthetic Rnt1p hairpin library to systematically titrate pathway enzyme levels while maintaining native cellular control strategies acting through transcriptional mechanisms.

In summary, we have developed a library of genetic control modules for yeast that can be implemented with different genetic targets and promoters to predictably tune gene expression levels. The Rnt1p library provides a key tool for synthetic biology applications in yeast, which can function to rationally dial in expression levels similar to well-developed control modules in bacteria such as ribosome binding sites (RBS). However, unlike RBS elements the structural and functional insulation of the synthetic Rnt1p controllers provide for more successful maintenance of regulatory activities across different genetic contexts[52]. We show here that the synthetic controllers can be applied to predictably modulate flux through metabolic pathways and probe regulation schemes in endogenous networks by introducing precise perturbations around major control points. With growing interests in eukaryotic hosts and complex networks in synthetic biology, and more specifically yeast in bioprocessing and biosynthesis applications, the synthetic controllers developed here will provide an important foundational toolset for the rapidly growing field.

2.4. Materials and Methods

2.4.1. Plasmid construction

Standard molecular biology techniques were utilized to construct all plasmids[53]. DNA synthesis was performed by Integrated DNA Technologies (Coralville, IA) or the Protein and Nucleic Acid Facility (Stanford, CA). All enzymes, including restriction enzymes and ligases, were obtained through New England Biolabs (Ipswich, MA) unless otherwise noted. Pfu polymerases were obtained through Stratagene. Ligation products were electroporated into *Escherichia coli* DH10B (Invitrogen, Carlsbad, CA), where cells

harboring cloned plasmids were maintained in Luria-Bertani media containing 50 mg/ml ampicillin (EMD Chemicals). Clones were initially verified through colony PCR and restriction mapping. All cloned constructs and chromosomal integrations were sequence verified by Laragen (Los Angeles, CA) or the Protein and Nucleic Acid Facility (Stanford, CA). Plasmid maps are available in Supplementary Figure 2.5.

A yeast-enhanced GFP gene, *yEGFP3*, was PCR-amplified from pSVA13[38] using forward and reverse primers GFP.mono.di.fwd (5' GCAAGCTTGGAGATCTAAAAGA AATAATGTCT) and GFP.mono.rev (5' CGCTCGAGGCCTAGGCTTTATTTGTACA ATT), respectively. The plasmid pCS182 was constructed by inserting the *yEGFP3* PCR product into a modified version of pRS316[54] harboring the GAL1-10 promoter via the unique restriction sites HindIII and XhoI located in the multiple cloning site (MCS) downstream of the GAL1-10 promoter. The ADH1 terminator was PCR-amplified from pSVA13[38] using the forward and reverse primers ADH1t_fwd (5' GCACCTCGAGAGG GCGCGCCACTTC) and ADH1t_rev (5' GCACGGTACCTATATTACCCTGTTATCC CTAGCGG), respectively. The base Rnt1p substrate characterization plasmid (pCS321) was constructed by inserting the ADH1 terminator PCR product into pCS182 via the unique restriction sites XhoI and KpnI located in the MCS. A *ymCherry* characterization plasmid, pCS1749, was constructed from pCS321 by replacing the GAL1-10 promoter with the endogenous TEF1 promoter and by replacing the *yEGFP3* open reading frame (ORF) the ADH1 terminator with the ORF of *ymCherry* and the CYC1 terminator (J. Liang et al., in preparation).

The endogenous *S. cerevisiae* gene *RNT1* was PCR-amplified directly from the yeast genome by colony PCR using forward and reverse primers Rnt1p_prmr_fwd (5'

GACCATGGGGATGGGCTCAAAAGTAGCAGGTAAAAAGAAAACC) and Rnt1p_ prmr_rev (5’ TGCTCGAGTCAGCTTGTATCTGAGAATTTTCTTTTCTTATTCTTTT GTGAG), respectively. The Rnt1p expression plasmid (pRNT1) was constructed by inserting the *RNT1* PCR product into the pProEx HT plasmid (Stratagene) via the unique restriction sites NcoI and XhoI located in the MCS downstream of the 5’ $(His)_6$ tag, the spacer region, and rTEV protease cleavage site.

The endogenous *S. cerevisiae* gene *ERG9* was PCR-amplified directly from the yeast genome by colony PCR using forward and reverse primers ERG9_321_prmr_fwd (5’ GCGAAGCTTGGAGATCTAAAAGAAATAATGGGAAAGCTATTACAATTGGC ATTGCATCC and ERG9_321_prmr_rev (5’ GGCTCGAGGCCTAGGCTTCACGCTCT GTGTAAAGTGTATATATAATAAAACCCAAGAAGA), respectively. The plasmid pCS321-ERG9 was constructed by replacing the *yEGFP3* sequence in pCS321 with *ERG9* by cloning the *ERG9* PCR product into the unique restriction sites HindIII and XhoI. The plasmid pUG6[45] was modified by removing the unique XhoI restriction site by site-directed mutagenesis via the oligonucleotides c1546g (5’ GTGTCGAAAACGAGCT CTGGAGAACCCTTAATATAAC) and c1546g_antisense (5’ GTTATATTAAGGGTT CTCCAGAGCTCGTTTTCGACAC) and the PfuUltra II polymerase (Stratagene). A PCR product harboring the full *ERG9* coding sequence through the 3’ end of the ADH1 terminator was amplified from pCS321-ERG9 using forward and reverse primers ERG9hpADH1t-SalI_fwd_prmr (5’ CAACGTCGACATGGGAAAGCTATTACAATTG GCA) and ERG9hpADH1t-SalI_rev_prmr (5’ AAGTGTCGACTATATTACCCTGTTA TCCCTAGCGG), respectively. The ERG9-RNT1 integration plasmid (pCS1813) was

constructed by inserting this PCR product into the modified pUG6 plasmid via the unique restriction site SalI located directly upstream of the first loxP site.

Insertion of engineered Rnt1p substrates and appropriate controls into the 3' UTR of pCS321, pCS1749, and pCS1813 was facilitated through either digestion with the appropriate restriction endonucleases and ligation-mediated cloning or homologous recombination-mediated gap-repair during transformation into *S. cerevisiae* strain W303 (*MATa, his3-11,15 trp1-1 leu2-3 ura3-1 ade2-1*) through standard lithium acetate procedures[55]. The Rnt1p substrates were amplified for insertion into pCS321 and pCS1813 with both techniques using the forward and reverse primers RntGap321_fwd (5' ACCCATGGTATGGATGAATTGTACAAATAAAGCCTAGGTCTAGAGGCG) and RntGap321_rev2 (5' TAAGAAATTCGCTTATTTAGAAGTGGCGCGCCCTCTCG AGGGCG), respectively. The Rnt1p substrates were amplified for insertion into pCS1749 by gap-repair using the forward and reverse primers mCherry_gap_fwd_prmr (5' GGTGGCATGGATGAACTATACAAATAATAAAGCCTAGGTCTAGAGGCG) and mCherry_gap_rev_prmr (5' TGACATAACTAATTACATGATGCGGCCCTCCCC TCTCGAGGGCG). In the case of digestion and ligation, the PCR products were digested with the unique restriction sites AvrII and XhoI, which are located 3 nts downstream of the *yEGFP3* or *ERG9* stop codon and upstream of the ADH1 terminator. Following construction and sequence verification of the desired vectors, 100–500 ng of each plasmid was transformed into W303. In the case of gap-repair (for pCS321 and pCS1749), 250–500 ng of the PCR product and 100 ng of pCS321 digested with AvrII and XhoI were transformed into the yeast strain. All yeast strains harboring cloned

plasmids were maintained on synthetic complete media with an uracil dropout solution containing 2% dextrose at 30°C.

2.4.2. 3' UTR replacement cassette and integration

The ERG9-RNT1 replacement cassettes were synthesized through PCR amplification from the appropriate pCS1813-based plasmids using forward and reverse primers ERG9-1150_fwd_prmr (5' AATTACCTCCTAACGTGAAGCCAAATGAAAC TCCAATTTTCTTGAAAGTT) and Rnt1p_cassette_rev_prmr2 (5' GGCCTCTACCTAT TATGTAAGTACTTAGTTATTGTTCGGAGTTGTTTGTTAATACGACTCACTATA GGGAGACCGGCAGA), respectively. These PCR products extend from 1150 nts into the *ERG9* gene to the end of the second loxP site with an overhang extension comprising 50 nts of homology to the native *ERG9* 3' UTR. Each integration cassette (~1–5 μg) was transformed into yeast as previously described. The integrants were selected and maintained on YPD plates with 200 mg/ml G418.

2.4.3. Rnt1p substrate characterization assays

S. cerevisiae cells harboring pCS321-based and pCS1749-based plasmids were grown on synthetic complete media with an uracil dropout solution and the appropriate sugars (2% raffinose and 1% sucrose for pCS321; 2% dextrose for pCS1749) overnight at 30ºC. The cells were back-diluted the following morning into fresh media (4.5 ml total volume in test tubes and 450 μl in deep-well plates) to an optical density at 600 nm (OD_{600}) of 0.1 and grown again at 30ºC. For pCS321-based plasmids, after 1 hr, 0.5 ml (test tubes) or 50 μl (plates) of 20% galactose (2% final concentration) or water (non-

induced control) was added to the cell cultures. The cells were grown for another 4.5 hr before measuring the fluorescence levels or collecting cells for RNA extraction.

S. cerevisiae integrated with Rnt1p hairpins or its controls were grown on YPD overnight at 30ºC. The cells were back-diluted the following morning into fresh media (5 ml total volume in test tubes) and grown again for 3 hr at 30ºC. After 3 hr, the cells were back-diluted to an OD_{600} of 0.1 (for RNA extraction and growth rate determination, 5 ml total volume) or 0.05 (for ergosterol quantification, 7 ml total volume) and grown for an appropriate length of time dependent on the application.

2.4.4. Fluorescence quantification

On a SAFIRE plate reader (TECAN, Männedorf, Switzerland), GFP fluorescence was read from 200 µl of cells with an excitation wavelength of 485 nm, an emission wavelength of 515 nm, and a gain of 100. The population-averaged fluorescence readings were normalized to the amount of cells by dividing the relative fluorescence units (RFU) by the OD_{600} of the sample. On the Quanta flow cytometer (Beckman Coulter, Fullerton, CA), the distribution of GFP fluorescence was measured with the following settings: 488-nm laser line, 525-nm bandpass filter, and photomultiplier tube setting of 5.83. Data were collected under low flow rates until 10,000 viable cell counts were collected. A non-induced cell population was used to set a gate to represent GFP-negative and GFP-positive populations. The median fluorescence of the positive population was measured from three identically grown samples. The LSRII flow cytometer (Becton Dickinson Immunocytometry Systems) was used to measure ymCherry fluorescence from p1749-based plasmids. ymCherry was excited at 532 nm and measured with a splitter of 600 nm

LP and a bandpass filter of 610/20 nm. A DAPI stain (excited at 405 nm and measured with a bandpass filter of 450/50 nm) was utilized to gate for cell viability. The mean fluorescence was measured from three identically grown samples and baseline-subtracted with an empty vector control. Reported values and their error are calculated from the mean and standard deviation from the triplicate data, respectively.

2.4.5. Quantification of cellular transcript levels

Total RNA from *S. cerevisiae* was collected by a standard hot acid phenol extraction method[56] and followed by DNase I (New England Biolabs) treatment to remove residual plasmid DNA according to manufacturer's instructions. cDNA was synthesized from 5 μg of total RNA with gene-specific primers for *yEGFP3*, *ymCherry*, *ERG9*, and *ACT1*[57] (rnt1p_rtpcr_rev2 and ACT1_rtpcr_rev, respectively) and SuperScript III Reverse Transcriptase (Invitrogen) according to manufacturer's instructions. The forward and reverse primers for *yEGFP3* quantification are rnt1p_rtpcr_fwd2 (5' CGGT GAAGGTGAAGGTGATGCTACT) and rnt1p_rtpcr_rev2 (5' GCTCTGGTCTTGTAGT TACCGTCATCTTTG), respectively; for *ymCherry* quantification are mCherry_qrtpcr_ fwd (5' AAGGGTTTAAGTGGGAGCGTGTGA) and mCherry_qrtpcr_rev (5' AAGGC ACCATCTTCAGGGTACATTCG), respectively; for *ERG9* quantification are erg9_ rtpcr_fwd (5' AACTGTTGAACTTGACCTCCAGATCGTTTG) and erg9_rtpcr_rev (5' GGCTCTGTCCTTCACATCGGGGGCATTTCC), respectively; for *ACT1* quantification are ACT1_rtpcr_fwd (5' GGCATCATACCTTCTACAACGAAT) and ACT1_rtpcr_rev (5' GGAATCCAAAACAATACCAGTAGTTCTA), respectively. Relative transcript levels and their error were quantified in triplicate from three identical reactions from the

cDNA samples by using an appropriate primer set and iQ SYBR Green Supermix (Bio-Rad, Hercules, CA) on an iCycler iQ qRT-PCR machine (Bio-Rad) according to the manufacturer's instructions. For each run, a standard curve was generated for either *yEGFP3*, *ymCherry*, or *ERG9* and a house-keeping gene, *ACT1*, using a dilution series for a control representing no insertion of an Rnt1p substrate. Relative transcript levels were first individually determined for each sample and then the values for *yEGFP3*, *ymCherry*, and *ERG9* were normalized by their corresponding *ACT1* values.

2.4.6. Cell growth rate determination

At multiple time points during a course of 7 hr, 200 μl were taken from a yeast culture and the OD_{600} measured on a SAFIRE plate reader. The growth rate, k, and its standard error were analyzed using Prism 5 by fitting the data to an exponential growth curve.

2.4.7. Cellular ergosterol quantification

The method for quantification of cellular ergosterol levels was adapted from previously developed protocols[58-59]. Briefly, yeast cells were harvested after 8 hr with the OD_{600} recorded and collected by centrifugation at 3,000 rpm for 5 min. The cells were washed with water and centrifuged again. 10 ml of 25% alcoholic KOH [25% KOH, 60% (v/v) ethanol] was added to the cell pellet and vortexed. The suspension was transferred to a 50 ml Falcon tube and saponified by incubating at 90°C for 3 hr. After cooling to room temperature, the nonsaponified sterols were extracted by adding 5 ml of heptane, vortexing, and collecting the heptane layer once it had clarified. The heptane layer was

directly applied to a 96-well UV plate (Greiner Bio-One) and its absorbance was read in the UV spectrum on a SAFIRE plate reader. A relative ergosterol value (REV) was calculated from the following equation:

$$REV = \frac{OD_{281.5}}{290} - \frac{OD_{230}}{518}$$

The reported value and error were determined from the mean and standard deviation, respectively, from three individual heptanes aliquots.

2.4.8. In vitro *transcription of Rnt1p substrates*

All Rnt1p substrates were PCR-amplified to include an upstream T7 promoter site and A-rich sequences flanking the hairpin using forward and reverse primers Rnt1p-T7-PCR_fwd_prmr (5' TTCTAATACGACTCACTATAGGGACCTAGGAAACAAACAA AGTTGGGC) and Rnt1p-T7-PCR_rev_prmr (5' CTCGAGTTTTTATTTTTCTTTTTGC CGGGCG), respectively. 1–2 μg of PCR product was transcribed with T7 Polymerase (New England Biolabs) in the presence and absence of α-P^{32}-GTP. The 25-μl reaction consisted of the following components: 1x RNA Pol Reaction Buffer (New England Biolabs), 3 mM rATP, 3 mM rCTP, 3 mM rUTP, 0.3 mM rGFP, 1 μl RNaseOUT (Invitrogen), 10 mM $MgCl_2$, 2 mM DTT, 1 μl T7 Polymerase, and 0.5 μCi α-P^{32}-GTP. Unincorporated nucleotides were removed from the reactions by running the samples through NucAway Spin Columns (Ambion, Austin, TX) according to the manufacturer's instructions.

2.4.9. Rnt1p expression and purification

The pRNT1 plasmid was transformed into *E. coli* strain BL21 using the Z-competent *E. coli* Transformation Kit and Buffer Set (Zymo Research, Orange, CA) according to manufacturer's instructions. Rnt1p was collected as a protein extract as previously described[60]. Briefly, an overnight culture of BL21 cells harboring pRNT1 was back-diluted to an OD_{600} of 0.5. Once the culture reached an OD_{600} of 1.1–1.4, it was induced with 1 mM IPTG and grown for an additional 3 hr. The cells were centrifuged at 2,500g for 12 min at 4°C and the resulting cell pellet was frozen in a -80°C freezer. After weighing the frozen cell pellet, the cells were resuspended in 4 ml Ni_{2+} buffer [25% (v/v) glycerol, 1 M NaCl, 30 mM Tris pH 8.0] per gram of harvested cells. The resuspension was sonicated (Heat Systems-Ultrasonics, Inc.) twice with the following settings: 2 x 30 sec, output control 5, and 50% duty cycle. Cellular debris was removed by centrifugation at 20,000g for 30 min at 4°C and the supernatant was filtered through a 0.2-μm pore size Acrodisc 25 mm syringe filter (Pall Life Sciences, Ann Arbor, MI).

Rnt1p was purified from the resulting supernatant with one 1-ml HisTrap HP column (GE Healthcare) on an AKTA FPLC machine (GE Healthcare). Elution of the protein was performed with an imidazole concentration of 150 mM in Ni_{2+} buffer and the protein was collected in 6 1-ml fractions. Protein purification was confirmed by analyzing an aliquot of each fraction on a SDS-PAGE gel (NuPAGE 4–12% Bis-Tris Gel, Invitrogen) and protein function was confirmed by incubating an aliquot of each fraction with a control Rnt1p substrate and analyzing the resulting cleavage products on an 8% denaturing polyacrylamide gel. Positive fractions were pooled and concentrated to less than a 3-ml volume using a Centricon Centrifugal Filter Device (10,000 MWCO;

Millipore) according to the manufacturer's instructions. The concentrated protein was then injected into a Slide-A-Lyzer Dialysis Cassette (10,000 MWCO; Pierce Biotechnology) and buffer-exchanged twice with Rnt1p Storage Buffer [50% (v/v) glycerol, 0.5 M KCl, 30 mM Tris pH 8.0, 0.1 M DTT, 0.1 M EDTA] at 4ºC. The first buffer exchange took place for 4 hr and the second buffer exchange occurred overnight. The purified Rnt1p was stored in aliquots at –20°C.

2.4.10. In vitro *Rnt1p substrate cleavage assay*

Cleavage assays were performed on Rnt1p substrates as previously described[39, 60]. Briefly, a 10-µl mixture of RNA and Rnt1p was incubated at 30ºC for 15 min in Rnt1p reaction buffer [30 mM Tris (pH 7.5), 150 mM KCl, 5 mM spermidine, 20 mM $MgCl_2$, 0.1 mM DTT, and 0.1 mM EDTA (pH 7.5)]. RNA concentrations were varied from 0.1 to 1.0 µM and the Rnt1p concentration was 2.3 µM. The cleavage reaction products were separated on an 8% denaturing polyacrylamide gel run at 35 W for 30 min. Gels were transferred to filter paper and analyzed for relative substrate and product levels through phosphorimaging analysis on a FX Molecular Imager (Bio-Rad). The levels of cleaved RNA product were determined and fit to a Michaelis-Menten model using Prism 5 (GraphPad), where a relative V_{max} was calculated and reported with the standard error determined by the fit of the model.

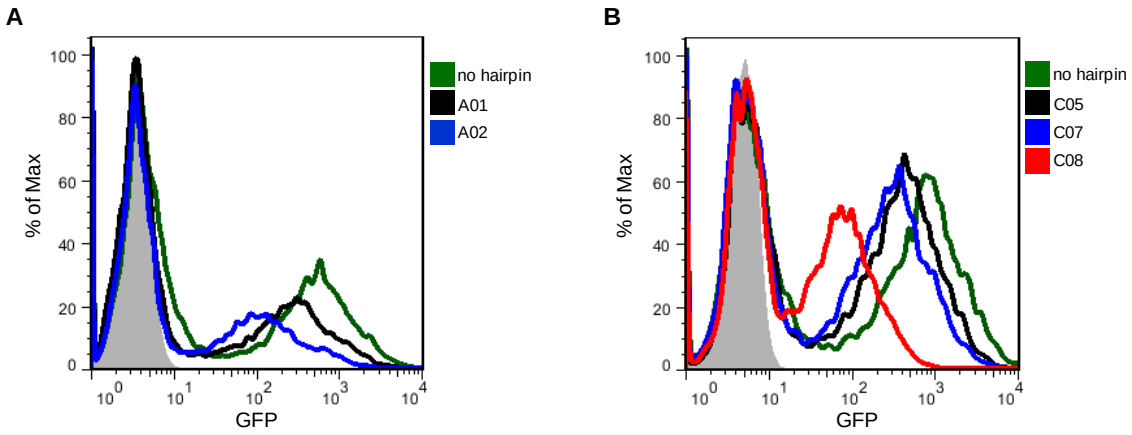
2.4.11. In vitro *Rnt1p substrate mobility shift assay*

Mobility shift assays were performed as previously described[39, 60]. Briefly, a 10-µl mixture of RNA and Rnt1p were incubated on ice for 10 min in Rnt1p binding buffer

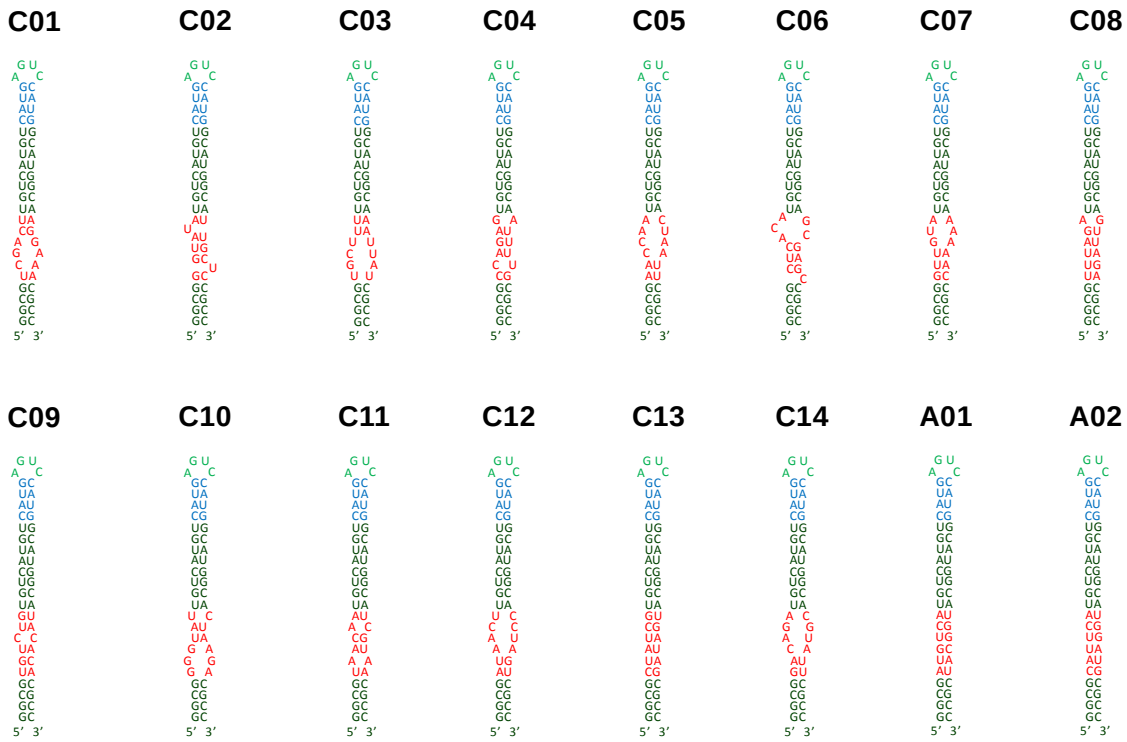
[20% (v/v) glycerol, 30 mM Tris (pH 7.5), 150 mM KCl, 5 mM spermidine, 0.1 mM DTT, and 0.1 mM EDTA (pH 7.5)]. The RNA concentration in all samples was 200 nM and the Rnt1p concentration ranged from 0 to 1.7 μM. The binding reaction products were separated on a 6% native polyacrylamide gel run at 350 V until the samples entered the gel and then at 150 V for 2 hr. Gels were transferred to filter paper and analyzed for free RNA and RNA-Rnt1p complex levels through phosphorimaging analysis on a FX Molecular Imager. The fraction of unbound RNA to total RNA was determined and fit to a modified Scatchard model using Prism 5, where a K_D value was calculated and reported with the standard error determined by the fit of the model.

2.5. Supplementary Information

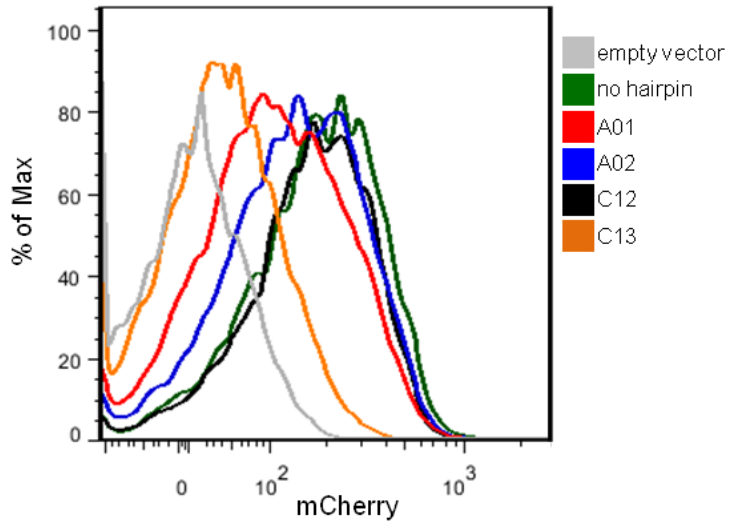
Supplementary Figures and Tables



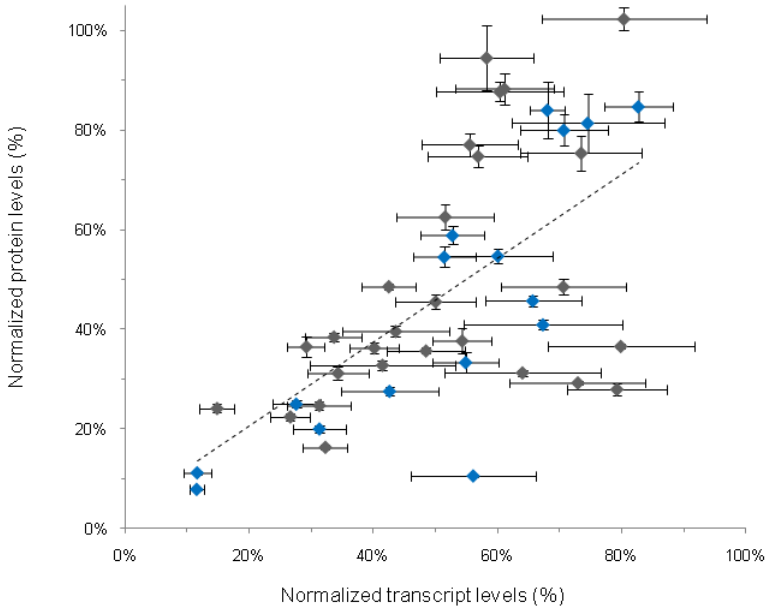
Supplementary Figure 2.1. Flow cytometry histograms of pCS321-based constructs bearing control and library Rnt1p hairpins. The shaded population (grey) indicates the noninduced cell population generated by analyzing cells harboring the 'no insert' control in the absence of galactose. Histograms are representative of three independent experiments. (A) Histograms of constructs bearing the A01 and A02 control hairpins. (B) Histograms of constructs bearing the C05, C07, and C08 library hairpins.



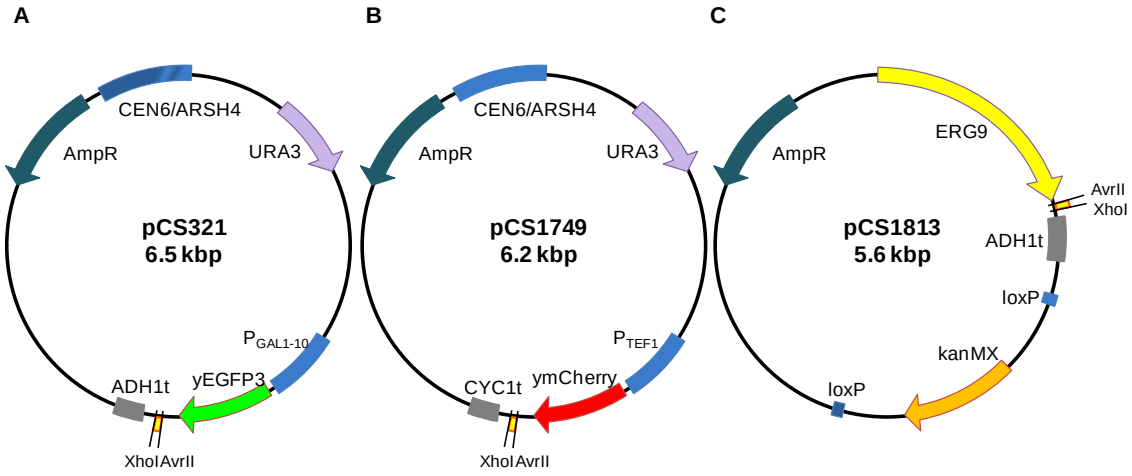
Supplementary Figure 2.2. Sequences and structures of the final Rnt1p cleavage library and the positive control Rnt1p hairpins.



Supplementary Figure 2.3. Flow cytometry histograms of pCS1749-based constructs bearing control and library Rnt1p hairpins. The empty vector histogram (grey) lacks a fluorescent gene and is indicative of the cell and media autofluorescence. Histograms are representative of three independent experiments.



Supplementary Figure 2.4. Correlation analysis of protein and transcript levels for all hairpins identified from the fluorescence-based *in vivo* screening assay. Color scheme is as follows: final library hairpins selected for structural stability to act as modular control elements, blue; library hairpins not selected for inclusion in the final Rnt1p hairpin library, black. The regression line is determined from the entire data set and indicates an r-value of 0.629. The r-value for an identical analysis performed on the data set in blue indicates a stronger correlation (r value = 0.817, Figure 2.3C). All normalized protein and transcript levels are determined as described in Figure 2.1D.



Supplementary Figure 2.5. Plasmid maps for key constructs used in this work. (A) Plasmid map of pCS321, the *yGFP3* characterization plasmid. (B) Plasmid map of pCS1749, the *ymCherry* characterization plasmid. (C) Plasmid map of pCS1813, the Rnt1p hairpin integration plasmid.

Supplementary Table 2.1. Sequence and i*n vivo* characterization data for all screened Rnt1p hairpins. The nucleotides of the cleavage efficiency box are indicated in red. Constructs labeled with 'C' or 'A' represent the final selected cleavage library and positive controls. Constructs labeled with 'c' represent screened hairpins not selected for inclusion in the final Rnt1p hairpin library. All normalized protein and transcript levels are determined as described in Figure 2.1D.

Substrate	Sequence	Normalized protein levels (%)	Normalized transcript levels (%)
C01	GGCGUCGACUUGUCAU GUCAUGAGUCCAUGGC AUGGCAAGGAAACGCC	84% ± 6%	68% ± 3%
C02	GGCGGGUAUAUGUCAU GUCAUGAGUCCAUGGC AUGGCAUUGCUCCGCC	80% ± 3%	71% ± 7%
C03	GGCGUGCUUUUGUCAU GUCAUGAGUCCAUGGC AUGGCAAAUUAUCGCC	55% ± 1%	60% ± 9%
C04	GGCGCCAGAGUGUCAU GUCAUGAGUCCAUGGC AUGGCAAUUUUGCGCC	20% ± 1%	31% ± 4%
C05	GGCGAACCAAUGUCAU GUCAUGAGUCCAUGGC AUGGCACUAAUUCGCC	55% ± 2%	51% ± 5%
C06	GGCGCUCACAUGUCAU GUCAUGAGUCCAUGGC AUGGCAGCGAGCCGCC	33% ± 2%	55% ± 5%
C07	GGCGGUUGUAUGUCAU GUCAUGAGUCCAUGGC AUGGCAAAAAACCGCC	41% ± 1%	67% ± 13%
C08	GGCGUUUAGAUGUCAU GUCAUGAGUCCAUGGC AUGGCAGUUAGACGCC	11% ± 0%	12% ± 2%
C09	GGCGUGUCUGUGUCAU GUCAUGAGUCCAUGGC AUGGCAUACACACGCC	25% ± 1%	28% ± 4%
C10	GGCGGGGUAUUGUCAU GUCAUGAGUCCAUGGC AUGGCACUAAGACGCC	46% ± 1%	66% ± 8%
C11	GGCGUAACAAUGUCAU GUCAUGAGUCCAUGGC AUGGCAUCGUAACGCC	11% ± 0%	56% ± 10%
C12	GGCGAUAACUUGUCAU GUCAUGAGUCCAUGGC AUGGCACCUAGUCGCC	81% ± 6%	75% ± 12%
C13	GGCGCUAUCGUGUCAU GUCAUGAGUCCAUGGC AUGGCAUGAUAGCGCC	8% ± 0%	12% ± 1%
C14	GGCGGACAGAUGUCAU GUCAUGAGUCCAUGGC AUGGCACGUAUUCGCC	85% ± 3%	83% ± 6%
A01	GGCGAUGUCAUGUCAU GUCAUGAGUCCAUGGC AUGGCAUGGCAUCGCC	59% ± 2%	53% ± 5%
A02	GGCGCAUUCAUGUCAU GUCAUGAGUCCAUGGC AUGGCAUGGAUGCGCC	28% ± 1%	43% ± 8%
c01	GGCGACUUACUGUCAU GUCAUGAGUCCAUGGC AUGGCAUAUGCCCGCC	102% ± 2%	80% ± 13%

c02	GGCGAUUCGCUGUCAU GUCAUGAGUCCAUGGC AUGGCAAAACGCCGCC	95% ± 6%	58% ± 8%
c03	GGCGAACUUAUGUCAU GUCAUGAGUCCAUGGC AUGGCAUAAUAACGCC	31% ± 1%	64% ± 13%
c04	GGCGGGACAGUGUCAU GUCAUGAGUCCAUGGC AUGGCAUAGUUGCGCC	22% ± 1%	27% ± 3%
c05	GGCGUUAUGAUGUCAU GUCAUGAGUCCAUGGC AUGGCAAUGUUGCGCC	36% ± 1%	40% ± 4%
c06	GGCGGUCGCAUGUCAU GUCAUGAGUCCAUGGC AUGGCAGUCUAACGCC	37% ± 2%	29% ± 3%
c07	GGCGUUUGGCUGUCAU GUCAUGAGUCCAUGGC AUGGCAAUCAUGCGCC	16% ± 0%	32% ± 4%
c08	GGCGAUGAAAUGUCAU GUCAUGAGUCCAUGGC AUGGCAUUUACGCGCC	38% ± 2%	54% ± 5%
c09	GGCGGUAAAGUGUCAU GUCAUGAGUCCAUGGC AUGGCAUUAUAGCGCC	46% ± 1%	50% ± 6%
c10	GGCGUUUUAGUGUCAU GUCAUGAGUCCAUGGC AUGGCAAAAAUGCGCC	75% ± 3%	73% ± 10%
c11	GGCGGGUAGUUGUCAU GUCAUGAGUCCAUGGC AUGGCAAGUGGACGCC	24% ± 1%	15% ± 3%
c12	GGCGAUUCAGUGUCAU GUCAUGAGUCCAUGGC AUGGCAGGGUAUCGCC	25% ± 1%	31% ± 5%
c13	GGCGAAGCCGUGUCAU GUCAUGAGUCCAUGGC AUGGCAUCUACACGCC	88% ± 3%	61% ± 8%
c14	GGCGCCGGAAUGUCAU GUCAUGAGUCCAUGGC AUGGCAGUUGAGCGCC	49% ± 2%	71% ± 10%
c15	GGCGACAUUGUGUCAU GUCAUGAGUCCAUGGC AUGGCAACGUUGCGCC	33% ± 1%	41% ± 12%
c16	GGCGCCUGCAUGUCAU GUCAUGAGUCCAUGGC AUGGCAGGCCAUCGCC	40% ± 1%	44% ± 9%
c17	GGCGGAUCCAUGUCAU GUCAUGAGUCCAUGGC AUGGCAAUGAAGCGCC	36% ± 1%	48% ± 6%
c18	GGCGGUAGGGUGUCAU GUCAUGAGUCCAUGGC AUGGCAACGUAGCGCC	37% ± 1%	80% ± 12%
c19	GGCGAGUAGGUGUCAU GUCAUGAGUCCAUGGC AUGGCACGCACACGCC	28% ± 1%	79% ± 8%
c20	GGCGGUUUGAUGUCAU GUCAUGAGUCCAUGGC AUGGCAGCAGUUCGCC	29% ± 0%	73% ± 11%
c21	GGCGUUUUAAUGUCAU GUCAUGAGUCCAUGGC AUGGCAAUUGUUCGCC	38% ± 1%	34% ± 5%
c22	GGCGAUUAUGUGUCAU GUCAUGAGUCCAUGGC AUGGCAGUUAUGCGCC	49% ± 1%	42% ± 4%

c23	GGCGAUGUGUUGUCAU GUCAUGAGUCCAUGGC AUGGCACAGACACGCC	88% ± 2%	60% ± 10%
c24	GGCGAAUUUUUGUCAU GUCAUGAGUCCAUGGC AUGGCAGUAGGUCGCC	63% ± 2%	52% ± 8%
c25	GGCGCUAUCAUGUCAU GUCAUGAGUCCAUGGC AUGGCACGUAAUCGCC	31% ± 1%	34% ± 5%
c26	GGCGUACACAUGUCAU GUCAUGAGUCCAUGGC AUGGCAUUCUUGCGCC	77% ± 2%	55% ± 8%
c27	GGCGACUUAUUGUCAU GUCAUGAGUCCAUGGC AUGGCAUCAAUACGCC	75% ± 2%	57% ± 8%

Supplementary Table 2.2. *In vivo* characterization data for the Rnt1p cleavage library in the context of *ymCherry* (pCS1749). All normalized protein and transcript levels are determined as described in Figure 2.1D.

Substrate	Normalized protein levels (%)	Normalized transcript levels (%)
C01	100% ± 5%	84% ± 10%
C02	87% ± 4%	90% ± 5%
C03	73% ± 5%	63% ± 5%
C04	2% ± 0%	40% ± 6%
C05	63% ± 5%	69% ± 9%
C06	89% ± 5%	91% ± 6%
C07	53% ± 2%	57% ± 9%
C08	33% ± 2%	57% ± 8%
C09	44% ± 2%	47% ± 5%
C10	72% ± 2%	89% ± 2%
C11	17% ± 1%	23% ± 3%
C12	97% ± 6%	109% ± 5%
C13	10% ± 1%	21% ± 2%
C14	77% ± 4%	101% ± 4%
A01	80% ± 7%	72% ± 5%
A02	56% ± 3%	71% ± 7%

Acknowledgements

We thank K Hoff, S Bastian, and FH Arnold for assistance in the purification of Rnt1p; and J Liang for assistance with *in vitro* assays. This work was supported by the National Science Foundation (CAREER award to CDS; CBET-0917705) and the Alfred P. Sloan Foundation (fellowship to CDS).

References

1. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338 (2000).
2. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339-342 (2000).
3. Basu, S., Mehreja, R., Thiberge, S., Chen, M.T. & Weiss, R. Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A* **101**, 6355-6360 (2004).
4. Pfleger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* **24**, 1027-1032 (2006).
5. Anderson, J.C., Voigt, C.A. & Arkin, A.P. Environmental signal integration by a modular AND gate. *Mol Syst Biol* **3**, 133 (2007).
6. Pitera, D.J., Paddon, C.J., Newman, J.D. & Keasling, J.D. Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. *Metab Eng* **9**, 193-207 (2007).
7. Peralta-Yahya, P.P. & Keasling, J.D. Advanced biofuel production in microbes. *Biotechnol J* **5**, 147-162 (2010).
8. Jin, Y.S., Ni, H., Laplaza, J.M. & Jeffries, T.W. Optimal growth and ethanol production from xylose by recombinant Saccharomyces cerevisiae require moderate D-xylulokinase activity. *Appl Environ Microbiol* **69**, 495-503 (2003).

9. Jones, K.L., Kim, S.W. & Keasling, J.D. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab Eng* **2**, 328-338 (2000).

10. Zhu, M.M., Skraly, F.A. & Cameron, D.C. Accumulation of methylglyoxal in anaerobically grown Escherichia coli and its detoxification by expression of the Pseudomonas putida glyoxalase I gene. *Metab Eng* **3**, 218-225 (2001).

11. Alper, H., Jin, Y.S., Moxley, J.F. & Stephanopoulos, G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. *Metab Eng* **7**, 155-164 (2005).

12. Paradise, E.M., Kirby, J., Chan, R. & Keasling, J.D. Redirection of flux through the FPP branch-point in Saccharomyces cerevisiae by down-regulating squalene synthase. *Biotechnol Bioeng* **100**, 371-378 (2008).

13. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* **102**, 12678-12683 (2005).

14. Carrier, T.A. & Keasling, J.D. Library of synthetic 5' secondary structures to manipulate mRNA stability in Escherichia coli. *Biotechnol Prog* **15**, 58-64 (1999).

15. Jensen, P.R. & Hammer, K. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl Environ Microbiol* **64**, 82-87 (1998).

16. Smolke, C., Martin, V. & Keasling, J. in Protein expression technologies: current status and future trends. (ed. F. Baneyx) (Horizon Bioscience, Wymondham, Norfolk; 2004).

17. Szczebara, F.M. et al. Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nat Biotechnol* **21**, 143-149 (2003).

18. Hawkins, K.M. & Smolke, C.D. Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat Chem Biol* **4**, 564-573 (2008).

19. Ro, D.K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943 (2006).

20. Nguyen, H.T. et al. Engineering of Saccharomyces cerevisiae for the production of L-glycerol 3-phosphate. *Metab Eng* **6**, 155-163 (2004).

21. Ostergaard, S., Olsson, L. & Nielsen, J. Metabolic engineering of Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **64**, 34-50 (2000).

22. Veen, M. & Lang, C. Production of lipid compounds in the yeast Saccharomyces cerevisiae. *Appl Microbiol Biotechnol* **63**, 635-646 (2004).

23. Louis, M. & Becskei, A. Binary and graded responses in gene networks. *Sci STKE* **2002**, pe33 (2002).

24. Hawkins, K.M. & Smolke, C.D. The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae. *J Biol Chem* **281**, 13485-13492 (2006).

25. Nevoigt, E. et al. Engineering promoter regulation. *Biotechnol Bioeng* **96**, 550-558 (2007).

26. Nevoigt, E. et al. Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. *Appl Environ Microbiol* **72**, 5266-5273 (2006).

27. Caponigro, G. & Parker, R. Mechanisms and control of mRNA turnover in Saccharomyces cerevisiae. *Microbiol Rev* **60**, 233-249 (1996).

28. Filippov, V., Solovyev, V., Filippova, M. & Gill, S.S. A novel type of RNase III family proteins in eukaryotes. *Gene* **245**, 213-221 (2000).

29. Lamontagne, B. et al. Sequence dependence of substrate recognition and cleavage by yeast RNase III. *J Mol Biol* **327**, 985-1000 (2003).

30. Wu, H., Henras, A., Chanfreau, G. & Feigon, J. Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc Natl Acad Sci U S A* **101**, 8307-8312 (2004).

31. Chanfreau, G., Elela, S.A., Ares, M., Jr. & Guthrie, C. Alternative 3'-end processing of U5 snRNA by RNase III. *Genes Dev* **11**, 2741-2751 (1997).

32. Chanfreau, G., Rotondo, G., Legrain, P. & Jacquier, A. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J* **17**, 3726-3737 (1998).

33. Elela, S.A., Igel, H. & Ares, M., Jr. RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell* **85**, 115-124 (1996).

34. Catala, M., Lamontagne, B., Larose, S., Ghazal, G. & Elela, S.A. Cell cycle-dependent nuclear localization of yeast RNase III is required for efficient cell division. *Mol Biol Cell* **15**, 3015-3030 (2004).

35. Lamontagne, B. & Abou Elela, S. Short RNA guides cleavage by eukaryotic RNase III. *PLoS One* **2**, e472 (2007).

36. Ge, D., Lamontagne, B. & Elela, S.A. RNase III-mediated silencing of a glucose-dependent repressor in yeast. *Curr Biol* **15**, 140-145 (2005).

37. Pelletier, J. & Sonenberg, N. Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell* **40**, 515-526 (1985).

38. Mateus, C. & Avery, S.V. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast* **16**, 1313-1323 (2000).

39. Lamontagne, B. & Elela, S.A. Evaluation of the RNA determinants for bacterial and yeast RNase III binding and cleavage. *J Biol Chem* **279**, 2231-2241 (2004).

40. Jeppsson, M., Johansson, B., Jensen, P.R., Hahn-Hagerdal, B. & Gorwa-Grauslund, M.F. The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant Saccharomyces cerevisiae strains. *Yeast* **20**, 1263-1272 (2003).

41. Lamontagne, B., Tremblay, A. & Abou Elela, S. The N-terminal domain that distinguishes yeast from bacterial RNase III contains a dimerization signal required for efficient double-stranded RNA cleavage. *Mol Cell Biol* **20**, 1104-1115 (2000).

42. Khosla, C. & Keasling, J.D. Metabolic engineering for drug discovery and development. *Nat Rev Drug Discov* **2**, 1019-1025 (2003).

43. Bach, T.J. Some New Aspects of Isoprenoid Biosynthesis in Plants - a Review. *Lipids* **30**, 191-202 (1995).

44. Poulter, C.D. & Rilling, H.C. in 413-441 (Wiley, New York; 1981).

45. Guldener, U., Heck, S., Fielder, T., Beinhauer, J. & Hegemann, J.H. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* **24**, 2519-2524 (1996).

46. Daum, G., Lees, N.D., Bard, M. & Dickson, R. Biochemistry, cell biology and molecular biology of lipids of Saccharomyces cerevisiae. *Yeast* **14**, 1471-1510 (1998).

47. Kennedy, M.A., Barbuch, R. & Bard, M. Transcriptional regulation of the squalene synthase gene (ERG9) in the yeast Saccharomyces cerevisiae. *Biochim Biophys Acta* **1445**, 110-122 (1999).

48. Lautz, T., Stahl, U. & Lang, C. The human c-fos and TNFalpha AU-rich elements show different effects on mRNA abundance and protein expression depending on the reporter in the yeast Pichia pastoris. *Yeast* **27**, 1-9 (2010).

49. Vasudevan, S. & Peltz, S.W. Regulated ARE-mediated mRNA decay in Saccharomyces cerevisiae. *Mol Cell* **7**, 1191-1200 (2001).

50. Zhou, W., Edelman, G.M. & Mauro, V.P. Transcript leader regions of two Saccharomyces cerevisiae mRNAs contain internal ribosome entry sites that function in living cells. *Proc Natl Acad Sci U S A* **98**, 1531-1536 (2001).

51. Zhou, W., Edelman, G.M. & Mauro, V.P. Isolation and identification of short nucleotide sequences that affect translation initiation in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **100**, 4457-4462 (2003).

52. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946-950 (2009).

53. Sambrook, J. & Russell, D.W. Molecular Cloning: A Laboratory Manual, Edn. 3rd. (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY; 2001).

54. Sikorski, R.S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122**, 19-27 (1989).

55. Gietz, R. & Woods, R. in Guide to Yeast Genetics and Molecular and Cell Biology, Part B, Vol. 350. (eds. C. Guthrie & G. Fink) 87-96 (Academic, San Diego, CA; 2002).

56. Caponigro, G., Muhlrad, D. & Parker, R. A small segment of the MAT alpha 1 transcript promotes mRNA decay in Saccharomyces cerevisiae: a stimulatory role for rare codons. *Mol Cell Biol* **13**, 5141-5148 (1993).

57. Ng, R. & Abelson, J. Isolation and sequence of the gene for actin in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **77**, 3912-3916 (1980).

58. Asadollahi, M.A. et al. Production of plant sesquiterpenes in Saccharomyces cerevisiae: effect of ERG9 repression on sesquiterpene biosynthesis. *Biotechnol Bioeng* **99**, 666-677 (2008).

59. Arthington-Skaggs, B.A., Jradi, H., Desai, T. & Morrison, C.J. Quantitation of ergosterol content: novel method for determination of fluconazole susceptibility of Candida albicans. *J Clin Microbiol* **37**, 3332-3337 (1999).

60. Lamontagne, B. & Elela, S.A. Purification and characterization of Saccharomyces cerevisiae Rnt1p nuclease. *Methods Enzymol* **342**, 159-167 (2001).

Chapter III: Synthetic RNA modules for precise control of expression levels in yeast by tuning RNase III activity

Abstract

The design of synthetic gene networks requires an extensive genetic toolbox to precisely control the activities and levels of protein components to achieve desired cellular functions. Recently, a novel class of RNA-based control modules, which act through posttranscriptional processing of transcripts by directed RNase III (Rnt1p) cleavage, were shown to provide predictable control over gene expression and unique properties for manipulating biological networks. Here, we increase the regulatory range of the Rnt1p control elements, by modifying a critical region for enzyme binding to its hairpin substrates, the binding stability box (BSB). We used a high-throughput, cell-based selection strategy to screen a BSB library for sequences that exhibit low fluorescence and thus high Rnt1p processing efficiencies. Sixteen unique BSBs were identified that cover an intermediate range of protein expression levels (25%-75%), due to the ability of the sequences to affect the hairpin cleavage rate and to form active cleavable complexes with Rnt1p. We further demonstrated that the activity of synthetic Rnt1p hairpins can be rationally programmed by combining the synthetic BSBs with a set of sequences located within a different region of the hairpin that directly modulate cleavage rates, providing a modular assembly strategy for this class of RNA-based control elements.

3.1. Introduction

The field of synthetic biology encompasses the engineering of new cellular functions through the design of synthetic gene networks. The precise tuning of protein levels is critical for proper functioning of integrated genetic networks. For example, the optimization of metabolic networks often requires the precise tuning and regulation of enzyme levels and activities to avoid undesired consequences associated with metabolic burden due to gene overexpression[1-2], the accumulation of toxic intermediates[3-5], and the redirection of metabolic flux from pathways critical to cell growth and viability[6-8]. Altered levels of protein components can be achieved by controlling transcription[9-13], posttranscriptional stability and translation[14-17], and protein stability[18-19]. In addition, libraries of genetic control elements have been generated to increase the precision with which protein levels can be modulated[3, 10-11, 16-17]. However, the majority of gene regulatory tools developed to date function in bacterial hosts, such as *Escherichia coli*. Therefore, extending toolsets of genetic control elements to other cellular chassis is essential to supporting the design of more complex, integrated genetic networks in those organisms.

The budding yeast, *Saccharomyces cerevisiae*, is a relevant cellular chassis in industrial bioprocessing[20-25]. The current genetic toolbox for *S. cerevisiae* gene regulation relies primarily on transcriptional control mechanisms such as inducible and constitutive promoter systems. Many inducible promoters depend on accurately controlling the level of the exogenously-applied inducer molecule, where intermediate expression levels are determined through the partitioning of cells in the population between either being fully repressed or expressing the desired protein[26]. While engineered variants have been

constructed that offer more tunable responses to varying inducer concentrations[11-12], these systems can exhibit other undesirable properties, such as pleiotropic effects of the inducer molecules, undesired effects of altering the natural regulatory networks associated with the native promoter system, and the cost associated with the inducing molecule in scale-up processes. RNA-based control modules based on posttranscriptional mechanisms may offer an advantage since their activities are independent of the choice of promoter. Moreover, RNA-based controllers can be combined with transcriptional controllers to expand the design of integrated regulatory networks and thus provide more sophisticated control strategies.

We previously developed a novel class of RNA control modules that act through posttranscriptional cleavage by the *S. cerevisiae* Rnt1p enzyme (Chapter II). Rnt1p recognizes RNA hairpins that contain a consensus AGNN tetraloop, which forms a predetermined fold that is recognized by the dsRNA-binding domain (dsRBD) of Rnt1p[27-29]. RNA hairpins cleaved by Rnt1p have three critical regions: the initial binding and position box (IBPB), comprising the tetraloop; the binding stability box (BSB), comprising the base-paired region immediately adjacent to the tetraloop; and the cleavage efficiency box (CEB), comprising the region containing and surrounding the cleavage site[29]. Rnt1p hairpins were inserted as genetic control elements within the 3' untranslated region (UTR) of a transcript in order to direct cleavage to that region, thereby inactivating the transcript and lowering target protein levels. We designed an initial library based on randomization of the Rnt1p substrate CEB and screened this library to identify a set of Rnt1p control modules that tune expression levels through differential Rnt1p processing rates (Chapter II). The utility of the Rnt1p control modules was

demonstrated for achieving predictable control over protein levels and manipulating biological networks.

Here, we examined the role of a different critical region of the Rnt1p substrate, the BSB, on Rnt1p processing efficiencies and thus gene regulatory activities. We generated a library of hairpins based on randomization of the BSB to identify sequences that modulated Rnt1p binding affinity. Rigid structural constraints imposed by the BSB resulted in a low percentage of sequence variants in the library that exhibit Rnt1p binding activity, and thus required the development of a selection strategy based upon fluorescence-activated cell sorting (FACS) by enriching for cells exhibiting low fluorescence. In total, 16 unique BSBs were identified that span an intermediate range of protein expression levels. *In vitro* characterization assays indicated that altered expression levels are due to the ability of BSBs to determine the hairpin cleavage rate and to form active cleavable complexes with Rnt1p. The integration of the synthetic BSB sequences with different synthetic CEB modules demonstrated that the BSB sequences function as modules that retain their relative activities under the context of different CEBs. Further characterization indicated that proportional deviation from the 'parent' BSB was inversely related to the strength of the coupled CEB. Our work establishes a set of BSB sequences and a previously developed set of CEB sequences as modular units that can be implemented combinatorially to build synthetic Rnt1p hairpins exhibiting precisely tuned processing properties and an extended range of gene regulatory activities.

3.2. Results

3.2.1. Design and selection of an Rnt1p binding library to achieve tunable gene regulatory control

Rnt1p is an RNase III enzyme that cleaves hairpin structures in *S. cerevisiae*. An Rnt1p substrate can be divided into three critical regions: the IBPB, the BSB, and the CEB [29] (Figure 3.1A). The BSB has a reported structural requirement in which the three nucleotides immediately below the tetraloop must form Watson-Crick base-pairs and the nucleotides in the fourth position must also base-pair, in either a Watson-Crick or wobble conformation, for optimal activity[30]. Rnt1p initially binds to the tetraloop and then cleaves the hairpin at two locations within the CEB: between the 14th and 15th nts upstream of the tetraloop and the 16th and 17th nts downstream of the tetraloop. Naturally-occurring Rnt1p hairpins have been identified in numerous noncoding RNAs, where Rnt1p plays a critical role in noncoding RNA processing and editing[31-33], and in transcripts, where Rnt1p was shown to play a role in controlling gene expression[34-36].

We previously developed a genetic system in which Rnt1p-mediated cleavage was used to regulate gene expression in yeast through the placement of Rnt1p hairpins in the 3' UTR of a target transcript (Figure 3.1B) (Chapter II). We developed a set of synthetic Rnt1p hairpins based on sequence modification within the CEB that exhibit a broad range of cleavage rates and thus gene regulatory activities. Since the BSB and CEB are required elements for Rnt1p binding and cleavage, respectively, we hypothesized that a similar library screening approach could be applied to generate synthetic BSBs exhibiting different Rnt1p binding affinities. The synthetic CEB and BSB elements are anticipated to act as modular units such that they could be implemented combinatorially

to build synthetic Rnt1p hairpins exhibiting more precise tuning and an extended range of gene regulatory activities.

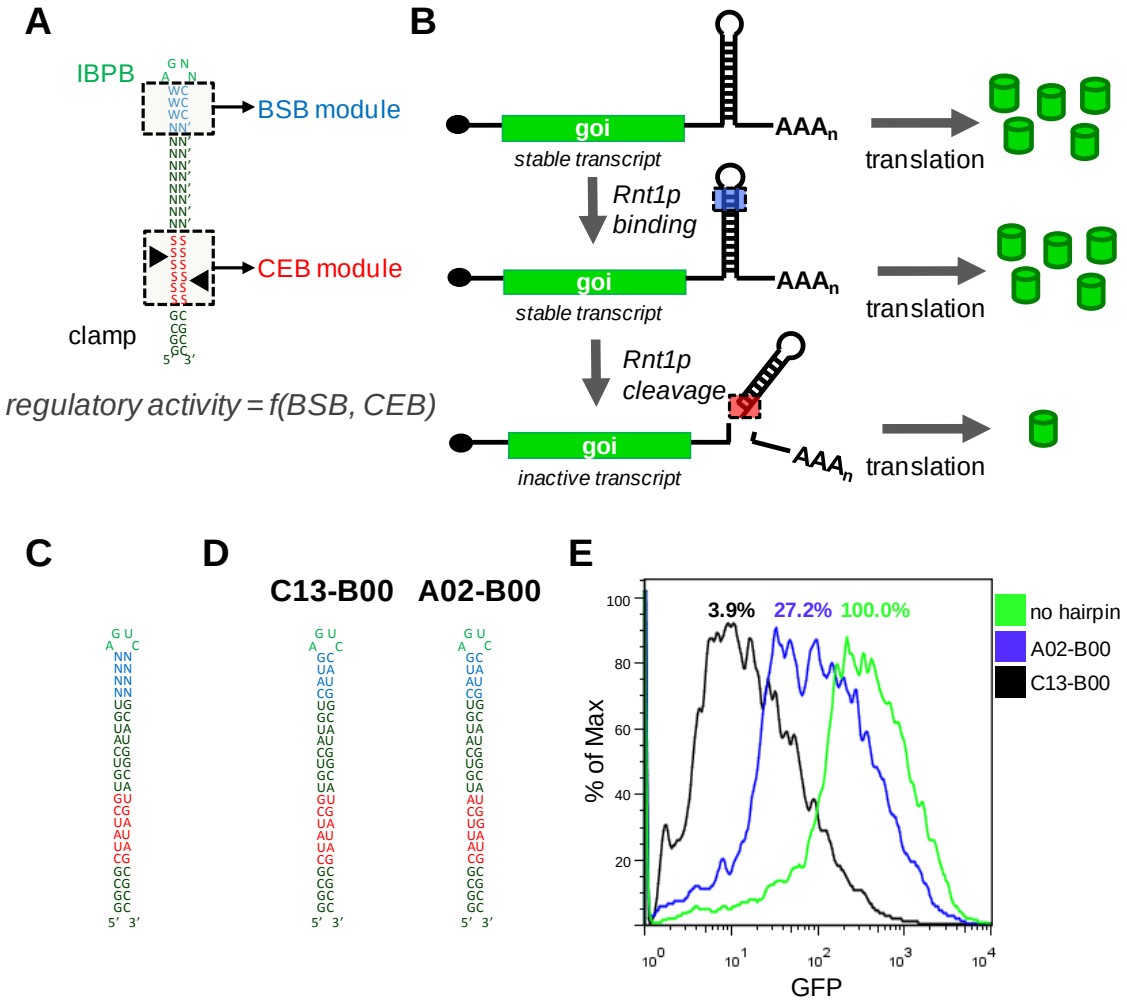


Figure 3.1. Implementation of Rnt1p hairpins as posttranscriptional genetic control elements and binding library design. (A) Consensus regions of an Rnt1p hairpin. Color scheme is as follows: cleavage efficiency box (CEB), red; binding stability box (BSB), blue; initial binding and positioning box (IBPB), green. Black triangles indicate locations of cleavage sites. The clamp region is a synthetic sequence that acts to insulate and maintain the structure of the control element. The regulatory activity of our synthetic hairpins is a function of the modular elements in the CEB and the BSB. (B) Rnt1p hairpins in the 3' UTR of a gene of interest (goi) reduce protein levels through transcript destabilization by endonucleolytic cleavage. Barrels represent protein molecules. Stable transcripts are exported out of the nucleus to the cytoplasm where translational processes occur. Unstable transcripts caused by Rnt1p cleavage will have reduced protein expression levels. (C) Sequence and structure of an Rnt1p binding library based on the randomization of 8 nts in the BSB and containing the C13 CEB. (D) Sequences and

structures of synthetic CEB sequences (Chapter II) used in the screening and characterization of the binding library. ‘B00’ refers to the ‘parent’ BSB used in the cleavage library. (E) Histograms of GFP fluorescence for cell populations containing the Rnt1p constructs based on the ‘parent’ BSB. Numbers above each peak represent the median GFP fluorescence for that sample normalized to a control sample with a construct containing no Rnt1p hairpin.

We designed an Rnt1p binding library based on randomizing the BSB (8 nt) to generate Rnt1p hairpins that exhibit different gene regulatory activities due to altered binding affinity between the hairpin and Rnt1p (Figure 3.1C). One of the variables in the design of the binding library was the CEB to place within the stem, as we had previously described a set of synthetic CEBs with modified gene regulatory activity (Chapter II). As it was unknown how the Rnt1p hairpins would respond to changes in the BSB, we selected a synthetic CEB contained in the hairpin that demonstrated the lowest level of gene expression (C13). All hairpins with active BSBs containing the CEB of C13 were expected to have comparable gene expression levels, improving their probability of being identified in the screen. C13 is also fully base-paired, such that its integration into an Rnt1p substrate stem results in a stable structure that is less susceptible to changes in flanking sequences and has a greater probability of maintaining the desired hairpin structure (Figure 3.1D). Flow cytometry analysis also indicates that C13 achieves the greatest population separation from a no hairpin control of all synthetic CEBs (Figure 3.1E). This separation is representative of that expected for the active and inactive binding populations in the binding library, thus, increasing the enrichment of the cell-based sort. We refer to the ‘parent’ BSB that was used in the cleavage library as B00, where synthetic Rnt1p hairpins are identified by their CEB and BSB as Cxx–Bxx (i.e., C13–B00) or Axx–Bxx (i.e., A02–B00).

Synthetic Rnt1p substrates with altered binding affinities were identified through a cell-based fluorescence screen. The designed library (N8) has a diversity of 65,536 different sequences. Due to the rigid structural requirements of the BSB, we predicted that at most 1,024 sequences (1.56%) would be actively bound and subsequently cleaved by Rnt1p. A manual plate-based screening strategy was previously used to efficiently identify active sequences from the cleavage library. However, due to the loose structural requirements for the CEB, the cleavage library had a substantially higher positive rate than that anticipated for the binding library, such that the plate-based screen was not feasible for the number of colonies that would have to be screened to identify a reasonable diversity of active BSBs. As a higher throughput method we employed FACS to efficiently identify cells with diminished fluorescence.

The binding library was transformed into yeast through a library-scale gap-repair strategy and clones exhibiting strong gene regulatory activity were selected through FACS by gating for cells with low fluorescence levels (Figure 3.2A). We performed two different single FACS screens based on a single-color (pCS1585; *yEGFP3*) system and a two-color (pCS1748; *ymCherry* and *yEGFP3*) system (Liang, J.C. et al., in preparation). Three fractions (A, B, and C) were collected for the pCS1585 system around the expression level of C13-B00. Following the initial sort, each fraction was regrown and subsequently characterized by flow cytometry (Figure 3.2B, Supplementary Figure 3.1). Only fraction A retained a low level of fluorescence, whereas the regrown populations in fractions B and C shifted substantially towards inactive (high fluorescence) levels. The results suggest that sorted fractions B and C contain a large percentage of false positives, or clones harboring plasmids containing inactive Rnt1p hairpins that exhibited low

fluorescence levels due to noise in gene expression profiles. Based on the fraction profiles, fraction A was selected for further testing. Most of the false positives due to genetic noise were removed with the two-color pCS1748 system. A gate representing diminished GFP fluorescence was determined based on a GFP positive control lacking an Rnt1p hairpin. All library clones with decreased GFP levels were collected into a single fraction (D). Fraction D was regrown, characterized by flow cytometry, and demonstrated to retain low fluorescence levels (Figure 3.2C, Supplementary Figure 3.2). Therefore, fraction D was also selected for further characterization.

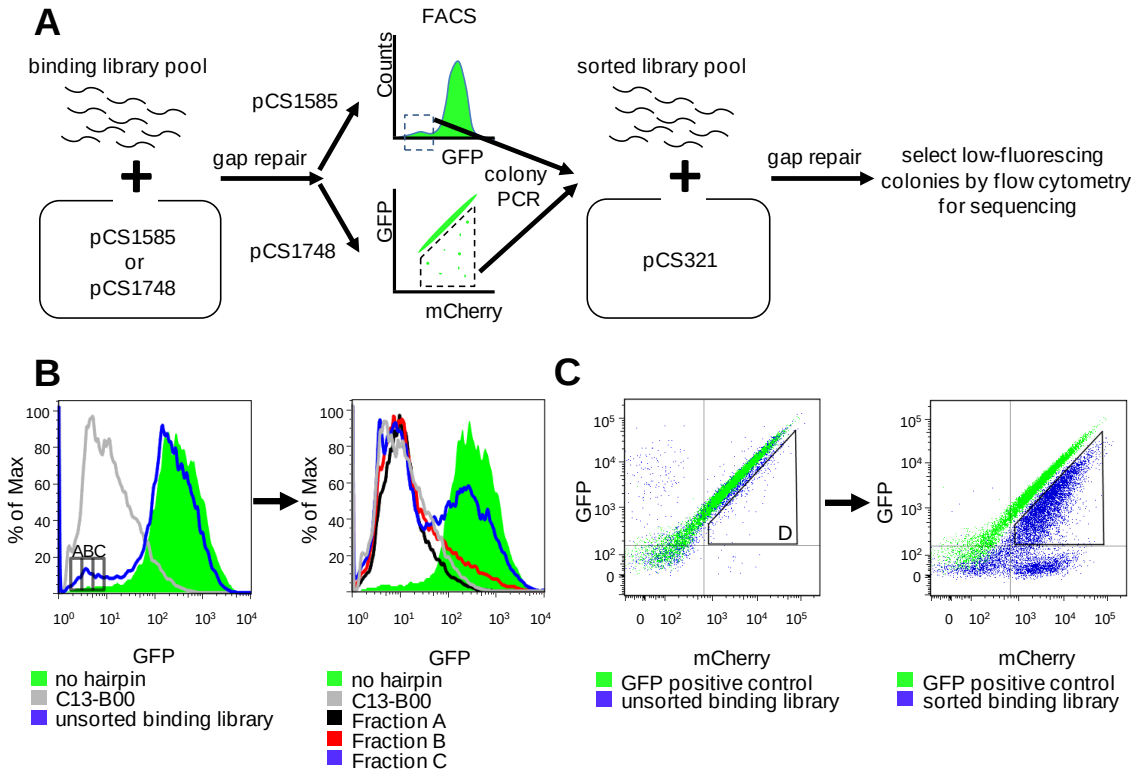


Figure 3.2. *In vivo* screening of an Rnt1p binding library. (A) A high-throughput, *in vivo*, fluorescence-based screen for Rnt1p hairpin activity. The library was cloned through gap-repair into yeast in two different plasmid systems. Clones exhibiting low GFP fluorescence were sorted from the population through FACS. A sorted library pool was generated through colony PCR from collected cellular fractions and gap-repaired into the characterization plasmid. Clones that maintained low GFP fluorescence levels were selected for sequencing and further characterization. (B) FACS procedure for the single-

color (pCS1585-based) system. Fractions A, B, and C were collected based on exhibiting GFP levels similar to the median fluorescence of C13-B00 (left panel). Only fraction A maintained a low level of GFP expression after the fractions were regrown (right panel). (C) FACS procedure for the two-color (pCS1748-based) system. Fraction D was collected based on a gate set to collect all cells exhibiting GFP fluorescence levels below a cells containing a positive GFP control construct lacking an Rnt1p hairpin module (left panel). Fraction D maintained a low level of GFP expression when regrown (right panel).

After the completion of the FACS screens, the sorted constructs were recloned to remove false positives due to mutations in the plasmid or the yeast background that would cause reduced GFP levels independent of Rnt1p activity. We retrieved the selected Rnt1p hairpin sequences from fractions A and D by colony PCR and gap-repaired the recovered hairpin constructs into pCS321. Individual clones were initially characterized for gene regulatory activity by measuring cellular fluorescence through a plate reader assay. Colonies positive for GFP knockdown were sequenced to determine the BSB sequence. In total 16 unique BSB sequences were identified including the 'parent' BSB (Supplementary Figure 3.3; Supplementary Table 3.1). The predicted secondary structure of the hairpins was determined by RNAstructure (http://rna.chem.rochester.edu/RNAstructure.html). The binding library structures deviate from structural requirements that were previously established through *in vitro* studies[30]. All of the BSB structures contain Watson-Crick base-pairing in the first three nucleotides below the tetraloop; however, in the fourth position certain sequences exhibit mismatching. The results from our cell-based BSB library screen suggest that the *in vivo* structural requirements for the BSB are not as stringent as previously described.

3.2.2. A synthetic Rnt1p binding library exhibits a range of gene regulatory activities **in vivo**

The initial screen for active BSB sequences was performed in the context of a CEB exhibiting high cleavage activity (C13). As a result, most of the recovered binding library hairpins exhibited low gene expression levels, such that differences in activity between the synthetic BSBs were difficult to resolve with the flow cytometry assay at low fluorescence levels (Supplementary Table 3.1). To gain better resolution on the differences in BSB activity and examine the activity of the synthetic BSBs in the context of a different CEB, we integrated the selected BSB sequences within the context of an Rnt1p hairpin containing a different synthetic CEB (A02) (Figure 3.1D). The range of regulatory activities spanned by the binding library in the context of the A02 CEB was measured at the protein and transcript levels. Flow cytometry analysis of the synthetic Rnt1p hairpins indicated that the selected set of hairpins spanned an intermediate gene regulatory range – from 25% (A02–B05) to 75% (A02–B01) (Table 3.1, Figure 3.3A). The regulatory activities of the selected hairpins are not evenly distributed across this range, with the majority exhibiting activities in the range of 25–45% relative protein levels. The results suggest that the binding library achieves a smaller regulatory range than that observed with the cleavage library and may be more appropriate for fine-tuning (Chapter II). We built negative controls for several binding library hairpins and for the ‘parent’ hairpins (A02–B00 and C13–B00) by mutating the tetraloop sequence (CAUC or GAAA) to impede Rnt1p activity while maintaining the secondary structure of the hairpins. The negative controls demonstrated that the majority of knockdown observed from each hairpin is due to Rnt1p processing (Figure 3.3B).

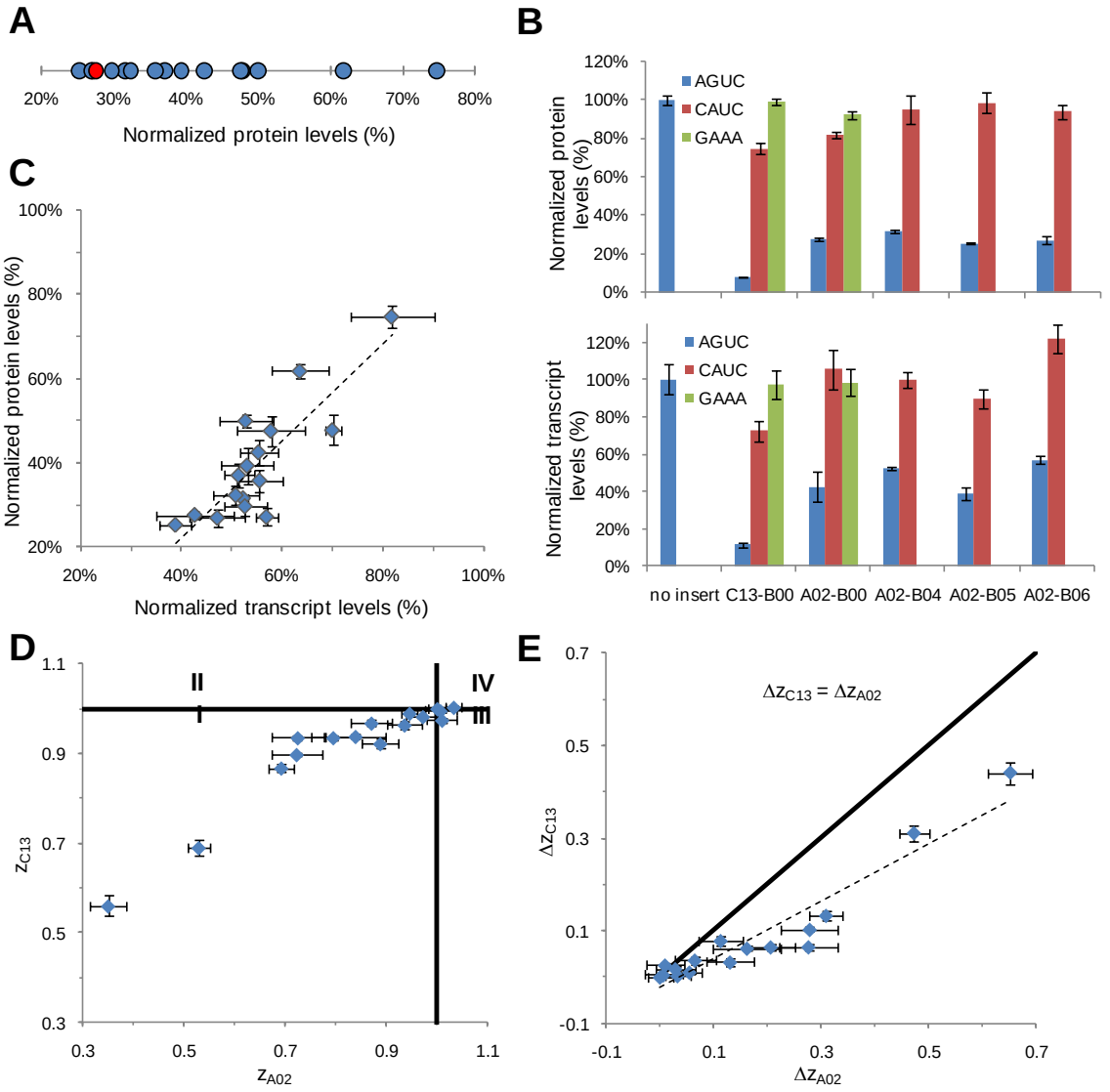


Figure 3.3. *In vivo* characterization of the selected Rnt1p binding library and demonstration of the modularity of the BSB sequences. (A) The gene regulatory activities of the binding library spans an intermediate range of protein expression levels. Normalized protein expression levels were determined by measuring the median GFP levels from a cell population containing the appropriate construct through flow cytometry analysis and values are reported relative to that from an identical construct lacking a hairpin module. The 'parent' BSB is indicated in red. (B) The transcript and protein levels associated with several binding library members and their corresponding mutated tetraloop (CAUC or GAAA) controls support that the observed gene regulatory activity is due to Rnt1p processing. Transcript levels were determined by measuring transcript levels of *yEGFP3* and a house-keeping gene, *ACT1*, through qRT-PCR and normalizing the *yEGFP3* levels with their corresponding *ACT1* levels. Normalized transcript levels for each construct are reported relative to that from an identical construct lacking a hairpin module ('no insert'). (C) Correlation analysis of protein and transcript levels from

the binding library members demonstrates a strong correlation between the two measures of gene regulatory activity. (D) The gene regulatory activity of synthetic BSBs is conserved in the context of different CEB modules. The ratio of the knockdown exhibited from a binding library hairpin to that exhibited from the 'parent' (B00 BSB), z_C, was determined in the context of two different CEBs and plotted against each other. Regions I and IV represent hairpins whose activities relative to 'parent' remain consistent when combined with different CEBs. Regions II and III represent hairpins whose activity varies relative to 'parent' when combined with different CEBs. (E) Synthetic BSBs modules generally exhibit higher relative activities in the context of weaker CEB modules. A variable representing the departure from 'parent' activity, Δz_C, was calculated in the context of two different CEBs and plotted against each other. The solid line indicates where the values of Δz_C are the same for both hairpins.

Table 3.1. *In vivo* characterization data for the binding library. All normalized protein and transcript levels were determined as described in Figure 3.3A and Figure 3.3B, respectively.

Substrate	Normalized protein levels (%)	Normalized transcript levels (%)
A02-B00	28% ± 1%	43% ± 8%
A02-B00 (CAUC)	81% ± 2%	106% ± 11%
A02-B01	75% ± 3%	82% ± 8%
A02-B02	62% ± 2%	64% ± 6%
A02-B03	50% ± 2%	53% ± 5%
A02-B04	32% ± 1%	52% ± 1%
A02-B05	25% ± 0%	39% ± 3%
A02-B06	27% ± 2%	57% ± 2%
A02-B07	37% ± 3%	51% ± 3%
A02-B08	30% ± 2%	53% ± 4%
A02-B09	36% ± 3%	56% ± 5%
A02-B10	42% ± 3%	55% ± 4%
A02-B11	32% ± 2%	51% ± 4%
A02-B12	27% ± 2%	47% ± 5%
A02-B13	39% ± 4%	53% ± 5%
A02-B14	48% ± 4%	70% ± 2%
A02-B15	48% ± 4%	58% ± 7%
no insert	100% ± 3%	100% ± 8%

The reduced protein expression levels observed from the Rnt1p binding library is expected to be due to a reduction in the steady-state transcript levels due to rapid degradation of the transcript following endonucleolytic cleavage by Rnt1p. We measured relative transcript levels for each Rnt1p hairpin by quantitative real-time PCR (qRT-PCR) (Table 3.1). A plot of normalized *yEGFP3* expression levels versus normalized *yEGFP3* transcript levels indicates that there is a strong positive correlation (r = 0.847) between the two measures of activity (Figure 3.3C). A preservation of rank order was also observed between protein and transcript levels as indicated by the Spearman rank correlation coefficient (ρ = 0.668). Specifically, with decreasing transcript levels a similar decrease in protein levels is generally observed, confirming that the fluorescence observed was due to changes in the steady-state transcript levels. The negative controls based on mutating the tetraloop confirmed that Rnt1p cleavage is the cause of the observed transcript knockdown (Figure 3.3B).

3.2.3. Synthetic BSBs exhibit modular activity with different CEBs in vivo

We next examined the gene expression data for the Rnt1p hairpins harboring the synthetic BSBs in the context of two CEBs (C13, A02) for trends in regulatory activity across the binding library. We defined a new variable z_C as the ratio of the knockdown from a binding library member (Bxx) to that of the 'parent' BSB (B00) for a specific CEB (Cxx or Axx):

$$z_C(Bxx) = \frac{knockdown_{C-Bxx}}{knockdown_{C-B00}}$$

A *z* value greater than unity indicates increased knockdown due to the synthetic BSB, whereas a *z* value less than unity indicates decreased knockdown due to the BSB. We

calculated *z* values for each BSB in the context of the CEBs C13 and A02 (z_{C13} and z_{A02}, respectively) and plotted the two variables against each other (Figure 3.3D). For ease in interpretation, we divided the graph into four regions with the point (1,1) at the intersection of the quadrants. Regions I and IV indicate BSBs for which activities relative to parent are conserved between the different CEBs, whereas regions II and III indicate BSBs that exhibit varying activities in the context of different CEBs. Nearly all BSBs are located in regions I and IV, with the majority falling in region I. The data indicate that if a BSB causes increased knockdown in the context of one CEB, it will likely exhibit the same activity in the context of another CEB.

To further examine the gene regulatory activities of the synthetic BSBs, we determined a new variable Δz_C, which is the difference between the z_C value of the 'parent' (which by definition is 1) and the *z* value of the BSB:

$$\Delta z_C(Bxx) = z_C(B00) - z_C(Bxx) = 1 - z_C(Bxx)$$

We calculated Δz for each BSB in the context of the CEBs C13 and A02 and plotted the variables against each other (Figure 3.3E). Data points that fall on the $\Delta z_{C13} = \Delta z_{A02}$ line would indicate BSBs that have the same proportional effect on knockdown for both CEBs. The data fall beneath the $\Delta z_{C13} = \Delta z_{A02}$ line in the region where Δz_{A02} is greater than Δz_{C13}, indicating that Rnt1p hairpins with weaker CEBs are affected more by changes in binding affinity through modification to the BSBs. The data exhibit a strong positive correlation ($r = 0.946$) and can be fit with a trendline by linear regression that passes close to the origin (0,0), suggesting that we see a consistent ratio between Δz_{C13} and Δz_{A02} values, where this ratio is dependent on the CEBs. The data also exhibit a strong preservation of rank order ($\rho = 0.929$), demonstrating that the relative activity

between BSB modules is maintained regardless of the CEB module present in the hairpin. Taken together, the results show the maintenance of BSB activity in connection with different CEB stems and also that the proportional deviation from the 'parent' BSB is determined by the strength of the CEB.

3.2.4. **In vitro** ***characterization demonstrates that Rnt1p binding library members achieve differential activity through alterations in Rnt1p cleavage rates and affinity***

We hypothesized that the variation in transcript processing and subsequent protein expression levels exhibited by the binding library is due to variations in binding affinity resulting from alterations in the BSB sequence and/or structure. To examine whether the synthetic Rnt1p binding library members exhibit differences in binding affinity to Rnt1p, we performed *in vitro* binding assays with purified Rnt1p. Binding reactions were ran with 20 nM of *in vitro* synthesized radiolabeled RNA encoding an Rnt1p hairpin and varying concentrations of purified Rnt1p. The reactions were ran in the absence of magnesium and other divalent metal ions that are essential for cleavage to allow Rnt1p to bind to the substrates without subsequent cleavage[37]. Bound products were separated by nondenaturing polyacrylamide gel electrophoresis and quantified through phosphorimaging analysis (Figure 3.4A). We analyzed the reaction through a modified Scatchard equation in which the fraction of unbound RNA (R) to total RNA (R_0) is plotted against the enzyme (E) concentration. The equation is as follows:

$$Z = \frac{R}{R_o} = \frac{K_D}{K_D + [E]}$$

The dissociation constant, K_D, for each synthetic Rnt1p hairpin was determined through this analysis method (Table 3.2). The data indicate that there is a moderate positive

correlation (r = 0.486) between K_D and *in vivo* gene regulatory activity (Figure 3.4B). While we observe several data points demonstrating similar transcript levels for different K_D values, hairpins that bind less tightly to Rnt1p (i.e., higher K_D) generally tend to have higher transcript levels as anticipated. The binding library has an expanded range of K_D values compared to the cleavage library, due to several library members having K_D values greater than those previously reported with the cleavage library (Babiskin, A.B. and Smolke, C.D., in submission). *In vivo*, we observe that most binding library members have increased gene expression levels greater than 'parent'. *In vitro*, this same phenomenon is experienced as most binding library members have decreased affinity for Rnt1p. Binding library members that have decreased gene expression levels than 'parent' *in vivo* also have K_D values comparable or less than 'parent' *in vitro*. However, we also observed that the mutant tetraloop control bound Rnt1p with a similar K_D as the library hairpins, although cleavage was not evident (Figure 3.3B). It has been previously reported that Rnt1p is able to bind its substrates in inactive and active conformations *in vitro*[38]. Therefore, it is plausible that the binding observed in these *in vitro* assays is due to both types of complexes with Rnt1p. Under this situation, the reported K_D may not be solely related to complexes that can be processed. It has also been shown that changes in the BSB affect both binding affinity and hairpin processing by Rnt1p *in vitro*[29]. Therefore, it is important to examine the effects of the binding library on Rnt1p processing rates.

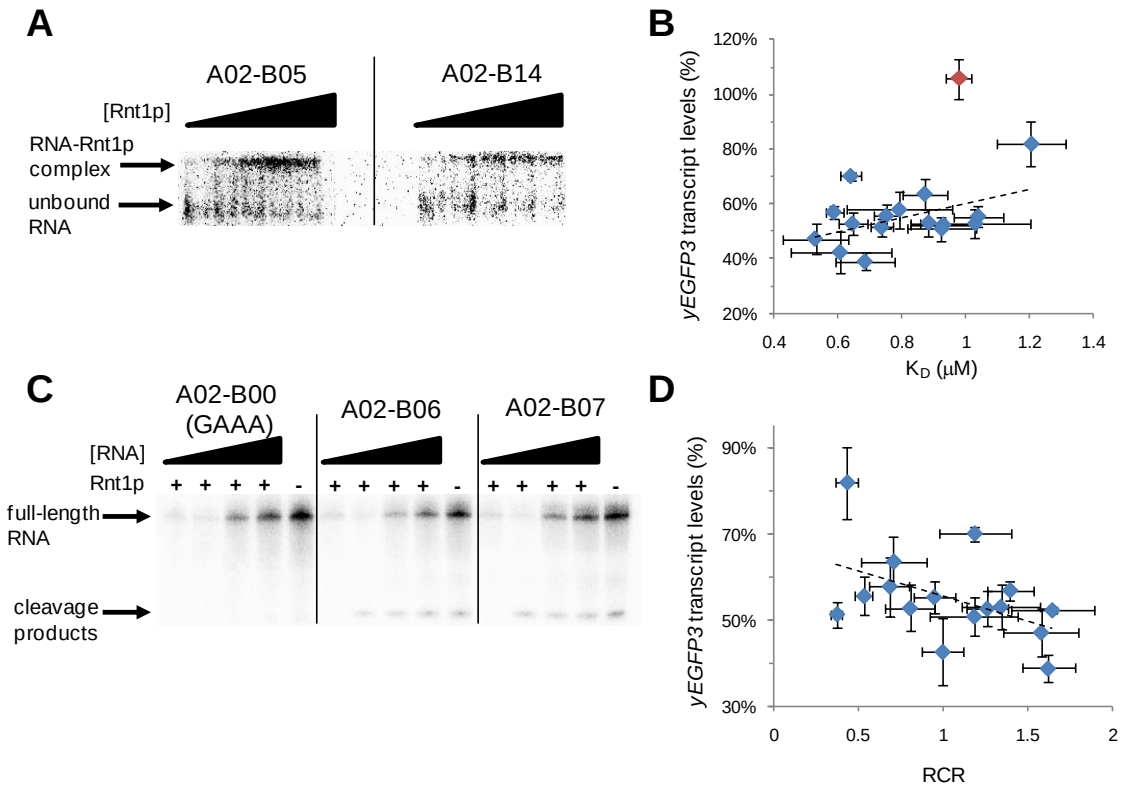


Figure 3.4. *In vitro* characterization of the binding library demonstrates that the observed tuning of gene regulatory activity is achieved through modulation of cleavage rates and binding affinities. (A) Representative mobility shift assays and analyses by nondenaturing polyacrylamide gel electrophoresis of two binding library members: A02-B05 and A02-B14. The top band corresponds to RNA-Rnt1p complexes; the bottom band corresponds to unbound RNA. Rnt1p was added to the following final concentrations in each reaction (left to right; in μM): 0, 0.42, 0.83, 1.25, 1.66. (B) Correlation analysis of binding affinity (K_D) and *yEGFP3* transcript levels indicates a moderate positive correlation between binding affinity and gene regulatory activity. The data point for the A02-B00 (GAAA) negative control is indicated in red. (C) Representative cleavage reaction assays and analyses by denaturing polyacrylamide gel electrophoresis on hairpins A02-B00 (GAAA), A02-B06, and A02-B07. The top band corresponds to unreacted full-length RNA; the bottom band corresponds to the three cleavage products expected from Rnt1p processing. The three cleavage products differ in size by 1 nt and cannot be resolved into individual bands under the conditions used for this assay. RNA was added to the following final concentrations in each reaction (left to right; in μM): 0.2, 0.5, 1.0, 2.0. Reactions lacking Rnt1p were performed with 0.2 μM of RNA. (D) Correlation analysis of relative cleavage rate (RCR) and *yEGFP3* transcript levels demonstrates a moderate positive correlation between cleavage rate and gene regulatory activity.

Table 3.2. *In vitro* characterization data for the binding library.

Substrate	RCR	K_D (μM)
A02-B00	1.00 ± 0.12	0.61 ± 0.16
A02-B00 (GAAA)	0*	0.98 ± 0.04
A02-B01	0.44 ± 0.07	1.21 ± 0.11
A02-B02	0.71 ± 0.20	0.87 ± 0.07
A02-B03	0.81 ± 0.15	1.03 ± 0.17
A02-B04	1.65 ± 0.25	0.93 ± 0.10
A02-B05	1.62 ± 0.15	0.69 ± 0.09
A02-B06	1.40 ± 0.14	0.59 ± 0.03
A02-B07	0.37 ± 0.03	0.74 ± 0.04
A02-B08	1.26 ± 0.12	0.65 ± 0.05
A02-B09	0.53 ± 0.05	0.75 ± 0.04
A02-B10	0.95 ± 0.12	1.04 ± 0.08
A02-B11	1.19 ± 0.26	0.93 ± 0.11
A02-B12	1.58 ± 0.22	0.53 ± 0.10
A02-B13	1.34 ± 0.23	0.88 ± 0.06
A02-B14	1.19 ± 0.21	0.64 ± 0.03
A02-B15	0.69 ± 0.12	0.79 ± 0.16

*Immeasurable due to lack of product formation

We analyzed the cleavage reaction between Rnt1p and the binding library through a Michaelis-Menten model, with the substrate (S) being the hairpin transcript, the enzyme (E) being Rnt1p, and the product (P) being the cleaved pieces of the transcript. Under these conditions, the following reaction occurs:

$$E + S \underset{k_{-1}}{\overset{k_1}{\leftrightharpoons}} ES \xrightarrow{k_2} E + P$$

The rate of product formation (V) is modeled as:

$$V = \frac{V_{max} * [S]}{K_M + [S]} = \frac{k_2 * [E]_0 * [S]}{K_M + [S]}$$

The maximum rate of product formation (V_{max}) is the product of the total enzyme concentration ($[E]_0$) and k_2. Alterations in the cleavage efficiency will have an effect on the value of k_2 and thus V_{max}. We performed *in vitro* RNA cleavage reactions with a constant concentration of purified Rnt1p against a range of *in vitro* synthesized radiolabeled Rnt1p hairpins to determine the relative values of k_2 for each synthetic Rnt1p hairpin. Reaction products were separated by denaturing polyacrylamide gel electrophoresis and quantified through phosphorimaging analysis (Figure 3.4C). The resulting data were fit to the Michaelis-Menten model to calculate a relative cleavage rate (RCR), which is directly proportional to V_{max}. The RCR value for A02–B00 is set to 1 and the rest of the reported values are normalized to A02–B00.

The RCR values for each synthetic Rnt1p hairpin were determined through this analysis method (Table 3.2). The results confirm that the mutant tetraloop control is not processed by Rnt1p *in vitro* (supporting *in vivo* observations). There is a moderate positive correlation ($r = 0.480$) between the measured RCR and gene regulatory activity for the synthetic Rnt1p hairpins (Figure 3.4D). Generally, increases in Rnt1p's ability to cleave a substrate result in greater transcript knockdown. Compared to the cleavage library (Babiskin, A.B. and Smolke, C.D., in submission), we observe a smaller range of RCR values with the binding library, due to the observed decreased range in transcript knockdown. In fact, we observe that cleavage library members exhibiting gene regulatory activities within the range exhibited by the binding library members have similar RCR values. However, a correlation analysis between the K_D and RCR values indicates that there is no correlation between the binding affinity and relative cleavage rates for the binding library members (data not shown). Changing the BSB does result in changes in

binding affinity, but it can also result in variation in the cleavage rate. For any given hairpin, the two properties contribute to the observed transcript levels. In particular, the BSB may affect the hairpin's ability to form active or inactive complexes or the BSB may affect the hairpin's processing rate.

3.3. Discussion

We utilized a cell-based library screening approach to develop a set of synthetic BSB sequences to modulate the gene regulatory activity of engineered Rnt1p hairpins. Previous *in vitro* studies showed that the BSB contains nucleotides critical to Rnt1p binding as mutations in the region resulted in reduced affinity[29]. These studies established a consensus BSB structure, where three Watson-Crick base-pairs were required immediately below the tetraloop followed by another base-pair that could include the wobble guanine-uracil pair. Based on these reported structural requirements, we estimated that a small percentage (~1.6%) of the randomized BSB library would contain hairpins cleavable by Rnt1p. Therefore, a high-throughput FACS-based screen was employed to enrich the library for hairpins resulting in reduced GFP fluorescence. The BSB library was designed in the context of a synthetic CEB from the cleavage library that produced the greatest amount of knockdown (C13) (Chapter II). This design biased the library such that any positive hits would exhibit the lowest possible expression levels to enhance the separation, and thus the selection, of the population of cells containing active hairpins versus the larger population containing inactive hairpins (~98.4% of the population). In total, 16 unique BSBs were identified. However, in contrast to previous *in*

vitro work[29], several of the synthetic BSBs did not contain a base-pair in the fourth position from the tetraloop, suggesting that this structural requirement is relaxed *in vivo*.

The selected BSB sequences were further characterized in the context of a CEB that exhibited weaker gene regulatory activity (A02) (Chapter II) to better resolve differences in the BSB activities. The regulatory activities of the BSB sequences are distributed across an intermediate range of 25% to 75% relative protein levels with the majority exhibiting activities in the range of 25-45%. The binding library exhibited a decreased range of activity relative to the cleavage library (Chapter II), suggesting that the binding library is more appropriate for the tuning of gene expression. For example, the cleavage library can be employed initially to first identify regulatory ranges of interest. More focused regulatory activities can then be explored through implementation of the synthetic BSBs with the appropriate synthetic CEBs. As such, the combinatorial application of the synthetic BSB and CEB elements can be used to extend and tune the regulatory range accessible through the engineered Rnt1p hairpins. To demonstrate the ability to predictably combine BSB and CEB sequences, we examined the synthetic BSB sequences in the context of two synthetic CEBs generated in the cleavage library. Our experimental results show that the synthetic CEB and BSB elements act as modular units, that the BSBs maintain their activity under the context of different CEBs, and that the proportional deviation from the 'parent' BSB is determined by the strength of the CEB.

In vitro characterization studies determined the relationship between binding affinity, cleavage rate, and gene-regulatory activity for the binding library. Changes in the BSB sequence are expected to result in changes in Rnt1p cleavage rate as well as to affect Rnt1p binding[29]. We observed a moderate correlation between the binding affinity

and transcript levels and between the cleavage rate and transcript levels for the binding library. In contrast, although slight changes in affinity were detected with the cleavage library, these changes were not correlated with gene regulatory activity (Chapter II). However, there was a stronger correlation observed between cleavage rate and transcript levels for the cleavage library. The data indicate that nucleotide modifications in the BSB cause changes in affinity and cleavage rate; however, the exact contribution of each of these variables to the observed gene regulatory activity is unclear. In addition, Rnt1p is known to bind in active and inactive complexes where the inactive complex is magnesium-independent[38-39]. The inactive complex may be more prevalent in the binding assay due to the absence of magnesium in the reaction buffer, which is also critical to the proper folding of RNA molecules[40]. The changes in K_D are reflective of how much total RNA is bound regardless of conformation. Thus, it is possible that the changes to the BSB are affecting the partitioning between inactive and active states.

This work extends the regulatory capacity of the first set of posttranscriptional control elements in yeast by developing a set of BSB modules that can be integrated with a previously described set of CEB modules to rationally design synthetic Rnt1p substrates. With 16 CEB and 16 BSB modules developed, 256 different Rnt1p hairpins can be generated. The two classes of modules can be combined to predictably set and precisely tune levels of gene expression in *S. cerevisiae*. The engineered Rnt1p control modules can be used in combination with engineered or native promoter systems, as well as other posttranscriptional elements, to provide a powerful tool for programming genetic regulatory networks in yeast, and thus advancing the application of this cellular chassis in biomanufacturing and biosynthesis processes.

3.4. Materials and Methods

3.4.1. Plasmid construction

Standard molecular biology techniques were utilized to construct all plasmids[41]. DNA synthesis was performed by Integrated DNA Technologies (Coralville, IA) or the Protein and Nucleic Acid Facility (Stanford, CA). All enzymes, including restriction enzymes and ligases, were obtained through New England Biolabs (Ipswich, MA) unless otherwise noted. Pfu polymerases were obtained through Stratagene. Ligation products were electroporated with a GenePulser XCell (Bio-Rad, Hercules, CA) into *Escherichia coli* DH10B (Invitrogen, Carlsbad, CA), where cells harboring cloned plasmids were maintained in Luria-Bertani media containing 50 mg/ml ampicillin (EMD Chemicals). Clones were initially verified through colony PCR and restriction mapping. All cloned constructs were sequence verified by Elim Biopharmaceuticals (Hayward, CA) or the Protein and Nucleic Acid Facility (Stanford, CA). Plasmid maps are available in Supplementary Figure 3.4.

The construction of the Rnt1p characterization plasmid, pCS321, and the Rnt1p expression plasmid, pRNT1, have been previously described (A Babiskin and C Smolke, in submission). A screening plasmid (pCS1585) was constructed from pCS321 by replacing the GAL1-10 promoter with the endogenous TEF1 promoter (JC Liang, unpublished data, 2008). A second screening plasmid (pCS1748) was constructed from pCS1585 by inserting an additional open reading frame (ORF) containing the yeast enhanced mCherry gene, *ymCherry*, flanked by a TEF1 promoter and a CYC1 terminator (JC Liang et al., in preparation).

Insertion of engineered Rnt1p substrates and appropriate controls into the 3’ UTR of *yEGFP3* in pCS321 and pCS1585 was performed through either digestion with appropriate restriction endonucleases and ligation-mediated cloning or homologous recombination-mediated gap-repair during transformation into *S. cerevisiae* strain W303 (*MATa, his3-11,15 trp1-1 leu2-3 ura3-1 ade2-1*) through standard lithium acetate procedures[42]. The Rnt1p substrates were amplified for insertion with both techniques using the forward and reverse primers RntGap321_fwd (5’ ACCCATGGTATGGATGA ATTGTACAAATAAAGCCTAGGTCTAGAGGCG) and RntGap321_rev2 (5’ TAAGA AATTCGCTTATTTAGAAGTGGCGCGCCCTCTCGAGGGCG), respectively. In the case of digestion and ligation, the PCR products were digested with the unique restriction sites AvrII and XhoI, which are located 3 nts downstream of the *yEGFP3* stop codon and upstream of the ADH1 terminator. Following construction and sequence verification of the desired vectors, 100–500 ng of each plasmid was transformed into strain W303. In the case of gap-repair, 250–500 ng of the PCR product and 100 ng of plasmid digested with AvrII and XhoI were transformed into the yeast strain. All yeast strains harboring cloned plasmids were maintained on synthetic complete media with an uracil dropout solution and 2% dextrose at 30°C.

3.4.2. Library-scale yeast transformation

Yeast transformations with the binding library were performed on Rnt1p substrates as previously described[43]. The C13-based binding library was amplified with template BndLib_C13 (5’ AGCCTAGGTCTAGAGGCGCTATCGTGTCATGTNNNNA GTCNNNNGCATGGCATGATAGCGCCCTCGAGAGGG) and forward and reverse

primers C13BindLibgap_fwd_prmr (5’ GTATTACCCATGGTATGGATGAATTGTAC AAATAAAGCCTAGGTCTAGAGGCGCTATC) and C13BindLibgap_rev_prmr (5’ AATCATAAGAAATTCGCTTATTTAGAAGTGGCGCGCCCTCTCGAGGGCGCTA TCA), respectively. The reaction was scaled-up to 800 μl to obtain roughly 40–50 μg of PCR product. 8 μg of plasmid (either pCS1585 or pCS1748) was digested overnight with AvrII and XhoI in 400 μl total reaction volume. Two tubes of DNA were made with 375 μl of PCR product (~20 μg) and 150 μl of digested plasmid (~3 μg). A third tube acting as a negative control contained 450 μl of water and 75 μl of digested plasmid (~1.5 μg). Each tube was extracted with phenol-chloroform (1:1) and ethanol-precipitated into fresh tubes.

500 μl of Tris-DTT [2.5 M DTT, 1 M Tris (pH 8.0)] was added to a 50-ml culture of yeast strain W303 that was grown in YPD to an OD_{600} of 1.3-1.5 at 30°C. Following 10–15 minutes of additional incubation, the cells were collected and washed in 25 ml of ice-cold Buffer E [10 mM Tris (pH 7.5), 2 mM $MgCl_2$, 270 mM sucrose] and washed again in 1 ml of Buffer E before being resuspended to a final volume of 300 μl in Buffer E. 60 μl of this cell mixture was added to the negative control tube and 120 μl was added to the two tubes containing digested plasmid and the library. After allowing the precipitated DNA to resuspend, 50 μl of the negative control or the library suspension was transferred to a chilled 2-mm gap cuvette and electroporated (540 V, 25 μF, infinite resistance, 2 mm gap). Each 120-μl tube of library suspension contained enough material for two electroporations. Following electroporation, the cells were resuspended in 1 ml of prewarmed YPD and added to a fresh 15-ml Falcon tube. The cuvette was washed a second time with a fresh 1-ml aliquot of YPD, which was added to the same Falcon tube.

Library electroporations were collected in the same Falcon tube (8 ml total). The Falcon tubes were incubated with shaking for 1 hour at 30°C. After incubation, the cells were collected and resuspended in 1 ml of synthetic complete media with an uracil dropout solution and 2% dextrose. The resuspension was added to 6 ml of fresh media to prepare for FACS.

3.4.3. FACS and sorted library retransformation

The transformed binding library was grown for 2–3 days in liquid culture. Following this growth period, the library and appropriate control cultures were collected and suspended in 1x PBS with 1% BSA and either 7-amino-actinomycin D (7-AAD; Invitrogen) or 4’,6-diamidino-2-phenylindole (DAPI; Invitrogen) was added as a viability stain. The cell suspension was passed through a 40-μm Cell Strainer (BD Falcon) prior to analysis on a FACSAria or FACSAria II flow cytometry cell sorter (Becton Dickinson Immunocytometry Systems, San Jose, CA). On the FACSAria, GFP was excited at 488 nm and measured with a bandpass filter of 530/30 nm. 7-AAD was excited at 488 nm and measured with a bandpass filter of 695/40 nm. On the FACS Aria II, GFP was excited at 488 nm and measured with a splitter of 505 nm and bandpass filter of 525/50 nm. mCherry was excited at 532 nm and measured with a splitter of 600 nm and a bandpass filter of 610/20 nm. DAPI was excited at 355 nm and measured with a bandpass filter of 450/50 nm. Detailed sorting procedures are presented in Supplementary Figures 3.1 and 3.2. The collected fractions were diluted to 100 ml in synthetic complete media with an uracil dropout solution and 2% dextrose and grown until an OD_{600} of approximately 1.5. The culture was continually back-diluted and grown in successively decreasing culture

volume for two days at which time freezer stocks were made of the fractions. 100 μl of the culture was collected and the library hairpins amplified by colony PCR with forward and reverse primers RntGap321_fwd and RntGap321_rev2, respectively. The PCR products representing the sorted binding library were recloned through a gap-repair method by transforming the DNA with the pCS321 plasmid in yeast strain W303.

3.4.4. Rnt1p substrate characterization assays

S. cerevisiae cells harboring pCS321-based plasmids were grown on synthetic complete media with an uracil dropout solution and the appropriate sugars (2% raffinose, 1% sucrose) overnight at 30ºC. The cells were back-diluted the following morning into fresh media (4.5 ml total volume in test tubes and 450 μl in deep-well plates) to an optical density at 600 nm (OD_{600}) of 0.1 and grown again at 30ºC. After 1 hr, 0.5 ml (test tubes) or 50 μl (plates) of 20% galactose (2% final concentration) or water (non-induced control) was added to the cell cultures. The cells were grown for another 4.5 hr before measuring the fluorescence levels or collecting cells for RNA extraction. Cells harboring pCS1585-based and pCS1748-based plasmids followed the same procedure as pCS321-based plasmids, except 2% dextrose was the only sugar in the media and no induction was required.

3.4.5. Fluorescence quantification

On the Quanta flow cytometer (Beckman Coulter, Fullerton, CA), the distribution of GFP fluorescence was measured with the following settings: 488-nm laser line, 525-nm bandpass filter, and photomultiplier tube setting of 5.83 (pCS321-based) of 4.50

(pCS1585-based). Data were collected under low flow rates until 10,000 viable cell counts were collected. For pCS321-based plasmids, a non-induced cell population was used to set a gate to represent GFP-negative and GFP-positive populations. For pCS1585-based plasmids, a plasmid harboring the same backbone as pCS1585 but with no fluorescence gene was used to set the GFP-negative and GFP-positive gates.

The LSRII flow cytometer (Becton Dickinson Immunocytometry Systems) was used to measure mCherry and GFP fluorescence from pmCh-Y-based plasmids. GFP was excited at 488 nm and measured with a splitter of 505 nm and a bandpass filter of 525/50. mCherry was excited at 532 nm and measured with a splitter of 600 nm LP and a bandpass filter of 610/20 nm. DAPI was excited at 405 nm and measured with a bandpass filter of 450/50 nm.

3.4.6. Quantification of cellular transcript levels

Total RNA from *S. cerevisiae* was collected by a standard hot acid phenol extraction method[44] and followed by DNase I (New England Biolabs) treatment to remove residual plasmid DNA according to manufacturer's instructions. cDNA was synthesized from 5 μg of total RNA with gene-specific primers for *yEGFP3* and *ACT1*[45] (rnt1p_rtpcr_rev2 and ACT1_rtpcr_rev, respectively) and SuperScript III Reverse Transcriptase (Invitrogen) according to manufacturer's instructions. The forward and reverse primers for *yEGFP3* quantification are rnt1p_rtpcr_fwd2 (5' CGGTGAAGGTGA AGGTGATGCTACT) and rnt1p_rtpcr_rev2 (5' GCTCTGGTCTTGTAGTTACCGTCA TCTTTG), respectively. The forward and reverse primers for *ACT1* quantification are ACT1_rtpcr_fwd (5' GGCATCATACCTTCTACAACGAAT) and ACT1_rtpcr_rev (5'

GGAATCCAAAACAATACCAGTAGTTCTA), respectively. Relative transcript levels were quantified in triplicate from three identical reactions from the cDNA samples by using an appropriate primer set and iQ SYBR Green Supermix (Bio-Rad) on an iCycler iQ qRT-PCR machine (Bio-Rad) according to the manufacturer's instructions. For each run, a standard curve was generated for *yEGFP3* and a house-keeping gene, *ACT1*, using a dilution series for a control representing no insertion of an Rnt1p substrate. Relative *yEGFP3* and *ACT1* levels were first individually determined for each sample and then the *yEGFP3* values were normalized by their corresponding *ACT1* values.

3.4.7. In vitro *transcription of Rnt1p substrates*

All Rnt1p substrates were PCR-amplified to include an upstream T7 promoter site using forward and reverse primers Rnt1p-T7-PCR_fwd_prmr (5' TTCTAATACGACTC ACTATAGGGACCTAGGAAACAAACAAAGTTGGGC) and Rnt1p-T7-PCR_rev_ prmr (5' CTCGAGTTTTTATTTTTCTTTTTGCCGGGCG), respectively. 1–2 μg of PCR product was transcribed with T7 RNA Polymerase (New England Biolabs) in the presence and absence of α-P^{32}-GTP. The 25-μl reaction consisted of the following components: 1x RNA Pol Reaction Buffer (New England Biolabs), 3 mM rATP, 3 mM rCTP, 3 mM rUTP, 0.3 mM rGFP, 1 μl RNaseOUT (Invitrogen), 10 mM $MgCl_2$, 2 mM DTT, 1 μl T7 Polymerase, and 0.5 μCi α-P^{32}-GTP. Unincorporated nucleotides were removed from the reactions by running the samples through NucAway Spin Columns (Ambion, Austin, TX) according to the manufacturer's instructions.

3.4.8. Rnt1p expression and purification

The pRNT1 plasmid was transformed into *E. coli* strain BL21 using the Z-competent *E. coli* Transformation Kit and Buffer Set (Zymo Research, Orange, CA) according to manufacturer's instructions. Rnt1p was collected as a protein extract as previously described[46]. Briefly, an overnight culture of BL21 cells harboring pRNT1 was back-diluted to an OD_{600} of 0.5. Once the culture reached an OD_{600} of 1.1–1.4, it was induced with 1 mM IPTG and grown for an additional 3 hr. The cells were centrifuged at 2,500g for 12 min at 4°C and the resulting cell pellet was frozen in a -80°C freezer. After weighing the frozen cell pellet, the cells were resuspended in 4 ml Ni_{2+} buffer [25% (v/v) glycerol, 1 M NaCl, 30 mM Tris pH 8.0] per gram of harvested cells. The resuspension was sonicated (Heat Systems-Ultrasonics, Inc.) twice with the following settings: 2 x 30 sec, output control 5, and 50% duty cycle. Cellular debris was removed by centrifugation at 20,000 g for 30 min at 4°C and the supernatant was filtered through a 0.2-μm pore size Acrodisc 25 mm syringe filter (Pall Life Sciences, Ann Arbor, MI).

Rnt1p was purified from the resulting supernatant with one 1-ml HisTrap HP column (GE Healthcare) on an AKTA FPLC machine (GE Healthcare). Elution of the protein was performed with an imidazole concentration of 150 mM in Ni_{2+} buffer and the protein was collected in 6 1-ml fractions. Protein purification was confirmed by analyzing an aliquot of each fraction on a SDS-PAGE gel (NuPAGE 4-12% Bis-Tris Gel, Invitrogen) and protein function was confirmed by incubating an aliquot of each fraction with a control Rnt1p substrate and analyzing the resulting cleavage products on an 8% denaturing polyacrylamide gel. Positive fractions were pooled and concentrated to less than a 3-ml volume using a Centricon Centrifugal Filter Device (10,000 MWCO;

Millipore) according to the manufacturer's instructions. The concentrated protein was then injected into a Slide-A-Lyzer Dialysis Cassette (10,000 MWCO; Pierce Biotechnology) and buffer-exchanged twice with Rnt1p Storage Buffer [50% (v/v) glycerol, 0.5 M KCl, 30 mM Tris pH 8.0, 0.1 M DTT, 0.1 M EDTA] at 4ºC. The first buffer exchange took place for 4 hr and the second buffer exchange occurred overnight. The purified Rnt1p was stored in aliquots at –20°C.

3.4.9. In vitro *Rnt1p substrate cleavage assay*

Cleavage assays were performed on Rnt1p substrates as previously described[38, 46]. Briefly, a 10-μl mixture of RNA and Rnt1p were incubated at 30ºC for 15 min in Rnt1p reaction buffer [30 mM Tris (pH 7.5), 150 mM KCl, 5 mM spermidine, 20 mM $MgCl_2$, 0.1 mM DTT, and 0.1 mM EDTA (pH 7.5)]. RNA concentrations were varied from 0.2 to 2.0 μM and the Rnt1p concentration was 2.3 μM. The cleavage reaction products were separated on an 8% denaturing polyacrylamide gel run at 35 W for 30 min. Gels were transferred to filter paper and analyzed for relative substrate and product levels through phosphorimaging analysis on a FX Molecular Imager (Bio-Rad). The levels of cleaved RNA product were determined and fit to a Michaelis-Menten model using Prism 5 (GraphPad), where a relative V_{max} was calculated and reported with the standard error determined by the fit of the model.

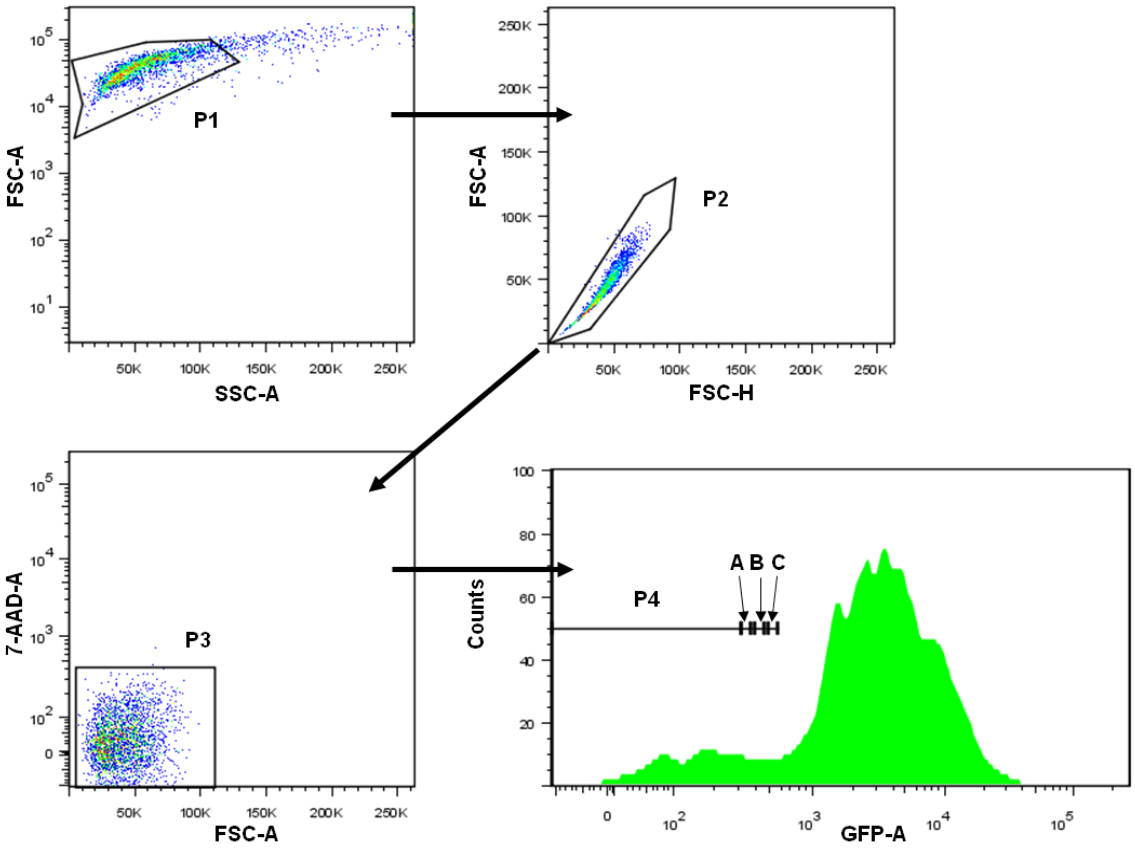
3.4.10. In vitro Rnt1p substrate mobility shift assay

Mobility shift assays were performed as previously described[38, 46]. Briefly, a 10-μl mixture of RNA and Rnt1p were incubated on ice for 10 min in Rnt1p binding buffer

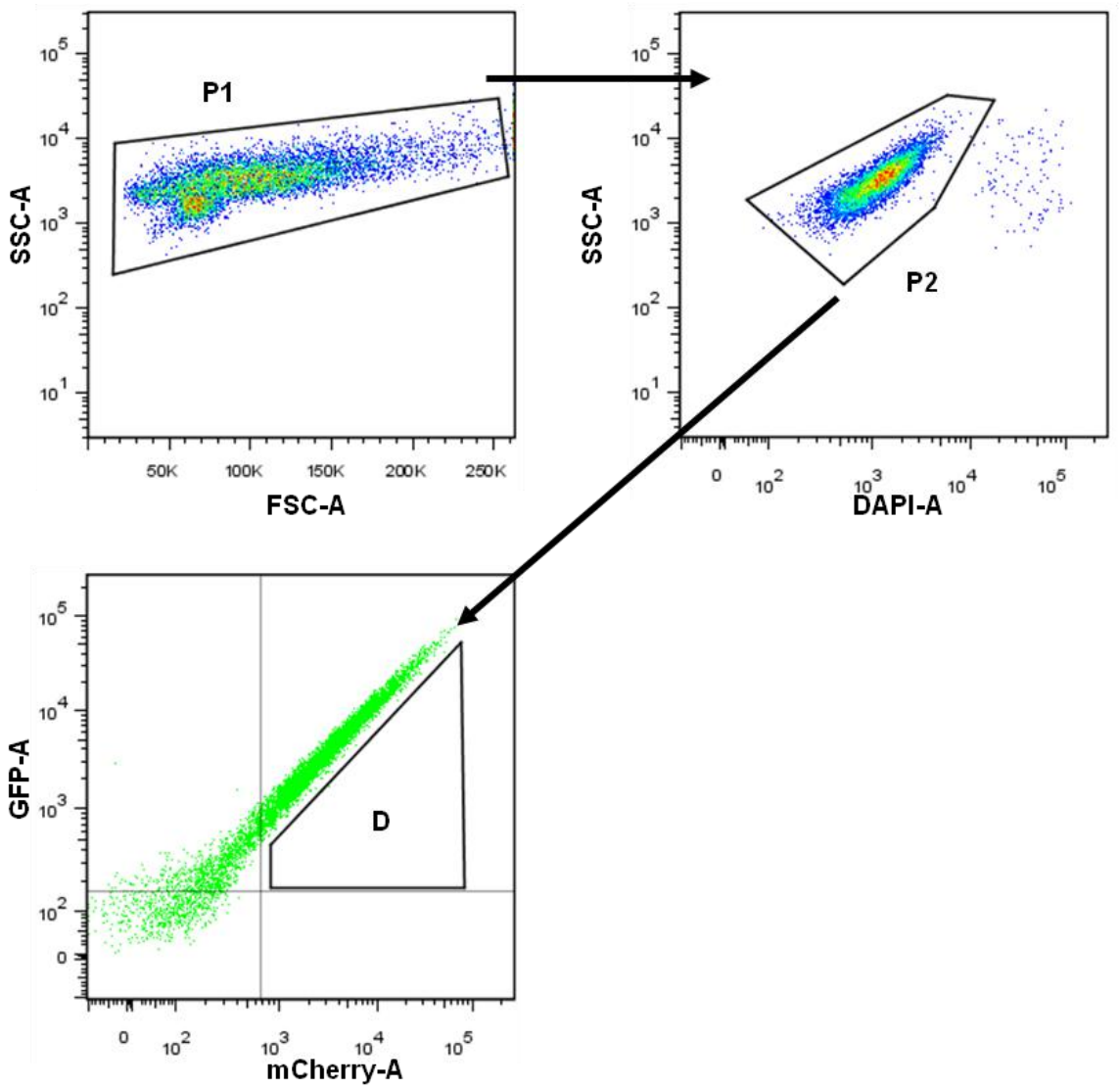
[20% (v/v) glycerol, 30 mM Tris (pH 7.5), 150 mM KCl, 5 mM spermidine, 0.1 mM DTT, and 0.1 mM EDTA (pH 7.5)]. The RNA concentration in all samples was 200 nM and the Rnt1p concentration ranged from 0 to 1.7 μM. The binding reaction products were separated on a 6% native polyacrylamide gel run at 350 V until the samples entered the gel and then at 150 V for 2 hr. Gels were transferred to filter paper and analyzed for free RNA and RNA-Rnt1p complex levels through phosphorimaging analysis on a FX Molecular Imager. The fraction of unbound RNA to total RNA was determined and fit to a modified Scatchard model using Prism 5, where a K_D value was calculated and reported with the standard error determined by the fit of the model.

3.5. Supplementary Information

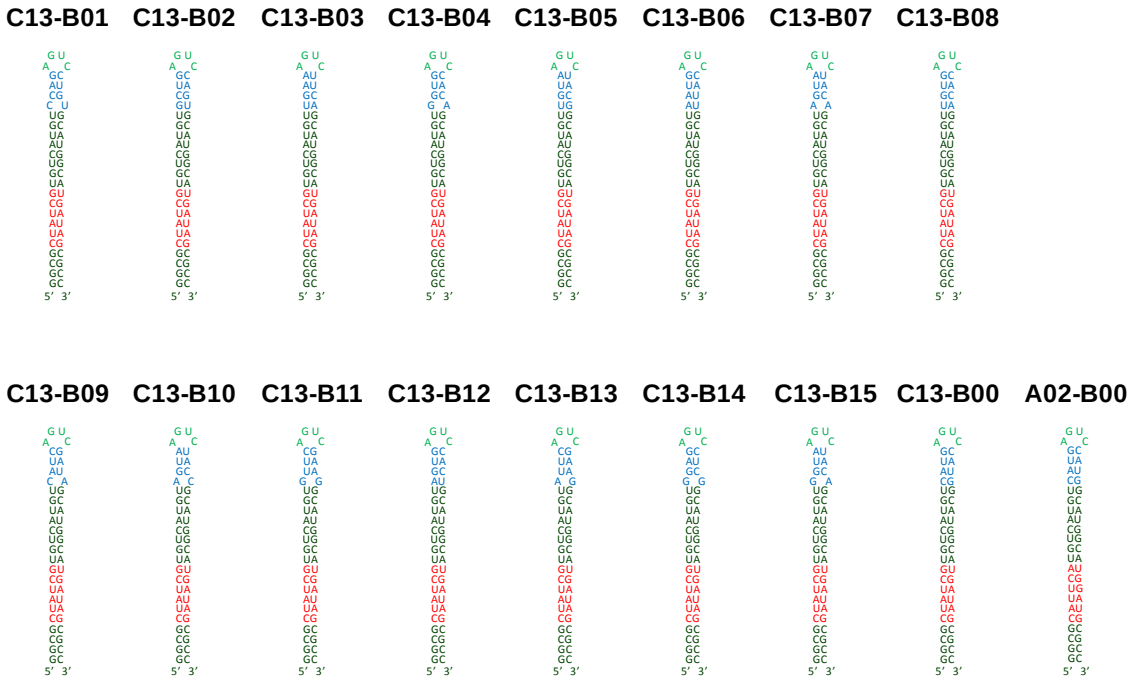
Supplementary Figures and Tables



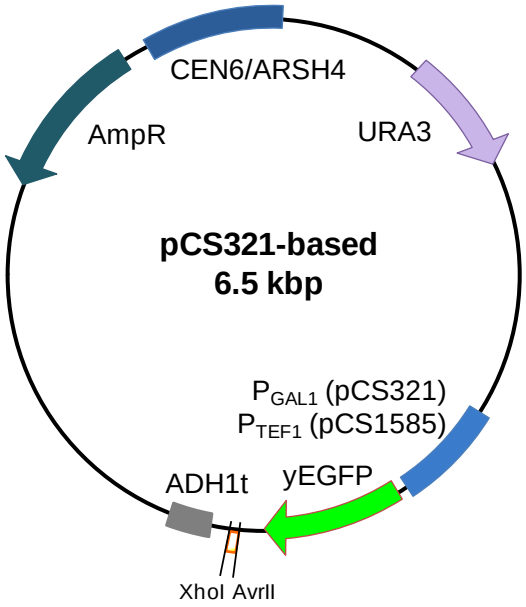
Supplementary Figure 3.1. FACS analysis and gating procedure for pCS1585 system on FACSAria. As an example, data for the construct baring no Rnt1p hairpin is presented. Dot plots show initial gating of stable cells (P1), followed by gating for cell uniformity (P2), and finally gating for live cells with the 7-AAD stain (P3). GFP-negative cells (P4) were gated initially with a construct lacking a fluorescent gene (empty vector). Cells outside of P4 represent GFP-positive cells. Fractions A, B, and C are represented on this graph, but collections were only performed with the binding library sample. The fractions cover the range of expression seen with the C13-B00 hairpin. With the binding library sample, ~120,000 cells were analyzed with 719, 841, and 943 cells collected in fractions A, B, and C respectively.



Supplementary Figure 3.2. FACS analysis and gating procedure for pCS1748 system on FACSAria II. As an example, data for the GFP positive construct (no Rnt1p hairpin) is presented. Dot plots show initial gating of stable cells (P1) and subsequent gating of uniform, live cells (P2) with DAPI used for the viability stain. A construct lacking fluorescent genes (empty vector) was used to set the gates for mCherry- and GFP-positive cells. A single gate (D) was set to collect all GFP-positive cells that exhibited lower GFP fluorescence than cells containing the positive construct. With the binding library sample, 1,000,000 cells were analyzed with 18,416 cells collected in fraction D.



Supplementary Figure 3.3. Sequences and structures of the selected Rnt1p binding library and control hairpins containing the 'parent' BSB. The binding library was initially sequenced when in the context of the C13 CEB.



Supplementary Figure 3.4. Plasmid maps of pCS321-based vectors. pCS321 is the characterization plasmid and GFP expression is driven by the GAL1 promoter. pCS1585 is a screening plasmid used with FACS and GFP expression is driven by the TEF1 promoter.

Supplementary Table 3.1. Sequence and *in vivo* characterization data of all tested Rnt1p hairpins. The nucleotides of the BSB are indicated in blue. The CEB sequences in the 'parent' hairpins (xx-B00) are indicated in red. All normalized protein and transcript levels were determined as described in Figure 3.3A and Figure 3.3B, respectively.

Substrate	Sequence	Normalized protein levels (%)	Normalized transcript levels (%)
A02-B00	GGCGCAUUCAUGUCAUGUCAUGAG UCCAUGGCAUGGCAUGGAUGCGCC	28% ± 1%	43% ± 8%
A02-B01	GGCGCAUUCAUGUCAUGUCCAGAG UCCUGUGCAUGGCAUGGAUGCGCC	75% ± 3%	78% ± 11%
A02-B02	GGCGCAUUCAUGUCAUGUGCUGAG UCCAGUGCAUGGCAUGGAUGCGCC	62% ± 2%	64% ± 6%
A02-B03	GGCGCAUUCAUGUCAUGUUGAAAG UCUUCAGCAUGGCAUGGAUGCGCC	50% ± 2%	53% ± 5%
A02-B04	GGCGCAUUCAUGUCAUGUGGUGAG UCCACAGCAUGGCAUGGAUGCGCC	32% ± 1%	57% ± 7%
A02-B05	GGCGCAUUCAUGUCAUGUUGUAAG UCUACGGCAUGGCAUGGAUGCGCC	25% ± 0%	36% ± 5%
A02-B06	GGCGCAUUCAUGUCAUGUAAUGAG UCCAUUGCAUGGCAUGGAUGCGCC	27% ± 2%	60% ± 6%
A02-B07	GGCGCAUUCAUGUCAUGUAGUAAG UCUACAGCAUGGCAUGGAUGCGCC	37% ± 3%	51% ± 3%
A02-B08	GGCGCAUUCAUGUCAUGUUGUGAG UCCACAGCAUGGCAUGGAUGCGCC	30% ± 2%	53% ± 4%
A02-B09	GGCGCAUUCAUGUCAUGUCAUCAG UCGAUAGCAUGGCAUGGAUGCGCC	36% ± 3%	60% ± 9%
A02-B10	GGCGCAUUCAUGUCAUGUAGUAAG UCUACCGCAUGGCAUGGAUGCGCC	42% ± 3%	55% ± 4%
A02-B11	GGCGCAUUCAUGUCAUGUGUUCAG UCGAAGGCAUGGCAUGGAUGCGCC	32% ± 2%	51% ± 4%
A02-B12	GGCGCAUUCAUGUCAUGUAGUGAG UCCACUGCAUGGCAUGGAUGCGCC	27% ± 2%	47% ± 5%
A02-B13	GGCGCAUUCAUGUCAUGUAUUCAG UCGAAGGCAUGGCAUGGAUGCGCC	39% ± 4%	53% ± 5%
A02-B14	GGCGCAUUCAUGUCAUGUGGAGAG UCCUCGGCAUGGCAUGGAUGCGCC	48% ± 4%	74% ± 8%
A02-B15	GGCGCAUUCAUGUCAUGUGGUAAG UCUACAGCAUGGCAUGGAUGCGCC	48% ± 4%	58% ± 7%
C13-B00	GGCGCUAUCGUGUCAUGUCAUGAG UCCAUGGCAUGGCAUGAUAGCGCC	8% ± 0%	12% ± 1%
C13-B01	GGCGCUAUCGUGUCAUGUCCAGAG UCCUGUGCAUGGCAUGAUAGCGCC	48% ± 2%	
C13-B02	GGCGCUAUCGUGUCAUGUGCUGAG UCCAGUGCAUGGCAUGAUAGCGCC	37% ± 2%	
C13-B03	GGCGCUAUCGUGUCAUGUUGAAAG UCUUCAGCAUGGCAUGAUAGCGCC	20% ± 1%	
C13-B04	GGCGCUAUCGUGUCAUGUGGUGAG UCCACAGCAUGGCAUGAUAGCGCC	9% ± 0%	
C13-B05	GGCGCUAUCGUGUCAUGUUGUAAG UCUACGGCAUGGCAUGAUAGCGCC	8% ± 0%	
C13-B06	GGCGCUAUCGUGUCAUGUAAUGAG UCCAUUGCAUGGCAUGAUAGCGCC	8% ± 0%	
C13-B07	GGCGCUAUCGUGUCAUGUAGUAAG UCUACAGCAUGGCAUGAUAGCGCC	11% ± 1%	
C13-B08	GGCGCUAUCGUGUCAUGUUGUGAG UCCACAGCAUGGCAUGAUAGCGCC	10% ± 0%	

C13-B09	GGCGCUAUCGUGUCAUGUCAUCAG UCGAUAGCAUGGCAUGAUAGCGCC	15% ± 1%
C13-B10	GGCGCUAUCGUGUCAUGUAGUAAG UCUACCGCAUGGCAUGAUAGCGCC	14% ± 1%
C13-B11	GGCGCUAUCGUGUCAUGUGUUCAG UCGAAGGCAUGGCAUGAUAGCGCC	11% ± 1%
C13-B12	GGCGCUAUCGUGUCAUGUAGUGAG UCCACUGCAUGGCAUGAUAGCGCC	10% ± 0%
C13-B13	GGCGCUAUCGUGUCAUGUAUUCAG UCGAAGGCAUGGCAUGAUAGCGCC	14% ± 0%
C13-B14	GGCGCUAUCGUGUCAUGUGGAGAG UCCUCGGCAUGGCAUGAUAGCGCC	17% ± 0%
C13-B15	GGCGCUAUCGUGUCAUGUGGUAAG UCUACAGCAUGGCAUGAUAGCGCC	14% ± 1%

Acknowledgements

We thank J. Liang and A. Chang for assistance with FACS and for providing the pCS1585 and pCS1748 plasmids; K. Hoff for assistance in the expression and purification of Rnt1p; and S. Bastian and F.H. Arnold for assistance in sonication and FPLC. This work was supported by the National Science Foundation (CAREER award CBET-0917705 to C.D.S.); and the Alfred P. Sloan Foundation, fellowship (to C.D.S.).

References

1. Jin, Y.S., Ni, H., Laplaza, J.M. & Jeffries, T.W. Optimal growth and ethanol production from xylose by recombinant Saccharomyces cerevisiae require moderate D-xylulokinase activity. *Appl Environ Microbiol* **69**, 495-503 (2003).
2. Jones, K.L., Kim, S.W. & Keasling, J.D. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab Eng* **2**, 328-338 (2000).

3. Pfleger, B.F., Pitera, D.J., Smolke, C.D. & Keasling, J.D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* **24**, 1027-1032 (2006).

4. Zhu, M.M., Lawman, P.D. & Cameron, D.C. Improving 1,3-propanediol production from glycerol in a metabolically engineered Escherichia coli by reducing accumulation of sn-glycerol-3-phosphate. *Biotechnol Prog* **18**, 694-699 (2002).

5. Zhu, M.M., Skraly, F.A. & Cameron, D.C. Accumulation of methylglyoxal in anaerobically grown Escherichia coli and its detoxification by expression of the Pseudomonas putida glyoxalase I gene. *Metab Eng* **3**, 218-225 (2001).

6. Alper, H., Jin, Y.S., Moxley, J.F. & Stephanopoulos, G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. *Metab Eng* **7**, 155-164 (2005).

7. Alper, H., Miyaoku, K. & Stephanopoulos, G. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* **23**, 612-616 (2005).

8. Paradise, E.M., Kirby, J., Chan, R. & Keasling, J.D. Redirection of flux through the FPP branch-point in Saccharomyces cerevisiae by down-regulating squalene synthase. *Biotechnol Bioeng* **100**, 371-378 (2008).

9. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**, 1203-1210 (1997).

10. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* **102**, 12678-12683 (2005).

11. Nevoigt, E. et al. Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. *Appl Environ Microbiol* **72**, 5266-5273 (2006).

12. Hawkins, K.M. & Smolke, C.D. The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae. *J Biol Chem* **281**, 13485-13492 (2006).

13. Jensen, P.R. & Hammer, K. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl Environ Microbiol* **64**, 82-87 (1998).

14. Win, M.N. & Smolke, C.D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A* **104**, 14283-14288 (2007).

15. Beisel, C.L., Bayer, T.S., Hoff, K.G. & Smolke, C.D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol Syst Biol* **4**, 224 (2008).

16. Carrier, T.A. & Keasling, J.D. Library of synthetic 5' secondary structures to manipulate mRNA stability in Escherichia coli. *Biotechnol Prog* **15**, 58-64 (1999).

17. Anderson, J.C., Clarke, E.J., Arkin, A.P. & Voigt, C.A. Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol* **355**, 619-627 (2006).

18. Keiler, K.C., Waller, P.R. & Sauer, R.T. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**, 990-993 (1996).

19. Mateus, C. & Avery, S.V. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast* **16**, 1313-1323 (2000).

20. Hawkins, K.M. & Smolke, C.D. Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat Chem Biol* **4**, 564-573 (2008).

21. Nguyen, H.T. et al. Engineering of Saccharomyces cerevisiae for the production of L-glycerol 3-phosphate. *Metab Eng* **6**, 155-163 (2004).

22. Ostergaard, S., Olsson, L. & Nielsen, J. Metabolic engineering of Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **64**, 34-50 (2000).

23. Ro, D.K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943 (2006).

24. Szczebara, F.M. et al. Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nat Biotechnol* **21**, 143-149 (2003).

25. Veen, M. & Lang, C. Production of lipid compounds in the yeast Saccharomyces cerevisiae. *Appl Microbiol Biotechnol* **63**, 635-646 (2004).

26. Louis, M. & Becskei, A. Binary and graded responses in gene networks. *Sci STKE* **2002**, pe33 (2002).

27. Wu, H., Henras, A., Chanfreau, G. & Feigon, J. Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc Natl Acad Sci U S A* **101**, 8307-8312 (2004).

28. Wu, H. et al. A novel family of RNA tetraloop structure forms the recognition site for Saccharomyces cerevisiae RNase III. *EMBO J* **20**, 7240-7249 (2001).

29. Lamontagne, B. et al. Sequence dependence of substrate recognition and cleavage by yeast RNase III. *J Mol Biol* **327**, 985-1000 (2003).

30. Lamontagne, B., Hannoush, R.N., Damha, M.J. & Abou Elela, S. Molecular requirements for duplex recognition and cleavage by eukaryotic RNase III: discovery of an RNA-dependent DNA cleavage activity of yeast Rnt1p. *J Mol Biol* **338**, 401-418 (2004).

31. Chanfreau, G., Elela, S.A., Ares, M., Jr. & Guthrie, C. Alternative 3'-end processing of U5 snRNA by RNase III. *Genes Dev* **11**, 2741-2751 (1997).

32. Chanfreau, G., Rotondo, G., Legrain, P. & Jacquier, A. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J* **17**, 3726-3737 (1998).

33. Elela, S.A., Igel, H. & Ares, M., Jr. RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell* **85**, 115-124 (1996).

34. Ge, D., Lamontagne, B. & Elela, S.A. RNase III-mediated silencing of a glucose-dependent repressor in yeast. *Curr Biol* **15**, 140-145 (2005).

35. Danin-Kreiselman, M., Lee, C.Y. & Chanfreau, G. RNAse III-mediated degradation of unspliced pre-mRNAs and lariat introns. *Mol Cell* **11**, 1279-1289 (2003).

36. Lee, A., Henras, A.K. & Chanfreau, G. Multiple RNA surveillance pathways limit aberrant expression of iron uptake mRNAs and prevent iron toxicity in S. cerevisiae. *Mol Cell* **19**, 39-51 (2005).

37. Lamontagne, B., Tremblay, A. & Abou Elela, S. The N-terminal domain that distinguishes yeast from bacterial RNase III contains a dimerization signal required for efficient double-stranded RNA cleavage. *Mol Cell Biol* **20**, 1104-1115 (2000).

38. Lamontagne, B. & Elela, S.A. Evaluation of the RNA determinants for bacterial and yeast RNase III binding and cleavage. *J Biol Chem* **279**, 2231-2241 (2004).

39. Lavoie, M. & Abou Elela, S. Yeast ribonuclease III uses a network of multiple hydrogen bonds for RNA binding and cleavage. *Biochemistry* **47**, 8514-8526 (2008).

40. Misra, V.K. & Draper, D.E. The linkage between magnesium binding and RNA folding. *J Mol Biol* **317**, 507-521 (2002).

41. Sambrook, J. & Russell, D.W. Molecular Cloning: A Laboratory Manual, Edn. 3rd. (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY; 2001).

42. Gietz, R. & Woods, R. in Guide to Yeast Genetics and Molecular and Cell Biology, Part B, Vol. 350. (eds. C. Guthrie & G. Fink) 87-96 (Academic, San Diego, CA; 2002).

43. Chao, G. et al. Isolating and engineering human antibodies using yeast surface display. *Nat Protoc* **1**, 755-768 (2006).

44. Caponigro, G., Muhlrad, D. & Parker, R. A small segment of the MAT alpha 1 transcript promotes mRNA decay in Saccharomyces cerevisiae: a stimulatory role for rare codons. *Mol Cell Biol* **13**, 5141-5148 (1993).

45. Ng, R. & Abelson, J. Isolation and sequence of the gene for actin in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **77**, 3912-3916 (1980).

46. Lamontagne, B. & Elela, S.A. Purification and characterization of Saccharomyces cerevisiae Rnt1p nuclease. *Methods Enzymol* **342**, 159-167 (2001).

Chapter IV: Engineering ligand-responsive RNA controllers in yeast through the assembly of RNase III tuning modules

Abstract

The programming of cellular networks to achieve new biological functions depends on the development of genetic tools that link the presence of a molecular signal to gene-regulatory activity. Recently, a set of engineered RNA controllers was described that enabled predictable tuning of gene expression in the yeast *Saccharomyces cerevisiae* through directed cleavage of transcripts by an RNase III enzyme, Rnt1p. Here, we describe a strategy for building a new class of RNA sensing-actuation devices based on direct integration of RNA aptamers into a region of the Rnt1p hairpin that modulates Rnt1p cleavage rates. We demonstrate that ligand binding to the integrated aptamer domain is associated with a structural change sufficient to inhibit Rnt1p processing. Two tuning strategies based on the incorporation of different functional modules into the Rnt1p switch platform were demonstrated to optimize switch dynamics and ligand responsiveness. We further demonstrated that these tuning modules can be implemented combinatorially in a predictable manner to further improve the regulatory response properties of the switch, which resulted in an increase in the fold-change from 1.93 to 2.47. A third tuning strategy was employed by placing multiple copies of the switch in tandem. Three copies of the single module switch resulted in a increase of fold-change from 1.93 to 5.57. The modularity and tunability of the Rnt1p switch platform will allow for rapid optimization and tailoring of this gene control device, thus providing a useful tool for the design of complex genetic networks in yeast.

4.1. Introduction

The field of synthetic biology encompasses the engineering of biological systems that exhibit new functions through the design of synthetic gene networks. The proper functioning of genetic networks encoding complex behaviors depends on the coordinated regulation of genetic responses, enzymatic activities, and protein levels[1-4]. As such, the genetic programming of biological systems depends on our ability to design genetic devices that can detect molecular signals and link these detection events to new types of genetic control and thus biological function.

Motivated by the versatility of sensing and actuation functions that RNA can exhibit and the relative ease with which RNA structures can be modeled and designed, researchers have engineered RNA-based devices that detect diverse molecular signals and link this information to the regulation of gene expression events[5]. RNA devices generally couple RNA components that exhibit sensing, actuation, and information transmission activities. By varying the regulatory RNA encoded within the actuator component, RNA devices have been developed that function in different organisms through a variety of gene-regulatory mechanisms, including translation[6-10], transcript degradation[11-12], transcriptional activation[13], and splicing[14-16]. Recently, several examples of RNA devices that act through the modulation of RNase III processing activities in mammalian cells have been described[17-19]. However, such regulatory strategies have not yet been extended to RNase III enzymes in microorganisms.

We previously developed a genetic control system based on directing transcript cleavage through the RNase III enzyme in the yeast *Saccharomyces cerevisiae* (Rnt1p), where Rnt1p hairpin substrates are placed in the 3’ untranslated region (UTR) of the

targeted gene (Chapter II). We utilized cell-based screening strategies to develop libraries of synthetic Rnt1p hairpins that exhibit predictable gene-regulatory activity based on modules that are inserted into two key regions of the hairpin – the cleavage efficiency box (CEB) and the binding stability box (BSB). These studies indicated that the CEB and BSB modules function independently and can be combinatorially integrated into a single Rnt1p hairpin to achieve a wide range of gene-regulatory activities (Chapters II and III).

Here, we describe the design of a new class of RNA devices based on the synthetic library of Rnt1p regulatory elements, which we call an Rnt1p switch. We have designed a strategy for coupling RNA aptamers to Rnt1p hairpins to build RNA control systems in yeast that exhibit integrated sensing and actuation functions. Specifically, theophylline aptamers were integrated into the CEB region of the Rnt1p hairpin, which plays a key role in modulating Rnt1p cleavage rates, where binding of the small molecule to the aptamer sequence resulted in inhibition of cleavage activity. We demonstrated three different strategies for tuning the quantitative properties of the device response curve based on the integration of different modules into the Rnt1p switch platform. Aptamer/CEB, BSB, and multiple switch modules were used to modulate the EC50 value and switching activity of the device. Finally, we demonstrated that these modules can be implemented combinatorially in a predictable manner to further improve the regulatory response properties of the switch. The application of these tuning strategies resulted in an increase of the fold-change from 1.93 to 2.47 in a single module switch and an increase from 1.93 to 5.57 when three copies of the single module switch were implemented.

4.2. Results

4.2.1. Design of a ligand-responsive RNA switch based on Rnt1p processing

We set out to utilize the synthetic Rnt1p genetic control elements developed in early work from our laboratory in the design of a new class of RNA devices based on the modulation of Rnt1p processing, called an Rnt1p switch. The first step in building an integrated RNA device platform is identifying a physical coupling strategy between the sensor (i.e., aptamer) and actuator (i.e., Rnt1p hairpin) components that maintains the activity of each component, but allows ligand binding to the sensor to modify the activity of the actuator. Rnt1p substrates adopt hairpin structures that contain three critical regions for enzyme binding and processing: the initial binding and positioning box (IBPB), comprising an AGNN tetraloop; the BSB, comprising the base-paired region immediately below the tetraloop; and the CEB, comprising the nucleotides adjacent to the two cleavage sites[20] (Figure 4.1A). Rnt1p initially binds to the tetraloop (the IBPB) and the interaction is stabilized by the BSB. The enzyme then cleaves the hairpin at two locations within the CEB: between the 14th and 15th nts upstream of the tetraloop and the 16th and 17th nts downstream of the tetraloop. While the CEB has no structural or sequence requirements, the BSB and IBPB have more rigid sequence and structural requirements[21]. Therefore, we identified the CEB as a possible region of the Rnt1p hairpin where the integration of aptamer sequences may not disrupt enzyme binding and processing.

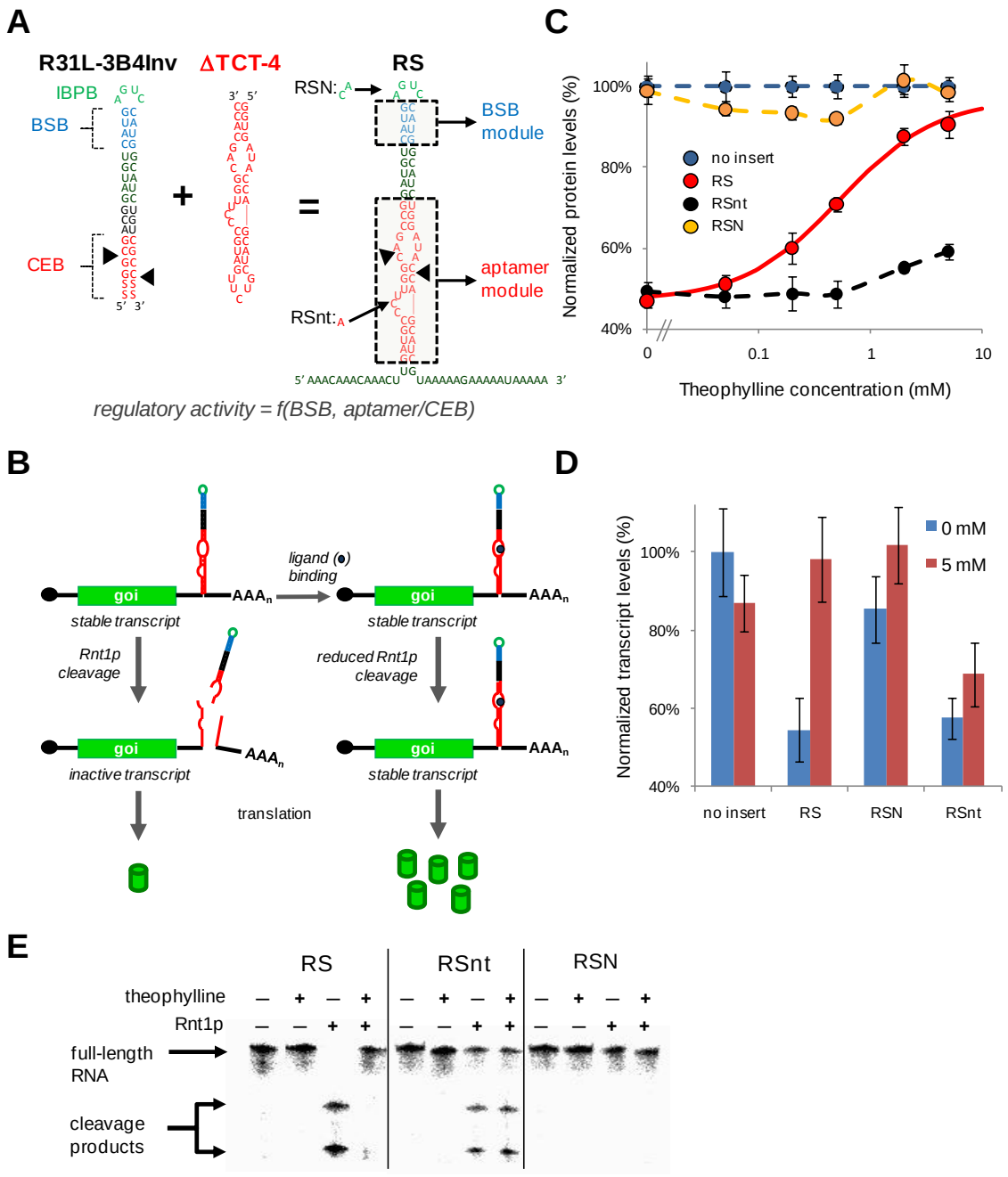


Figure 4.1. Design and implementation of Rnt1p switches as posttranscriptional genetic control elements. (A) The design of an Rnt1p switch (RS) by the integration of a sensor component (ΔTCT-4 aptamer) into the actuator component (R31L-3B4Inv Rnt1p hairpin). R31L-3B4Inv contains the consensus regions of a Rnt1p substrate: the cleavage efficiency box (CEB), the binding stability box (BSB), and the initial binding and positioning box (IBPB). The nucleotide modifications associated with the Rnt1p and ligand binding controls (RSN and RSnt, respectively) are indicated on RS. Color scheme is as follows: CEB and aptamer, red; BSB, blue; IBPB, green. Black triangles indicate locations of cleavage sites by Rnt1p. The regulatory activity of the synthetic Rnt1p

switches is a function of the modular elements in the BSB and CEB/aptamer, which are indicated by the dashed boxes on RS. (B) The synthetic Rnt1p switches control the destabilization of target transcripts by controlling Rnt1p cleavage in the 3' UTR of transcripts encoding a gene of interest (goi). Barrels represent protein molecules and dark blue circles represent the ligand molecule. In the absence of ligand binding to the aptamer module, the transcript is inactivated by Rnt1p cleavage and translation inhibited. Ligand binding to the aptamer module in the CEB inhibits Rnt1p cleavage activity and stabilizes the transcript, resulting in increased protein production. (C) The dose response curves of RS and the Rnt1p and ligand binding controls, RSN and RSnt, respectively, indicate that decreased gene expression is caused by Rnt1p cleavage and that cleavage is inhibited by theophylline. The following concentrations of theophylline are used for all response curves (in mM): 0, 0.05, 0.2, 0.5, 2, 5. Normalized protein levels were determined by measuring the median GFP levels from cells harboring constructs with the indicated switch through flow cytometry analysis, and values are reported relative to a construct lacking a hairpin module (set to 100%). Solid curves indicate the theoretically-determined model fit. The model parameters for the curve fit are provided in Table 4.1. Dashed curves are utilized for control constructs that are not fit to the model and are generated through Microsoft Excel's smooth line option. (D) Analysis of transcript levels of the Rnt1p switch, RS, and the controls, RSN and RSnt, supports the proposed mechanism of inhibition of Rnt1p processing due to ligand binding. Relative transcript levels are determined by measuring transcript levels of *yEGFP3* and a house-keeping gene, *ACT1*, through qRT-PCR and normalizing the *yEGFP3* levels with their corresponding *ACT1* levels. Normalized transcript levels for each construct are reported relative to that from an identical construct lacking a hairpin module. (E) Cleavage reaction assays and analyses by denaturing polyacrylamide gel electrophoresis on the Rnt1p switch RS and the controls, RSnt and RSN, support the proposed mechanism of inhibition of Rnt1p processing due to ligand binding. The top band corresponds to uncleaved full-length RNA; the bottom bands correspond to the three cleavage products expected from Rnt1p processing. Two of the expected cleavage products differ in size by 2 nt and cannot be resolved into individual bands under the conditions used for this assay. RNA was added to the final concentration of 0.05 μM in each reaction. When present, Rnt1p was added to the final concentration of 20.7 μM and theophylline was added to the final concentration of 10 mM.

We designed an Rnt1p switch platform based on direct replacement of the CEB with an aptamer sequence (Figure 4.1A). Ligand binding in the CEB can inhibit Rnt1p cleavage by inducing secondary or tertiary structural changes in the region that Rnt1p is unable to process effectively or by sequestering the CEB nucleotides. The Rnt1p switch is located in the 3' UTR of a target transcript, where it acts to modulate the stability of that transcript (Figure 4.1B). Specifically, in the absence of ligand the hairpin is

processed by Rnt1p, resulting in transcript destabilization and a decrease in gene expression. However, ligand binding at the integrated aptamer component will result in inhibition of Rnt1p processing of the hairpin structure, resulting in transcript stabilization and an increase in expression. Thus, the Rnt1p switch acts as a gene expression 'ON' switch, where gene expression is expected to increase with increasing ligand concentrations. In addition, A-rich spacer sequences were placed around the device to more readily enable the extension of the system to the integration of multiple devices in the 3' UTR (Figure 4.1A).

The theophylline aptamer ΔTCT-4[22] was selected for the initial development of the Rnt1p switch platform due to this sequence having the highest reported affinity for the small molecule. However, the placement of this aptamer in the previously characterized Rnt1p substrate R31-27[21], which is the same base hairpin that was used in the development of the Rnt1p cleavage library (Chapter II), resulted in the disruption of effective Rnt1p cleavage of the hairpin in the absence of ligand (data not shown). To lower the baseline of the 'OFF' state and thus provide a greater potential switching range, we utilized a different Rnt1p base hairpin in the switch design that had been previously reported to have the highest rate of Rnt1p processing *in vitro* (R31L-3B4Inv)[20]. For all experiments, the highest external theophylline concentration applied to the yeast cells was 5 mM. Above this concentration, theophylline begins to become toxic to yeast and the reliability of the data decreases. All ligand titration data were fit to a simple binding model with Prism 5 (GraphPad) as:

$$Y(t) = b + \frac{(M - b) * t}{EC50 + t}$$

where *Y* represents the normalized gene expression data for a given theophylline concentration (*t*), *b* represents the amount of gene expression in the absence of theophylline, and *M* is the theoretical maximal 'ON' state determined by the model. The value of *b* was determined experimentally and not fit in the nonlinear regression. The regulatory range of the switch is characterized by two measures, the fold-change, determined as the ratio of the 'ON' state to the 'OFF' state, and the dynamic range, determined as the difference between the 'ON' and the 'OFF' states. The responsiveness of the switch was determined as the EC50 value of the dose response curve (reported in terms of extracellular theophylline concentration). *M* gives a measure of the regulatory range of the switch in the case where the ligand is not cytotoxic at concentrations above 5 mM. The theoretical fold-change and dynamic range based on the measured values of *b* and the model-calculated parameter *M* are reported for all switches in Supplementary Table 4.1.

We placed the theophylline aptamer in several locations within the CEB to screen for the most effective aptamer integration site (Supplementary Figure 4.1). One integration site location resulted in a functioning RNA switch (RS-theo1-B00), which retained the ability to be effectively cleaved by Rnt1p in the 'OFF' state and exhibited theophylline dependent increases in target gene expression levels (Figure 4.1C). In the switch naming system, the 'theo1' indicates the identity of the integrated aptamer (ΔTCT-4 aptamer) and 'B00' indicates the identity of the BSB module contained in R31L-3B4Inv, which is the same as the BSB module used in the development of the cleavage library (Chapter II). We refer to the parent Rnt1p switch RS-theo1-B00 as RS, and for other switches we will only indicate modules different than the parent switch for

simplicity. The RS module switches from 47% to 91% as theophylline concentrations increase from 0 to 5 mM, resulting in a fold-change of 1.93 and a dynamic range of 44% (Table 4.1). The EC50 value for RS was determined to be 0.54 mM, where previous studies have indicated a substantial drop in theophylline across the cell membrane ranging from ~300- to 1,500-fold[18, 23](Liang, J.C., Michener, J.K., and Smolke, C.D., unpublished data, 2008). To verify that the reduced gene expression levels from RS were due to Rnt1p cleavage and that the theophylline dependent response was due to the integrated aptamer sequence, we characterized two control hairpins: RSN, containing a mutant CAUC tetraloop that had been previously shown to impede Rnt1p activity *in vivo* (Chapter II), and RSnt, containing a single nucleotide mutation in the theophylline aptamer that obstructs the associated ligand binding activity. The dose response curve of RSN demonstrates gene expression levels at or around 100% for all theophylline concentrations, whereas the dose response curve of RSnt demonstrates a consistent level of reduced gene expression independent of theophylline (Figure 4.1C).

Rnt1p processing is expected to destabilize the transcript and thus reduce transcript levels. We monitored relative steady-state transcript levels of RS and its controls, RSN and RSnt, in the presence and absence of ligand by qRT-PCR and confirmed that decreased transcript levels are associated with decreased protein expression (Figure 4.1D). In order to further verify the competition between ligand-binding and Rnt1p cleavage, we performed *in vitro* cleavage assays with purified Rnt1p and radiolabeled Rnt1p substrates in the presence and absence of theophylline (Figure 4.1E). We observed no Rnt1p activity on the mutated tetraloop control (RSN) and the inability of theophylline to impact Rnt1p activity with the inactive aptamer control

(RSnt). In contrast, Rnt1p was able to cleave the switch (RS) in the absence of theophylline, but that activity was inhibited when theophylline was added. We also performed identical reactions without Rnt1p to examine any nonspecific effects of theophylline on the Rnt1p substrates, which confirmed that theophylline has no effect on how the RNA runs in the gel (Figure 4.1E). Taken together, the results indicate that the knockdown in expression levels and the theophylline response exhibited by RS are due to Rnt1p cleavage and ligand binding to the integrated aptamer, respectively.

Table 4.1. Relevant parameters for all RS-based switches and the Rnt1p and ligand binding controls, RSN and RSnt. *b* and *Y* are experimentally determined values corresponding to the normalized protein levels at 0 mM and 5 mM theophylline, respectively. The EC50 and theoretical maximal output (*M*) are parameters determined by fitting the dose response data to the binding model. The fold-change is the ratio of *Y* to *b* and the dynamic range is the difference between these two values. The theoretical fold-change and dynamic range determined by *M* instead of *Y* are reported in Supplementary Table 4.1.

switch	EC50 (mM)	b=Y(0mM)	Y(5mM)	fold-change	dynamic range	M
RS	0.54±0.05	47±2%	91±3%	1.93±0.10	44±4%	97±1%
RSN		98±3%	101±5%	1.03±0.06	3±6%	
RSnt		49±2%	59±2%	1.20±0.07	10±3%	
RS-theo2	1.47±0.30	52±2%	84±2%	1.60±0.06	32±3%	94±3%
RS-theo3	1.17±0.09	41±4%	101±4%	2.47±0.26	60±6%	116±2%
RS-B03	0.48±0.08	79±4%	102±4%	1.29±0.08	23±5%	105±1%
RS-B05	0.56±0.07	51±1%	97±4%	1.91±0.10	46±4%	104±2%
RS-B06	0.59±0.11	56±5%	101±6%	1.82±0.19	46±7%	108±3%
RS-B07	0.84±0.04	44±2%	100±4%	2.30±0.13	57±5%	110±1%
RS-B12	0.47±0.08	44±2%	94±4%	2.11±0.12	50±4%	99±2%
RSx2	1.07±0.02	20±1%	79±3%	3.91±0.19	59±3%	92±0%
RSx3	1.81±0.06	10±0%	57±1%	5.57±0.26	47±1%	74±1%
RS-B07x2	1.81±0.11	16±1%	69±4%	4.24±0.36	53±4%	89±2%
RS-B12x2	0.89±0.07	22±0%	73±1%	3.30±0.08	51±1%	81±2%
RS-theo3-B07	1.13±0.10	37±2%	91±2%	2.47±0.14	54±3%	103±2%
RS-theo3-B12	1.14±0.13	36±1%	84±1%	2.31±0.09	48±2%	95±2%

4.2.2. Replacement of the aptamer sequence modulates ligand responsiveness and Rnt1p processing

The response curve of the Rnt1p switch can be modulated by three parameters associated with the genetic device: the strength of the ligand-aptamer interaction, the binding affinity of Rnt1p for the hairpin, and the cleavage rate of Rnt1p on the hairpin. The ligand-aptamer interaction will primarily affect the EC50 of the dose response curve, where aptamers with stronger affinities (lower K_D values) will result in decreased EC50 values. However, modifications that result in stronger interactions between the hairpin and Rnt1p will result in increased EC50 values. Finally, modifications to the cleavage rate of Rnt1p on the hairpin will result in changes to the dynamic range or fold-change of the switch based on Michaelis-Menten enzyme kinetics, where increased Rnt1p activity corresponds to an increase in V_{max} resulting in a lowering of the baseline in the absence of theophylline. The two regions of an Rnt1p hairpin that can be experimentally altered to achieve these affects are the CEB (or integrated aptamer sequence) and the Rnt1p binding regions (BSB and IBPB). We first examined the ability to modulate the EC50 value associated with the Rnt1p switch by incorporating aptamers with different ligand affinities. However, because the aptamer sequence is located within the CEB there may be unpredicted effects of integrating alternative aptamer sequences on the Rnt1p processing rates.

We replaced the aptamer component in RS with two alternative theophylline aptamer sequences: Δ-33 (RS-theo2) and a ΔTCT-4 variant (RS-theo3)[22] (Figure 4.2A). The Δ-33 aptamer (K_D = 0.32 μM) exhibits a slightly decreased affinity relative to the ΔTCT-4 aptamer (K_D = 0.29 μM)[22]. However, the effect of the varied aptamer sequences

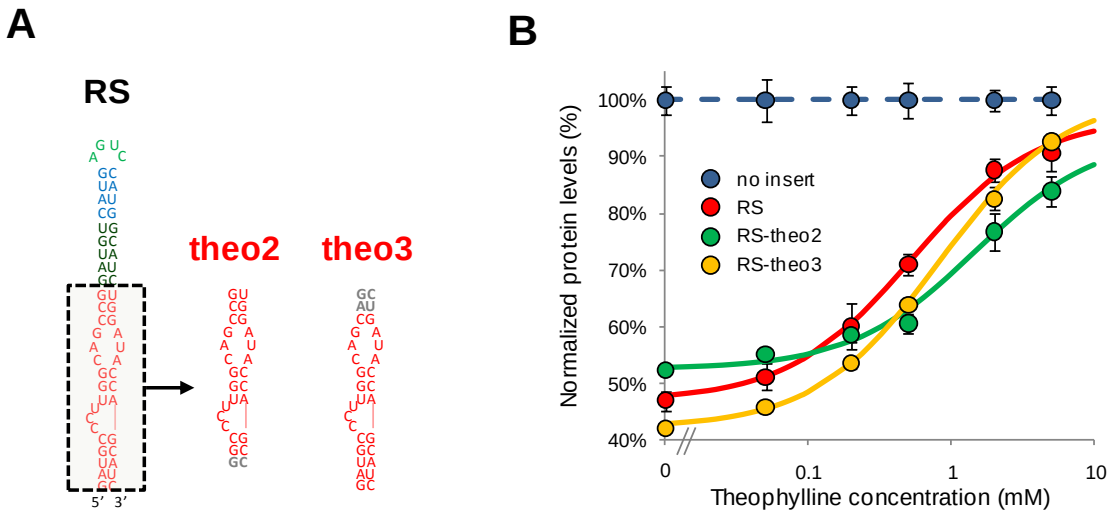


Figure 4.2. Tuning the response curve of the Rnt1p switch through the integration of different theophylline aptamers. (A) Switch and aptamer module sequences illustrating the replacement of the aptamer sequence in RS with different theophylline aptamers: theo2 and theo3. Gray lettering is used to indicate the nucleotides in theo2 and theo3 that differ from theo1. (B) The dose response curves of RS, RS-theo2, and RS-theo3 indicate a shift in EC50 values and variations in the baseline levels. Data are reported as indicated in Figure 4.1C. The model parameters for the curve fit are provided in Table 4.1.

on the switch response curve was unknown, as aptamer integration required removal of terminal loops and different numbers of base-pairs at the base of the switch. The EC50 values of RS-theo2 (1.47 mM) and RS-theo3 (0.83 mM) were greater than that exhibited by RS (0.54 mM), and both switches exhibited changes in the baseline expression levels of the dose response curves (Figure 4.2B, Table 4.1). The increased EC50 values for both switches indicate that the aptamer modifications decreased ligand affinity. The effects of the new aptamer sequences on Rnt1p cleavage varied in an unpredictable manner. Specifically, RS-theo2 exhibited decreased Rnt1p processing and RS-theo3 exhibited increased processing as determined from the baseline levels in the absence of ligand (RS: 47%; RS-theo2: 52%; RS-theo3: 42%). Therefore, RS-theo3 exhibited a larger dynamic range (DR) and fold-change (fold) than RS (RS-theo3: 51% DR, 2.21 fold; RS: 44% DR,

1.93 fold), although its EC50 value was approximately 50% greater than that of RS, whereas RS-theo2 exhibited a decreased dynamic range and fold-change (RS-theo2: 32% DR, 1.60 fold). These results indicate that the sequence and structure of aptamer components, in addition to their binding properties, can impact the regulatory range exhibited by Rnt1p switches.

4.2.3. Incorporation of synthetic BSBs modulates ligand responsiveness and processing of the Rnt1p switch

Modifications to the binding region of the Rnt1p hairpin can also be used to tune the switch response. A set of synthetic BSB modules was previously described (Chapter III), which can be used to modulate Rnt1p binding affinities. However, the synthetic BSB modules were shown to affect Rnt1p binding and processing (Chapter III), such that changes in both the EC50 value and baseline expression level of the dose response curve are expected from their integration into the Rnt1p switch.

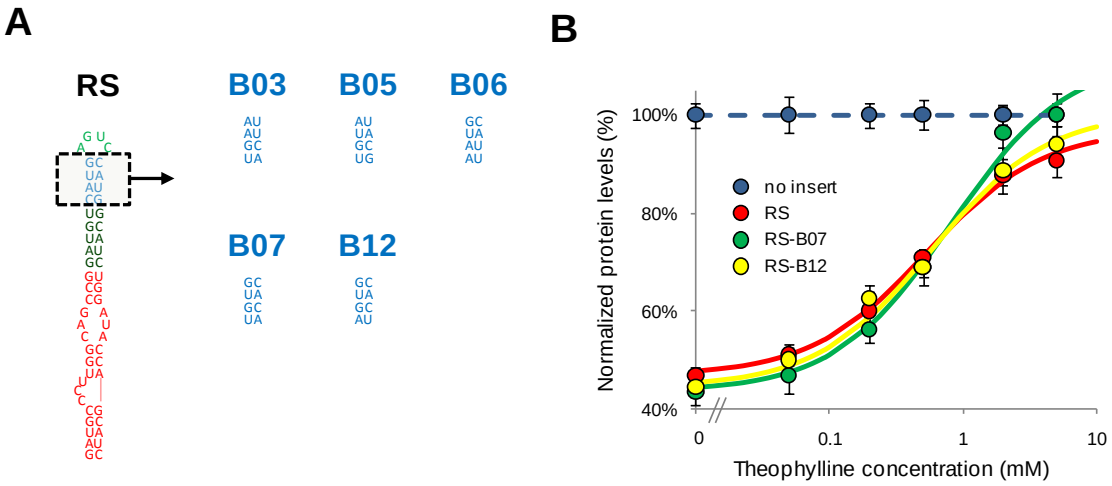


Figure 4.3. Tuning the response curve of the Rnt1p switch through the integration of different synthetic BSB modules. (A) Switch and BSB module sequences illustrating the replacement of the BSB sequence with synthetic BSB modules (B03, B05, B06, B07, and

B12). The gene-regulatory activities of the synthetic BSB modules as previously determined in the context of an Rnt1p hairpin control element are provided in Supplementary Table 4.2. (B) The dose response curves of RS, RS-B07, and RS-B12 indicate that the two synthetic BSB modules decrease baseline levels and exhibit minor effects on the EC50 of the response curve. The dose response curves of the other BSB-incorporated switches are presented in Supplementary Figure 4.2. Data are reported as indicated in Figure 4.1C. The model parameters for the curve fit are provided in Table 4.1.

We incorporated several synthetic BSB modules (B03, B05, B06, B07, B12) into the BSB of RS (Figure 4.3A, Supplementary Table 4.2). We observed a shift in the baseline expression levels for the dose response curves of all BSB-modified Rnt1p switches, where expression levels at 5 mM theophylline were restored to ~90–100% normalized protein levels (Figure 4.3B, Table 4.1, Supplementary Figure 4.2). The observed shifts in the baseline levels with the incorporation of different BSB modules were likely a result of altered processing rates by Rnt1p. Two of the switches, RS-B07 and RS-B12, exhibited decreases in the baseline expression levels (44%) relative to RS (47%), thereby resulting in an increased dynamic range and fold-change (RS-B07: 57% DR, 2.30 fold; RS-B12: 50% DR, 2.11 fold; RS: 44% DR, 1.93 fold). The remainder of the switches exhibited increases in the baseline levels, resulting in slightly reduced or similar dynamic range and fold-change. The majority of the modified BSB Rnt1p switches exhibited slight shifts in the EC50 values (from 0.48 to 0.59 mM). However, RS-B07 exhibited a substantially higher EC50 value of 0.81 mM. The improvement in switch activity for RS-B07 and RS-B12 as determined by the fold-change was due to the increased processing by Rnt1p. However, RS-B12 exhibited greater sensitivity to lower concentrations of theophylline relative to RS-B07 as determined by the EC50 values for these switches (RS-B12: 0.58 mM; RS-B07: 0.81 mM). These results indicate that the

synthetic BSB modules can be used to tune the regulatory range exhibited by Rnt1p switches.

4.2.4. The application of multiple switch modules decreases theophylline responsiveness and increases fold-change

We examined a third tuning strategy for the Rnt1p switch regulatory response based on integrating multiple copies of a switch in the 3' UTR of a target transcript. The integration of multiple copies of *cis*-acting RNA switches in the 3' UTR of target transcripts has previously been demonstrated to modulate the regulatory response by decreasing baseline expression levels[11, 24]. However, it is also expected that this tuning strategy will result in decreased sensitivity to the ligand or an increased EC50 value for a given regulatory system. Based on a mechanism of competition inhibition, we anticipated that the RS switch design would demonstrate close to full restoration of gene activity under maximal ligand concentrations, resulting in significant increases in the fold-change and dynamic range of the system.

We built two-copy (RSx2) and three-copy (RSx3) RS switch constructs (Figure 4.4A). We observed an increase in the EC50 value and a decrease in baseline expression levels with increasing switch copy number (Figure 4.4B, Table 4.1). The effect of multiple Rnt1p switches on the EC50 value of the system was determined to be nearly additive. However, the effect of multiple switches on the baseline level (*b*) was determined to be multiplicative:

$$b_x = (b_1)^x = (47\%)^x$$

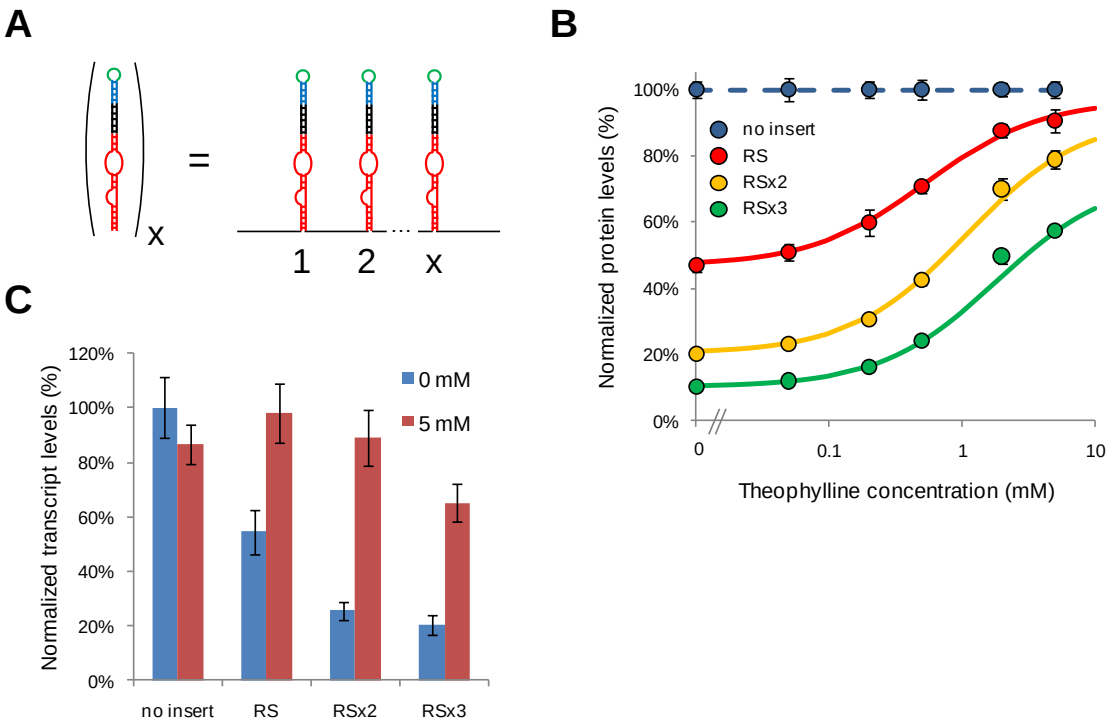


Figure 4.4. Predictive tuning of the Rnt1p switch response curve through the integration of multiple copies of the switch module. (A) Schematic representing the integration of multiple switch modules. Each Rnt1p switch is insulated with spacer sequences as indicated in Figure 4.1A. Two and three copies of the original RS switch were examined for their effects on the regulatory response. (B) The dose response curves of RS, RSx2, and RSx3 indicate increasing dependence on theophylline concentration and decreasing baseline levels with each added switch module. Data are reported as indicated in Figure 4.1C. The model parameters for the curve fit are provided in Table 4.1. (C) The transcript levels of RS, RSx2, and RSx3 support increased Rnt1p processing of the multiple switch modules. Data are reported as indicated in Figure 4.1D.

where *x* is the number of modules. At the highest concentration of theophylline added to the system (5 mM), the gene expression levels decreased with increasing switch number. We expected the ON state expression levels to be multiplicative with the theoretical maximal state (*M*) (Supplementary Table 4.1); while this relationship was observed for RSx2, it was not observed for RSx3. With three copies of RS, the full switch dynamic range cannot be observed when limited to 5 mM theophylline, as evidenced by the lack of saturation of the dose response curve. The fold-change of the system increased with

increasing switch number from 1.93 (RS) to 3.91 (RSx2) to 5.57 (RSx3). We verified that the observed trends in protein expression were consistent with transcript levels, supporting increased Rnt1p processing of the multiple switch modules (Figure 4.4C). Based on the observed trends, we expect that we would likely observe a decrease in regulatory activity when over three copies of the Rnt1p switch are implemented, as further decreases in baseline levels would be limited and the EC50 value would be well above 2 mM theophylline, thus substantially limiting the ON state expression levels that could be accessed. These results indicate that the integration of multiple switch copies can be used to predictably tune the dynamic range and EC50 value exhibited by a given Rnt1p switch.

4.2.5. Combined tuning strategies support the rational design of Rnt1p switch control systems with enhanced regulatory properties

Our studies identified three different strategies for tuning the regulatory response of an Rnt1p switch based on altering the aptamer/CEB, BSB, and number of switch modules. We propose that these tuning modules can be combined to rationally design Rnt1p switches with enhanced regulatory properties. We first examined the combined implementation of synthetic BSB modules with multiple copies of the modified Rnt1p switches. We built genetic control systems that contained two copies of Rnt1p switches incorporating the BSB modules that resulted in the largest single-copy switch dynamic ranges (RS-B07, RS-B12). Both two-copy switch systems (RS-B07x2, RS-B12x2) exhibited decreased baseline expression levels that followed the previously described multiplicative trends (Figure 4.5A, Table 4.1). Although both switches exhibited the

expected increased dependence on theophylline, the EC50 values were not additive as had been observed with RS. Limitations in theophylline concentration had a more substantial impact on the experimentally attainable dynamic range of RS-B07x2 compared to that of RSx2, as RS-B07x2 exhibits an EC50 value that is nearly twice that of RSx2. Even with this limitation, RS-B07x2 exhibits a larger fold-change than that of RSx2. However, the theoretical fold-change (based on the theoretical maximal output of the switch (*M*) from the model fit) of RS-B07x2 increases from 4.24 to 5.42, whereas that of RSx2 only increases to 4.54. RS-B12 experienced the lowest level of gene expression in its ON state (at 5 mM theophylline and *M*) when compared to the other switches incorporating synthetic BSB modules (Table 4.1, Supplementary Table 4.1). Therefore, due to the multiplicative effect associated with multiple switch modules, we expected the ON state for RS-B12x2 to be diminished greater than the other two module switches. Although the baseline was reduced multiplicatively for RS-B12x2 and the theoretical fold-change enhanced, the decrease to the ON state resulted in smaller improvements in the fold-change relative to RSx2 and RS-B07x2 from their respective single module switch.

We next examined the combined implementation of synthetic BSB modules with alternative aptamer modules to improve Rnt1p switch regulatory range. We identified two synthetic BSB modules (B07, B12) and one alternative aptamer module (theo3) that resulted in increases in dynamic range and fold-change when integrated individually into the Rnt1p switch platform. Therefore, we designed two new Rnt1p switches that combined these modules: RS-theo3-B07 and RS-theo3-B12 (Figure 4.5B). The combined

module switches exhibited changes in EC50 values and dynamic ranges that were

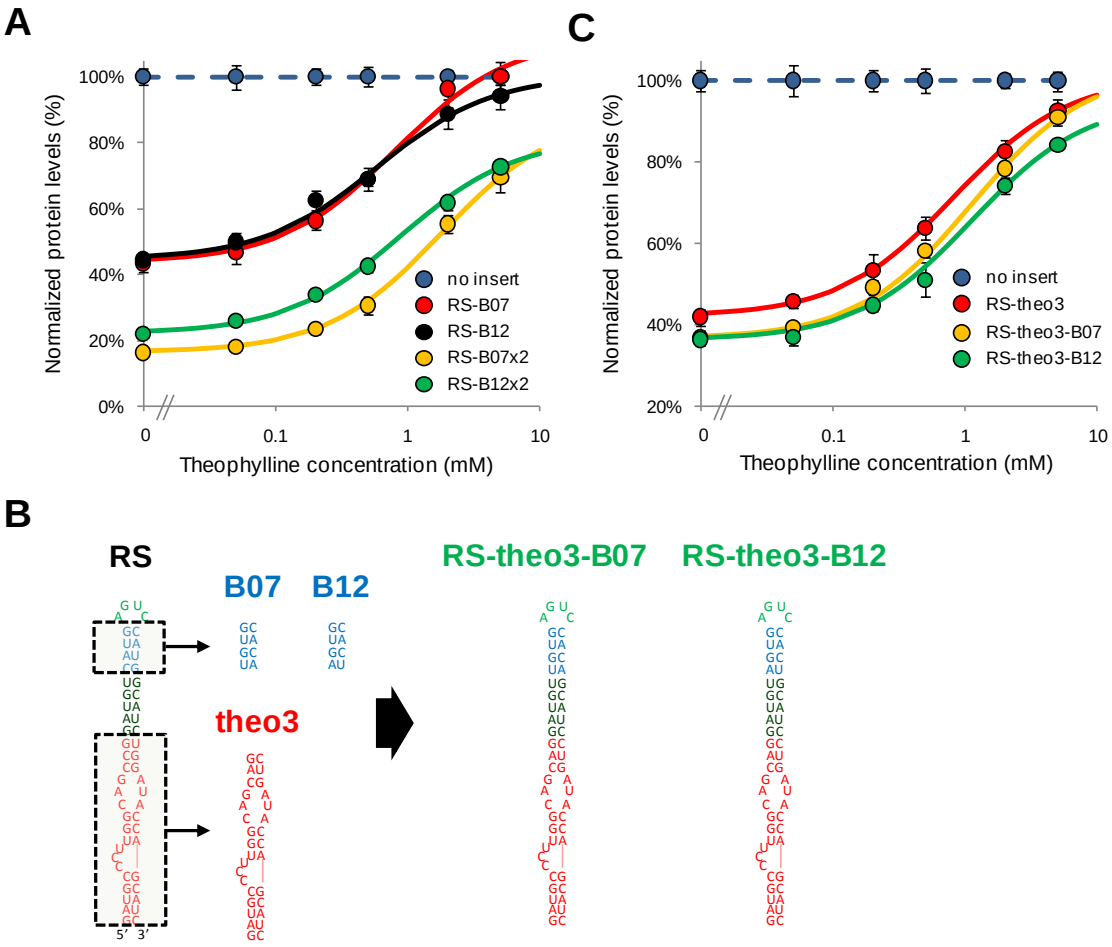


Figure 4.5. Combinatorial implementation of multiple tuning modules results in predictive tuning of the Rnt1p switch regulatory response curve. (A) The dose response curves of switches that incorporate synthetic BSB and multiple switch modules (RS-B07, RS-B12, RS-B07x2, and RS-B12x2) indicate increased dependence on theophylline concentration and decreased baseline levels based on the combined activity of the two incorporated tuning modules. The model parameters for the curve fit are provided in Table 4.1. (B) Optimization of single-copy switch designs by incorporating synthetic BSB modules (B07 and B12) with the aptamer/CEB module (theo3) that exhibited the most improved switch response curves. (C) The dose response curves of RS3-theo3, RS-theo3-B07, and RS-theo3-B12 indicate that integration of optimized BSB and aptamer/CEB modules results in Rnt1p switches exhibiting improved dynamic range and fold-change over that observed with the individual modules. Data are reported as indicated in Figure 4.1C. The model parameters for the curve fit are provided in Table 4.1.

improved from the effects of the individual tuning modules (Figure 4.5C, Table 4.1). Specifically, RS-theo3-B07 and RS-theo3-B12 exhibited increased EC50 values (1.13 mM and 1.14 mM, respectively) from those exhibited by RS-theo3 (0.83 mM), RS-B07 (0.81 mM), and RS-B12 (0.58 mM). In addition, both RS-theo3-B07 and RS-theo3-B12 exhibit baseline levels (37% and 36%, respectively) that are lower than the switches harboring the individual tuning modules (RS-theo3: 42%; RS-B07 and RS-B12: 44%) (Figure 4.5C). The rationally tuned Rnt1p switches, RS-theo3-B07 and RS-theo3-B12, exhibit the greatest fold-change of the single-copy switches generated in this study (RS-theo3-B07: 2.47; RS-theo3-B12: 2.31). The results demonstrate that the rational combination of independent tuning modules can be used to improve switch regulatory ranges in a manner that is predictable from the individual module effects.

4.3. Discussion

We have developed a new class of synthetic RNA devices based on ligand-responsive modulation of RNase III cleavage activity in *S. cerevisiae*. Our design strategy incorporated a sensor component, encoded within an aptamer sequence, directly within a region of Rnt1p substrates known to effect enzyme cleavage – the CEB. This direct integration design strategy differs from commonly implemented design strategies for synthetic riboswitches that couple the aptamer component to the regulatory component through a linker sequence[5, 25-26]. As such the Rnt1p switch does not undergo substantial secondary structure rearrangements between gene-regulatory active and inactive states as is observed with many RNA switch platforms, but instead functions through inhibiting Rnt1p processing as a result of the slight structural changes or sequestering of CEB

nucleotides associated with ligand binding at the aptamer/CEB domain. A similar direct integration design strategy was recently described for a synthetic RNA switch platform based on modulating Drosha (an RNase III enzyme in the RNAi pathway) processing of its substrates by integration of aptamer sequences into the basal segment region of pri-miRNA hairpins[19]. In addition, a direct coupling strategy was previously described for the design of an RNA switch that modulated Dicer (another RNase III enzyme in the RNAi pathway) processing of shRNA substrates by coupling aptamer sequences to the terminal loop of the shRNA hairpin[17]. These examples indicate that the direct integration of sensor and actuator components may present an effective design strategy for riboswitches based on the modulation of RNase III processing activities.

We had previously demonstrated that the CEB region has no rigid structural or sequence requirements for efficient Rnt1p cleavage (Chapter II). Therefore, the described Rnt1p switch platform is likely amenable to diverse aptamer sequences, supporting the ability to tailor Rnt1p switches to diverse molecular effectors. However, our earlier work also demonstrated that different nucleotide sequences in the CEB resulted in altered processing rates by Rnt1p *in vivo* (Chapter II). Therefore, it is also expected that changing the aptamer sequence within the CEB will result in unpredictable changes to the baseline levels exhibited by the switch due to effects of the aptamer sequence on Rnt1p cleavage activity. This was observed when integrating different theophylline aptamer sequences into the Rnt1p switch platform (Figure 4.2B, Table 4.1). Ultimately, the activity of an Rnt1p switch will be determined by two factors: 1) the ability of the aptamer sequence and structure to be effectively processed by Rnt1p; and 2) the ability of the ligand to compete with Rnt1p binding and cleavage or decrease Rnt1p's ability to

cleave the ligand-bound CEB. For example, in our initial studies we examined different placements of the theophylline aptamer in the Rnt1p hairpin stem and observed several designs that were processed efficiently by Rnt1p, but where Rnt1p activity was not inhibited under ligand addition (data not shown). It is likely that the sequences and structures of certain aptamers will inhibit Rnt1p cleavage activity when placed in the CEB. Therefore, the successful design of Rnt1p switches responsive to new effector molecules will likely require screening of potential aptamer sequences within the switch platform for those sequences and structures that can be effectively cleaved by Rnt1p. In addition, newer *in vivo* screening methods will likely be effective in identifying the best aptamer candidates from *in vitro* enriched aptamer pools[27].

Optimization of the Rnt1p switch regulatory response was significantly aided by earlier foundational work conducted in our laboratory on developing synthetic CEB and BSB modules that can be implemented individually or in combination to rationally build tailored Rnt1p control elements (Chapters II and III). We demonstrated that the synthetic BSB modules can be integrated into an Rnt1p switch to tune switch dynamics. Using this approach we identified two synthetic BSB modules (B07 and B12) that increase the dynamic range and fold-change of the switch response. However, it was also observed that the rank order activities of the BSB modules as previously characterized in the Rnt1p hairpin substrate were different than that observed in the Rnt1p switch hairpin (RS) as determined by the baseline level in the absence of theophylline. This difference is likely due to the fact that the intervening stem between the CEB and BSB modules are different in the original Rnt1p hairpin substrate and the hairpin used for the Rnt1p switch. In addition, the aptamer sequence integrated into the CEB region also extends beyond the

CEB module, introducing additional changes to that stem region. Therefore, the data indicate that the relative activities of the BSB elements may be affected by the stem sequence between the CEB and BSB modules.

We also demonstrated that the different tuning modules can be implemented combinatorially to predictably tailor the switch response curve based on the individual activities of each module. For example, by integrating the BSB modules (B07 and B12) with the aptamer/CEB module (theo3) that most improved the switch response curve into a single Rnt1p switch, we were able to build single-copy switches exhibiting a greater dynamic range, fold-change, and lowered baseline expression level than any of the switches harboring one of these modules (Figure 4.5C, Table 4.1). However, as one drawback of this tuning strategy, modules that improved the switching activity of the response curve generally also resulted in increased EC50 values. This trade-off between switching activity and sensitivity highlights the requirement for developing higher affinity aptamers or aptamers with appropriate affinities for the intended application [28].

A third tuning strategy was examined to improve switching activity based on the implementation of multiple copies of Rnt1p switches in a single target transcript. Two-copy switch systems were examined for the original Rnt1p switch (RS) and two BSB-tuned switches that exhibited improved regulatory responses (RS-B07 and RS-B17). In general, the implementation of a second copy of a given switch resulted in decreased baseline levels, increased fold-change, and increased EC50 values (Figure 4.4B, Figure 4.5A, Table 4.1), where the observed effects on baseline levels were multiplicative. However, the observed dynamic range (difference between the ON and the OFF states) did not increase with increasing switch number, due to the response curves not saturating

at 5 mM theophylline. In addition, the maximal ON state (*M*), determined by extrapolating the data with the binding model curve fit, indicated that the theoretical dynamic range only slightly increased for RS-B07x2 and RS-B17x2 (Supplementary Table 4.1). This limited improvement in theoretical dynamic range was due to substantial decreases in *M* with increasing switch number, potentially due to nonspecific effects resulting from the additional structures such as interference with translation or transcript destabilization or due to increased residual cleavage by Rnt1p of the substrate in the ligand-bound form. We also built a three-copy version of RS (RSx3), which exhibited additional increases in the fold-change over the two-copy system. However, the implementation of three Rnt1p switch modules exhibited limiting returns in the decreased baseline levels and increased the EC50 of the response curve to a level approaching that which would not allow proper observation of switch activity due to limitations in theophylline concentrations that can be applied to the yeast culture.

This work extends the utility of Rnt1p control elements for cellular engineering applications by developing an integrated RNA platform for building ligand-responsive Rnt1p-based control devices in yeast. By taking a synthetic biology approach to the development of a tool for designing precise genetic control elements based on building Rnt1p controllers from synthetic modules (Chapters II and III), we were able to identify readily implementable tuning and optimization strategies for the more complex Rnt1p-based regulatory device. Specifically, we demonstrated that three tuning strategies, based on the implementation of synthetic aptamer/CEB, BSB, and switch modules independently or in combination, can be used to optimize the regulatory response of the engineered control system. Importantly, the effects of the tuning modules when

implemented combinatorially were predictable from their individual activities. An interesting aspect of this engineered riboswitch platform is that our single-copy switches generally exhibit fully restored gene expression activity at maximum ligand concentration compared to constructs lacking a switch, such that the observed differences between switch activities are largely due to changes in baseline levels. The integration of compatible aptamers into this switch platform that respond to broad classes of molecules of interest, including nontoxic exogenous chemicals, primary metabolites, and chemicals of industrial interest, will advance the engineering of genetic circuits in yeast for diverse biotechnological applications.

4.4 Materials and Methods

4.4.1. Plasmid construction

Standard molecular biology techniques were utilized to construct all plasmids[29]. DNA synthesis was performed by Integrated DNA Technologies (Coralville, IA) or the Protein and Nucleic Acid Facility (Stanford, CA). All enzymes, including restriction enzymes and ligases, were obtained through New England Biolabs (Ipswich, MA) unless otherwise noted. Pfu polymerase was obtained through Stratagene. Ligation products were electroporated with a GenePulser XCell (Bio-Rad, Hercules, CA) into *Escherichia coli* DH10B (Invitrogen, Carlsbad, CA), where cells harboring cloned plasmids were maintained in Luria-Bertani media containing 50 mg/ml ampicillin (EMD Chemicals). Clones were initially verified through colony PCR and restriction mapping. All cloned constructs were sequence verified by Laragen (Los Angeles, CA) or Elim Biopharmaceuticals (Hayward, CA).

The construction of the Rnt1p characterization plasmid, pCS321, and the Rnt1p expression plasmid, pRNT1, have been previously described (Chapter II). The plasmid map for pCS321 is available in Supplementary Figure 4.3. Insertion of engineered Rnt1p substrates and appropriate controls into the 3' UTR of *yEGFP3* in pCS321 was performed through either digestion with appropriate restriction endonucleases and ligation-mediated cloning or homologous recombination-mediated gap-repair during transformation into *S. cerevisiae* strain W303 (*MATa, his3-11,15 trp1-1 leu2-3 ura3-1 ade2-1*) through standard lithium acetate procedures[30]. The single module Rnt1p switch (RS), its mutant tetraloop control, the BSB variants, and the theophylline aptamer variant RS-theo3 with its own BSB variants were amplified for insertion with both techniques using the forward and reverse primers RS_fwd (5' ATGGTATGGATGAATTGTACAA ATAAAGAGCCTAGGAAACAAACAAACTTGATGCCCTTGG) and RS_rev (5' AA ATTCGCTTATTTAGAAGTGGCGCGCCCTCTCGAGTTTTTATTTTTCTTTTTACG ATGCTGGTATC), respectively. Two of the theophylline aptamer variants required individually designed primers: RSnt_fwd (5' ATGGTATGGATGAATTGTACAAATA AAGAGCCTAGGAAACAAACAAACTTGATGCCATTGG) and RS_rev for RSnt; and RS-theo2_fwd (5' ATGGTATGGATGAATTGTACAAATAAAGAGCCTAGGAAACA AACAAACTTGGCCCTTGGCA) and RS-theo2_rev (5' AAATTCGCTTATTTAGAAG TGGCGCGCCCTCTCGAGTTTTTATTTTTCTTTTTACGGCTGGTATCCA) for RS-theo2. A second switch module was amplified for insertion using the forward and reverse primers RSx2_fwd (5' GTGCTCGAGAAACAAACAAACTTGATGCCCTTGGCA) and RSx2_rev (5' CAGCTCGAGTTTTTATTTTTCTTTTTACGATGCTGGTATCCAGA TG). A third switch module was amplified for insertion using the forward and reverse

primers RSx3_fwd (5’ GTGCCTAGGAAACAAACAAACTTGATGCCCTTGGCA) and RSx3_rev (5’ CAGCCTAGGTTTTTATTTTTCTTTTTACGATGCTGGTATCCAGA TG). In the case of digestion and ligation, the PCR products were digested with the unique restriction sites AvrII and/or XhoI, which are located 3 nts downstream of the *yEGFP3* stop codon and upstream of the ADH1 terminator. Following construction and sequence verification of the desired vectors, 100–500 ng of each plasmid was transformed into strain W303. In the case of gap-repair, 250–500 ng of the PCR product and 100 ng of plasmid digested with AvrII and XhoI were transformed into the yeast strain. All yeast strains harboring cloned plasmids were maintained on synthetic complete media with an uracil dropout solution and 2% dextrose at 30°C. A table of primers and template sequences used for each switch module is provided in Supplementary Table 4.3.

4.4.2. Rnt1p substrate characterization assays

S. cerevisiae cells harboring pCS321-based plasmids were grown on synthetic complete (SC) media with an uracil dropout solution and the appropriate sugars (2% raffinose, 1% sucrose) overnight at 30ºC. The cells were back-diluted the following morning into a 4.3-ml combination of SC media and 25 mM theophylline dissolved directly into the same SC media to an optical density at 600 nm (OD_{600}) of 0.1 (~200 µl) and grown again at 30°C. The theophylline solution and SC media were mixed to achieve the appropriate concentration of theophylline (0 mM to 5 mM) in the test tube at 5 ml. After 1 hr, the test tube volume was brought up the final volume of 5 ml by adding 0.5 ml of 20% galactose (2% final concentration) for induction or water (non-induced control) to

the cell cultures. The cells were grown for another 4.5 hr before measuring the fluorescence levels.

4.4.3. Fluorescence quantification

On the Quanta flow cytometer (Beckman Coulter, Fullerton, CA), the distribution of GFP fluorescence was measured with the following settings: 488-nm laser line, 525-nm bandpass filter, and photomultiplier tube setting of 5.83 (pCS321-based) or 4.50 (pCS1585-based). Data were collected under low flow rates until 5,000 viable cell counts were collected. A non-induced cell population was used to set a gate to represent GFP-negative and GFP-positive populations. Final data values are reported as the average of the median GFP-positive fluorescence from three independently-grown samples and normalized to a construct baring no module inserts ('no insert') at the same theophylline concentration. Standard error is determined from the standard deviation of the triplicate samples.

4.4.4. Quantification of cellular transcript levels

Total RNA from *S. cerevisiae* was collected by a standard hot acid phenol extraction method[31] and followed by DNase I (New England Biolabs) treatment to remove residual plasmid DNA according to manufacturer's instructions. cDNA was synthesized from 5 μg of total RNA with gene-specific primers for *yEGFP3* and *ACT1* [32] (rnt1p_rtpcr_rev2 and ACT1_rtpcr_rev, respectively) and SuperScript III Reverse Transcriptase (Invitrogen) according to manufacturer's instructions. The forward and reverse primers for *yEGFP3* quantification are rnt1p_rtpcr_fwd2 (5' CGGTGAAGGTGA

AGGTGATGCTACT) and rnt1p_rtpcr_rev2 (5’ GCTCTGGTCTTGTAGTTACCGTCA TCTTTG), respectively. The forward and reverse primers for *ACT1* quantification are ACT1_rtpcr_fwd (5’ GGCATCATACCTTCTACAACGAAT) and ACT1_rtpcr_rev (5’ GGAATCCAAAACAATACCAGTAGTTCTA), respectively. Relative transcript levels were quantified in triplicate from three identical reactions from the cDNA samples by using an appropriate primer set and iQ SYBR Green Supermix (Bio-Rad, Hercules, CA) on an iCycler iQ quantitative real-time PCR (qRT-PCR) machine (Bio-Rad) according to the manufacturer’s instructions. For each run, a standard curve was generated for *yEGFP3* and a house-keeping gene, *ACT1*, using a dilution series for a control representing no insertion of an Rnt1p substrate. Relative *yEGFP3* and *ACT1* levels were first individually determined for each sample and then the *yEGFP3* values were normalized by their corresponding *ACT1* values.

4.4.5. In vitro *transcription of Rnt1p substrates*

Rnt1p switches were PCR-amplified to include an upstream T7 promoter site using forward and reverse primers Rnt1p_col-T7_fwd_prmr (5’ TTCTAATACGACTCA CTATAGGATGGTATGGATGAATTGTACAAATAAAGCCTA) and Rnt1p_col-T7_ rev_prmr (5’ AAATTCGCTTATTTAGAAGTGGCGC), respectively. 500 ng of PCR product was transcribed with T7 RNA Polymerase (New England Biolabs) in the presence and absence of α-P^{32}-GTP. The 25-µl reaction consisted of the following components: 1x RNA Pol Reaction Buffer (New England Biolabs), 3 mM rATP, 3 mM rCTP, 3 mM rUTP, 0.3 mM rGFP, 1 µl RNaseOUT (Invitrogen), 10 mM $MgCl_2$, 2 mM DTT, 1 µl T7 Polymerase, and 0.5 µCi α-P^{32}-GTP. Unincorporated nucleotides were

removed from the reactions by running the samples through NucAway Spin Columns (Ambion, Austin, TX) according to the manufacturer's instructions.

4.4.6. Rnt1p expression and purification

The pRNT1 plasmid was transformed into *E. coli* strain BL21 using the Z-competent *E. coli* Transformation Kit and Buffer Set (Zymo Research, Orange, CA) according to manufacturer's instructions. Rnt1p was collected as a protein extract as previously described[33]. Briefly, an overnight culture of BL21 cells harboring pRNT1 was back-diluted to an OD_{600} of 0.5. Once the culture reached an OD_{600} of 1.1–1.4, it was induced with 1 mM IPTG and grown for an additional 3 hr. The cells were centrifuged at 2,500g for 12 min at 4°C and the resulting cell pellet was frozen in a –80°C freezer. After weighing the frozen cell pellet, the cells were resuspended in 4 ml Ni_{2+} buffer [25% (v/v) glycerol, 1 M NaCl, 30 mM Tris pH 8.0] per gram of harvested cells. The resuspension was sonicated (Heat Systems-Ultrasonics, Inc.) twice with the following settings: 2 x 30 sec, output control 5, and 50% duty cycle. Cellular debris was removed by centrifugation at 20,000g for 30 min at 4°C and the supernatant was filtered through a 0.2-μm pore size Acrodisc 25-mm syringe filter (Pall Life Sciences, Ann Arbor, MI).

Rnt1p was purified from the resulting supernatant with one 1-ml HisTrap HP column (GE Healthcare) on an AKTA FPLC machine (GE Healthcare). Elution of the protein was performed with an imidazole concentration of 150 mM in Ni_{2+} buffer and the protein was collected in 6 1-ml fractions. Protein purification was confirmed by analyzing an aliquot of each fraction on a SDS-PAGE gel (NuPAGE 4-12% Bis-Tris Gel, Invitrogen) and protein function was confirmed by incubating an aliquot of each fraction

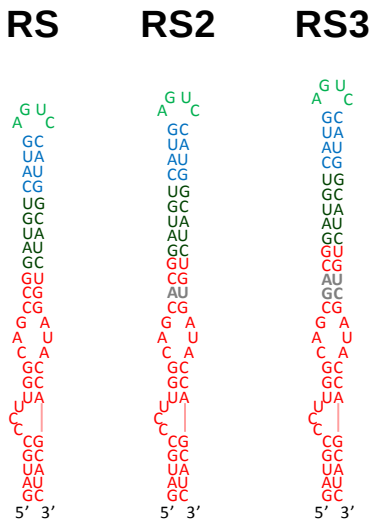
with a control Rnt1p substrate and analyzing the resulting cleavage products on an 8% denaturing polyacrylamide gel. Positive fractions were pooled and concentrated to less than a 3-ml volume using a Centricon Centrifugal Filter Device (10,000 MWCO; Millipore) according to the manufacturer's instructions. The concentrated protein was then injected into a Slide-A-Lyzer Dialysis Cassette (10,000 MWCO; Pierce Biotechnology) and buffer-exchanged twice with Rnt1p Storage Buffer [50% (v/v) glycerol, 0.5 M KCl, 30 mM Tris pH 8.0, 0.1 M DTT, 0.1 M EDTA] at 4ºC. The first buffer exchange took place for 4 hr and the second buffer exchange occurred overnight. The purified Rnt1p was stored in aliquots at –20°C.

4.4.7. In vitro *Rnt1p substrate cleavage assay*

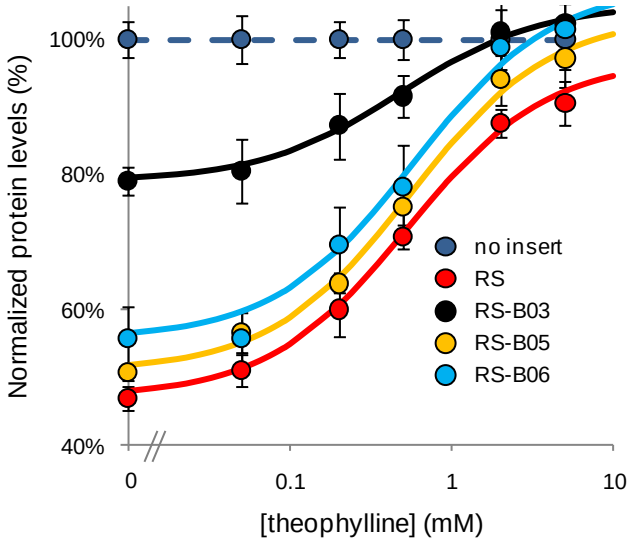
Cleavage assays were performed on Rnt1p substrates as previously described[33-34]. Briefly, a 10-µl mixture of RNA and Rnt1p were incubated at 30ºC for 30 min in Rnt1p reaction buffer [30 mM Tris (pH 7.5), 150 mM KCl, 5 mM spermidine, 20 mM $MgCl_2$, 0.1 mM DTT, and 0.1 mM EDTA (pH 7.5)] in the presence or absence of 10 mM theophylline. The RNA concentration was 0.05 µM and the Rnt1p concentration was 20.7 µM. The cleavage reaction products were separated on an 8% denaturing polyacrylamide gel run at 35 W for 30 min. Gels were transferred to filter paper and analyzed for relative substrate and product levels through phosphorimaging analysis on a FX Molecular Imager (Bio-Rad).

4.5. Supplementary Information

Supplementary Figures and Tables

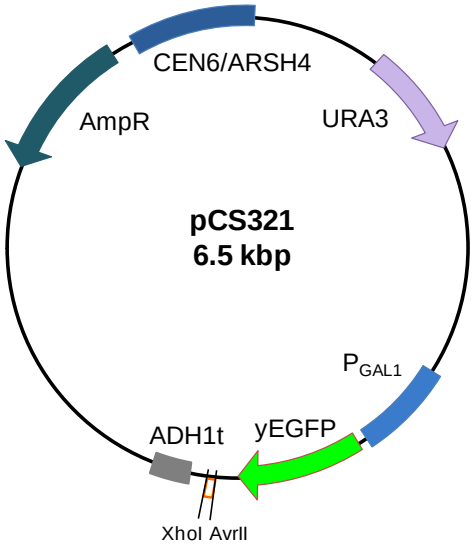


Supplementary Figure 4.1. Sequences illustrating the placement of the ΔTCT-4 aptamer within R31L-3B4Inv at multiple locations. RS2 and RS3 resulted in nonfunctional switches that were unresponsive to theophylline (data not shown), while RS was functional and utilized as the base Rnt1p switch design in this study. Gray lettering is used to indicate the nucleotides in RS2 and RS3 that differ from RS.



Supplementary Figure 4.2. The dose response curves of RS, RS-B03, RS-B05, and RS-B06 indicate that these synthetic BSB modules increase baseline expression relative to

the original Rnt1p switch (RS). These switches exhibit a reduced fold-change relative to that exhibited by RS. Data are reported as indicated in Figure 4.1C. The model parameters for the curve fit are provided in Table 4.1.



Supplementary Figure 4.3. Plasmid map of pCS321, the Rnt1p hairpin characterization plasmid.

Supplementary Table 4.1. The theoretical fold-change and dynamic range of all Rnt1p switches examined in this study as determined from experimentally measured baseline expression at 0 mM theophylline (*b*) and the theoretical maximal output (*M*) calculated by fitting the dose response data to the binding model.

switch	b=Y(0mM)	M	theorectical fold-change	theorectical dynamic range
SR	47% ± 2%	97% ± 1%	2.07 ± 0.08	50% ± 2%
SR-theo2	52% ± 2%	94% ± 3%	1.79 ± 0.08	41% ± 3%
SR-theo3	42% ± 2%	101% ± 1%	2.40 ± 0.11	59% ± 2%
SR-B03	79% ± 4%	105% ± 1%	1.33 ± 0.07	26% ± 4%
SR-B05	51% ± 1%	104% ± 2%	2.04 ± 0.06	53% ± 2%
SR-B06	56% ± 5%	108% ± 3%	1.94 ± 0.17	52% ± 6%
SR-B07	44% ± 2%	112% ± 4%	2.56 ± 0.12	68% ± 4%
SR-B12	44% ± 2%	101% ± 2%	2.26 ± 0.11	56% ± 3%
SRx2	20% ± 1%	92% ± 0%	4.54 ± 0.16	71% ± 1%
SRx3	10% ± 0%	74% ± 1%	7.18 ± 0.32	64% ± 1%
SR-B07x2	16% ± 1%	89% ± 2%	5.42 ± 0.33	72% ± 2%
SR-B12x2	22% ± 0%	81% ± 2%	3.69 ± 0.10	59% ± 2%
SR-theo3-B07	37% ± 2%	103% ± 2%	2.79 ± 0.15	66% ± 3%
SR-theo3-B12	36% ± 1%	95% ± 2%	2.61 ± 0.11	59% ± 2%

Supplementary Table 4.2. The previously reported gene-regulatory activity of the synthetic BSB modules selected for use in this study in the context of the Rnt1p hairpin genetic control element (A02). The data is taken from previous work (Chapter III).

substrate	Normalized protein levels (%)	Normalized transcript levels (%)
A02-B00	28% ± 1%	43% ± 8%
A02-B03	50% ± 2%	53% ± 5%
A02-B05	25% ± 0%	39% ± 3%
A02-B06	27% ± 2%	57% ± 2%
A02-B07	37% ± 3%	51% ± 3%
A02-B12	27% ± 2%	47% ± 5%

Supplementary Table 4.3. Oligonucleotide template sequences for all switches built in this study. The sequences of the indicated primers are provided in the Materials and Methods section.

switch	forward primer	reverse primer	template
SR	SR_fwd	SR_rev	AAACAAACTTGATGCCCTTGGCAGCCGG ATGTCATGAGTCCATGGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA
SRN	SR_fwd	SR_rev	AAACAAACTTGATGCCCTTGGCAGCCGG ATGTCATGCATCCATGGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA
SRnt	SRnt_fwd	SR_rev	AAACAAACTTGATGCCATTGGCAGCCGG ATGTCATGAGTCCATGGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA
SR-theo2	SR-theo2_fwd	SR-theo2_rev	AAACAAACTTGGCCCTTGGCAGCCGGAT GTCATGAGTCCATGGCATCTGGATACCA GCCGTAAAAAGAAAAATAAA
SR-theo3	SR_fwd	SR_rev	AAACAAACTTGATGCCCTTGGCAGCACG ATGTCATGAGTCCATGGCATCGTGATAC CAGCATCGTAAAAAGAAAAATAAA
SR-B03	SR_fwd	SR_rev	AAACAAACTTGATGCCCTTGGCAGCCGG ATGTTGAAAGTCTTCAGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA
SR-B05	SR_fwd	SR_rev	AAACAAACTTGATGCCCTTGGCAGCCGG ATGTTGTAAGTCTACGGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA
SR-B06	SR_fwd	SR_rev	AAACAAACTTGATGCCCTTGGCAGCCGG ATGTAATGAGTCCATTGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA
SR-B07	SR_fwd	SR_rev	AAACAAACTTGATGCCCTTGGCAGCCGG ATGTTGTGAGTCCACAGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA
SR-B12	SR_fwd	SR_rev	AAACAAACTTGATGCCCTTGGCAGCCGG ATGTAGTGAGTCCACTGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA
SRx2	SRx2_fwd	SRx2_rev	AAACAAACTTGATGCCCTTGGCAGCCGG ATGTCATGAGTCCATGGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA

SRx3	SRx3_fwd	SRx3_rev	AAACAAACTTGATGCCCTTGGCAGCCGG ATGTCATGAGTCCATGGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA
SR-B07x2	SRx2_fwd	SRx2_rev	AAACAAACTTGATGCCCTTGGCAGCCGG ATGTTGTGAGTCCACAGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA
SR-B12x2	SRx2_fwd	SRx2_rev	AAACAAACTTGATGCCCTTGGCAGCCGG ATGTAGTGAGTCCACTGCATCTGGATAC CAGCATCGTAAAAAGAAAAATAAA
SR-theo3-B07	SR_fwd	SR_rev	AAACAAACTTGATGCCCTTGGCAGCACG ATGTTGTGAGTCCACAGCATCGTGATAC CAGCATCGTAAAAAGAAAAATAAA
SR-theo3-B12	SR_fwd	SR_rev	AAACAAACTTGATGCCCTTGGCAGCACG ATGTAGTGAGTCCACTGCATCGTGATAC CAGCATCGTAAAAAGAAAAATAAA

Acknowledgements

We thank K. Hoff for assistance in the expression and purification of Rnt1p; and S. Bastian and F.H. Arnold for assistance in sonication and FPLC. This work was supported by the National Science Foundation (CAREER award CBET-0917705 to C.D.S.); and the Alfred P. Sloan Foundation, fellowship (to C.D.S.).

References

1. Andrianantoandro, E., Basu, S., Karig, D.K. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Mol Syst Biol* **2**, 2006 0028 (2006).
2. Basu, S., Mehreja, R., Thiberge, S., Chen, M.T. & Weiss, R. Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A* **101**, 6355-6360 (2004).
3. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338 (2000).
4. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339-342 (2000).

5. Win, M.N., Liang, J.C. & Smolke, C.D. Frameworks for programming biological function through RNA parts and devices. *Chem Biol* **16**, 298-310 (2009).

6. Grate, D. & Wilson, C. Inducible regulation of the S. cerevisiae cell cycle mediated by an RNA aptamer-ligand complex. *Bioorg Med Chem* **9**, 2565-2570 (2001).

7. Bayer, T.S. & Smolke, C.D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* **23**, 337-343 (2005).

8. Desai, S.K. & Gallivan, J.P. Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J Am Chem Soc* **126**, 13247-13254 (2004).

9. Suess, B., Fink, B., Berens, C., Stentz, R. & Hillen, W. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res* **32**, 1610-1614 (2004).

10. Sharma, V., Nomura, Y. & Yokobayashi, Y. Engineering complex riboswitch regulation by dual genetic selection. *J Am Chem Soc* **130**, 16310-16315 (2008).

11. Chen, Y.Y., Jensen, M.C. & Smolke, C.D. Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc Natl Acad Sci U S A* **107**, 8531-8536 (2010).

12. Win, M.N. & Smolke, C.D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A* **104**, 14283-14288 (2007).

13. Buskirk, A.R., Landrigan, A. & Liu, D.R. Engineering a ligand-dependent RNA transcriptional activator. *Chem Biol* **11**, 1157-1163 (2004).

14. Thompson, K.M., Syrett, H.A., Knudsen, S.M. & Ellington, A.D. Group I aptazymes as genetic regulatory switches. *BMC Biotechnol* **2**, 21 (2002).

15. Weigand, J.E. & Suess, B. Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. *Nucleic Acids Res* **35**, 4179-4185 (2007).

16. Kim, D.S., Gusti, V., Dery, K.J. & Gaur, R.K. Ligand-induced sequestering of branchpoint sequence allows conditional control of splicing. *BMC Mol Biol* **9**, 23 (2008).

17. An, C.I., Trinh, V.B. & Yokobayashi, Y. Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. *RNA* **12**, 710-716 (2006).

18. Beisel, C.L., Bayer, T.S., Hoff, K.G. & Smolke, C.D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol Syst Biol* **4**, 224 (2008).

19. Beisel, C.L., Chen, Y.Y., Culler, S.J., Hoff, K.G. & Smolke, C.D. Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. *Nucleic Acids Res* (2010).

20. Lamontagne, B. et al. Sequence dependence of substrate recognition and cleavage by yeast RNase III. *J Mol Biol* **327**, 985-1000 (2003).

21. Lamontagne, B., Hannoush, R.N., Damha, M.J. & Abou Elela, S. Molecular requirements for duplex recognition and cleavage by eukaryotic RNase III: discovery of an RNA-dependent DNA cleavage activity of yeast Rnt1p. *J Mol Biol* **338**, 401-418 (2004).

22. Zimmermann, G.R., Wick, C.L., Shields, T.P., Jenison, R.D. & Pardi, A. Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *RNA* **6**, 659-667 (2000).

23. Koch, A.L. The metabolism of methylpurines by Escherichia coli. I. Tracer studies. *J Biol Chem* **219**, 181-188 (1956).

24. Win, M.N. & Smolke, C.D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456-460 (2008).

25. Isaacs, F.J., Dwyer, D.J. & Collins, J.J. RNA synthetic biology. *Nat Biotechnol* **24**, 545-554 (2006).

26. Suess, B. & Weigand, J.E. Engineered riboswitches: overview, problems and trends. *RNA Biol* **5**, 24-29 (2008).

27. Sinha, J., Reyes, S.J. & Gallivan, J.P. Reprogramming bacteria to seek and destroy an herbicide. *Nat Chem Biol* **6**, 464-470 (2010).

28. Beisel, C.L. & Smolke, C.D. Design principles for riboswitch function. *PLoS Comput Biol* **5**, e1000363 (2009).

29. Sambrook, J. & Russell, D.W. Molecular Cloning: A Laboratory Manual, Edn. 3rd. (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY; 2001).

30. Gietz, R. & Woods, R. in Guide to Yeast Genetics and Molecular and Cell Biology, Part B, Vol. 350. (eds. C. Guthrie & G. Fink) 87-96 (Academic, San Diego, CA; 2002).

31. Caponigro, G., Muhlrad, D. & Parker, R. A small segment of the MAT alpha 1 transcript promotes mRNA decay in Saccharomyces cerevisiae: a stimulatory role for rare codons. *Mol Cell Biol* **13**, 5141-5148 (1993).

32. Ng, R. & Abelson, J. Isolation and sequence of the gene for actin in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **77**, 3912-3916 (1980).

33. Lamontagne, B. & Elela, S.A. Purification and characterization of Saccharomyces cerevisiae Rnt1p nuclease. *Methods Enzymol* **342**, 159-167 (2001).

34. Lamontagne, B. & Elela, S.A. Evaluation of the RNA determinants for bacterial and yeast RNase III binding and cleavage. *J Biol Chem* **279**, 2231-2241 (2004).

Chapter V: Synthetic IRESes promoting translation under normal physiological conditions in *S. cerevisiae*

Abstract

We examined the ability to fine-tune gene expression levels with control elements that impact translation initiation by the ribosomal complex to facilitate the construction of multicistronic vectors in yeast. Internal ribosome entry sites (IRESes) act independent of cap-based translation initiation and function through direct association with the 18S ribosomal RNA (rRNA). This mechanism of translation initiation is analogous to Shine-Dalgarno sequences in prokaryotes where direct association with the 16S rRNA occurs. Previous work had developed a library of short IRES sequences that base-pair with sections of the 18S rRNA under stressful conditions where the viability of the cell was dependent upon the IRES initiating translation of an auxotrophic gene product. We built a dicistronic vector to assess these IRES sequences under normal physiological conditions and discovered the lack of IRES activity. A 10-nt portion of an active IRES derived under stressful conditions was isolated and placed in tandem with multiple copies of itself. Seven copies of this module were required before IRES activity was visualized on a colorimetric plate-based assay. We propose designing a library screen for short IRES sequences based on placing a 10-nt randomized sequence adjacent to six modules of the known active IRES. Only positive IRES sequences will demonstrate the expected color change and will be selected for further characterization.

5.1. Introduction

Synthetic biology is advancing capabilities for the design of biological systems exhibiting desired functions. The proper functioning of synthetic genetic circuits often relies on the coordination of expression levels for the key protein components[1-3]. In prokaryotes, numerous genetic regulation schemes have been developed based on the control of translation initiation at the ribsome binding site (RBS), the Shine-Dalgarno sequence[4]. Modulation of gene activity has been achieved through the screening of a library of RBSes[5-6] and through the integration of the RBS into a riboswitch platform where the accessibility of the ribosome to the RBS is regulated through effector concentration[7-9]. Eukaryotic organisms utilize a different mechanism of translation initiation based upon the primary association of the ribosome at the 5' cap structure[10]. The fundamental difference in mechanisms between prokaryotic and eukaryotic organisms has resulted in the development of genetic tools that do not translate to the regulation of translation initiation in eukaryotes analogous to the prokaryotic RBS modules. However, a less common mechanism of translation initiation exists in eukaryotes based on the association of the ribosome to specific transcript structures and sequences known as internal ribosome entry sites (IRESes). IRESes are thus able to initiate translation via a cap-independent mechanism.

IRESes were initially discovered as a control element in the translation of coding viral RNA[11]. Subsequently, cellular IRESes were discovered in transcripts from viral-infected cells in which cap-dependent translation was effectively shut down[12]. Synthetic IRES modules were generated in mouse lines through deletional studies of the *Gtx* IRES that resulted in a 9-nt sequence that internally initiated translation and was

complementary to the 18S ribosomal RNA (rRNA), a critical component of the ribosomal complex[13]. In addition, when multiple modules of this 9-nt sequence were placed in tandem, the amount of translation was greatly enhanced as the avidity of the region with the 18S rRNA increased. It was hypothesized that this complementarity with the 18S rRNA initiates translation in a prokaryote-like manner by acting analogous to Shine-Dalgarno sequences that base-pair to the corresponding rRNA in prokaryotes, the 16S rRNA[4]. Synthetic IRES modules were also generated through library screening of a short random nucleotide sequence in mammalian cells[14] and in yeast cells[15]. In both studies, a dicistronic vector was constructed and the randomized library was placed in the intercistronic region (IR) such that positive IRESes could be selected through expression of the second cistron.

Prokaryotes naturally have several genes under the control of a single promoter that results in the synthesis of a multicistronic transcript where translation initiation is mediated through upstream RBSes. Eukaryotic transcripts are monocistronic and generally do not use internal initiation sequences to start translation and instead rely on a cap-dependent translation process. However, multicistronic transcripts can be generated through the introduction of an IRES element before each gene. Previously, retroviral multicistronic vectors had been developed in mammalian systems where multiple viral IRESes were incorporated[16-17], but these systems do not allow for the tuning of the gene components besides through the relocation of the gene behind different IRESes on the vector. These vectors are not available for use in yeast due to the viral IRESes not being able to initiate translation in the microorganism[18]. Recently, a reporter construct harboring two fluorescent genes was constructed for insertion of IRES elements in the IR

in order to control the ratio of expression between the two genes[19]. This work attempted to insert two previously described yeast cellular IRESes, *p150* and *YAP1*[18], with only *p150* resulting in expression of the second cistron. A library of small sequential IRES modules, acting analogous to Shine-Dalgarno sequences, with various translational efficiencies can lead to the improved development of multicistronic vectors in yeast where the ratio of gene expression between the cistrons can be modulated with appropriate library IRES sequences.

Here, we describe initial studies to develop an IRES library in *S. cerevisiae* that will have activity at normal physiological levels. Through the usage of the yeast α-galactosidase, MEL1[20], we developed a visual plate-based assay for screening a library of IRES modules. Visual confirmation of IRES activity was achieved only when seven modules of a modified synthetic IRES, IRES47, were placed in tandem; however, the overall IRES strength could not be quantified through a colorimetric assay of MEL1 activity. We propose a method for developing a set of synthetic IRES modules by placing six copies of IRES47 in tandem followed by a randomized seventh module. This library can then be screened in yeast for active sequences that would constitute the synthetic IRES module set.

5.2. Results

5.2.1. Implementing internal ribosome entry sites as RNA-based gene regulatory elements in dicistronic vectors

An 18-nt library was recently screened for IRES activity in *Saccharomyces cerevisiae* in a dicistronic vector where the IRES would drive the translation of the second cistron encoding the auxotrophic marker, *HIS3*[15]. The reported IRESes had varying degrees of complementarity to the 18S rRNA and interacted with various locations along the rRNA. One drawback of using an auxotrophic selection marker for the identification of active IRES sequences is that the cells are under stress when IRES-driven translation is occurring. The activity of cellular IRESes had been found to be induced under a multitude of cellular stresses[12]. The authors of the study did not directly compare the expression activities of the selected 18-nt IRESes to a cap-dependent control which consists of *HIS3* expressed from the same promoter as the dicistronic transcript (*ADH* promoter). However, from the reported data, when the plasmids containing the IRESes were transformed and plated on media lacking histidine, it took anywhere between 5 days to 4 weeks for colonies to form. We performed similar plate-based experiments with *HIS3* in yeast where the levels of this enzyme was set at ~10% of expression from the GAL1 promoter. Under these conditions, we observed that colonies developed within 1–2 days (A.H. Babiskin and C.D. Smolke, unpublished work, 2008). Thus, in comparing the results from these assays the selected IRESes are only producing very modest amount of proteins, barely above background levels. Therefore, these elements would likely be of limited use in many cellular engineering applications.

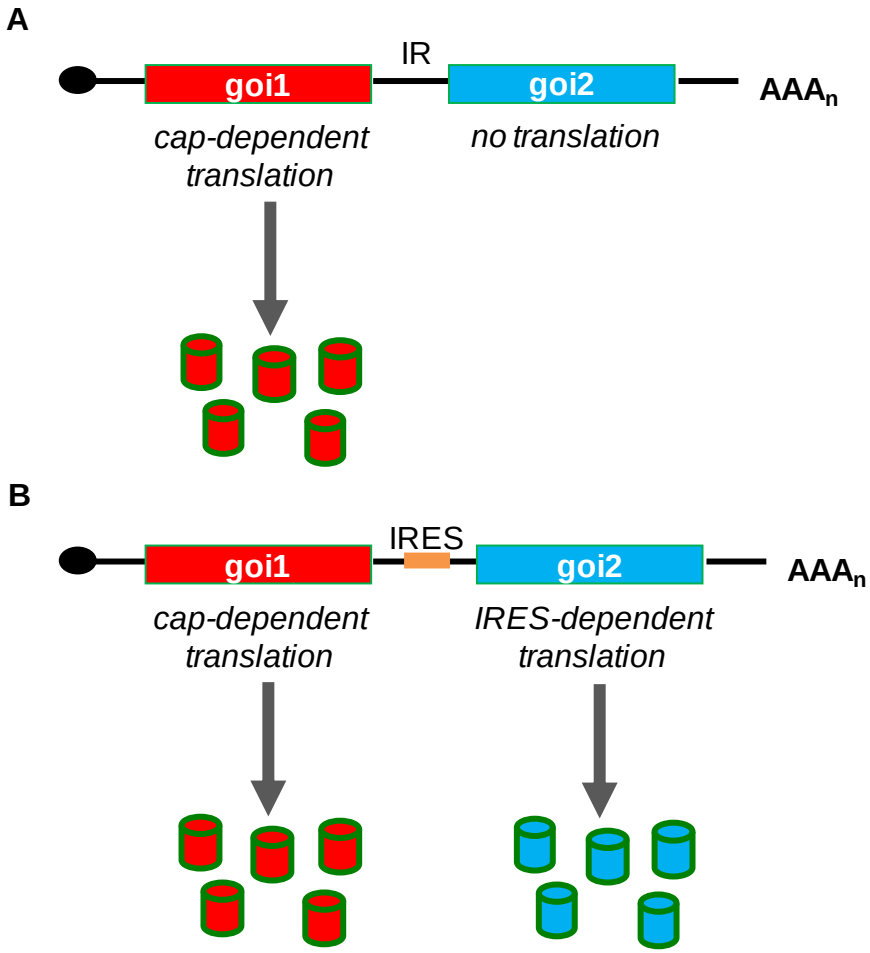


Figure 5.1. A yeast dicistronic vector based on insertion of an internal ribosome entry site (IRES) into the intercistronic region (IR) between the two genes of interest (goi1 and goi2). The expression pattern of the vector from both cistrons is demonstrated in the absence (A) and presence (B) of an IRES module. Barrels represent protein molecules.

We set out to develop a screen for IRES activity at normal physiological conditions for yeast. Initially, we began our studies with fluorescent dicistronic vectors containing RFP and GFP. In the developed vectors, translation of the first cistron is cap-dependent and translation of the second cistron is dependent on an active IRES being placed in the intergenic region (IR) (Figure 5.1). The distance between the start codon of the second cistron and IRES insertion site was determined through numerous previously studies with short sequential IRESes[13-15]. In addition, the vectors are high-copy to

increase the amount of transcripts in the cell, such that more IRES activity can be observed. *mRFP1*[21] and *yEGFP3*[22] were placed in two conformations, where we altered their order along the transcript (*mRFP1-yEGFP3*; *yEGFP3-mRFP1*), and we initially tested several previously selected 18-nt IRESes. It was our intention to screen a new library of short sequential IRESes based on increased fluorescence from the second cistron. However, we could not verify IRES activity in these vectors due to the inability to measure any changes in fluorescence with the previously selected 18-nt IRESes (data not shown).

5.2.2. Development of a plate-based screen for IRES activity

To improve the library screen, we decided to switch from a fluorescence-based assay of gene expression to a more sensitive enzyme-based reporter assay of gene expression based on MEL1[20]. Similar to standard LacZ assays, MEL1 assays associate a colorimetric change that can be observed from colonies on plates with changes in gene expression levels. An advantage of MEL1 is that the protein is secreted by yeast and requires no additional steps besides the actual plating to see the blue color formed, which is the product of the MEL1 enzyme acting on the substrate 5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside (X-α-gal). We placed *MEL1* in a dicistronic vector with *mRFP1* as the first cistron and *MEL1* as the second cistron to construct pRM. A *MEL1* monocistronic control, pMEL1, was also created in order to compare IRES-driven MEL1 activity to that caused by cap-dependent translation. As an initial control for IRES activity, the *YAP1* 5’ UTR was inserted in the IR of the dicistronic vector to create pR-YAP1-M. The 5’ UTR of *YAP1* had been demonstrated to contain a structured IRES with

two regions of complementarity to the 18S rRNA[18]. pRM, pMEL1, and pR-YAP1-M were tested for IRES activity by streaking out yeast cells harboring those plasmids on X-α-gal plates (Figure 5.2). The *YAP1* IRES was able to drive expression of MEL1 at physiological conditions. pRM, lacking an IRES, showed no color development, while the pMEL1 demonstrated higher MEL1 expression than pR-YAP1-M as observed by a stronger blue color.

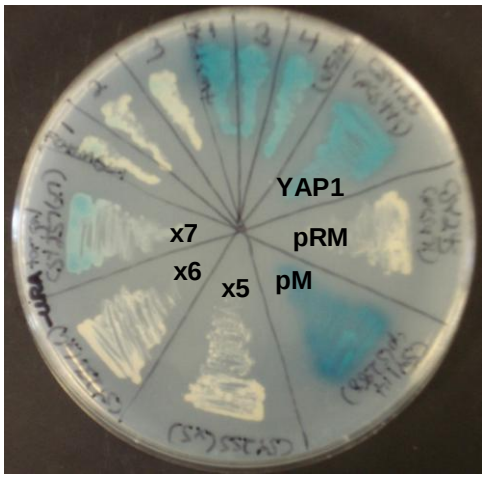


Figure 5.2. Visualization of MEL1 activity on X-α-gal plates. The cleaved products of X-α-gal by MEL1 form a blue color. The strength of the blue color is an indication of the amount of MEL1 protein being translated. Colonies from uracil dropout plates were streaked out on X-α-gal plates and allowed to incubate for two days at 30°C. pM, monocistronic *MEL1* control; pRM, *mRFP1-MEL1* dicistronic vector; YAP1, pR-YAP1-M; x5, pR-IRES47x5-M; x6, pR-IRES47x6-M; x7, pR-IRES47x7-M.

5.2.3. Implementation of short sequential IRESes in tandem drives translation initiation of MEL1

We next selected two of the strongest IRESes from the library of 18-nt IRESes for study in our dicistronic vector: IRES41 and IRES47[15]. The sequences were shortened to 10 nts based on the complementarity to the 18S rRNA (IRES41: TGCTGGGGTT; IRES47: CTGGTTGCTA) and inserted into the IR, where neither sequence exhibited any

blue color formation (data not shown). Additional nucleotides were included in and removed from the IRES sequences to respectively strength or weaken the base-pairing with the 18S rRNA; however, no blue color was observed with any of these constructs (data not shown). Previous studies with short sequential IRESes had determined that placing multiple IRES modules in tandem increased overall IRES activity[13]. We placed five copies of our shortened versions of IRES41 and IRES47 in tandem in the IR. 9-nt linker sequences were used to separate individual modules and no linker sequence was used more than once in a single design. The linker sequences were either A-rich or contained elements of the β-globin 5' UTR, which previously had been used as a negative control for IRES activity[13]. A construct harboring five modules of IRES47, IRES47x5, exhibited the faintest blue color. Therefore, we also built and examined six and seven copies of the IRES47 module. Seven modules of IRES47, IRES47x7, demonstrated substantial blue color formation, albeit less than that seen from the *YAP1* IRES (Figure 5.2).

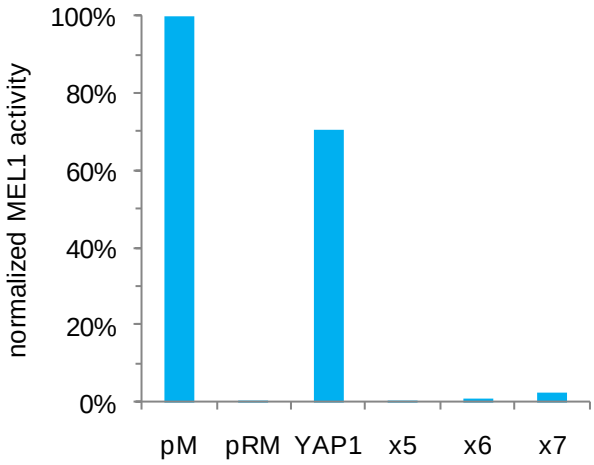


Figure 5.3. Quantification of MEL1 activity of constructs baring IRES modules. Spectrophotometry is used to measure the yellow cleavage product of PNPG caused by MEL1 activity. Activities are normalized to the cap-dependent control, pM. Cultures

were inoculated, allowed to grow overnight, and then harvested in the morning. pM, monocistronic *MEL1* control; pRM, *mRFP1-MEL1* dicistronic vector; YAP1, pR-YAP1-M;. x5, pR-IRES47x5-M; .x6, pR-IRES47x6-M; x7, pR-IRES47x7-M.

We next developed a protocol to quantify MEL1 levels from our expression vectors. An enzyme assays was performed by collecting yeast cells and resuspending them in an acidic buffer containing p-nitrophenyl-α-D-galactopyranoside (PNPG). A basic solution was then added to allow for yellow color generation (color formation cannot occur in the acidic buffer), which was then measured by spectrophotometry. From this protocol, we determined that the *YAP1* IRES had 71% of the activity as the cap-dependent control (pMEL1) from cultures grown overnight (Figure 5.3). However, only slight increases in activity over background could be measured with the multiple module IRES samples, particularly with IRES47x7. These solution-based results were not consistent with the amount of color formation observed on the plates with IRES47x7. It is a possibility that the assay is less sensitive for low expression levels.

While it is our goal to demonstrate IRES activity at physiological conditions, we had experienced issues with recovering cells from the plates with X-α-gal. The cells had no growth issues when initially grown on synthetic complete media lacking X-α-gal. It is possible that the substrate itself or the product from the MEL1-catalyzed reaction may be toxic to these cells. If that is the case, then we did not eliminate cellular stresses with this plate-based assay. The quantification assay does not have this issue since activity is measured from cells collected from synthetic complete media lacking X-α-gal. The removal of cellular stress may explain the reduction of IRES activity in the quantification of the multiple IRES modules through the solution-based assay.

5.2.4. Design of an IRES library to achieve tunable gene regulatory control

We observed increasing translational activity with increased IRES module number. Robust IRES activity was only observed with seven copies of IRES47 on the X-α-gal plates, whereas 6 copies exhibited only a very slight blue color. Based on these observations, we proposed that a library of 10-nt IRESes could be screened by randomizing the seventh module (Figure 5.4). The template sequence of this design can be found in Supplementary Table 5.1. In this design, when an inactive IRES sequence is placed in the seventh position, translation levels will be comparable to the 6-module IRES and hence virtually no color formation. An active IRES in the seventh position will drive blue color formation and will be selected for further characterization and study. As an additional control to this library design, we built a 7-module IRES where we included AT repeats in the seventh position. This construct displayed no visual difference from the 6-module IRES (data not shown).

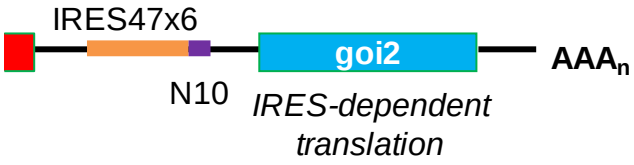


Figure 5.4. Proposed design for selection of a 10-nt IRES library in a dicistronic vector. Six modules of the IRES47 is positioned upstream of a 10-nt randomized region. Translation initiation mediated through the IR drives expression of the second cistron, goi2.

In preparation for the IRES library selection, the multiple module strains were restreaked from freezer stocks to serve as positive and negative controls and plated on X-α-gal plates. We observed that the IRES47x7 no longer produced a blue color, even though previous streaks from the same freezer stock demonstrated this behavior (data not shown). The sample was resequenced and the presence of the seven modules could not be

verified. It is possible that the homologous recombination was occurring between these elements and removing IRES modules over time due to the number of IRES repeats. In addition, optimization of the quantification assay for MEL1 activity failed to generate a strong signal for IRES47x7 (Figure 5.3), even though this construct generated a distinguishable blue color on X-α-gal plates (Figure 5.2). Due to the inability of the MEL1 assay to measure low expression levels and also the difficulty of recovering colonies from X-α-gal plates due to apparent toxicity of the plate-based assay, we decided to generate an alternative dicistronic vector with CyPET and YPET[23], since both proteins could be read simultaneously on the flow cytometer available in the laboratory (Cell Lab Quanta SC; Beckman Coulter). Based on preliminary tests for fluorescent signal strength, we decided to place CyPET in the first cistron and YPET in the second cistron to be controlled by IRES-dependent translation. We planned to use this two-fluorescence reporter construct in a high-throughput library screen based on fluorescence-activated cell sorting (FACS) for YPET-positive cells. The dicistronic vector (pCyY) and the monocistronic vectors (pCyPET and pYPET) initially displayed expected results (Figure 5.5), but problems arose once the IRESes were cloned into the dicistronic vector due to the sequence homology between CyPET and YPET. Sequencing of the IR proved to be very challenging and the dicistronic vector was very susceptible to homologous recombination between the gene pairs, effectively destroying the vectors. At this stage, we decided to put the project on hold until a resolution for these stability issues and a proper dual-gene system were found.

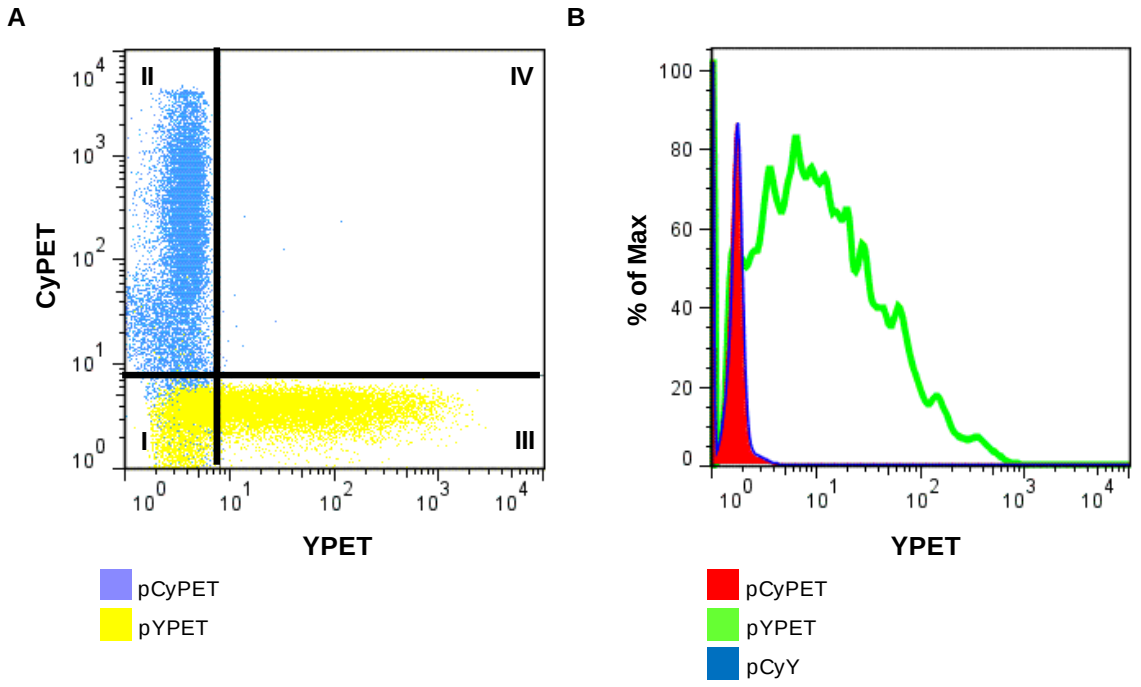


Figure 5.5. Expression profiles for the CyPET- and YPET-based constructs. (A) The monocistronic controls for CyPET and YPET correctly express their respective protein. pCyPET served as the negative control for the gating of YPET expression. pYPET served as the negative control for the gating of CyPET expression. Quadrants II and III represent positive CyPET and YPET populations, respectively. (B) The dicistronic CyPET-YPET vector, pCyY, exhibits no YPET expression from the second cistron. For simplification, only YPET expression is shown. pCyY is overlaid with pCyPET to further demonstrate the lack of YPET expression.

5.3. Discussion and Future Work

By controlling translation of the *MEL1* gene through IRES activity, we were able to develop a visual screen for IRES activity that allows for the rapid selection of IRES sequences. Preliminary experiments with our high-copy dicistronic vector determined that single-module small sequential IRESes were not effective in causing gene expression at substantial levels. The application of multiple IRES modules in tandem resulted in increased translation from *MEL1* and the effect could be visually observed with seven modules, albeit the level of activity was only a fraction of that observed with the *YAP1*

positive IRES control (Figure 5.3). The *YAP1* IRES contains two sequences that are complementary to the 18S rRNA that are positioned in a complex secondary structure[18]. Even though previous research had determined that the secondary structure was not critical for the function of cellular IRESes[13-14, 24], the substantial difference in activity may be due to the ability of the IRESes with large secondary structures, like *YAP1*, to form a conformational shape that allows for enhanced interaction with its complementary regions of the 18S rRNA. The mechanism by which increased activity can be seen with multiple small sequential IRES modules is believed to be due to increased synergy between the modules and components of the ribosomal machinery[13].

The strength of each individual IRES as well as the number of copies and spacing between modules have been determined to be factors in determining the overall strength of an IRES module[13]. In the development of the library, we would like to keep two of these factors (spacing and module number) constant while varying the third (strength of IRES module). The potency of an IRES is based on the location of base-pairing with the 18S rRNA and the actual strength of that base-pairing. Increasing the base-pairing interactions of an IRES segment does not necessarily lead to improved strength because that interaction can inhibit ribosome scanning[13]. On the other hand, decreased complementarity will also decrease the likelihood of interaction with the 18S rRNA. Because of these restrictions, the rational design of additional IRESes by modulating base-pairing has not been successful. It is necessary to develop a library of IRES from a randomized nucleotide region that will be able to access various regions of the 18S rRNA as well as alter the complementarity of each region. Since our work has established that single-module IRESes do not produce quantifiable levels of protein expression, we

proposed the development of a library in the context of multiple modules. Our design is based on the 7-module IRES, IRES47, where the seventh module would be randomized and screened for IRES activity (Figure 5.4). The effectiveness of this design is based on the observations that six modules of IRES47 and a 7-module negative control, in which the seventh position harbored a negative control sequence, produced no substantial blue color in a *MEL1* plate assay, in contrast to the 7-module version (Figure 5.2). Utilizing the same plate-based assay, positive IRES modules would be identified based on the generation of blue colonies. These clones would then be sequenced and built into 7-module versions for direct comparison against IRES47. Once a library of IRES sequences had been determined, these modules would be integrated combinatorially to extend the range of IRES activity.

The initially proposed system consisting of an *mRFP1-MEL1* dicistronic vector failed to be suitable for performing the IRES selection due to recombination issues and the lack of a strong measurable IRES signal (Figure 5.3). A second proposed system consisting of a *CyPET-YPET* dicistronic vector also encountered recombination issues. Recently, a new plasmid has been developed in the Smolke laboratory that contains two transgenes, *ymCherry* and *yEGPF3* (J.C. Liang et al., in preparation). While the fluorescence from ymCherry cannot be read on the Cell Lab Quanta SC, its fluorescence can be measured by the LSRII flow cytometer (Becton Dickinson Immunocytometry Systems) available at the Stanford Shared FACS Facility at Stanford University. In this dual-gene vector, both fluorescent genes are under control of the TEF1 promoter with the open reading frame (ORF) of *ymCherry* preceding that of *yEGFP3*. To create an *ymCherry/yEGFP3* dicistronic vector, the current plasmid can be modified by replacing

the region between and including CYC1 terminator of *ymCherry* and the TEF1 promoter of *yEGFP3* with an intercistronic sequence. Translation of *yEGFP3* will now be dependent upon placement of an active IRES in the IR. This dual fluoresecent reporter construct offers several advantages over the *mRFP1/yEGFP3* variants initially tested such as the general clarity of flow cytometry data over plate reader data, the ability to read fluorescence levels of both genes simultaneously, and the option of selecting the library through FACS.

The usage of our previous IRES designs in the *ymCherry/yEGFP3* dicistronic vector will not remove the possibility of homologous recombination occurring with the individual modules. We propose removing gap-repair as a method of building these multiple module IRESes and building everything directly by cloning in *E. coli*. We also propose redesigning the linker sequences to only contain adenine nucleotides. The main reason for the original design of various linker sequences and the cloning strategy by gap-repair was to build IRESes from smaller oligonucleotides to decrease the expense and the mutation rate associated with the synthesis of larger oligonucleotides. By ordering the entire IRES sequence in a single piece of DNA, we will be able to keep the linker sequences constant and possibly remove some of the homologous recombination occurring in yeast due to gap-repair by transforming a pure plasmid instead.

One of the potential applications of an IRES library is the generation of prokaryotic-like ‘operons’ in yeast. Multicistronic vectors have already been constructed for mammalian systems through the incorporation of viral IRESes[16-17]. One such vector was utilized to reconstitute the tetrahydrobiopterin (BH_4) pathway in BH_4-deficient human fibroblast cells[25]. Here, production of BH_4 was restored and modulated through

placement of the genes of the missing enzyme components at different locations under the control of different viral IRESes in the retroviral multicistronic vector. Similarly, the constitution of a heterologous pathway in yeast can be mediated through a single multicistronic vector. Optimization of product yield can be achieved through a combinatorial approach where the yeast IRES library is utilized to control relative expression of each genetic component. Another benefit of such a system is that the entire 'operon' is under the control of a single promoter and a promoter can be selected that allows global regulation of the foreign pathway due to single or multiple signals. One possible limitation of the IRES library is whether it can initiate translation at levels comparable to cap-dependent translation or natural yeast IRESes such as *YAP1* (Figure 5.3). It remains to be seen whether the current levels achieved from the sequential IRESes are sufficient to register a phenotypic response in the system.

5.4. Materials and Methods

5.4.1. Plasmid and strain construction

Standard molecular biology techniques were utilized to construct all plasmids[26]. DNA synthesis was performed by Integrated DNA Technologies (Coralville, IA). All enzymes, including restriction enzymes and ligases, were obtained through New England Biolabs (Ipswich, MA) unless otherwise noted. Pfu polymerases were obtained through Stratagene. Ligation products were electroporated into *Escherichia coli* DH10B (Invitrogen, Carlsbad, CA), where cells harboring cloned plasmids were maintained in Luria-Bertani media containing 50 mg/ml ampicillin (EMD Chemicals). Clones were

initially verified through colony PCR and restriction mapping. All cloned constructs and chromosomal integrations were sequence verified by Laragen (Los Angeles, CA).

A monomeric RFP gene, *mRFP1*, was PCR-amplified from pRSETB/mRFP1[27] using forward and reverse primers RFPmono.fwd2 (5' GCAAGCTTGGAGATCTAAA AGAAATAATGGCCTCCTCCGAGGACGT) and RFP_di_rev (5' GCGGTTGTCTAC ATGACTGACGCGTCCACTAGTCTTTAGGCGCCGGTGGAGTGG). A yeast-enhanced GFP gene, *yEGFP3*, was PCR-amplified from pSVA13[22] using forward and reverse primers GFP_di_fwd (5' GCGTCAGTCATGTAGACAACCGCGGGCACGTG AAAAGAAATAATGTCTAAAGGTGAAGAA) and GFP.mono.rev (5' CGCTCGAGG CCTAGGCTTTATTTGTACAATTCATCCATACCATGG), respectively. The *mRFP1* and *yEGFP3* products were spliced by overlap extension (SOE)[28] together using the forward and reserves primers RFPmono.fwd2 and GFP.mono.rev. The plasmid pCS101 was constructed by inserting the *mRFP1/yEGFP3* SOE by PCR product into pCS59, a modified version of pKW430[29] in which the second transcription start site of the ADH1 promoter was mutated to a NheI restriction site (M Win, C Smolke, unpublished data, 2004), via the unique restriction sites HindIII and XhoI which removes the NLS-NES *GFP* gene originally contained on pKW430. The yeast α-galactosidase, *MEL1*, was PCR-amplified from pMELα[20] using forward and reverse primers MEL1.di.fwd (5' GCGTCA GTCATGTAGACAACCGCG) and MEL1.di.rev (5' CGCTCGAGGCCTAGGCTTTAA GAAGAGGGTCTCAACCTATAGAG). The plasmid pCS165 was constructed by inserting the *MEL1* PCR product into pCS101 via the unique restriction sites SacII and AvrII, replacing *yEGFP3* with *MEL1*. Further sequencing of pCS165 revealed that there was unintended mutation in the intercistronic region between *mRFP1* and *MEL1*. The

mutation was corrected through direct insertion of a correct PCR product by gap-repair in the intercistronic generated with the forward and reverse primers IRfix_fwd (5’ CGAGGGCCGCCACTCCACCGGCGCCTAAAGACTAGTGGACGCGTCAGTCAT) and IRfix_MEL1_rev (5’ AGAAAGTAGAAAGCAAACATTATTTCTTTTCACGTGC CCGCGG) and template IRfix_temp (5’ ACTAGTGGACGCGTCAGTCATGTAGACA ACCGCGGGCACGTG) into pCS165 digested with MluI and SacII. The resultant corrected dicistronic plasmid is named pRM and its plasmid map is available in Supplementary Figure 5.1.

Monocistronic controls of pRM were generated with *mRFP1* and *MEL1*. *mRFP1* was again PCR-amplified from pRSETB/mRFP1 using forward and reverse primers RFPmono.fwd2 and RFP.mono.di.rev (5’ CGCTCGAGCCCTAGGCTTTAGG CGCCGGTGGAGTGG). The monocistronic *mRFP1* plasmid pRFP was constructed by inserting the *mRFP1* PCR product into pCS59 via the unique restriction sites HindIII and XhoI. *MEL1* was PCR-amplified from pMELα using the forward and reverse primers MEL1.mono.fwd (5’ GCAAGCTTGGAGATCTAAAAGAAATAATGTTTGCTTTCTA CTTTCTCACCG) and MEL1.di.rev. The monocistronic *MEL1* plasmid pMEL1 was constructed by inserting the *MEL1* PCR product into pCS59 via the unique restriction sites HindIII and XhoI.

The *YAP1* 5’ UTR was obtained through PCR from the yeast genome using forward and reverse primers YAP1_fwd (5’ GTCCGCGGTTGGTGTTTAGCTTTTTTT CCTGAGC) and YAP1_rev (5’ GGCTGGGTTTAAGAAACAACTTTTCCTTCTTTAA ACGT). The *YAP1* control plasmid pR-YAP1-M was constructed by gap-repairing the *YAP1* 5’ UTR PCR product into pCS165 via the unique restriction sites MluI and SacII.

Even though the erroneous pCS165 was used in the creation of the pCS449, the primers for the *YAP1* 5' UTR product contained the correct intercistronic sequence.

All short nucleotide IRESes with five or less modules were amplified by PCR using forward and reverse primers IRES_Gap_fwd2 (5' CGAGGGCCGCCACTCCACC GGCGCCTAAAGACTAGTGGACGCGTCAGTCATGTAG) and IRES_Gap_rev2 (5' ATGCATGCGGTGAGAAAGTAGAAAGCAAACATTATTTCTTTTCACGTG) except for the two- to four-module versions of IRES47, which used forward and reverse primers IRES47x1-5_prmr_fwd (5' GACTAGTGGACGCGTCAGTCATGTAGACAACCGCG G) and IRES47x1-5_prmr_rev (5' AGAAAGTAGAAAGCAAACATTATTTCTTTTCA CGTGTAGCA). The templates for amplification can be found in Supplementary Table 5.1. The PCR products were gap-repaired into pRM via the unique restriction sites MulI and SacII. The plasmids containing these IRESes are referred to as pR-'IRES name'-M. For example, the plasmid containing IRES47x5 is named pR-IRES47x5-M. The six and seven module versions of IRES47 were amplified by PCR using the forward and reverse primers IRES47x6-7_prmr_fwd (5' AAAAAAAAACTGGTTGCTAAAATTTAAACTG GTTGCTAATTTAATAACTGGTT) and IRES47x6-7_prmr_rev (5' AGAAAGTAGAA AGCAAACATTATTTCTTTTCACGTGTAGC). The templates for amplification can be found in Supplementary Table 5.1. The PCR products were gap-repaired into pR-IRES47x5-M via the unique restriction site PmlI.

A yellow fluorescent gene, *YPET*, was PCR-amplified from pBAD33/YPET[23] using forward and reverse primers YPET_di_fwd_prmr (5' GTCAAATAGACAACCGC GGGCACGTGAAAAGAAATAATGTCTAAAGGTGAAGAATTATTCACTGGTGT) and YPET_di_rev_prmr (5' GCCGAGCAGCAGCAAAACTCGAGCCCTAGGCTTTA

GTGGTGGTGGTGGTGGT). The plasmid pCS1192 was constructed by inserting the *YPET* PCR product into pRM, via the unique restriction sites SacII and XhoI, replacing *MEL1* with *YPET*. A cyan fluorescent gene, *CyPET*, was PCR-amplified from pBAD33/CyPET[23] using forward and reverse primers CyPET_di_fwd_prmr (5’ CCGCT GGAATAAGCTTGGAGATCTAAAAGAAATAATGTCTAAAGGTGAAGAATTATT CGGCGG) and CyPET_di_rev_prmr (5’ CCGCGGTTGTCTATTTGACTGACGCGTC CACTAGTCTTTAGTGGTGGTGGTGGTGGT). The plasmid pCyY was constructed by inserting the *CyPET* PCR product into pCS1192, via the unique restriction sites HindIII and SpeI, replacing *mRFP1* with *CyPET*.

Monocistronic controls of pCyY were generated with *CyPET* and *YPET*. *CyPET* was again PCR-amplified from pBAD33/CyPET using forward and reverse primers CyPET_di_fwd_prmr and CyPET_mono_rev_prmr (5’ AAACTCGAGCCCTAGGCTTT AGTGGTGGTGGTGGTGGT). The monocistronic *CyPET* plasmid pCyPET was constructed by inserting the *CyPET* PCR product into pCS59 via the unique restriction sites HindIII and XhoI. *YPET* was PCR-amplified from pBAD33/YPET using the forward and reverse primers YPET_mono_fwd_prmr (5’ AATAAGCTTGGAGATCTA AAAGAAATAATGTCTAAAGGTGAAGAATTATTCACTGGTGT) and YPET_di_ rev_prmr. The monocistronic *YPET* plasmid pYPET was constructed by inserting the *YPET* PCR product into pCS59 via the unique restriction sites HindIII and XhoI.

Following construction and sequence verification of the desired vectors, 100–500 ng of each plasmid was transformed into W303. In the case of gap-repair, 250–500 ng of the PCR product and 100 ng of plasmid digested with the appropriate restriction sites were transformed into the yeast strain. All yeast strains harboring cloned plasmids were

maintained on synthetic complete media with an uracil dropout solution containing 2% dextrose at 30°C. For the visualization of *MEL1* activity, yeast colonies were streaked on synthetic complete plates made with an uracil dropout solution containing 2% dextrose and 0.1 mg/ml 5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside (X-α-gal) (Glycosynth) dissolved in DMF. Cells expressing MEL1 will turn blue on these plates.

5.4.2. MEL1 quantification

The method for quantification of cellular MEL1 levels was adapted from previously developed protocols[30-31]. The reaction products of the cleavage of p-nitrophenyl-α-D-galactopyranoside (PNPG) by MEL1 form a characteristic yellow color. Briefly, yeast cells were grown overnight on synthetic complete media with an uracil dropout solution containing 2% dextrose. In the morning, various volumes were collected with the OD_{600} and volume recorded. The cell pellet was resuspended in 200 μl HSD Buffer [20 mM HEPES, pH 7.5, 0.002% (w/v) SDS, 10 mM DTT]. 60 μl of chloroform was added to the cell suspension and the sample was vortexed for 10 seconds. The sample was pre-equilibrated by incubating at 30ºC for 5 minutes. At 5 minutes, 800 μl of Z-PNPG [7 mM PNPG (Alfa Aesar), 61 mM citric acid, 77 mM Na_2HP0_4] was added to the sample. At various times, 100-μl aliquots were removed from the sample and the reaction stopped with 900 μl of 0.1 M Na_2CO_3. The cleaved product of PNPG remains colorless at a pH of 4.0 (the pH of Z-PNPG). The addition of basic Na_2CO_3 allows the yellow color to form. The terminated reaction products were centrifuged for 5 minutes at max speed to help clear cellular debris. 900 μl of the terminated reaction products were run on a Life Science UV/Vis Spectrophotometer (Beckman Coulter Fullerton, CA) with

the OD_{400} and OD_{550} was measured and recorded. MEL1 activity was calculated from the following equation:

$$MEL1\ activity = \frac{OD_{400} - (.9 * OD_{550})}{OD_{600} * volume * time}$$

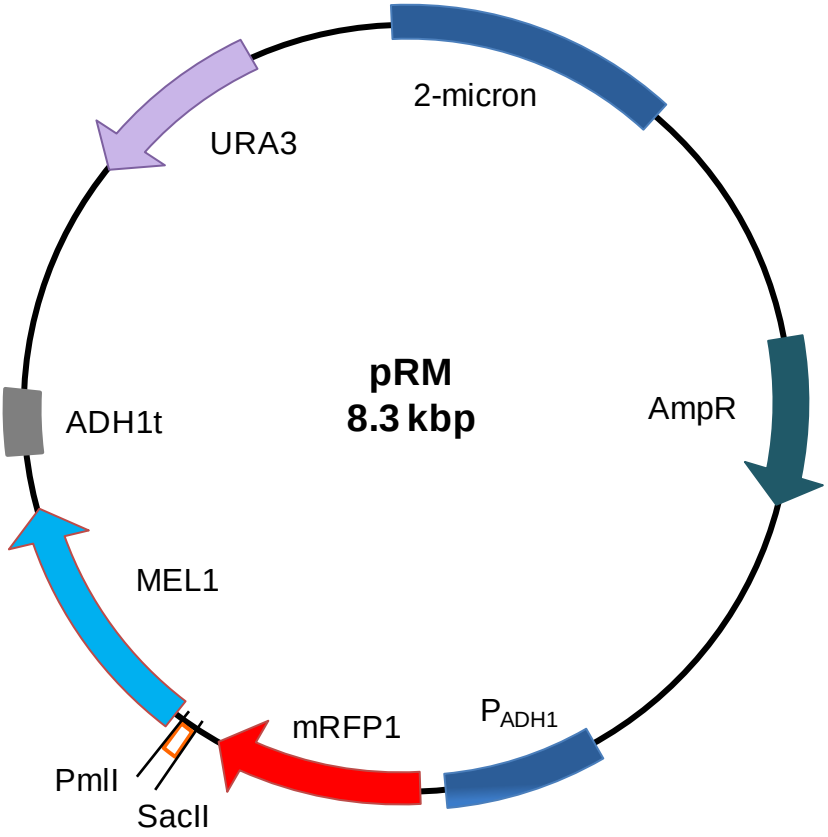
Only samples with OD_{400} values less than 3.0 are valid.

5.4.3. CyPET and YPET fluorescence distribution

S. cerevisiae cells harboring the pCyPET, pYPET, and pCyY plasmids were grown on synthetic complete media with an uracil dropout solution and 2% dextrose overnight at 30ºC. The cells were back-diluted the following morning into fresh media (5.0 ml total volume in test tubes) to an optical density at 600 nm (OD_{600}) of 0.1 and grown for 6 hours at 30ºC. On the Quanta flow cytometer (Beckman Coulter, Fullerton, CA) equipped with a 488-nm laser and UV arc lamp, the distribution of CyPET and YPET fluorescence was measured through 480/40-nm band-pass and 535/30 band-pass filters, respectively, and photomultiplier tube settings of 5.83 and 3.23, respectively. Data were collected under low flow rates until 10,000 viable cell counts were collected. pCyPET was used to set a gate to represent YPET-negative and YPET-positive populations. pYPET was used to set a gate to represent CyPET-negative and CyPET-positive populations.

5.5. Supplementary Information

Supplementary Figures and Tables



Supplementary Figure 5.1. Plasmid map of pRM, the dicistronic IRES characterization and screening plasmid. IRES modules are placed directly upstream of the *MEL1* gene.

Supplementary Table 5.1. The oligonucleotide template sequences of all synthetic IRESes tested in this study. For the forward and reverse primers used for PCR amplification, see Materials and Methods.

IRES	template	
IRES41x1	GGACGCGTCAGTCATGTAGACAACCGCGGATGCATTGCTGGGGTTCACG TGAAAAGAAATAATGTTTGCTTTCTAC	
IRES41x1 strong	GGACGCGTCAGTCATGTAGACAACCGCGGATGCATTGCTGGCACCCACGT GAAAAGAAATAATGTTTGCTTTCTA	
IRES41x5	forward	GCGTCAGTCATGTAGACAACCGCGGATGCATTGCTGGGGTT TTCTGACATTGCTGGGGTTTTCTGTTCTTGCTGGGGTT
	reverse	CAAACATTATTTCTTTTCACGTGAACCCCAGCAATGTCAGA AAACCCCAGCAATGTCAGAAAACCCCAGCAAGAA
IRES47x1	GGACGCGTCAGTCATGTAGACAACCGCGGATGCATCTGGTTGCTACACGT GAAAAGAAATAATGTTTGCTTTCTA	
IRES47x2	TCATGTAGACAACCGCGGCTGGTTGCTATTCTGACATCTGGTTGCTACAC GTGAAAAGAAATAATGTTTG	
IRES47x3	forward	TCATGTAGACAACCGCGGCTGGTTGCTATTCTGACATCTGG TTGCTAAGTTGTGTTCTGGT
	reverse	CAAACATTATTTCTTTTCACGTGTAGCAACCAGAACACAAC TTAGCAACC
IRES47x4	forward	TCATGTAGACAACCGCGGCTGGTTGCTATTCTGACATCTGG TTGCTAAGTTGTGTTCTGGT
	reverse	CAAACATTATTTCTTTTCACGTGTAGCAACCAGTTTTTTTT TTAGCAACCAGAACACAACTTAGCAACC
IRES47x5	forward	GCGTCAGTCATGTAGACAACCGCGGATGCATCTGGTTGCTA TTCTGACATCTGGTTGCTATTCTGTCTGCTGGTTGCTA
	reverse	CAAACATTATTTCTTTTCACGTGTAGCAACCAGATGTCAGA ATAGCAACCAGATGTCAGAATAGCAACCAGCAGAC
IRES47x6	AATTTAAACTGGTTGCTAATTTAATAACTGGTTGCTACACGTGAAAAGAA ATAATGTTTGC	
IRES47x7	AATTTAAACTGGTTGCTAATTTAATAACTGGTTGCTAATATATATACTGG TTGCTACACGTGAAAAGAAATAATGTTTGC	
IRES47x6+ni7	TTAAACTGGTTGCTAATTTAATAACTGGTTGCTAATATATATAATATATA TACACGTGAAAAGAAATAATGTTTGCTTTC	
IRES47x7 library	TTAAACTGGTTGCTAATTTAATAACTGGTTGCTAATATATATANNNNNNN NNNCACGTGAAAAGAAATAATGTTTGCTTTC	

Acknowledgements

This work was supported by the National Science Foundation (CAREER award to C.D.S.; CBET-0917705) and the Alfred P. Sloan Foundation (fellowship to C.D.S.).

References

1. Basu, S., Mehreja, R., Thiberge, S., Chen, M.T. & Weiss, R. Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A* **101**, 6355-6360 (2004).
2. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338 (2000).
3. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339-342 (2000).
4. Shine, J. & Dalgarno, L. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci U S A* **71**, 1342-1346 (1974).
5. Anderson, J.C., Clarke, E.J., Arkin, A.P. & Voigt, C.A. Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol* **355**, 619-627 (2006).
6. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946-950 (2009).
7. Desai, S.K. & Gallivan, J.P. Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J Am Chem Soc* **126**, 13247-13254 (2004).
8. Topp, S. & Gallivan, J.P. Random walks to synthetic riboswitches--a high-throughput selection based on cell motility. *Chembiochem* **9**, 210-213 (2008).
9. Lynch, S.A. & Gallivan, J.P. A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res* **37**, 184-192 (2009).

10. Kozak, M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187-208 (1999).

11. Nomoto, A., Kitamura, N., Golini, F. & Wimmer, E. The 5'-terminal structures of poliovirion RNA and poliovirus mRNA differ only in the genome-linked protein VPg. *Proc Natl Acad Sci U S A* **74**, 5345-5349 (1977).

12. Hellen, C.U. & Sarnow, P. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev* **15**, 1593-1612 (2001).

13. Chappell, S.A., Edelman, G.M. & Mauro, V.P. A 9-nt segment of a cellular mRNA can function as an internal ribosome entry site (IRES) and when present in linked multiple copies greatly enhances IRES activity. *Proc Natl Acad Sci U S A* **97**, 1536-1541 (2000).

14. Owens, G.C., Chappell, S.A., Mauro, V.P. & Edelman, G.M. Identification of two short internal ribosome entry sites selected from libraries of random oligonucleotides. *Proc Natl Acad Sci U S A* **98**, 1471-1476 (2001).

15. Zhou, W., Edelman, G.M. & Mauro, V.P. Isolation and identification of short nucleotide sequences that affect translation initiation in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **100**, 4457-4462 (2003).

16. De Felipe, P. & Izquierdo, M. Tricistronic and tetracistronic retroviral vectors for gene transfer. *Hum Gene Ther* **11**, 1921-1931 (2000).

17. de Felipe, P. & Izquierdo, M. Construction and characterization of pentacistronic retrovirus vectors. *J Gen Virol* **84**, 1281-1285 (2003).

18. Zhou, W., Edelman, G.M. & Mauro, V.P. Transcript leader regions of two Saccharomyces cerevisiae mRNAs contain internal ribosome entry sites that function in living cells. *Proc Natl Acad Sci U S A* **98**, 1531-1536 (2001).

19. Edwards, S.R. & Wandless, T.J. Dicistronic regulation of fluorescent proteins in the budding yeast Saccharomyces cerevisiae. *Yeast* **27**, 229-236 (2010).

20. Melcher, K., Sharma, B., Ding, W.V. & Nolden, M. Zero background yeast reporter plasmids. *Gene* **247**, 53-61 (2000).

21. Campbell, R.E. et al. A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* **99**, 7877-7882 (2002).

22. Mateus, C. & Avery, S.V. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast* **16**, 1313-1323 (2000).

23. Nguyen, A.W. & Daugherty, P.S. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat Biotechnol* **23**, 355-360 (2005).

24. Le Quesne, J.P., Stoneley, M., Fraser, G.A. & Willis, A.E. Derivation of a structural model for the c-myc IRES. *J Mol Biol* **310**, 111-126 (2001).

25. Laufs, S., Kim, S.H., Kim, S., Blau, N. & Thony, B. Reconstitution of a metabolic pathway with triple-cistronic IRES-containing retroviral vectors for correction of tetrahydrobiopterin deficiency. *J Gene Med* **2**, 22-31 (2000).

26. Sambrook, J. & Russell, D.W. Molecular Cloning: A Laboratory Manual, Edn. 3rd. (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY; 2001).

27. Shaner, N.C. et al. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* **22**, 1567-1572 (2004).

28. Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. & Pease, L.R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61-68 (1989).

29. Stade, K., Ford, C.S., Guthrie, C. & Weis, K. Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90**, 1041-1050 (1997).

30. Ryan, M.P., Jones, R. & Morse, R.H. SWI-SNF complex participation in transcriptional activation at a step subsequent to activator binding. *Mol Cell Biol* **18**, 1774-1782 (1998).

31. Rupp, S. LacZ assays in yeast. *Methods Enzymol* **350**, 112-131 (2002).

Chapter VI: Conclusions

For many applications of synthetic biology, such as metabolic engineering, there is a need for the development of genetic tools that allow for the fine-tuning of gene expression in *Saccharomyces cerevisiae*. The majority of tools currently employed are limited to the use of native yeast promoter systems with or without the regulation of transcription factors and the use of synthetic promoters engineered rationally or through the screening of libraries[1]. Given constant culture conditions, output from constitutive promoters cannot be altered unless used in plasmid systems with different copy numbers. However, the difference in expression between low- and high-copy plasmids is considerable, with high-copy plasmids often resulting in overexpression of a desired plamid. Constitutive promoters can be used interchangeably, but they are limited in number and thus limited in what gene expression levels can be achieved. The regulatory range of a constitutive promoter can be modulated through the library of a mutatnt promoter library, such as the TEF1 promoter library created through random mutagenesis of the wild-type promoter[2]. Inducible promoter systems have an advantage in that expression output from the promoter can be controlled by setting the concentration of the small molecule inducers. However, many inducible promoter systems do not exhibit substantial titratable regimes, such that fine tuning expression levels can be difficult from such systems. Often times, inducible promoter systems exhibit heterogenous behavior, where a given cell in the population is either fully expressing or repressed in these systems and a change in the ratio of these two states across the population is observed by varying the inducer concentration[3]. Inducible promoters are also not desirable for industrial applications due to the additional cost and to nonspecific or pleotropic effects

associated with the inducing molecule. Synthetic promoter systems are also limiting in applications where expression of an endogenous gene is to be modulated, but it is desired to retain the cellular regulatory mechanisms associated with the native promoter.

RNA-based postranscriptional elements have the advantage of being coupled to any promoter of choice, allowing the utilization of enhanced control strategies. RNA is a diverse molecule with structural, enzymatic, and ligand-binding properties[4]. Once a coding RNA is transcribed, it goes through two processes: translation and degradation. Previously described RNA devices have controlled these process through antisense or ribozyme technology, respectively, in response to a small molecule effector[5-7]. In my thesis research, we developed novel genetic control modules to precisely tune gene expression in *S. cerevisiae* through posttranscriptional regulation. Two libraries of synthetic RNA hairpins placed in the 3' untranslated region (UTR) of transcripts mediated endonucleolytic cleavage by the RNase III enzyme Rnt1p. Processing by the enzyme leads to a reduction in transcript levels that ultimately results in decreased protein levels. Various intermediate levels of gene expression were achieved through the randomization of nucleotides associated with the cleavage efficiency box (CEB) and binding stability box (BSB) of Rnt1p substrates (Chapter II and Chapter III, respectively). These CEB and BSB modules were integrated combinatorially to further extend the range of the overall Rnt1p hairpin library (Chapter III). In addition, the ligand-sensing properties of RNA aptamers were incorporated into Rnt1p hairpins to develop an Rnt1p switch (Chapter IV). Here, the binding of ligand directly interfered with Rnt1p cleavage, resulting in increased gene expression. Switch dynamics were rationally modulated and improved through incorporating aptamers with differing affinities, BSB modules, and

switch modules. Lastly, we attempted to screen for a library of small sequential internal ribosomal entry sites (IRESes) at physiological conditions that differ in their ability to initiate translation (Chapter V). Such synthetic IRES elements will allow for the construction of yeast 'operons', where appropriate gene ratios can be achieved through the usage of IRESes of various strengths.

The Rnt1p hairpin library and the Rnt1p switches provides an efficient way to control gene expression levels in *S. cerevisiae*. The RNA modules take advantage of an unique property of the Rnt1p enzyme - its specific recognition of hairpins containing an AGNN tetraloop[8]. Other RNase III enzymes characterized to date have not demonstrated this property. As such, these hairpins will only function in *S. cerevisiae*, where Rnt1p is a critical enzyme involved in the processing of non-coding structural and functional RNAs[9-12]. It is possible that these hairpins can be extended as orthogonal control elements in additional yeast species, other eukaryotes, and prokaryotes by heterologously-expressing Rnt1p. The transfer of RNase-based regulatory strategies has been recently demonstrated in studies where a yeast variant of the RNase III Dicer enyzme from *S. castellii* and other factors were expressed in *S. cerevisiae* to introduce the mechanism of RNA interference (RNAi)[13].

There are several challenges with implementation of the Rnt1p hairpin that should be addressed in future studies. The dependence of the hairpins' regulatory activity on Rnt1p may present disadvantages as gene expression is dependent on the cellular machinary of the host organism. For example, reductions in cellular Rnt1p levels may ultimately limit the regulatory range of the library and, in the case of Rnt1p switches, decrease the dynamic range of the switches, while increasing responsiveness to the

effector. On the other hand, it may be possible to counteract natural decreases of Rnt1p or to increase the silencing activity by Rnt1p by expressing more Rnt1p in the yeast strain. The Rnt1p substrates generated in this thesis generally act as modular units as demonstrated through the placement of the cleavage library in multiple genetic contexts (Chapter II) and the multiplicative baseline observed in the multiple module Rnt1p switches (Chapter IV). The mulitplicative baseline also suggests that the Rnt1p hairpins derived from the cleavage library can be placed in tandem to increase silencing activity. The initial attempts to place two hairpins in tandem in the 3' UTR were what led to the eventual removal of structurally weak structures from the cleavage library as hairpin structures appeared to be affecting each others' activity. The multiple module switches did not appear to have this issue as their structures were insulated by A-rich flanking sequences. The inclusion of such flanking sequences could help enhance the modularity of the cleavage library hairpins, but will likely also affect the knockdown observed from the hairpins leading to a recharacterization of the library or perhaps a screening of a new cleavage library with the flanking sequences included. Due to the small percentage of cleavage hairpins that lack modularity in a particular system, the described additional studies will only result in minor improvements. Instead, if multiple hairpin constructs are desired, it may be more prudent to perform research on linker sequences in the intervening regions between hairpins to ensure maintenance of activity.

The Rnt1p switches will serve as important tools in metabolic engineering by allowing the construction of synthetic feedback systems based on the sensing of metabolites and other molecules directly involved in the metabolic network and regulation of key enzyme activities. The development of tailored metabolite-responsive

Rnt1p switches is currently limited by the number of aptamers available that respond to ligands of interest and the subset of those aptamers that can be successfully integrated into the Rnt1p switch design due to structural constraints. In the development of the Rnt1p switch with the theophylline aptamer, the aptamer was positioned at different locations in the CEB (Supplementary Figure 4.1). While each design demonstrated the ability to be cleaved by Rnt1p, only one (RS) featured inhibition of cleavage upon the addition of ligand. This work highlights that the binding interactions of the ligand with the aptamer and its location in the CEB is critical for the ligand-dependent impedance of Rnt1p activity. The xanthine aptamer[14] and the tetracycline aptamer[15] were also built into the switch platform at multiple locations in the CEB, albeit unsuccessfully. Incorporation of both aptamers did not lead to any significant knockdown in the absence of ligand, i.e. the aptamer structure itself was inhibiting Rnt1p activity. These aptamers and the failed theophylline aptamers highlight the importance of the aptamer structure being processable by Rnt1p and the ability of the ligand binding interactions to interfere with Rnt1p activity.

The added structural requirements for the Rnt1p switch platform likely reduce the probability that an *in vitro* selected aptamer will respond in this platform. Therefore, future efforts to expand the Rnt1p switch platform to respond to alternative stimili should focus on the selection of aptamers with the switch platform *in vivo* in an appropriate reporter plasmid. For example, libraries can be designed with randomized aptamer/CEB region and functional sequences can be selected for decreased fluorescence in the absence of ligand (for ability to be cleaved by Rnt1p) and for an increase in fluorescence upon ligand addition (for selection of functional aptamers in which ligand binding interferes

with Rnt1p activity). Similar efforts are already under way in the Smolke laboratory with the selection of aptamers in the ribozyme-based riboswitch platform (J.C. Liang et. al., in preparation), and these strategies can be directly extended to the Rnt1p switch platform.

The Rnt1p hairpin library and the Rnt1p switches can be applied to a wide range of applications. We demonstrated the library's ability to examine the architecture of endogenous networks by pertubating metabolic flux around a major control point. Specifically, the expresson of ERG9, the first enzyme involved in the ergosterol biosynthesis, was modulated to reduce flux through the pathway (Chapter II). Native feedback control around this pathway was observed as ERG9 expression levels were maintained at a certain threshold value. In metabolic engineering applications, the Rnt1p-based modules can be employed to reduce metabolic burden, balance cytotoxic intermediate, and redirect cellular resources from native pathways.

We envision that the Rnt1p library can be employed by two methods. First, with the 16 CEB and 16 BSB modules described, 256 different Rnt1p hairpins can be constructed and integrated into the 3’ UTR of a gene of interest to screen for a desired phenotype. Ordering the oligonucleotides necessary to construct the 256 hairpins will likely be expensive. As an alternative option, since a wide range of gene expression was observed from the cleavage library, a library of randomized CEB sequences could be screened for a desired phenotype in a construct appropriate for the application of interest. Second, the two libraries can be applied rationally to limit the amount of constructs tested. The cleavage library was observed to have the largest range of gene regulatory activity and also the best distribution, while the binding library had a decreased range more suitable for the tuning of gene expression. The cleavage library can be employed

first to identify regulatory regions of interest. Then, more focused expression levels can be explored through the implementation of the binding library BSB modules with the appropriate cleavage library CEB modules.

A synthetic IRES library, once successfully developed, will allow the construction of multicistronic vectors in yeast where the relative ratios of the individual gene products can be altered through the integration of the library sequences. While a structured, native IRES in yeast provides similar levels of gene expression as cap-dependent translation (Figure 5.3), single copies of small IRES sequences with complementarity to the 18S rRNA are ineffective at producing substantial levels of gene product. Seven linked copies of these IRES modules were able to produce a visual phenotypic response due to MEL1 reponse (Figure 5.2), but levels were measured to be much lower compared to cap-dependent translation (Figure 5.3). Questions about the sensitivity of the MEL1 assay and issues with homologous recombination led to the attempt to construct a dual fluorescense reporter system with *CyPET* and *YPET*. However, homology between the gene pairs destabilized the resulting vectors. Recently, a new low-copy dual fluorescence reporter plasmid has been characterized in the Smolke laboratory that contains two transgenes, *ymCherry* and *yEGPF3* (J.C. Liang et al., in preparation). A future direction for the IRES project would be to rearrange this vector to produce an *ymCherry-yEGFP3* dicistronic vector. The ability of *yEGFP3* expression to report on IRES activity should first be tested through the placement of multiple copies of the IRES47 sequence in the intergenic region (IR). It is recommended that the method of gap repair be removed from this project since it depends on homologous recombination for successful cloning. It is also recommended that the entire IRES module be cloned from a single oligonucleotide template so that the

linker sequences are constant. Since the response from MEL1 and yEGFP3 may be different, there may need to be additional optimization around the number of IRES copies necessary for activity to be observed with yEGFP3. Once the dual-fluorescence reporter is optimized, the proposed library strategy can be employed to select active IRES modules through fluorescence-activated cell sorting (FACS).

In summary, we have developed synthetic genetic control modules for *S. cerevisiae* acting through posttranscriptional mechanisms that can be implemented with different genetic targets and promoters. The Rnt1p hairpin library provides a key tool for synthetic biology applications in yeast where the predictable tuning of gene expression is necessitated. Aptamer integration allows for the construction of ligand-responsive Rnt1p-based control devices. The development of synthetic IRESes will be useful in the creation of yeast 'operons' where various ratios of expression for several genes can be linked through a single transcriptional event. The synthetic control modules developed in this thesis will provide an important toolset for advancing yeast as a microbial host for bioprocessing and biosynthesis applications.

References

1. Nevoigt, E. Progress in metabolic engineering of Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **72**, 379-412 (2008).
2. Nevoigt, E. et al. Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. *Appl Environ Microbiol* **72**, 5266-5273 (2006).

3. Louis, M. & Becskei, A. Binary and graded responses in gene networks. *Sci STKE* **2002**, pe33 (2002).

4. Win, M.N., Liang, J.C. & Smolke, C.D. Frameworks for programming biological function through RNA parts and devices. *Chem Biol* **16**, 298-310 (2009).

5. Bayer, T.S. & Smolke, C.D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* **23**, 337-343 (2005).

6. Win, M.N. & Smolke, C.D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A* **104**, 14283-14288 (2007).

7. Win, M.N. & Smolke, C.D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456-460 (2008).

8. Lamontagne, B. et al. Sequence dependence of substrate recognition and cleavage by yeast RNase III. *J Mol Biol* **327**, 985-1000 (2003).

9. Catala, M., Lamontagne, B., Larose, S., Ghazal, G. & Elela, S.A. Cell cycle-dependent nuclear localization of yeast RNase III is required for efficient cell division. *Mol Biol Cell* **15**, 3015-3030 (2004).

10. Chanfreau, G., Elela, S.A., Ares, M., Jr. & Guthrie, C. Alternative 3'-end processing of U5 snRNA by RNase III. *Genes Dev* **11**, 2741-2751 (1997).

11. Chanfreau, G., Rotondo, G., Legrain, P. & Jacquier, A. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J* **17**, 3726-3737 (1998).

12. Elela, S.A., Igel, H. & Ares, M., Jr. RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell* **85**, 115-124 (1996).

13. Drinnenberg, I.A. et al. RNAi in budding yeast. *Science* **326**, 544-550 (2009).

14. Kiga, D., Futamura, Y., Sakamoto, K. & Yokoyama, S. An RNA aptamer to the xanthine/guanine base with a distinctive mode of purine recognition. *Nucleic Acids Res* **26**, 1755-1760 (1998).

15. Berens, C., Thain, A. & Schroeder, R. A tetracycline-binding RNA aptamer. *Bioorg Med Chem* **9**, 2549-2556 (2001).

Appendix

Table of plasmids and yeast strains and their relation to the laboratory database

All plasmids (maintained in bacteria) and yeast strains described in this thesis are listed in the following table. They appear with their respective plasmid stock (pCS) and yeast strain (CSY) numbers in the Smolke laboratory database. Unless indicated in the database, all plasmid stocks are maintained in the *E. coli* DH10B strain and/or the *S. cerevisiae* W303 strain. Different nomenclatures were utilized in the initial creation of the plasmids/strains (as they appear in the database) and in the text of this thesis. The name as they appear in the thesis is found in the "Name in thesis" column and the name as they appear in the database can be found in the "Database description" column. The "Thesis chapter" column contains the chapters in which the indicated plasmids/strains appear in the thesis.

Plasmid (pCS)	Strain (CSY)	Name in thesis	Thesis chapter	Database description
4		modified pRS316	II	pRS316 (low-copy, URA3 selection) with GAL1 promoter
8		pRS316	II	low-copy, URA3 selection
13	11	pKW430	V	high-copy, ADH1 promoter, URA3 selection, contains NLS-NES-GFP
37		pSVA13	II, V	yEGFP3 source
55		pRSETB/ mRFP1	V	mRFP1 source
59		pCS59	V	modified ADH1 promoter of pCS13 (modified to have one transcript start site), created by Maung Win
65	15	pRFP	V	pCS59/mRFP1: monocistronic mRFP1, high-copy, ADH1 promoter, URA3 selection
66	10	yEGFP3-mRFP1	V	pCS59/GFP/RFP: dicistronic yEGFP3-mRFP1, high-copy, ADH1 promoter, URA3 selection
101		pCS101, mRFP1-yEGFP3	V	pCS59/RFP/GFP: dicistronic mRFP1-yEGFP3, high-copy, ADH1 promoter, URA3 selection, incorrect IR sequence
141		pMELa	V	MEL1 source
165	100	pCS165	V	pCS59/mRFP1/MEL1: dicistronic mRFP1-MEL1, high-copy, ADH1 promoter, URA3 selection, incorrect IR sequence
182	77	pCS182	II	pCS4/yEGFP3: monocistronic yEGFP3, low-copy, GAL1 promoter, no terminator, URA3 selection
270		pUG6	II	integration cassette with kanamycin resistance selection marker
288	114	pMEL1, pM	V	pCS59/MEL1: monocistronic MEL1 vector, high-copy, ADH1 promoter, URA3 selection
321	121	pCS321, no insert	II, III, IV	pCS182 with ADH1 terminator included downstream of yEGFP3
449	133	pR-YAP1-M, YAP1	V	pCS165/YAP1: YAP1 IRES placed in IR (correct IR sequence)
471	245	pRM	V	pCS59/mRFP1/MEL1: dicistronic mRFP1-MEL1, high-copy, ADH1 promoter, URA3 selection (corrected IR sequence)
809	255	pR-IRES47x5-M, x5	V	pCS471/IRES47x5: 5 modules of IRES47 in IR
	256	pR-IRES47x6-M, x6	V	pCS471/IRES47x6: 6 modules of IRES47 in IR
	257	pR-IRES47x7-M, x7	V	pCS471/IRES47x7: 7 modules of IRES47 in IR
1135	204	RS	IV	pCS321/TR2_12: working theophyline Rnt1p switch

Plasmid (pCS)	Strain (CSY)	Name in thesis	Thesis chapter	Database description
	205	RSN	IV	pCS321/TR2_12_neg: mutated tetraloop (CAUC)
	206	RSnt	IV	pCS321/TR2_12_no_theo: inactive theophylline aptamer
1192		pCS1192	V	pCS59/RFP/YPET: dicistronic mRFP1-YPET, high-copy, ADH1 promoter, URA3 selection
1193	431	pCyPET	V	pCS59/CyPET: monocistronic CyPET, high-copy, ADH1 promoter, URA3 selection
1194	432	pYPET	V	pCS59/YPET: monocistronic YPET, high-copy, ADH1 promoter, URA3 selection
1195	430	RSx2	IV	pCS321/TR2_12x2: two modules of TR2_12 in tandem
1254	445	pCyY	V	pCS59/CyPET/YPET: dicistronic CyPET-YPET, high-copy, ADH1 promoter, URA3 selection
1332	462	RSx3	IV	pCS321/TR2_12x3: three modules of TR2_12 in tandem
1333		pBAD33/YPET	V	YPET source
1334		pBAD33/CyPET	V	CyPET source
1418		pRNT1	II, III, IV	pPROEX/Rnt1p: Rnt1p expression plasmid
1682		pCS321-ERG9	II	pCS321/ERG9: yEGFP3 replaced by ERG9
1717		modified pUG6	II	pUG6 XhoI mutant (XhoI site changed to CTGGAG)
1813		pCS1813	II	pCS1717/ERG9: ERG9 cloned upstream of first loxP site
1960	683	RS-B05	IV	pCS321/TR2-bind4
1961	684	RS-B06	IV	pCS321/TR2-bind5
1962	685	RS-B07	IV	pCS321/TR2-bind7
1963	686	RS-B17	IV	pCS321/TR2-bind17
2037	748	RS-B07x2	IV	pCS321/TR2-bind7x2
2038	749	RS-B17x2	IV	pCS321/TR2-bind17x2
	3	wild-type strain, W303	II, III, IV, V	W303 (MATa, his3-11,15 trp1-1 leu2-3 ura3-1 ade2-1)
	329	A01 (GFP)	II	pCS321/R2
	330	A02, A02-B00 (GFP)	II, III, IV	pCS321/R3
	791	C01 (GFP)	II	pCS321/A1
	792	C02 (GFP)	II	pCS321/A8
	793	C03 (GFP)	II	pCS321/B1
	794	C04 (GFP)	II	pCS321/B4
	795	C05 (GFP)	II	pCS321/B7
	796	C06 (GFP)	II	pCS321/B10

Plasmid (pCS)	Strain (CSY)	Name in thesis	Thesis chapter	Database description
	797	C07 (GFP)	II	pCS321/B11
	798	C08 (GFP)	II	pCS321/C2
	799	C09 (GFP)	II	pCS321/C6
	800	C10 (GFP)	II	pCS321/C10
	801	C11 (GFP)	II	pCS321/C11
	802	C12 (GFP)	II	pCS321/D3
	331	C13, C13-B00 (GFP)	II, III	pCS321/D6
	803	C14 (GFP)	II	pCS321/D10
	467	A01 (GFP) (CAUC)	II	pCS321/R2_neg
	468	A02, A02-B00 (GFP) (CAUC)	II, III	pCS321/R3_neg
	469	C01 (GFP) (CAUC)	II	pCS321/D6_neg
	753	C02 (GFP) (CAUC)	II	pCS321/A1_neg
	754	C03 (GFP) (CAUC)	II	pCS321/A8_neg
	755	C04 (GFP) (CAUC)	II	pCS321/B1_neg
	756	C05 (GFP) (CAUC)	II	pCS321/B4_neg
	474	C06 (GFP) (CAUC)	II	pCS321/B7_neg
	757	C07 (GFP) (CAUC)	II	pCS321/B10_neg
	475	C08 (GFP) (CAUC)	II	pCS321/B11_neg
	758	C09 (GFP) (CAUC)	II	pCS321/C2_neg
	759	C10 (GFP) (CAUC)	II	pCS321/C6_neg
	760	C11 (GFP) (CAUC)	II	pCS321/C10_neg
	479	C12 (GFP) (CAUC)	II	pCS321/C11_neg
	761	C13, C13-B00 (GFP) (CAUC)	II, III	pCS321/D3_neg
	762	C14 (GFP) (CAUC)	II	pCS321/D10_neg
	605	no insert (ERG9)	II	ERG9::ERG9-AvrII-XhoI-ADH1t-kanMX (ERG9 3' UTR replacement)
	606	C13-B01	III	pCS321/D6_P7-1
	607	C13-B02	III	pCS321/D6_P7-3
	608	C13-B03	III	pCS321/D6_P7-6
	609	A02-B01	III	pCS321/R3_P7-1
	610	A02-B02	III	pCS321/R3_P7-3
	611	A02-B03	III, IV	pCS321/R3_P7-6
	645	RS-B03	IV	pCS321/TR2_P7-6
	648	A02-B04	III	pCS321/R3-bind2
	649	A02-B05	III, IV	pCS321/R3-bind4
	650	A02-B06	III, IV	pCS321/R3-bind5
	651	A02-B07	III, IV	pCS321/R3-bind6

Plasmid (pCS)	Strain (CSY)	Name in thesis	Thesis chapter	Database description
	652	A02-B08	III	pCS321/R3-bind7
	653	A02-B09	III	pCS321/R3-bind8
	654	A02-B10	III	pCS321/R3-bind11
	655	A02-B11	III	pCS321/R3-bind15
	656	A02-B12	III, IV	pCS321/R3-bind17
	657	A02-B13	III	pCS321/R3-bind18
	658	A02-B14	III	pCS321/R3-bind20
	659	A02-B15	III	pCS321/R3-bind22
	660	A02-B04 (CAUC)	III	pCS321/R3-bind2N
	661	A02-B05 (CAUC)	III	pCS321/R3-bind4N
	662	A02-B06 (CAUC)	III	pCS321/R3-bind5N
	667	A01 (ERG9)	II	ERG9::ERG9-R2-ADH1t-kanMX
	668	C13-GA3 (ERG9)	II	ERG9::ERG9-GA3-ADH1t-kanMX
	669	C06 (ERG9)	II	ERG9::ERG9-B10-ADH1t-kanMX
	670	C07 (ERG9)	II	ERG9::ERG9-B11-ADH1t-kanMX
	671	C13-B04	III	pCS321/D6-bind2
	672	C13-B05	III	pCS321/D6-bind4
	673	C13-B06	III	pCS321/D6-bind5
	674	C13-B07	III	pCS321/D6-bind6
	675	C13-B08	III	pCS321/D6-bind7
	676	C13-B09	III	pCS321/D6-bind8
	677	C13-B10	III	pCS321/D6-bind11
	678	C13-B11	III	pCS321/D6-bind15
	679	C13-B12	III	pCS321/D6-bind17
	680	C13-B13	III	pCS321/D6-bind18
	681	C13-B14	III	pCS321/D6-bind20
	682	C13-B15	III	pCS321/D6-bind22
	691	A02 (ERG9)	II	ERG9::ERG9-R3-ADH1t-kanMX
	692	C10 (ERG9)	II	ERG9::ERG9-C10-ADH1t-kanMX
	693	C08 (ERG9)	II	ERG9::ERG9-C2-ADH1t-kanMX
	694	C13-B00 (GAAA)	III	pCS321/D6 (GA3)
	695	A02-B00 (GAAA)	III	pCS321/R3 (GA3)
	702	RS-theo2	IV	pCS321/TR2_theo2
	704	RS-theo3	IV	pCS321/TR2_theo4
	763	A01 (mCherry)	II	pCS1749/R2
	764	A02 (mCherry)	II	pCS1749/R3
	765	C01 (mCherry)	II	pCS1749/A1
	766	C02 (mCherry)	II	pCS1749/A8

Plasmid (pCS)	Strain (CSY)	Name in thesis	Thesis chapter	Database description
	767	C03 (mCherry)	II	pCS1749/B1
	768	C04 (mCherry)	II	pCS1749/B4
	769	C05 (mCherry)	II	pCS1749/B7
	770	C06 (mCherry)	II	pCS1749/B10
	771	C07 (mCherry)	II	pCS1749/B11
	772	C08 (mCherry)	II	pCS1749/C2
	773	C09 (mCherry)	II	pCS1749/C6
	774	C10 (mCherry)	II	pCS1749/C10
	775	C11 (mCherry)	II	pCS1749/C11
	776	C12 (mCherry)	II	pCS1749/D3
	777	C13 (mCherry)	II	pCS1749/D6
	778	C14 (mCherry)	II	pCS1749/D10
1749	781	pCS1749	II	monocistronic ymCherry, TEF1 promoter, CYC1 terminator, low-copy, URA3 selection, created by Joe Liang
1585		pCS1585	III	monocistronic yEGFP3, TEF1 promoter, ADH1 terminator, low-copy, URA3 selection, created by Joe Liang
1748		pCS1748	III	contains the ymCherry and yEGFP3 ORFs of pCS1749 and yEGFP3, low-copy, URA3 selection, created by Joe Liang