

GENERATING TEMPERATURE SENSITIVE INTEINS FOR
STUDYING GENE FUNCTIONS

A Thesis
presented to
the Faculty of the Graduate School
at the University of Missouri

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
MING CHEN
Dr. Change Tan, Thesis Supervisor

December 2008

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

GENERATING TEMPERATURE SENSITIVE INTEINS FOR
STUDYING GENE FUNCTIONS

presented by Ming Chen,

a candidate for the degree of Master of Arts,

and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Change Tan

Dr. Karen Cone

Dr. Troy Zars

Dr. Karen Bennett

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Dr. Change Tan for her guidance, support, and instruction throughout my work in her lab; without her kindness and understanding, this thesis could never be accomplished.

My sincere thanks go to Drs. Troy Zars, Karen Cone, and Karen Bennett for serving on my graduate committee, and also Dr. Anand Chandrasekhar for his advice and instruction during my graduate study.

I am also grateful to the other members of our lab. Dr. Guihong Tan for his instruction and help all through the experiments, Mr. Sengkai Ong for his advice and friendship in the last one and a half years, and also our former member Dr. Chris Foote for his helpful advice on the experiments.

Thank all my friends in Columbia for their companionship and advice on my life and career development; they brought much fun into my life.

Besides, I would like to thank my parents for their love, patience, support and encouragement throughout my pursuit of this degree.

Lastly but not least, I would like to give my unending thanks to my beautiful wife Liyi Yang for her love, support and advice.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF TABLES	v
LIST OF FIGURES	vi
ABSTRACT	vii
Chapter 1: INTRODUCTION	1
1.1 Overview.....	1
1.2 Genetic strategies for generating mutants and studying gene functions	2
1.2.1 Gene targeting.....	2
1.2.2 Dominant negative mutants	5
1.2.3 RNA interference	6
1.2.4 Temperature-sensitive mutants	8
1.3 Protein splicing and intein	11
1.3.1 Definitions and history.....	11
1.3.2 Intein nomenclature	12
1.3.3 Intein distribution.....	13
1.3.4 Intein classification and conserved motifs.....	13
1.3.5 Molecular mechanism of protein splicing.....	18
1.3.6 Endonuclease activity of inteins	20
1.3.7 Intein function and application	21
1.4 Reporter systems: GFP and Gal4	22
1.4.1 GFP	22
1.4.2 Gal4.....	24
1.5 Goals of this research	25
Chapter 2: GENERATION OF NEW TEMPERATURE-SENSITIVE INTEINS ..	26
2.1 Introduction	26
2.2 Materials and methods.....	27
2.2.1 Error-prone PCR.....	27
2.2.2 Cloning pS5-Gal4-Intein.....	29
2.2.3 Identifying temperature-sensitive mutants.....	30
2.2.4 Shifting “ <i>gal4-intein</i> ” from pS5 to pU.....	32
2.2.5 Eliminating “leaky” TS alleles by pU-Gal4-Intein yeast growth	33
2.3 Results	33
2.3.2 Screening for temperature-sensitive inteins.....	34
2.3.3 Eliminating not tightly controlled TS alleles.....	37
2.4 Summary	45
Chapter 3: ANALYSIS OF TEMPERATURE-SENSITIVE INTEIN SPLICING..	46
3.1 Introduction	46
3.2 Materials and methods.....	47
3.2.1 Constructing pMet-Gal4-intein.....	47

3.2.2 Protein expression, splicing, and extraction	48
3.2.3 Western blot	50
3.2.4 Constructing pMet-eGFP-intein.....	50
3.2.5 eGFP rotein expression, splicing, extraction, and western blot analysis	52
3.3 Results	53
3.3.1 Gal4-Intein splicing	53
3.3.2 eGFP-Intein splicing.....	56
3.3.3 Time course analysis of eGFP-Intein splicing	58
3.4 Discussion	59
Chapter 4: CONCLUSIONS AND FUTURE STUDIES	61
REFERENCES.....	66
APPENDICES.....	76
6.1 Appendix 1: media for yeast and bacterial culturing	76
6.2 Appendix 2: Solutions and buffers.....	80
6.3 Appendix 3: Detailed protocols for experiments	83
6.3.1 Preparation of plasmid DNA from bacteria.....	83
6.3.2 Rescue of yeast plasmids into <i>E. Coli</i>	84
6.3.3 Yeast transformation.....	85
6.3.4 Protein extraction from yeast for western blot analysis.....	86
6.3.5 Western blot.....	87
6.3.6 Bacterial transformation.....	89
6.3.7 Electro-transformation	89
6.4 Appendix 4: Amino acids abbreviations and linear structure formula.....	90
6.5 Appendix 5: PCR primers list	91

LIST OF TABLES

Table 1.1 Comparison of different methods for studying gene functions	11
Table 1.2 Conserved amino acids in inteins	18
Table 2.1 Grouping intein alleles based on growth profile.....	41

LIST OF FIGURES

Figure 1.1 Schematic structures of intein motifs and conserved residues	15
Figure 1.2 Inteins and protein splicing	19
Figure 2.1 Cloning scheme of pS5-Gal4-Intein.....	28
Figure 2.2 Error-prone PCR to introduce mutations into VMA intein sequence	34
Figure 2.3 Temperature-sensitive intein selection	35
Figure 2.4 Serial dilution assay to confirm the TS splicing.....	36
Figure 2.5 Shifting intein from pS5 to pU	38
Figure 2.6 Eliminating not tightly controlled intein alleles	39
Figure 2.7 Generating and identifying new TS inteins	40
Figure 2.8 Yeast growth profile of pU-Gal4-Intein transformed into FY761	41
Figure 2.9 Point mutations of TS inteins.	44
Figure 3.1 Protocol for protein expression and splicing	49
Figure 3.2 Cloning scheme of pMet-eGFP-Intein	52
Figure 3.3 Gal4-Intein splicing at different temperatures.....	54
Figure 3.4 Comparison of different assay methods for S42	56
Figure 3.5 eGFP-Intein splicing at different temperatures	57
Figure 3.6 Comparison of different assay methods for S36	58
Figure 3.7 Time course analysis of intein splicing	59

ABSTRACT

Reversible and easy to use, temperature-sensitive (TS) mutations, are powerful tools to study the functions of all genes, including essential and pleiotropic ones. However, the rarity of TS alleles and the difficulty of generating and identifying them have limited their use. A novel approach to generate TS mutations efficiently utilizes a conditionally-splicing intein, an “intein switch”. Unmodified, an intein, which is a self-splicing protein sequence embedded in-frame within a precursor protein, excises itself precisely from the precursor protein while simultaneously ligating the flanking sequences (exteins). The intein switch is a TS-intein that splices itself only at the permissive temperature (18°C) to generate a wild-type host protein. At the non-permissive temperature (30°C), it fails to splice and remains within the host protein, leading to loss-of-function of the host protein.

To extend the versatility of the TS-intein switch, we designed a scheme to look for intein switches that would function at different temperature ranges and that would splice with a high efficiency. We have generated and characterized 44 TS intein alleles. The permissive temperatures of these TS-intein switches differ from 18°C to 30°C. They function similarly in different host proteins. This makes it possible to choose a TS-intein switch according to the specific temperature determined by the optimal growth of an organism or a special experimental design.

CHAPTER 1

INTRODUCTION

1.1 Overview

In the post-genomic era, one important task is to annotate the biological functions of each gene encoded in the genomes of different organisms. The extraordinarily large amount of genes to be studied, together with the fact that a significant number of them are either essential for viability, or function at multiple places and times during development, demands methods that can generate conditional mutations efficiently and that can be readily adapted in different organisms.

About one in four genes is essential for the viability of an organism for many species. For example, the *D. melanogaster* genome has 14,066 protein coding genes (<http://flybase.net>), and 3,600 (24%) of these genes are indispensable for the viability of flies (Miklos 1996). Similar percentage of genes are required for the viability of other organisms, including mice, worms, and yeast.

Pleiotropy is the phenomenon that one gene product (RNA or protein) functions at different locations in an organism and/or different times during the life cycle of the organism. *D. melanogaster* provides a good example of pleiotropy, since almost every gene studied is expressed in more than one place or at more than one developmental stage. Of the nearly 600 randomly selected enhancer trap lines expressed in the *D. melanogaster* larval brain, only two are found exclusively expressed in the nervous system (Datta 1993).

To define the functions of pleiotropic and essential genes, we need to reversibly turn a gene "on" and "off" at a particular time and within specific cell types. The extraordinary large amount of genes to be studied also requires the tool can be applied to the genome-wide scale. Unfortunately, there is no ideal method available yet to meet all these requirements.

We have developed and optimized a novel approach that can be used to eliminate gene function conditionally. We used a temperature-sensitive (TS) splicing intein, which splices and generates functional host proteins only at permissive temperature. At non-permissive temperature, it can not be spliced out and remains inserted in the host protein, thus renders the host protein nonfunctional. The TS intein switches can be used to conditionally control gene function by shifting between temperatures. In addition, this system can be applied universally in large scale functional gene annotation.

1.2 Genetic strategies for generating mutants and studying gene functions

1.2.1 Gene targeting

Since the generation of the first knock-out mouse (Doetschman 1987; Thomas 1987) about two decades ago, gene targeting has become a very powerful and routine technique for functional genetic studies. It is sometimes referred as site-specific homologous recombination, the principles of which were first developed in the 1970s in yeast (Moore 1975). In yeast, most recombinations between ectopic vector DNA and genomic DNA occur by homologous recombination, as opposed to the random integration that occurs in mammalian systems. Homologous recombination in mammalian systems was first achieved by Smithies et al in 1985 (Smithies 1985), this group successfully inserted DNA sequences into the human chromosomal beta-globin

locus by homologous recombination. What essentially led to the application and expansion of homologous recombination technique to mice was the successful derivation of pluripotent embryonic stem cell lines from mouse blastocysts in the early 1980s (Evans 1981; Martin 1981). This technique paved the way to modify the mouse genome by homologous recombination in embryonic stem cells, and it was first successful for the selectable hypoxanthine phosphoribosyl transferase (HPRT) gene locus (Doetschman 1987; Thomas 1987). Shortly after that, due to the development of the enrichment strategies for correct homologous recombination, targeting of non-selectable genes such as *int-2* and *c-abl* became possible (Mansour 1988; Schwartzberg 1989). Using flanked homologous "arm"s, gene targeting has become a very powerful and informative tool for dissecting gene functions. To study gene functions by gene targeting, one introduces some extra DNA modifications to prevent the formation of a functional gene product that is of interest. For instance, one can introduce point mutations in some amino acids that are critical for the protein function, or delete one exon that is essential to the biological function of the peptide of interest (Tronche 2002). The phenotype is then examined and compared with wild-type individuals; the differences are explained by the loss of function effect of the gene being targeted.

Traditional gene targeting is problematic when we want to study genes that are indispensable for the viability of that organism, as the individuals may not survive to the stage we need in order to study the phenotypes caused by that gene mutation. The development of conditional gene targeting made the study of essential genes possible; it also allowed gene activities to be controlled in a time- and tissue-dependent manner. This is especially helpful for studying the functions of pleiotropic genes in a specific tissue or

in a certain developmental stage. The Cre-LoxP system has been widely used to switch on/off gene expression in a given cell population or for a certain time period (Lakso 1992; Orban 1992; Gu 1994; Kwan 2002). Cre is a 38 kDa recombinase protein derived from bacteriophage P1, which mediates intramolecular (excisive or inversional) and intermolecular (integrative) site specific recombination between LoxP sites without the need for accessory factors (Sauer 1993). The LoxP site is a 34bp DNA segment, which consists of two 13-bp inverted repeats flanking an asymmetric 8-bp core region that offers the loxP site directionality. The manner of recombination depends on the orientations of the two LoxP sites flanking a DNA fragment; inverted LoxP sites will cause an inversion, whereas LoxP sites in direct orientation will cause a deletion. Using the Cre-LoxP system, the recombination of the DNA fragment flanked by LoxP sites can be spatially and temporally controlled by regulating the expression of the enzyme Cre (Lakso 1992).

Twenty years of gene targeting techniques have revolutionized the field of mouse genetics and allowed the analysis of gene functions in the context of the whole animal. The temporospatial control of Cre expression has allowed the conditional targeting of genes of interest, which is very valuable for investigating gene functions in a specific tissue or at a certain developmental stage. Some concerns arise with using the Cre-LoxP system to conditionally control gene functions. First, this method is irreversible. Second, the sensitivity of the LoxP target to recombination may vary from locus to locus (Vooijs 2001). A gene locus can become very insensitive to recombination due to methylation (Rassoulzadegan 2002)). Third, the recombination may occur at a stage that is not designed for or wanted (Kellendonk 1999).

1.2.2 Dominant negative mutants

The expression of dominant negative mutant forms of a gene of interest is a very popular method used in the current study of gene functions. A dominant negative mutant functions at the protein level. Dominant negative mutation involves the manipulation of the cloned gene to encode mutant polypeptides that when over expressed can disrupt the activity of the wild-type protein. Such inhibitory variants of a wild-type product can be designed because proteins have multiple functional domains that can be mutated independently: for instance, domains for oligomerization, substrate-binding region, sites for catalysis, membrane association, and so forth. For example, if a protein is multimeric, a derivative capable of interacting with wild-type polypeptide chains but defective in other biological activities will be inhibitory if it causes the formation of non-functional multimers. A very classic way of achieving the dominant negative effect is to introduce mutations that can switch a transcriptional activator into a repressor; this can be accomplished if the mutation leaves the DNA-binding domain intact, but destroys the activity of the transcriptional activation. For example, the dominant negative form of the yeast GCN4 protein lacks the domain necessary for transcriptional activation and blocks the function of the wild-type protein (Hope 1986). The expression of these derivatives can be put under the control of a regulated promoter, such as the glucocorticoid response element (GRE) of mouse mammary tumor virus, which confers inducibility by glucocorticoids (Chandler 1983). Alternatively a cell-type-specific regulatory element has been used (Swift 1984).

The application of dominant negative mutant is very useful for studying gene function, but it is hard to use it for genome-wide level studies. It is unrealistic and

impossible to design dominant negative forms for every gene product in the whole genome.

1.2.3 RNA interference

RNA interference (RNAi), in which a double-stranded ribonucleic acid (dsRNA) inhibits the expression of gene(s) with complementary nucleotide sequences, is a highly conserved mechanism in eukaryotes for regulating gene expressions. The discovery of RNAi was preceded by unexpected outcomes in the work of plant scientists in the early 1990s (Napolic 1990), in which their attempt to generate darker flowers by introducing additional copies of a gene encoding chalcone synthase (the enzyme for flower pigmentation) failed. Instead these extra copies produced even less pigmented flowers, indicating the chalcone synthase was in fact decreased. Actually, both the endogenous genes and the transgenes were down regulated in the white flowers. This phenomenon was first referred to as "co-suppression". After this initial finding in plants, the existence of this phenomenon was observed in other organisms. In 1995, a study with *C. elegans* demonstrated that sense RNA was as effective as antisense RNA in suppressing gene expression (Guo 1995), shortly after that, many similar phenomenon were observed in other organisms and they were termed as post transcriptional gene silencing (Pal-Bhadra 1997; Cogoni 1996; Ratcliff 1997). The landmark work in the discovery of RNAi was done by Craig C. Mello and Andrew Fire in 1998 (Fire 1998). They observed what actually affected the gene expression was dsRNA, not mRNA or antisense RNA, and they named the process RNA interference (RNAi). It was later discovered that small interfering RNAs (siRNAs) mediate mRNA degradation during the RNAi process (Martinez 2004).

Many details of the RNAi process have been uncovered, which opened new avenues for genetic manipulation in cells and organisms. RNAi is nowadays a very popular tool for knocking down gene expression. To use RNAi to study gene function, the first step is to synthesize a double-stranded RNA that has sequence complementary to the gene of interest, and then introduce it into a cell or an organism, where it is recognized as exogenous genetic material. This activates the RNAi pathway and inhibits the expression of the targeted gene (Daneholt 2007). Since RNAi normally do not totally abolish gene expression, it is sometimes referred as "knockdown", to distinguish it from "knockout" procedures where expression of a gene is entirely eliminated. Using RNAi for functional genetic studies, we normally transfect cells with a plasmid that encodes the sequence from which siRNAs can be generated, which will be processed inside cells and initiate the RNAi effect (Brummelkamp 2002). Lentiviral vector systems have been used to produce conditional RNAi (Tiscornia 2004); that is, the activation or deactivation of transcription can be made inducible, and therefore the knockdown effect is also controllable.

Using RNAi to knock down the expression of a gene of interest has also been extended into the global scale, allowing genome-wide loss-of-function screening (Kamath 2003; Boutros 2004; Cullen 2005). Strategies to design a genome-wide RNAi library can be much more complicated than the design of a single siRNA for a specific gene. To overcome this problem, people use computational modeling to help design siRNA libraries as well as to predict the efficiency of gene knockdown (Huesken 2005; Ge 2005). Microarray technology has also been applied to assist the large-scale screening (Janitz 2006; Vanhecke 2005). Although using RNAi for genome-wide screen is

relatively fast and easy, it also has several disadvantages, such as the variability and incompleteness of knockdown (Harborth 2003; Gonczy 2000). Also, it has been found that siRNAs almost never fully deplete the target mRNA, and normally several different RNA oligos must be screened before an effective siRNA can be identified. Another aspect of the complexity of using RNAi for functional genetic and genomic studies is the specificity of siRNA to degrade a particular mRNA; siRNAs can recognize and interfere with the expression of mRNAs that are partially homologous with the designated target mRNA, although the efficiency of knockdown effect on off-target mRNAs is normally much less (Jackson 2003). There are reports that siRNAs can tolerate some mismatches between the designed siRNA and the target mRNA to be knocked-down (Amarzguioui 2003; Saxena 2003); this complicates the interpretation of phenotypes caused by a knockdown effect of a particular gene, since the effect may be due to the decreased activity of other genes of which one is not aware.

1.2.4 Temperature-sensitive mutants

A TS (or heat-sensitive) mutation can be defined as a conditional mutation that produces the mutant phenotype at one temperature (restrictive or non-permissive), but not at another (permissive) temperature. Such mutations allow reversible ablation of protein function by switching between temperatures. TS mutations have contributed much to our understanding of many biological processes. One classical example is the understanding of cell division from the TS mutations in yeast (Pringle 1975; Yanangida 1998). Crucial molecules mediating cell cycle progression were identified through the use of TS mutants to study the cell division process under two different temperatures. TS mutants were also applied in the study of multi-cellular organisms. One good example is the use of a TS

allele of Notch that was studied in *D. melanogaster* (Shellenbarger 1978; Costa 2005).

The traditional method to generate TS mutations for specific genes uses chemical mutagens, which cause random mutations in a host organism. After mutagenesis, one screens for the desired TS phenotype and identifies the gene causing the phenotypes. This requires screening a large number of progeny; thus it is quite laborious. Another drawback of using this technology for functional genetic studies is that some proteins are not easily to be mutated to TS, since small changes in their amino acid sequences may not ablate the function of those proteins in a temperature-dependent manner (Harris 1991; Suzuki 1971). A more advanced mutagenesis approach is to introduce mutations in vitro. One can clone the gene of interest and randomly mutagenizes it using low-fidelity PCR. The mutagenized DNA is then introduced into a host organism and its ability to cause a TS phenotype is screened in a homozygous background. Still, the frequency of generating TS mutants using this approach is low and therefore it requires screening a large number of progeny. Since both methods introduced above provide a very low frequency of mutations, scientists have been looking for more direct approaches. One way is to replace charged amino acid clusters with alanines. The charged clusters are hydrophilic and therefore not generally incorporated into the hydrophobic cores of proteins, so replacing the charged clusters with alanines (hydrophobic amino acid) generally affects domain-domain interactions, but the overall secondary structure of the protein is not affected, so this protein can possibly act as TS. Using this approach to the actin gene of *S. cerevisiae* resulted in an amazingly 44% of generated mutations behaving in a TS maner in vivo (Wertman 1992). The efficiency of directly mutating specific amino acids yielding TS alleles is therefore much greater than that from random mutagenesis.

A more advanced way to generate TS alleles was reported recently, in which, the protein of interest is fused to an amino-terminal element, called a “degron”. At 37°C, the degron, together with the tagged protein, will degrade via an ubiquitin-mediated "N-end rule" pathway that can recognize aberrantly-folded proteins. At lower temperatures (23°C), the degron does not activate the N-end rule pathway and therefore the fusion protein remains intact and is not subject to degradation (Dohmen 1994; Dohmen 2005; Kanemaki 2003). This approach can be generalized theoretically for generating TS mutants for any proteins, while in fact it turns out not all fusion proteins are able to be efficiently degraded. In Kanemaki's study (Kanemaki 2003), 60% of the 103 essential genes of yeast resulted in loss of viability while the other 40% remain viable even under non-permissive temperature. This suggests that the element degron is not universal in rendering a host protein TS. Also, extending the application of this approach to multi-cellular eukaryotes has not been successful; efficient proteolysis through N-end pathway has not been achieved for degron-tagged proteins in mammalian systems (Levy 1999; Lindner 2002). Another reason which prohibits the application of this approach is the non-permissive temperature, 37°C, is too high for many organisms, to survive for phenotype analysis.

All the methods discussed above are very helpful in dissecting gene functions; however, to use them to interfere gene functions conditionally on the genome-wide scale; each has its drawbacks (Summarized in Table 1.1). We are trying to develop a more widely applicable technique: TS intein, for conditionally eliminating gene function, this technique could presumably be applied to different organisms for every gene on the genome-wide scale.

Comparison of Different Methods for Gene Function Study			
Methods	Conditional Possible?	Genome-wide possible?	Comments and Drawbacks
Gene Targeting	Yes	Not likely	Too laborious
Dominant negative forms	Yes	No	Impossible to generate dominant negative forms for every gene in the genome
Small inhibitors	Yes	No	Impossible to generate for every gene in the genome
Temperature-sensitive mutants	Yes	No	Practically not able to find TS mutants for all genes
RNAi	Yes	Yes	Mismatch is a problem. Gene function “knockdown”, not totally eliminated
TS-Intein	Yes	Yes	Can not be extended to warm-blooded species, as individuals remain constant body temperatures.

Table 1.1 Comparison of different methods for studying gene functions

1.3 Protein splicing and intein

1.3.1 Definitions and history

Protein splicing is a post-translational event that involves the precise excision of an internal polypeptide segment, named intein, with the concomitant ligation of its flanking sequences called exteins (Perler 1994). The term protein splicing was not introduced until 1990 (Kane 1990; Hirata 1990) in the study of the yeast *vma1* gene, which encodes for a subunit of vacuolar ATPase. Anraku's lab (Hirata 1990) first isolated the cDNA for the yeast vacuolar ATPase A-subunit and found the two ends of the cDNA are very conserved when compared with other known ATPase, but the central region had no similarity to any of them. Instead, the central part had some sequence similarity to endonucleases. What was surprising was that SDS-PAGE analysis of the isolated yeast

VMA1 protein had a molecular weight of only 70kDa, the same as expected for the subunit without the central region. This inconsistency was solved by Kane et al (Kane 1990). Their work showed that the central region was still present in the transcribed mRNA, and also translated into protein. But after that, there was a posttranslational processing step that splices the central region out of the precursor protein. They proposed to use the term intein (**internal protein**) for the central region, which was excised; N and C extein (**external protein**) were given for the flanking segments, and protein splicing was the term used to describe the whole process.

1.3.2 Intein nomenclature

Inteins are named after the organism and the name of the host protein in which they are found (Perler 1994, 2000). For example, the above mentioned Vma1 intein from yeast *Saccharomyces cerevisiae* is named Sce VMA, indicating it is from *Saccharomyces cerevisiae* and the name of that intein-containing protein is VMA. When there are more than one intein in a host protein, they are distinguished by Roman numerals: like TerRIR-I, TerRIR-II, etc. Many inteins are bifunctional; they also have endonuclease activity. The product of an intein with endonuclease activity receives a name following the conventions for endonuclease naming (Belfort 1997). In this case, the name of these endonucleases will begin with the prefix "PI" (for protein insert), followed by a three letter species indicator and a Roman numeral to indicate the order of identification of the intein endonuclease in that organism. For instance, PI-MgaI denotes the endonuclease activity of the intein in an ABC transporter from *Mycobacterium gastris* (Mga Pps1) (Saves 2001). Endonuclease PI-SceI is encoded by the intein in the yeast vacuolar ATPase catalytic subunit Sce VMA1.

1.3.3 Intein distribution

As summarized in InBase (<http://www.neb.com/neb/inteins.html>) (Perler 2002), more than 200 inteins have been identified in more than 100 different species and strains, residing in more than 50 various families of host proteins. These inteins are from 128 to 1650 amino acids long and they are found in members of all three domains of life: Eukaryotes, Bacteria, and Archaea, with the overwhelming majority of inteins were found in the genomes of prokaryotes, mostly archaeobacteria and eubacteria. Inteins are found in proteins with diverse functions, including metabolic enzymes, DNA and RNA polymerases, ribonucleotide reductases, proteases, and the ATPase, but with a special preference in enzymes involved in DNA replication and repair (Liu 2000). Strangely, the phylogenetic distribution of inteins is sporadic, that is, we can not predict the presence of an intein within an extein based on the existence of intein in a particular gene from a closely related species or strain (Perler 2002).

For the position of inteins within the host proteins, it is found that inteins appear to prefer conserved regions, for instance, nucleotide-binding domains (Pietrokovski 2001). There are exceptions to this rule. For example, the ATPase and the replication factor C intron insertion sites that are not restricted to the conserved parts of the host protein.

1.3.4 Intein classification and conserved motifs

Inteins are conventionally divided into two large groups: classical inteins and mini-inteins. Most inteins belong to the former group. Classical inteins, normally called large inteins, consist of two domains: a self-splicing domain and an endonuclease domain (Liu 2000). The N- and C-terminal regions of the large inteins contain the essential

elements for splicing. These terminal regions are therefore called splicing domains. The endonuclease domain is not essential for protein splicing, since deletion of this domain from the large intein does not affect protein splicing (Derbyshire 1997; Chong 1997). The endonuclease domain in the large inteins is responsible for the spread of inteins. In mini-inteins, however, instead of having a central endonuclease domain, they contain a linker sequence that does not have any catalytic activity.

By comparing the amino acid sequences from different inteins, it is observed that most large inteins consist of ten conserved motifs: A, N2, B, N4, C, D, E, H, F, and G. These motifs are structurally heterogeneous. Mini-inteins lack the central motifs C, D, E, and H (Figure 1.1).

The Ser and Cys in motif A, the His in Motif B, the His, Asn and Ser/Cys/Thr in motif G are the most conserved residues in the splicing motifs. Many of the conserved residues are essential for intein splicing or endonuclease activity. Point mutations of those amino acids abolish the corresponding activities. Mini-inteins can be as small as 134 amino acids, but still are able to be spliced out from the precursor proteins (Evans 1999), confirming the endonuclease domain is not essential for intein splicing. On the other hand, the N-terminal splicing region and the C-terminal splicing region are necessary for intein splicing (Duan 1997; Ichiyanagi 2000; Perler 1998).

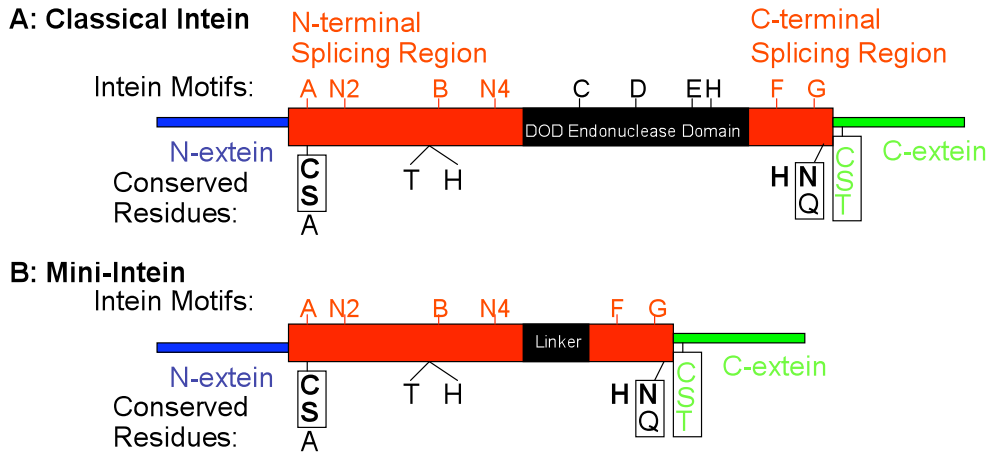


Figure 1.1 Schematic structures of intein motifs and conserved residues

(A) Classical Intein; (B) Mini-Intein

Modified from: Perler, F. B. (2002). InBase, the Intein Database. *Nucleic Acids Res.* 30, 383-384.

Motifs A, N2, B, and N4 constitute the N-terminal splicing domain, of approximately 150 to 200 amino acids. The function of the N-terminal splicing domain is to facilitate the disruption of the peptide bond at the N end of the intein (Pietrokovski 1998).

Motif A is the most N-terminal segment of inteins. It consists of 13 residues, two of which, the first and the last, are highly conserved. It is suggested that these two conserved amino acids are essential for the initiation and completion of intein splicing. The first amino acid is almost always a Cys, with very rare cases an Ala, Gln, or Ser. The last amino acid of the motif A is mostly a Gly, or, in extremely rare cases, an Ala, Lys, Thr, Arg, Tyr, or Asn (Perler 1997; Pietrokovski 1994; Perler 2002).

Motif N2 consists of 7 amino acids, of which the fifth is the most highly

conserved and is either Asp or Glu. Very often the fourth amino acid is a Gly (Petrokovski 1998; Perler 2002).

Motif B has 14 residues; the 10th position is always occupied by a His in all the identified inteins, while the 7th residue is mostly Thr. These two conserved amino acids are involved in the initiation of protein splicing (Perler 1997; Petrokovski 1994; Perler 2002).

Motif N4 is found in most inteins (although Sce VMA does not have a motif N4). Motif N4 consists of 16 residues, of which the 11th is very conserved; it is either an Asp or Glu, and is usually preceded by a Gly (Petrokovski 1998; Perler 2002).

Motifs C, D, E, and H form the DOD endonuclease domain; however, it should be noted that the DOD domain is not necessary for protein splicing (Guhan 2003; Ding 2003; Perler 2002). Besides the common DOD motifs within the core endonuclease domain, some specific inteins have additional segments. For example, the Sce VMA intein also has a DNA recognition region (DRR) before the DOD motifs (Duan 1997). Another additional domain is found in the Pfu-RIR1-1 intein; there is an extra domain inserted between the DOD motif and the C-terminal splicing domain (Ichiyanagi 2000).

Motifs C and E are parts of the DOD endonuclease domain (Matsumura 2006). The DOD endonuclease is one of four classes of homing endonucleases that are defined by their conserved sequence motifs (Belfort 1997; Jurica 1999). Like other known DOD endonucleases (also known as dodecapeptide or LAGLIDADG), these motifs have sequences of 9 to 10 residues, which form the DNA recognizing center and are separated by a linker (motif D, discussed later) of about 90-130 residues. Some conserved Gly residues are found at position 3 and 9 of motif C and at position 4 and 10 of motif E

(Perler 1997, Perler 2002).

Motif D is in the linker between motif C and E. It consists of 8 residues. The Lys at position 2 is essential for the endonuclease activity of the DOD domain. At least in some cases, substitution of the lysine will abolish the endonuclease activity (Perler 2002). This suggests that motif D contributes to the formation of the active endonuclease domain.

Motif H is a 19 amino acid segment that follows motif E, Leu is found rather conserved at position 13 and 14 in both. It is suspected that these Leu residues are involved in the intein-DNA interaction process (Petrokovski 1994; Perler 2002).

Motifs F and G: These two form the C-terminal splicing domain, which are 25-40 amino acids in length (Petrokovski 1998; Perler 2002). Motif F consists of 16 residues, most of them are highly conserved (Table 2). Motif G is an 8-amino acid fragment, 7 of which belong to intein, while the other one is the N-terminal residue of C-extein. There is a linker (normally 2 to 5 amino acids) that separates the motifs F and G. The last two amino acids (with the last one a very conserved Asn, and the second to last a His) of motif G play essential roles in hydrolyzing the peptide bond at the C end of the intein (Petrokovski 1998; Perler 2002), whereas the first amino acid of the C-extein (very conserved, either Ser, Thr, or Cys) is critical for extein ligation.

Table 1. 2 summarizes the conserved amino acids of inteins.

Domain	Motif	Amino acid sequence	Legend
N-terminal splicing domain	A	Ch..Dp.hhh..G	C, cysteine; D, aspartic acid; G, glycine; T, threonine; L, leucine; K, lysine; I, isoleucine; P, proline; F, phenylalanine; S, serine; V, valine; Y, tyrosine; H, histidine; N, asparagine h, hydrophobic residue (G, V, L, I, A, M) a, acidic residue (D, E) r, aromatic residue (F, Y, W) p, polar residue (S, T, C) . , nonconserved residue p ⁺¹ , residue +1 of the C-extein (S, T, C) ▲, C-terminal cleavage site
	N2	...GD..	
	B	G..h.hT..H.hhh	
	N4GD.....	
Endonuclease domain	C	LhG..hhaG	h, hydrophobic residue (G, V, L, I, A, M) a, acidic residue (D, E) r, aromatic residue (F, Y, W) p, polar residue (S, T, C) . , nonconserved residue p ⁺¹ , residue +1 of the C-extein (S, T, C) ▲, C-terminal cleavage site
	D	.K.IP..h	
	E	.L.GhFahDG	
	H	p.S..hh..h.LL..hGI	
C-terminal splicing domain	F	rVYDLpV..a..HNF	r, aromatic residue (F, Y, W) p, polar residue (S, T, C) . , nonconserved residue p ⁺¹ , residue +1 of the C-extein (S, T, C) ▲, C-terminal cleavage site
	G	h NGhhhHN▲ p ⁺¹	

Table 1.2 Conserved amino acids in inteins

Modified from: Starokadomskyy, P.L. (2007) Protein splicing. Mol Biol 41(2): 314-30.

1.3.5 Molecular mechanism of protein splicing

Protein splicing is an autocatalytic process that does not require any additional cofactors or enzymes (Anraku 2005; Noren 2000; Perler 2002). However, the detailed mechanism mediating this process remained unknown for a long time because the splicing is so rapid that the precursor protein as well as the intermediates is not easy to be detected and studied. This problem was solved when it was found that substitution of some of the conserved residuals could block protein splicing at various steps, leading to the accumulation of the splicing intermediates (Pietrokovski 1998; Noren 2000).

The most popularly accepted theory about protein splicing states that protein splicing involves four consecutive nucleophilic displacement reactions, via three conserved splice junction residues (Pietrokovski 1998; Noren 2000, Choi 2006; Clarke

1994). (Figure 1.2)

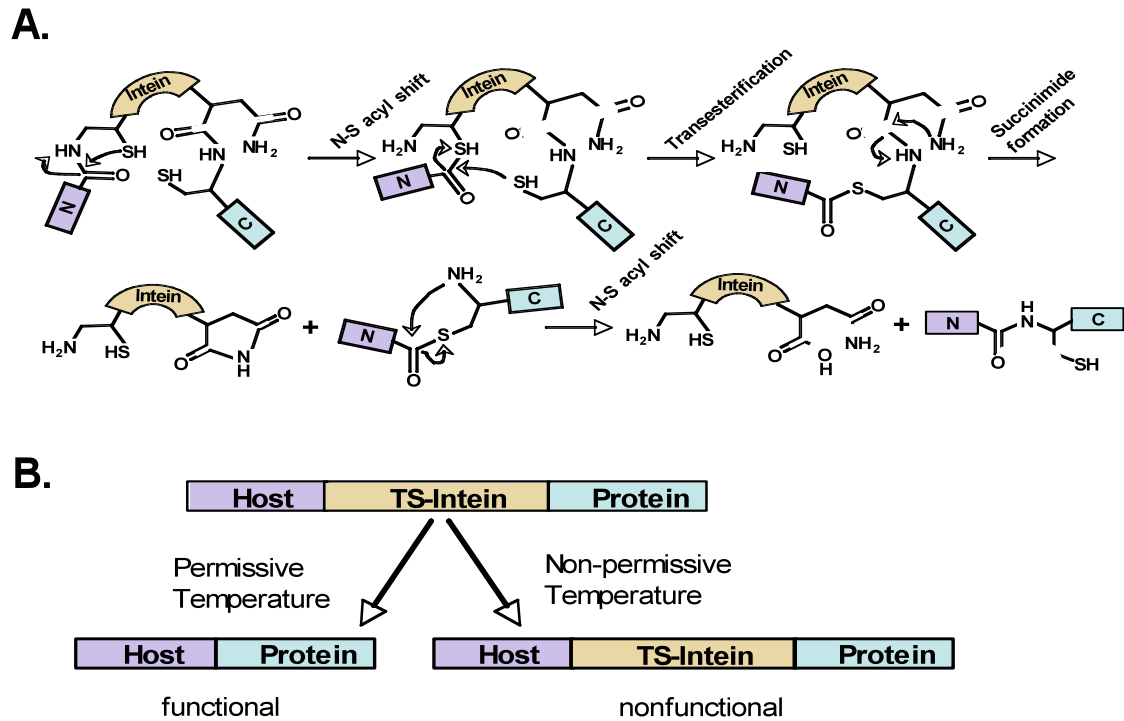


Figure 1.2 Inteins and protein splicing

(A) mechanism of protein splicing; (B) principle of TS-intein function.

Step 1: N-O/S Shift

The trigger of intein splicing is an autocatalytic N-O or N-S shift that occurs at the N-terminal splicing site, which involves the first N-terminal residue of the intein. This autocatalytic reaction moves the N-extein to the hydroxy/thiol side chain of the Ser/Thr/Cys, and results in a very reactive ester or thioester bond. This N-O/S shift is not easy from a chemical energy perspective, since the peptide bond is disrupted to form a highly reactive ester/thioester bond. This is possibly facilitated by the conformational advantage provided by the N-extein and the first amino-acid residue of the intein, which renders the original peptide bond itself energetically disadvantageous, thus the splicing can be initiated (Klabunde 1998).

Step 2: Transesterification

The hydroxy/thiol group of the first residue of C-extein attacks the resulting ester/thioester bond from step 1, which leads to the cleavage of the N-terminal splicing junction, and the N-extein is transferred to the side chain of the first residue of the C-extein. The second step yields a branched protein with two N ends.

Step 3: Asn cyclization

The branched intermediate formed in the second step has a charge rearrangement that tends to induce the cyclization of the conserved Asn at the C end of the intein, yielding succinimide (Xu 1994). The cyclization then leads to the disruption of the peptide bond between the intein and C-extein, leaving the free intein and the exteins linked via a ester/thioester bond.

Step 4: O/S-N shift

As mentioned before, the ester/thioester bond from step 3 is energetically disadvantageous, so the reverse of N-O/S shift occurs and yields a native peptide bond between two exteins.

The mechanism of intein-mediated protein splicing has been extensively studied. There are several detailed reviews describing this intricate process (Perler 1998, Paulus 2000, Ding 2003).

1.3.6 Endonuclease activity of inteins

Most inteins (except mini-inteins) have endonuclease activity, which transfers the intein-coding sequence from its host gene into host protein alleles that lack the intein sequence (Noren 2000; Yang 2004). This is referred as intein homing, suggesting the similarity with the intron homing (Khan 2005).

Homing endonucleases are grouped into four classes according to the structure of their active centers: LAGLIDADG (or DOD), GIY-YIG, His-Cys, and HNH (Shao 1996). Most inteins are DOD endonucleases with two highly conserved LAGLIDADG motifs in the active center (Perler 2002), each of which resides in one α -helix, and two antiparallel α -helix forming the active center (Mills 2005). The intein endonuclease domains usually recognize nonpalindromic DNA fragments of 14-40 bp (Bakhrat 2004; Ding 2003). Besides the DOD domain, the intein endonuclease activity requires the DNA recognition region (DRR) within the N-terminal splicing domain, since point mutations of particular residues in that region impair the specificity and activity of the DOD domain (Lykke-Andersen 1996; Bakhrat 2004). During the intein homing process, intein first introduces a double-strand break at the insertion site, which produces a 4-bp 3'-overhang (He 1998; Bakhrat 2004), then the intein-coding sequence is used as a template to repair the double-strand (Perler 2002). Mini-inteins (which constitute about 15% of all inteins) lack the endonuclease domain (Perler 2002; Pietrokovski 1998).

1.3.7 Intein function and application

Although intein has been widely and extensively studied, its biological role in cells is unclear. Its function was initially considered as the selfish-DNA (Gogarten 2002). Intein utilizes the regulatory elements of host genes for expression; and the DOD domain ensures its spreading over the genome, like other mobile elements, such as transposons and retrotransposons (Pietrokovski 2001). The difference is that whereas retrotransposons occur in many copies in the genome, intein sequences only exist in certain sites of particular genes. Besides, it is suggested that free inteins (after spliced out from the host proteins) perform some unknown regulatory and enzymatic functions in the cell

(Klabunde 1998). For instance, inteins can regulate the activity of host enzymes. It is reported that enzyme activity can differ due to the presence or absence of an internal intein sequence (Pietrokovski 2001).

Inteins are widely applied in biotechnological processes. For example, inteins are used for protein purification. To do this, the gene of interest is first ligated in frame into the intein tag and, in conjunction with a chitin-binding domain from *Bacillus circulans* as an affinity tag. After purification by adsorption onto the chitin column, the immobilized fusion protein is induced to undergo self-cleavage, resulting in the release of the intact protein of interest (Chong 1997). This method has proved to be able to increase the yield and purity of proteins compared with other protein purification methods, and is commercialized by New England Biolabs (IMPACT™).

1.4 Reporter systems: GFP and Gal4

1.4.1 GFP

The Green Fluorescent Protein was first discovered by Shimomura et al (Shimomura 1962) as a companion protein to aequorin, a chemiluminescent protein derived from *Aequorea* jellyfish. Later on, the emission spectrum of GFP was published and found to peak at 508 nm, whereas the pure aequorin peaked near 470 nm with blue chemiluminescence (Johnson 1962). It was thereafter found that GFP converts the blue emission of aequorin to green; in *Aequorea victoria*, GFP fluorescence occurs when aequorin interacts with Ca^{2+} ions, which induces the blue glow. When some of the luminescent energy is transferred to GFP, the overall visible color is shifted towards green (Morise 1974).

The application of GFP as a tool by biologists did not start until it was cloned and

sequenced by Prasher (Prasher 1992) and then expressed by several labs (Chalfie 1994; Inouye 1994), both in bacteria and nematodes. The GFP molecule fluoresces in the absence of any exogenous cofactors from jellyfish. GFP is composed of 238 amino acids (26.9 kDa), organized in an 11-stranded β -barrel threaded by an α -helix running through the axis of the barrel shell (Ormö 1996; Yang 1996). The chromophore is attached to the alpha-helix and buried in the center of the cylinder structure. This protects the fluorophore from being quenched by the surrounding environment. Several amino acids are required for the formation of the fluorophore, including: Ser65, Tyr66, and Gly67. The inside-facing sidechains of the barrel structure induce the cyclization reaction between these three amino acids and lead to the formation of the fluorophore.

The crystal structure obtained by Ormö and Yang has provided a much detailed understanding of the chromophore formation and the neighboring residue interactions. This also assisted researchers in modifying these residues to produce GFP derivatives, which were prompted by the drawbacks, including poor photo-stability and folding when at 37°C and dual peaked excitation wave lengths of wild-type GFP. To date, many different mutant types of GFP have been generated (Shaner 2005). The first big change was mutating the S65 to T by Roger Tsien (Heim 1995), which increased fluorescence, photo-stability and also shifted the excitation peak to 488, making GFP match the spectral characteristics of FITC filter sets. The second big improvement was the point mutation (F64L) which generated enhanced GFP (eGFP). This improved folding efficiency at 37°C (Shelley 2005). Many other mutant forms of the wild-type GFP have been made, including: blue fluorescent protein, cyan fluorescent protein, and yellow fluorescent protein, each with its specific mutations and unique colors. All of these

variances have been extensively applied in life science research.

The green fluorescent protein (GFP) has become one of the most widely studied and exploited proteins in life science research and beyond. The intrinsic property of its ability to generate a highly visible, efficiently emitting internal fluorophore makes it one of the incomparable tools in biochemistry and cell biology studies. It is not only used as a marker for studying gene expression and protein localization in cells or organisms (Cheng 1996 Rizzuto 1996), but also as a physiological indicator (Stryler 1978) and a biosensor (Romoser 1997).

1.4.2 Gal4

Gal4p (881 amino acids, 99kDa) is a DNA-binding transcription factor required for the activation of the yeast gal genes. Gal4p is inhibited by Gal80p and activated by Gal3p (Bhat 2001). Gal genes, such as *gal1*, *gal10*, *gal2*, and *gal3*, encode proteins required for the transport of galactose into the yeast cell and its metabolism through the glycolytic pathway. In response to galactose, Gal4p activates transcription by recognizing a 17 base-pair long sequence in the upstream activating sequence (UAS) (5'-cggrnrcynyncnccg-3') of these gal genes (Lohr 1995). There are about 300 potential Gal4-binding sites identified in the whole yeast genome. The number of UAS sites and their affinity to Gal4p vary among different gal genes, and therefore cause differential transcriptional activation (Lohr 1995). The UAS itself is sufficient for Gal4-activated gene transcription. It is independent of the genes to be controlled, and does not depend on species specific molecules. This makes Gal4-UAS system widely used for generating inducible ectopic gene expression. The Gal4p consists of two parts: a DNA binding domain and a transcriptional activation domain. These two domains can be separated and

function independently (Lohr 1995), which is the principle behind the yeast two-hybrid technique and many gene transcription studies (Brent 1997).

1.5 Goals of this research

The intein switch is a very promising tool for conditionally controlling gene functions, since the regulation can be easily achieved by shifting between low and high temperatures. A major advantage of this system is that it can be used as a universal tool for genomic scale functional annotation. To do this, we just need to embed the intein sequence in-frame within the genes of interest. Since protein splicing is a very fast process and is modulated at the post-translational level, the lag time for phenotypic response will be very minimal.

Although nine TS alleles have been reported, they have not been fully characterized or optimized; the splicing activities of the TS inteins reported had only been tested at 18°C and 30°C, which is too wide a temperature range. Besides, the existing inteins have “leaky” splicing at non-permissive temperatures.

To facilitate the application of the TS intein system and increase its versatility, we have generated 37 new intein switches that can work at multiple temperature ranges. These intein switches can be applied to different organisms that live under various preferable temperatures. We characterized both the original alleles and the newly identified alleles that have tightly controlled splicing activities; they do not splice at all at non-permissive temperature, a property essential for studying gene functions.

CHAPTER 2

GENERATION OF NEW TEMPERATURE-SENSITIVE INTEINS

2.1 Introduction

To generate more TS inteins, we used low-fidelity PCR to mutate one of the original TS alleles (Zeidler 2004). The idea was to introduce more mutations into the intein sequence and then screen for intein mutants with tighter controllability and functioning at various temperatures.

These further mutated inteins were first cloned into *gal4* gene, and their ability to support the growth of yeast FY760 whose endogenous *gal4* was deleted was tested on galactose medium at both 18°C and 30°C. Galactose provides the carbon source for yeast growth and its metabolism requires functional Gal4p; therefore, the TS inteins we were looking for should support yeast growth at 18°C but not at 30°C. The selected TS alleles were verified by serial dilution assay. To eliminate those alleles that have leaky splicing at non-permissive temperature, we shifted “*gal4-intein*” into a high-copy number (~50 per yeast cell) plasmid pU, and tested for TS splicing in the same way as we did for pS5 (~2 per yeast cell). pU-Gal4-Intein produces more Gal4 transcripts and proteins than pS5-Gal4-Intein, consequently the leakiness can be amplified. This allowed us to identify and eliminate those that do not tightly turn off intein splicing at non-permissive temperatures.

2.2 Materials and methods

2.2.1 Error-prone PCR

Intein fragments were amplified using pS5-Gal4-TS¹⁹ (Zeidler 2004) as a template, using Swap1F and Swap11R as primers (Appendix 5: PCR Primers List) to add sequences homologous to the regions flanking the HindIII recognition sites within the *gal4* gene (HindIII cuts *gal4* gene after the 60th nucleotide). The homologous sequence added is required for yeast homologous recombination to clone the intein DNA fragment into the pS5-Gal4 vector (Figure 2.1). The error-prone PCR differs from normal PCR in recipe constituents; error-prone PCR uses increased amount of MgCl₂, dTTP and dCTP, reduced dATP and dGTP, as well as MnCl₂. This modification increases the chance of mismatch when incorporating nucleotides for double-strand DNA synthesis.

Cloning scheme of pS5-Gal4-Intein

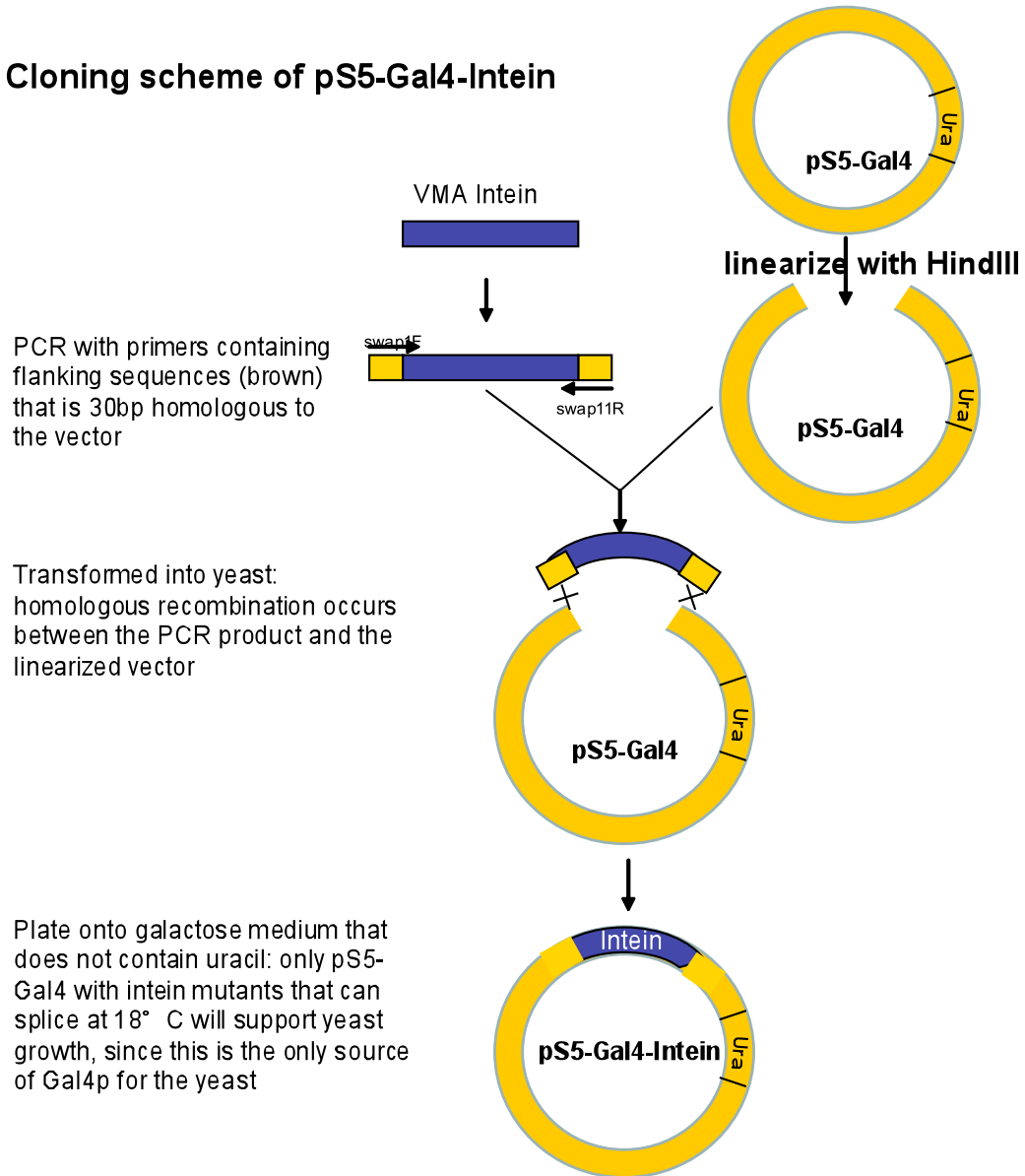


Figure 2.1 Cloning scheme of pS5-Gal4-Intein

Recipe for error-prone PCR:

MnCl ₂ (0.01M)	5 µl
MgCl ₂ (0.1M)	13 µl
Taq	4 µl
dNTP*	10 µl

10X PCR Buffer	10 μ l
Template	0.5 μ l
Primer1 (Swap1F)	6 μ l
Primer2 (swap11R)	6 μ l
ddH ₂ O	45.5 μ l
Total:	100 μ l

Note: the dNTP* mixture used for error-Prone PCR does not have equal amounts of all four nucleotides; instead, we made this special 10X dNTP* as : 2mM dGTP, 2mM dATP, 10mM dCTP, and 10mM dTTP.

The thermocycler program for the PCR reaction is:

Step 1:	94°C	2 minutes	
Step 2:	94°C	20 seconds	} (Step 2, 3, and 4 cycles for 40 times)
Step 3:	55°C	15 seconds	
Step 4:	72°C	30 seconds	
Step 5:	72°C	2 minutes	

Before used for yeast transformation, the PCR products were purified using the Promega Wizard SV Gel and PCR Clean-Up system (Cat. #A9281).

2.2.2 Cloning pS5-Gal4-Intein

We used yeast homologous recombination to clone the PCR product of intein fragment into pS5-Gal4. Briefly, the pS5-Gal4 was linearized by HindIII restriction digestion, which cuts the *gal4* gene after the 60th nucleotide, the “gapped” pS5-Gal4 was then transformed together with the purified error-prone PCR product into the yeast strain *S. cerevisiae* FY760 using lithium acetate technique (Appendix 3: Yeast transformation).

Since the two ends of PCR products have homologous region with the *gal4* gene in the vector, homologous recombination will occur to repair the gap of the vector, forming pS5-Gal4-Intein (Figure 2.1) with the intein inserted into the Gal4 sequence after the 60th nucleotide.

2.2.3 Identifying temperature-sensitive mutants

The pS5-Gal4-Intein yeast transformations were plated on Uracil drop-out selective medium containing 2% galactose as the carbon source (SG-Ura, Appendix 1: Media for yeast and bacterial culturing), and incubated in an 18°C oven. The Uracil can be transcribed and translated from the *ura3* coding sequence within the pS5 vector, which provides a selection marker. The yeast recipient *S. cerevisiae* FY760 we used here is a Gal4-Knock-Out; therefore, without other carbon sources, the yeast growth will need exogenous functional Gal4p to absorb and metabolize galactose. Only yeast with the expected homologous recombination which generates Gal4N-Intein-Gal4C that splices properly at 18°C to form functional Gal4p can grow. By this stage, over 1,000 clones were recovered.

These colonies were then tested for temperature dependent growth; yeast colonies were picked using a toothpick and diluted into 400 µl of sterilized ddH₂O. 2µl of the dilute was dropped onto two identical SG-Ura plates, of which one was incubated at 18°C and the other at 30°C.

Clones that grow at 18°C but not 30°C were selected as TS alleles and were further verified by serial dilution assay. First we cultured 3 ml of yeast with SD-Ura medium (glucose as carbon source) in a 15 ml culturing tube overnight. When the yeast reached the exponential growth stage (O.D.₆₀₀=0.8), the yeast was spun down at 3,000

rpm (Eppendorf, Centrifuge 5415D) for 1 minute, and the pellet was washed twice with ddH₂O. The final yeast pellets were then resuspended using 100 µl of ddH₂O. This concentration was then used as the most concentrated yeast for a serial dilution assay. To reach a 1:10 dilution, 10 µl of the concentrated yeast was added to a 96-well plate with 90µl of ddH₂O. The 1:100 dilution was arrived at by diluting this 1:10 diluted yeast again with a 1/10 dilution. In the same way, three consecutive dilutions gave the 1:1000 dilution. 2µl of each of the yeast serial dilutions were then seeded onto SG-Ura plates that were incubated in ovens set at temperatures of 18°C, 22°C, 25°C and 30°C to monitor yeast growth.

Together with those selected TS inteins we generated, two controls were included for the serial dilution assay. One is the pS5-Gal4 embeded with wild-type (WT) Sce VMA intein (pS5-Gal4-WT), which was directly derived from the intein in the host *vma* gene from yeast *S. cerevisiae*. It splices correctly at all temperatures between 18°C and 30°C, and therefore the yeast growth is temperature independent. The other control we used is pS5-Gal4 disrupted by Dead (Dd) intein (pS5-Gal4-Dd). In contrast to the wild-type intein, the “Dead” intein was created by a point mutation (N454Q) that abolishes splicing activity at all temperature. Yeast bearing pS5-Gal4-Dd therefore can not produce functional Gal4p to support FY760 growth on SG-Ura plates, regardless of the temperature. These two controls served as the marker for splicing and non-splicing respectively. Furthermore, the wild-type control was also used as a timer for different SG-Ura plates; since yeast grows at different rates under different temperatures, when comparing the intein splicing at different temperatures by looking at the yeast growth, the wild-type (seeded on all plates) can serve as a standard marker.

2.2.4 Shifting “*gal4-intein*” from pS5 to pU

We used the “cut and paste” method to shift the intein segment from pS5 to pU, first we recovered the plasmid pS5-Gal4-Intein from yeast FY760 (Appendix 3: Rescue of yeast plasmids into *E. coli*). DNAs directly isolated from yeast are not good for restriction digestion, due to both the quality and quantity concerns. So we transformed the DNA prepared from yeast into bacteria MC1061 using electro-transformation (Appendix 3: Electro-transformation); thereafter we recovered plasmid from MC1061 by a modified alkaline-lysis/PEG precipitation procedure (Appendix 3: Preparation of plasmid DNA from bacteria). These rescued pS5-Gal4-Intein plasmids were then ready for molecular cloning manipulations.

To shift the “Intein” fragment from pS5 to pU, we cut the “Intein” using two restriction enzymes BamHI and XhoI: In a 20 μ l double digestion system, we added 1 μ l of each enzyme, BamHI and XhoI, 2 μ l of NEB Buffer2, 4 μ l (~200ng/ μ l) of DNA plasmid pS5-Gal4-Intein, and added ddH₂O up to 20 μ l. The reaction took 2 hours at 37°C. While digesting the intein from pS5-Gal4-Intein, we also digested the pU-Gal4 vector using the same two enzymes, which would be used as the plasmid backbone for ligation. The intein and pU-Gal4 (both after BamHI and XhoI digestion) were purified using the Promega Wizard SV Gel and PCR Clean-Up System. DNAs were then quantified using the Nanodrop spectrophotometer ND-1000 and then were ready for ligation. In the Fermentas ligation system, we always adjusted the volumes of “Intein” and “pU-Gal4” to make the final molar amount ratio to be 3:1. After 3 hours ligation at room temperature, the ligation products were transformed into *E. Coli* top10 strain (appendix 3: Bacterial transformation.), and the transformations were then plated on LB

plates with ampicillin. Positive clones were screened using PCR and enzyme digestion. We first used Swap1F and Swap11R primers for the first-step screening. The PCR recipe is as following (per 100 μ l): 10X PCR Buffer 10 μ l, MgCl₂ (25mM) 10 μ l, dNTP (2.5mM) 8 μ l, Swap1F (10 μ M) 4 μ l, Swap11R(10 μ M) 4 μ l, Taq 1 μ l, with ddH₂O up to 100 μ l. The thermocycler program used is depicted below:

Step 1:	94°C	2 minutes	
Step 2:	94°C	20 Seconds	} (Step 2, 3, and 4 cycle for 35 times)
Step 3:	55°C	15 Seconds	
Step 4:	72°C	30 Seconds	
Step 5:	72°C	2 minutes	

Together with the newly generated alleles, we also shifted the 9 old ones into pU.

2.2.5 Eliminating “leaky” TS alleles by pU-Gal4-Intein yeast growth

Plasmids pU-Gal4-Intein from confirmed positive clones were then transformed into yeast *S. cerevisiae* FY760, using the lithium acetate technique (Appendix 3: Yeast transformation). For each intein allele transformation, a normal-sized colony grown on SD-Ura plates was picked to test for TS splicing. Using the same protocol as for pS5-Gal4-Inteins, serial dilution assay was performed to compare at the TS intein splicing on SG-Ura plates incubated at 18 °C and 30 °C.

2.3 Results

2.3.1 Error-prone PCR of Sce VMA intein

Before being purified to use for yeast transformation, the error-prone PCR products were checked with a 1% agarose gel, which showed a ~1.4kb band (Fig. 2.2).

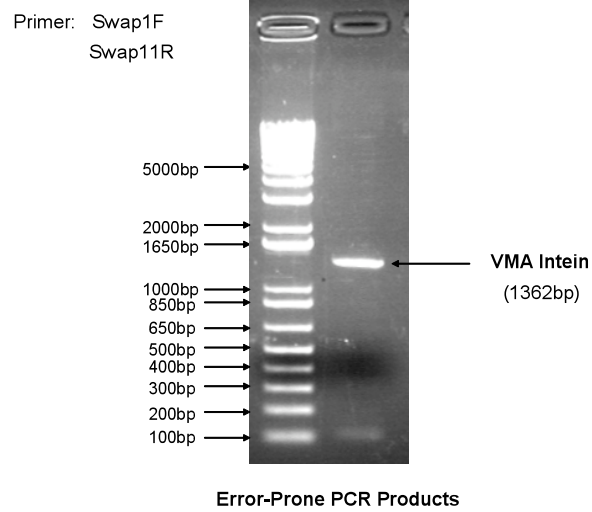


Figure 2.2 Error-prone PCR to introduce mutations into VMA intein sequence

2.3.2 Screening for temperature-sensitive inteins

Yeast cells transformed with pS5-Gal4-Intein were tested for TS growth at both 18°C and 30°C (Figure 2.3). Colonies that grow at 18°C but not 30°C were selected as TS ones for verification using serial dilution assay.

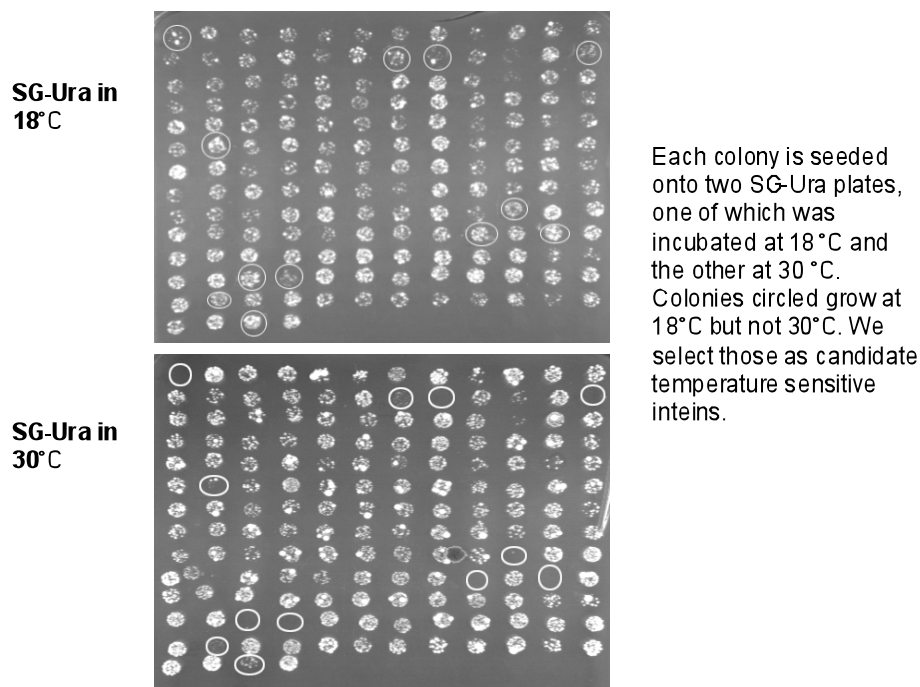


Figure 2.3 Temperature-sensitive intein selection

Note: Each clone was tested for growth at two temperatures: 18°C and 30°C on SG-Ura plates.

Result for serial dilution assay is shown in Fig. 2.3. Most of these selected inteins did support TS yeast growth, suggesting the splicing of precursor protein Gal4-Intein indeed relies on the temperature it is cultured. Also, it should be noted that these inteins switch on/off splicing at different temperature ranges, for example, some alleles grew well at 18°C but not above that, while some others grew normally at 25°C but not when incubated at 30°C. There were also some other alleles that did not grow well in the serial test and were eliminated from our further test.

We tested more than 1,000 colonies and found 148 clones behave temperature-dependent growth. After serial dilution assay, 94 (including 9 old TS alleles we had before) clones were retained as TS ones, by looking for those that grow well at 18°C but

not 30°C.

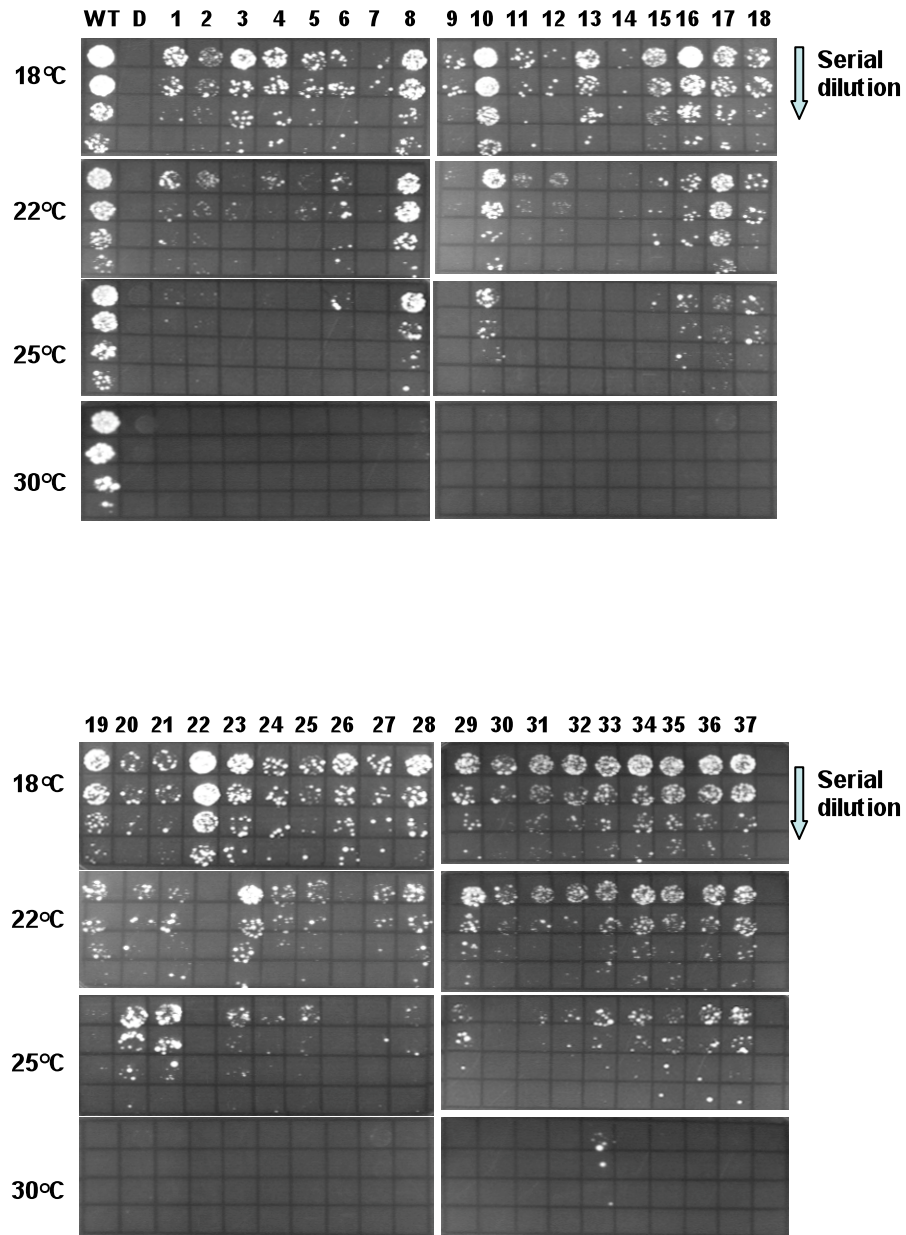


Figure 2.4 Serial dilution assay to confirm the TS splicing

Note: The numbers labeled in this figure are not the name we later assigned for each

allele. At this stage, we just gave each one a number according to the order its serial dilution assay was performed.

2.3.3 Eliminating not tightly controlled TS alleles.

To eliminate those alleles that have leaky splicing at non-permissive temperatures, we transferred “*gal4-intein*” from the low copy number (~2 per yeast cell) plasmid pS5 to the high copy number (~50 per yeast cell) plasmid pU. Briefly, we digested pS5-Gal4-Intein with BamHI and XhoI (Fig. 2.5A). The intein fragment was then purified and ligated with pU-Gal4 digested with the same enzymes. Positive clones were screened by colony PCR (Fig. 2.5B) and enzyme digestion (Fig. 2.5C). Fig. 2.5B shows the agarose gel image for colony PCR of 6 clones picked from pU-Gal4-TS1 ligation transformation. As expected, the water control did not produce any band at the expected size, while the positive control (use DNA pS5-Gal4-Intein as PCR template) generated a very strong signal. Of the 6 colonies tested, 4 have band at expected size (1352bp), although with varying amount of DNA. Plasmids from these 4 colonies were then further checked by enzyme digestion (XhoI, BamHI) for confirmation and the result was then checked by running on 1% agarose gel. Fig. 2.5C suggests all the four clones tested were positive.

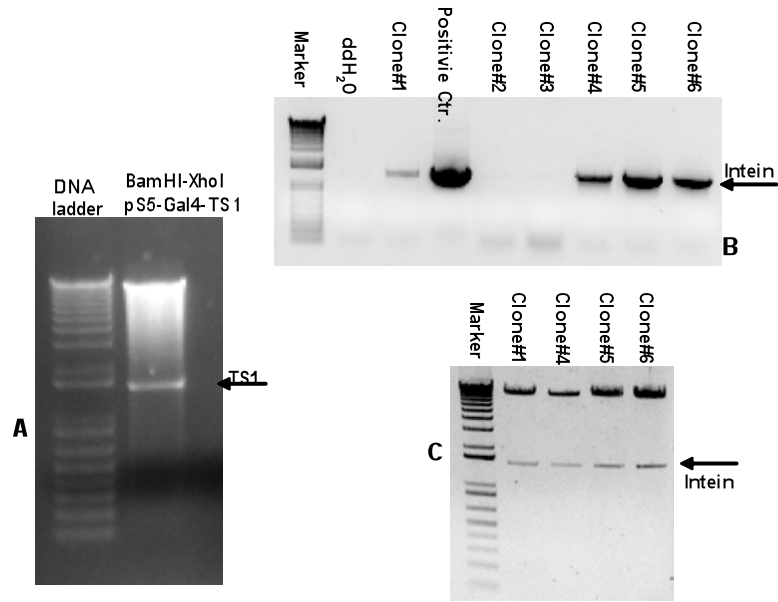


Figure 2. 5 Shifting intein from pS5 to pU

(A) BamH-XhoI digestion of pS5-Gal4-TS1, intein fragment is then gel purified and ligated into pU vector; (B) Colony PCR to screen positive clones, primers: swap1F&swap11R; (C) BamHI-XhoI digestion of plasmid rescued from PCR-positive clones.

Figure 2.6 displays the direct comparison of the growth of yeast with pS5-Gal4-Intein or pU-Gal4-Intein. Some of the intein alleles rendered the yeast to grow differentially in the pU vector, while others did not. Those alleles (in Figure 2.6, S16 and S22) that supported yeast growth at both temperatures were regarded as “leaky” ones. They were not included in further experiments. This leaves us with 48 intein alleles (including 7 old ones).

To identify the nature of the mutations of different intein alleles, we sequenced the intein part in both pS5-Gal4-Intein and pU-Gal4-Intein. For the same alleles, same sequences were obtained with both plasmids. We found that five intein alleles contain the same mutations. Therefore, we actually identified 44 different intein alleles (including the

7 old ones). Figure 2.7 summarizes the steps to identify TS inteins. The number of alleles remained at each step were indicated on the right.

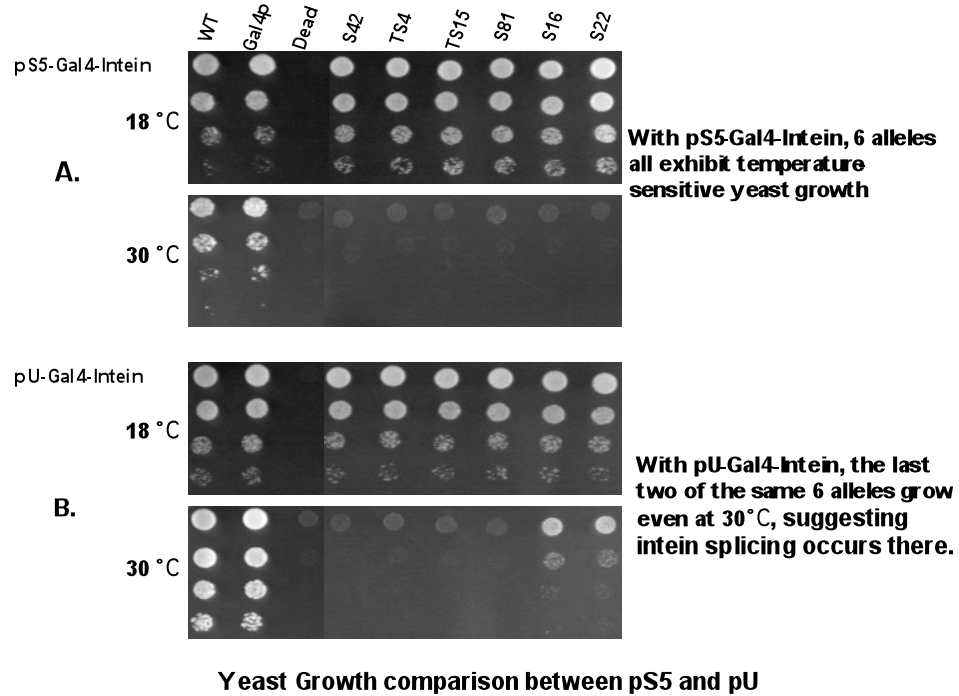


Figure 2. 6 Eliminating not tightly controlled intein alleles

(A) Yeast growth on SG-Ura plate when transformed with pS5-Gal4-Intein

(B) Yeast growth on SG-Ura plate when transformed with pU-Gal4-Intein

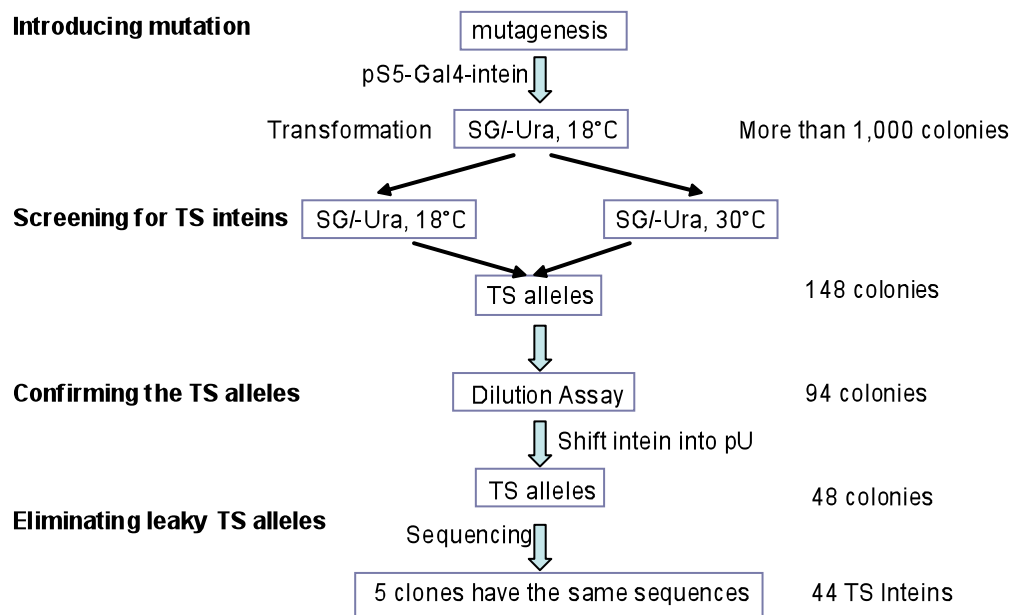


Figure 2. 7 Generating and identifying new TS inteins

We organized these 44 TS intein alleles according to the growth of yeast containing pU-Gal4-Intein, as shown in Fig. 2.8. From the left to right, intein alleles were ordered from lowest permissive temperature (18 °C) to highest (30 °C), except the first three columns, which are controls, pU-Gal4, pU-Gal4-WT, and pU-Gal4-Dd. The first row (with 1:10 serial dilutions) shows the yeast growth on the SD-Ura plate, where glucose was provided as a carbon source; therefore, yeasts with all intein alleles grew with the same rate. From the second row, the yeast were incubated under different temperatures on SG-Ura plates where galactose is the carbon source, from 18°C to 35°C. Old intein alleles were named in the format of “TS” followed by a number; while new intein mutants generated in this work have a name starting with the initial “S” and then the number we assigned after the serial dilution assay of pS5-Gal4-Intein.

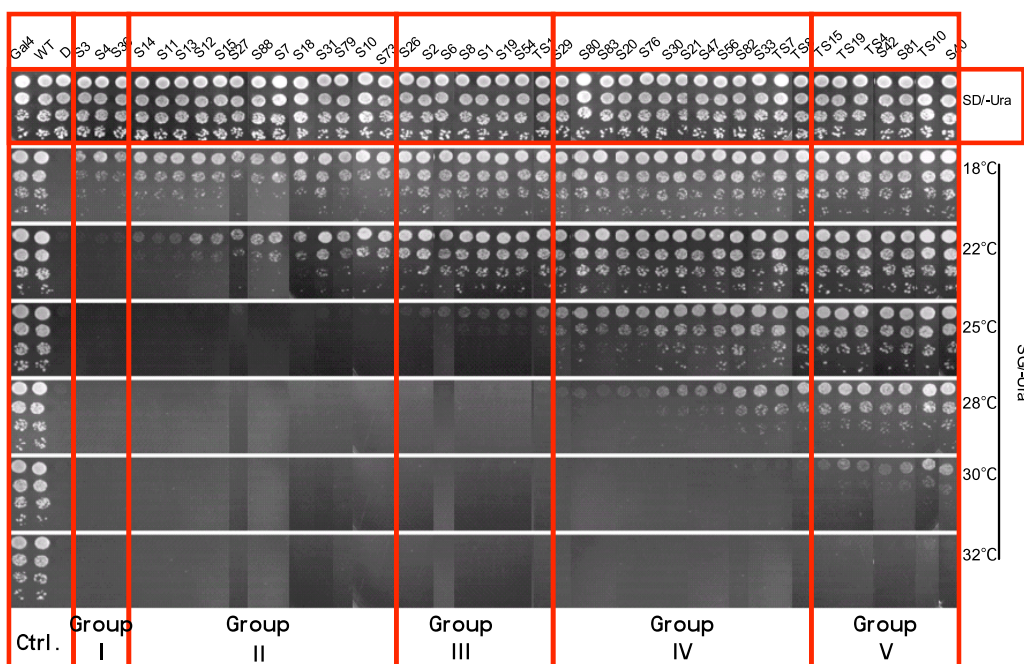


Figure 2. 8 Yeast growth profile of pU-Gal4-Intein transformed into FY761

These different inteins were grouped according to the growth profiles of yeast carrying them (table 2.1). Based on the highest temperature of a TS intein-Gal4 allowing for yeast growth, it is referred to with permissive temperature of 18°C, 22°C, 25°C, 28°C, or 30°C.

Group	Permissive Temperature	Non-permissive Temperature	Members
I	$\leq 18^{\circ}\text{C}$	$>18^{\circ}\text{C}$	S3, S4, S36
II	$\leq 22^{\circ}\text{C}$	$>22^{\circ}\text{C}$	S14, S11, S13, S12, S15, S27, S88, S7, S18, S31, S79, S10, S73
III	$\leq 25^{\circ}\text{C}$	$>25^{\circ}\text{C}$	S26, S2, S6, S8, S1, S19, S54, TS1
IV	$\leq 28^{\circ}\text{C}$	$>28^{\circ}\text{C}$	S29, S80, S83, S20, S76, S30, S21, S47, S56, S82, S33, TS7, TS8
V	$\leq 30^{\circ}\text{C}$	$>30^{\circ}\text{C}$	TS15, TS19, TS4, S42, S81, TS10, S40

Table 2. 1 Grouping intein alleles based on growth profile

We also analyzed the point mutations that were introduced into the intein sequence, which are what essentially causes the intein splicing temperature-sensitively.

Figure 2.9 displays the amino acids sequence changes for each intein allele. Besides the two point mutations from the PCR template (TS19), these intein alleles all have 1 to 4 new point mutations, with a total of 83 point mutations found in 43 alleles (not including TS19). We did not find any hot-spots that when mutated can cause temperature sensitivity; these 83 point mutations are located randomly along the whole intein sequence, and many of them are within the non-conserved regions, or in the endonuclease domain. Among the 86 point mutations found from the 43 intein alleles, we did see some hot mutations, there are 7 Glu to Lys switches, and 6 Iles are mutated to be Val, but we are not sure if these two types of mutations are causing temperature-sensitive splicing. The reason they are occurring more frequently may be they are more frequently generated in the PCR process, especially since all the 7 Glu to Lys switches happen on the 22nd amino acid, which also locates in a non-conserved region. Besides, there is no preference on specific types of mutations (for example, side chain polarity change, or changes on side chain charge). We can not exclude the possibility that mutations introduced into the intein sequences cause an overall 3D structure change and render its splicing temperature-sensitively. Further analysis of 3D structure of these alleles might reveal why these intein alleles are TS. That may provide clues on how we can more efficiently design TS inteins using site-directed mutation technique, and may generate better intein mutants.



To be continued

Continued:



Figure 2. 9 Point mutations of TS inteins.

Note: Each dot represents one point mutation. For example, “W157R” means the 157th amino acid, tryptophan, is mutated to arginine. The yellow and red dots are two point mutations derived from TS19.

2.4 Summary

By introducing mutations into the intein sequence, we generated and identified 94 (including 7 old ones) intein alleles that support TS yeast growth in the pS5 plasmid. After shifting these alleles into the high copy number plasmid pU, we eliminated some not tightly controlled ones, and finally 44 alleles were kept as our TS inteins after sequencing. Yeast transformed with these intein mutants grow at different permissive temperatures, suggesting functional Gal4p is generated within different temperature ranges.

CHAPTER 3

ANALYSIS OF TEMPERATURE-SENSITIVE INTEIN SPLICING

3.1 Introduction

In the previous chapter, we got a set of intein mutants. As tested in an intein-splicing dependent growth system, these intein alleles splice and generate functional host protein in a temperature-dependent way. In this chapter, we designed experiments to test the intein splicing directly using western blot, since whether protein splices or not would generate proteins with different sizes (intact host protein vs. intein embedded host protein).

We first looked at the intein splicing in the context of Gal4p. This also provided us with information about the reliability of the yeast growth assays we used in screening TS inteins; whether the temperature-dependent yeast growth we saw in chapter 2 really reflects TS intein splicing. We cloned the “*gal4-intein*” into protein expression vector pMet, making it under the control of Met25 promoter. Met25 is methionine-suppressive; the promoter is active when placed in a methionine-free environment, while in the presence of methionine, transcription is turned off. Met25 promoter makes the expression of protein controllable by adding or taking away methionine. With this pMet-Gal4-Intein, we induced transcription for a fixed time period (to make sure same amount of transcripts were generated each time), and then split the yeast cultures and incubated them at

different temperatures, allowing protein translation and splicing to occur.

As part of our purposes to search for TS inteins that can control gene functions universally, i.e. the TS splicing does not depend on the host proteins flanking it, we tested selected TS inteins in the context of another protein, eGFP. Using the same procedure for the Gal4p, we expressed eGFP-Intein protein and analyzed the TS splicing using western blot assay.

To determine the splicing kinetics of these TS mutants, we also performed a time course analysis for several intein alleles. Information from kinetic study would be valuable when choosing intein alleles for each specific experiment.

3.2 Materials and methods

3.2.1 Constructing pMet-Gal4-intein

To make the construct pMet-Gal4-intein, we used “copy and paste” to clone the “Gal4-Intein” fragment into pMet vector. To generate the “Gal4-Intein” DNA fragment, we digested the plasmid pU-Gal4-Intein using KpnI (from NEB) and blunted the sticky end using T4 DNA polymerase (NEB). The polymerase activity was inactivated by incubating at 65°C for 20 minutes. The KpnI digested and blunted plasmid was then digested with BamHI, which would produce the 5’ sticky end of Gal4-Intein. At the same time, the pMet vector was treated with HindIII digestion and blunted, followed by BamHI digestion. The “Gal4-Intein” fragment was then ligated into the BamHI/HindIII(blunted) restriction sites of pMet using T4 DNA ligase (Fermentas), and transformed into the *E. Coli* Top10 (Appendix 3: Bacterial transformation). Positive clones containing pMet-Gal4-Intein were confirmed by colony PCR using primers swap1F and swap11R.

3.2.2 Protein expression, splicing, and extraction

The pMet-Gal4-Intein mini-prepared from *E.coli* Top10 cells were then transformed into the yeast *S. cerevisiae* strain FY761 (which differs from FY760 as genomic VMA intein is deleted in FY761).

Yeast cells transformed with pMet-Gal4-intein were cultured overnight in a 15ml culturing tube with 3 ml of SD-Ura media (dextrose as carbon source), with 400 μ M methionine added to repress the Met25 promoter activity, at 30°C with 300 rpm shaking. The next day, when the yeast culture had reached O.D. 600 of 0.8 (it normally takes 20 hours), the yeast cells were spun down and washed twice with ddH₂O to eliminate all residual methionine, and the pellets were then resuspended to a final O.D.600 of 0.7 using SD-Ura-Met. This methionine-free medium activates the pMet25 promoter activity and induces the Gal4-Intein transcription. After 30 minutes of transcription induction in the shaker set at 30°C and 300 rpm, the transcription was stopped by adding back sterilized 100X methionine to make a final concentration of 400 μ M. The yeast cultures (after adding methionine) were then separated into Eppendorf tubes, 1 ml each. These tubes were incubated at different temperatures allowing for intein splicing. We tested the intein splicing at different temperatures for different intein alleles, according to its permissive temperature ranges defined by yeast growth profile (Figure 2.8). Those with low permissive temperatures (first 12 alleles) were tested at 10°C, 15°C, 20°C, and 25°C; while intein alleles that showed higher permissive temperatures were incubated at 20°C, 25°C, 30°C, and 35°C. As the intein splicing itself is a chemical reaction of which the splicing velocity varies with temperatures, we gave different time periods for splicing before yeast cells were processed for protein extraction. Splicing incubated at 10°C and

15°C were given 4 hours for splicing, while 20°C and 25°C 2 hours, and the 30°C and 35°C 1 hour. Figure 3.1 summarizes the protocol for protein expression and splicing.

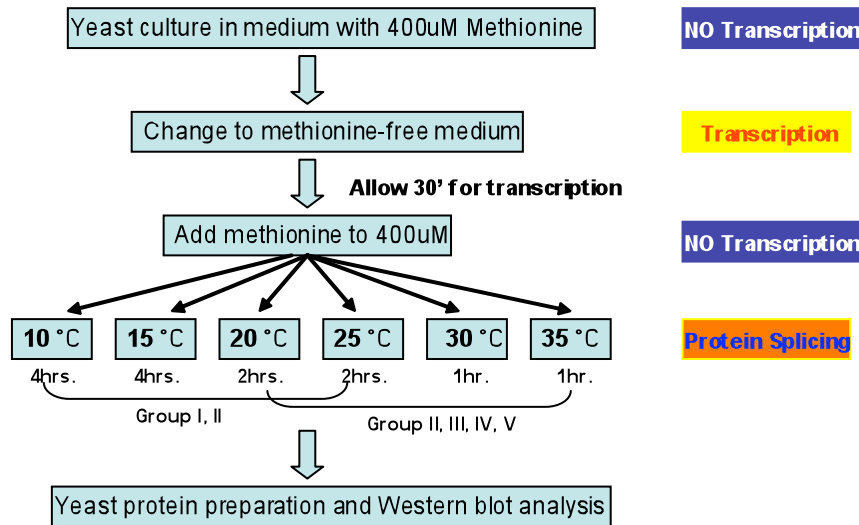


Figure 3. 1 Protocol for protein expression and splicing

Yeast proteins were extracted with glassbeads and Thorner buffer containing 8M Urea, 5% SDS, 50mM Tris-HCl (PH 6.8), 5% beta-mercaptoethanol and 1% protease inhibitor cocktail (Appendix 3: Protein extraction from yeast for western blot analysis). Yeast cells were first spun down at 7,500 rpm (Eppendorf, Centrifuge 5415D) for 1 minute and the pellets were then resuspended in 60µl of Thorner buffer pre-heated to 70°C heat block. Immediately after the yeast cells were resuspended, the yeast was transferred into another tube with 30 µg of glassbeads, also preheated to 70°C. This process was completed as quickly as possible to inactivate the endogenous proteases activity. After 2 minutes at 70 °C, the samples were vortexed at high speed for 4 minutes and the tubes were transferred back to the 70°C for 5 more minutes to denature the proteins. After that, the debris was spun down for 2 minutes at full speed of 13,200 rpm

(Eppendorf, Centrifuge 5415D) and the supernatant was transferred to a fresh tube. The proteins were then stored in -80°C until they were analyzed in the western blot assay.

3.2.3 Western blot

Samples stored in -80°C were quickly thawed by incubating on heat blocker set at 70°C for 1 minute. 7µl of each yeast extract was loaded for SDS-PAGE. SDS-PAGE gels were pre-made according to recipes listed in Appendix2: Solutions and buffers. We used 10% acrylamide for the resolving layer and 4% acrylamide for the stacking layer. Vacant lanes were filled with the same volume of Thorner buffer. Proteins were separated by running the gel for 1 hour with a constant voltage of 100 volts in the running buffer (Tris base, SDS, Glycine. For recipe see Appendix 3: Buffers and Solutions). Proteins were transferred to nitrocellulose membranes (GE Water &Process Technologies) with constant current (350mA) for 2 hours. The buffer for membrane transfer differs from running buffer with a reduced amount of glycine as well as tris base; it is also devoid of SDS (for recipe see Appendix 3: Buffers and Solutions). The membranes bearing proteins were first blocked with 5% powder milk dissolved in PBST, and thereafter probed with anti-Gal4p antibody (rabbit origin, Sigma G9293, 1:5000 dilution), followed by three washes using PBST. The secondary anti-Rabbit IgG HRP-conjugated antibody (Invitrogen G21234, goat origin) was also diluted (1:5000) in PBST and allowed to react for 45 minutes. After three PBST washes, the membranes were developed with enhanced chemiluminescence reagent (Pierce #32106), images were captured using the Fujifilm LAS-3000 Imaging System.

3.2.4 Constructing pMet-eGFP-intein

We used yeast homologous recombination to construct pMet-eGFP-intein. Figure

3.2 displays how the construct was generated: eGFP-intein (intein disrupted eGFP) was formed by three separate parts: N-eGFP, intein, C-eGFP, which were cloned into the pMet vector under the control of the Met25 promoter. The pMet vector was linearized using EcoRI; all the other three pieces were obtained by PCR using high fidelity Taq enzyme (Promega M829A). The N-eGFP was amplified from the vector pFA6a-eGFP (S65T)-KanMx6 as a template, using primers Met-GFP-f and GFP108_pr-npct, while the C-eGFP came from PCR using GFP108_pf-npct and Met-GFP-r as primers. The Intein fragments were directly amplified from pU-Gal-Inteins, using GFP108f-npct and GFP108r-npct as primers. All primers were designed to have 30 nucleotides overlap with its adjacent fragment, so that homologous recombination can occur. The EcoRI digested pMet vector, and the three gel purified PCR products were co-transformed into yeast *S. cerevisiae* FY760 using the lithium acetate technique (Appendix 3: Yeast transformation), and positive clones were selected using uracil drop-out selective medium SD-Ura (dextrose as carbon source) and incubated at 30°C. Normal-sized colonies from each transformation were picked, from which plasmids with the desired fusions were isolated and confirmed by HindIII and BamHI double digestion before inducing protein expression and splicing.

Cloning scheme of pMet-eGFP-Intein

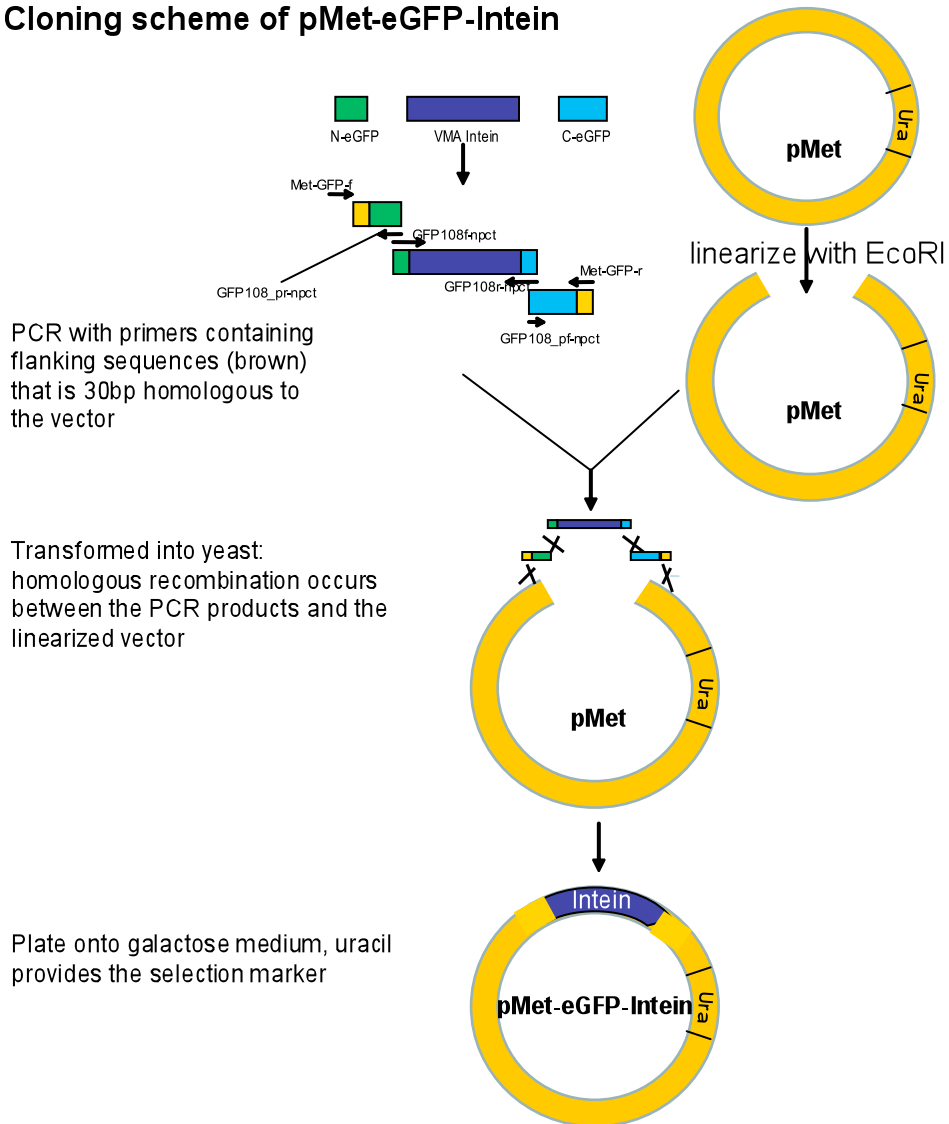


Figure 3. 2 Cloning scheme of pMet-eGFP-Intein

3.2.5 eGFP protein expression, splicing, extraction, and western blot analysis

Protein expression and splicing for eGFP was basically the same as we used for Gal4p, except that, with eGFP, we tested the intein splicing at the temperatures used for yeast growth assay: 18 °C, 22 °C, 25 °C, and 30 °C. Samples at all temperatures were incubated for 2 hours for splicing.

Protocol for eGFP western blot is the same as for Gal4p, with the only difference at the membrane transfer step. Instead of using constant current (350mA) for 2 hours with

Gal4p, we used same current for 1 hour in eGFP, since the protein size of eGFP is much smaller than Gal4p. The primary antibody used is anti-GFP N-terminal (rabbit origin, Sigma G1544, 1:4000 dilution).

3.2.6 Time-course analysis of intein splicing

Time course analysis was conducted for some chosen intein alleles (WT, TS4, S33, and S56). After 30 minutes of transcription induction, we let protein splicing occur for different time periods—30, 45, 60, 80, 100, and 120 minutes, at both 18°C and 30°C. The protein samples were then extracted and analyzed by western blot.

3.3 Results

3.3.1 Gal4-Intein splicing

Gal-Intein splicing at different temperatures was directly analyzed using western blot assay (Figure 3.3). Different intein alleles were ordered according to the yeast growth profile (Figure.2.8), starting with S3 in the first column and row one, and then moving down along the first column for the second intein. The first column displays six intein alleles and column two row one is the seventh intein allele.

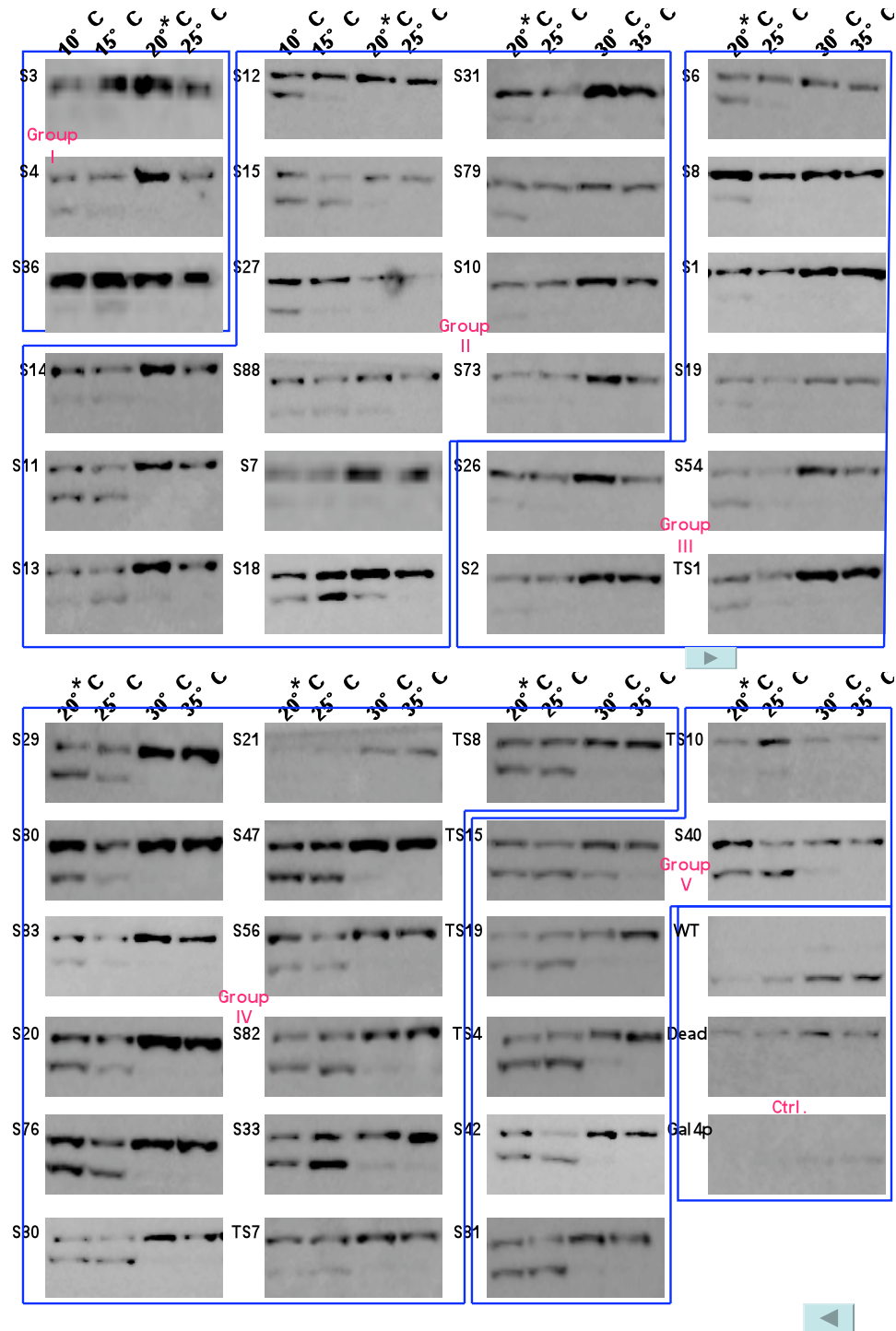


Figure 3.3 Gal4-Intein splicing at different temperatures

Note: Different splicing durations was used for different temperatures tested: 4 hrs. for splicing at 10°C and 15°C, 2 hrs. for 20°C and 25°C, and 1 hr. for 30°C and 35°C.

For each panel, the upper band represents the unspliced protein, while the lower one is the spliced functional protein.

Western blot of Gal4-Intein splicing shows that these inteins do splice in a TS way in the Gal4p context, demonstrating the temperature-dependent yeast growth indeed reflects TS Gal4-Intein splicing. The TS property is consistent between the growth profile and western blot, alleles belonging to low permissive temperature groups (for example: group I) only splices at low temperatures in the western blot assay. We noticed that temperatures at which the spliced form can be detected in the western blot is lower than those defined by yeast growth profiles for all the intein alleles tested. For example, S42, in the western blot (Figure 3.4B), we can see splicing only at 20°C and 25°C, but not 30°C. While with the growth profile (Figure 3.4A), the yeast can grow at temperatures as high as 30°C. This discrepancy can be caused by two reasons. First, splicing time for the two assays is different, several days for the yeast growth but only a couple of hours (vary from 1 to 4) for the western blot. The other explanation can be the sensitivity difference between the two assay methods, since the yeast may need much less amount of functional Gal4p for growth than the level detectable by western blot.

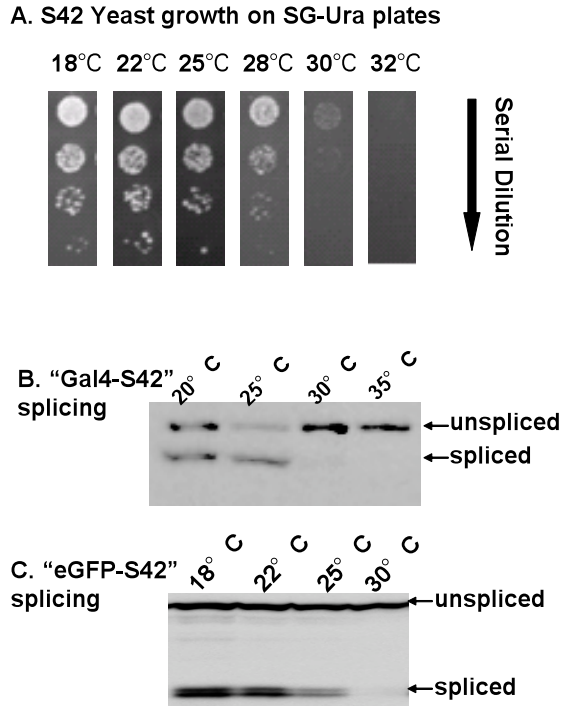


Figure 3. 4 Comparison of different assay methods for S42

(A) S42 Serial dilution assay of yeast growth on SG-Ura plate; (B) Western blot of “Gal4-S42” splicing; (C) Western blot of “eGFP-S42” splicing. Note: Data for this figure were extracted from Figure 2.8, 3.3, and 3.5 for the convenience of comparison.

3.3.2 eGFP-Intein splicing

To determine the effect of host protein contexts on TS-intein splicing, we analyzed TS intein splicing in eGFP. Western blot result for eGFP-Intein splicing was organized in the same way as for Gal4p (Figure 3.5).

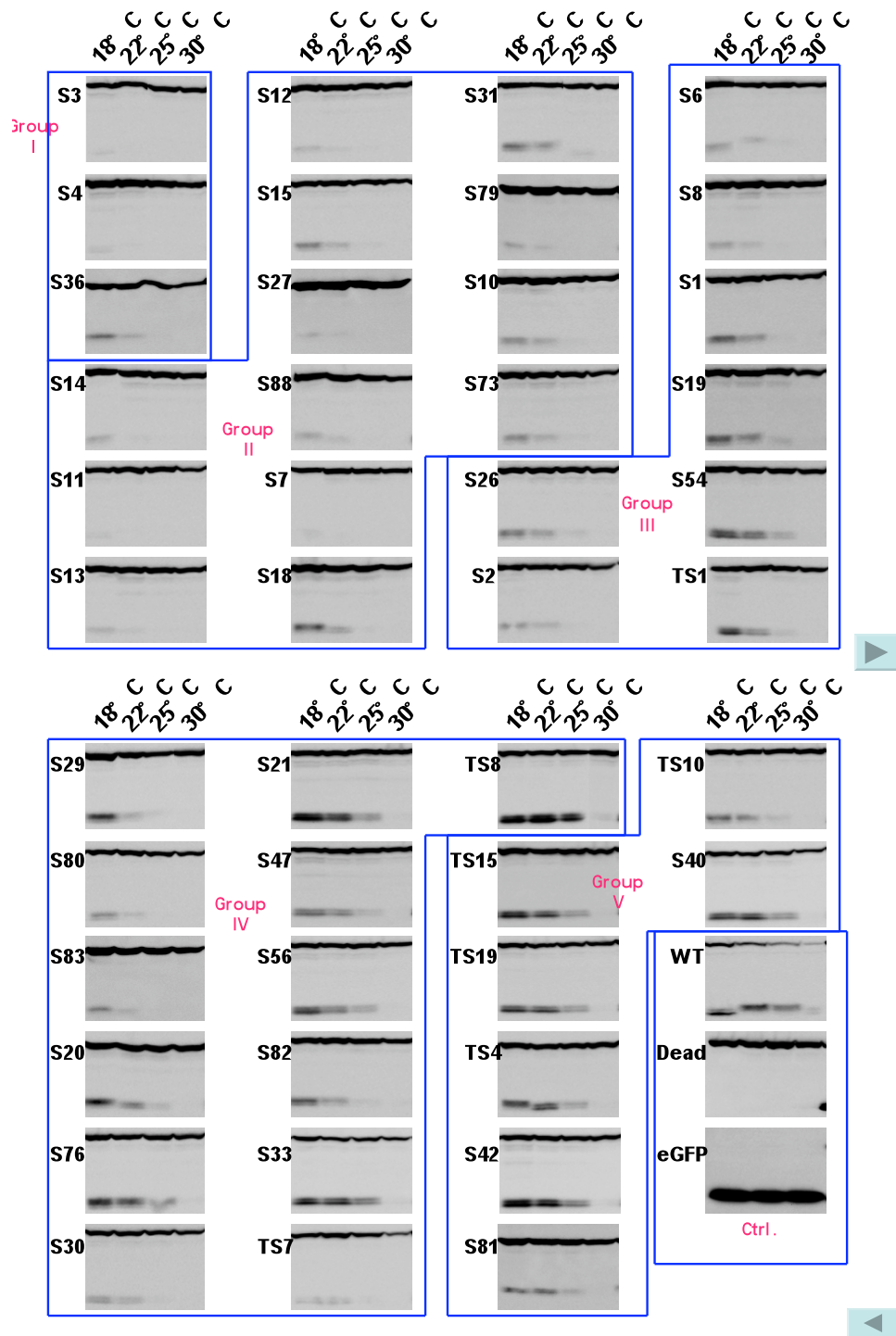


Figure 3. 5 eGFP-Intein splicing at different temperatures

Note: For all TS-intein alleles, transcription and splicing time were 30' and 2 hrs, respectively. For each panel, the upper band represents the unspliced protein, while

the lower one is the spliced functional protein.

As expected, these inteins do still behave TS splicing inside eGFP, suggesting the TS splicing is not host protein dependent. Also, for most intein alleles, the switching on/off temperatures are the same in Gal4p and in eGFP. For example, with S42, the on/off temperatures are the same in Gal4p and in eGFP. For example, with S42, the highest temperature we can see spliced form for both eGFP and Gal4p is 25°C (Figure 3.4B and C). There are, however, some exceptions. For example, eGFP-S36 splices at temperatures as high as 22°C (Figure 3.6A), whereas we can not see any Gal4-S36 splicing or yeast growth at 22°C (Figure 3.6B and C), suggesting some alleles splice in a host protein dependent manner. Therefore, for each specific protein, one may need to test a couple alleles. Fortunately, we have plenty of alleles to choose from and the vast majority of these behave the same in both Gal4p and eGFP.

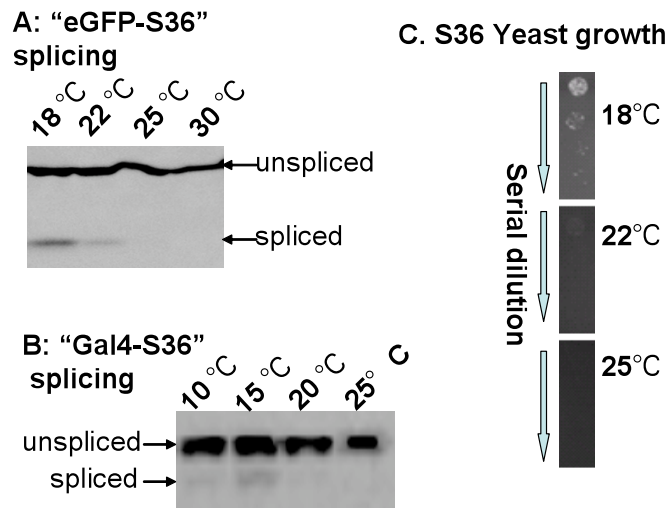


Figure 3. 6 Comparison of different assay methods for S36

3.3.3 Time course analysis of eGFP-Intein splicing

To further compare the different intein alleles, we analyzed the rate of the splicing of several alleles (Figure 3.7).

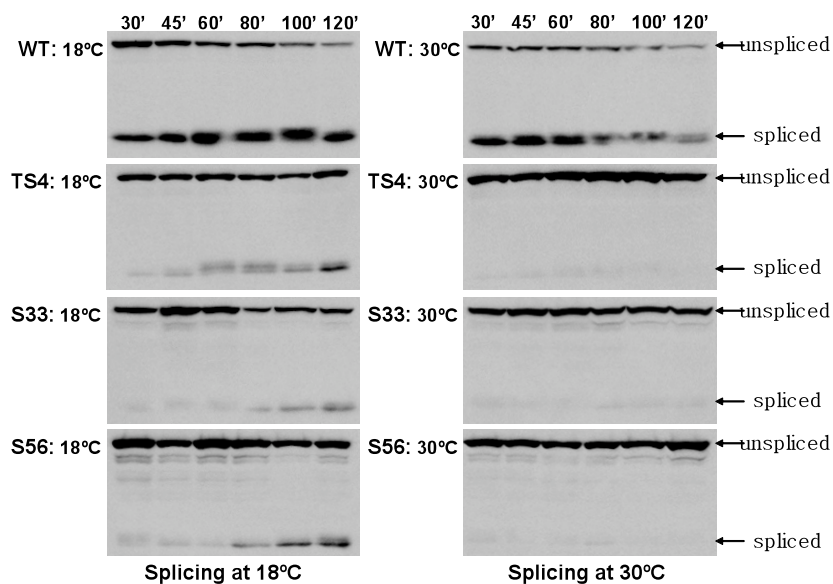


Figure 3. 7 Time course analysis of intein splicing

Note: After 30' of transcription induction, intein splicing was tested for different time periods: 30' 45' 60' 80' 100' 120'

Time course analysis suggests these inteins have different splicing kinetics. For example, S33 does not splice very much until 80 minutes, while we can see considerable amounts of TS4 splicing even at 60 minute time point.

3.4 Discussion

In this chapter, we used two reporter systems: Gal4p and eGFP, to directly look at the splicing of the selected TS intein mutants. The Gal4-Intein splicing assay proves that the temperature-dependent yeast growth is indeed a reflection of the TS intein splicing. Those belonging to low permissive temperature groups also splice only at low temperatures, those belonging to high permissive temperature groups splice at high temperatures. For most of the intein alleles, the temperature at which that intein allele switches on/off splicing are the same for eGFP and Gal4p. Results for splicing of eGFP-

intein suggest the TS splicing does not depend on the host proteins flanking them, this would be very beneficial as we want to make it a tool for universally controlling functions of different gene products.

CHAPTER 4

CONCLUSIONS AND FUTURE STUDIES

TS intein is a very promising tool for functional genetic and genomic studies. This work aimed to generate and optimize TS inteins that can be applied to different proteins, as well as to look for tightly controlled inteins with different on/off switch temperatures.

By using low-fidelity error-prone PCR, we introduced mutations into the intein sequences. TS mutants were isolated based on Gal4-intein-splicing-dependent yeast growth and verified by serial dilution assay. Tightly controlled alleles were then identified from this pool by magnifying the possible “leaky” splicing at non-permissive temperatures. When the “*gal4-intein*” gene is put into a high-copy-number plasmid pU, even very slight splicing generates functional Gal4p enough to support yeast growth in the selective media. After this second-round screening, about half of the candidate intein alleles from the primary screen were dropped and finally we had 44 TS mutants (including 7 old ones we had before). As organized in Fig. 2.8, yeast carrying these intein alleles grew in the temperature dependent manner. They essentially have the same property, in the sense that they splice normally at low permissive temperatures (at least in 18 °C), but when the temperature increases, the splicing efficiency decreases until it reaches a temperature point when it is not able to generate sufficient amount of Gal4p for yeast growth. That point is what we defined as the non-permissive temperature for each allele, and the temperature points below non-permissive temperature are called

permissive temperatures. We grouped these 44 alleles according to their permissive temperatures defined by the yeast growth profile.

The Gal4p western blot experiment was designed to directly examine the Gal4-Intein splicing. Not surprisingly, the result is consistent with the yeast growth assay. Higher permissive temperature alleles (defined by yeast growth profile) splice better in the western blot analysis. There are a few exceptions, like S21, which does not splice at all based on our western blot. Since the protein expression of S21 is also very low, it is possible that there are some abnormalities with the pMet-Gal4-S21 construct (We sequenced the “*gal4-s21*” portion. It is correct, but maybe something is wrong with the vector backbone).

To study whether this TS splicing applies to other host proteins, these 44 TS intein alleles were further tested for their splicing in the context of eGFP. All intein mutants showed temperature-dependent splicing in both host proteins. Overall, the TS inteins splice similarly in Gal4p and eGFP. For those with higher permissive temperatures in Gal4p, the eGFP-inteins splice better, and they splice at higher temperatures (not just at 18 °C as low permissive temperature ones). For example, with S42 (Figure 3.4), the splicing permissive temperature for both Gal4-Intein and eGFP-Intein is 25 °C. This suggests that, for most alleles, TS intein splicing behaves the same when in the context of different proteins, at least for Gal4p and eGFP. There are exceptions. For example, in eGFP, S81 splices very slightly at 25 °C, but within Gal4p, the splicing signal can be very strong at 25°C. This should be due to different host proteins: some intein mutants do not always behave constant TS property within different host proteins. S36 (Figure 3.6) is another exception, with eGFP it splices at 18°C and

22°C, but we did not see any splicing either in the Gal4p western blot or the Gal4-dependent growth at 22°C.

In addition, for most intein alleles, the temperatures at which we observed the splicing were lower than that defined by yeast growth profile. S42 is an example for this phenomenon (Figure 3.4). In the western blot assay for both Gal4p and eGFP, S42 splices at 25°C, but not at 30°C, while from the yeast growth profile, S42 containing yeast grows at temperatures between 18°C and 30°C, although not very well at 30°C. This difference can be explained by either (or both) different time period used for splicing, or due to the sensitivity of different assay methods, because more time was allowed for intein splicing during yeast growth than for the western blot assay, and the western blot assay is much less sensitive than yeast growth.

Study on the splicing kinetics suggests different intein alleles splice at different rates. For example, S33 does not splice very much until 80 minutes, while we can see considerable amounts of TS4 splicing even at the 60 minute time point. Information about the splicing kinetics of different alleles is very valuable to assist choosing TS intein allele for each specific experiment, since some studies need the on/off switch of protein function to be fast, while for others this may not be important. Note our splicing kinetics assay is not measuring intein splicing *per se*. Several other processes are occurring at the same time, including mRNA turn over, protein translation, and turnover of both the unspliced and spliced proteins. Thus the western blot reflects a collective effects of all these processes. The turnover rate for the spliced form is the same since it is the same protein. We have not tested the rate difference of different intein alleles in terms of the mRNA turnover, translation, and turnover of the intein containing precursor proteins.

Nonetheless, for controlling the activity of a gene and thus the phenotype of an organism, it is the final spliced, functional product that counts.

We found that, while almost all old TS mutants have relatively high permissive temperature, with the newly generated ones, the permissive temperature varies between 18°C and 30°C, suggesting the success of our strategy to generate more versatile intein mutants. This improvement provides us much more flexibility of using intein switch for gene function studies, since some experiments require the test to be conducted in a not high temperature, and also different species prefer different living temperatures.

Based on the temperature at which the splicing is turned off and splicing efficiencies of these TS-inteins, three alleles were selected as the best TS mutants for future use, they are: S42, S8, and S13, which functions between 25°C and 30°C, 20°C and 25°C, 15°C and 20°C, respectively. The splicing of these intein mutants is stringently regulated on its respective temperature range, and is not host protein dependent.

Taken together, we have identified a set of TS intein alleles that can potentially be used to conditionally control gene functions. They can be used as on/off switches working at different temperature ranges; therefore, we can choose from this pool based upon our specific needs. While there are a couple alleles that splice differently in the context of eGFP and Gal4p, most of the alleles behave essentially the same in splicing for both host proteins, suggesting the intein splicing is promiscuous. However, to use these intein alleles as a tool for studying gene functions, it is worthwhile that we first verify the TS splicing on the host protein to be controlled. Also, it is important for us to make sure when the intein is inserted into a host protein, it would disrupt the function of that protein. And we may need to try different sites to insert the intein to find one site that makes the

host protein nonfunctional at non-permissive temperatures.

Further work can also be planned on testing how well these intein mutants can be used for functional genetic studies. Several factors would affect the usability of this technique. For example, the amount of spliced protein for biological functions can be an issue, since biological system may require different amounts of different proteins. Use intein switch to regulate gene function, we need to make sure when the protein splicing is turned on, the amount of functional protein generated can meet the requirement of that biological system. This is also important because the protein concentration might affect the splicing efficiency. Finally, to make this a tool to generate genomic-scale library of TS intein mutations, we need to develop a strategy on how we can most efficiently insert these intein mutants into different genes in genomes.

REFERENCES

- Amarzguioui, M., Holen, T., Babaie, E., and Prydz, H. (2003). Tolerance for mutations and chemical modifications in a siRNA. *Nucleic Acids Research* 31, 589-595.
- Anraku, Y., Mizutani, R., and Satow, Y. (2005). Protein splicing: its discovery and structural insight into novel chemical mechanisms. *IUBMB Life* 57, 563-574.
- Bakhrat, A., Jurica, M.S., Stoddard, B.L., and Raveh, D. (2004). Homology modeling and mutational analysis of Ho endonuclease of yeast. *Genetics* 166, 721-728.
- Belfort, M., and Roberts, R.J. (1997). Homing endonucleases: keeping the house in order. *Nucleic Acids Research* 25, 3379-3388.
- Bhat, P.J., and Murthy, T.V. (2001). Transcriptional control of the GAL/MEL regulon of yeast *Saccharomyces cerevisiae*: mechanism of galactose-mediated signal transduction. *Molecular Microbiology* 40, 1059-1066.
- Boutros, M., Kiger, A.A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Paro, R., and Perrimon, N. (2004). Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* 303, 832-835.
- Brent, R., and Finley, R.L., Jr. (1997). Understanding gene and allele function with two-hybrid methods. *Annual Review of Genetics* 31, 663-704.
- Brummelkamp, T.R., Bernards, R., and Agami, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. *Science* 296, 550-553.
- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. *Science* 263, 802-805.
- Chandler, V.L., Maler, B.A., and Yamamoto, K.R. (1983). DNA sequences bound specifically by glucocorticoid receptor in vitro render a heterologous promoter hormone responsive in vivo. *Cell* 33, 489-499.
- Cheng, L., Fu, J., Tsukamoto, A., and Hawley, R.G. (1996). Use of green fluorescent protein variants to monitor gene transfer and expression in mammalian cells. *Nature Biotechnology* 14, 606-609.
- Choi, J.J., Nam, K.H., Min, B., Kim, S.J., Soll, D., and Kwon, S.T. (2006). Protein trans-splicing and characterization of a split family B-type DNA polymerase from the hyperthermophilic archaeal parasite *Nanoarchaeum equitans*. *Journal of Molecular Biology* 356, 1093-1106.
- Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A., et al. (1997). Single-column

- purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* 192, 271-281.
- Chong, S., and Xu, M.Q. (1997). Protein splicing of the *Saccharomyces cerevisiae* VMA intein without the endonuclease motifs. *The Journal of Biological Chemistry* 272, 15587-15590.
- Clarke, N.D. (1994). A proposed mechanism for the self-splicing of proteins. *Proceedings of the National Academy of Sciences (USA)* 91, 11084-11088.
- Cogoni, C., Irelan, J.T., Schumacher, M., Schmidhauser, T.J., Selker, E.U., and Macino, G. (1996). Transgene silencing of the *al-1* gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. *The EMBO Journal* 15, 3153-3163.
- Costa, R.M., Drew, C., and Silva, A.J. (2005). Notch to remember. *Trends in Neurosciences* 28, 429-435.
- Cullen, L.M., and Arndt, G.M. (2005). Genome-wide screening for gene function using RNAi in mammalian cells. *Immunology and Cell Biology* 83, 217-223.
- Daneholt, B. (2006). Advanced Information: RNA interference. *The Nobel Prize in Physiology or Medicine 2006*.
- Datta, S., Stark, K., and Kankel, D.R. (1993). Enhancer detector analysis of the extent of genomic involvement in nervous system development in *Drosophila melanogaster*. *Journal of Neurobiology* 24, 824-841.
- Derbyshire, V., Wood, D.W., Wu, W., Dansereau, J.T., Dalgaard, J.Z., and Belfort, M. (1997). Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution. *Proceedings of the National Academy of Sciences of the United States of America* 94, 11466-11471.
- Ding, Y., Xu, M.Q., Ghosh, I., Chen, X., Ferrandon, S., Lesage, G., and Rao, Z. (2003). Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *The Journal of Biological Chemistry* 278, 39133-39142.
- Ding, Y., Xu, M.Q., Ghosh, I., Chen, X., Ferrandon, S., Lesage, G., and Rao, Z. (2003). Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *The Journal of Biological Chemistry* 278, 39133-39142.
- Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* 330, 576-578.

- Dohmen, R.J., and Varshavsky, A. (2005). Heat-inducible degron and the making of conditional mutants. *Methods in Enzymology* 399, 799-822.
- Dohmen, R.J., Wu, P., and Varshavsky, A. (1994). Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science* 263, 1273-1276.
- Duan, X., Gimble, F.S., and Quioco, F.A. (1997). Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell* 89, 555-564.
- Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154-156.
- Evans, T.C., Jr., Benner, J., and Xu, M.Q. (1999). The in vitro ligation of bacterially expressed proteins using an intein from *Methanobacterium thermoautotrophicum*. *The Journal of Biological Chemistry* 274, 3923-3926.
- Evans, T.J.T., and Xu, M.Q. (2002). Mechanistic and kinetic considerations of protein splicing. *Chemical Reviews* 102, 4869-4884.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806-811.
- Ge, G., Wong, G.W., and Luo, B. (2005). Prediction of siRNA knockdown efficiency using artificial neural network models. *Biochemical and Biophysical Research Communications* 336, 723-728.
- Gogarten, J.P., Senejani, A.G., Zhaxybayeva, O., Olendzenski, L., and Hilario, E. (2002). Inteins: structure, function, and evolution. *Annual Review of Microbiology* 56, 263-287.
- Gu, H., Marth, J.D., Orban, P.C., Mossmann, H., and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* 265, 103-106.
- Guhan, N., and Muniyappa, K. (2003). *Mycobacterium tuberculosis* RecA intein, a LAGLIDADG homing endonuclease, displays Mn(2+) and DNA-dependent ATPase activity. *Nucleic Acids Research* 31, 4184-4191.
- Guo, S., and Kemphues, K.J. (1995). *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* 81, 611-620.
- Harborth, J., Elbashir, S.M., Vandeburgh, K., Manninga, H., Scaringe, S.A., Weber, K., and Tuschl, T. (2003). Sequence, chemical, and structural variation of small

- interfering RNAs and short hairpin RNAs and the effect on mammalian gene silencing. *Antisense & Nucleic Acid Drug Development* 13, 83-105.
- Harris, S.D., and Pringle, J.R. (1991). Genetic analysis of *Saccharomyces cerevisiae* chromosome I: on the role of mutagen specificity in delimiting the set of genes identifiable using temperature-sensitive-lethal mutations. *Genetics* 127, 279-285.
- He, Z., Crist, M., Yen, H., Duan, X., Quioco, F.A., and Gimble, F.S. (1998). Amino acid residues in both the protein splicing and endonuclease domains of the PI-SceI intein mediate DNA binding. *The Journal of biological chemistry* 273, 4607-4615.
- Heim, R., Cubitt, A.B., and Tsien, R.Y. (1995). Improved green fluorescence. *Nature* 373, 663-664.
- Hirata, R., Ohsumk, Y., Nakano, A., Kawasaki, H., Suzuki, K., and Anraku, Y. (1990). Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *The Journal of biological chemistry* 265, 6726-6733.
- Hope, I.A., and Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* 46, 885-894.
- Huesken, D., Lange, J., Mickanin, C., Weiler, J., Asselbergs, F., Warner, J., Meloon, B., Engel, S., Rosenberg, A., Cohen, D., et al. (2005). Design of a genome-wide siRNA library using an artificial neural network. *Nature Biotechnology* 23, 995-1001.
- Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K., and Morikawa, K. (2000). Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. *Journal of Molecular Biology* 300, 889-901.
- Inouye, S., and Tsuji, F.I. (1994). Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein. *FEBS Letters* 341, 277-280.
- Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnology* 21, 635-637.
- Janitz, M., Vanhecke, D., and Lehrach, H. (2006). High-throughput RNA interference in functional genomics. *Handbook of Experimental Pharmacology*, 97-104.
- Johnson, F.H., Shimomura, O., and Saiga, Y. (1962). Action of cyanide on *Cypridina luciferin*. *Journal of Cellular and Comparative Physiology* 59, 265-272.

- Jurica, M.S., and Stoddard, B.L. (1999). Homing endonucleases: structure, function and evolution. *Cell Mol Life Sci* 55, 1304-1326.
- Kamath, R.S., and Ahringer, J. (2003). Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods* 30, 313-321.
- Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebel, M., and Stevens, T.H. (1990). Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* 250, 651-657.
- Kanemaki, M., Sanchez-Diaz, A., Gambus, A., and Labib, K. (2003). Functional proteomic identification of DNA replication proteins by induced proteolysis in vivo. *Nature* 423, 720-724.
- Kellendonk, C., Tronche, F., Casanova, E., Anlag, K., Opherk, C., and Schutz, G. (1999). Inducible site-specific recombination in the brain. *Journal of Molecular Biology* 285, 175-182.
- Khan, M.S., Khalid, A.M., and Malik, K.A. (2005). Intein-mediated protein trans-splicing and transgene containment in plastids. *Trends in Biotechnology* 23, 217-220.
- Klabunde, T., Sharma, S., Telenti, A., Jacobs, W.R., Jr., and Sacchettini, J.C. (1998). Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nature Structural Biology* 5, 31-36.
- Kwan, K.M. (2002). Conditional alleles in mice: practical considerations for tissue-specific knockouts. *Genesis* 32, 49-62.
- Lakso, M., Sauer, B., Mosinger, B., Jr., Lee, E.J., Manning, R.W., Yu, S.H., Mulder, K.L., and Westphal, H. (1992). Targeted oncogene activation by site-specific recombination in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* 89, 6232-6236.
- Levy, F., Johnston, J.A., and Varshavsky, A. (1999). Analysis of a conditional degradation signal in yeast and mammalian cells. *European Journal of Biochemistry / FEBS* 259, 244-252.
- Lindner, K., Gregan, J., Montgomery, S., and Kearsey, S.E. (2002). Essential role of MCM proteins in premeiotic DNA replication. *Molecular Biology of the Cell* 13, 435-444.
- Liu, X.Q. (2000). Protein-splicing intein: Genetic mobility, origin, and evolution. *Annual Review of Genetics* 34, 61-76.
- Lohr, D., Venkov, P., and Zlatanova, J. (1995). Transcriptional regulation in the yeast GAL gene family: a complex genetic network. *Faseb J* 9, 777-787.

- Lykke-Andersen, J., Garrett, R.A., and Kjems, J. (1996). Protein footprinting approach to mapping DNA binding sites of two archaeal homing enzymes: evidence for a two-domain protein structure. *Nucleic Acids Research* 24, 3982-3989.
- Mansour, S.L., Thomas, K.R., and Capecchi, M.R. (1988). Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* 336, 348-352.
- Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences (USA)* 78, 7634-7638.
- Martinez, J., and Tuschl, T. (2004). RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes & Development* 18, 975-980.
- Matsumura, H., Takahashi, H., Inoue, T., Yamamoto, T., Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., and Kai, Y. (2006). Crystal structure of intein homing endonuclease II encoded in DNA polymerase gene from hyperthermophilic archaeon *Thermococcus kodakaraensis* strain KOD1. *Proteins* 63, 711-715.
- McRae, S.R., Brown, C.L., and Bushell, G.R. (2005). Rapid purification of EGFP, EYFP, and ECFP with high yield and purity. *Protein Expression and Purification* 41, 121-127.
- Miklos, G.L., and Rubin, G.M. (1996). The role of the genome project in determining gene function: insights from model organisms. *Cell* 86, 521-529.
- Mills, K.V., and Perler, F.B. (2005). The mechanism of intein-mediated protein splicing: variations on a theme. *Protein and Peptide Letters* 12, 751-755.
- Moore, C.W., and Sherman, F. (1975). Role of DNA sequences in genetic recombination in the *iso-1-cytochrome c* gene of yeast. I. Discrepancies between physical distances and genetic distances determined by five mapping procedures. *Genetics* 79, 397-418.
- Morise, H., Shimomura, O., Johnson, F.H., and Winant, J. (1974). Intermolecular energy transfer in the bioluminescent system of *Aequorea*. *Biochemistry* 13, 2656-2662.
- Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into *Petunia* Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant Cell* 2, 279-289.
- Noren, C.J., Wang, J., and Perler, F.B. (2000). Dissecting the Chemistry of Protein Splicing and Its Applications. *Angewandte Chemie (International ed)* 39, 450-466.

- Orban, P.C., Chui, D., and Marth, J.D. (1992). Tissue- and site-specific DNA recombination in transgenic mice. *Proceedings of the National Academy of Sciences (USA)* 89, 6861-6865.
- Ormo, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y., and Remington, S.J. (1996). Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* 273, 1392-1395.
- Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. (1997). Cosuppression in *Drosophila*: gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. *Cell* 90, 479-490.
- Paulus, H. (2000). Protein splicing and related forms of protein autoprocessing. *Annual Review of Biochemistry* 69, 447-496.
- Perler, F.B. (2000). InBase, the Intein Database. *Nucleic Acids Research* 28, 344-345.
- Perler, F.B. (2002). InBase: the Intein Database. *Nucleic acids research* 30, 383-384.
- Perler, F.B. (2005). Protein splicing mechanisms and applications. *IUBMB Life* 57, 469-476.
- Perler, F.B., Davis, E.O., Dean, G.E., Gimble, F.S., Jack, W.E., Neff, N., Noren, C.J., Thorner, J., and Belfort, M. (1994). Protein splicing elements: inteins and exteins-a definition of terms and recommended nomenclature. *Nucleic Acids Research* 22, 1125-1127.
- Perler, F.B., Olsen, G.J., and Adam, E. (1997). Compilation and analysis of intein sequences. *Nucleic Acids Research* 25, 1087-1093.
- Perler, F.B., Olsen, G.J., and Adam, E. (1997). Compilation and analysis of intein sequences. *Nucleic Acids Research* 25, 1087-1093.
- Petrokovski, S. (1994). Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci* 3, 2340-2350.
- Petrokovski, S. (1998). Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci* 7, 64-71.
- Petrokovski, S. (2001). Intein spread and extinction in evolution. *Trends Genet* 17, 465-472.
- Petrokovski, S. (2001). Intein spread and extinction in evolution. *Trends Genet* 17, 465-472.

- Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G., and Cormier, M.J. (1992). Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* 111, 229-233.
- Pringle, J.R. (1975). Induction, selection, and experimental uses of temperature-sensitive and other conditional mutants of yeast. *Methods in Cell Biology* 12, 233-272.
- Rassoulzadegan, M., Magliano, M., and Cuzin, F. (2002). Transvection effects involving DNA methylation during meiosis in the mouse. *The EMBO Journal* 21, 440-450.
- Ratcliff, F., Harrison, B.D., and Baulcombe, D.C. (1997). A similarity between viral defense and gene silencing in plants. *Science* 276, 1558-1560.
- Rizzuto, R., Brini, M., De Giorgi, F., Rossi, R., Heim, R., Tsien, R.Y., and Pozzan, T. (1996). Double labelling of subcellular structures with organelle-targeted GFP mutants in vivo. *Curr Biol* 6, 183-188.
- Romoser, V.A., Hinkle, P.M., and Persechini, A. (1997). Detection in living cells of Ca²⁺-dependent changes in the fluorescence emission of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence. A new class of fluorescent indicators. *The Journal of Biological Chemistry* 272, 13270-13274.
- Sauer, B. (1993). Manipulation of transgenes by site-specific recombination: use of Cre recombinase. *Methods in Enzymology* 225, 890-900.
- Saves, I., Westrelin, F., Daffe, M., and Masson, J.M. (2001). Identification of the first eubacterial endonuclease coded by an intein allele in the *pps1* gene of mycobacteria. *Nucleic Acids Research* 29, 4310-4318.
- Saxena, S., Jonsson, Z.O., and Dutta, A. (2003). Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *The Journal of Biological Chemistry* 278, 44312-44319.
- Schwartzberg, P.L., Goff, S.P., and Robertson, E.J. (1989). Germ-line transmission of a *c-abl* mutation produced by targeted gene disruption in ES cells. *Science* 246, 799-803.
- Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005). A guide to choosing fluorescent proteins. *Nature Methods* 2, 905-909.
- Shao, Y., Xu, M.Q., and Paulus, H. (1996). Protein splicing: evidence for an N-O acyl rearrangement as the initial step in the splicing process. *Biochemistry* 35, 3810-3815.

- Shellenbarger, D.L., and Mohler, J.D. (1978). Temperature-sensitive periods and autonomy of pleiotropic effects of l(1)Nts1, a conditional notch lethal in *Drosophila*. *Developmental Biology* 62, 432-446.
- Shimomura, O., Johnson, F.H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *Journal of Cellular and Comparative Physiology* 59, 223-239.
- Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., and Kucherlapati, R.S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* 317, 230-234.
- Stryer, L. (1978). Fluorescence energy transfer as a spectroscopic ruler. *Annual Review of Biochemistry* 47, 819-846.
- Suzuki, D.T., Grigliatti, T., and Williamson, R. (1971). Temperature-sensitive mutations in *Drosophila melanogaster*. VII. A mutation (para-ts) causing reversible adult paralysis. *Proceedings of the National Academy of Sciences of the United States of America* 68, 890-893.
- Swift, G.H., Hammer, R.E., MacDonald, R.J., and Brinster, R.L. (1984). Tissue-specific expression of the rat pancreatic elastase I gene in transgenic mice. *Cell* 38, 639-646.
- Thomas, K.R., and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51, 503-512.
- Tiscornia, G., Tergaonkar, V., Galimi, F., and Verma, I.M. (2004). CRE recombinase-inducible RNA interference mediated by lentiviral vectors. *Proceedings of the National Academy of Sciences (USA)* 101, 7347-7351.
- Tronche, F., Casanova, E., Turiault, M., Sahly, I., and Kellendonk, C. (2002). When reverse genetics meets physiology: the use of site-specific recombinases in mice. *FEBS letters* 529, 116-121.
- Vanhecke, D., and Janitz, M. (2005). Functional genomics using high-throughput RNA interference. *Drug Discovery Today* 10, 205-212.
- Vooijs, M., Jonkers, J., and Berns, A. (2001). A highly efficient ligand-regulated Cre recombinase mouse line shows that LoxP recombination is position dependent. *EMBO Reports* 2, 292-297.
- Wachsman, W., Cann, A.J., Williams, J.L., Slamon, D.J., Souza, L., Shah, N.P., and Chen, I.S. (1987). HTLV x gene mutants exhibit novel transcriptional regulatory phenotypes. *Science* 235, 674-677.

- Wertman, K.F., Drubin, D.G., and Botstein, D. (1992). Systematic mutational analysis of the yeast ACT1 gene. *Genetics* 132, 337-350.
- Xu, M.Q., Comb, D.G., Paulus, H., Noren, C.J., Shao, Y., and Perler, F.B. (1994). Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation. *The EMBO Journal* 13, 5517-5522.
- Yanagida, M. (1998). Fission yeast cut mutations revisited: control of anaphase. *Trends in Cell Biology* 8, 144-149.
- Yang, F., Moss, L.G., and Phillips, G.N., Jr. (1996). The molecular structure of green fluorescent protein. *Nature Biotechnology* 14, 1246-1251.
- Yang, J., Meng, Q., and Liu, X.Q. (2004). Intein harbouring large tandem repeats in replicative DNA helicase of *Trichodesmium erythraeum*. *Molecular Microbiology* 51, 1185-1192.
- Zeidler, M.P., Tan, C., Bellaiche, Y., Cherry, S., Hader, S., Gayko, U., and Perrimon, N. (2004). Temperature-sensitive control of protein activity by conditionally splicing inteins. *Nature Biotechnology* 22, 871-876.

APPENDICES

6.1 Appendix 1: Media for yeast and bacterial culturing

Synthetic Complete Drop-Out Mix

<u>Amount in mix</u>	<u>Product Name</u>	<u>Sigma Catalog NO.</u>
0.2g	Adenine hemisulfate	A-9126
0.2g	Arginine HCl	A-5131
0.2g	Histidine HCl	H-8000
0.2g	Isoleucine	I-2752
0.4g	Leucine	L-8912
0.2g	Lysine HCl	L-6027
0.2g	Methionine	M-9625
0.3g	Phenylalanine	P-2126
0.2g	Serine	S-4500
0.2g	Threonine	T-8625
0.3g	Tryptophan	T-0254
0.2g	Tyrosine	T-3754
0.12g	Uracil	U-0750
0.9g	Valine	V-0500

Combine the ingredients in a clean 1000 ml bottle, add deionized water to 600ml and let it dissolve, it will be the 10X Synthetic Complete Drop Out mix. We autoclave and store it in the 4°C refrigerator. The above quantity will be sufficient for 10 batches (each with 600ml) of SD drop-out medium. Components in bold can be omitted as needed to make drop-out mixes, e.g. SD-Ura means excluding uracil in the mixture. For this experiments,

as we always need the medium to be deficient of uracil and sometimes require methionine-free as well, so I made the mix without uracil and methionine, and we call it SD-Ura-Met mix.

SD-Ura Plates

- 60ml SD-Ura-Met mix (pre-made)
- 12g Dextrose (Labscientific Cat# FLY8012-10)
- 4.02g Difco Yeast Nitrogen Base w/o Amino Acids (BD 291940)
- 0.02g Methionine (Sigma M-9625)
- 12g Difco granulated Agar (BD 214510)

Add ddH₂O to 600ml, using stir bar to help dissolve, autoclave and pour the plates when the medium cools down to approximately 50°C. Plates are kept at 4°C.

SG-Ura Plates

- 60ml SD-Ura-Met mix (pre-made)
- 12g D-Galactose (Fisher Scientific Cat# BP 656-500)
- 4.02g Difco Yeast Nitrogen Base w/o Amino Acids (BD 291940)
- 0.02g Methionine (Sigma M-9625)
- 12g Difco granulated Agar (BD 214510)

Add ddH₂O to 600ml, using stir bar to help dissolve, autoclave and pour the plates when the medium cools down to approximately 50°C. Plates are kept at 4°C.

YPD

- 10g Bacto Yeast extracts (BD 212720)
- 20g Bacto-peptone (BD 211820)
- 20g Dextrose (Labscientific Cat# FLY8012-10)

Add ddH₂O to 1000ml, using stir bar to help dissolve, autoclave.

(For pouring plates, add 20g of Difco Granulated Agar before autoclaving)

SD-Ura-Met

60ml SD-Ura-Met mix (pre-made)

12g Dextrose (Labscientific Cat# FLY8012-10)

4.02g Difco Yeast Nitrogen Base w/o Amino Acids (BD 291940)

Add ddH₂O to 600ml, using stir bar to help dissolve, autoclave.

100X Methionine Solution

1.2g Methionine (Sigma M-9625)

Add ddH₂O to 200ml and dissolve, autoclave and kept at 4°C.

SD-Ura (with 400µM Methionine)

60ml SD-Ura-Met mix (pre-made)

12g Dextrose (Labscientific Cat# FLY8012-10)

4.02g Difco Yeast Nitrogen Base w/o Amino Acids (BD 291940)

6ml 100X Methionine Solution

Add ddH₂O to 600ml, using stir bar to help dissolve, autoclave.

Terrific Broth

12g Tryptone (Fisher Scientific Cat# BP14212)

24g Bacto Yeast extracts (BD 212720)

4ml Glycerol

Add ddH₂O up to 900 ml, then autoclave and the base broth is made. To prepare for the Terrific Broth, add 100 mL of a sterile solution of 0.17 M KH₂PO₄ and 0.72 M K₂HPO₄ to 900 mL of base broth.

Thorner Buffer (100ml in total)

- 48g Urea (Fisher Scientific U15500)
- 5g SDS (Fisher Scientific BP166500)
- 50ml 100mM Tris, PH6.8 (Fisher Scientific BP1512)

Add ddH₂O to final volume of 100ml

---before use, add 2-Mercaptoethanol (Fisher Scientific BP176100) to 5%, and protease inhibitors (Roche 11836153-001) (1:100), DTT (Fisher Scientific BP1725) with a final concentration of 1mM.

SOB Media

- 20g Tryptone (Fisher Scientific Cat# BP14212)
- 5g Bacto Yeast extracts (BD 212720)
- 0.584g NaCl (Fisher Scientific Cat# S271-10)
- 0.186g KCl (Fisher Scientific Cat# BP366-5000)
- 2.033g MgCl₂.6H₂O (Fisher Scientific Cat# S93292)
- 2.468g MgSO₄.7H₂O (Fisher Scientific Cat# AC42390-5000)

Add ddH₂O to 1000ml, using stir bar to help dissolve. Filter Sterilize (Do not autoclave!)

SOC Media

- 20g Tryptone (Fisher Scientific Cat# BP14212)
- 5g Bacto Yeast extracts (BD 212720)
- 0.584g NaCl (Fisher Scientific Cat# S271-10)
- 0.186g KCl (Fisher Scientific Cat# BP366-5000)
- 2.033g MgCl₂.6H₂O (Fisher Scientific Cat# S93292)
- 2.468g MgSO₄.7H₂O (Fisher Scientific Cat# AC42390-5000)

5g Dextrose (Labscientific Cat# FLY8012-10)

Add ddH₂O to 1000ml, using stir bar to help dissolve. Filter Sterilize (Do not autoclave!)

TB (for making competent cells)

1.675g Pipes (Fisher Scientific Cat#BP304-500)

5.44g MnCl₂ · 4H₂O (Fisher Scientific Cat# S80070)

1.1025g CaCl₂ · 2H₂O (Fisher Scientific Cat#BP510500)

9.32g KCl (Fisher Scientific Cat# BP366-5000)

Add ddH₂O to 500 ml, filter sterilize and store at 4 °C.

6.2 Appendix 2: Solutions and buffers

Solutions needed for plasmid DNA preparation from bacteria

- GTE Buffer (50 mM glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 200 µL/tube)
- 0.2 N NaOH / 1% SDS (freshly made) (300 µL/tube)
- 3.0 M Potassium acetate (pH 4.8) (300 µL/tube)
- RNase A (DNase-free) (10 mg/mL) (2 µL/tube)
- Chloroform (400 µL/tube)
- Isopropanol, 100%
- Ethanol, 70%
- 4.0 M NaCl
- 13% PEG₈₀₀₀ (sterilized by autoclaving, rather than by filtration)
- Deionized H₂O
- Terrific Broth (1 L)

Yeast plasmid rescue buffer:

2% Triton X-100

1% SDS

100mM NaCl

10mM Tris, PH8.0

1mM EDTA

Buffers for Western Blot

SDS PAGE RECIPES

Resolving Layer					
For 5mls:	Water	30%ACRYL	5X LGS	30%APS	TEMED
7.5%	2.75ml	1.25ml	1ml	15µl	15µl
10.00%	2.3ml	1.7ml	1ml	15µl	15µl
12%	2ml	2ml	1ml	15µl	15µl
15%	1.5ml	2.5ml	1ml	15µl	15µl
Stacking Layer					
	Water	30%ACRYL	10XUGS	30%APS	TEMED
4.0%	2.4ml	0.4ml	0.3ml	10µl	10µl

5X LGS

192g Tris Base

46g Tris HCL

50ml of 20% SDS

----adjust using water to 1L

10X UGS

180g Tris HCL

12.5g Tris Base

50ml 20% SDS

----adjust using water to 1L

10X Running Buffer (1L)

30g Tris Base

144g Glycine

10g SDS

Add ddH₂O to 1 liter

10X Transfer Buffer (1L)

18.9g Tris Base

90g Glycine

Add ddH₂O to 1 liter, do not adjust pH, and do not use MeOH in 1X)

10X PBS

200 NaCl

5g KCl

5g KH₂PO₄

27.8g Na₂HPO₄·2H₂O

Add ddH₂O to 1 liter.

Western Wash Buffer (PBST) (1L)

100ml 10X PBS

2ml Tween 20 (Fisher Scientific BP337-100)

900ml ddH₂O

6.3 Appendix 3: Detailed protocols for experiments

6.3.1 Preparation of plasmid DNA from bacteria

1. Incubate bacteria cultures overnight at 37 °C in Terrific Broth added with antibiotic, in flasks or tubes. (To maintain adequate aeration in the flasks or tubes, restrict the culture volume to be no more than one quarter of the total flask volume, or one fifth of the total tube volume.)

2. Pellet 1.5 mL aliquots of culture for 1 min in a microcentrifuge.

3. Remove the supernatant by aspiration and resuspend the bacterial pellet in 200 µL of GTE buffer by pipetting up and down.

4. Add 300 µL of freshly prepared 0.2 N NaOH / 1% SDS and then mix the contents of the tube by inversion until the solution clears. Then incubate on ice for 5 min.

Note: Throughout this procedure, the use of a vortex must be avoided so as to minimize shearing of the contaminating chromosomal DNA.

5. Neutralize the solution by adding 300 µL of 3.0 M potassium acetate, pH 4.8, mix by inverting the tube, and incubate on ice for 5 min.

6. Remove cellular debris by centrifuging for 10 min at room temperature, and then transfer the supernatant to a clean tube.

7. Add RNase A (DNase-free) to a final concentration of 20 µg/mL and incubate the tube at 37°C for 20 min.

8. After the RNase A treatment, extract the supernatant twice with 400 µL of chloroform. Mix the layers by hand for 30s after each extraction. Centrifuge the tube for 1 min to separate the phases and remove the aqueous phase to a clean tube.

9. Precipitate the total DNA by adding an equal volume of 100% isopropanol and immediately centrifuging the tube for 10 min at room temperature, at high speed.
10. Wash the DNA pellet with 500 μ L of 70% ethanol and then dry for 3 min.
11. Dissolve the pellet in 32 μ L of deionized H₂O, and precipitate the plasmid DNA by first adding 8.0 μ L of 4 M NaCl, and then adding 40 μ L of autoclaved 13% PEG8000.
12. After thorough mixing, incubate the sample on ice for 20 min, and then pellet the plasmid DNA by centrifugation for 15 min at 4°C in a fixed-angle rotor. The temperature parameter here is very important; adhere to the recommended 4°C.
13. Carefully remove the supernatant and rinse the pellet with 500 μ L of 70% ethanol. Then dry the pellet under vacuum for 3 min, resuspend in 20 μ L of deionized H₂O, and store at -20°C.

6.3.2 Rescue of yeast plasmids into *E. Coli*

1. Transfer a loop-full of yeast from a plate into a screw cap tube containing 250 μ l glass beads and 400 μ l of plasmid rescue buffer. It is best to use freshly grown up strains.
2. Add 200 μ l of phenol/chloroform.
3. Vortex for 3 minutes at full speed (use the vortex with the foam top)
4. Spin in micro-centrifuge at room temperature for 5 minutes.
5. Transfer aqueous phase to a snap-cap eppendorf tube containing 800 μ l of 100% EtOH. Cool at -20°C for at least 15 minutes.
6. Spin for 15 minutes at 4 °C and aspirate supernatant carefully avoiding pellet.
7. Add 500 μ l of ice-cold 70% EtOH to the tube, invert once, and spin for 5 minutes. Aspirate off the supernatant again and dry the pellet.

8. Resuspend in 40 μ l of sterile dH₂O. DNA prepared in this way can also be used as a template for PCR (1 μ l of a 5X dilution per 20 μ l of reaction volume)
9. Add 3.5 μ l of DNA to 40 μ l of E.Coli prepared for electroporation. Immediately transfer to a clean, sterile cuvette and electroporate (2.5kEV, 25 μ F, 200 ohms). Then add 300 μ l of SOC to resuspend the cells out of the cuvette and transfer to a sterile tube. Incubate at 37 °C for 30 minutes and plate out 30 and 300 μ l onto LB/AMP plates.

6.3.3 Yeast transformation

1. Overnight culture yeast at 30 °C in YPD medium
2. Dilute the overnight culture by 1:5 to 1:10 (using new YPD medium) and incubate for another 2 hours.
3. Harvest the culture by centrifuge at 3,500rpm (Eppendorf, Centrifuge 5702R) for 5 minutes, decant the culture medium.
4. Resuspend the yeast pellet with 1X TE buffer, centrifuge again at 3, 500 rpm (Eppendorf, Centrifuge 5702R) for 5 minutes.
5. Decant the TE solution, transfer the pellet to a 1.5 ml Eppendorf tube, centrifuge for another minute at 3,500rpm (Eppendorf, Centrifuge 5702R).
6. Take away the supernatant using pipet, add 20 μ l TE/LiAc/H₂O (with volume ration of 2:2:8).
7. Add 3.8 μ l of Sonic Salmon DNA (Stratagene #201190), plasmid (or DNA fragment, volume amount varies with concentration), and the competent cells we made in the previous steps. With a total final volume to be 25 μ l.
8. Add 125 μ l of PEG4000/LiAc/TE (8:1:1), vortex immediately.

9. Incubate at 30 °C for half an hour; shake the tubes every 10 minutes.
10. Add 17 µl of DMSO, heat-shock in 42 °C waterbath for 15 minutes.
11. Transfer the yeast into a 15 ml tube, add 5 ml of YPD and incubate at 30 °C for 2 hours.
12. Centrifuge the yeast; resuspend the pellet using 1X TE Buffer, then plate on appropriate plates.
13. Incubate plates for 2-4 days at 30 °C to recover the transformants. Positive clones can be further identified and tested.

6.3.4 Protein extraction from yeast for western blot analysis

1. Overnight culture up to 3ml of yeast cells in the SD-Ura (400µM Methionine) to an OD600 of approximately 0.7, measure the exact concentrations.
2. Spin cells at 2500X g for 1minute at room temperature.
3. Resuspend with 5 ml of ddH₂O, spin cells at 2500X g for 1minute at room temperature
4. Repeat step 3.
5. After the second wash, use SD-Ura-Met to resuspend the yeast pellet, decide the volume to be used according to the concentration measured at step 1, to make the resuspended yeast with OD600 of 0.7 per milliliter.
6. Keep the yeast at 30 °C shaker for 30 minutes to induce transcription.
7. Stop the transcription by adding corresponding amount of 100X methionine (pre-made).

8. Separate the yeast into different Eppendorf tubes, keep the tubes in the oven set at specified temperatures, wait until time is reached (depends on the temperatures to be tested, can be 1hr, 2hrs, or 3hrs), then it is ready to extract the protein for analysis.
9. Spin down yeast cells with 7,500X rpm (Eppendorf, Centrifuge 5415D) for 1 minute, decant the supernatant. Meanwhile, add 30 μg of glass beads to a second set of tubes preheated on 70 $^{\circ}\text{C}$ heating block. Also preheat a tube containing Thorner Buffer supplemented with 5% 2-Mercaptoethanol at 70 $^{\circ}\text{C}$.
10. Resuspend the cell pellets from step 8 with Thorner Buffer, and then transfer the cell suspension to the tubes with glass beads.
11. Incubate at 70 $^{\circ}\text{C}$ for 10 minutes. As soon as the sample is resuspended in the thorner buffer it should be transferred immediately to a tube in the heating block to inactivate proteases as quickly as possible.
12. Vortex for 4 minutes at full speed, transfer the tubes to the 70 $^{\circ}\text{C}$ heater block for another 5 minutes.
13. Spin down the debris by centrifuging for 2 minutes at full speed (13,200 rpm) (Eppendorf, Centrifuge 5415D), transfer the supernatants to a fresh tube, the samples are now ready to be checked or store in the -80 $^{\circ}\text{C}$ freezer.

6.3.5 Western blot

Nitrocellulose membrane is from GE Water & Process Technologies, catalog no. EP4HY00010

1. Run standard SDS-PAGE, run the gel at constant 100 volts. Meanwhile pre-chill the transfer buffer.

2. Assemble transfer sandwich under buffer: Pour enough 1X transfer buffer into dish to submerge gel and filter paper. Add 4 pieces of Whatman paper, sponges, piece of nitrocellulose (cut to size of gel), and gel (with extra acryl. cut away)
3. Place one piece of sponge on either side of cassette with one piece of Whatman paper each side.
4. While under solution, place gel on top of nitrocellulose and place on Whatman paper. Remove from solution and place this on positive side spong/paper.
5. Place one more piece of Whatman paper on top of gel, then place final piece of Whatman and second sponge (negative side) and close cassette gently.
6. Place positive side of cassette (red) facing positive side of cell (also red).
7. Set at 350mA, for Gal4p western blot, run for 2 hours at cold room. While for eGFP, 1 hour is enough. (in both cases, make sure current is the limiting factor)
8. Take filter out and place in PBS.
9. Pour off PBS, block for 45minutes using 5% milk dissolved in wash buffer (PBST).
10. Incubate with primary antibody (both from Sigma, rabbit origin. use 1:4000 dilution for anti-N-eGFP, and 1:5000 dilution for anti-Gal4p) in block solution (add sodium azide to 0.05%) at room temperature for one hour or overnight in the cold room.
11. Wash 3X, 10 minutes each, with wash buffer at room temperature.
12. Use HRP-anti-rabbit IgG (Invitrogen) at 1:5000 dilution in block buffer, incubate for 45 minutes, at room temperature.
13. Wash 3X, 10 minutes each, with wash buffer at room temperature.
14. Mix ECL detection reagents (pierce #1859698, 1859701) (1:1) in a 15ml tube.
15. Dry the membrane, and transfer it to a plastic wrap, add the mixed ECL solutions,

normally, I use 1ml for each membrane, keep the solution on the membrane for 1 minute, use filter paper to remove the reagents, then the membrane is ready to be exposed. We use the Fujifilm LAS-3000 Imaging System for taking pictures.

6.3.6 Bacterial transformation

1. Pipette the DNA solution from the DNA tube into E. coli bacteria tube and make sure it is labeled.
2. Put tubes on ice for 5 minutes.
3. Put tubes directly from ice into 42°C water bath for 60 seconds.
4. Put DNA tubes directly from water bath onto ice for 2 minutes.
5. With a large transfer pipet, add 0.25 ml LB broth into your DNA tube (bring solution to the first mark on the pipet.) Incubate at 37°C water bath for 10 minutes.
6. With the small transfer pipet, pipet 0.10 ml from your DNA tube onto your LB agar plate with antibiotic.
7. Spread the solutions on the plates using glass beads.
8. Put your plates in a 37°C incubator for 24 hours.

6.3.7 Electro-transformation

1. Thaw electro-competent cells on ice, keep on ice once thawed
2. Add DNA (< 5 µl) with 40ul electro-competent cells in an eppendorf tube, keep on ice for a couple of minutes.

Then transfer the cell/DNA mix into an electroporation cuvette

Note: the gene pulser should already be set properly as follows:

- time constant = 4.5 - 5.0 ms

- resistance = 200 W

- capacitance = 25 mFD

4. Pulse the cells once; the voltage display blinks, and the gene pulser beeps

5. Quickly transfer 37 °C SOC to cuvette, mix by gently pipetting up and down, and transfer SOC/cells back to culture tube; put back in ice for one or two minutes.

7. Incubate for 30 min. in 37 °C water bath.

8. Plate cells on LB plates with appropriate antibiotic.

6.4 Appendix 4: Amino acids abbreviations and linear structure formula

Name	Abbr.	Short	Linear structure formula
Alanine	Ala	A	CH ₃ -CH(NH ₂)-COOH
Arginine	Arg	R	HN=C(NH ₂)-NH-(CH ₂) ₃ -CH(NH ₂)-COOH
Asparagine	Asn	N	H ₂ N-CO-CH ₂ -CH(NH ₂)-COOH
Aspartic acid	Asp	D	HOOC-CH ₂ -CH(NH ₂)-COOH
Cysteine	Cys	C	HS-CH ₂ -CH(NH ₂)-COOH
Glutamine	Gln	Q	H ₂ N-CO-(CH ₂) ₂ -CH(NH ₂)-COOH
Glutamic acid	Glu	E	HOOC-(CH ₂) ₂ -CH(NH ₂)-COOH
Glycine	Gly	G	NH ₂ -CH ₂ -COOH
Histidine	His	H	NH-CH=N-CH=C-CH ₂ -CH(NH ₂)-COOH
Isoleucine	Ile	I	CH ₃ -CH ₂ -CH(CH ₃)-CH(NH ₂)-COOH
Leucine	Leu	L	(CH ₃) ₂ -CH-CH ₂ -CH(NH ₂)-COOH
Lysine	Lys	K	H ₂ N-(CH ₂) ₄ -CH(NH ₂)-COOH

Methionine	Met	M	CH ₃ -S-(CH ₂) ₂ -CH(NH ₂)-COOH
Phenylalanine	Phe	F	Ph-CH ₂ -CH(NH ₂)-COOH
Proline	Pro	P	NH-(CH ₂) ₃ -CH-COOH
Serine	Ser	S	HO-CH ₂ -CH(NH ₂)-COOH
Threonine	Thr	T	CH ₃ -CH(OH)-CH(NH ₂)-COOH
Tryptophan	Trp	W	Ph-NH-CH=C-CH ₂ -CH(NH ₂)-COOH
Tyrosine	Tyr	Y	HO-p-Ph-CH ₂ -CH(NH ₂)-COOH
Valine	Val	V	(CH ₃) ₂ -CH-CH(NH ₂)-COOH

6.5 Appendix 5: PCR primers list

Swap1F: 5'-

TGCTTTGCCAAGGGTACCAATGTTTTAATGGCGGATGGGTCTATTGAATG-3'

Swap1R: 5'-ATTATGGACGACAACCTGGT-3'

Met-GFP-r: 5'-ATT ACA TGA CTC GAG GTC GAC GGT ATC GAT AAG CTT CTA
TTT GTA TAG TTC ATC CAT GC-3'

Met-GFP-f: 5'-TAG ATA CAA TTC TAT TAC CCC CAT CCA TAC TGC AAA GAT
GAA CAG TAA AGG AGA AGA AC-3'

GFP108r-npct: 5'-

TGCCCTTCAGCTCGATGCGGTTCCACCAGGGTGTCCGCTCGAACTTCACCTCG
GCGCGgcaATTATGGACGACAACCTGGTTG-3'

GFP108f-npct: 5'-

ATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA

ACTACGCATGCTTTGCCAAGGGAACGAATG-3'

GFP108_pf-npct: 5'-CGCGCCGAGGTGAAGTTCGAG-3'

GFP108_pr-npct: 5'-GTAGTTGCCGTCGTCCTTGAAG-3'