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.

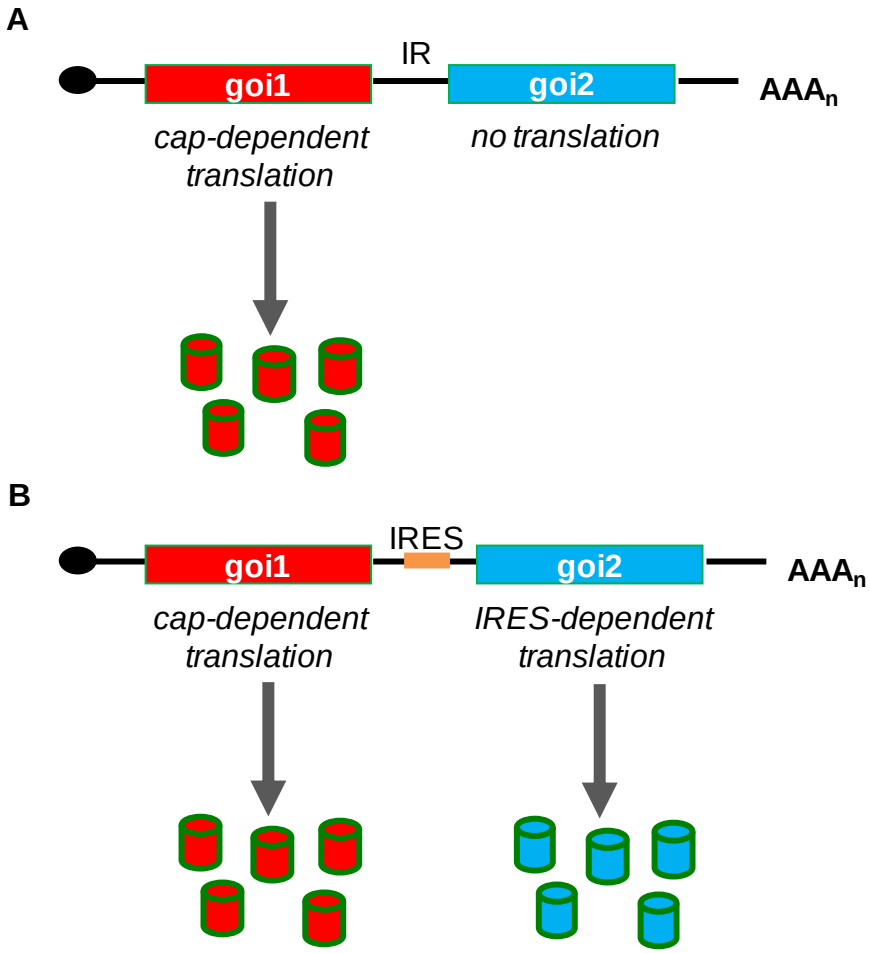


Figure 5.1. A yeast dicistronic vector based on insertion of an internal ribosome entry site (IRES) into the intercistronic region (IR) between the two genes of interest (goi1 and goi2). The expression pattern of the vector from both cistrons is demonstrated in the absence (A) and presence (B) of an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

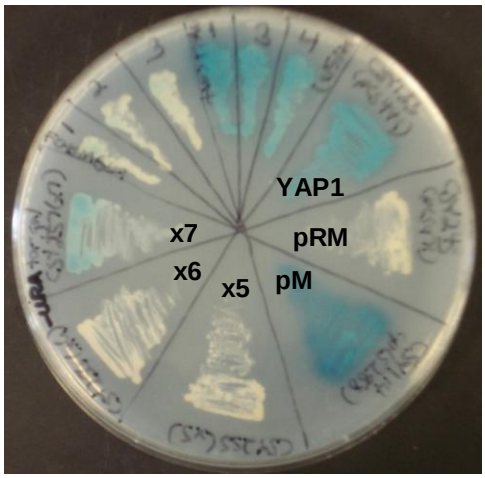


Figure 5.2. Visualization of MEL1 activity on X-α-gal plates. The cleaved products of X-α-gal by MEL1 form a blue color. The strength of the blue color is an indication of the amount of MEL1 protein being translated. Colonies from uracil dropout plates were streaked out on X-α-gal plates and allowed 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).

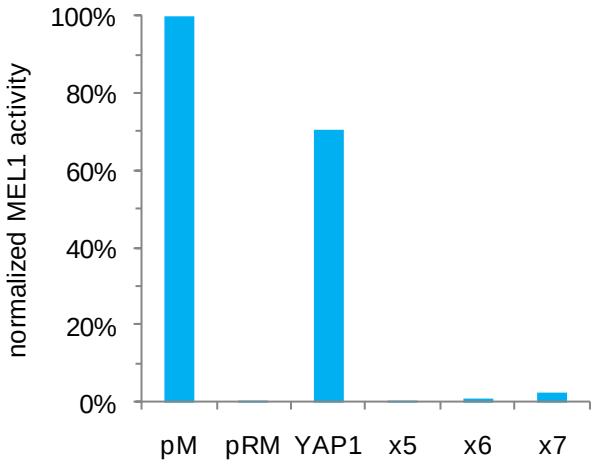


Figure 5.3. Quantification of MEL1 activity of constructs baring IRES modules. Spectrophotometry is used to measure the yellow cleavage product of PNPG caused by MEL1 activity. Activities are normalized to the cap-dependent control, pM. Cultures

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).

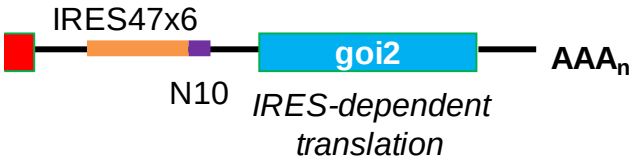


Figure 5.4. Proposed design for selection of a 10-nt IRES library in a dicistronic vector. Six modules of the IRES47 is positioned upstream of a 10-nt randomized region. Translation initiation mediated through the IR drives expression of the second cistron, goi2.

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.

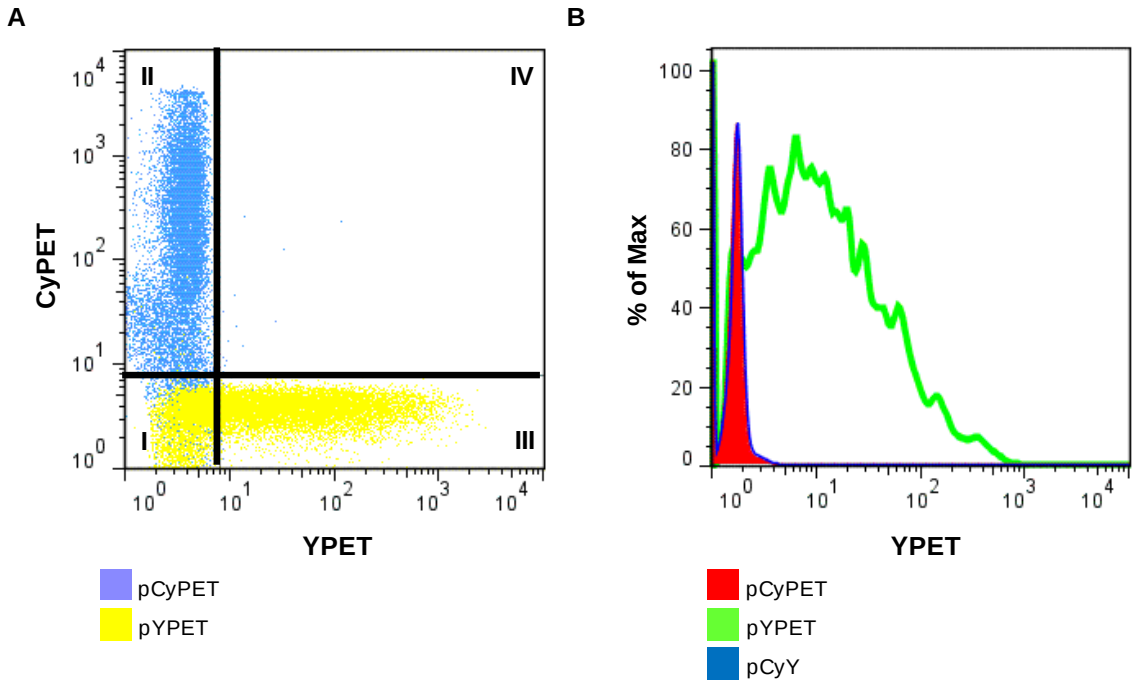


Figure 5.5. Expression profiles for the CyPET- and YPET-based constructs. (A) The monocistronic controls for CyPET and YPET correctly express their respective protein. pCyPET served as the negative control for the gating of YPET expression. pYPET served as the negative control for the gating of CyPET 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:

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

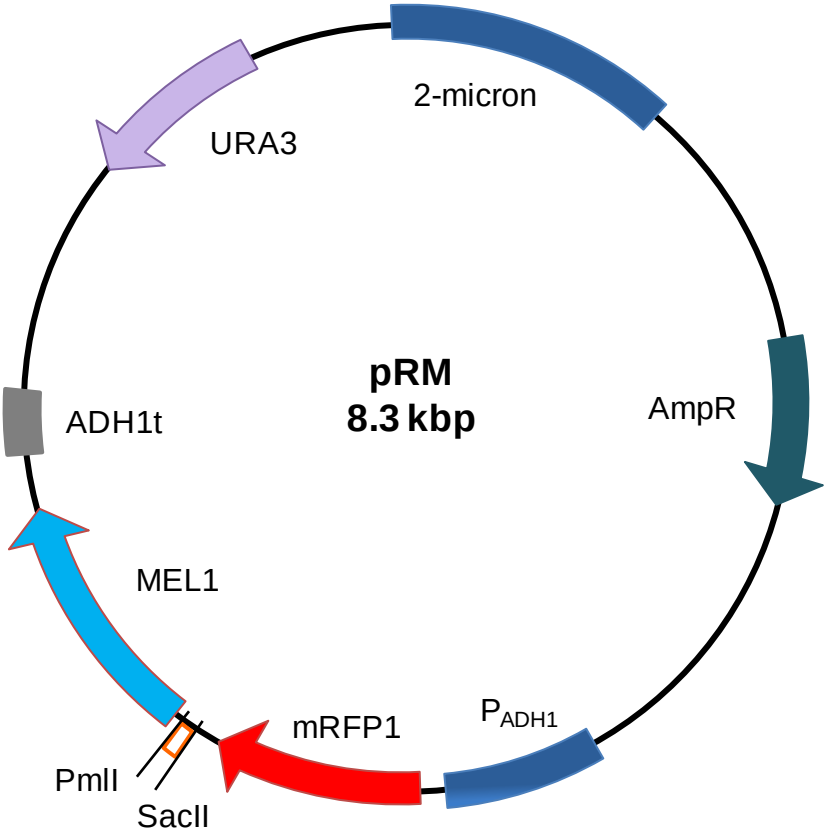
Only samples with OD_{400} values less than 3.0 are valid.

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).