Distribution Agreement

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing material of this type. I agree that permission to copy from, or to publish, this dissertation may be granted by the Professor under whose direction it was written, or in his absence, by the Dean of the Graduate School when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

_Weilin Peng_____ 01/30/2012_____

Weilin Peng Date

I: Expanding Genetic Code in *S. cevevisiae* with an Orthogonal $tRNA^{Trp/UCA}$/ Tryptophanyl tRNA Synthetase Pair

II: Synthetic Yeast Prions based on a Non-NQ-rich Amyloidogenic Sequence Derived from the NAC Protein Sequence of α-Synuclein.

By

Weilin Peng

Doctor of Philosophy, Chemistry

_____

Vincent P. Conticello, Ph.D.
Adviser

_____

David G. Lynn, Ph.D
Committee Member

_____

Dale E. Edmondson, Ph.D.
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____ Date

Abstract

I: Expanding Genetic Code in *S. cevevisiae* with an Orthogonal tRNA$^{Trp/UCA}$/ Tryptophanyl tRNA Synthetase Pair

II: Synthetic Yeast Prions based on a Non-NQ-rich Amyloidogenic Sequence Derived from the NAC Protein Sequence of α-Synuclein.

By

Weilin Peng

In an approach to expand the genetic code, a general strategy has been developed that utilizes an orthogonal tRNA and aminoacyl-tRNA-synthetase (aaRS) pair. Given the high homology of the translational machinery for eukaryotic cells, *Saccharomyces cerevisiae* is an ideal gateway host to develop an orthogonal pair for expansion of the genetic code in eukaryotic organisms. This process will facilitate the incorporation of biophysical probes into proteins for biophysical analysis or provide another pathway in the evolution of novel protein function. Therefore, in Chapter 2 of this thesis, we describe the methodology of constructing an orthogonal *Mycoplasma genitalium* tryptophanyl-tRNA synthetase (*Mgen*TrpRS)/*Mgen*tRNA^{Trp} pair for incorporation of non-canonical amino acids in *S. cerevisiae* cell in response to opal (UGA) codon. In chapter 3, a library selection scheme to identify mutant *Mgen*TrpRS variants with altered selectivity for the non-canonical 5-hydroxy-tryptophan and 5-methoxy-tryptophan is investigated and the spectroscopic properties of enhanced cyan fluorescent protein (ECFP) containing these non-canonical amino acids in the chromophore are described.

Studies of the yeast prions [*PSI*$^{+}$] and [*URE3*] have revealed the mechanism of prion formation and propagation in *S. cerevisiae*. It was recently suggested that the domains of Sup35p can be analyzed via dissection of the individual domains, such that

substituting the N/Q-rich subdomain with other high Q/N content sequences can construct novel yeast prions. The interchangeability of the NQ domain with other aggregation-prone sequences not only helps illuminate the principles of yeast prion domain architecture, but also allows prion [PSI^+] to serve as a model to explore the controllable epigenetic regulators. In Chapter 4, a non-Aβ-component (NAC) fragment from α-synuclein, which lacks the compositional bias towards Gln/Asn that is characteristic of the yeast prion domains described thus far, is explored as a structural alternative for the NQ domain of Sup35p to generate a novel chimeric yeast prion. The results demonstrated that a novel prion [NAC^+] could arise from domain substitution within Sup35p, and that the sequence of the prion-forming domain was not restricted to Asn/Gln-rich proteins. This model system might be useful for investigation of the potential for *in vivo* transmission of amyloidogenic sequences associated with protein mis-folding diseases.

I: Expanding Genetic Code in *S. cevevisiae* with an Orthogonal $tRNA^{Trp/UCA}$/ Tryptophanyl tRNA Synthetase Pair

II: Synthetic Yeast Prions based on a Non-NQ-rich Amyloidogenic Sequence Derived from the NAC Protein Sequence of α-Synuclein.

By

Weilin Peng

B.S., Hunan University, 2001

M.S., Nanjing University, 2004

Adviser

Vincent P. Conticello, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Dorctor of Philosophy in Chemistry

2012

Acknowledgements

I would first like to gratefully and sincerely thank my adviser, Vince Conticello, for his steady support, understanding and encouragement throughout my graduate research. I greatly appreciate his enthusiasm for science and for his assistance in pathing me to think independently to overcome the difficulty in my projects. The research skills that I have learned through his guidance will certainly help me throughout my future career and throughout life. Thank you so much, Vince.

I would also like to thank my committee members Dr. David Lynn and Dr. Edmondson, for their helpful guidance and assistance in both my research and my life. With the passion to science, Dr. Lynn always gave me many suggestions in my annual research report. Also, I remember no matter when I step to Dr. Demondson's office and ask for helping in research, he always be patient and discuss the details in the experiments with me.

I would like to express my deep gratitude to Jet and Holly for guiding me everything when I started my research project in the lab. A huge thank-you is given to I-lin and Tao, for their valuable efforts in my research projects. I also thank the members of Conticello's lab, Melissa Patterson, Wookhyun Kim, Steven Dublin, and Chunfu (Chris) Xu, for their supporting and friendship in the lab.

I want to thank Dr. Yuri Chernoff for providing experimental material and protocol. I wish to thank Dr. Songli Xu and Dr. Alexa L. Mattheyses for their assistance in confocal experiment. Especially thanks to Dr. Frederick H. Strobel for his technical expertise and helpful suggestion in Masspec experiment.

Thanks to all of my wonderful friends who have listened to me, supported me, and encouraged me over these long years. For all the friends I met at Emory, Yu, Yimin, Yihan, Jin Liu, Jin Wang, Fuchang, Haipeng, Guannan, Zhi, Sophie, Weiqiang, Lin, thank you so much.

Finally, special thanks are given to my family, my grandma, Ruimei, my parents, Wenming and Wenji, my husband, Yi, and the coming angel boy. Thank you for all understanding, encourage, surporting and huge love you gave me. I love you all.

Table of Contents

CHAPTER **Page**

Chapter I: Introduction 1

Expanding the Genetic Code: Adding Unnatural Amino Acids into Proteins 2

Design of a chimeric yeast prion in *S. cerevisiae* 13

Reference 22

Chapter II: Creation of an orthogonal tRNA$^{Trp/UCA}$/ tryptophanyl tRNA synthetase pair for incorporation of amino acid analogues in *S. cevevisiae* 32

Introduction 33

Experimental Procedures 40

Results and Discussion 49

An orthogonal *Mgen*tRNA$^{Trp/UCA}$ for *S. cerevisiae* 49

In vivo aminoacylation assay confirms the *Mgen*tRNA$^{Trp/UCA}$ / *Mgen*TrpRS pair is orthogonal in *S. cerevisiae* 50

The *Mgen*TrpRS specifically recognize the *Mgen*tRNA$^{Trp/UCA}$ for the incorporation of Tryptophan to the UGA mutation on the Trp66 of ECFP 51

Optimizing the suppression efficiency of the *Mgen*tRNA^{Trp} */Mgen*TrpRS pair 53

Conclusion 58

Reference 76

Chapter III: Development of a Screening System for the Expanding Genetic Code in Yeast to Incorporate Tryptophan Analogues 81

Introduction 82

Experimental Procedures 85

Result and Discussion 99

Analysis of the screening effect of *ura3-14* reporter gene 99

Site-saturation mutagenesis of *Mgen*TrpRS 101

Library screening of *Mgen*TrpRS mutants with activity toward 5-OH-Trp 103

Library screening of *Mgen*TrpRS mutants with activity toward 5-MeO-Trp 106

Spectral analysis of ECFP and UAA incorporated ECFP variants 107

Confocal microscopy image analysis of cells expressing ECFP and UAA incorporated ECFP variants 108

Conclusion 108

Reference 127

Chapter IV: Synthetic Yeast Prions Based on a Non-NQ-rich Amyloidogenic Sequence Derived from the NAC Protein Sequence of α-Synuclein. 131

Introduction 132

Experimental Procedures 135

Results 150

Cells containing the SYN1-G-Sup35 *and* NAC-G-Sup35 fusion proteins both showed two phenotypes 150

Aggregation state of NACp and SYN1p are different 152

[NAC^{+}] was reversibly curable 154

Overproduction of the NAC-NR-M-GFP induced [NAC+] phenotype 156

[*NAC*$^{+}$] was induced by cytoduction 156

[NAC+] represents an extra-chromosomal genetic element that can be confirmed with mating and sporulation experiment 159

Discussion 160

Reference 182

LIST OF TABLES

TABLE **PAGE**

Table 2-1 Table 2-1. *E. coli* and *S. cerevisiae* strains. 38

Table 2-2. The oligonucleotides for PCR amplification and DNA construction. 59

Table 2-3. The oligonucleotides used for PCR amplification and DNA construction. 60

Table 2-4. *E. coli* and *S. cerevisiae* plasmids. 61

Table 2-5. Expression schemes of orthogonal pair component carried by different plasmids for optimizing suppression efficiency. 62

Table 2-6. Mutations introduced into the eRF1 coding sequence. 63

Table 3-1. The oligonucleotides used for PCR amplification and DNA construction. .. 110

Table 3-2. The oligonucleotides used for PCR amplification and DNA construction. .. 111

Table 3-3. Plasmids used in chapter 3. 112

Table 3-4. The *Mgen*TrpRS mutations used in the chapter 3. 113

Table 4-1. Strain used in Chapter 4. 139

Table 4-2. The oligonucleotides used in Chapter 4. 140

Table 4-3. Plasmid used in Chapter 4 141

LIST OF ILLUSTRATIONS

FIGURE **PAGE**

Figure 1-1 Schematic representation of residue-specific incorporation of non-canonical amino acids (ncAAs) into proteins. 5

Figure 1-2. Schematic representation of the general method for genetically encoding non-canonical amino acids in living cells. 7

Figure 1-3. A general positive and negative selection scheme for evolving synthetase variants specific for an unnatural amino acid in *E. coli*..... 9

Figure 1-4. Selection scheme for genetically encoding non-canonical amino acids in *S. cerevisiae*. 12

Figure1-5. A nucleation event stabilizes protein conformers in an altered self-replicating prion conformation..... 15

Figure 1-6. The yeast prions [PSI^{+}] and [*URE3*] are the result of self-propagating protein conformations. 18

Figure 2-1 Proposed strategy for development of an orthogonal system for incorporation of tryptophan analogues in *S. cerevisiae* strains 39

Figure 2-2. Two-dimensional (cloverleaf) structure of the *M. genitalium* tryptophan tRNA (*Mgen*$tRNA^{Trp/UCA}$) 64

Figure 2-3. Internal promoter sequence of tRNAs..... 64

Figure 2-4. Creating a functional *Mgen*$tRNA^{Trp/UCA}$ in yeast..... 65

Figure 2-5. The *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS dependent phenotype of the *ade1-14* reporter gene. 66

Figure 2-6. Western analysis result of "mix and match" experiment I..... 67

Figure 2-7. Dependence of suppression efficiency for four different arrangements of tRNA gene expression. 68

Figure 2- 8. Phenotypic analysis of the *ade1-14* reporter gene on YPD and adenine drop-out media..... 69

Figure 2- 9. Western blot analysis of ECFP expression level in each suppression system.. 70

Figure 2- 10. FACS analysis of ECFP fluorescent strength. 71

Figure 2- 11. A set of plasmids in "Mix and Match" experiment II. 72

Figure 2- 12. The "Mix and Match" experiment II..... 73

Figure 2- 13. Phenotypic analysis of Yeast strain GT17K12 carrying wild-type or mutant *SUP45*. 74

Figure 2- 14. Western blot analysis of eRF1 mutant effects on suppresion 75

Figure 3-1. The structure of *Bst*TrpRS (PDB code: 1MB2) 114

Figure 3-2. ClustalW alignment between the sequences of *Bst*TrpRS and *Mgen*TrpRS 115

Figure 3-3. Plasmids for the construction of *Mgen* TrpRS site-saturation mutagenesis library. 116

Figure 3-4. The scheme of the first-generation library screening strategy, in which the *ura3-14* reporter is located on a low-copy plasmid pLEU2ura3-14. 117

Figure 3-5. Phentotypic assay of positive (i) and negative (ii) selection for the *ura3-14* reporter on pLEU2ura3-14 for *S. cerevisiae* strain (74D-694). 118

Figure 3-6. The second-generation strategy for library screening in which the *ura3-14* reporter is cloned into a 2μ plasmid. 119

Figure 3-7. The *Mgen* pair-dependent cellular phenotypes at 21 °C. 120

Figure 3-8. The *Mgen* pair-dependent cellular phenotypes at 30 °C. 121

Figure 3-9. The 5-OH-Trp dependent phenotype of cells carrying *MgentRNA*$^{Trp/UCA}$ */Mgen*TrpRS mutant on selective media by suppression at the *ade1-14* allele. 122

Figure 3-10. Western blot analysis of 5-hydroxy-tryptophan incorporation in S. cerevisiae strain (74D-694) containing the indicated mutant *Mgen*TrpRSs. 123

Figure 3-11. Western blot analysis of 5-MeO-Trp incorporation in *S. cerevisiae* strain 74D-694 with the *Mgen* TrpRSs mutants MeO1 and MeO7. 124

Figure 4-1. Comparison of the amino acid sequences of α-synuclein and β-synuclein.. 165

Figure 4-2. Schematic representations of the modified proteins NAC-G-Sup35**.** 165

Figure 4-3. Electron microscopy image of fibers formed by NAC-NR-M, NQ-NR-M and Syn1-NR-M. 166

Figure 4-4 (a). MALDI-TOF mass spectrum of purified NAC-NR-M protein. 167

Figure 4-4 (b). MALDI-TOF mass spectrum of purified SYN1-NR-M protein. 167

Figure 4-4 (c). MALDI-TOF mass spectrum of purified Sup35-NM protein. 167

Figure 4-5. SDS solubility analysis of His-tagged NAC-NR-M fibers. 170

Figure 4-6. Protein sedimentation assay showing the protein distribution in the supernatant (s) and pellet (p) portions. 170

Figure 4-7. The cells expressing NACp grow on YPD (a) and SD-Ade (b). 171

Figure 4-8. The fluorescent microscopy image of [NAC^{+}], [nac^{-}], [$SYN1^{+}$], [PSI^{+}] and [psi^{-}] cells. 172

Figure 4-9. Spontaneous conversion rates estimated on SD-Ade plate. 173

Figure 4-10. Deletion or inhibition of Hsp104 cures [NAC] prion. 174

Figure 4-11. Overexpression of NAC-M-GFP fusion protein induced the formation of [NAC] prion.. 175

Figure 4-12 Diffusion and dynamics of GFP-fused NACp, SYN1p and Sup35p in the stationary phase cells. 176

Figure 4-13. Diffusion and dynamics of GFP-fused NACp, SYN1p and Sup35p in the mid-log phase cells. 177

Figure 4- 14. Cytoduction experiment of [NAC] prion. 178

Figure 4- 15. Kinetics of aggregation by NAC peptide. 179

Figure 4- 16. The electron microscopy image of ulta-soniced fibers. 180

Figure 4-17. The mating and sporulation experiment of [NAC] prion. 181

ABBREVIATIONS

Amp^R	ampicillin resistance gene
bp	base pair
CRBB	Congo red binding buffer
cyh	cycloheximide
DTT	dithiothreitol
x *g*	rcf or relative centrifugal force
G418	geneticin
GuHCl	guanidine hydrochloride
h	hour
IPTG	isopropyl β-D-thiogalactopyranoside
KanMX	G418 resistance gene
kb	kilo bases
LB	Luria-Bertani broth
µg	micro gram
µl	micro liter
µM	micro molar
mg	milligram
mL	milliliter
mM	millimolar
min	minute
OD	optical density

PAGE	polyarylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
[*xxx*$^{-}$] (lower case)	non-prion state
[*XXX*$^{+}$] (upper case)	prion state
rpm	rotations per minute
SD medium	Synthetic Define minimal medium
SDS	sodium dodecylsulfate
XXX (upper case)	wild type xxx gene
xxx (lower case)	mutant xxx gene
XXXp	protein product of XXX gene
xxx::KanMX (*::*)	mutation by insertion of *KanMX*
TBS	Tris buffered saline
Tris	tris-(hydroxyl-methyl) aminomethane
aaRS	aminoacyl-tRNA synthetases
NAC	non-Aβ-component

Chapter I: Introduction

I. Expanding the Genetic Code: Adding Unnatural Amino Acids into Proteins

With the building blocks for the ribosomal protein synthesis restricted to the twenty canonical amino acids, proteins can carry out a marvelous diversity of functions. Such building blocks include a limited number of functional groups, including carboxylic acids, amides, thiol, thiol ether, alcohol, alkyl and aryl groups. (Cropp and Schultz, 2004) However, native proteins still need additional chemistry beyond the functional groups included in canonical amino acids for proper function in cells, which includes some cofactor (metal ions, thiamine and flavin) and post-translational modifications (glycosylation, methylation and phosporylation). (Walsh et al., 2005) Although many arguments have been proposed to explain the strikingly small set of the amino acids, (Di Giulio and Medugno, 1999; Ronneberg et al., 2000; Wong, 1975), it is clear that the additional building blocks can introduce more chemistries into proteins, which may allow the design of peptides and proteins with novel activities. Moreover, extra building blocks would facilitate the incorporation of biophysical probes into proteins for analysis or provide another pathway in the evolution of new organism function. (Liu and Schultz, 2010) Therefore, many approaches have been pursued to add new chemistries into the genetic code.

Chemical methodology

A number of chemical methods were employed to overcome the chemical limitation of twenty canonical amino acids. The first of these approaches is chemical

modification of proteins, which has been employed to introduce reactive molecules, biophysical probes and other chemical groups into proteins. Due to its nucleophilic thiol group and low abundance in proteins, cysteine is the most common target residue for chemical modifications, such as oxidation or alkylation. However, such methods depend on the unique reactivity of a specific amino acid, such that the overall selectivity of the chemical functionalization is limited.(Wang and Schultz, 2005)

Another well-known method for the chemical modification of proteins is solid-phase synthesis, which permits the manipulation of polypeptide structure with both canonical amino acids and unnatural amino acids. (Kent, 1988; Merrifie.Rb, 1969) Given the size limitation of this method (<100 amino acids), more efficient strategies, including semi-synthetic techniques such as chemoenzymatic synthesis and chemical ligations, were developed to produce larger full-length artificial proteins containing unnatural amino acids. (Carne, 1994; Dawson and Kent, 2000; Jackson et al., 1994) In addition, a technique named expressed protein ligation (EPL) further facilitates the protein semi-synthesis by enabling the linkage of a recombinantly expressed protein to a synthetic peptide containing one or more non-canical amino acids. (Muir, 2003; Muir et al., 1998) Nevertheless, the general application of such techniques is limited by the need for chemically orthogonal protection, the requirement of appropriate sites for cleavage and ligation, and difficulty of incorporation of unnatural amino acids at the internal sites.

In vitro biosynthetic approaches

An *in vitro* method has been developed to produce proteins containing non-canonical amino acids by taking advantage of the existing protein biosynthetic machinery

in a cell-free translation system. Because of the high fidelity of aminoacyl-tRNA synthetases (aaRS), mis-aminoacylation of the tRNAs with non-canonical amino acids is a considerable synthetic challenge. Moreover, the large number of chemically reactive ribosyl hydroxyl groups within the tRNAs presents a problem in selectivity of chemical aminoacylation. To overcome these difficulties, truncated tRNAs (with the 3'-terminal mono or dinucleotide removed) are enzymatically ligated to nucleotides that have been chemically aminoacylated with non-canonical amino acids. This method has been used successfully in the incorporation of unnatural amino acids with modified backbone, fluorophores, and photochemically reactive side chains.(Kanamori et al., 1997; Koh et al., 1997; Murakami et al., 2000) Over fifty different amino acid analogues have been incorporated into recombinant proteins using this *in vitro* biosynthetic technique, for investigation of protein structure, protein folding, enzyme mechanism and signal transduction. (Ellman et al., 1992) However, the major limitation of this technique is the scale of the process is relatively low and is not easily amenable to the production of significant amounts of the chemically modified proteins.

***In vivo* protein synthesis with auxotrophic *E. coli* strains**

For the incorporation of unnatural amino acids into proteins, the *in vivo* methods displays obvious advantages compared to the *in vitro* techniques, which include high yields of protein, technical facility and the possibility to study chemically modified proteins in the cellular environment. In the1950s, Cohen and coworkers used an auxotrophic bacterial strain that could not biosynthesize methionine to incorporate selenomethionine into the proteins. (Cohen and Munier, 1956) This selective pressure

incorporation (SPI) method, based on the use of auxotrophic *E. coli* host strain, relies on the relaxation of the substrate specificity of the native aaRSs of the host organism.

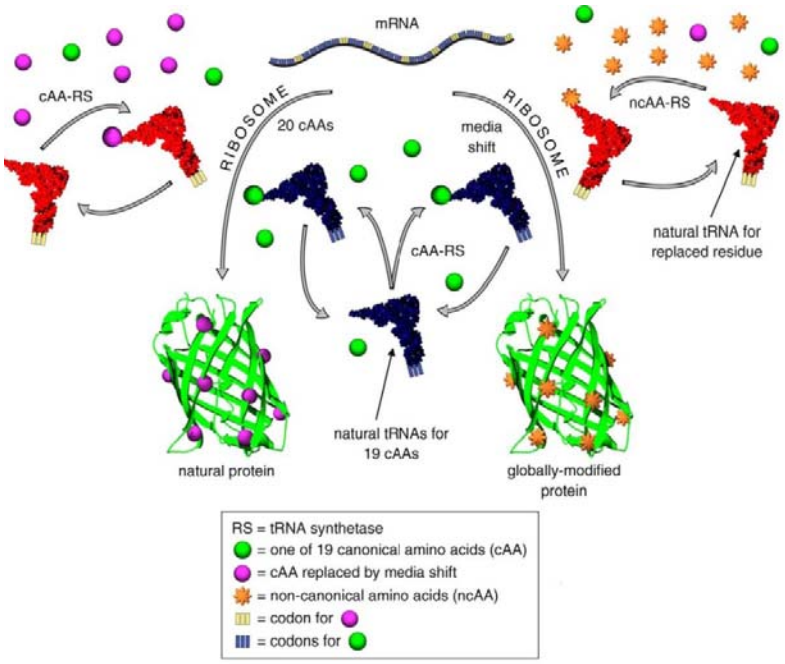


Figure 1-1 Schematic representation of residue-specific incorporation of non-canonical amino acids (ncAAs) into proteins. (Johnson et al., 2010)

A natural mRNA contains codons for the twenty canonical amino acids (cAAs). A cAA (purple sphere) assigned to one of those codons (yellow) is replaced with an ncAA (orange star). A medium shift is performed to remove the cAA to be replaced (purple sphere) and to introduce the ncAA (orange star) along with the remaining 19 cAAs (green spheres). The ncAA is charged to the appropriate tRNA (red) by either the wild-type or a mutant aminoacyl-tRNA synthetase (aaRS). The correctly aminoacylated $tRNA^{CAA}$ (blue with green sphere) and the misacylated tRNA (red with orange star) are processed by the ribosome to give a globally modified protein. The left path depicts normal protein synthesis with the twenty cAAs for comparison.

As shown in Figure 1-1, with the SPI method, when the cells are switched to the new media, in which the common amino acid is depleted and replaced with its analogue, the induction of expression results in protein that are globally-modified with unnatural amino acids with residue specificity. (Link and Tirrell, 2003; van Hest and Tirrell, 1998) In fact, SPI is a useful method with relatively high synthetic efficiency for multi-site incorporation in a target protein and the diversity of unnatural amino acids for incorporation can be increased by aaRS over expression, active-site mutagenesis, binding-domain engineering and proof-reading attenuation. (Datta et al., 2002; Doring et al., 2001; Kiick et al., 2000; Tang and Tirrell, 2002) The global incorporation of non-canonical amino acids provided powerful tools for protein structural biology. For example, replacement of the canonical methionine residues of human annexin V with the structurally similar analogues norleucine, selenomethionine and telluromethionine was used to probe the effects of side-chain packing interaction on protein stability.(Budisa et al., 1998) Substitution of tryptophan with 4-aminotryotophan in enhanced cyan fluorescent protein (ECFP) yielded proteins with novel fluorescent properties. (Bae et al., 2003) Though very useful, this method still has some limitations, lack of positional selectivity in substitution, cellular toxicity of the non-canonical amino acid, and a limitation to analogues with structural similarity to the canonical amino acids. (Liu and Schultz, 2010)

Genetic encoding of non-canonical amino acids

The site-specific incorporation of non-canonical amino acids into proteins requires the presence of a unique codon that can be decoded selectively by a tRNA that has been charged with the non-canonical amino acid *in viv*o. Unlike the replication of

DNA and the transcription of RNA, in which both the template and product have the same basic recognition elements, translation is a specific biological process in which the template (mRNA) and product (protein) are connected by an adaptor molecule (tRNA). The correlation between the two structures is the basis of the genetic code. The fidelity and efficiency of co-translational incorporation of non-canonical amino acids relies on critical molecular recognition steps in translation. First, since there are 64 possibilities for a triplet codon system and 61 are decoded by the 20 canonical amino acids, 3 "blank" codons remain for non-canonical amino acid assignment. Next, an exogenous tRNA needs to be aminoacylated with a specific non-canonical amino acid by the mutant aaRS. Finally, such aminoacyl-tRNAs can enter the ribosome of host organism and pair with the unique mRNA codon that is assigned to the non-canonical amino acids. In addition, another key requirement is that the combination of codon, tRNA and aaRS, termed as orthogonal set, must be functional in the host organism and not cross-react with other tRNAs, codons, and aaRSs. (Liu and Schultz, 2010; Wang et al., 2009) (Figure 1-2)

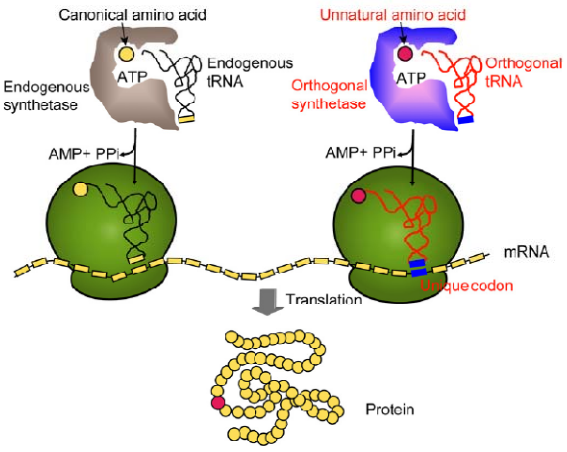


Figure 1-2. Schematic representation of the general method for genetically encoding non-canonical amino acids in living cells. (Wang et al., 2009)

Incorporation of unnatural amino acids into protein in prokaryotic cells

This general approach to site-specific introduction of non-canonical amino acids *in vivo* was developed initially using *E. coli*, which is easily genetically manipulated and has well-established protocols for selections and screens of combinatorial gene libraries. Because TAG is the least frequently employed termination codon in the *E. coli* genome, it is mostly commonly reassigned to non-canonical amino acids. Initial efforts to encode non-canonical amino acids focused on modification of *E. coli* glutaminyl-tRNA synthetase (GlnRS)/$tRNA^{Gln}$ to be orthogonal in *E. coli*. However, approximately 10% mis-aminoacylation of native $tRNA^{Gln}$ occurred using the mutant GlnRS, which precluded this approach for non-canonical amino acid incorporation. (Liu et al., 1997a; Liu et al., 1997b) Subsequently, Schultz and co-workers found that an engineered tyrosyl-tRNA synthetase and cognate nonsense suppressor tRNA from the archaea *Methanocaldococcus jannaschii* (*Mj*TyrRS/*Mj*$tRNA^{Tyr}$) was suitable as an orthogonal pair that can be used to successfully incorporate unnatural amino acids in *E. coli*. (Wang et al., 2001) To alter the substrate specificity of the orthogonal *Mj*TyrRS, a general strategy was developed for the selection of mutant synthetases displaying the desired selectivity from a combinatorial library. As shown in Figure 1-3, the initial positive selection was based on suppression of an amber stop codon in the chloramphenicol acetyltransferase (CAT) gene in the presence of the desired non-canonical amino acid A negative selection was applied subsequently to surviving cells, which was based on the suppression of an in-frame amber codon within the coding sequence of the toxic barnase gene. In this negative selection step, cells are grown in the absence of unnatural amino

acids so that all cells with mutant *Mj*TyrRS that can aminoacylate endogenous amino acids die and only cells with *Mj*TyrRS variants that charge unnatural amino acids leave. Using this selection strategy, more than 30 novel amino acids have been successfully added to the genetic code of *E. coli*. (Wang et al., 2006) In addition to the *Mj*TyrRS, other orthogonal pairs have been developed for *E. coli* as well, including the leucyl-tRNA synthetase from *Methanobacterium thermoautotrophicum*, the glutamyl-tRNA synthetase from *Pyrococcus horikoshii,* E188K mutant aspartyl-tRNA synthetase from yeast, and the pyrrolysyl-tRNA synthetase from *M. barkeri* or *M. mazei*. (Ambrogelly et al., 2007; Anderson and Schultz, 2003; Blight et al., 2004; Pastrnak et al., 2000; Santoro et al., 2003)

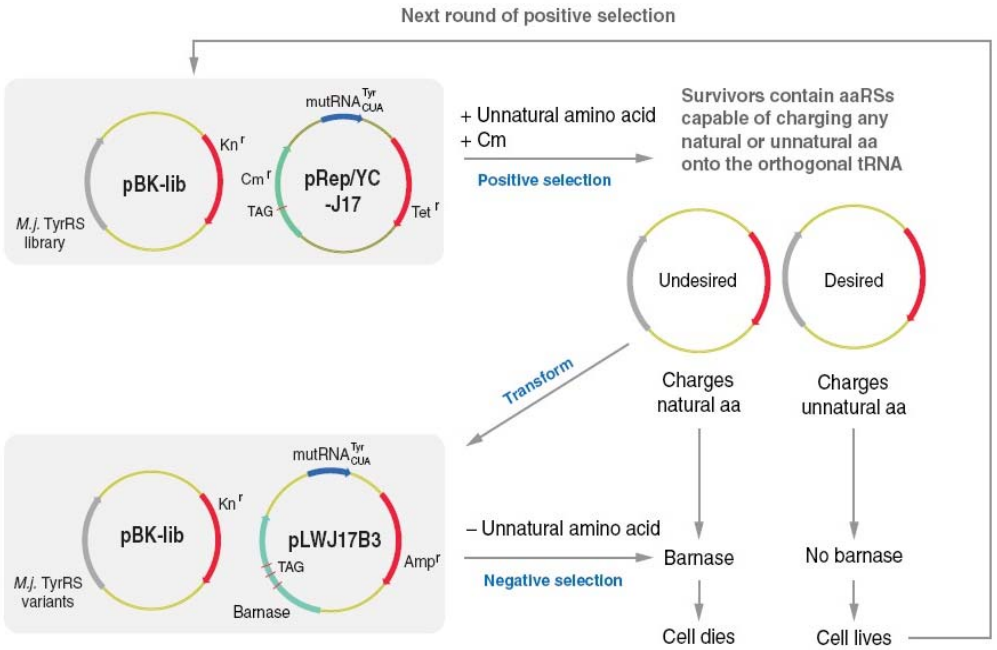


Figure 1-3. A general positive and negative selection scheme for evolving synthetase variants specific for an unnatural amino acid in *E. coli*. (Cm, chloramphenicol) (Wang et al., 2006)

Genetically encoding unnatural amino acids in yeast

In order to extend this orthogonal pair strategy for genetic code expansion to eukaryotic organisms, the yeast *S. cerevisiae* was examined as a possible host organism. This yeast has several advantages as a host organism, including rapid doubling time, ease of genetic manipulation, a reasonably well-defined, proteome and transcriptome, and high genetic homology to other eukaryotic cells. Several bacterial or archaeal tRNA/aaRS pairs, including the *E.coli* TyrRS/$tRNA^{Tyr}$, *E.coli* LeuRS/$tRNA^{Leu}$ and *Methanosarcina barkeri* PylRS/$tRNA^{Pyl}$ pairs have been shown to be orthogonal in *S. cerevisiae*. (Blight et al., 2004; Chin et al., 2003; Hancock et al., 2010; Mukai et al., 2008) Unlike *E. coli*, the transcription of a tRNA in yeast depends on RNA polymerase III and its associated factors that recognize A and B box sequences, which are internal in the tRNA structure. (Galli et al., 1981) Therefore, the exogenous bacterial tRNAs that lack the A and B box sequences cannot be efficiently expressed in *S. cerevisiae* and must be modified accordingly. (Drabkin et al., 1996)

In an approach to change the amino acid specificity of the orthogonal aaRS for the incorporation of unnatural amino acids in response to an amber codon, a two-step selection scheme that is similar to the one used in *E. coli* has been employed. (Chin et al., 2003) (Figure 1-4) An *S. cerevisiae* strain (MaV203: pGADGAL) that contains the transcriptional activator protein GAL4 with two amber mutated permissive sites was used for library screening. In the positive selection, because of a *GAL4*-controlled *URA3* gene, cells that can grow in the presence of non-canonical amino acids and in the absence of uracil are the ones carrying active aaRS mutants, which can aminoacylate tRNA with non-canonical amino acids and/or canonical amino acids. Consequently, addition of 5-

FOA and removal of the non-canonical amino acids in the negative selection, cells that survived are expressing aaRS specific to the desired analogue because 5-FOA is converted to a toxic product with the expression of *URA3* gene. In this way, more than 20 unnatural amino acids have been added to the yeast genetic code in response to UAG codon. (Liu and Schultz, 2010)

The addition of unnatural amino acids to the genetic code provides another possibility in the evolution of proteins with novel properties. However, in the approach of expanding genetic code in yeast, the scope of current orthogonal pairs is still limited to mutant *E. coli* $tRNA^{Tyr}$/TyrRS, a human initiator $tRNA^{fMet}$/*E. coli* GlnRS and *M. barkeri* tRNAPyl/PylRS pairs. (Chin et al., 2003; Hancock et al., 2010; Kowal et al., 2001). In addition, non-canonical amino acids have been inserted into proteins only in response to the amber (UAG) codon. Therefore, in Chapter 2 of this thesis, we describe the methodology of constructing an orthogonal *Mycoplasma genetalium* tryptophanyl-tRNA synthetase (*Mgen*TrpRS)/*Mgen*$tRNA^{Trp}$) pair for incorporation of non-canonical amino acids in *S. cerevisiae* cell in response to opal (UGA) codon. In chapter 3, a library selection scheme for the identification of mutant *Mgen*TrpRS variants with altered selectivity for the non-canonical amino acids, 5-hydroxy-tryptophan and 5-methoxy-tryptophan, is investigated and the spectroscopic properties of enhanced cyan fluorescent protein (ECFP) containing these non-canonical amino acids in the chromophore is described.

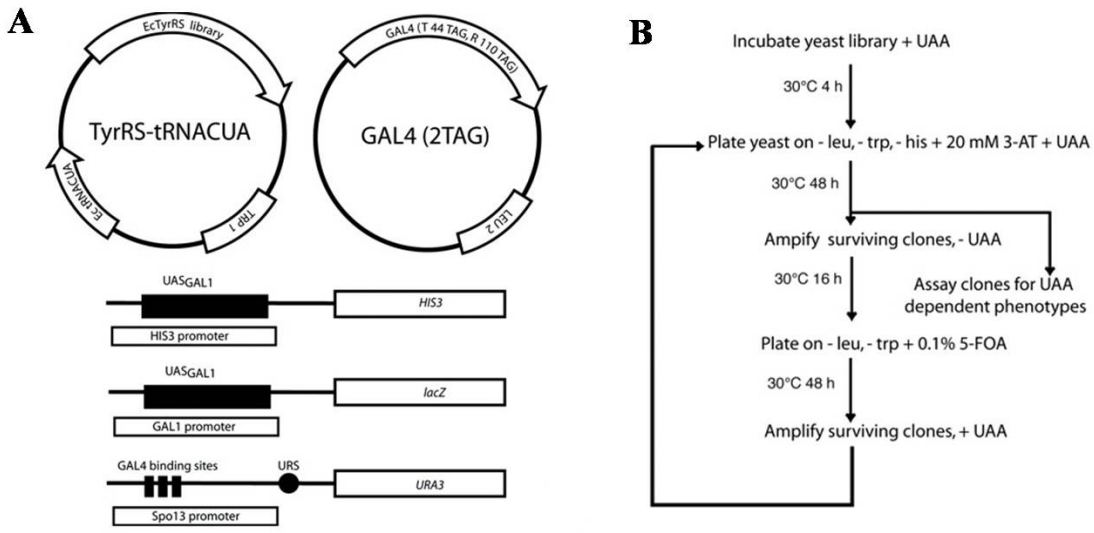


Figure 1- 4. Selection scheme for genetically encoding non-canonical amino acids in *S. cerevisiae*. (Chin et al., 2003) (A) Plasmids and reporters for the positive and negative selections. (B) The scheme used to select mutant synthetases that encode additional amino acids in *S. cerevisiae*. (UAA, unnatural amino acid)

II. Design of a chimeric yeast prion in *S. cerevisiae*

Protein mis-folding and amyloid aggregation cause diseases

Proteins are synthesized in cells on the ribosome from the genetic information encoded within DNA. The health of all organisms relies mainly on the correct folding of their constituent proteins into the three-dimensional structures that is related to protein function. Consequently, the mechanism whereby proteins adopt the appropriate three-dimensional folded structures is one of the most important questions in biological research. It is well-accepted that the amino acid sequence and the thermodynamic laws are intrinsic universal factors to determine protein folding. Whereas, the *in vivo* folding of nascent polypeptides is a co-translational process, in which protein mis-folding is a consequence of the stochastic nature of co-translational protein folding, because some regions of the incompletely folded proteins are inevitably exposed to the cytoplasm and then prone to inappropriate interaction with molecules in the extremely crowded environment of the cell,. (Dobson, 2003; Ellis, 2001) Therefore, cells develop a range of strategies to cope with such events, including the degradation using the quality control system of ubiquitin-proteasome. In addition, the presences of many molecular chaperons in the cells, which are able not only to protect proteins as they fold but also to rescue mis-folded and aggregated proteins and enable them to fold correctly. (Bukau and Horwich, 1998; Hartl and Hayer-Hartl, 2002; Taipale et al., 2010) However, the ability of the cells to handle mis-folded proteins is limited by the efficiency of the degradation machinery and the molecular chaperones. In some cases, proteins with a high propensity to mis-fold can escape from the cellular protective protein-folding mechanism. As a result, mis-folded and aggregated proteins can accumulate intracellularly or extracellularly.

Protein folding and mis-folding are the final ways of generating and abolishing specific types of cellular activity. An impairment in the folding of a given protein usually results in a reduction of the quantity and function of the protein, which may affect its cellular role and lead to disease.(Chiti and Dobson, 2006; Dobson, 2001b; Thomas et al., 1995) The largest group of mis-folding diseases is associated with the conversion of soluble proteins into fibrillar aggregates, which are generally described as amyloid fibrils, i.e., highly ordered β-sheet-rich protein assemblies.(Westermark et al., 2005) These disorders include systemic amyloidoses, such as insulin-related amyloid and lysozyme amyloidosis, and organ-limited amyloidosis, like type II diabetes. In addition, a range of protein mis-folding diseases are associated with neurodegenerative conditions, in which aggregation occurs in the brain, like Alzheimer's disease, Parkinson's disease and transmissible spongiform encephalopathies, i.e., prion diseases. (Chiti and Dobson, 2006)

The prion, an infection extra-chromosomal element

Prion proteins share an unusual property as they adopt two distinct states in which proteins of identical amino acid sequence can exist either in a functional (non-prion) or dysfunctional (prion) conformations. The prion form can serve as a template for the conformational conversion of the non-prion form. This templating effect accelerates the conversion and serves as the basis for cell-to-cell transmission of the prion. (Figure 1-5) The self-replication of conformational information enables prion proteins to act as infectious epigenetic elements. (Prusiner, 1982; Shorter and Lindquist, 2005)

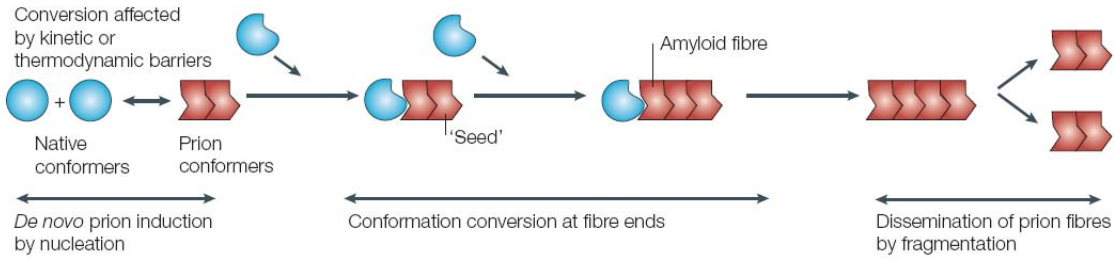


Figure1-5. A nucleation event stabilizes protein conformers in an altered self-replicating prion conformation. (Shorter and Lindquist, 2005)

The probability of nucleation is dictated by kinetic and thermodynamic considerations such that spontaneous conversion of the soluble native protein to the prion assembly occurs at a very low rate under physiological conditions. Once formed, the nucleus, or 'seed', recruits other conformers (that are probably in a transiently unfolded state) and converts them to the self-replicating conformation. The nucleus then increases in size to become an amyloid fiber and continues to convert other conformers to the self-replicating conformation at the fiber ends. Amplification of conformational replication is achieved by the fragmentation of fibres to liberate new ends. Fragmentation also allows the dissemination of infectious material.

Mammalian prion

The transmissible spongiform encephalopathies (TSE), including Creutzfeldt-Jakob disease (CJD) and bovine spongiform encephalopathy ('mad cow' deisease) are transmissible diseases associated with prion. Usually, transmissible dieases are nucleic

acid-based, but the causative agent of these prion diseases is not nucleic-acid based because it is extraordinarily resistant to nucleases and UV radiation. (Alper et al., 1967; Prusiner, 1982) Consequently, it was revealed that the biochemical purification of the infectious agent yielded a component that consisted solely of protein. This observation underlies the well-known 'protein-only' hypothesis for prion diseases. (Prusiner et al., 1983) The protein that mediates the infective self-propagation is termed the prion protein (PrP). In its normal state, PrP adopts a soluble cellular form (PrP^{C}), which undergo a conformational conversion to the abnormal protease-resistant insoluble protein form (PrP^{Sc}).(Prusiner, 1991) These two forms of the PrP protein have been characterized biochemically. PrP^{C} protein, as shown by NMR and X-ray crystallography, is rich in α-helical structure and contains little β-sheet structure (James et al., 1997; Riek et al., 1996); however, the PrP^{Sc} form of the protein has mainly β-sheet structure (Post, 1998). It is believed that the infectiousness of prion disease comes from the ability of the amyloidogenic PrP^{Sc} to serve as a template and recruit soluble PrP^{C} into PrP^{Sc} assemblies, resulting in a conformational conversion In the prion disease, such self-propagation of amyloid conformation is critical to disease progression. (Lansbury, 1999)

Fungal prions

Unlinke PrP in mammalian cells, which is disease-causing, prions found in *Saccharomyces cerevisiae* are non-pathogenic. In 1994, Wickner stated that the extra-chromosomal inheritance in yeast and other fungi could be caused by the same mechanism as in prion diseases. (Wickner, 1994a). Several different classes of yeast prions have been described that may be functionally related, but differ in the identity of the cellular protein that underlies the prion phenotype. [*URE3*], the prion form of the

Ure2 protein, is a cytoplasmically transferable element that affects nitrogen catabolism repression in yeast. Normally, yeast growing on rich nitrogen sources such as ammonium ion represses the production of proteins that are needed to metabolize poor nitrogen sources such as ureidosuccinate.(ter Schure et al., 2000) Ure2p is an antagonist of the transcriptional activator Gln3, thus, when it is bound in the [*URE3*] state, it loses its native function. Gln3 is constitutively active and the nitrogen catabolic repression is attenuated, enabling yeast to utilize ureidosuccinate as a nitrogen source, even in the presence of ammonium ion.(Uptain and Lindquist, 2002; Wickner, 1994b) (Figure 1-6)

Another well understood cytoplasmically transferable element in yeast is [PSI^+], which was identified as a yeast non-Mendelian factor in 1965 (Cox, 1965a). [PSI^+] stands for the prion form of Sup35p, a GTP-dependent polypeptide chain release factor, eRF3. In the prion state, the majority of Sup35p molecules are aggregated and inactive, resulting in higher levels of nonsense codon suppression. Therefore, the detection of [PSI^+] is usually based on the suppression of nonsense-codon mutations in auxotrophic markers in the *ade1* gene, in which the UGG codon for Trp at position 244 is replaced by the termination codon UGA (Cox et al., 1988).

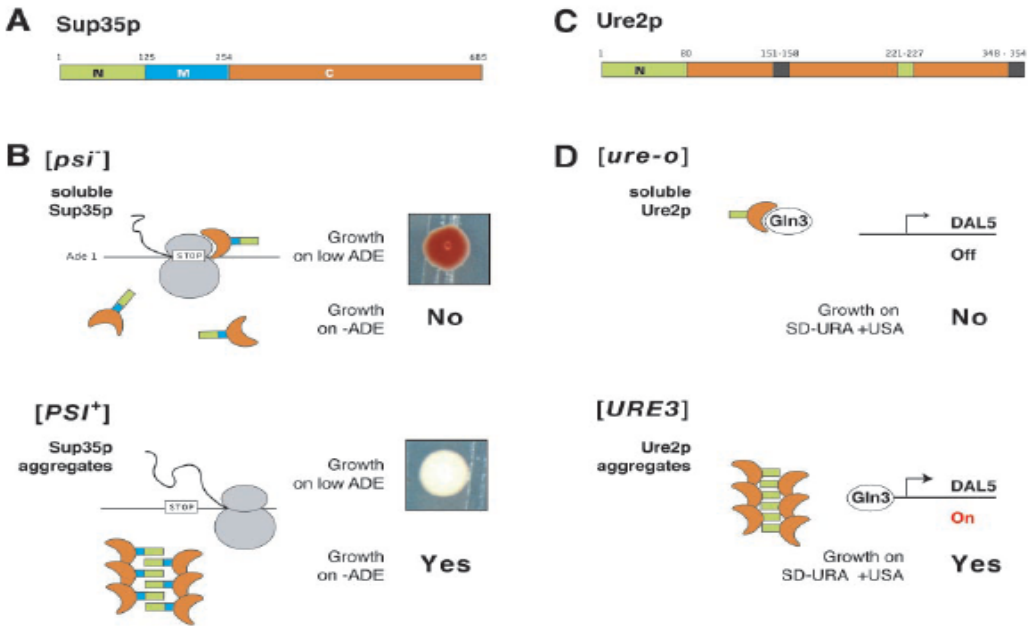


Figure 1-6. The yeast prions [*PSI*$^{+}$] and [*URE3*] are the result of self-propagating protein conformations. (Chien et al., 2004)

(*A*) Sup35p is a modular protein involved in translation termination, in which protein self-association of a specific domain is responsible for the [*PSI*$^{+}$] phenotype. The amino-terminal prion-forming domain, N (*green*), is glutamine- and asparagine-rich. The middle domain, M (*blue*), is rich in charged residues. The carboxy-terminal domain, C (*orange*), contains the essential translation-termination function of the protein. (*B*) Sup35p is soluble in [*psi*$^{-}$] yeast and able to facilitate translation termination while in [*PSI*$^{+}$] yeast; Sup35p is aggregated, resulting in suppression of nonsense codons. Translation termination can be monitored using an *ade1* reporter harboring a premature stop codon. [*PSI*$^{+}$] cells are white and capable of growth on media lacking adenine, whereas [*psi*$^{-}$] yeast accumulate a red pigment caused by lack of Ade1p and are incapable of growth on adenine-deficient media. (*C*) Ure2p is a modular protein involved in regulation of

nitrogen catabolism; self-propagating aggregation of Ure2p is responsible for the [*URE3*] phenotype. In addition to the glutamine/asparagine-rich amino terminus (*green*), Ure2p also contains another region that facilitates prion behavior (*green*) as well as segments that antagonize prion formation (*black*). The remainder of the protein (*orange*) resembles glutathione *S*-transferase and is necessary for Ure2p signaling of the presence of high-quality nitrogen sources through Gln3p. (*D*) Normally Ure2p binds the transcription factor Gln3p, preventing the up-regulation of genes, such as *DAL5*, required for uptake of poor nitrogen sources. Serendipitously, Dal5p imports not only the poor nitrogen source allantoate, but also ureidosuccinate (USA), an intermediate in uracil biosynthesis. Thus [*ure-o*] yeast cannot grow on USA-containing media lacking uracil. In [*URE3*] yeast, Ure2p is aggregated and inactive, leading to constitutive activation of Dal5p and enabling growth on USA media lacking uracil.

Cells harboring the non-prion form of Sup35 cannot suppress the *ade1-14* gene and are auxotrophic for adenine biosynthesis and unable to grow on adenine-deficient synthetic media (SD-Ade). Meanwhile, a red pigment accumulates as a byproduct of adenine biosynthesis when the cells grow on rich media, causing the formation of red colonies. In contrast, [PSI^{+}] cells are able to grow on SD-Ade media and are white on rich media as suppression of the nonsense codon results in the biosynthesis of full-length Ade1p. (Figure 1-6)

Prion domains at Sup35p

Sup35 is a 685-amino acid protein composed of three distinct domains. At the carboxy terminus, the evolutionarily conserved C domain (residue 254-685) is a GTP-

binding domain that is homologous to eukaryotic translation elongation factors and essential for protein activity and cell survival. The central M domain (residue 124-253) is a highly charged middle region that may confer solubility in the non-prion state and stabilize the Sup35 during mitosis and meiosis.(Liu et al., 2002b) The N-terminal domain (residue 1-253) is a glutamine/asparagines-rich (Q/N rich) region, which is essential for converting Sup35p to the prion state *in vivo*. (Krishnan and Lindquist, 2005; Uptain and Lindquist, 2002)

When the C domain is removed and the NM domain was fused to the glucocorticoid receptor (GR), the new fusion protein can still undergo all of the physical and epigenetic prion-like behavior of the native prion. (Li and Lindquist, 2000) However, the cellular effects were confined to genes under the control of the glucocorticoid responsive elements (GRE) that are recognized the receptor domain. Therefore, the N and M domain, which are also referred as the prion -orming domain (PrD), are sufficient to encode prion function. The critical elements for prion induction and propagation have been further narrowed to the N domain, because an M domain deletion mutant of Sup35p was able to present both prion and non-prion states, which indicates the dispensability of the M domain. (Liu et al., 2002a) There are two distinct regions in the N domains: a Q/N rich tract (residues 1-39, NQ subdomain) and an oligopeptide repeat (residue 40-112, NR subdomain) that comprise five imperfect repeats (R1-R5) and one partial repeat (R6) PQGGYQQYN, a consensus sequence to PrP.(DePace et al., 1998; Liu and Lindquist, 1999; Parham et al., 2001) Deletion of two Sup35 repeats impairs the spontaneous formation of [*PSI*$^{+}$], while addition of two extra copies of R2 increase the rate of appearance of [*PSI*$^{+}$] 5000-fold. (Liu and Lindquist, 1999) Therefore, the NR subdomain

might facilitate the correct alignment of intermolecular contact between NM molecules and interact with Hsp104, a chaperone protein that is essential for the stable inheritance of the [*PSI*$^{+}$] phenotype.(Shorter and Lindquist, 2005)

Substitution of the NQ subdomain with other high Q/N content sequence can be used to construct novel yeast prions. Santoso, *et al.*, identified a protein New1p in *S. cerevisiae* that contained a Sup35p-like *N*-terminal domain, including a Q/N-rich tract (NYN repeat) and oligopeptide repeats. This domain, comprising the first 153 residues of New1p, could be fused to the M and C domain of Sup35p to generate a new yeast prion *[NU*$^{+}$*]*. (Santoso et al., 2000) Similarly, the substitution of the NYN repeat from New1p for the NQ subdomain in Sup35p also generated a chimeric protein F that is capable of forming a prion state *[F*$^{+}$*]*. (Osherovich et al., 2004) The interchangeability of the NQ domains with other aggregation-prone sequences not only helps illuminate the principles of yeast prion domain architecture, but also allows prion [*PSI*$^{+}$] to serve as a model to explore the controllable epigenetic regulators with such prion switches. In Chapter 4, a non-Aβ-component (NAC) fragment from α-synuclein is explored as a structural alternative for the NQ domain of Sup35p to generate a novel chimeric yeast prion. This sequence lacks the compositional bias towards Gln/Asn that is characteristic of the yeast prion domains described thus far. However, the NAC peptide has been demonstrated to undergo self-association into amyloid fibers and has been isolated in the aggregated state from brain tissue of patients suffering from Alzheimer's disease and Parkinson's disease.

Reference

Alper, T., Cramp, W.A., Haig, D.A., and Clarke, M.C. (1967). Does the agent of scrapie replicate without nucleic acid? Nature *214*, 764-766.

Ambrogelly, A., Gundllapalli, S., Herring, S., Polycarpo, C., Frauer, C., and Soll, D. (2007). Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. Proc Natl Acad Sci U S A *104*, 3141-3146.

Anderson, J.C., and Schultz, P.G. (2003). Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. Biochemistry *42*, 9598-9608.

Bae, J.H., Rubini, M., Jung, G., Wiegand, G., Seifert, M.H., Azim, M.K., Kim, J.S., Zumbusch, A., Holak, T.A., Moroder, L.*, et al.* (2003). Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. J Mol Biol *328*, 1071-1081.

Blight, S.K., Larue, R.C., Mahapatra, A., Longstaff, D.G., Chang, E., Zhao, G., Kang, P.T., Green-Church, K.B., Chan, M.K., and Krzycki, J.A. (2004). Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. Nature *431*, 333-335.

Budisa, N., Huber, R., Golbik, R., Minks, C., Weyher, E., and Moroder, L. (1998). Atomic mutations in annexin V--thermodynamic studies of isomorphous protein variants. Eur J Biochem *253*, 1-9.

Bukau, B., and Horwich, A.L. (1998). The Hsp70 and Hsp60 chaperone machines. Cell *92*, 351-366.

Carne, A.F. (1994). Chemical modification of proteins. Methods Mol Biol *32*, 311-320.

Chien, P., Weissman, J.S., and DePace, A.H. (2004). Emerging principles of conformation-based prion inheritance. Annu Rev Biochem *73*, 617-656.

Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z., and Schultz, P.G. (2003). An expanded eukaryotic genetic code. Science *301*, 964-967.

Chiti, F., and Dobson, C.M. (2006). Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem *75*, 333-366.

Cohen, G.N., and Munier, R. (1956). [Incorporation of structural analogues of amino acids in bacterial proteins]. Biochim Biophys Acta *21*, 592-593.

Cox, B. (1965). [PSI], a cytoplasmic suppressor of super-suppression in yeast. Heredity *20*, 505.

Cox, B., Tuite, M., and McLaughlin, C. (1988). The psi factor of yeast: a problem in inheritance. Yeast *4*, 159-178.

Cropp, T.A., and Schultz, P.G. (2004). An expanding genetic code. Trends Genet *20*, 625-630.

Datta, D., Wang, P., Carrico, I.S., Mayo, S.L., and Tirrell, D.A. (2002). A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins. J Am Chem Soc *124*, 5652-5653.

Dawson, P.E., and Kent, S.B. (2000). Synthesis of native proteins by chemical ligation. Annu Rev Biochem *69*, 923-960.

DePace, A.H., Santoso, A., Hillner, P., and Weissman, J.S. (1998). A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. Cell *93*, 1241-1252.

Di Giulio, M., and Medugno, M. (1999). Physicochemical optimization in the genetic code origin as the number of codified amino acids increases. J Mol Evol *49*, 1-10.

Dobson, C.M. (2001). The structural basis of protein folding and its links with human disease. Philos Trans R Soc Lond B Biol Sci *356*, 133-145.

Dobson, C.M. (2003). Protein folding and misfolding. Nature *426*, 884-890.

Doring, V., Mootz, H.D., Nangle, L.A., Hendrickson, T.L., de Crecy-Lagard, V., Schimmel, P., and Marliere, P. (2001). Enlarging the amino acid set of Escherichia coli by infiltration of the valine coding pathway. Science *292*, 501-504.

Drabkin, H.J., Park, H.J., and RajBhandary, U.L. (1996). Amber suppression in mammalian cells dependent upon expression of an Escherichia coli aminoacyl-tRNA synthetase gene. Mol Cell Biol *16*, 907-913.

Ellis, R.J. (2001). Macromolecular crowding: an important but neglected aspect of the intracellular environment. Curr Opin Struct Biol *11*, 114-119.

Ellman, J.A., Mendel, D., and Schultz, P.G. (1992). Site-specific incorporation of novel backbone structures into proteins. Science *255*, 197-200.

Galli, G., Hofstetter, H., and Birnstiel, M.L. (1981). Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature *294*, 626-631.

Hancock, S.M., Uprety, R., Deiters, A., and Chin, J.W. (2010). Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. J Am Chem Soc *132*, 14819-14824.

Hartl, F.U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. Science *295*, 1852-1858.

Jackson, D.Y., Burnier, J., Quan, C., Stanley, M., Tom, J., and Wells, J.A. (1994). A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. Science *266*, 243-247.

James, T., Liu, H., Ulyanov, N., Farr-Jones, S., Zhang, H., Donne, D., Kaneko, K., Groth, D., Mehlhorn, I., and Prusiner, S. (1997). Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform (National Acad Sciences).

Johnson, J.A., Lu, Y.Y., Van Deventer, J.A., and Tirrell, D.A. (2010). Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. Curr Opin Chem Biol *14*, 774-780.

Kanamori, T., Nishikawa, S., Shin, I., Schultz, P.G., and Endo, T. (1997). Probing the environment along the protein import pathways in yeast mitochondria by site-specific photocrosslinking. Proc Natl Acad Sci U S A *94*, 485-490.

Kent, S.B. (1988). Chemical synthesis of peptides and proteins. Annu Rev Biochem *57*, 957-989.

Kiick, K.L., van Hest, J.C., and Tirrell, D.A. (2000). Expanding the Scope of Protein Biosynthesis by Altering the Methionyl-tRNA Synthetase Activity of a Bacterial Expression Host. Angew Chem Int Ed Engl *39*, 2148-2152.

Koh, J.T., Cornish, V.W., and Schultz, P.G. (1997). An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. Biochemistry *36*, 11314-11322.

Kowal, A.K., Kohrer, C., and RajBhandary, U.L. (2001). Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. Proc Natl Acad Sci U S A *98*, 2268-2273.

Krishnan, R., and Lindquist, S.L. (2005). Structural insights into a yeast prion illuminate nucleation and strain diversity. Nature *435*, 765-772.

Lansbury, P.T., Jr. (1999). Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease. Proc Natl Acad Sci U S A *96*, 3342-3344.

Li, L., and Lindquist, S. (2000). Creating a protein-based element of inheritance. Science *287*, 661-664.

Link, A.J., and Tirrell, D.A. (2003). Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition. J Am Chem Soc *125*, 11164-11165.

Liu, C.C., and Schultz, P.G. (2010). Adding new chemistries to the genetic code. Annu Rev Biochem *79*, 413-444.

Liu, D.R., Magliery, T.J., Pastrnak, M., and Schultz, P.G. (1997a). Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. Proc Natl Acad Sci U S A *94*, 10092-10097.

Liu, D.R., Magliery, T.J., and Schultz, P.G. (1997b). Characterization of an 'orthogonal' suppressor tRNA derived from E. coli tRNA2(Gln). Chem Biol *4*, 685-691.

Liu, J.-J., Sondheimer, N., and Lindquist, S.L. (2002a). Changes in the middle region of Sup35 profoundly alter the nature of epigenetic inheritance for the yeast prion [PSI+]

10.1073/pnas.252652099. PNAS *99*, 16446-16453.

Liu, J.J., and Lindquist, S. (1999). Oligopeptide-repeat expansions modulate 'protein-only' inheritance in yeast. Nature *400*, 573-576.

Liu, J.J., Sondheimer, N., and Lindquist, S.L. (2002b). Changes in the middle region of Sup35 profoundly alter the nature of epigenetic inheritance for the yeast prion [PSI+]. Proc Natl Acad Sci U S A *99 Suppl 4*, 16446-16453.

Merrifie.Rb (1969). Solid-Phase Peptide Synthesis. Adv Enzymol Ramb *32*, 221-&.

Muir, T.W. (2003). Semisynthesis of proteins by expressed protein ligation. Annu Rev Biochem *72*, 249-289.

Muir, T.W., Sondhi, D., and Cole, P.A. (1998). Expressed protein ligation: a general method for protein engineering. Proc Natl Acad Sci U S A *95*, 6705-6710.

Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K., and Yokoyama, S. (2008). Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. Biochem Biophys Res Commun *371*, 818-822.

Murakami, H., Hohsaka, T., Ashizuka, Y., Hashimoto, K., and Sisido, M. (2000). Site-directed incorporation of fluorescent nonnatural amino acids into streptavidin for highly sensitive detection of biotin. Biomacromolecules *1*, 118-125.

Nemecek, J., Nakayashiki, T., and Wickner, R.B. (2009). A prion of yeast metacaspase homolog (Mca1p) detected by a genetic screen. Proc Natl Acad Sci U S A *106*, 1892-1896.

Osherovich, L.Z., Cox, B.S., Tuite, M.F., and Weissman, J.S. (2004). Dissection and design of yeast prions. PLoS Biol *2*, E86.

Parham, S.N., Resende, C.G., and Tuite, M.F. (2001). Oligopeptide repeats in the yeast protein Sup35p stabilize intermolecular prion interactions. EMBO J *20*, 2111-2119.

Pastrnak, M., Magliery, T.J., and Schultz, P.G. (2000). A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code. Helv Chim Acta *83*, 2277-2286.

Post, K. (1998). Rapid acquisition of [beta]-sheet structure in the prion protein prior to multimer formation. Biol Chem *379*, 1307-1317.

Prusiner, S.B. (1982). Novel proteinaceous infectious particles cause scrapie. Science *216*, 136-144.

Prusiner, S.B. (1991). Molecular biology of prion diseases. Science *252*, 1515-1522.

Prusiner, S.B., McKinley, M.P., Bowman, K.A., Bolton, D.C., Bendheim, P.E., Groth, D.F., and Glenner, G.G. (1983). Scrapie prions aggregate to form amyloid-like birefringent rods. Cell *35*, 349-358.

Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R., and Wuethrich, K. (1996). NMR structure of the mouse prion protein domain PrP (121–231). Nature *382*, 180-182.

Ronneberg, T.A., Landweber, L.F., and Freeland, S.J. (2000). Testing a biosynthetic theory of the genetic code: fact or artifact? Proc Natl Acad Sci U S A *97*, 13690-13695.

Santoro, S.W., Anderson, J.C., Lakshman, V., and Schultz, P.G. (2003). An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in Escherichia coli. Nucleic Acids Res *31*, 6700-6709.

Santoso, A., Chien, P., Osherovich, L.Z., and Weissman, J.S. (2000). Molecular basis of a yeast prion species barrier. Cell *100*, 277-288.

Shorter, J., and Lindquist, S. (2005). Prions as adaptive conduits of memory and inheritance. Nat Rev Genet *6*, 435-450.

Taipale, M., Jarosz, D.F., and Lindquist, S. (2010). HSP90 at the hub of protein homeostasis: emerging mechanistic insights. Nat Rev Mol Cell Biol *11*, 515-528.

Tang, Y., and Tirrell, D.A. (2002). Attenuation of the editing activity of the Escherichia coli leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. Biochemistry *41*, 10635-10645.

ter Schure, E.G., van Riel, N.A., and Verrips, C.T. (2000). The role of ammonia metabolism in nitrogen catabolite repression in Saccharomyces cerevisiae. FEMS Microbiol Rev *24*, 67-83.

Thomas, P.J., Qu, B.H., and Pedersen, P.L. (1995). Defective protein folding as a basis of human disease. Trends Biochem Sci *20*, 456-459.

Uptain, S.M., and Lindquist, S. (2002). Prions as protein-based genetic elements. Annu Rev Microbiol *56*, 703-741.

van Hest, J.C., and Tirrell, D.A. (1998). Efficient introduction of alkene functionality into proteins in vivo. FEBS Lett *428*, 68-70.

Walsh, C.T., Garneau-Tsodikova, S., and Gatto, G.J., Jr. (2005). Protein posttranslational modifications: the chemistry of proteome diversifications. Angew Chem Int Ed Engl *44*, 7342-7372.

Wang, L., Brock, A., Herberich, B., and Schultz, P.G. (2001). Expanding the genetic code of Escherichia coli. Science *292*, 498-500.

Wang, L., and Schultz, P.G. (2005). Expanding the genetic code. Angew Chem Int Edit *44*, 34-66.

Wang, L., Xie, J., and Schultz, P.G. (2006). Expanding the genetic code. Annu Rev Biophys Biomol Struct *35*, 225-249.

Wang, Q., Parrish, A.R., and Wang, L. (2009). Expanding the genetic code for biological studies. Chem Biol *16*, 323-336.

Westermark, P., Benson, M.D., Buxbaum, J.N., Cohen, A.S., Frangione, B., Ikeda, S., Masters, C.L., Merlini, G., Saraiva, M.J., and Sipe, J.D. (2005). Amyloid: toward terminology clarification. Report from the Nomenclature Committee of the International Society of Amyloidosis. Amyloid *12*, 1-4.

Wickner, R.B. (1994a). [URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae. Science *264*, 566-569.

Wickner, R.B. (1994b). [URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae. Science *264*, 566-569.

Wong, J.T. (1975). A co-evolution theory of the genetic code. Proc Natl Acad Sci U S A *72*, 1909-1912.

Chapter II: Creation of an orthogonal $tRNA^{Trp/UCA}$/ tryptophanyl tRNA synthetase pair for incorporation of amino acid analogues in *S. cevevisiae*

Introduction

For all organisms, the genetic code defines the relationship between nucleic acid sequence and the amino acid sequence of the corresponding polypeptides. Triplet codons specify the positions of the twenty canonical amino acids that are the basic building blocks of proteins. Although these amino acids are sufficient to sustain life, it is clear that introducing unnatural amino acids into proteins can open a window for protein engineering and overcome the limitations of functional groups that are implicit within the canonical amino acids. Chemical modification and *in vitro* biosynthetic strategies have been applied to site-specifically incorporate non-canonical amino acids into proteins (Carne, 1994; Hohsaka et al., 1999), but genetically encoding unnatural amino acids into proteins provides an advantage over post-translational modifications of polypeptides due to the high fidelity of co-translational incorporation *in vivo*. To incorporate non-canonical amino acids into the existing genetic code system, it is essential to re-define the genetic code to re-assign codon(s) to the unique amino acid. A strategy of introducing a tRNA/aminoacyl-tRNA-synthetase orthogonal pair in a host organism in response to a specific DNA coding sequence was developed (Figure 2-1). In particular, suppressor tRNAs and four-base-codon tRNAs have been used after chemically or enzymatically aminoacylation to incorporate an non-canonical amino acid into specific, defined sites within the target protein (Hohsaka and Sisido, 2002). Accordingly, novel orthogonal tRNA/aminoacyl tRNA synthetase pairs have been selected and evolved from leucyl, lysyl, glutaminyl, and tyrosyl tRNA-synthethase pairs for incorporation of amino acid analogues that can be genetically encoded in bacteria, yeast and mammalian cells by

reassigning amber non-sense or frame-shift codons. (Anderson and Schultz, 2003; Kowal et al., 2001; Liu and Schultz, 1999)

Among various non-canonical amino acids, tryptophan (Trp) analogues are especially useful in the area of protein fluorescence. As a well-established method, fluorescence spectroscopy is powerful for studying protein structure, function, dynamics, and intermolecular interactions. (Beechem and Brand, 1985) The canonical amino acid tryptophan, encoded by a single UGG triplet codon, is the main reporter of intrinsic fluorescence properties of most natural proteins and peptides, and its fluorescence and phosphorescence emissions are sensitive to the microenvironment and provide high-resolution, site-specific information about protein structure and dynamics. (Budisa and Pal, 2004) Consequently, tryptophan is an especially attractive substitution target for designing and engineering proteins on the basis of several considerations. First, the diverse and rich indole chemistry offers numerous analogues. (Fig. 1) Secondly, tryptophan only has a single triplet genetic codon (UGG); it is relatively rare in native protein sequences. Thirdly, tryptophan participates in numerous interactions in proteins, such as π-π stacking, hydrogen bonding, and cation-π interaction. (Budisa and Pal, 2004) Site-directed mutagenesis has been used to study the functional roles of tryptophan in proteins. In this approach, tryptophan residues are mutated to phenylalanine or tyrosine in the attempt to minimize structural perturbations by replacing one aromatic, planar moiety with another. However, this strategy is often limited because tryptophan residues might be important for appropriate function of a protein. For example, Trp57 in green fluorescent protein cannot be replaced by any of the remaining 19 canonical amino acids. (Palm and Wlodawer, 1999) Even if such replacements were possible, local structural

perturbations would change the spectral properties. Furthermore, the replacements of Trp with Phe or Tyr may result in structural alterations due to differing side-chain volumes between the amino acids. (Hasselbacher et al., 1995) Therefore, substitution of tryptophan with structurally non-canonical similar analogues is an attractive method for the modification of protein properties with minimal structural perturbation.

Trp analogues were introduced as early as 1992 as minimally perturbing spectroscopic probes. Subsequent research efforts have focused on the incorporation of Trp analogues into recombinant proteins using the auxotrophic method. (Ross et al., 1992; Ross et al., 1997) However, the auxotrophic method is restricted to global substitution of the canonical amino acid with a structurally similar non-canonical amino acid. In an effort to create a better system for site-specific substitution, Tirrell and colleagues have employed mutant version of the *S. cerevisiae* PheRS (T415G) and a mutant yeast amber suppressor tRNA in *E. coli* strain to enable site-specific incorporation of Trp analogues into recombinantly expressed polypeptides. (Kwon and Tirrell, 2007) Because the translational machinery of eukaryotic cells differs from prokaryotic cells, a new orthogonal $tRNA^{Trp}$/Tryptophany-tRNA synthetase pair is necessary to facilitate in vivo incorporation of Trp analogues. Zhang, *et. al.*, have described the construction of a mutant suppressor tRNA in which the 5' and 3' flanking sequences of the $tRNA^{Trp}$ gene from *Arabidopsis* were employed to control the transcription of a gene encoding a modified version of *B. subtilis* $tRNA^{Trp}$ in which the native anti-codon was replaced with a UCA sequence that recognized the opal (UGA) codon. Coexpression of the suppressor tRNA with a mutant TrpRS from *B. subtilis* provided a mechanism to incorporate the non-canonical amino acid 5-hydroxytryptophan in response to opal codons encoded with

recombinant polypeptides that were expressed in mammalian cells. (Zhang et al., 2004) This method relied on conversion of a sense tRNA into a suppressor tRNA, which can potentially influence the suppression efficiency and, hence, incorporation of the Trp analogues. In contrast, we sought to employ the natural promiscuity in the genetic code to create an orthogonal $tRNA^{Trp}$-TrpRS pair that could decode the opal (UGA) codon as tryptophan, and potentially serve as a mechanism for site-specific incorporation of Trp analogues with high efficiency. Mollicutes such as *Mycoplasma* and *Spiroplasma* display a variant genetic code in which the UGA codon is decoded as tryptophan through the use of specific tRNA that recognizes the opal codon. We envisioned that the $tRNA^{Trp/UCA}$ /TrpRS pairs from these organisms might serve as the basis for the design of an orthogonal system for non-canonical amino acid incorporation in a eukaryote such as *S. cerevisiae* on the basis of sequence comparisons with bacterial aminoacyl-tRNA synthetases, which are known to be orthogonal to the corresponding eukaryotic enzymes.

The development of the orthogonal pair system required the use of a soluble protein with well-known structure as a test substrate to evaluate the efficacy of *in vivo* incorporation of the non-canonical amino acid. The cyan variant of green fluorescent protein (GFP) was chosen due to its wide application as a reporter in cellular and molecular biology. In wild-type GFP, one tryptophan residue is encoded as UGG in the gene sequence at position 57 of the corresponding polypeptide. However, mutagenesis of the UAC (Tyr) codon to the UGG (Trp) codon at the position 66 can convert the emission color from green to cyan to obtain enhanced cyan fluorescent protein (ECFP). Bae *et. al* found that substitution of both Trp57 and Trp66 with tryptophan analogues could be induced in auxotrophic strains of *E. coli* using selective pressure incorporation (SPI).

This approach enabled the substitution of the native tryptophan residues with (4-amino)-Trp *in vivo* to afford a gold fluorescent protein (GdFP), which exhibited a red-shift in emission about 50 nm compared to enhanced yellow fluorescent protein (EYFP), the most red-shifted mutant yellow fluorescent protein (YFP) (λ_{em} = 527 nm) obtained from conventional mutagenesis methods. (Bae et al., 2003) In addition, Trp66 was proved to be the only tryptophan that affects the spectroscopic properties of the protein chromophore. Therefore, single-site incorporation of tryptophan analogues to the position 66 of ECFP might be a potential pathway to extend the absorbance and emission properties of fluorescent proteins. In this approach, we describe the design of an orthogonal $tRNA^{Trp}$/tryptophanyl tRNA synthetase (TrpRS) pairs from *Mycoplasma genetalium* for the selectively incorporation of a non-natural tryptophan analogue in the ECFP in *S. cerevisiae*.

Table2-1 Table 2-1. *E. coli* and *S. cerevisiae* strains.

Strain	**Genotype**	**Reference**
TOP10F'	*recA1 araD139 endA1 (Str^R) F'{ $lacI^q$, Tn10 (Tet^R) }*	Invitrogen
GT17 ([*psi*$^-$] 74-D694)	***MAT*a** *ade1-14$_{UGA}$ his3 leu2 trp1-289$_{UAG}$ ura3* [*psi*$^-$ *pin*$^-$]	(Derkatch et al., 1997)
GT17K12	***MAT*a** *ade1-14$_{UGA}$ his3 leu2 trp1-289$_{UAG}$ ura3 sup35::KanMX* [*psi*$^-$ *pin*$^-$]	This study
YSC1021 (YBR143C)	***MATa/α*** *his3Δ1 leu2Δ0 met15Δ0 lys2Δ0 ura3Δ0 sup45::KanMX*	openbiosystem

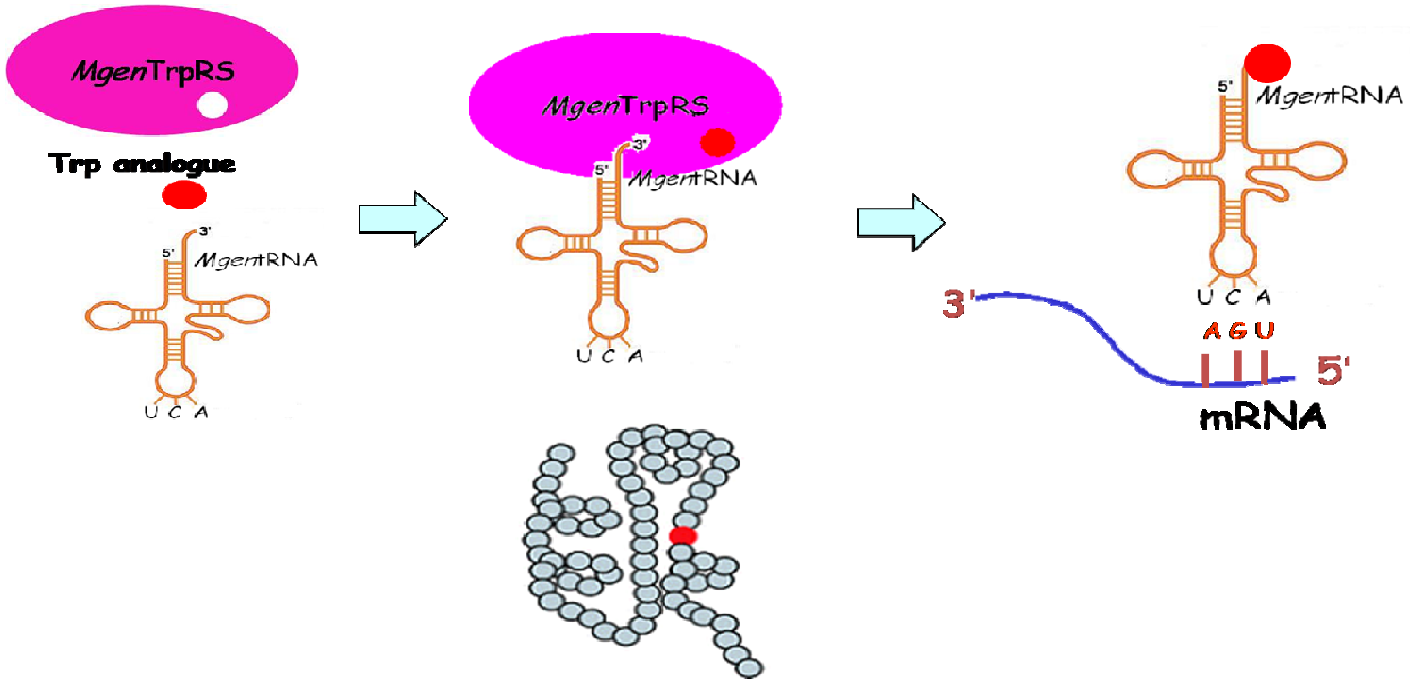


Figure 2-1 Proposed strategy for development of an orthogonal system for incorporation of tryptophan analogues in *S. cerevisiae* strains

The UGA in-frame stop codon is assigned to tryptophan analogues and decoded by the *Mgen*tRNA$^{Trp/UCA}$, in which the anti-codon loop can pair with it.

Experimental Procedures

Materials

All chemicals were purchased from Sigma-Aldrich (St Louis, MO) and Fisher Scientific Co. (Fair Lawn, NJ) unless otherwise noted. Restriction endonuclease enzymes, T4 DNA ligase, and T4 kinase were purchased from New England Biolabs (Beverly, MA). The shrimp alkaline phosphatase was obtained from Roche Applied Science (Indianapolis, IN), and Platinum *Pfx* DNA polymerase was obtained from Invitrogen Corp. (Carlsbad, CA).). Plasmids preparation kits were purchased from Qiagen (Valencia, CA). Yeast transformation kits were purchased from either Invitrogen (Carlsbad, CA) or Zymo Research (Orange, CA). The yeast shuttle expression vectors were purchased from American Type Culture Collection (ATCC#87669) and the plasmid pUC18-MgenTrpRS was purchased from ATCC#623677 as well. The pECFP-C1 vector was purchased from Clontech, Inc. (Mountain View, CA). Synthetic oligonucleotides were purchased from either Sigma-Genosys, Inc (The Woodlands, TX) or Integrated DNA Technologies (Coralville, IA) and were used as received. The prestain protein standard was purchased from Bio-Rad Laboratories, Inc. (Hercules, CA) and used for SDS-PAGE and western blotting analysis. The primary antibody used for ECFP was the Living Colors A.v. Monoclonal Antibody (JL-8) purchased from Clontech, Inc. (Mountain View, CA). The secondary antibody used was the goat anti-mouse secondary antibody, also from Clontech, Inc. The full-length ECFP were visualized by chemiluminescent detection using the 1-step NBT/BCIP reagent mixture from Pierce Biotechnology (Thermo Scientific) (Rockford, IL).

Media and growth conditions

YPD (1%, w/v, yeast extract; 2%, w/v, peptone; 2%, w/v, glucose) and 1/4YPD (0.25%, w/v, yeast extract; 2%, w/v, peptone; 4%, w/v, glucose) media were prepared in ddH2O and autoclaved for sterilization. To prepare solid media, agar (1.5%) was added to the media before sterilization. Synthetic Define (SD) minimal medium were prepared using standard protocols and minimal drop-out media lacking appropriate amino acids were prepared as needed, containing glucose or galactose as carbon source. LB media containing appropriate antibiotics were prepared using standard protocols. *S. cerevisiae* strains were grown at 30 °C and *E. coli* strains were grown at 37 °C with sufficient aeration.

Plasmids construction and strains

Construction of the Mgen*RNA$^{Trp/UCA}$ *expression vector

The gene encoding *Mgen*tRNA$^{Trp/UCA}$ cassette was assembled from three DNA segments: a 5'-leader sequence from a native *S. cerevisiae* tRNA^{Arg}, a synthetic *Mgen*tRNA$^{Trp/UCA}$ construct, and a 3' terminator from a native *S. cerevisiae* tRNA^{Arg}-tRNA^{Asp} sequence. In general, the control elements associated with bacterial tRNA sequences will not be recognized in eukaryotic hosts such as *S. cerevisiae*, therefore, the 5'-leader sequence was amplified via polymerase chain reaction (PCR) from the plasmid pScM that encoded the 5'-flanking sequence of yeast tRNA^{Arg} that directly links to a downstream tRNA sequence. [33] Restriction sites for *Hin*D III, *Acc65* I and *Bam*H I were introduced into pScM by PCR with the primers pScM-F (5'-GCGCAAGCTT ATCGATGTGTGTTATATGTACCTCTG C-3') and pScM-R (5'-CCATTGGATCC GGTACCCGTCTCATACTATTTGTTGAAGGTTTT ACTACTTTATTATAG-3'). The

PCR product was cleaved with *Hin*D III and *Bam*H I and inserted into the polylinker of pZErO-1 to get the plasmid pZErO.5linker.

Synthesis of the 3'-trailer sequence was obtained by annealing the synthetic oligonucleotides; Sc-trailer-F (GTACCCGTCTCAATTTTTTTGGCTACTCCTGTA GTTATTCTTCATTAATGCTT TGTTAACG) and Sc-trailer-R (GATCCGTTAAC AAAGCATTAATGAAGAATAACTACAGGA GTAGCCAAAAAAATTGAGACGG), which is designed to introduce *BsmB* I site upstream of 3'-trailer and downstream of the 5'-leader for incorporation of the tRNA gene. Phosphorylation of the annealed oligonucleotides with T4 polynucleotide kinase was followed by digestion of the pZErO-1 with *Acc65* I and *Bam*H I for ligation of the 3'-trailer sequence into the leader-modified pZErO-1 plasmid and get the plasmid pZErO.adapter.

The mutant opal suppressor *M. genitalium* tRNA $^{Trp/UCA}$ (*Mgen*tRNA$^{Trp/UCA}$) was accomplished by annealing the synthetic oligonucleotides *Mgen*tRNA-F (AGTAAG GGGTATAGTTCAAAGGTAGAACATCTGTCTTCAAAATAGAGTGTTGTGGGTT CGAGTCCTGCTACCCCTG-3') and *Mgen*tRNA-R (AAATCAGGGGTAGCAGG ACTCGAACCCACAACACTCTATTTTGAAGACAGATGTTCTACCTTTGAACTAT ACCCCT-3'). Annealing was carried out by dissolving the primers in distilled, de-ionized water (ddH_2O) to a final concentration of 0.5 µg/µL. Aliquots (10 µL each) of the two primers were mixed together with 5 M NaCl (4 µL), 50mM $MgCl_2$ (4 µL), and ddH_2O (172 µL). By gradually decreasing the temperature of the reactions from 99 °C to 30 ºC (decreased by 5 °C every 3 min), the DNA strands were annealed together and visualized by DNA agarose gel electrophoresis (4 % NuSieve agarose). The annealed oligonucleotides were phosphorylated with T4 polynucleotide kinase. The pZErO.adapter

with the leader and trailer was cleaved with *BsmB* I to afford acceptor for the *Mgen*tRNA$^{Trp/UCA}$ between the 5'-leader sequence and the 3'-trailer sequence. The entire *Mgen*tRNA$^{Trp/UCA}$ expression cassette was liberated from pZErO-1 by cleavage at the *Hin*D III and *Bam*H I sites flanking the *Mgen*tRNA$^{Trp/UCA}$ cassette to construct the plasmid pZErO.cassette. Next, the *Mgen*tRNA$^{Trp/UCA}$ cassette was cloned into compatible sites in the *S. cerevisiae* centromeric plasmid pRS315 (*LEU2* marker), to afford new plasmids p315.tRNA that were competent for expression of the *M. genitalium* tRNA in the host system. In addition, the *Mgen*tRNA$^{Trp/UCA}$ cassette was released by *Sal* I and *Bam*H from pRS315.tRNA and clone to pRS314 (*TRP1* marker) for constructing plasmid p314.tRNA.

Construction of the MgenTrpRS plasmid

The *M. genitalium* tryptophanyl-tRNA synthetase (*Mgen*TrpRS) gene was amplified by PCR from the plasmid pUC18, which carries the gene for *M. genitalium* tryptophanyl-tRNA synthetase. (Fraser et al., 1995) The primers *Mgen*TrpRS-F (GCTTAGGATCCAATAATGATAAAGCGCGC AATTACAGGGATTCAAGC) and *Mgen*TrpS-R (GGACCTCGAGTTAATCAAAAAGCTGATTAGATGTTAAACCA AATGC) were used to introduce *Bam*H I and *Xho* I restriction sites at the 5'- and 3'-termini respectively of the *Mgen*TrpRS gene. The *Mgen*TrpRS gene was cloned into the compatible sites within the polylinker of the *S. cerevisiae* 2μ expression plasmid p426 ADH to afford plasmid p426-TrpRS.The expression of *M. genitalium* TrpRS was placed under the control of the strong constitutive promoter from alcohol dehydrogenase (ADH). The orthogonal *Mgen*tRNA$^{Trp/UCA}$ / *Mgen*TrpRS system was constructed with the plasmids p315.tRNA/p426.TrpRS or p314.tRNA/p426.TrpRS.

Construction of the orthogonal pair plasmid p426.pair-F and p426.pair-R

It was first necessary to introduce *Avr* II and *Nhe* I restriction endonuclease sites in the plamid pRS426 ADH for cloning the tRNA gene cassettes. Inverse PCR amplification was used to introduce the sites into the plasmid using the ADH-F (CGAATCACCGGTCCTAGGGGCGACGTGGCGAGAAAGG) and ADH-R (AGCTACCGGTGCTAGCGGCTTTCCCCGTCAAGCTCTAAATCG) primers. These primers were designed to introduce an *Age* I cleavage site between the *Nhe* I and *Avr* II sites in the plasmid for linearization and subsequent re-ligation of the plasmid after inverse PCR amplification of the entire plasmid. The final product, p426-AgeI, contained the *Nhe* I and *Avr* II sites in a non-coding region of the plasmid for cloning of the *Mgen*tRNA$^{Trp/UCA}$ cassette. Because the *Nhe* I and *Avr* II sites provide identical sticking end, the tRNA cassettes was cloned to p426-AgeI with forward or reverse tracks to construct p426-tRNAF and p426-tRNAR and the direction of tRNA were confirmed by DNA sequencing analysis. Sequentially, the *Mgen*TrpRS gene was introduced to the tRNA-modified plasmids with *Xho* I and *Bam*H I restriction sites to generate p426.pair-F and p426.pair-R; in this case, the whole orthogonal pair were located on a high-copy 2μ yeast vector for constitutive expression.

Construction of ECFP expression plasmids

The Trp66UGA$_{STOP}$ mutant was amplified from the pECFP-C1 by PCR-based quick change site-directed mutagenesis method with primers CFPW66stop-F (CTCGTGACC ACCCTGACCTGAGGCGTGCAGTGCTTCAGC) and CFPW66stop-R (GCTGAAGC ACTGCACGCCTCAGGTCAGGG TGGTCACGAG). And then, the gene encoding the wild-type ECFP and the opal mutant, W66, was amplified with the primers CFP-F (AGGTTAGGATCCAATAATGGTGAGCAAGGGCGAG) and CFP-R (GCTTCC TCGAGATTACTTGTACAGCTCGTCCATGCCG), which are designed to

generate *Xho* I and *BamH* I restriction sites to flank the gene of ECFP variants for cloning into the inducible expression vector p423 GAL1, generating plasmid p423.ECFP and p423.W66.

Knockout chromosomal eRF1allele

A PCR-based gene deletion strategy was employed to the strain GT17 to disrupt and substitute the SUP45 gene at the chromosome, which encodes the eRF1. (Baudin et al., 1993; Carne, 1994) The maintenance plasmid pUKC802 was transformed into the strains before the chromosomal gene disruption, as the presence of SUP45 gene or eRF1 activity is essential for cells to survive. In the first round of PCR, UPTAG-PRIMER and DNTAG-PRIMER were used to amplify *sup45::KanMX* from strain YSC1021 (ORF, YBR143C); sequentially, the SUP45-UP-F and SUP45-DN-R primers were used for the second round amplification to the PCR product in the first-round reaction. The final integrative PCR product carrying the *sup45::kanMX* cassette was transformed into GT17 strain harboring pUKC802 to replace the chromosomal SUP45 gene. Transformants were then selected for the successful gene replacement, resulting from the spontaneous DNA recombination in yeast, on YPD medium containing 0.2 mg/mL geneticin (G418 sulfate). The replacement of endogenous gene was verified in mutant cells that grew on YPD+G418 by the presence of genomic PCR products of the expected size using primer sets SUP45-UP-F/KanB and KanC/ SUP45-DN-R, which span the left and right junctions of the deletion module within the genome. The mutant strain was named as GT17K12 to represent GT17 with chromosomal *SUP45* gene knockout.

A plasmid shuffling procedure was used to exchange wild-type *SUP45* alleles harbored by pUKC802. All mutants listed in Table 2-5 were obtained by Quickchange mutagenesis PCR using pDB800 as template and then subcloned back to pDB800 after sequencing to

generate pDB800.Mutant. The plasmids carrying mutant *SUP45* alleles were transformed to GT17K12. While the transformants were selected on SD-Leu medium, they were plated to SD-Leu medium containing 5-fluoroorotic acid (5-FOA) to select for cells that had lost the URA3-based plasmid pUKC802. Single colonies were then picked from the 5-FOA-containing medium for further characterization.

Strians

Unless specified, *E. coli* strain TOP10F' was used as the bacterial host to carry out the molecular cloning. *Saccharomyces cerevisiae* strains GT17 ([*psi*$^-$], [*pin*$^-$]), which is an epigenetic variant of *S. cerevisiae* strain 74D-694 (*MATa ade1-14UGA his3 leu2 trp1-289UAG ura3*), were kindly provided by Professor Yuri Chernoff and used for *in vivo* opal suppression of ECFP with the orthogonal *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS system. GT17K12 was used for screening effect of mutant eRF1 variants on the suppression efficiency.

In vivo aminoacylation assay

A series of plasmids (Table 2-4) were constructed to harbor the components of the orthogonal pair and transformed to yeast strain GT17 using the Frozen-EZ Yeast Transformation II Kit from Zymo Research. Single colonies were picked out from selective amino acid dropout transformation plate and streaked on YPD or additional adenine dropout SD plates to validate amimoacylation.

Protein expression in GT17

The plasmids carried orthogonal pair or relative empty plasmid were co-transformed with plasmids holding target protein gene (wild-type or W66stop mutant ECFP), as listed in Table 2-4, to the yeast strain GT17. Cells were grown in synthetic dropout medium (Table 2-4) + 2 % glucose at 30 °C overnight as starter culture and then diluted 1 to 20 volumes to a final volume of 5 mL of culture. When the OD_{600} reading reached 0.8-1 A.U., the cell culture were centrifuged at 1000*g* for 10 min at 4 °C and washed twice with equal volume sterile cold 0.9% NaCl. Next, the cells were re-suspended in SC dropout medium + 2% galactose to induce the expression of the ECFP variants was induced. Cells were harvested after 8 h or 16 h of expression.

Western-Blot Analysis of ECFP expression

The ECFP expression was induced consistently when the OD_{600} reading reached 0.8-1.0. After 8-16 hours expression, 2 mL cells were collected by centrifugation at 2000*g* for 10 min at 4 °C, and washed with 2 mL sterile cold 0.9% NaCl. Approximately 6×10^7 cells (OD_{600} = 1.0 A.U., 2mL) were collected and then lysed by Y-PER Yeast Protein Extraction Reagent (Thermo Scientific) to get the whole-cell soluble proteins. Alternatively, the whole cellular protein can be extracted with yeast lysis buffer (50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 10 mM KCl, 0.1 mM EDTA, 1mM DTT) at a concentration of ~3 x 10^6 cells/μl and the protease inhibitor cocktail for yeast (Sigma) was added at a concentration suggested by the supplier. An equal volume of 425- to 600-μm acid-washed glass beads were added to the cell suspension followed by vigorous agitation to lyse cells. After the lysate was centrifuged at 2000*g* for 10 min at 4 °C to pellet unbroken cells and cellular debris, the supernatant was carefully transferred to a

new tube to centrifuge at 16000*g* for 20 min at 4 °C to separate the soluble and pellet portions. The soluble portion was use for SDS-PAGE analysis.

Cell lysates (10 μL) were analyzed by SDS-PAGE (12%) using a Biorad Mini-PROTEAN II electrophoresis system, and transferred to nitrocellulose membrane. The ECFP level was probed by the Living Colors A.v. Monoclonal Antibody (JL-8) as primary and followed by the alkaline phosphatase-conjugated goat anti-mouse as secondary antibody. The blots were visualized with 1-step NBT/BCIP reagent mixture.

Fluorescence-activated cell sorting (FACS) and analysis

The expression level of ECFP in living yeast cells was quantitatively monitored by FACS analysis, which was carried out on a DSB LSR II system manufactured by Becton Dickinson (BD Biosciences, San Jose, CA). The 405 nm Violet laser was used for excitation of ECFP with a 470/15 filters.

After expression, aliquots (2 mL) of each culture were centrifuged at 1000*g* for 10 min, at 4 °C. The cell pellets were washed twice and suspended in PBS (8 g/L NaCl, 0.2 g/L KCl, 0.24 g/L KH_2PO_4, 1.44 g/L Na_2HPO_4, pH 7.4). Before FACS analysis, the cell density was adjusted with PBS to OD_{600} = 0.5 A.U. and the cell suspension were briefly sonicated. In general, high-throughput sampling module was use and 10,000 cells were counted for each analysis.

Results and Discussion

An orthogonal *Mgen*tRNA$^{Trp/UCA}$ for *S. cerevisiae*

A common method to genetically encode unnatural amino acids requires introducing a heterologous orthogonal tRNA that is not recognized by the endogenous aminoacyl-tRNA synthetases (aaRSs) of the host system. We employed the tRNA^{Trp} from *M. genitalium* (*Mgen*tRNA$^{Trp/UCA}$) because it encodes tryptophan using the opal (UGA) codon as well as the more common UGG codon. Moreover, two distinct tryptophanyl-tRNAs are employed to decode the respective codons. Since this bacterial tRNA is potentially orthogonal to the eukaryotic host *S. cerevisiae*, we envisioned that *Mgen*tRNA$^{Trp/UCA}$ might function effectively as an opal suppressor for co-translational insertion of tryptophan analogues. In the yeast cells, the *in vivo* expression of a heterologous tRNA depends on the transcription by RNA polymerase III, which recognizes intragenic elements in the tRNA coding sequence, the A and B box sequence from +8 to +19 and +52 to +62. (Francis and Rajbhandary, 1990; Murphy and Baralle, 1983) In addition, a 5'-flanking sequence is also necessary to the transcription of tRNA in eukaryotic system. (Geiduschek and Tocchini-Valentini, 1988) Although the wild-type *Mgen*tRNA$^{Trp/UCA}$ contain both A and B box, it still relies on 5'-flanking sequence to be efficiently transcribed in yeast. Previous research has shown that the precursor tRNA^{Arg} in the yeast polycistronic tRNA^{Arg}-tRNA^{Asp} gene can activate a down-stream transcriptionally inactive human initial tRNA or a yeast tyrosine suppressor tRNA gene, which can substitute for the tRNA^{Asp} gene. (Francis and Rajbhandary, 1990; Straby, 1988) Thus, the yeast tRNA^{Arg} sheds light on the 5'-leader sequence and the 3' trailer sequence characteristics that are required for efficient transcription of the di-cistronic tRNA^{Arg}-tRNA^{Asp} gene. These elements were employed direct the transcription of a synthetic

DNA cassette encoding *Mgen*tRNA$^{Trp/UCA}$. The entire transcriptional cassette for *Mgen*tRNA$^{Trp/UCA}$ was 363 bp in length (Fig. 2-4). In order to assess the suppressor-tRNA properties of *Mgen*tRNA$^{Trp/UCA}$, we began the analysis with an orthogonal system in which was cloned into a relatively low copy number centromeric plasmid pRS315. Subsequently, the *Mgen*tRNA$^{Trp/UCA}$ cassette was cloned in conjunction with the corresponding *Mgen*TrpRS gene into the high copy-number 2μ yeast expression vector. The synthetase gene was expressed under the control of constitutive ADH promoter to selectively aminoacylate *Mgen*tRNA$^{Trp/UCA}$ in yeast.

In vivo aminoacylation assay confirms the *Mgen*tRNA$^{Trp/UCA}$ / *Mgen*TrpRS pair is orthogonal in *S. cerevisiae*

In the *S. cerevisiae* strain GT17 (*MATa ade1-14 his3-Δ200 leu2-3, 112 trp1-289 ura3-52 Ψ pin*$^{-}$), the *ade1*-14 allele contains a mutation at codon 244 that converts a Trp codon (UGG) to an opal termination signal (UGA). (Chernoff et al., 1995b) If the stop codon UGA in *ade1-14* is successfully suppressed, tryptophan can be incorporated at this position. The stain is able to grow in SD–Ade medium and exhibits a white color on YPD solid media. In contrast, if the UGA codon cannot be suppressed, the stain cannot grow in SD–Ade medium and exhibits a red color on YPD solid media. Therefore, suppression of the *ade1-14* nonsense (UGA) mutant is a convenient method to screen for aminoacylation of *Mgen*tRNA$^{Trp/UCA}$ by *Mgen*TrpRS. No detectable suppression of *ade1-14* was observed over the genomic background of *S. cerevisiae* strain GT17 in cells harboring empty plasmids, which indicates that the opal codon cannot be suppressed in a blank cellular background. (Fig. 2-5) Conversely, the read-through of the opal mutation in

ade1-14 was observed only when the GT17 strain was transformed with plasmids carrying both *Mgen*tRNA$^{Trp/UCA}$ and *Mgen*TrpRS gene. This result showed that both *Mgen*tRNA$^{Trp/UCA}$ and *Mgen*TrpRS can be expressed at a proper level to suppress the opal mutation. (Fig. 2-5) To confirm that the exogenous *Mgen* pair determines the suppression and is orthogonal to the yeast cells, further mix-match comparisons were tested. As shown in Figure 2-5, when only one component of the *Mgen* pair is present in the GT17 strain, the UGA mutation in *ade1-14* cannot be suppressed and the dysfunctional truncated version of the protein *N*-succinyl-5-aminoimidazole-4-carboxamide ribotide (SAICAR) synthetase, encoded by *ade1*, interrupts the purine synthetic pathway. The GT17 strains display a red color phenotype on YPD solid media, an inability to propagate on SD-Ade media. These mix-match experiments confirm that the *Mgen*tRNA$^{Trp/UCA}$ is not recognized by any of the endogenous yeast aaRS and the *Mgen*TrpRS does not cross-charge any tRNAs in yeast cells. In this way, we identified the *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS pair is orthogonal in the *S. cerevisiae* cells, and, thus, is suitable as a system to decode the UGA codon and incorporate tryptophan or tryptophan analogues into protein.

The *Mgen*TrpRS specifically recognize the *Mgen*tRNA$^{Trp/UCA}$ for the incorporation of Tryptophan to the UGA mutation on the Trp66 of ECFP

The efficacy of suppression depends strongly on nucleotide sequence flanking the termination codons (i.e., codon context), such that wide variations in suppression efficiency, hence the co-translational insertion of non-canonical amino acids, can be observed for different sites within the protein coding sequence. (Bertram et al., 2000;

Bonetti et al., 1995) The opal stop codon in *ade1-14* allele has been previously selected as a phenotypic marker on the basis of ease of suppression within the encoded nucleotide context, which has an important effect on the recognition of translation release factor in *S. cerevisiae*. In order to further analyze the suppression efficiency of the orthogonal *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS pair, we employed a reporter gene encoding a modified version of the enhanced cyan fluorescent protein (ECFP). A Trp66UGA mutation was introduced into the coding sequence of ECFP, such that suppression efficiency could be monitored not only via expression level of the recombinant protein but also from recovery of the fluorescent signal. The wild-type and mutant ECFP ($ECFP_{W66UGA}$) genes were amplified from pECFP-C1 vector and cloned into compatible restriction sites within the polylinker region of the high-copy 2μ yeast expression vector p423 under the control of the inducible promoter GAL1. Western blot analysis of the whole-cell lysates indicated that expression of full-length ECFP was observed from the plasmid p423.W66STOP only in *S. cerevisiae* cell lines that supported co-expression of both the *Mgen*tRNA$^{Trp/UCA}$ and the *Mgen*TrpRS genes. In the absence of the latter, the S. cerevisiae strains could not produce full-length ECFP. (Figure 2-6) Taken together, these results indicate that the expression of full-length ECFP depended on the presence of the *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS pair. The data suggest that the *Mgen*tRNA^{Trp} is successfully transcribed with the yeast transcriptional apparatus and aminoacylated specifically by the cognate *Mgen*TrpRS but not by other aaRS in yeast, the opal mutation is efficiently suppressed, which resulted in expression of full-length ECFP. Although $ECFP_{UGA}$ expression level is noticeably lower than that of wild-type ECFP, the protein yield may be affected by competition with the eukaryotic release factors (*vide infra*). Nevertheless, the efficiency of suppression mediated by *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS

pair is sufficient to produce detectable levels of target protein expression. This experiment further demonstrates that the *Mgen*tRNA^{Trp}/*Mgen*TrpRS pair is able to function with the translation machinery of yeast to genetically decode the UGA codon with site-specific fidelity.

Optimizing the suppression efficiency of the *Mgen*tRNA^{Trp} /*Mgen*TrpRS pair

Many factors may affect the incorporation of unnatural amino acids with *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS, including the *Mgen*tRNA$^{Trp/UCA}$ expression level and the competition of translational release factors with *Mgen*tRNA$^{Trp/UCA}$ for binding to UGA stop codon ((Kohrer et al., 2004; Nakamura et al., 2001). As reference, we assigned the first expression system mentioned above as system A. In system B, the *Mgen*tRNA$^{Trp/UCA}$ was cloned to the centromeric plasmid pRS314 that carries *TRP1* maker to enable that the cells can accumulate tryptophan endogenously without transport from the culture medium. Interestingly, when the ECFP$_{UGA}$ expression had been introduced for 16 hours, the cells harboring the *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS pair constructed in system B exhibited a color that was redder than those in system A, in which the cells take up tryptophan exogenously. (Figure 2-8) Thus, endogenous production of tryptophan doesn't improve the suppression efficiency. In this system, the production of cellular tryptophan depends on the efficacy of *TRP1* gene expression, which may not as good as the effect of transporting tryptophan directly from medium. To increase the amount of *Mgen*tRNA$^{Trp/UCA}$, the tRNA gene was introduced to the high copy-number 2μ plasmid p426ADH.AgeI. Although overproduction of the suppressor tRNA gene might lead to leaky suppression of the UGA codon, (Kim et al., 1990) the increased gene dosage of *Mgen*tRNA$^{Trp/UCA}$ due to the higher plasmid copy-number did not cause detectable expression of the target gene in the absence of the corresponding tryptophanyl-tRNA

synthetase. As shown in Figure 2-10, cells harboring the plasmid p426.tRNAF failed to grown on the SD-Ura-Ade solid media and exhibited a red color on YPD solid media, which indicated that the UGA nonsense codon in the sequence context of *ade1-14* cannot be suppressed under these conditions.

The gene encoding *Mgen*tRNA$^{Trp/UCA}$ can be cloned into the p426.Age1 plasmid in either of two orientations. The adjacent *Mgen*tRNA$^{Trp/UCA}$ and *Mgen*TrpRS ORFs were cloned in a configuration in which the sense strands had a convergent (→←) orientation or a tandem (→→) orientation. The plasmid in which the genes were organized in a convergent orientation was named as p426.pair-R (expression system D), while the plasmid in which the genes were oriented in a tandem configuration was designated p426.pair-F (expression system C). We observed the growth of cells carrying p426.pair-F (Figure 2-8 C) or p426.pair-R (Figure 2-8 D) in the absence of adenine; however, the colony color of cells maintaining plasmid p426.pair-R appeared more on YPD solid media, in contrast to the white colony color of cells harboring p426.pair-F. This result suggested that the suppression efficiency of the *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS was more efficient when the corresponding transcriptional units had a tandem orientation as in plasmid p426.pair-F. Notably, the colony color cells harboring p426.pair-F was observed to be whiter than that of cells maintaining the p315-tRNA/p426-TrpRS combination of plasmids, which suggested that the higher copy number of the 2μ plasmid increased the gene dosage of *Mgen*tRNA$^{Trp/UCA}$ and effectively increased the suppression efficiency for the UGA codon in the *ade1-14* reporter gene. (Figure 2-8) However, we observed that the colony color of cells with p426.pair-R appeared pinker than that of cells carrying p315-tRNA/p426-TrpRS despite the higher copy number of *Mgen*tRNA$^{Trp/UCA}$. This result might arise from transcriptional interference in the convergent gene.(Callen et al., 2004;

Prescott and Proudfoot, 2002) The accumulated results suggested that the best suppression efficiency was obtained with the expression system C, in which the transcriptional units for the synthetase and tRNA were oriented in tandem within the expression plasmid.

Next, we assessed the efficacy of the orthogonal pair constructs listed in Table 2-4 using the expression of the gene encoding $ECFP_{UGA}$ as a test system. The sequence context of the UGA codon differs from that in the *ade1-14* gene, which provides a method to evaluate the generality of our approach we hypothesized that the expression level of full-length ECFP in induced cultures could be correlated with suppression efficiency for the opal codon in the different plasmid constructs. Expression cultures were divided into two aliquots; one of which was subjected to FACS analysis and the other one was employed for SDS-PAGE and Western blot analysis with the monoclonal anti-GFP antibody. (Figure 2-9 and 2-10) Consistent with the *ade1-14* suppression assay, the cell line harboring the p426.pair-F plasmid exhibited the highest mean fluorescence in the FACS analysis and displayed the strongest band on Western blot, which indicated that this cellular construct produced the highest level of ECFP among the orthogonal pair expression systems. Consequently, we conclude that this construct displays the highest suppression efficiency for the opal codon. Construct C, in which the cells harbor the p426.pair-F, was chosen as the host system for screening of mutant TrpRS libraries for non-canonical amino acid incorporation studies (see Chapter 3).

In addition to the *Mgen*$tRNA^{Trp/UCA}$ level, the competition between *Mgen*$tRNA^{Trp/UCA}$ and the eukaryotic release factor (eRF1) is another important consideration for suppression efficiency. Eukaryotic release factor (eRF1) mutants have

been described that exhibit a bias towards recognition of UAA/UAG rather than UGA as termination signals. (Bertram et al., 2000) These uni-potent release factor mutants might display a lowered ability to compete with the suppressor tRNA *Mgen*tRNA$^{Trp/UCA}$ and facilitate the co-translational incorporation of tryptophan (and tryptophan analogues) in response to genetically encoded UGA codons within a target protein sequence. In *S. cerevisiae*, the *sup45* gene encodes the eRF1 release factor; accordingly, a yeast strain, GT17K12, was constructed that carries a knock-out of the chromosomal copy of the *sup45* gene. Since *sup45* is an essential gene, an extra-chromosomal copy was maintained on the *URA3* plasmid pUKC802 in which the expression of sup45 was placed under the control of its native *SUP45* promoter. (Stansfield et al., 1992; Stansfield et al., 1995). Previous research indicated that several erf1 mutants (Table 2-5) display attenuated recognition of the UGA codon, and, hence, may be employed to enhance suppression of internal opal codons within protein reading frames. (Bertram et al., 2000; Liang et al., 2005) A plasmid shuffle strategy was used to substitute pUKC802 with the plasmids carrying mutant versions of the *sup45* gene (Table 2-5) and the cells harboring pDB800, a plasmid carrying wild-type *sup45* gene. The erf1 mutants were screened for growth after plasmid exchange within the *S. cerevisiae* knock-out strain GT17K12, which suggested that the eRF1 mutants were viable. We examined whether the presence of the attenuated eRF1 alleles resulted in detectable suppression of the UGA codon without the presence of orthogonal pair. As shown in Figure 2-13, transformants that carry eRF1 mutants, but not the orthogonal pair, were spread onto selective SD-Leu-Ura-Ade solid media, did not support cell growth. Conversely, cells growth was observed for transformants harboring both the eRF1 mutants and the orthogonal pair. Such observation indicated that in the presence of mutant release factor, the yeast strain GT17K12 is still

auxotrophic for adenine and the eRF1 mutants we used display inability to suppress the opal codon in *ade1-14* in the absence of a functional suppression system.

To quantitatively access the effects of the eRF1 mutants on the suppression efficiency related to the in-frame opal codon in the ECFP context, we employed western blot analysis to check the level of $ECFP_{UGA}$ expression, which is proportional to the suppression efficiency for the opal codon. The expression level of ECFP was observed to depend strongly on the time of induction. After four hours, Western blot analysis of whole-cell lysates after SDS-PAGE indicated significantly higher levels of full-length protein expression for cell lines that that contained the eRF1 mutants in comparison to that with a wild-type erf1 background (Figure 2-14A). Extending the expression time to 8 hours, consistent with the *ade1-14* suppression assay, no detectable full length ECFP was observed with only the presence of eRF1 defective mutants, but not the functional orthogonal pair. However, in the same induction period, none of eRF mutant vaiants displayed better assistance to the suppression effect by the orthogonal pair than the wild-type eRF1 (Figure 2-14B).

Notably, in a shorter induction time, the expression level was observed to vary among the erf1 mutants with the V68I mutant displaying noticeably highest ECFP expression level compared to wild-type erf1 or the other mutant eRF1 variants. This experiment assessed only the suppression of a single opal codon within a specific context in ECFP, however an enhanced suppression effect was observed for the mutants under shorter induction times. Over longer induction periods, no significant advantage was observed for the mutant erf1 variants over the wild-type system, and, indeed, the expression level of most of the mutants was lower than that of the wild-type. This

situation may reflect that the mutants are under significantly more stress than the wild-type strain due to the partially impaired activity of erf1. While this approach may not realize a benefit to single-site suppression studies, the mutant eRF1 strategy may find prove valuable for multi-site suppression in which opal codons are dispersed through a protein coding sequence.

Conclusion

An orthogonal *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS pair was successfully generated for site-specifically incorporation of tryptophan analogues to the in-frame UGA stop codon in *S. cerevisiae*. Several orthogonal pair gene constructs have been examined to optimize the suppression efficiency and the gene assignment of p426.pair-F was selected. (Need better conclusions).

Table 2-1. The oligonucleotides for PCR amplification and DNA construction.

Primer Name	Sequence
pScM-F	5’-GCGCAAGCTTATCGATGTGTGTTATATGTACCTCTGC-3’
pScM-R	5’-CCATTGGATCCGGTACCCGTCTCATACTATTTGTTGAA GGTTTTACTACTTTATTATAG-3’
Sc-trailer-F	5’-GTACCCGTCTCAATTTTTTTGGCTACTCCTGTAGTTATT CTTCATTAATGCTTTGTTAACG-3’
Sc-trailer-R	5’-GATCCGTTAACAAAGCATTAATGAAGAATAACTACAG GAGTAGCCAAAAAAATTGAGACGG-3’
MgentRNA-F	5’-AGTAAGGGGTATAGTTCAAAGGTAGAACATCTGTCTT CAAAATAGAGTGTTGTGGGTTCGAGTCCTGCTACCCCTG-3’
MgentRNA-R	5’-AAATCAGGGGTAGCAGGACTCGAACCCACAACACTCT ATTTTGAAGACAGATGTTCTACCTTTGAACTATACCCCT-3’
MgenTrpS-F	5’-GCTTAGGATCCAATAATGATAAAGCGCGCAATTACAG GGATTCAAGC-3’
MgenTrpS-R	5’-GGACCTCGAGTTAATCAAAAAGCTGATTAGATGTTAA ACCAAATGC-3’
CFPW66stop-F	5’-CTCGTGACCACCCTGACCTGAGGCGTGCAGTGCTTCAG C-3’
CFPW66stop-R	5’- GCTGAAGCACTGCACGCCTCAGGTCAGGGTGGTCACG AG-3’
CFP-F	5’- AGGTTAGGATCCAATAATGGTGAGCAAG GGCGAG-3’
CFP-R	5’-GCTTCCTCGAGATTACTTGTACAGCTCGTCCATGCCG-3’
ADH-F	5’-CGAATCACCGGTCCTAGGGGCGACGTGGCGAGAAAG G-3’
ADH-R	5’-AGCTACCGGTGCTAGCGGCTTTCCCCGTCAAGCTCTA AATCG-3’
UPTAG-Primer	5’-GGAAATACTTCAATAATGGATGTCCACGAGGTCTCTA AATACGCGGTCGCAGTGATCGTACGCTGCAGGTCGAC-3’
DNTAG-Primer	5’-CCCCCTTTTATTTATTTACGGTGTCGGTCTCGTAGACT CGACGCTCATTAGGTCTATCGATGAATTCGAGCTCG-3’

Table 2-2. The oligonucleotides used for PCR amplification and DNA construction.

Primer Name	Sequence
V68I-F	5’-CTAGGGTTAATCGTCTTTCCATTTTATCTGCTATCACT TCCACCCAAC -3’
V68I-R	5’-GTTGGGTGGAAGTGATAGCAGATAAAATGGAAAGAC GATTAACCCTAG -3’
L123V-F	5’-CCTATCAACACATCCTTATATGTTTGTGATAACAAAT TTCATACAGAAG -3’
L123V-R	5’- CTTCTGTATGAAATTTGTTATCACAAACATATAAGGA TGTGTTGATAGG -3’
L123I-F	5’- CCTATCAACACATCCTTATATATTTGTGATAACAAAT TTCATACAGAAG -3’
L123I-R	5’- CTTCTGTATGAAATTTGTTATCACAAATATATAAGGA TGTGTTGATAGG -3’
H129R-F	5’- CTTATATTTGTGTGATAACAAATTTAGAACAGAAGTT CTTTCGGAATTGCTTC -3’
H129R-R	5’- GAAGCAATTCCGAAAGAACTTCTGTTCTAAATTTGTT ATCACACAAATATAAG -3’
S67A-F	5’-CTAGGGTTAATCGTCTTGCTGTTTTATCTGCTATCACT TCCACCCAACAAAAG-3’
S67A-R	5’-CTTTTGTTGGGTGGAAGTGATAGCAGATAAAACAGCA AGACGATTAACCCTAG-3’
SUP45-UP45-F	5’-CTTATTAAGACTACAGAAATAGACAAAGGAAATACTTC AATAATG-3’
SUP45-DN45-R	5’-AAATTCTTTTTGATTCGATTTTTTTCTCCCCCTTTTATTT ATTTA-3’
KanB primer	5’-CTGCAGCGAGGAGCCGTAAT-3’
KanC primer	5’-TGATTTTGATGACGAGCGTAAT-3’

Table 2-3. *E. coli* and *S. cerevisiae* plasmids.

Plasmid	Character	Reference
pZErO-1	E.coli cloning vector, Zeocin^R	Invitrogen
pRS314	Yeast shuttle vector, *CEN, TRP1,* Amp^R	(Sikorski and Hieter, 1989)
pRS315	Yeast shuttle vector, *CEN, LEU2,* Amp^R	(Sikorski and Hieter, 1989)
p426 ADH	Yeast shuttle vector, *2μ, URA3,* Amp^R	(Mumberg et al., 1994)
p423 GAL1	Yeast shuttle vector, *2μ, HIS3,* Amp^R	(Mumberg et al., 1994)
pZErO.5linker	pZErO-1, 5'leader sequence pZErO-1←5'polylinker PCR (*Hin*D III / *Bam*H I)	this study
pZErO.adapter	pZErO-1, 5' and 3'flanking sequence pZErO.5linker ←3'polylinker PCR (*Acc65* I /*Bam*H I)	this study
pZErO.tRNA	pZErO-1, MgentRNA cassette pZErO.adapter ←MgentRNA PCR (*BsmB* I)	this study
p315.tRNA	pRS315, MgentRNA cassette pRS315← pZErO.tRNA (*Hin*D III / *Bam*H I)	this study
p314.tRNA	pRS314, MgentRNA cassette pRS314← p315.tRNA (*Sal* I / *Bam*H I)	this study
p426.TrpRS	p426 ADH, MgenTrpRS p426 ADH← MgenTrpRS PCR (*Xho* I/ *Bam*H I)	this study
p423. ECFP	p423 GAL1, wtECFP p423 GAL1← ECFP PCR (*Xho* I/ *Bam*H I)	this study
p423. W66STOP	p423 GAL1, mutECFP (W66STOP) p423 GAL1← W66STOP PCR (*Xho* I/ *Bam*H I)	this study
p426-AgeI	Inverse PCR amplification of p426 ADH with ADH-F/ ADH-R primers	this study
p426.tRNA-F	p426-Age I, MgentRNA cassette p426-Age I ← tRNA PCR (*Nhe* I / *Avr* II)	this study
p426.tRNA-R	p426-Age I, MgentRNA cassette p426-Age I ← tRNA PCR (*Nhe* I / *Avr* II)	this study
p426.pair-F	p426.tRNA-F, MgenTrpRS p426tRNA-F← p426.TrpRS (*Xho* I/ *Bam*H I)	this study
p426.pair-R	p426.tRNA-F, MgenTrpRS p426tRNA-F← p426.TrpRS (*Xho* I/ *Bam*H I)	this study
pDB800	pRS315, SUP45	(Salas-Marco et al., 2006; Stansfield et al., 1992)
pDB800.Mutant	pDB800, mutant SUP45 listed in Table 2-5	this study

Table 2-4. Expression schemes of orthogonal pair component carried by different plasmids for optimizing suppression efficiency.

Construction	***Mgen*tRNA$^{Trp/UCA}$**	***Mgen*TrpRS**	**Target Protein**	**SD Medium**
A1	pRS315	p426ADH	p423.ECFP	SD-Ura-Leu-His
A2	pRS315	p426ADH	p423. W66stop	SD-Ura-Leu-His
A3	p315.tRNA	p426ADH	p423. W66stop	SD-Ura-Leu-His
A4	pRS315	p426.TrpRS	p423. W66stop	SD-Ura-Leu-His
A5	pRS315.tRNA	p426.TrpRS	p423. W66stop	SD-Ura-Leu-His
B1	pRS314	p426ADH	p423.ECFP	SD-Ura-Trp-His
B2	pRS314	p426ADH	p423. W66stop	SD-Ura-Trp-His
B3	p314.tRNA	p426ADH	p423. W66stop	SD-Ura-Trp-His
B4	pRS314	p426.TrpRS	p423. W66stop	SD-Ura-Trp-His
B5	pRS314.tRNA	p426.TrpRS	p423. W66stop	SD-Ura-Trp-His
C1	p426ADH	p426ADH	p423. W66stop	SD-Ura -His
C2	p426ADH	p426ADH	p423. ECFP	SD-Ura-His
C3	p426.tRNA-R	N/A	p423. W66stop	SD-Ura-His
C4	p426.pair-R	p426.pair-R	p423. W66stop	SD-Ura-His
D3	p426.tRNA-F	N/A	p423. W66stop	SD-Ura-His
D4	p426.pair-F	p426.pair-F	p423. W66stop	SD-Ura-His

Table 2-5. Mutations introduced into the eRF1 coding sequence.

Mutant *sup45* Allele	**Residue Substitution in plasmid**	**Nucleotide Change**
1	pDB800.V68I	GTT→ATT
2	pDB800.S67A/L123V	TCC→GCT/TTG→GTT
3	pDB800.L123I	TTG→ATT
4	pDB800.L123V	TTG→GTT
5	pDB800.H129R	CAT→AGA
6	pDB800	Wild-type

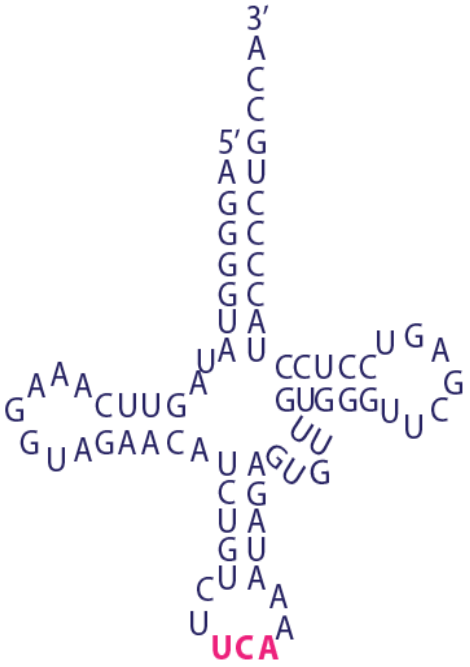


Figure 2-2. Two-dimensional (cloverleaf) structure of the *M. genitalium* tryptophan tRNA (*Mgent*RNA$^{Trp/UCA}$)

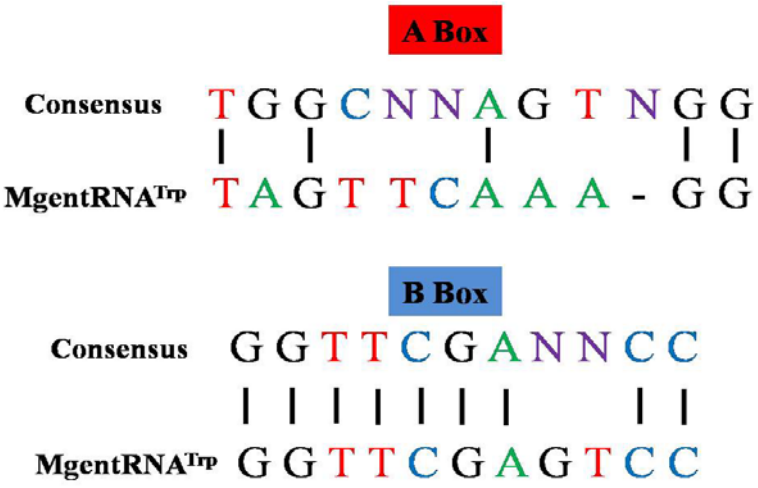
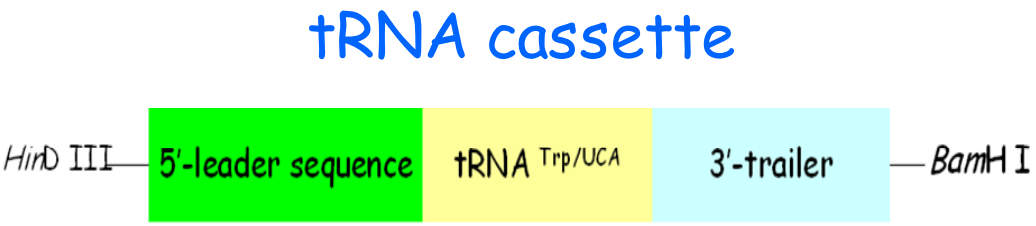


Figure 2-3. Internal promoter sequence of tRNAs. The *Mgent*RNA$^{Trp/UCA}$ and consensus eukaryotic internal tRNA promoter sequences (A and B boxes) are shown.

A



B

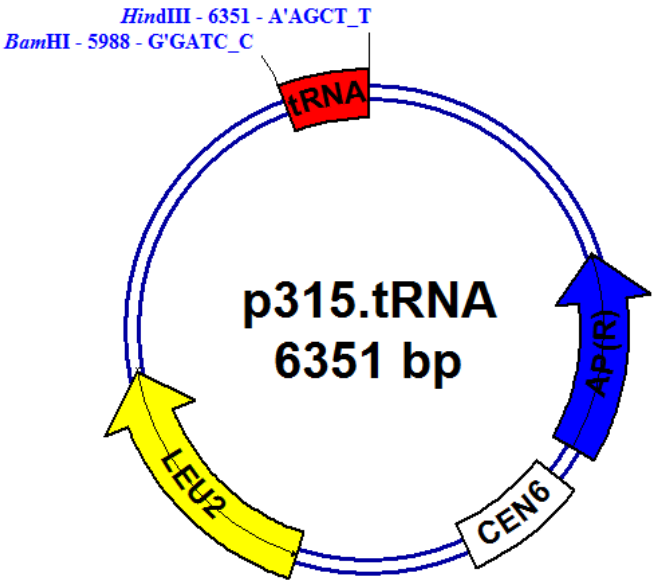
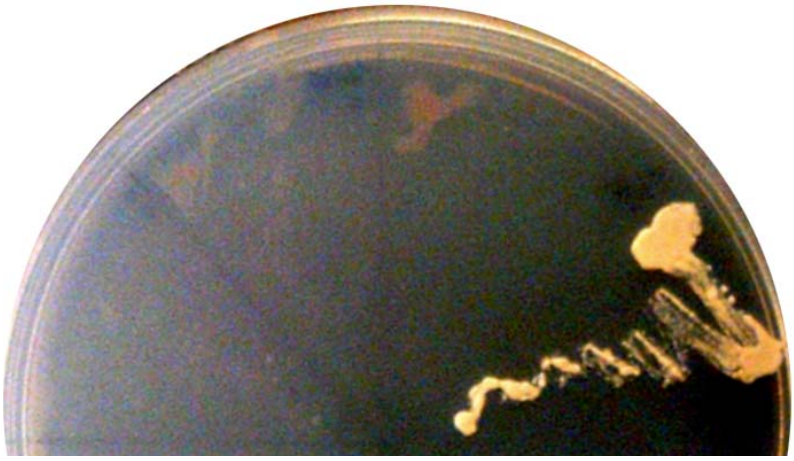


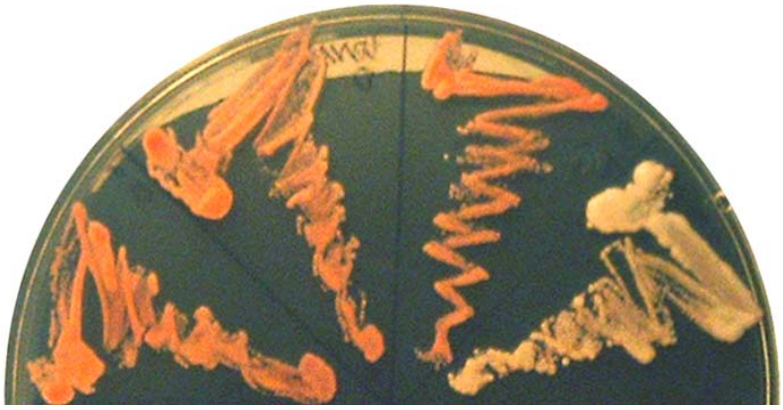
Figure 2-4. Creating a functional *Mgen*tRNA$^{Trp/UCA}$ in yeast.

(A) The *Mgen*tRNA$^{Trp/UCA}$ expression construct is created using the 5' and 3' flanking sequence from the yeast polycistronic tRNA^{Arg}-tRNA^{Asp} gene. (B) The *Mgen*tRNA$^{Trp/UCA}$ cassette is cloned to a centromeric plasmid pRS315 with *Hin*D III and *Bam*H I restriction endonuclease sites.

A



B



	1	2	3	4
***Mgen*tRNA$^{Trp/UCA}$**	-	+	-	+
***Mgen*TrpRS**	-	-	+	+

Figure 2-5. The *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS dependent phenotype of the *ade1-14* reporter gene on (A) SD-Leu-Ura-Ade solid medium and (B) YPD solid medium.

	Marker	1	2	3	4	5
***Mgen*tRNA$^{Trp/UCA}$**		-	-	+	-	+
***Mgen*TrpRS**		-	-	-	+	+
ECFP$_{W66stop}$		**WT ECFP**	+	+	+	+

Figure 2-6. Western analysis results of "mix and match" experiment I

The *Mgen*tRNA$^{Trp/UCA}$ cassette is located at plasmid pRS315 and the *Mgen*TrpRS gene is carried by p426ADH. In the absence of either suppressor *Mgen*tRNA$^{Trp/UCA}$ (lane 3) or *Mgen*TrpRS (lane 4), no full-length ECFP was detected by Western blot analysis. In contrast, when both *Mgen*tRNA$^{Trp/UCA}$ and *Mgen*TrpRS were presented in the cells, the W66stop opal mutation was suppressed and full length ECFP was expressed. (lane 5).

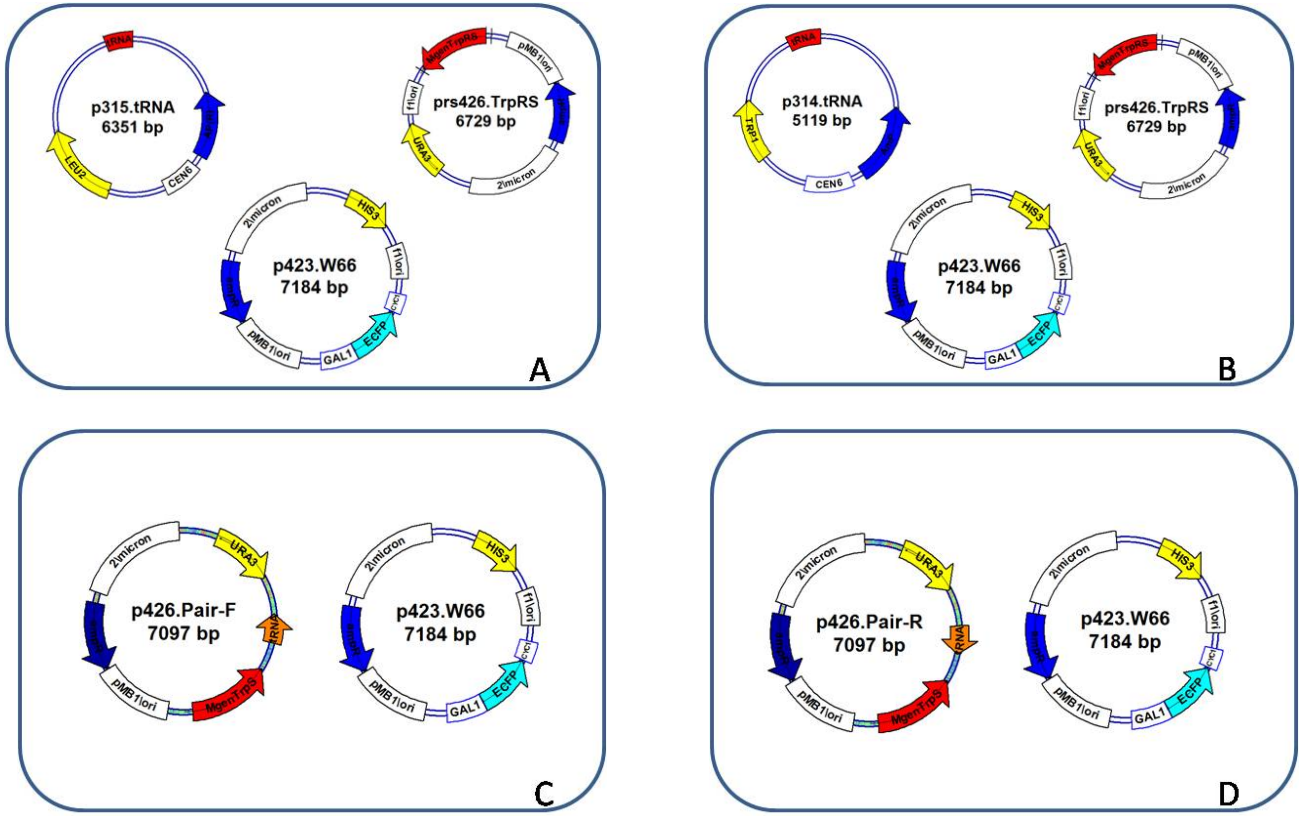


Figure 2-7. Dependence of suppression efficiency for four different arrangements of tRNA gene expression.

In (A) and (B), the *Mgen*tRNA$^{Trp/UCA}$ cassettes were harbored by centromeric plasmid, in which the only difference is that the (B) construct can provide tryptophan to host cell endogenously. In (C) and (D), both the orthogonal *Mgen*tRNA$^{Trp/UCA}$ cassettes and the *Mgen*TrpRS were hosted on the same 2μ plasmid, in which the two recombinant genes are oriented tandem in (C) and convergently in (D).

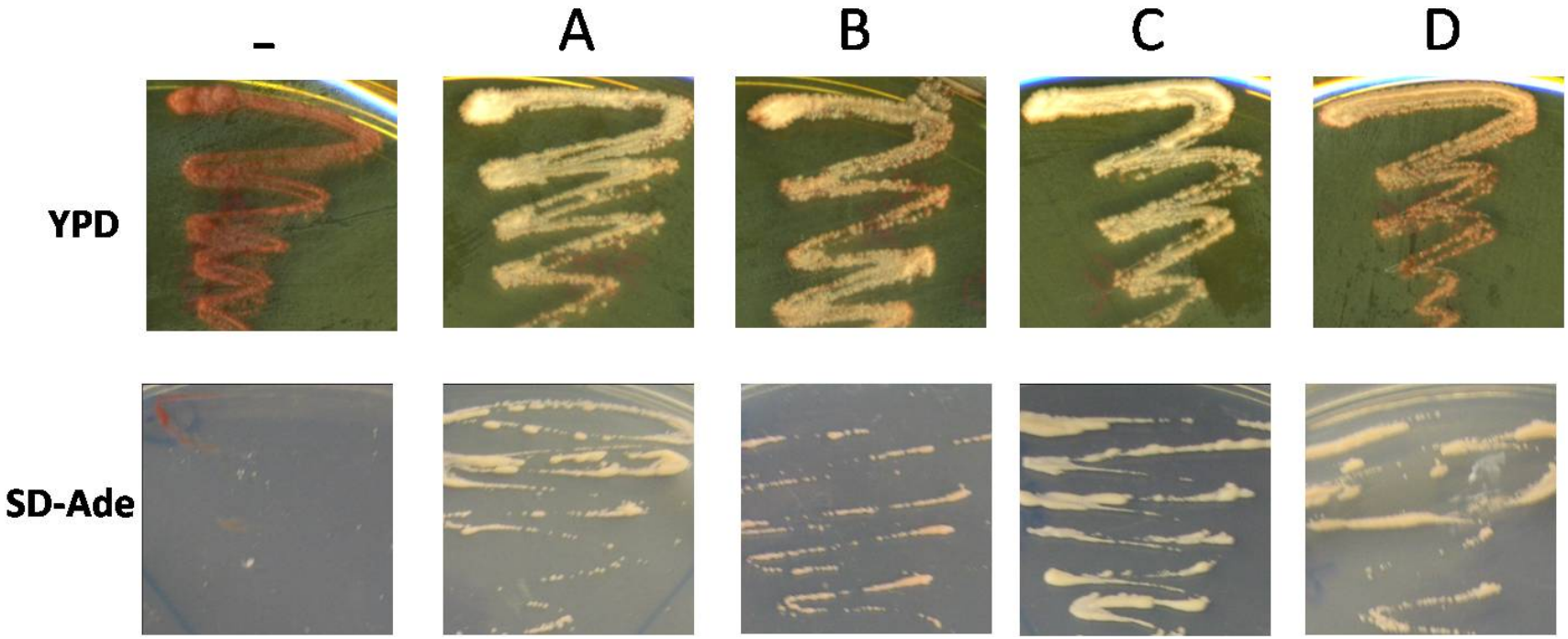


Figure 2- 8. Phenotypic analysis of the *ade1-14* reporter gene on YPD and adenine drop-out media.

The negative control used cells without orthogonal pair component. A-D represented the phenotypes of cells harboring corresponding expression system listed in Figure 2-7. Although the expression constructions of orthogonal components vary with respect to efficiency, each arrangement displayed a phenotype that was consistent with suppression.

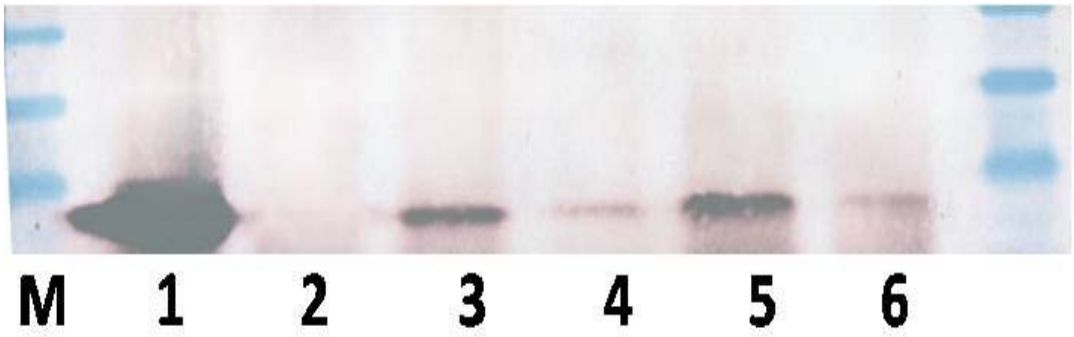
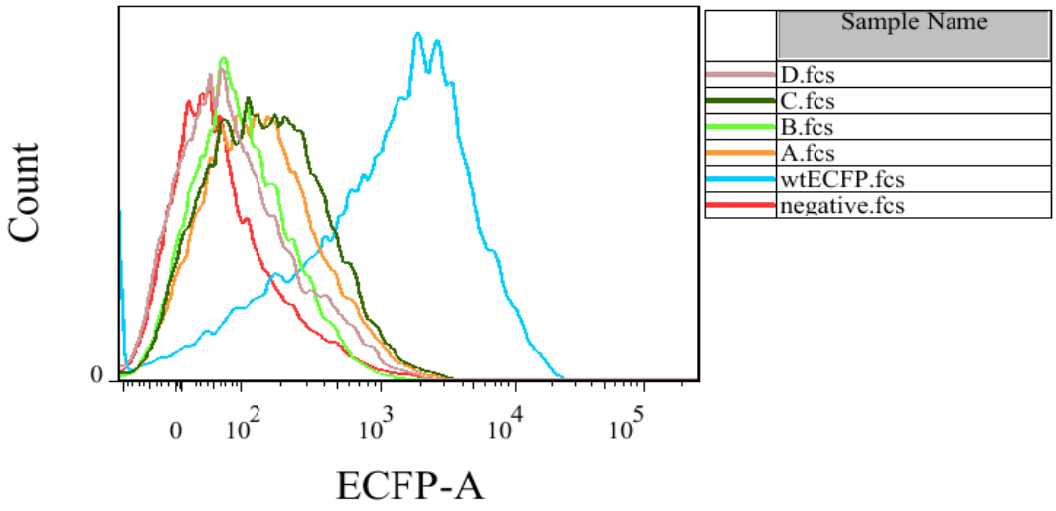


Figure 2- 9. Western blot analysis of ECFP expression level in each suppression system. Lane 1, wt-ECFP (cells carrying p426 ADH/p423.wtECFP); Lane 2, negative control (cells carrying p426ADH/p423.W66); Lane 3, expression system A; Lane 4, expression system B; Lane 5, expression system C; Lane 6, expression system D.



Sample Name	Statistic Mean F_{ECFP}	Number of Cells
Negative control	102	10000
A	208	10000
B	126	10000
C	233	10000
D	142	10000
wtECFP	2137	10000

Figure 2- 10. FACS analysis of ECFP fluorescent strength.

Samples A, B, C, and D correspond to the expression systems listed in Figure 2-7; the negative control uses cells harboring p426ADH/p423.W66 and the wtECFP sample employed cells with p426ADH/p423.wtECFP as positive control. The inserted table indicated the statistic mean fluorescent strength relative to each sample with 10000 counting cells.

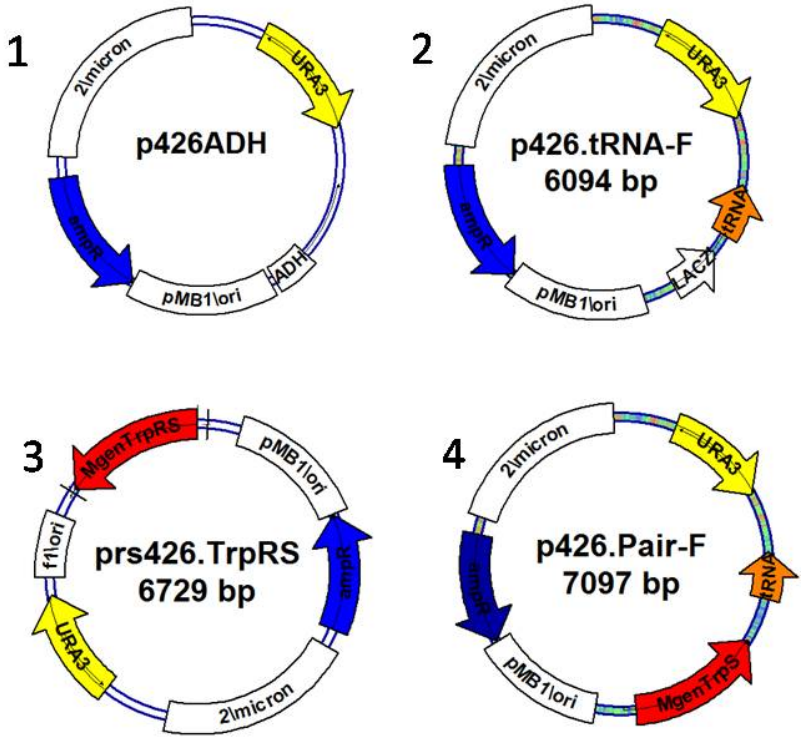
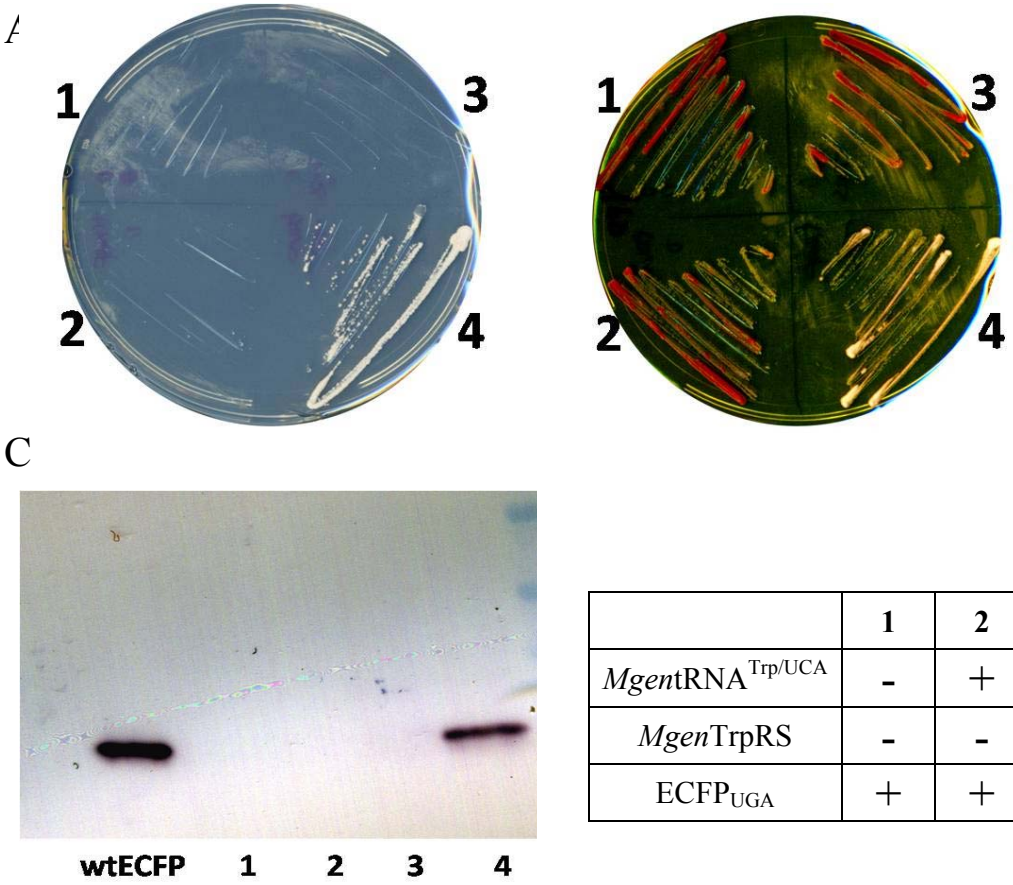


Figure 2- 11. A set of plasmids in "Mix and Match" experiment II.

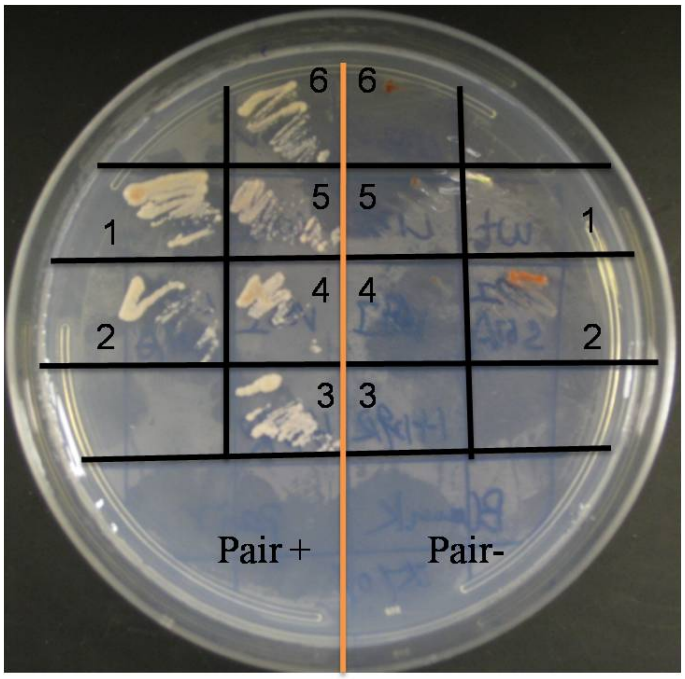
In order to examine the orthogonality and rule out the cross reaction between the optimized expression system C (a system with higher transcription level of suppression tRNA) and the host yeast cells, plasmids were constructed for testing the orthogonality in another "Mix and Match" experiment. A series of plasmids were constructed to carry 1, none component of the plasmids, 2, the *Mgen*tRNA$^{Trp/UCA}$ cassettes, 3, the *Mgen*TrpRS, 4 the *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS pair.



	1	2	3	4
*Mgen*tRNA$^{Trp/UCA}$	-	+	-	+
*Mgen*TrpRS	-	-	+	+
ECFP$_{UGA}$	+	+	+	+

Figure 2- 12. The "Mix and Match" experiment II

The results confirm that in the expression system C, the *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS pair is orthogonal to the yeast while the expression level of *Mgen*tRNA$^{Trp/UCA}$ was boosted by 2μ plasmid.(A) SD-Ura-Ade solid media, (B) YPD media, (C) Western blot analysis of the suppression of the W66stop mutation within ECFP and the orthogonality of the *Mgen* pair to yeast. (The inserted table indicated the orthogonal pair component harbored in the 2μ *URE3* marked plasmid.)

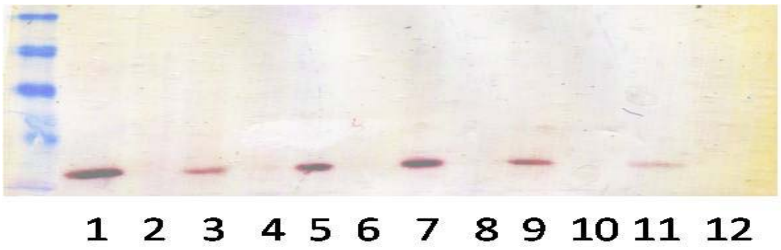


	SUP45
1	Wild-type
2	S67I/L123I
3	H129R
4	V68I
5	L123I
6	L123V

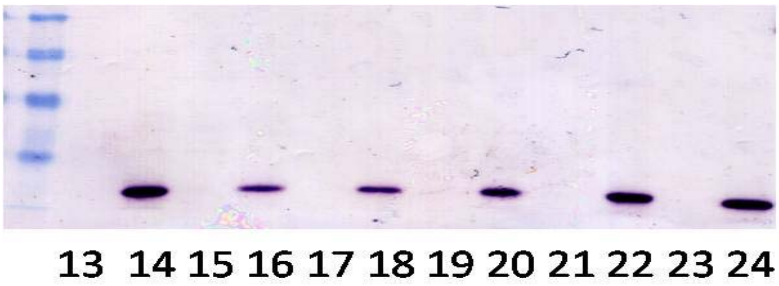
Figure 2-13. Phenotypic analysis of Yeast strain GT17K12 carrying wild-type or mutant *SUP45*.

Yeast strain GT17K12 carrying the wild-type or mutant *SUP45* and the plasmid p426.pair-F (Pair+) or p426ADH (Pair-) were streaked on SD-Leu-Ura-Ade solid media. To the *ade1-14* allele, the lack of suppression was observed with mutant *SUP45* only, but the growth of cells indicated that the opal suppression was supported by the presence of orthogonal *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS pair.

A



B



4h	eRF1 Mutant	Orthogonal Pair	8h	eRF1 Mutant	Orthogonal Pair
1	V68I	+	13	V68I	-
2	V68I	-	14	V68I	+
3	S67A/L123I	+	15	S67A/L123I	-
4	S67A/L123I	-	16	S67A/L123I	+
5	L123I	+	17	L123I	-
6	L123I	-	18	L123I	+
7	L123V	+	19	L123V	-
8	L123V	-	20	L123V	+
9	H129R	+	21	H129R	-
10	H129R	-	22	H129R	+
11	wt eRF1	+	23	wt eRF1	-
12	wt eRF1	-	24	wt eRF1	+

Figure 2- 14. Western blot analysis of eRF1 mutant effects on suppresion

Western blot analysis showed that eRF1 mutants facilitated the orthogonal *Mgen*tRNA$^{Trp/UCA}$/*Mgen*TrpRS pair to suppress the opal mutation when the ECFP$_{UGA}$ had been induced for 4 hours (A), but when the expression time had been elongated to 8 hours (B), the level of suppression of the wild-type eRF1 is similar to that of the eRF1 mutants.

Reference

Anderson, J.C., and Schultz, P.G. (2003). Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. Biochemistry *42*, 9598-9608.

Bae, J.H., Rubini, M., Jung, G., Wiegand, G., Seifert, M.H., Azim, M.K., Kim, J.S., Zumbusch, A., Holak, T.A., Moroder, L.*, et al.* (2003). Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. J Mol Biol *328*, 1071-1081.

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C. (1993). A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. Nucleic Acids Res *21*, 3329-3330.

Beechem, J.M., and Brand, L. (1985). Time-resolved fluorescence of proteins. Annu Rev Biochem *54*, 43-71.

Bertram, G., Bell, H.A., Ritchie, D.W., Fullerton, G., and Stansfield, I. (2000). Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. RNA *6*, 1236-1247.

Bonetti, B., Fu, L., Moon, J., and Bedwell, D.M. (1995). The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in Saccharomyces cerevisiae. J Mol Biol *251*, 334-345.

Budisa, N., and Pal, P.P. (2004). Designing novel spectral classes of proteins with a tryptophan-expanded genetic code. Biol Chem *385*, 893-904.

Callen, B.P., Shearwin, K.E., and Egan, J.B. (2004). Transcriptional interference between convergent promoters caused by elongation over the promoter. Mol Cell *14*, 647-656.

Carne, A.F. (1994). Chemical modification of proteins. Methods Mol Biol *32*, 311-320.

Chernoff, Y.O., Lindquist, S.L., Ono, B., Inge-Vechtomov, S.G., and Liebman, S.W. (1995). Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. Science *268*, 880-884.

Derkatch, I.L., Bradley, M.E., Zhou, P., Chernoff, Y.O., and Liebman, S.W. (1997). Genetic and Environmental Factors Affecting the de novo Appearance of the [PSI(+)] Prion in Saccharomyces cerevisiae. Genetics *147*, 507-519.

Francis, M.A., and Rajbhandary, U.L. (1990). Expression and function of a human initiator tRNA gene in the yeast Saccharomyces cerevisiae. Mol Cell Biol *10*, 4486-4494.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M.*, et al.* (1995). The minimal gene complement of Mycoplasma genitalium. Science *270*, 397-403.

Geiduschek, E.P., and Tocchini-Valentini, G.P. (1988). Transcription by RNA polymerase III. Annu Rev Biochem *57*, 873-914.

Hasselbacher, C.A., Rusinova, E., Waxman, E., Rusinova, R., Kohanski, R.A., Lam, W., Guha, A., Du, J., Lin, T.C., Polikarpov, I.*, et al.* (1995). Environments of the four tryptophans in the extracellular domain of human tissue factor: comparison of results from absorption and fluorescence difference spectra of tryptophan replacement mutants with the crystal structure of the wild-type protein. Biophys J *69*, 20-29.

Hohsaka, T., Ashizuka, Y., and Sisido, M. (1999). Incorporation of two nonnatural amino acids into proteins through extension of the genetic code. Nucleic Acids Symp Ser, 79-80.

Hohsaka, T., and Sisido, M. (2002). Incorporation of non-natural amino acids into proteins. Curr Opin Chem Biol *6*, 809-815.

Kim, D., Raymond, G.J., Clark, S.D., Vranka, J.A., and Johnson, J.D. (1990). Yeast tRNATrp genes with anticodons corresponding to UAA and UGA nonsense codons. Nucleic Acids Res *18*, 4215-4221.

Kohrer, C., Sullivan, E.L., and RajBhandary, U.L. (2004). Complete set of orthogonal 21st aminoacyl-tRNA synthetase-amber, ochre and opal suppressor tRNA pairs: concomitant suppression of three different termination codons in an mRNA in mammalian cells. Nucleic Acids Res *32*, 6200-6211.

Kowal, A.K., Kohrer, C., and RajBhandary, U.L. (2001). Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria. Proc Natl Acad Sci U S A *98*, 2268-2273.

Kwon, I., and Tirrell, D.A. (2007). Site-specific incorporation of tryptophan analogues into recombinant proteins in bacterial cells. J Am Chem Soc *129*, 10431-10437.

Liang, H., Wong, J.Y., Bao, Q., Cavalcanti, A.R., and Landweber, L.F. (2005). Decoding the decoding region: analysis of eukaryotic release factor (eRF1) stop codon-binding residues. J Mol Evol *60*, 337-344.

Liu, D.R., and Schultz, P.G. (1999). Progress toward the evolution of an organism with an expanded genetic code. Proc Natl Acad Sci U S A *96*, 4780-4785.

Mumberg, D., Muller, R., and Funk, M. (1994). Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res *22*, 5767-5768.

Murphy, M.H., and Baralle, F.E. (1983). Directed semisynthetic point mutational analysis of an RNA polymerase III promoter. Nucleic Acids Res *11*, 7695-7700.

Nakamura, Y., Uno, M., Toyoda, T., Fujiwara, T., and Ito, K. (2001). Protein tRNA mimicry in translation termination. Cold Spring Harb Symp Quant Biol *66*, 469-475.

Palm, G.J., and Wlodawer, A. (1999). Spectral variants of green fluorescent protein. Methods Enzymol *302*, 378-394.

Prescott, E.M., and Proudfoot, N.J. (2002). Transcriptional collision between convergent genes in budding yeast. Proc Natl Acad Sci U S A *99*, 8796-8801.

Ross, J.B., Senear, D.F., Waxman, E., Kombo, B.B., Rusinova, E., Huang, Y.T., Laws, W.R., and Hasselbacher, C.A. (1992). Spectral enhancement of proteins: biological incorporation and fluorescence characterization of 5-hydroxytryptophan in bacteriophage lambda cI repressor. Proc Natl Acad Sci U S A *89*, 12023-12027.

Ross, J.B., Szabo, A.G., and Hogue, C.W. (1997). Enhancement of protein spectra with tryptophan analogs: fluorescence spectroscopy of protein-protein and protein-nucleic acid interactions. Methods Enzymol *278*, 151-190.

Salas-Marco, J., Fan-Minogue, H., Kallmeyer, A.K., Klobutcher, L.A., Farabaugh, P.J., and Bedwell, D.M. (2006). Distinct paths to stop codon reassignment by the variant-code organisms Tetrahymena and Euplotes. Mol Cell Biol *26*, 438-447.

Sikorski, R., and Hieter, P. (1989). A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19-27.

Stansfield, I., Grant, G.M., Akhmaloka, and Tuite, M.F. (1992). Ribosomal association of the yeast SAL4 (SUP45) gene product: implications for its role in translation fidelity and termination. Mol Microbiol *6*, 3469-3478.

Stansfield, I., Jones, K.M., Kushnirov, V.V., Dagkesamanskaya, A.R., Poznyakovski, A.I., Paushkin, S.V., Nierras, C.R., Cox, B.S., Ter-Avanesyan, M.D., and Tuite, M.F. (1995). The products of the SUP45 (eRF1) and SUP35 genes interact to mediate translation termination in Saccharomyces cerevisiae. EMBO J *14*, 4365-4373.

Straby, K.B. (1988). A yeast tRNA(Arg) gene can act as promoter for a 5' flank deficient, non-transcribable tRNA(SUP)6 gene to produce biologically active suppressor tRNA. Nucleic Acids Res *16*, 2841-2857.

Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D.S., and Schultz, P.G. (2004). Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. Proc Natl Acad Sci U S A *101*, 8882-8887.

Chapter III: Development of a Screening System for the Expanding Genetic Code in Yeast to Incorporate Tryptophan Analogues

Introduction

In an approach to expand the genetic code, a general strategy has been developed that utilizes an orthogonal tRNA and aminoacyl-tRNA-stythetase (aaRS) pair, both of which must be expressed and processed into a functional form. Ideally, cross-reaction of the recombinant pair with the corresponding translational elements of the host system should be minimized to prevent mis-incorporation. Using this method, more than seventy unnatural amino acids (UAAs) have been co-translationally inserted at specific sites in recombinant polypeptides with high fidelity. (Cropp and Schultz, 2004; Liu and Schultz, 2010) Most of these UAAs were encoded in *E. coli* in response to the amber UAG codon. (Cellitti et al., 2008; Chin et al., 2002; Lee et al., 2009; Sakamoto et al., 2009; Xie et al., 2004) The extension of this methodology to higher eukaryotic cells is attractive because it provides powerful tools to analyze and engineer proteins in the more complicated biological systems that are more functionally relevant to human biochemistry.

Saccharomyces cerevisiae is an ideal host organism to develop this gateway. Because the translation machinery of eukaryotes is well conserved, orthogonal tRNA/aaRS pairs that have been developed in *S. cerevisiae* may be portable to more complex eukaryotic systems. In addition, yeast has advantages such as a relatively rapid growth rate, a fully sequenced genome, well-defined proteomic analysis, and facile genetic manipulation. However, given that the processing, expression and recognition of tRNAS in yeast are different from those in bacteria such as *E. coli*, the expression systems used to encode UAAs in *E. coli* are not suitable for the eukaryotes, therefore, other orthogonal pairs were evolved in order to add UAAs to eukaryotic systems, e.g., mutant *E. coli* $tRNA^{Tyr}$/TyrRS and *M. barkeri* $tRNA^{Pyl}$/PylRS pairs. (Chin et al., 2003;

Hancock et al., 2010) However, the scope of UAAs that have been incorporated in yeast is still limited and the development of new orthogonal pairs may be necessary to permit incorporation of a greater range of UAAs. Toward this objective, we have constructed a *Mycoplasma genitalium* $tRNA^{Trp/UCA}$/TrpRS pair (*Mgen* pair) to expand genetic code in *S. cerevisiae*. (Chapter 2)

Expansion of the genetic code in an target organism requires the following criteria to be satisfied: transport of the UAA into the cell, introduction of a unique codon to specify the UAA, design of a tRNA that can be charged to the UAA and decode the assigned codon, and development of a mutant aaRS that specifically recognizes and charges the UAA to the cognate tRNA. In this approach, many UAAs are cell-permeable and the unique codon can be selected from among the three termination codons. Thus, the key challenge in this approach is the identification of an orthogonal tRNA/aaRS pair that can selectively recognize the desired UAA. In Chapter 2, we have proved that the *Mgen* pair is orthogonal to *S. cerevisiae*. In this chapter, we develop a strategy to generate and screen libraries of *Mycoplasma genitalium* tryptophan tRNA-synthetase (*Mgen*TrpRS) mutants to identify candidates that can recognize a UAA of interest.

Schultz and colleagues have employed two-step selections to screen libraries of TyrRS mutants to identify variants that facilitate the incorporation of different UAAs. (Chin et al., 2003; Liu and Schultz, 1999) To broaden the substrate specificity of *Mgen*TrpRS, we developed a comparable “double sieve” screening method that can be used to generally discover *Mgen*TrpRS for UAAs. (Figure 3-6) In *S.cerevisiae*, fewer genes have been used for negative selection, but the *URA3* gene that is required for uracil biosynthesis is attractive because it can be utilized as a positive selection for uracil

prototrophy and as a negative selection for sensitivity to the anti-metabolite 5-fluroorotic acid (5-FOA). (Boeke et al., 1984; Vidal et al., 1996a; Vidal et al., 1996b) To use *URA3* as reporter gene in "double sieve" screening, the UGA codon that we assigned to the unique UAAs need to be placed within the open reading frame (ORF) of the *URA3* gene. It is known that fusions to the *N*-terminus of the *URA3* gene do not affect its function, (Ishioka et al., 1997), thus, attaching a nucleotide fragment with the UGA codon to the 5' end of URA3 gene has been employed as a method to screen for non-sense suppression *in vivo*. Previous studies have shown that the sequence context flanking the nonsense codon affected the efficiency of read-through. The *ade1-14* gene contained an easily suppressed nucleotide context at the UGA codon; therefore, Manogaran, *et al.,* fused a the DNA sequence flanking the UGA codon of the *ade1-14* gene (containing the nonsense mutation and 10 adjacent upstream and downstream codons) to the 5' end of *URA3* gene to construct the *ura3-14* allele for the detection of *[PSI$^+$]* phenotype using a non-sense suppression approach. (Manogaran et al., 2006) Thus, we can take advantage of this construct as a reporter in the "double sieve" screening.

The enhanced cyan fluorescent protein (ECFP) is a green fluorescent protein (GFP) variant that contains an indole side-chain in the chromophore due to the presence of Trp66. (Bae et al., 2003) Due to its unique distinct spectroscopic properties compared to tryptophan, (Wong and Eftink, 1997) 5-hydroxyl-tryptophan (5-OH-Trp) is an interesting analogue to substitute the Trp66 at the ECFP. Although it is possible to introduce 5-OH-Trp into protein barstar with the selective pressure incorporation (SPI) method in *E.coli*, the SPI approch cannot incorporate 5-OH-Trp into GFP protein, which might be due to certain unknown editing mechanisms during the translation. In this case, a mutant tryptophanyl-tRNA-synthetase (TrpRS) with altered substrate selectivity may

facilitate incorporation of 5-OH-Trp into to specific sites within target proteins such as ECFP

Chapter 2 provide detail concerning the construction and evaluation of the orthogonal system, while in this chapter, I demonstrate the development of a library screening system for selecting *Mgen*TrpRS variants for the incorporation of 5-hydroxyl tryptophan and methoxyl group at the counterparts of tryptophan (5-MeO-Trp) into protein.

Experimental Procedures

Materials

All chemicals were purchased from Sigma-Aldrich (St Louis, MO) and Fisher Scientific Co. (Fair Lawn, NJ) unless otherwise noted. Restriction endonuclease enzymes, T4 DNA ligase, and T4 kinase were purchased from New England Biolabs (Beverly, MA). The shrimp alkaline phosphatase was obtained from Roche Applied Science (Indianapolis, IN), and Platinum Pfx DNA polymerase was obtained from Invitrogen Corp. (Carlsbad, CA).). Plasmids preparation kits were purchased from Qiagen (Valencia, CA). Yeast transformation kits were purchased from either Invitrogen (Carlsbad, CA) or Zymo Research (Orange, CA). The yeast shuttle expression vectors were purchased from American Type Culture Collection (ATCC#87669) and the plasmid pUC18-MgenTrpRS was purchased from American Type Culture Collection (ATCC#623677) as well. Synthetic oligonucleotides were purchased from either Sigma-Genosys, Inc (The Woodlands, TX) or Integrated DNA Technologies (Coralville, IA) and

were used as received. The prestained protein molecular weight standard was purchased from Bio-Rad Laboratories, Inc. (Hercules, CA) and used for SDS-PAGE and western blot analysis. The primary antibody used for ECFP was the Living Colors A.v. Monoclonal Antibody (JL-8) purchased from Clontech, Inc. (Mountain View, CA) and the primary antibody used for binding His-tagged protein was the penta-histidine monoclonal antibody purchased from EMD Chemical Group (Gibbstown, NJ). The secondary antibody used was the goat anti-mouse secondary antibody, also from EMD Chemical Group. The full-length His-tagged ECFP were visualized by chemiluminescent detection using the 1-step NBT/BCIP reagent mixture from Pierce Biotechnology (Thermo Scientific) (Rockford, IL).

Media and growth conditions

YPD (1%, w/v, yeast extract; 2%, w/v, peptone; 2%, w/v, glucose) and 1/4YPD (0.25%, w/v, yeast extract; 2%, w/v, peptone; 4%, w/v, glucose) media were prepared in ddH_2O and autoclaved for sterilization. To prepare solid media, agar (1.5%) was added to the media before sterilization. Synthetic Defined (SD) minimal medium were prepared using standard protocols and minimal drop-out media lacking appropriate amino acids were prepared as needed, containing glucose or galactose as carbon source. LB media containing appropriate antibiotics were prepared using standard protocols. *S. cerevisiae* strains were grown at 30 °C and *E. coli* strains were grown at 37 °C with sufficient aeration.

Plasmids construction and strains

Construction of the expression plasmids for library report system

A single *Xho* I site was removed from p423 GPD plasmid before cloning the *Mgen*-pair cassette into it. *Xho* I was used to digest p423 GPD, and the sticky ends of digested vector were filled with Klenow fragment. The Klenow reaction was set up as follows. The final concentrations of each component included 1X NEB buffer 2 (1 μL, 10X stock), 1 mM each dNTP (1.0 μL, 10 mM dNTP mixture), 2 μg digested DNA, 5 units DNA Polymerase I, Klenow fragment and water to fill to 10 μL total volume in a 1.5 mL tube. The reaction was at 25 ºC for 30 min and stopped with heat inactivation of Klenow fragment at 75 ºC for 20 min. Following, the blunt-end DNA was re-circularized with T4 DNA ligase to generate the plasmid p423GPD-XhoI.

The *Mgen*tRNA$^{Trp/UCA}$/ *Mgen*TrpRS orthogonal pair cassette was amplified by PCR from the plasmid p426.pair-F (Chapter 2) with primers *Mgen*TrpRS-F and *Mgen*TrpRS-tRNA-R-R and cloned into p423GPD-XhoI with *Bam*H I and *Sal* I restriction sites at the 5'- and 3'- termini respectively of the *Mgen*-pair cassette to generate plasmid, p423.pair.

The reporter gene used for library screening is the *ura3-14* allele, which an 81 nucleotide region containing the nonsense codon (TGA) from the *ade1-14* allele was placed in-frame with the widely used *URA3* gene. (Manogaran et al., 2006) The plasmid pLEU2ura3-14 was kindly provided as a gift from the lab of Dr. Susan W. Liebman. The open reading frame (ORF) of *ura3-14* allele was amplified from the pLEU2ura3-14 plasmid using the primers Ura3-14-F-SpeI and Ura3-14-R and then introduced into p425 GPD with *Spe* I and *Xho* I sites to generate p425GPD.Ura3, expressing under the strong constitutive promoter from glyceraldehyde-3-phosphate dehydrogenase (GPD). Alternatively, using designed primers PGK-F and Ura3-14-R to amplify of both the PGK promoter and the ORF of *ura3-14*, the PCR product was double-digested with *Sac* I and

Xho I endonucleases and cloned into p425GPD to construct plasmid p425PGK.Ura3, in which the expression of *ura3-14* allele was under the control of the constitutive promoter from phosphoglycerate kinase (PGK).

Library construction: Site saturation mutagenesis of MgenTrpRS

The site-saturation mutagenesis is achieved by using primers for amplification of a gene where selectively chosen amino acid residues are mutated to all of the 20 possible amino acids to generate a library of enzymes. Using the X-ray crystal structure data of the homologous protein, *Bacillus stearothermophilus (Bst)* trypophanyl-tRNA-synthetase (*Bst*TrpRS PDB code: 1MB2), the residues (Met137, Ile141, Val149 and Val151) in the substrate binding pocket of *Mgen*TrpRS with close contacts to the 5-position of the indole ring at the L-tryptophan (5 Å) was chosen as targets for altering substrate specificity of the enzyme with randomizing mutagenesis. In this approach, we used an overlapping fragment PCR strategy to generate the *Mgen*TrpRS saturation mutagenesis library with NNK randomization at the 4 residues mentioned before. The first-round of PCR was employed to amplify the *Mgen*TrpRS gene from the plasmid p426.pair-F with the 2 groups of following primers: (N = A + G + T+ C, K = G + T and M= C + A)

I, MgenTrpS-F (Mutation at M137 and I141)

5'- GCTTAGGATCCAAAATGATAAAGCGCGCAATTACAGGGATTCAAGC-3' and

Overlap M137-I141-R (Mutation at M137 and I141)

5'-AATATCAGGTTGATAAAGCAAMNNATCAGCAGCMNNTAACACTGGGTAA G-3')

II, Overlap V149-151-F (Mutation at V149 and V151)

5'- GCTTTATCAACCTGATATTNNKCCANNKG GTAATGATCAGAAG-3'

and MgenTrpRS-R-PstI (Mutation at V149 and V151)

5’-GTGACCTGCAGGGACCTCGAGTTAATCAAAAAGCTGATTAGATGTTAAAC CAAATGC-3’

The Reactions for mutagenesis were prepared as follows. The final concentrations of each component included 1X Pfx50 amplification buffer (5 μL, 10X stock), 0.3 mM each dNTP (1.5 μL, 10 mM dNTP mixture), 0.3 μM each primer (1.5 μL, 10 μM stock each), 500 ng template DNA, 1.0-2.0 units Pfx50 enzyme and water to fill to 50 μL total volume in a thin-walled PCR tube. The solutions were mixed by pipette and pulsed to bring the solutions to the bottom of the tube. Care was taken not to introduce bubbles in the solutions. The tube(s) were placed in the heat block of a thermocycler and subjected to the following amplification regime; Step 1: 94 ºC for 5 min, Step 2: 94 ºC for 30 s, Step 3: 55 ºC for 30 s, Step 4: 68 ºC for 1 min (1 min/kb), Step 5: Repeat steps 2-4 30 times, Step 6: 68 ºC for 7 min, Step 7: 4 ºC forever.

The PCR products in the first-round were run on 1% agarose gel to confirm the size of the product generated and purified with Zymoclean Gel DNA recovery kit. Sequentially, the primers MgenTrpS-F and MgenTrpRS-R-PstI to amplify the mixture of the first-round PCR products were used for the second-round of PCR and then the DNA was purified following amplification using the QIAGEN PCR purification kit. Because directly connecting PCR product with large size yeast expression shuttle vector affected the ligation efficiency, the first ligation was designed between library PCR product and a common used *E.Coli* cloning vector pUC19. After digestion with *Bam*H I and *Pst* I restriction endonucleases, the final library PCR products were ligated to pUC19 vector to generate plasmids, pUC19.Lib. The ligation reaction was incubated overnight at 16 ºC. The DNA in ligation reaction was purified by butanol precipitation and Top10F′ high efficiency electrocompetent cells were prepared the same day for transformation. All of

the cells were plated on large library plates (~200 mL solid LB media supplemented with ampicillin (Amp 100 µg / mL)). One µL of each transformation suspension was added to 99 µL LB and the diluted culture was plated on a small LB+Amp Petri plate in order to evaluate the transformation efficiency. Plates were incubated overnight at 37 °C and the number of colonies was counted the next day. Ten clones were selected at random for sequencing to proof the diversity of the library and the combined library size was evaluated approximately as 2.6×10^5.

The large library plate was scraped using a sterile spreader after addition of 10 mL LB media on each one. The 10 mL suspension was added to 90 mL LB supplemented with 100 µg/mL Amp and grown 4 hours at 37 ºC with shaking at 225 rpm. Two 5 mL aliquots were taken from total culture and the DNA, pUC19.Lib, was isolated from the cells using the QIAGEN spin miniprep kit. Following, the concentrated DNA sample was digested with *Bam*H I and *Xho*I restriction endonucleases and the mutagenesis *Mgen*TrpRS library genes were cloned to plasmid p423.pair to construct the plasmid, p423.Lib. Frozen stocks were prepared by adding 800 µL cell cultures to 200 µL 80 % glycerol solution and storage at -80 ºC.

Similarly, the DNA in ligation reaction was purified by butanol precipitation and transformed into same-day-prepared Top10F´ high efficiency electrocompetent cells. Plating all cell culture on large library plates (LB+Amp) and around 3×10^5 colonies were collected on the next day. The plasmid p423.Lib was isolated from 2× 5mL colony suspension culture with the QIAGEN spin miniprep kit and used for high efficiency yeast library transformation.

High efficiency yeast library transformation using DNA as a carrier

Single colony of yeast strain GT17 (p425GPD.Ura3) was inoculated in 2 mL SD-Leu media overnight and sub-cultured into 20mL SD-Leu media for 24 hours. The 10 mL suspension was diluted to 400 mL same culture and the yeast cells were grown to the mid-log phase (a suspension containing 1×10^7 cells will give an OD_{600} of 0.6-0.7). The 400mL aliquots were centrifuged at 1000 x g for 10 min, 4 °C, in 2 Falcon tubes. The supernatant was discarded, and the cell pellets were resuspended in 25 mL sterile cold water, and then the cells were repelleted at 1000 x g for 10 min, 4 °C. The supernatant was discarded, and the cell pellets were resuspended in 20mL sterile cold LiTE buffer (0.1 M Lithium Acetate, 0.01M Tirs-HCl, 1mM EDTA, pH=7.5). Next, the cells were centrifuged at 1000 x g for 10 min, 4 °C again, and then resuspended in 5mL of LiSORB buffer (0.1 M Lithium Acetate, 0.01M Tirs-HCl, 1mM EDTA, 1M Sorbitol, pH=7.5) for every 200mL of original culture. The cell suspension was incubated at 30 °C for 1 hour with shaking. (At this step, 1 hour is optimal and 1.5 hours is dangerous to cells) The cell culture were harvested and resuspended in 250 µL of Carrier Resuspension Buffer (4 mg/mL single stand carrier DNA, 0.1 M Lithium Acetate, 0.01 M Tirs-HCl, 1mM EDTA, 1 M Sorbitol) for cells from 200mL of original culture. Sequentially, adding 2 µg of DNA p423.Lib into every 100 µL of cell suspension, the whole transformation culture was incubated at 30 °C for 30 minutes and then mixed with 900 µL of LiPEG (40% PEG_{3350}, 0.1 M Lithium Acetate, 0.01 M Tris-HCl, 1mM EDTA, pH 7.5) for each 100 µL of cells. It was important to divide the cell mixture in Eppendorf tubes (1 mL/tube) at this step, because it took too long time to heat a large volume of PEG uniformly. After that, incubating the mixture at 30 °C for 20 minutes and then the cell mixture were briefly heated at 42 °C for 10 min, followed by immediately plating the cells on large library plate (SD-Leu-His). One µL of transformation suspension was added to 99 µL sterile

water and the 20 μL of diluted culture was plated on a small SD-Leu-His Petri plate in order to evaluate the transformation efficiency. It was critical not to pellet the cells through the PEG since such action lessened the transformation frequency. In addition, waiting too long to plate cells also affect transformation due to the toxicity of PEG to the cells. Followed scraping all cells from library plates by adding 10 mL of sterile 0.9 % NaCl solution to each plate, sterile glycerol were added to a final concentration of 20 % and the cells were divided into several tubes for frozen stock at -80 ºC

Selection and screen of mutant MgenTrpRS

To select mutant *Mgen*TrpRS suitable for the incorporation of UAA, a double-sieve selection strategy was employed. This method consists of a combination of positive and negative selections with a mutant *Mgen* TrpRS library in the presence and absence of UAA. The reporter gene used for both positive and negative selection is the *ura3-14* allele, which an opal codon was placed in frame at the *N*-terminal of *URA3* gene.

A 100 μl aliquot of the thawed culture from the yeast library cells, which expressed the *ura3-14* allele and mutant variants of the *Mgen*TrpRS gene, was grown in 5 mL SD-Leu-His media overnight at 30 ºC with shaking at 225 rpm. The cells were harvested at 1000 x g for 10 min at room temperature, and washed twice with 10 mL sterile 0.9 % NaCl solution. The cell pellet was resuspended in 10 mL sterile 0.9 % NaCl solution and 2-3 mL of cell suspension was plated on SD-Leu-His-Ura+UAA solid medium for positive selection. The selection plates were incubated at 30 ºC for 2-3 days until small colonies appear.

All colonies from first positive selection plates were scraped with 10 mL sterile 0.9% NaCl solution each plate and harvested at 1000 x g for 10 min at room temperature. The cell pellet was resuspended in 10 mL of SD-Leu-His medium, and 5 mL of cell

suspension was kept in 20% glycerol for storage at – 80 ºC. An aliquot (1mL) of culture was added to 50 mL of SD-Leu-His medium for overnight incubation at 30 ºC with shaking at 225 rpm. The overnight culture was spun down at 1000 x g for 10 min at room temperature and washed twice with equal volume of sterile 0.9% NaCl solution. The cell pellet was re-suspended in 10 mL of sterile 0.9% NaCl solution and 2-3 mL of cell culture was plated on SD-Leu-His+0.1% 5-FOA solid medium for the negative selection. After incubation at 30 ºC for 48-72 h, small colonies had grown on the negative selection plate. Similarly, cells were scraped from negative plate with 10 mL sterile 0.9% NaCl solution for each plate and harvested at 1000 x g for 10 min at room temperature. The cell pellet was resuspended in 10 mL of SD-Leu-His medium, in which 5 mL of cell suspension was kept in 20% glycerol for storage at –80 ºC and 1 mL aliquot culture was added to 50 mL of SD-Leu-His medium for overnight incubation at 30 ºC with shaking at 225 rpm. An additional two rounds of positive and negative selection were carried on the cells and each time the products of the previous round were used for future selections.

Identification of the mutant MgenTrpRS

To identify the colonies that survive from the selection process, the plasmid DNA was extracted from yeast cells. Single colonies were picked from the selection plate and incubated in 2 mL SD-Leu-His medium overnight at 30 ºC with shaking at 225 rpm. The yeast plasmid DNA was extracted using the zymoprep yeast plasmid miniprep kit and the mutant *Mgen* TrpRS variant genes were amplified by PCR with primer MgenTrpRS-F and MgenTrpS-R. The final concentrations of each component in the PCR reaction included 1X *Pfx*50 amplification buffer (5 μL, 10X stock), 0.3 mM each dNTP (1.5 μL, 10 mM dNTP mixture), 0.3 μM each primer (1.5 μL, 10 μM stock each), 500 ng template

DNA, 1.0-2.0 units Pfx50 enzyme and water to fill to 50 μL total volume in a thin-walled PCR tube. The PCR product was analyzed with agarose gel electrophoresis to confirm the size and then and purified with Zymoclean Gel DNA recovery kit. For full DNA sequencing of the synthetase gene, oligonucleotides MgenTrpRS-F and MgenTrpS-R were used as sequence primers.

Site-directed mutagenesis of MgenTrpRS

An overlapping fragment PCR strategy was employed to generate site-directed mutant *Mgen*TrpRSs that were isolated from library screening with specificity to 5-hydroxyl-Tryptophan. At the first round of PCR, the mutant *Mgen*TrpRS variant gene fragments were amplified from the plasmid p426.TrpRS with the following primers,

Mutant A: MgenTrpRS-F (5'- GCTTAGGATCCAATAATGATAAAGCGCGCAATT ACAGGGATTCAAGC-3') / MutA-R (5'- CAATATCAGGTTGATAAAGCAAAATA TCAGCAGCACA TAACACTGGGTAAG -3') and MutA-B-F (5'- GCTTTATCAAC CTGATATTG-3') / MgenTrpS-R (5'-GGACCTCGAGTTAATCAAAAAGCTGATT AGATGTTAA ACCAAATGC-3').

Mutant B: MgenTrpRS-F (5'- GCTTAGGATCCAATAATGATAAAGCGCGCAATT ACAGGGATTCAAGC-3') / MutB-R (5'- CAATATCAGGTTGATAAAGCAACACA TCAGCAGCCA CTAACACTGGGTAAG -3') and MutA-B-F (5'- GCTTTATCAAC CTGATATTG-3') / MgenTrpS-R (5'-GGACCTCGAGTTAATCAAAAAGCTGATT AGATGTTAA ACCAAATGC-3').

Mutant C: MgenTrpRS-F (5'- GCTTAGGATCCAATAATGATAAAGCGCGCAATT ACAGGGATTCAAGC-3') / MutC-R (5'-CAATATCAGGTTGATAAAGCAACGCAT CAGCAGCCAATAACACTGGG-3') and MutC-F (5'-CTTTATCAACCTGATATTG

TGCCAGCTGGTAATGATCA GAAGCAG) / MgenTrpS-R (5'-GGACCTCGAGTTA ATCAAAAAGCTGATT AGATGTTAAACCAAATGC-3').

Mutant D: MgenTrpRS-F (5'- GCTTAGGATCCAATAATGATAAAGCGCGCAATT ACAGGGATTCAAGC-3') / MutD-R (5'-GTGCTGCTTCTGATCATTACCAATTG GCGCAATATCAG GTTGATAAAGCAAG-3') and MutD-F (5'-CTTGCTTTATC AACCTGATATTGCGCCAATTGGTAATG ATCAGAAGCAGCAC-3') / MgenTrpS-R (5'-GGACCTCGAGTTA ATCAAAAAGCTGATT AGATGTTAAACCAAATGC-3').

Mutant E: MgenTrpRS-F (5'- GCTTAGGATCCAATAATGATAAAGCGCGCAATT ACAGGGATTCAAGC-3') / MutE-R (5'-GATATCAGCAGCAATTAACACTGGGTA AG-3') and MutE-F-F (5'-GCTGATATCTTGCTTTATC-3') / MgenTrpS-R (5'-GGACCTCGAGTTA ATCAAAAAGCTGATT AGATGTTAAACCAAATGC-3').

Mutant F: MgenTrpRS-F (5'- GCTTAGGATCCAATAATGATAAAGCGCGCAATT ACAGGGATTCAAGC-3') / MutF-R (5'-GATATCAGCAGCGCTTAACACTGGGTA AG-3') and MutE-F-F (5'-GCTGATATCTTGCTTTATC-3') / MgenTrpS-R (5'-GGACCTCGAGTTA ATCAAAAAGCTGATT AGATGTTAAACCAAATGC-3').

In the second round of PCR, primers MgenTrpRS-F and MgenTrpRS-R were used to amplify the resulting PCR product from the first round. All PCR products were analyzed with agarose gel electrophoresis to confirm the size and then and purified with Zymoclean Gel DNA recovery kit. The *Bam*H I/ *Xho*I restriction sites were used to cloned the mutant *Mgen*TrpRS genes to plasmid p426.pair-F and generate plasmids p426.pair-Mut.X (X=A-F).

The mutant variant genes of *Mgen*TrpRS, which was screened from the library to 5-Methoxyl-Tryptophan, was amplified from the plasmid p426.TrpRS with QuickChange PCR strategy. The primers MeO1-F (5'-CTTTATCAACCTGATATTGTTCCACA

TGGTAATGATCAGAAGCA-3') and MeO1-R (5'-GTTGATAAAGCAAGATA TCAGCAGCCAATAACACTGGGTAAG-3') were used for Mutant 1; the primers MeO7-F (5'-CTTGCTTTATCAACCTGATATTGTTCCAACTGGTAATGATCAGAA G-3') and MeO7-R (5'-CTTCTGATCATTACCAGTTGGAACAATATCAGGTTGAT AAAGCAAG-3') were used for Mutant 7. The PCR reactions were subjected to the following amplification regime; Step 1: 94 ºC for 5 min, Step 2: 94 ºC for 30 s, Step 3: 55 ºC for 30 s, Step 4: 68 ºC for 8 min (1 min/kb), Step 5: Repeat steps 2-4 20 times, Step 6: 68 ºC for 7 min, Step 7: 4 ºC forever. The DNA was purified following amplification using the QIAGEN PCR purification kit and the DNA was run on agarose gel to confirm the size. The parent plasmid DNA (methylated) was removed by digestion with the *Dpn* I enzyme for 1 h at 37 ºC, and then the resulting PCR product was transform to TOP10F' cells. After sequencing, the correct mutant *Mgen*TrpRS gene was cloned back into the p426.Pair-F plasmid using the *BamH* I/*Xho* I restriction sites to construct plasmid p426.pair-Mut.1 and p426.pair-Mut.7.

Test protein expression with unnatural amino acid

The yeast cells hosting plasmid p423. N-W66 and p426.pair-Mut.X were grown in 2 mL of SD-Ura-His (2% glucose) medium overnight at 30 ºC with shaking at 225 rpm. On the next day, the starter culture was diluted 1 part in 20 to a final volume of 10 mL culture. When the OD_{600} reading reached 0.8-1 A.U., the cell culture were centrifuged at 1000 x g for 10 min at 4 °C and washed twice with equal volumes of sterile cold 0.9% NaCl. Next, the cells were resuspended in 200 µL of sterile cold ddH_2O, and, to each 100 µL of cell suspension, was added to 4.9 mL of SD-His-Ura (2% galactose + 2% raffinose) medium and SD-His-Ura+1mM UAA (2% galactose + 2% Raffinose) medium,

respectively, to induce the expression of ECFP. Cells were harvested after 24 to 48 h after induction with agitation.

Aliquots (5 mL) of each cell line were collected by centrifugation at 2000 × g for 10 min at 4 °C, and washed with 2 mL sterile cold 0.9% NaCl. Approximately 6×10^7 cells (OD_{600} = 1.0 A.U., 2 mL) were collected and then lysed by Y-PER Yeast Protein Extraction Reagent (Thermo Scientific) to get the whole-cell soluble proteins. Afterwards, 30 μL of the cellular protein extracts were separated using SDS-PAGE (12%), and transferred to nitrocellulose membrane. The ECFP level was probed by the Living Colors A.v. Monoclonal Antibody (JL-8) as primary antibody, followed by the alkaline phosphatase-conjugated goat anti-mouse as secondary antibody. The blots were visualized with 1-step NBT/BCIP reagent mixture.

Large scale protein expression with unnatural amino acid

The yeast cells hosting plasmid p423. N-W66 and p426.pair-Mut.X were grown in 5 mL of SD-Ura-His (2% glucose) medium overnight at 30 ºC with shaking at 225 rpm. The starter culture was used inoculate 50 mL liquid culture of SD-Ura-His (2% glucose) medium and the cells were grown at 30 ºC for 24 h with shaking, and then the 50 mL cell culture was added into 1L of SD-Ura-His (2% glucose) medium and grew to an OD_{600} = 1.0 A.U.. The cells were harvested by centrifugation of the culture at 2000*g* for 10 min at room temperature and washed twice with sterile 0.9% NaCl solution. The cell pellet was fully re-suspended in 1L medium of SD-Ura-His+2mM UAA (2% galactose + 2% raffinose) and grew for 48 hours at 30 ºC with shaking. The cells were collected with centrifugation at 4000*g* for 10 min at room temperature and the cell pellet was either subjected to lysis or frozen at –80 ºC.

Purification of ECFP

The purification of ECFP was facilitated by an *N*-terminal His_6 tag. An aliquot (3 mL) of YPER protein lysis reagent containing the protease inhibitor cocktail (EDTA-free) was used to fully resuspend each gram of yeast cell pellet. The cell suspension was shaken at room temperature for 30-60 min and clarified by centrifugation at 10,000*g* for 10 min at 4 ºC. The supernatant was removed to a Falcon tube (50 mL) loaded onto Qiagen Ni-NTA agarose column. At least 15 column volumes of washing buffer (50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, pH 7.4) was used to wash off the non-specifically bound proteins and the ECFP was eluted by 10 column volumes of elution buffer (50 mM Tris-HCl, 300 mM NaCl, 250 mM imidazole, pH 7.4)

Fluorescent spectra analysis

The fluorescence values of ECFP and ECFP variants were determined at 20 ºC in aqueous 100 mM NaCl and 50 mM Tris-HCl buffer at pH 7.4. Excitation wavelengths were set at 433 nm and 450 nm respectively, which are the major and minor excitation peak at the fluorescence spectra of ECFP. The fluorescence spectra were recorded on a Horiba Jobin Yvon Fluoromax-3 spectrometer equipped with digital software.

Fluorescence microscopy

Fluorescent images were obtained using a Zeiss LSM510 UV confocal laser scanning microscope (Carl Zeiss Inc., New York, NY) at wavelength 405 nm, and image analysis was conducted using a Zeiss LSM image browser (Carl Zeiss). When the ECFP had been induced to express for 24 h, cells were washed with 0.9% NaCl, and re-suspended in PBS, and subjected to fluorescence microscopy.

Result and Discussion

Analysis of the screening effect of *ura3-14* reporter gene

To isolate the *Mgen*TrpRS mutants with changed amino acid specificity to tryptophan analogues in response to the UGA opal codon, a two-step "double sieve" screening strategy was used with the *ura3-14* gene. In the positive selection, cells with a library of *Mgen*TrpRS were grown in the presence of tryptophan analogue and in the absence of uracil. Cells expressing *Mgen*TrpRS mutants that can aminoacylate *Mgen*tRNA with tryptophan and/or a tryptophan analogue will produce full length *URA3* gene product by suppression of the UGA mutation in the *ura3-14* allele. The cells containing these functional mutants are prototrophic for uracil biosynthesis and capable of survival on uracil drop-out medium. Subsequently, the surviving cell population was subjected to a negative selection, in which tryptophan analogue was absent in the medium but the anti-metabolite 5-fluoroorotic acid was present. If the *Mgen*TrpRS mutants recognize and charge the canonical amino acid tryptophan to its cognate tRNA, the *ura3-14* gene will be expressed as the full-length product. This situation results in the conversion of 5-FOA into the toxic compound 5-fluorouracil, which inhibits cell growth.(Boeke et al., 1984) Using this selection scheme, we anticipated that, after several rounds of selection, cells harboring expressing *Mgen*TrpRS mutants that can aminoacylate *Mgen*tRNA with tryptophan analogues would be culled from the library.

Before attempting to screen the library *Mgen*TrpRS mutants, the screening efficiency of the *ura3-14* gene was examined. It was reported that the suppression of the *ura3-14* gene by [*PSI*+] appeared to be less efficient at 30 °C than at 21 °C. (Manogaran et al., 2006) In the first generation screening method, the plasmid p423.pair and a low-copy reporter

plasmid (containing a single UGA mutation within the *N*-terminal leader sequence to the *URA3* gene) were co-transformed into *S. cerevisiae* strain 74D-694. The resulting transformants grew on SD-Leu-His-Ura solid medium at 30 °C and 21 °C, which indicated that the orthogonal *Mgen* pair effectively suppressed the UGA codon within the *ura3-14* allele. In contrast, when the same *S. cerevisiae* cell line was transformed with the pLEU2ura3-14 reporter plasmid and the plasmid p423GPD, a construct lacking the *Mgen* Pair, no growth was observed on SD-Leu-His-Ura plate at 21 °C but significant growth was shown at 30 °C. This experiment established that the *Mgen* pair was expressed in a functional form and suppressed the UGA codon in the *ura3-14* gene and supported cell growth on uracil drop-out media, but the positive selection worked better at 21 °C. (Figure 3-5(i)) However, when the negative selection was tested, cells containing *Mgen* Pair and a low-copy reporter pLEU2ura3-14 exhibited unexpectedly growth on SD-Leu-His+0.1% 5-FOA plate at both 30 °C and 21 °C. In the negative selection, the growth of cells with *Mgen* pair indicated that the suppression efficiency was insufficient to support a level of expression of *ura3-14* gene that permitted effective counter-selection in the presence of 5-FOA. (Figure 3-5 (ii))

To further improve the dynamic range of the ura3-*14* reporter gene, a high-copy 2μ plasmid, which has a copy number 10-30 times of the low-copy plasmid employed for pLEU2ura3-14, was used to increase the expression level of *ura3-14*. Two constructions of the reporter gene were tested at 30 °C and 21 °C, one with the PGK promoter from the pLEU2ura3-14 (p425PGK.Ura3) and the other with constitutive GPD promoter (p425GPD.Ura3). At 21 °C, when the reporter *ura3-14* was under the control of PGK promoter, cells grew only on the SD-Leu-His-Ura plate in the presence of the *Mgen* pair; however, a high background of cell growth was observed for cells harboring empty

plasmid and p425GPD.Ura3 on the same solid medium. (Figure 3-7A) Similarly, for the negative selection, the p425PGK.Ura3 construct also displayed better selectivity than the p425GPD.Ura3, in which more cells survived with the absence of the *Mgen* pair. (Figure 3-7B) Nevertheless, at 21 °C, even the reporter gene that had been cloned to high-copy plasmid, the negative selection still showed high background that limited the dynamic selection range and was not suitable for library screening.

Next, the selection systems were evaluated at 30 °C, a temperature optimal for yeast growth. As shown in Figure3-8A, at the positive selection, cells grew in the presence of active *Mgen* pair and no detectable cell growth was observed with empty plasmid; in contrast, at the negative selection, cells survived in the absence of *Mgen* pair. It was noticeable that in the presence of the *Mgen* pair, on the 5-FOA plate, the ratio of surviving cells was slightly higher for cells maintaining the p425PGK.Ura3 plasmid in comparison to those containing p425GPD.Ura3. (Figure3-8B) Given that the negative selection gave more reliable results for the *ura3-14* reporter gene expressed at 30 °C with GPD promoter, we employed these conditions for further library selection. The experimental results indicated that an increase in the copy number of the plasmid can extend the dynamic range of expression of the *ura3-14* gene at 30 °C.

Site-saturation mutagenesis of *Mgen*TrpRS

5-Hydroxy-tryptophan, a tryptophan analogue exhibiting red-shift absorbance with respect to tryptophan, was selected as an initial substrate for experiments directed towards selection of a TrpRS with novel selectivity. Zhang, *et. al.*, have described a single mutation (V144P) that enabled the *Bacillus. subtilis* tryptophan-aminoacyl-tRNA-synthetase (*Bs*TrpRS) to incorporate 5-OH-Trp into proteins in mammalian cells. (Zhang

et al., 2004) Based on the highly homology among active site of TrpRSs in different organisms, we subjected the corresponding residue, Val151, of *Mgen*TrpRS at a position that was homologous to residue Val144 of the *Bs*TrpRS to mutagenesis to proline. However, when the *Mgen*TrpRS$_{V151P}$ was co-expressed with the *Mgen*tRNA and ECFP$_{UGA}$, no full-length ECFP could be detected by Western-Blot analysis. Therefore, the experimental results indicated that the *Mgen*TrpRS$_{V151P}$ is not suitable to incorporate 5-OH-Trp into protein in yeast. Consequently, site-saturation mutagenesis was pursued as an alternative method to alter the substrate specificity of the *Mgen*TrpRS enzyme. Since the crystal structure of the *Mgen* TrpRS has not been reported, we used ClustalW to analysis the primary sequence of *Mgen*TrpRS and *Bst*TrpRS, the latter of which has been structurally characterized. (Figure 3-3) Thus, based on the crystal structure of the homologous *Bst*TrpRS (PDB 1MB2), four residues at the active site of *Mgen*TrpRS were chosen for site-saturation mutagenesis (Met137, Ile141, V149 and V151). Each of the four residues selected as mutation target were in close contacts (~5 Å) with the 5-position of the indole ring of the *L*-tryptophan. (Figure 3-1)

In the first attempt to randomize the four residues, a library of *Mgen*TrpRS gene was created using an overlapping PCR approach and then the final PCR product was directly ligated to the *Bam*H I/*Xho* I double-digested plasmid vector p423.pair. However, direct ligation between the PCR products and yeast expression vector (7 kb) limited the ligation efficiency, resulting in low number of independent clones (2×10^3 colonies with 1 μg of DNA). To increase the ligation efficiency, we initially ligated the library of PCR products to a smaller *E. coli* cloning vector, pUC19, using the *Bam*H I/ *Pst* I restriction sites within the polylinker of the plasmid. Transformation of the ligation product to TOP10F' competent cells resulted in approximately 2.6×10^6 independent clones, in

which the mutagenesis of all four sites was confirmed by DNA sequence analysis of ten randomly selected clones. Given that four residues were selected for saturation mutatagenesis with the twenty proteinogenic amino acids, the theoretical library size at the protein level is 1.6×10^5 (20^4), thus the independent clones obtained can cover theoretical library size. In the next step, plasmids were collected from the clones and the pool of *Mgen*TrpRS mutants was sub-cloned back to the yeast expression vector p423.pair to construct library expression plasmid, p423.Lib. The ligation efficiency between cell-produced DNA inserts and yeast expression vector were higher, thus 3×10^5 independent clones were obtained and transformed into GT17 cells for library screening.

Library screening of *Mgen*TrpRS mutants with activity toward 5-OH-Trp

After transformation, the GT17 cells harboring 3×10^5 independent *Mgen*TrpRS mutagenesis variants were subjected to three rounds of positive and negative selection. Thirty-two colonies were randomly picked out for DNA sequence analysis and 4 mutants of *Mgen*TrpRS were isolated, which were identified as Mut A (M137C), Mut B (M137V/I141V), Mut C (M137L/I141A/V151A) and Mut D (V149A/V151I). Because the mutation of Met137 appeared within most of the mutants, we decided to investigate the role of the Met137 mutation. A site-saturation library of *Mgen*TrpRS mutants was generated in which the Met137 was replaced with all 20 of the common amino acids. After positive and negative selection, 2 additional mutants was obtained, which were Mut E (M137I) and Mut F (M137S).

The mutant TrpRS genes that we isolated from the library screening were amplified using PCR and cloned into the expression plasmid to generate plasmids p426.pair-Mut.X (X=A-F). Each *MgentRNA*$^{Trp/UCA}$/*Mgen*TrpRS mutant was co-transformed with p423.N-W66 into GT17 cells, converting the yeast strains to histidine

and uracil prototrophy. Consequently, each transformant was assayed for growth dependence on 5-OH-Trp by suppression of UGA nonsense codon at the *ade1-14* allele. In the presence of 5-OH-Trp, a selectively active *Mgen*TrpRS mutant can mediate the suppression of *ade1-14*, resulting in growth of the strain on SD-His-Ura-Ade+5OH-Trp solid medium. Failure of the strain containing the same *Mgen*TrpRS mutant to grow on SD-His-Ura-Ade plate provides indirect evidence of attenuated activity of the synthetase toward tryptophan. As expected from the growth of the *Mgen*TrpRS mutants assayed on SD-His-Ura-Ade+5-OH-Trp solid medium (Figure 3-9), the activity of Mut F in cells was relatively low. The light pink color results from a low level of pigment accumulation due to a partially disrupted adenine synthetic pathway, i.e., attenuated activity of protein *N*-succinyl-5-aminoimidazole-4-carboxamide ribotide (SAICAR) synthetase due to a lower level of non-sense suppression. (Chernoff et al., 1995b; Cox, 1965b) However, when all of the corresponding strains were assayed on the SD-His-Ura-Ade plate, Mut C exhibited the most robust growth in the absence of 5-OH-Trp, which suggested that this *Mgen*TrpRS variant didn't have a bias to activation of 5-OH-Trp over Trp. The Mut A, Mut D and Mut E also grew on the SD-His-Ura-Ade plate, but obviously worse than the growth on SD-His-Ura-Ade+5-OH-Trp plate, indicating these mutants prefer 5-OH-Trp as substrate rather than tryptophan. The cleanest preference for 5-OH-Trp-dependent activity was observed for the *Mgen*TrpRS mutant, Mut B, as only trace cell growth was observed on the SD-His-Ura-Ade plate. (Figure 3-9)

In addition, the selectivity for 5-OH-Trp of all *Mgen*TrpRS mutants was subjected to characterization using an assay based on suppression of W66STOP mutation in the target protein $ECFP_{UGA}$, of which the expression level can be revealed by Western blot analysis. When the expression of $ECFP_{UGA}$ was induced for 16 hours, although all

*Mgen*TrpRS variants still exhibited moderate fidelity to Trp, the Mut A and Mut B showed noticeably higher activity toward 5-OH-Trp compared to other mutants. In this case, 5-OH-Trp is an active substrate for the mutant *Mgen*TrpRS variants A and B. (Figure 3-10 A) In such *Mgen*TrpRS variants, the mutagenesis substituted the large hydrophobic residues Met137 and Ile141 with smaller ones (A: M137C; B: M137V/I141V), therefore, the binding pocket was enlarged to adopt a larger substrate 5-OH-Trp. Given that target protein expression depended on suppression of UGA nonsense codon, the protein expression level was limited over 16 hours, therefore extending expression time is useful to increase the accumulation level of target protein. Unexpectedly, a significant elevation of Trp incorporation is observed at Mut A with extending expression time to 48 hours. Despite the fact that Mut B also displayed a rise of taking Trp, the incorporation efficiency to 5-OH-Trp was still noticeably higher than that of Trp. (Figure 3-10B) Taken together with the result from –Ade plate assay, these experiments suggested that Mut B was a better substrate with which to carry out large scale expressions directed toward incorporation of 5-OH-Trp.

A possible explanation to rationalize the retention of modest activity of the *Mgen*TrpRS mutants toward tryptophan is that the Asp140 residue was left intact in all of these mutants. Although the Asp140 is not in the vicinity (~5Å) of the 5-position of the indole moiety of tryptophan, this residue is homologous to residue Asp132 of *Bst*TrpRS that was suggested as the determinant of tryptophan specify for TrpRS.(Antonczak et al., 2011; Carter, 1993; Doublie et al., 1995; Retailleau et al., 2007) Therefore, the intact Asp140 in Mut A and Mut B might be the reason associated with the reserved activity to tryptophan. However, since the Asp side-chain forms a hydrogen-bonded interaction with

the NH group of the indole ring, its removal may also abrogate activity of the mutant synthetases to 5-hydroxy-tryptophan as well.

Library screening of *Mgen*TrpRS mutants with activity toward 5-MeO-Trp

The GT17 cells harboring 3×10^5 independent *Mgen*TrpRS mutagenesis variants pool that was the same as previously described were also subjected to 2 rounds of selection (Positive-Negative-Positive). Nine *Mgen*TrpRS mutants were isolated, which were designated MeOx (X=1-9); however, when these mutation were cloned into the expression plasmide after PCR amplification to generate plasmids p426.pair-Mut.X (X=1-9), only the MeO1 and MeO7 *Mgen*TrpRS mutants supported suppression of the UGA codon in *S. cerevisiae* strain 74D-694. Consequently, these two mutants were subjected to characterize the 5-MEO-Trp incorporation with the suppression of W66STOP mutation at the target protein $ECFP_{UGA}$ by western blotting analysis. In the presence of MeO1 mutant, the $ECFP_{UGA}$ expression level of the target protein is similar in media supplemented with 5-MeO-Trp and Trp and with Trp alone. However, the MeO7 mutant displayed a preference toward 5-MeO-Trp over Trp because the $ECFP_{UGA}$ expression level is high with the presence of 5-Meo-Trp and Trp than with Trp alone. (Figure 3-11) Therefore, the MeO7 mutant was selected for large-scale expression of $ECFP_{UGA}$ directed toward 5-MeO-Trp incorporation. The reason for the remaining activity to tryptophan in MeO7 might be the same as that postulated earlier for the 5-OH-Trp-selective *Mgen*TrpRS mutants A and B, i.e., that the Asp140 residue is left intact after mutagenesis. We anticipate that its removal might abrogate activity towards the non-canonical substrate; however this hypothesis remains to be tested.

Spectral analysis of ECFP and UAA incorporated ECFP variants

The spectroscopic properties of ECFP and UAA incorporated ECFPs were subjected to fluorescence analysis. Upon the excitation of 433 nm, the fluorescence emission spectrum of wild-type ECFP (wtECFP) displayed a major peak at 475 nm and a minor peak at 501 nm. Compared to the spectrum of wild-type ECFP, the incorporation of 5-MeO-Trp at the UGA66 codon resulted in a slightly red-shift of the major emission peak, from 475 nm to 481 nm. (Figure 3-12A) In a simple model, it is expected that an increase in the size of the delocalized π-system leads to red-shift of the wavelength, (Palm and Wlodawer, 1999). The lone-pair electrons of the methoxyl oxygen atom are conjugated to the delocalized π-system of the chromophore at the ECFP, resulting a slight red-shift of the emission peak. The major emission peak was also observed at a similar position upon incorporation of 5-OH-Trp into ECFP (5-OH-ECFP), but, in addition, it was observed that the 5-OH-ECFP exhibited a significant high intensity peak at the 503 nm. (Figure 3-12A) To further investigate the spectroscopic change associated with 5-OH-Trp substitutions, the excitation wavelength was tuned to 450 nm, which corresponds to the shoulder excitation peak with ECFP. As expected, the intensity of emission peak at 503 nm increased, showing a stronger signal at 503 nm and weaker at 475 nm. (Figure 3-12B) The presence of double emission peaks in ECFP fluorescence is likely due to two different conformations found in the crystal structure of ECFP, which result from alternative displacement of two hydrophobic residues (Tyr145 and His 148) to the surface of the protein, exposed to solvent. (Bae et al., 2003; Rizzo et al., 2004; Seifert et al., 2002) Therefore, the emission peak at 503 nm of 5-OH-ECFP may be associated with the minor conformation, in which the His148 residue flips toward the chromophore. The 5-hydroxyl

substituent on the indole ring may stabilize the minor conformation of ECFP and result in a stronger peak at the 503 nm.

Confocal microscopy image analysis of cells expressing ECFP and UAA incorporated ECFP variants

Yeast cells expressing ECFP and 5OH-ECFP were excited with a 405 nm laser and imaged on a Zeiss LSM 510 META confocal microscope with multi-track mode. NFT 490 nm was used as secondary emission dichroic bean splitter, which reflects wavelengths less than 490 nm 90 degree. For the ECFP, the fluorescence mainly appeared at the 475 nm filter that indicated as the cyan channel; (Figure 3-13A) in contrast, the cells expressing 5-OH-ECFP appeared at both cyan channel and the channel with 505 nm long-pass filter, which was indicated as yellow channel in this chapter. (Figure 3-13B) The images from multi-fluorescent channels could be overlaid with those obtained with the transmitting light channel (ChD) and displayed the fluorescent property of cells at the 405 nm laser excitation. As shown in Figure 3-13, the cells expressing ECFP exhibited a cyan color, while the cells expressing 5-OH-ECFP were imaged as yellow. These results are consistent with the respective fluorescent emission spectra of ECFP and 5-OH-ECFP.

Conclusion

A double sieve selection strategy was developed for screening *Mgen*TrpRS mutation variants to incorporate 5-OH-Trp and 5-MeO-Trp in response to the UGA codon in yeast. With site-saturation mutagenesis at the active sites of *Mgen*TrpRS, approximately 3×10^5 independent *Mgen*TrpRS mutants were subjected to multiple rounds

of positive and negative selection, which resulted in mutant *Mgen*TrpRSs that were selective for 5-OH-Trp (MutB) and 5-MeO-Trp (MeO7) incorporation. Although the mutants still showed modest activity to the original substrate, tryptophan, their activity still exhibited a bias for 5-OH-Trp or 5-MeO-Trp, respectively, in the presence of tryptophan. In addition, introducing another UGA codon at the *ura3-14* allele might also be helpful to expand the library selection dynamics, but the selection of the position for the nonsense codon needs further investigation as sequence context and distance between nonsense codons can have a large effect on suppression efficiency. Incorporation of 5-OH-Trp or 5-MeO-Trp in place of Trp66 in ECFP resulted in alterations in the fluorescent emission spectra that could be understood in terms of the influence of the substituent. Therefore, site-specific incorporation of tryptophan analogues would afford the possibility to create novel fluorescent protein variants.

Table 3-1. The oligonucleotides used for PCR amplification and DNA construction.

Primer Name	**Sequence**
MgenTrpRS-F	5’- GCTTAGGATCCAATAATGATAAAGCGCGCAATTACAG GGATTCAAGC-3’
MgenTrpRS-tRNA-R-R	5’- GCGCGTCGACGTTAACAAAGCATTAATGAAGAATA-3’
MgenTrpRS-R-PstI	5’- GTGACCTGCAGGGACCTCGAGTTAATCAAAAAGCTGA TTAGATGTTAAACCAAATGC-3’
Overlap M137-I141-R	5’- AATATCAGGTTGATAAAGCAAMNNATCAGCAGCMNN TAACACTGGGTAAG-3’
Overlap V149-151-F	5’- GCTTTATCAACCTGATATTNNKCCANNKGGTAATGAT CAGAAG-3’
MgenTrpS-R	5’-GGACCTCGAGTTAATCAAAAAGCTGATTAGATGTTAA ACCAAATGC-3’
Mgen-LibSeq-F	5’- CTTTATTAGCACTAGGAC-3’
Ura3-14-F-SpeI	5’- ACTGACTGACTAGTATGGGATCCGAGGTGCTAACGCC AGACTCC -3’
Ura3-14-R	5’- ACTGACTGCTCGAGTTAGTTTTGCTGGCCGCATCTTCT-3’
PGK-F	5’- ACTGACTGGAGCTCTATTTTAGATTCCTGACTTCAACT CAAGAC -3’
MutA-R	5’- CAATATCAGGTTGATAAAGCAAAATATCAGCAGCACA TAACACTGGGTAAG -3’
MutA-B-F	5’- GCTTTATCAACCTGATATTG-3’
(N = A + G + T+ C, K = G + T and M= C + A)	

Table 3-2. The oligonucleotides used for PCR amplification and DNA construction.

Primer Name	**Sequence**
MutB-R	5’- CAATATCAGGTTGATAAAGCAACACATCAGCAGCCACTAACACTGGGTAAG -3’
MutC- R	5’-CAATATCAGGTTGATAAAGCAACGCATCAGCAGCCAATAACACTGGG
MutC-F	5’-CTTTATCAACCTGATATTGTGCCAGCTGGTAATGATCAGAAGCAG
MutD-F	5’-CTTGCTTTATCAACCTGATATTGCGCCAATTGGTAATGATCAGAAGCAGCAC-3’
MutD-R	5’-GTGCTGCTTCTGATCATTACCAATTGGCGCAATATCAGGTTGATAAAGCAAG-3’
MutE-R	5’-GATATCAGCAGCAATTAACACTGGGTAAG-3’
MutE-F-F	5’-GCTGATATCTTGCTTTATC-3’
Mut-F-R	5’-GATATCAGCAGCGCTTAACACTGGGTAAG-3’
MEO1- R	5’-GTTGATAAAGCAAGATATCAGCAGCCAATAACACTGGGTAAG-3’
MEO1- F	5’-CTTTATCAACCTGATATTGTTCCACATGGTAATGATCAGAAGCA-3’
MEO7-F	5’-CTTGCTTTATCAACCTGATATTGTTCCAACTGGTAATGATCAGAAG-3’
MEO7-R	5’-CTTCTGATCATTACCAGTTGGAACAATATCAGGTTGATAAAGCAAG-3’
CFP-His-BamHI-F	5’-AGGTTAGGATCCATAATGCATCATCACCATCACCATGTGAGCAAGGGCGAG-3’
CFP-XhoI-R	5’-GCTTCCTCGAGTTACTTGTACAGCTCGTCCATG CCG-3’

Table 3-3. Plasmids used in chapter 3.

Plasmid	Character	Reference
pUC19	*E.coli* cloning vector, Amp^R	NEB
P423 GPD	Yeast shuttle vector, *2μ, HIS3,* Amp^R	(Mumberg et al., 1994)
P425 GPD	Yeast shuttle vector, *2μ, LEU2,* Amp^R	(Mumberg et al., 1994)
p426 ADH	Yeast shuttle vector, *2μ, URA3,* Amp^R	(Mumberg et al., 1994)
p423 GAL1	Yeast shuttle vector, *2μ, HIS3,* Amp^R	(Mumberg et al., 1994)
p423GPD-XhoI	Deleted *Xho* I site at p423 GPD	this study
p423.pair	p423GPD-XhoI, *Mgen*Pair p423GPD-XhoI ← *Mgen*Pair PCR (*Bam*H I/ *Sal* I)	this study
pUC19.TrpRS	pUC19, *Mgen*TrpRS pUC19←*Mgen*TrpRS PCR (*Bam*H I/ *Pst* I)	this study
pUC19.Lib	pUC19, *Mgen*TrpRS Library pUC19←*Mgen*TrpRS Library PCR (*Bam*H I/ *Pst* I)	this study
p423.Lib	p423GPD-XhoI, *Mgen*TrpRS Library p423GPD-XhoI ← pUC19.Lib (*Bam*H I/ *Xho* I)	this study
p426.TrpRS	p426 ADH, MgenTrpRS p426 ADH← MgenTrpRS PCR (*Xho* I/ *Bam*H I)	this study
p423. N-W66	p423 GAL1, mutECFP (N-Histag-W66STOP) p423 GAL1← N-His-W66 PCR (*Xho* I/ *Bam*H I)	this study
pLEU2ura3-14	Yeast Plsmid, *Amp^R CEN ARS LEU2 URA3*	(Manogaran et al., 2006)
p425GPD.Ura3	p425 GPD, ura3-14 allele p425 GPD ← ura3-14 PCR (*Spe* I /*Xho* I)	this study
p425PGK.Ura3	p425 GPD, ura3-14 allele p425 GPD ← PGK-ura3-14 PCR (*Sac* I /*Xho* I)	this study
p426.tRNA-F	p426-Age I, MgentRNA cassette p426-Age I ← tRNA PCR (*Nhe* I / *Avr* II)	this study
p426.pair-F	p426.tRNA-F, MgenTrpRS p426tRNA-F← p426.TrpRS (*Xho* I/ *Bam*H I)	this study
p426.pair-Mut.X	p426.pair-F, mutant *Mgen*TrpRS listed in Table 3-5	this study

Table 3-4. The *Mgen*TrpRS mutations used in the chapter 3.

Mutant *Mgen*TrpRS	**Residue Substitution in plasmid**	**Nucleotide Change**
A	p426.M137C	ATG→TGT
B	p426.M137V/I141V	ATG→GTG/ATC→GTG
C	p426.M137L/I141A/V151A	ATG→TTG/ATC→GCG/GTT→GCT
D	p426.V149A/V151I	GTT→GCG/GTT→ATT
E	p426.M137I	ATG→ATT
F	p426.M137S	ATG→AGC
1	p426.M137L/V151H	ATG→TTG/GTT→CAT
7	p426. V151T	GTT→ATC

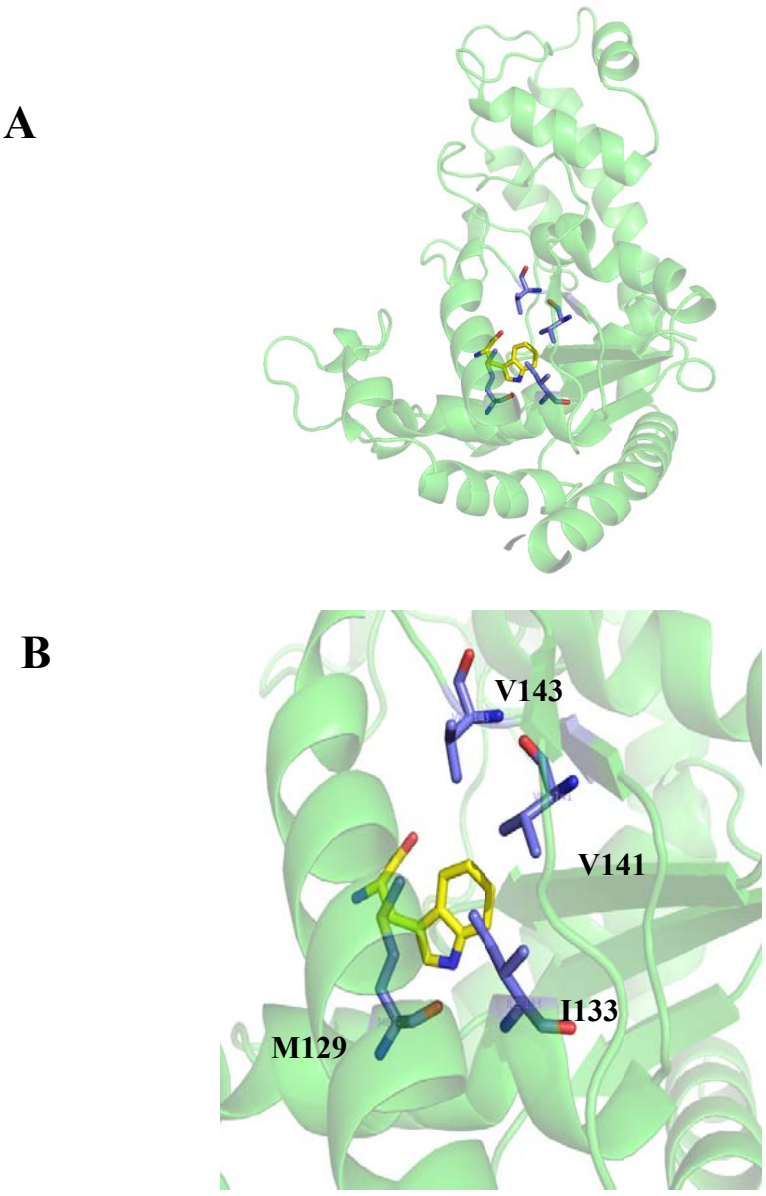


Figure 3-1. The structure of *Bst*TrpRS (PDB code: 1MB2)

(A) The structure of chain A of *Bst*TrpRS. (B) A view of the active site at the *Bst*TrpRS with residues chosen for site-saturation mutagenesis. Each of the four residues selected is within 3-5 Å of the 5-position of indole moiety of tryptophan.

Figure 3-2. ClustalW alignment between the sequences of *Bst*TrpRS and *Mgen*TrpRS

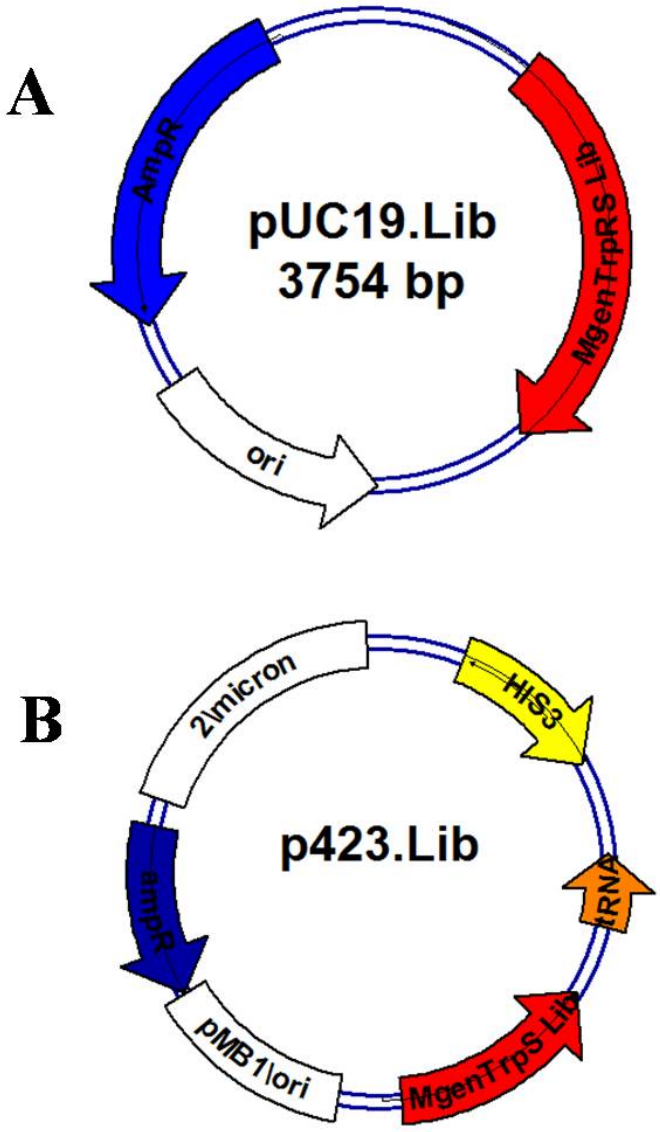


Figure 3-3. Plasmids for the construction of *Mgen* TrpRS site-saturation mutagenesis library.

The mutant *Mgen*TrpRS PCR products were cloned into pUC19 vector initially to generate pUC.Lib (A) and then the *E. coli*- amplified *Mgen*TrpRS gene library was sub-cloned into p423.Pair to construct plasmid p423.Lib that used for library screening.

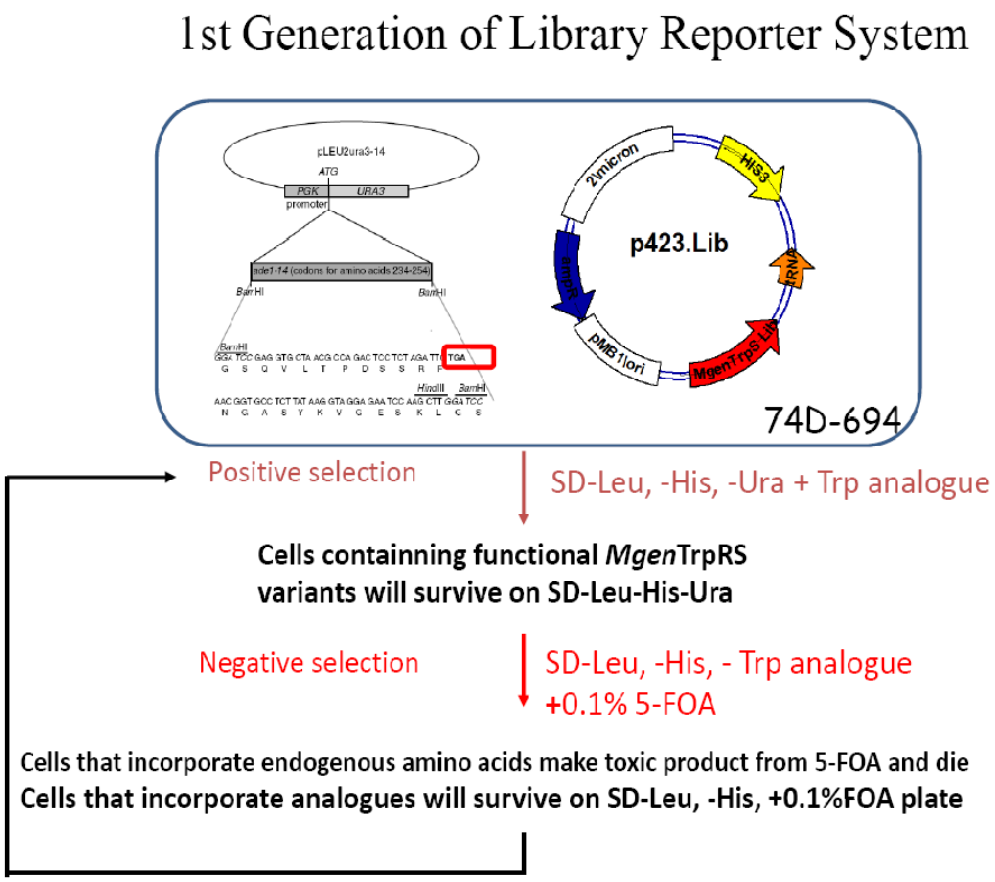


Figure 3-4. The scheme of the first-generation library screening strategy, in which the *ura3-14* reporter is located on a low-copy plasmid pLEU2ura3-14.

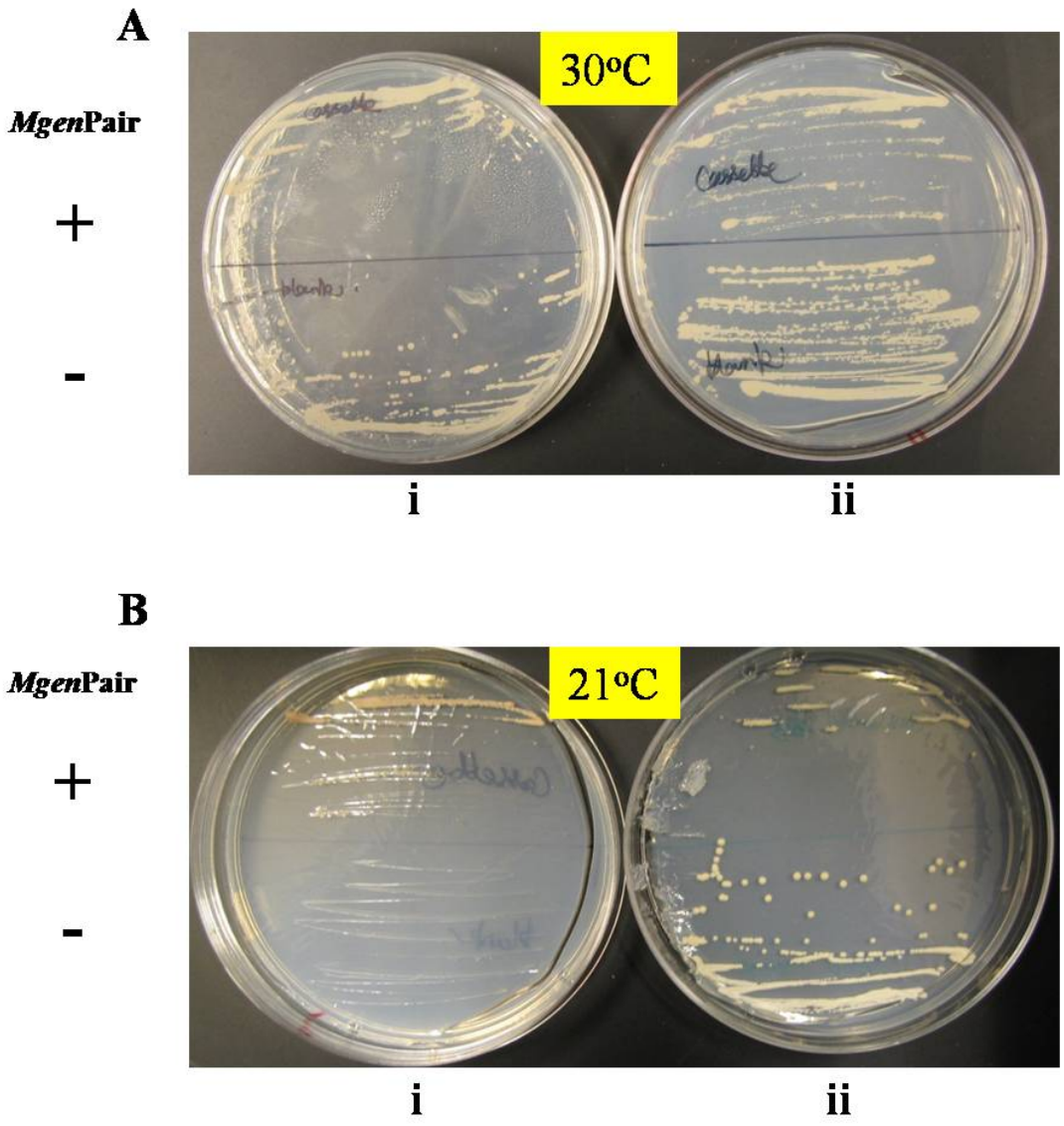


Figure 3-5. Phentotypic assay of positive (i) and negative (ii) selection for the *ura3-14* reporter on pLEU2ura3-14 for *S. cerevisiae* strain (74D-694) in the presence (+) and absence (–) of co-expression of the wild-type *Mgen*RS/tRNA pair at 30 °C (A) and 21 °C (B). Culture media for positive selection (i) is SD-Leu-His-Ura plate and for negative selection (ii) is SD-Leu-His+5-FOA plate.

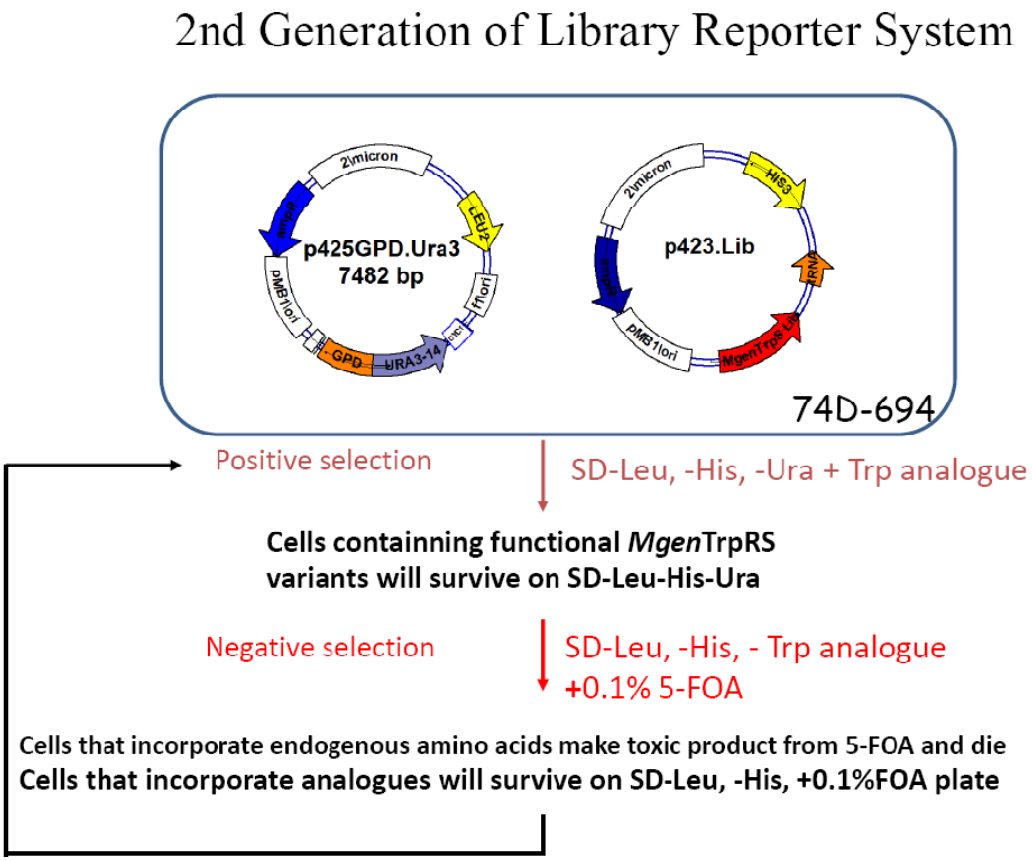


Figure 3-6. The second-generation strategy for library screening in which the *ura3-14* reporter is cloned into a 2μ plasmid.

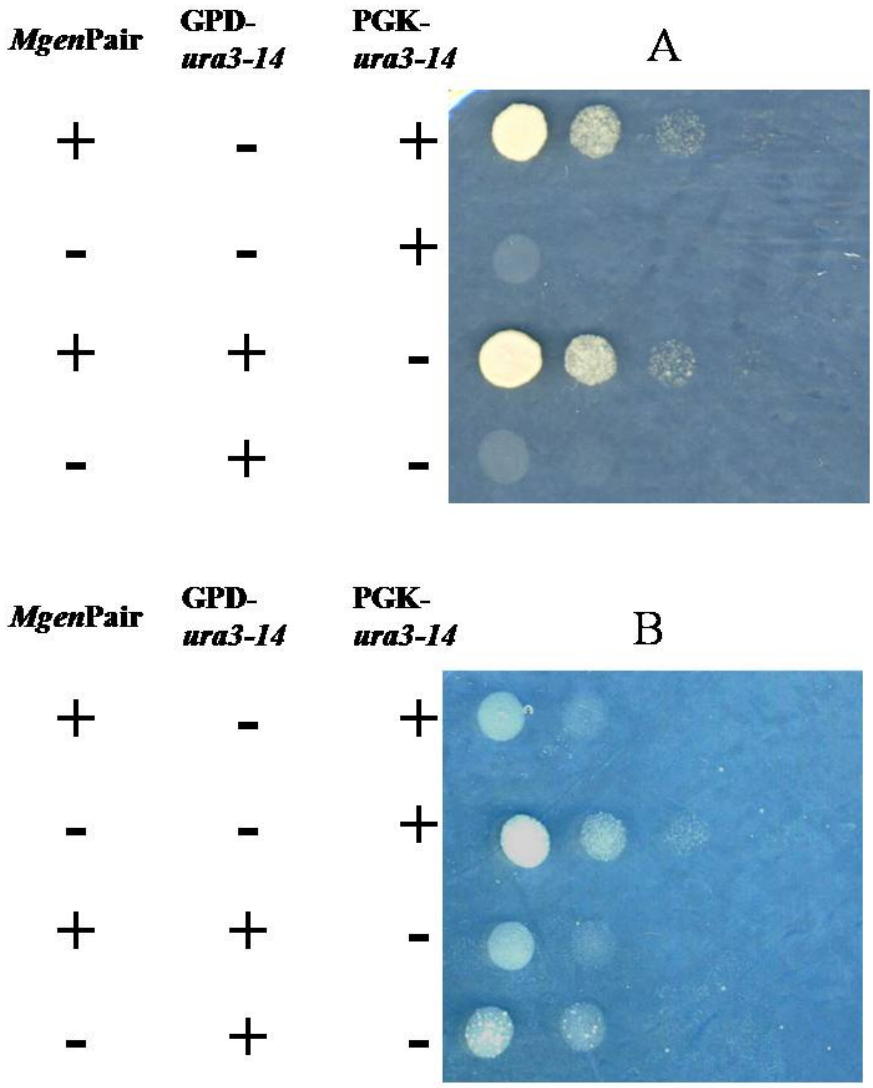


Figure 3-7. The *Mgen* pair-dependent cellular phenotypes on (A) SD-Leu-His-Ura media and (B) SD-Leu-His+5-FOA media at 21 °C.

In the second-generation strategy, the *ura3-14* reporter gene was cloned into 2μ plasmids and expressed under the control of GPD promoter and PGK promoter separately in *S. cerevisiae* strain 74D-694.

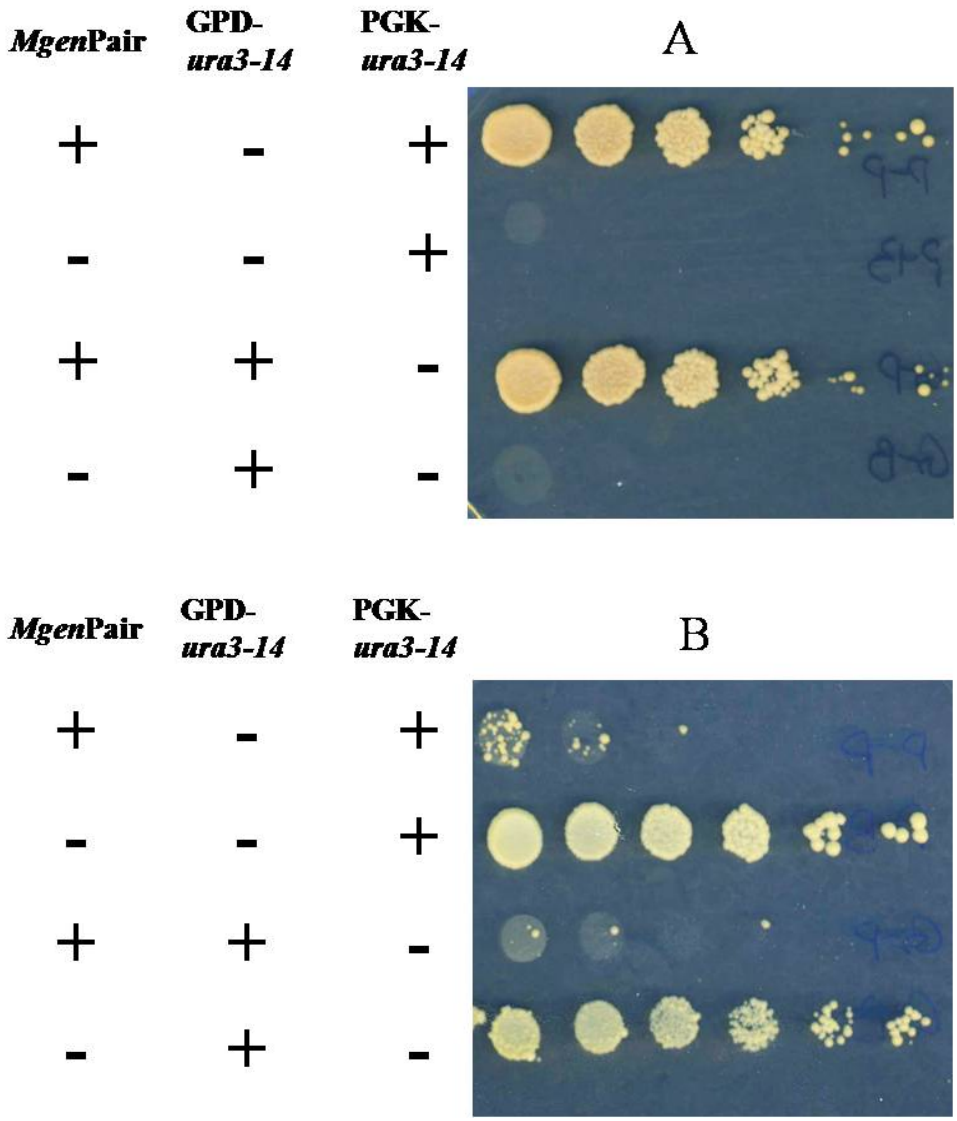


Figure 3-8. The *Mgen* pair-dependent cellular phenotypes on (A) SD-Leu-His-Ura media and (B) SD-Leu-His+5-FOA media at 30 °C.

In the second-generation strategy, the *ura3-14* reporter gene was cloned into a 2μ plasmids and expressed under the control of GPD promoter and PGK promoter separately in *S. cerevisiae* strain 74D-694.

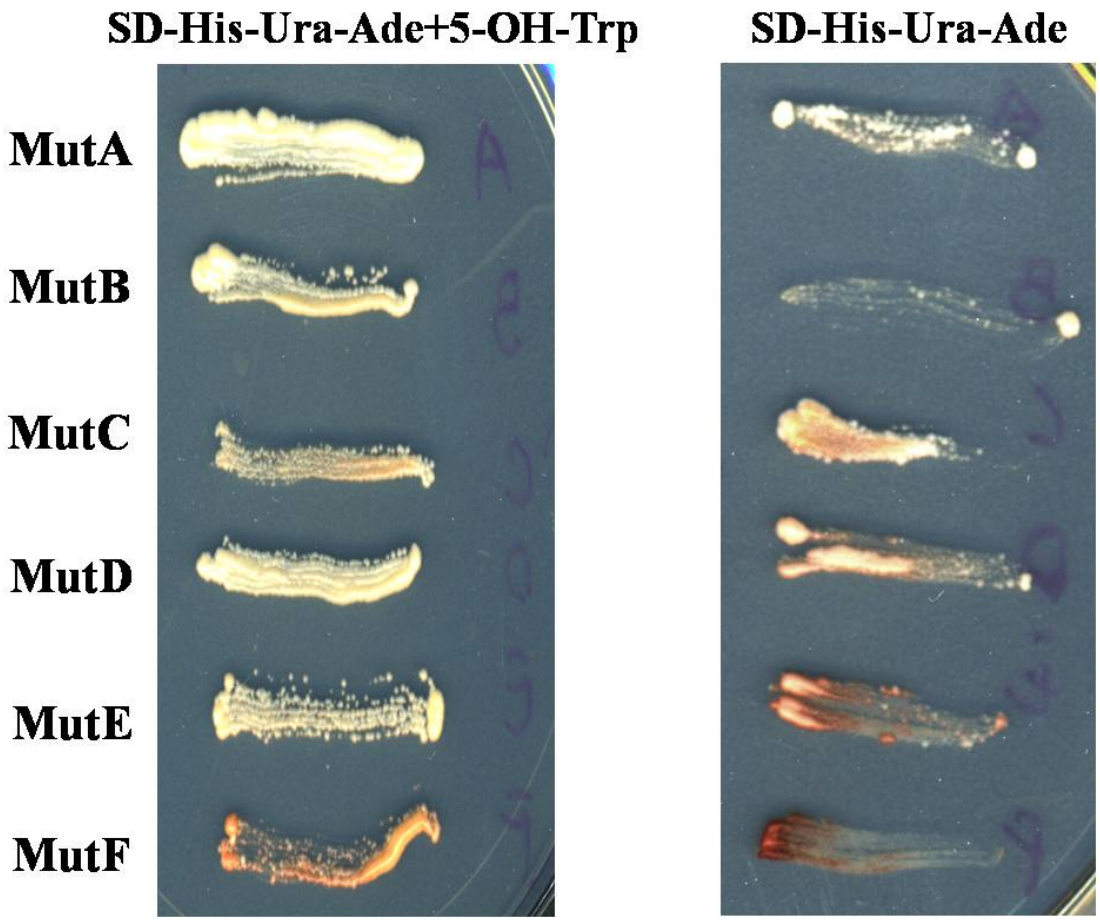
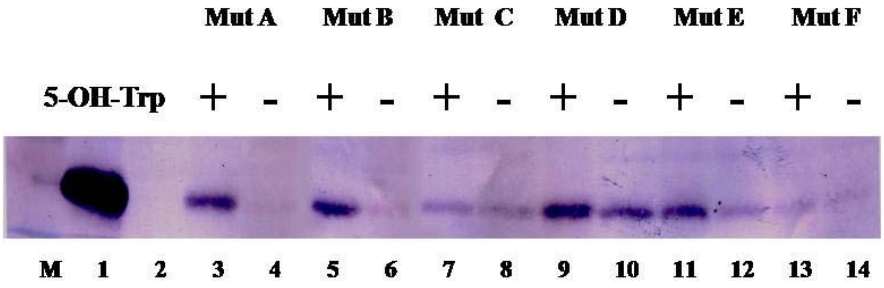


Figure 3-9. The 5-OH-Trp dependent phenotype of cells carrying *MgentRNA*$^{Trp/UCA}$ */Mgen*TrpRS mutant on selective media by suppression at the *ade1-14* allele.

On the left column, the cells containing the indicated mutants were assayed on adenine drop-out plate in the presence of tryptophan and 5-OH-tryptophan in the solid medium; cells on the right column were assayed on adenine drop-out plate in the presence of tryptophan only in *S. cerevisiae* strain 74D-694.



	Plasmid	5-OH-Trp
1	p426.pair-F + p423. N-W66	-
2	p426 ADH+ p423. N-W66	-
3	p426.pair-Mut.A+ p423. N-W66	+
4	p426 ADH+ p423. N-W66	-
5	p426.pair-Mut.B+ p423. N-W66	+
6	p426 ADH+ p423. N-W66	-
7	p426.pair-Mut.C+ p423. N-W66	+
8	p426 ADH+ p423. N-W66	-
9	p426.pair-Mut.D+ p423. N-W66	+
10	p426 ADH+ p423. N-W66	-
11	p426.pair-Mut.E+ p423. N-W66	+
12	p426 ADH+ p423. N-W66	-
13	p426.pair-Mut.F+ p423. N-W66	+
14	p426 ADH+ p423. N-W66	-

B.

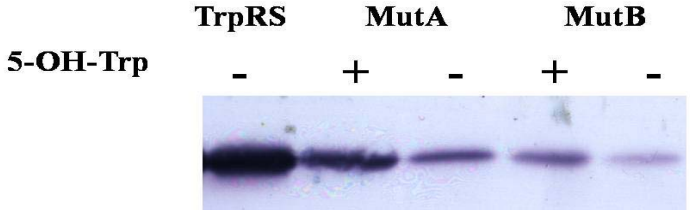
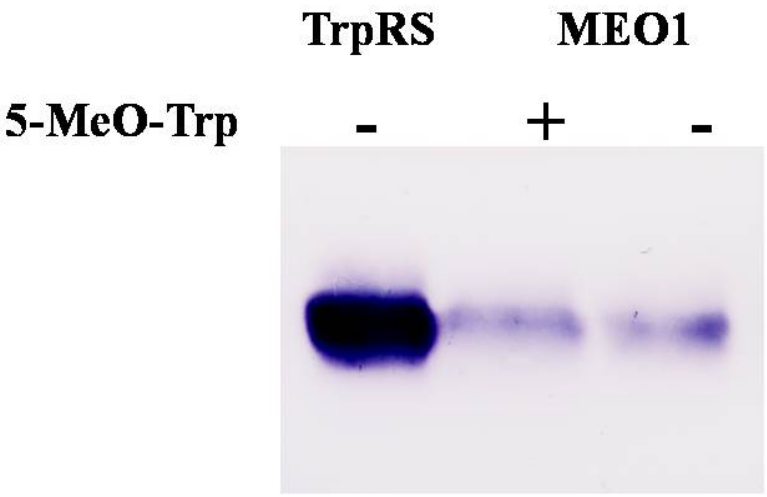


Figure 3-10. Western blot analysis of 5-hydroxy-tryptophan incorporation in S. cerevisiae strain (74D-694) containing the indicated mutant *Mgen*TrpRSs.

(A) The opal suppression efficiency of $ECFP_{UGA}$ in yeast in the presence and absence of 5-OH-Trp with the mutant *Mgen* pairs. The $ECFP_{UGA}$ was induced to express for 16 hours. The plasmids in each type of cell are listed the table. (B) Expression level of $ECFP_{UGA}$ after 48 hours for TrpRS mutants A and B in *S. cerevisiae* strain 74D-694. (The concentration of 5-OH-Trp is 1 mM and the concentration of Trp is 0.2 mM)



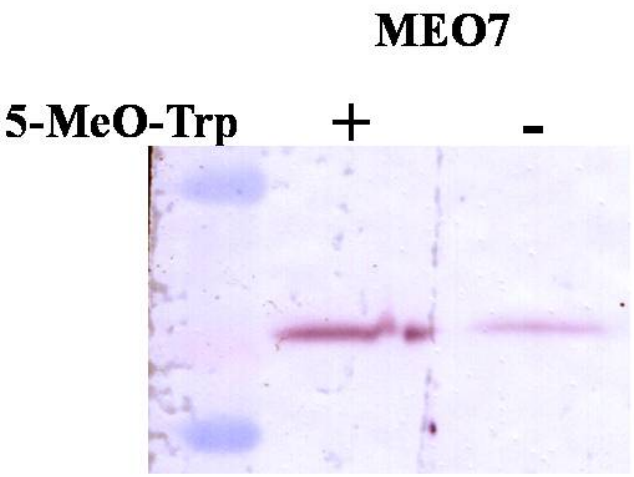


Figure 3-11. Western blot analysis of 5-MeO-Trp incorporation in *S. cerevisiae* strain 74D-694 with the *Mgen* TrpRSs mutants MeO1 and MeO7.

The mutant MeO1did not display selectivity for 5-MeO-Trp (A), but the mutant MeO7 exhibited better expression level for $ECFP_{UGA}$ in the presence of 5-MeO-Trp (B). (The concentration of 5-MeO-Trp is 1 mM and the concentration of Trp is 0.2 mM)

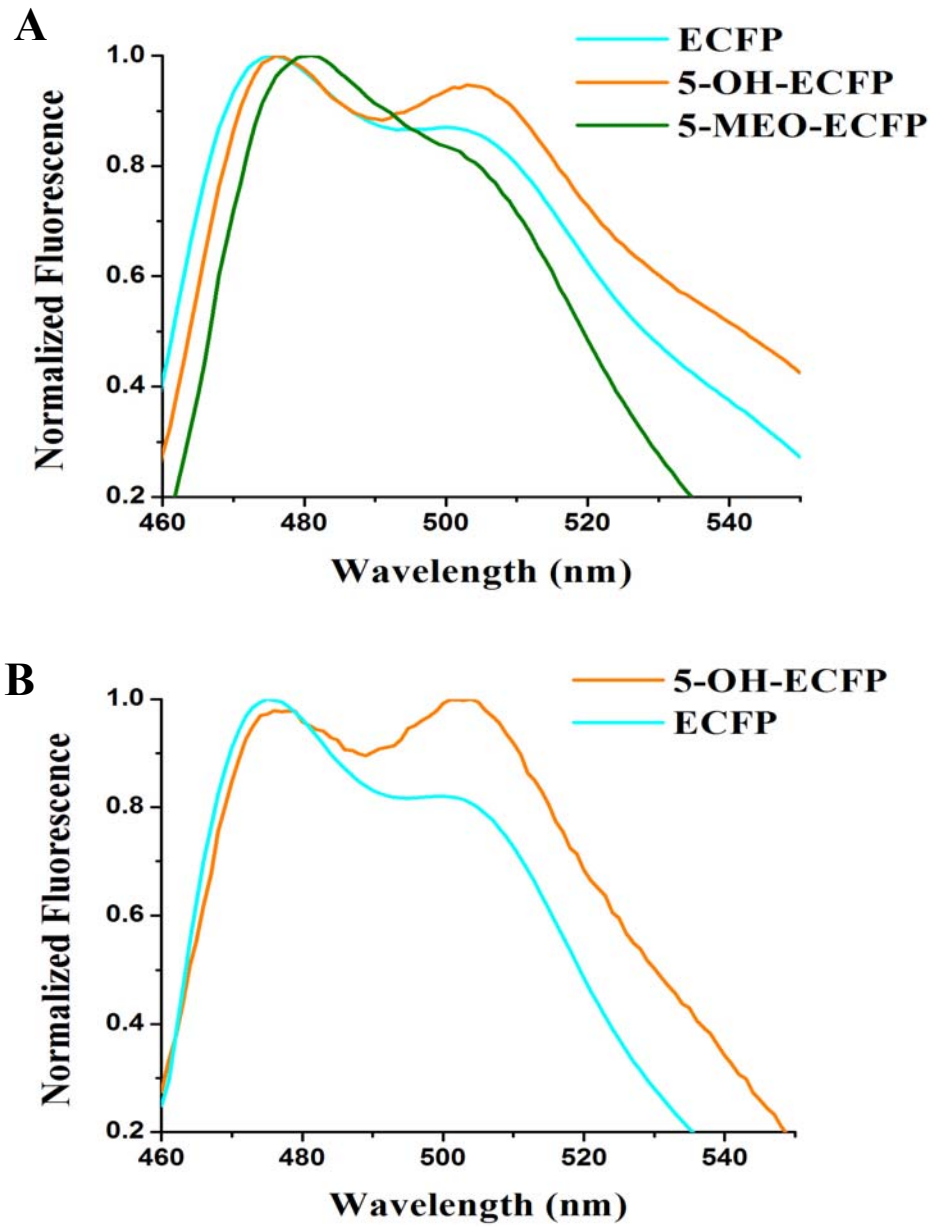


Figure 3-12. Fluorescence emission of ECFP and ECFP variants were characterized at two excitation wavelengths. When the λ_{Ex}=433 nm (A), the emission maximum of 5-MeO-ECFP was slightly red-shifted to 481 nm and the shoulder peak of ECFP was increased with respect to the 5-OH-Trp incorporation. With the excitation at 450 nm, (B) the emission maximum of 5-OH-ECFP was significantly changed at 503 nm.

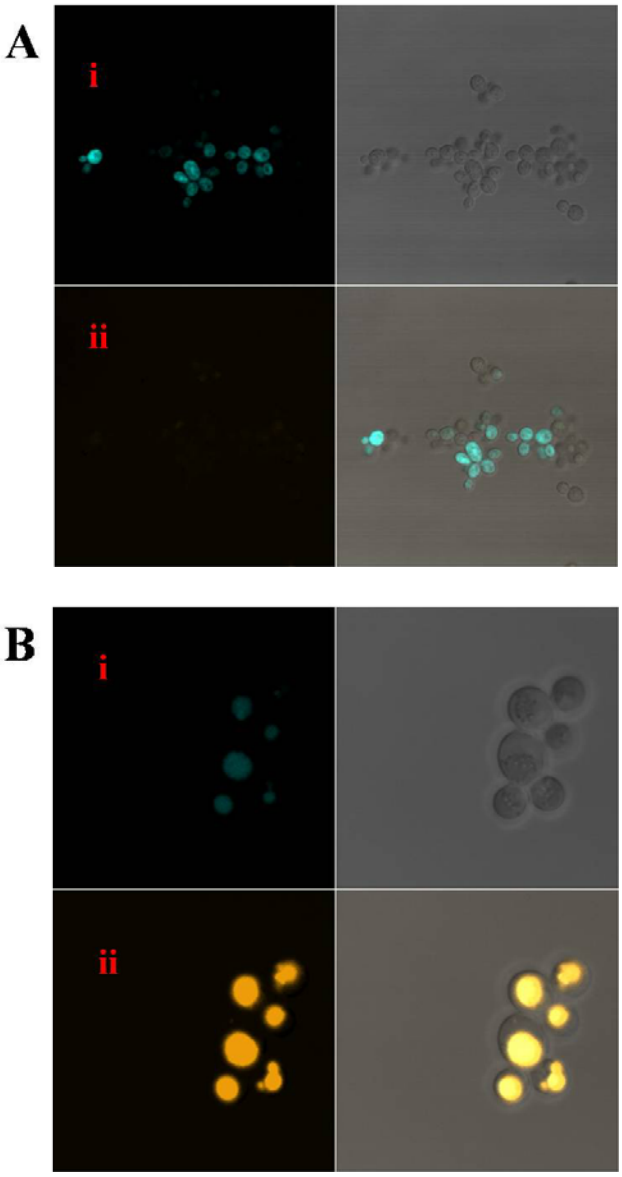


Figure 3-13. Confocal fluorescence microscopy of cells of *S. cerevisiae* strain 74D-694 expressing ECFP (A) and 5-OH-ECFP (B) were excited with the laser tuned to 405 nm and imaged with a Zeiss LSM 510 confocal microscope in multichannel mode. Channel (i) detected emission wavelength between 475 nm and 490 nm, while channel (ii) detected emission wavelength longer than 505 nm. (ECFP and 5-OH-ECFP were express in the presence of 1mM 5-OH-Trp and 0.2 mM Trp)

Reference

Antonczak, A.K., Simova, Z., Yonemoto, I.T., Bochtler, M., Piasecka, A., Czapinska, H., Brancale, A., and Tippmann, E.M. (2011). Importance of single molecular determinants in the fidelity of expanded genetic codes. Proc Natl Acad Sci U S A *108*, 1320-1325.

Bae, J.H., Rubini, M., Jung, G., Wiegand, G., Seifert, M.H., Azim, M.K., Kim, J.S., Zumbusch, A., Holak, T.A., Moroder, L.*, et al.* (2003). Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. J Mol Biol *328*, 1071-1081.

Boeke, J.D., LaCroute, F., and Fink, G.R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol Gen Genet *197*, 345-346.

Carter, C.W., Jr. (1993). Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. Annu Rev Biochem *62*, 715-748.

Cellitti, S.E., Jones, D.H., Lagpacan, L., Hao, X., Zhang, Q., Hu, H., Brittain, S.M., Brinker, A., Caldwell, J., Bursulaya, B.*, et al.* (2008). In vivo incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. J Am Chem Soc *130*, 9268-9281.

Chernoff, Y.O., Lindquist, S.L., Ono, B., Inge-Vechtomov, S.G., and Liebman, S.W. (1995). Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. Science *268*, 880-884.

Chin, J.W., Cropp, T.A., Anderson, J.C., Mukherji, M., Zhang, Z., and Schultz, P.G. (2003). An expanded eukaryotic genetic code. Science *301*, 964-967.

Chin, J.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002). Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. Proc Natl Acad Sci U S A *99*, 11020-11024.

Cox, B.S. (1965). A Cytoplasmic Suppressor of Super-Suppressor in Yeast. Heredity *20*, 505-+.

Cropp, T.A., and Schultz, P.G. (2004). An expanding genetic code. Trends Genet *20*, 625-630.

Doublie, S., Bricogne, G., Gilmore, C., and Carter, C.W., Jr. (1995). Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. Structure *3*, 17-31.

Hancock, S.M., Uprety, R., Deiters, A., and Chin, J.W. (2010). Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. J Am Chem Soc *132*, 14819-14824.

Ishioka, C., Suzuki, T., FitzGerald, M., Krainer, M., Shimodaira, H., Shimada, A., Nomizu, T., Isselbacher, K.J., Haber, D., and Kanamaru, R. (1997). Detection of heterozygous truncating mutations in the BRCA1 and APC genes by using a rapid screening assay in yeast. Proc Natl Acad Sci U S A *94*, 2449-2453.

Lee, H.S., Spraggon, G., Schultz, P.G., and Wang, F. (2009). Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe. J Am Chem Soc *131*, 2481-2483.

Liu, C.C., and Schultz, P.G. (2010). Adding new chemistries to the genetic code. Annu Rev Biochem *79*, 413-444.

Liu, D.R., and Schultz, P.G. (1999). Progress toward the evolution of an organism with an expanded genetic code. Proc Natl Acad Sci U S A *96*, 4780-4785.

Manogaran, A.L., Kirkland, K.T., and Liebman, S.W. (2006). An engineered nonsense URA3 allele provides a versatile system to detect the presence, absence and appearance of the [PSI+] prion in Saccharomyces cerevisiae. Yeast *23*, 141-147.

Mumberg, D., Muller, R., and Funk, M. (1994). Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res *22*, 5767-5768.

Palm, G.J., and Wlodawer, A. (1999). Spectral variants of green fluorescent protein. Methods Enzymol *302*, 378-394.

Retailleau, P., Weinreb, V., Hu, M., and Carter, C.W., Jr. (2007). Crystal structure of tryptophanyl-tRNA synthetase complexed with adenosine-5' tetraphosphate: evidence for distributed use of catalytic binding energy in amino acid activation by class I aminoacyl-tRNA synthetases. J Mol Biol *369*, 108-128.

Rizzo, M.A., Springer, G.H., Granada, B., and Piston, D.W. (2004). An improved cyan fluorescent protein variant useful for FRET. Nat Biotechnol *22*, 445-449.

Sakamoto, K., Murayama, K., Oki, K., Iraha, F., Kato-Murayama, M., Takahashi, M., Ohtake, K., Kobayashi, T., Kuramitsu, S., Shirouzu, M., *et al.* (2009). Genetic encoding of 3-iodo-L-tyrosine in Escherichia coli for single-wavelength anomalous dispersion phasing in protein crystallography. Structure *17*, 335-344.

Seifert, M.H., Ksiazek, D., Azim, M.K., Smialowski, P., Budisa, N., and Holak, T.A. (2002). Slow exchange in the chromophore of a green fluorescent protein variant. Journal of the American Chemical Society *124*, 7932-7942.

Vidal, M., Brachmann, R.K., Fattaey, A., Harlow, E., and Boeke, J.D. (1996a). Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. Proc Natl Acad Sci U S A *93*, 10315-10320.

Vidal, M., Braun, P., Chen, E., Boeke, J.D., and Harlow, E. (1996b). Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. Proc Natl Acad Sci U S A *93*, 10321-10326.

Wong, C.Y., and Eftink, M.R. (1997). Biosynthetic incorporation of tryptophan analogues into staphylococcal nuclease: effect of 5-hydroxytryptophan and 7-azatryptophan on structure and stability. Protein Sci *6*, 689-697.

Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G., and Schultz, P.G. (2004). The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. Nat Biotechnol *22*, 1297-1301.

Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D.S., and Schultz, P.G. (2004). Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. Proc Natl Acad Sci U S A *101*, 8882-8887.

Chapter IV: Synthetic Yeast Prions Based on a Non-NQ-rich Amyloidogenic Sequence Derived from the NAC Protein Sequence of α-Synuclein.

Introduction

Protein folding and misfolding are the ultimate ways of generating and abolishing specific cellular activity. In some case, proteins with high propensity to misfold can escape from protective folding-control regulation and form intractable aggregates in the cells or in extracellular space (Dobson, 2003). Protein misfolding disorders, including Parkinson's disease, Alzheimer's diseases, the spongiform encephalopathies and type II diabetes, are associated with protein aggregation deposit in tissues (Dobson, 2001a; Kelly, 1998).

The roles of α-synuclein (α-syn) in the pathogenesis of neurodegenerative diseases, especially in Parkinson's disease (PD), have been extensively studied since 1997 (Polymeropoulos et al., 1997). Human α-syn is a 140 amino acid protein that is found in neurons and is the major component of several pathological lesions diagnostic of specific neurological disorders, including PD, dementia with Lewy body and the Lewy body variant of Alzheimer's disease (Maries et al., 2003). In solution, α-synuclein does not adopt a uniform secondary structure; however, it displays an α-helical conformation in association with lipids. (Davidson et al., 1998; Weinreb et al., 1996). Although the function of α-synuclein is unclear, misfolded and deposited α-synuclein, which adopts a non-native β-sheet like structure, appears to be critical determinant for the development of Parkinson's disease (Cookson, 2005).

The α-syn is homologous to three other synuclein family members termed β-synuclein (β-syn), γ-synuclein (γ-syn) and synoretins. Among them, only α-syn and β-syn are expressed in the mammalian brain and α-syn has the greatest sequence homology to β-syn (Lavedan, 1998). Despite the homology between α- and β-syn, only α-syn is found

in fibrillar pathological lesions (Spillantini et al., 1998).The critical difference between α- and β-syn is a hydrophobic stretch of 12 amino acid residues (71-VTGVTAVAQKTV-82) in the middle of α-syn, which is absent in β-syn (Giasson et al., 2001). Furthermore, deletion or disruption of this domain through the introduction of charged amino acids (A76E, or A76R) reduces the rate of α-syn's fibrillization *in vitro*. Thus, the hydrophobic fragment 71-82 appears to play a critical role in the polymerization of α-synuclein (Giasson et al., 2001). Another well-known peptide segment that may play a role in α-syn aggregation is the non-Aβ-component (NAC), which is the second major component in Alzheimer's Disease amyloid (Ueda et al., 1993). NAC corresponds to residues 61-95 of α-syn and includes the 71-82 hydrophobic stretch. Subsequently, NAC was found to self-aggregate and form an amyloid fibril that has toxic effect to cells *in vitro* (El-Agnaf and Irvine, 2000; El-Agnaf et al., 1998). Although protein folding should follow the same physical rules *in vivo* and *in vitro*, the extremely crowded intracellular environment is hard to mimic in solution. To better understand the relationship between aggregation-prone sequences and the transmissibility of the aggregation phenotype, *in vivo* investigations of peptide self-association are necessary.

In order to investigate the *in vivo* self-propagation of aggregation-prone fragments in α-syn, the yeast *Saccharomyces cerevisiae* can be employed as a model system because it retains many fundamental aspects of eukaryotic biology (Khurana and Lindquist, 2010). In *S. cerevisiae*, self-associating and self-propagating protein sequences, i.e., prions, occur. The best characterized yeast prion is [PSI^{+}], an aggregated form of the protein Sup35p, a factor controlling the efficiency of translation termination (Uptain and Lindquist, 2002). Protein domain analysis has assigned the terminator factor function to the C-terminal (C) domain. The *N*-terminal region of Sup35p, consisting of *N*-domain (1-

123) and *M*-domain (124-253), is not essential for the cell growth, but is required for prion propagation. The *N*-domain has been further divided into an asparagine/glutamine-rich domain NQ (1-39) and an oligo-peptide repeat domain NR (40-123). Mutational analysis has shown that the NQ_{1-39} domain was critical for prion behavior (DePace et al., 1998; Paushkin et al., 1996). Although many sequences have been observed in *S. cerevisiae* proteins with N/Q content as high as that of Sup35p, only a limited number of such sequences can form prions with ability to propagate (Osherovich et al., 2004; Sherman, 2004). To distinguish heritable aggregates from general N/Q rich aggregation-prone sequences, the NQ domain of Sup35p was substituted with other N/Q-rich sequences from native yeast proteins. A new yeast prion, [NU^{+}], was discovered using this approach, in which a high NQ content sequence from the corresponding protein New1p was used to replace the NQ domain of Sup35p, was demonstrated to support a prion phenotype (Osherovich et al., 2004; Santoso et al., 2000). These data also suggest that it may be possible to design novel artificial yeast prions based on substitution of N/Q rich domain with other aggregation-prone sequences. However, it remains a subject of conjecture whether non-N/Q-rich amyloidogenic sequence can support prion formation following a similar approach within a sequence context that is favorable for prion formation.

Because α-syn fragments can polymerize and subsequently induce the assembly of the full-length protein, it would be potentially capable of replacing the NQ domain in Sup35p and serving as the aggregation unit. In this study, two aggregation-prone peptides from α-syn, SYN1 (65-NVGGAVVTGVTAVA-78) encompassing the main fibrillogenic sequence 71- 82 and NAC (61- EQVTNVGGAVVTGVTAVAQKTVEGAGSIAAATG-93), were examined for their ability to sustain a prion phenotype when substituted for the

native NQ domain of Sup35p. The experimental results indicate that both Sup35p variants display phenotypes that are compatible with prion (associated) and non-prion (un-associated) states of the corresponding protein, but only the NAC-substituted Sup35p variant behaves as a novel yeast prion. This study represents the first instance in which a non-N/Q-rich sequence has been demonstrated to serve as the aggregation domain in the formation of yeast prion. Moreover, these studies suggest that other aggregation-prone peptide sequences may be employed to create synthetic yeast prions within the sequence context of Sup35p and may serve as models to study protein aggregation *in vivo*. Furthermore, the [NAC^{+}] prion might be particularly useful in exploring the factors, such as the residue substitutions, that affect the aggregation of α-syn.

Experimental Procedures

Materials

All chemicals were purchased from Sigma-Aldrich (St Louis, MO) and Fisher Scientific Co. (Fair Lawn, NJ) unless otherwise noted. Restriction endonuclease enzymes, T4 DNA ligase, and T4 kinase were purchased from New England Biolabs (Beverly, MA). Plasmids preparation kits were purchased from Qiagen (Valencia, CA). High-fidelity DNA polymerase *Pfx-50* was purchased from Invitrogen (Carlsbad, CA). Yeast transformation kits were purchased from either Invitrogen (Carlsbad, CA) or Zymo Research (Orange, CA). The yeast shuttle expression vectors were purchased from American Type Culture Collection (ATCC#87669). Synthetic oligonucleotides were purchased from either Sigma-Genosys, Inc (The Woodlands, TX) or Integrated DNA

Technologies (Coralville, IA) and were used as received. TALON metal affinity resin was purchased from Clontech, Inc. (Mountain View, CA).

Media and growth conditions

YPD (1%, w/v, yeast extract; 2%, w/v, peptone; 2%, w/v, glucose) and 1/4YPD (0.25%, w/v, yeast extract; 2%, w/v, peptone; 4%, w/v, glucose) media were prepared in ddH_2O and autoclaved for sterilization. To prepare solid media, agar (1.5%) was added to the media before sterilization. Synthetic minimal media were prepared using standard protocols and minimal drop-out media lacking appropriate amino acids were prepared as needed. LB media containing appropriate antibiotics were prepared using standard protocols. *S. cerevisiae* strains were grown at 30 °C and *E. coli* strains were grown at 37 °C with sufficient aeration.

Plasmids construction and strains

The plasmids and oligonucleotides used in this study are listed in Table 4-3. Two sets of oligonucleotides NAC-F/NAC-R and SYN1-F/SYN1-R were designed so that (1) DNA generated by annealing and mutually primed extension of the oligonucleotide pairs afforded genes that encoded the NAC and SYN1 peptides, respectively; (2) *Bam*H I and *Pst* I recognition sequences were included at the ends of the oligonucleotides to facilitate cloning into appropriate expression vectors; (3) for expression in yeast, the sequences encoding NAC and SYN1 were optimized based on the preferred codon usage in *S. cerevisiae*. To anneal the DNA, the mixture of NAC-F/NAC-R or SYN1-F/SYN1-R was cooled down by 1 °C per 5 minutes from 99 °C to 30 °C. The annealed DNA was consisted of the double-stranded DNA in the middle region and two long 5'-overhangs at

both ends. The overhangs were filled in with Klenow fragment to generate the full length DNA, which was then cloned into the *Bam*H I/*Pst* I-digested vector plasmid pZEro-1-NM. The generated plasmid pZErO-1-NM.NAC and pZErO-1-NM.SYN1, with the sequence confirmed by DNA sequencing, was subsequently used to sub-clone the fragment NAC-NR and SYN1-NR into p313-NGMC at the *Bam*H I/*Pfl*M I sites, generating plasmid p149.NAC and p149.SYN1.

The expression plasmid for NAC-M protein was made on backbone of pET21d. The gene encoding NAC-M protein was amplified using primers NAColi-F and NAColi-R. For the expression in *E. coli*, the sequence encoding NAC was optimized based on the preferred codon usage in *E. coli*. To anneal the DNA, the mixture of NAColi-F and NAColi-R was cooled down by 1 °C per 5 minutes from 99 °C to 30 °C. The annealed DNA was consisted of the double-stranded DNA in the middle region and two long 5'-overhangs at both ends. The overhangs were filled in with Klenow fragment to generate the full length DNA, which was then cloned into the *Bam*H I/*Pst* I-digested vector plasmid pZEro-1-NM. After confirmation of sequence, the NAColi-M gene were amplified from the plasmid pZEro-1-NM.NAColi by PCR with primers NAC-NcoI-F and NAC-M-R, and then the NAColi-M gene was introduced to pET21d with *Nco*I/*Xho*I sites. The expression plasmid for NAC-M protein was made on backbone of pET21d. The gene encoding NAC-M protein was amplified using primers NAColi-F and NAColi-R. For the expression in *E. coli*, the sequence encoding NAC was optimized based on the preferred codon usage in *E. coli*. To anneal the DNA, the mixture of NAColi-F and NAColi-R was cooled down by 1 °C per 5 minutes from 99 °C to 30 °C. The annealed DNA was consisted of the double-stranded DNA in the middle region and two long 5'-overhangs at both ends. The overhangs were filled in with Klenow fragment to generate the full length

DNA, which was then cloned into the *Bam*H I/*Pst* I-digested vector plasmid pZEro-1-NM. After confirmation of sequence, the NAColi-M gene were amplified from the plasmid pZEro-1-NM.NAColi by PCR with primers NAC-NcoI-F and NAC-M-R, and then the NAColi-M gene was introduced to pET21d with *Nco*I/*Xho*I sites.

Saccharomyces cerevisiae strains OT56 ([PSI^{+}], [PIN^{+}]), OT60 ([psi^{-}], [PIN^{+}]), and GT17 ([psi^{-}], [pin^{-}]), which are the epigenetic variants of *S. cerevisiae* strain 74D-694 (*MATa ade1-14UGA his3 leu2 trp1-289UAG ura3*), were kindly provided by Professor Yuri Chernoff. Additional information about strains is listed in Table. 1. The strains were modified to disrupt the native SUP35 gene (*sup35::KanMX*) located on the chromosome, using the method previously reported (Giaever et al., 2002). The maintenance plasmid p316-Sp-Sup35 was transformed into the strains before the chromosomal gene disruption, as the presence of SUP35 gene or Sup35p activity is essential for cell survival.

Table 4-1. Strain used in Chapter 4.

Strain	**Genotype**	**Reference**
TOP10F'	*rec*A1 *ara*D139 *end*A1 (Str^R) F'[*lac*I^q, Tn*10* (Tet^R)]	Invitrogen
OT56 ([*PSI*$^+$] 74-D694)	*MAT***a** *ade1-14*$_{UGA}$ *his3 leu2 trp1-289*$_{UAG}$ *ura3* [*PSI*$^+$]	(Derkatch et al., 1996a)
OT60 ([*psi*$^-$] 74-D694)	*MAT***a** *ade1-14*$_{UGA}$ *his3 leu2 trp1-289*$_{UAG}$ *ura3* [*psi*$^-$ *PIN*$^+$]	(Chernoff et al., 1995a)
GT17([*psi*$^-$] 74-D694)	*MAT***a** *ade1-14*$_{UGA}$ *his3 leu2 trp1-289*$_{UAG}$ *ura3* [*psi*$^-$ *pin*$^-$]	(Derkatch et al., 1997)
OT56k	OT56, *sup35::KanMX*	this study
OT60k	OT60, *sup35::KanMX*	this study
GT17k	GT17, *sup35::KanMX*	this study
74-D694-Δsup35-Δhsp104	*MAT***a** *ade1-14*$_{UGA}$ *his3 leu2 trp1-289*$_{UAG}$ *ura3* [*psi*$^-$ *pin*$^-$], *sup35::TRP1* [*HIS3, SUP35*] *hsp104::URA3*	(Hara et al., 2003)
GT17k NAC-Δhsp104	GT17, *sup35::KanMX* [*HIS3, NAC-sup35*] *hsp104::URA3*	this study
GT388	*MAT*α *lys1-1 his3 leu1 ura3 ade2-1 SUQ5 kar1-1 cyh*^{R} [*psi*$^-$ rho$^-$]	(Borchsenius et al., 2001)
GT388k	GT388, *sup35::KanMX*	this study
GT650	*MAT*a *ade1-14 his3 leu2 lys2 ura3 trp1 sup35::HIS3* [*psi*$^-$ *PIN*$^+$] [CEN ARS URA3 SUP35]	Provided by Dr. Chernoff
GT810	the same as GT650 but *MAT*α	Provided by Dr. Chernoff

Table 4- 2. The oligonucleotides used in Chapter 4.

Name	**Sequence**
NAC-F	5'CTTAGGATCCAACAATGTCGGAGCAAGTAACTAATGTAGGTGGAGCGGTTGTAACTGGTGTGACTGCTGTTGCGCAAAA 3'
NAC-R	5'GACCTGCAGGCTGACCAGTAGCCGCTGCAATGGACCCTGCACCCTCAACTGTTTTTTGCGCAACAGCAGTCACACCAG 3'
NAColi-F	5'CTTAGGATCCATGGGCTCTGAACAAGTAACCAACGTTGGCGGCGCAGTAGTAACCGGCGTCACTGCTGTTGCACAGAAA 3'
NAColi-R	5'GACCTGCAGGCTGACCAGTCGCTGCTGCGATGCTGCCCGCACCTTCGACAGTTTTCTGTGCAACAGCAGTGACGCCGG 3'
Syn1-F	5'ATCCAACAATGTCTAACGTTGGTGGTGCTGTTGTTACTGGTGTTACTGCTGTTGCTCAACCTGCA 3'
Syn1-R	5'GTTGAGCAACAGCAGTAACACCAGTAACAACAGCACCACCAACGTTAGACATTGTTG 3'
Hsp-F	5' GGGGTCGACATGAACGACCAAACGCAATTT 3'
Hsp-R	5' GGGGTCGACTTTTAGATTATTCACAGC 3'
MATa	5' AAATAAACGTATGAGATCTA 3'
MATalpha	5' GCAGCACGGAATATGGGACT 3'
MATdistal	5' ATGTGAACCGCATGGGCAGT 3'

Table 4-3. Plasmid used in Chapter 4

Plasmid	**Characteristics**	**Reference**
pZErO-1	*E. coli* cloning vector, Zeocin^R	Invitrogen
pET21d	*E. coli* expression vector, C-terminal 6XHis tag	Novagen
pRS313	Yeast shuttle vector, *CEN, HIS3,* Amp^R	(Sikorski and Hieter, 1989)
pRS314	Yeast shuttle vector, *CEN, TRP1,* Amp^R	(Sikorski and Hieter, 1989)
pRS315	Yeast shuttle vector, *CEN, LEU2,* Amp^R	(Sikorski and Hieter, 1989)
pRS316	Yeast shuttle vector, *CEN, URA3,* Amp^R	(Sikorski and Hieter, 1989)
pZEro-1+NM	pZErO-1, *sup35-NM* pZErO-1←Sup35-NM PCR (*Bam*H I/ *Xho* I)	this study (Qi, 2007)
p106. SYN1	pZErO-1, *sup35-NM(Δ3-39::syn1)* pZErO-1+NM←SYN1 (*Bam*H I/ *Pst* I)	(Qi, 2007)
p106. NAC	pZErO-1, *sup35-NM(Δ3-39::nac)* pZErO-1+NM←NAC (*Bam*H I/ *Pst* I)	this study
p106. NAColi	pZErO-1, *sup35-NM(Δ3-39::nacoli)* pZErO-1+NM←NAColi (*Bam*H I/ *Pst* I)	this study
p316-Sp-Sup35	pRS316, P_{SUP35}, *SUP35*	(DePace et al., 1998)

Plasmid	**Characteristics**	**Reference**
p149.WT	pRS313, P_{SUP35},*sup35-N-GFP-MC*	this study
p149.SYN1	pRS313, P_{SUP35},*sup35-N(Δ3-39::syn1)-GFP-MC* p149.WT← p106. SYN1 (*Bam*H I/ *PflM* I)	this study
p149.NAC	pRS313, P_{SUP35},*sup35-N(Δ3-39::nac)-GFP-MC* p149.WT← p106. NAC(*Bam*H I/ *PflM* I)	this study
pET21d.NAColi	pET21d, *sup35-NM* (*(Δ3-39::nacoli)-(His)$_6$* pET21d←NAColi PCR *(Nco* I/*Xho* I*)*	this study
pET21a.Sup35NM	pET21a, *sup35-NM-(His)$_6$*	(Qi, 2007)
pET21a. *ΔNQ* -M	pET21a, *sup35(ΔNQ)M-(His)$_6$*	(Qi, 2007)
p108. WT	pRS426, P_{CUP1}, *sup35NM-GFP*	this study
p108.NAC	pRS426, P_{CUP1},*sup35-N(Δ3-39::nac)-M-GFP* p108.WT← p106. NAC(*Bam*H I/ *PflM* I)	this study
pJet114	pRS426, P_{CUP1}, *HSP104* p426-P_{CUP1}-Sup35NM-GFP ← Hsp104 PCR (*Bam*H I/*Sac* I)	(Qi, 2007)
p314.NAC	pRS314, P_{SUP35}, *sup35N(Δ3-39::nac)-GFP-MC* pRS314 ← p149.NAC (*Xho* I/*Sac* I)	this study
p315.NAC	pRS315, P_{SUP35}, *sup35N(Δ3-39::nac)-GFP-MC* pRS315 ← p149.NAC (*Xho* I/*Sac* I)	this study

Unless specified, *E. coli* strain TOP10F' was used as the bacterial host to carry out the molecular cloning. *E. coli* strain BL21 [DE3] was used for the expression of the recombinant proteins.

Gene replacement by plasmids shuffling

A plasmid shuffling procedure was used to exchange SUP35 alleles in the *sup35::KanMX* mutation strains, including OT56k, OT60k, GT17k, GT388k, GT650 and GT810**.** For example, recombinant p313-NAC-sup35 genes were transformed into GT17K harboring p316-Sp-Sup35. Transformants were selected on SD-His medium and then plated to SD-His medium containing 5-fluoroorotic acid (5-FOA) to counter-select against cells that retained the URA3-based plasmid p316-Sp-Sup35. Single colonies were then picked from the 5-FOA-containing medium and further characterized.

Hsp104 gene knockout construction

The Hsp104 gene knockout was conducted by transforming *hsp104::URA3*, amplified from the genomic DNA of *S. cerevisiae* strain 74D-694-Δsup35-Δhsp104, into the desired prion strains and selecting the knockout strains on minimal medium lacking uracil. First, the gene was amplified using primers Hsp-F and Hsp-R, and purified on an agarose gel. The linear DNA was then transformed into the competent cells of the desired yeast strains. The transformants were incubated in rich medium for 2 h or overnight to allow the DNA recombination, and then spread on minimal medium without uracil for knockout selection. Colonies of cells with *hsp104::URA3* integration appeared in about 3 days, and they were further tested for hsp104 deletion. PCR amplification of the chromosome containing wild type HSP104 generates a ~3-kb product, whereas that of

hsp104::URA3 generates a ~4-kb product. In addition, the PCR product of *hsp104::URA3*, but not of hsp104, can be cut into two fragments by *Nco* I.

Prion assays *in vivo*

The suppression of *ade1-14* gene was monitored by color assay on YPD medium or growth assay on SD-Ade as described previously(Chernoff et al., 2002b). The ability to cure the prion phenotype was assayed by passage of the cells on YPD medium containing 5 mM guanidinium chloride (GuHCl) for three successive rounds of solid-phase culture.

Hsp104 protein overexpression was achieved by the addition of 50 μM $CuSO_4$, which induced the P_{CUP1}:HSP104 expression in plasmid pJET114 in [PRION+] cells. Cells were incubated in the induction medium for 1 day and then subjected to the color assay on 1/4YPD medium and growth assay on SD-Ade solid medium. The Hsp104 knockout was introduced into [$PRION^+$] stains as mentioned above and then all colonies from SD-Ura plate were streaked on YPD and SD-Ade for prion assays.

For prion induction experiments, yeast strains harboring plasmid P108.NAC fusion protein NAC-M-GFP were incubated in the minimal medium SD-Ura-His containing 50 μM $CuSO_4$ for 1 day and then spread on 1/4YPD for the color assay. The culture was diluted such that around 200-300 cells were spread on each plate for the convenience in counting the colonies. The induction rate was calculated by dividing the number of the white colonies by that of the total colonies. Some white colonies were randomly picked and tested for their GuHCl curability.

Differential centrifugation analysis of yeast extract

Total cellular proteins were extracted from yeast as previously described (Chernoff et al., 2002a). Yeast cells were grown to mid-log phase (~2 OD_{600}) in YPD medium at 30 °C with constant shaking. After cooling to 0 °C for 15 min, cells were harvested at 2000 x *g* for 5 min at 4 °C. The pellet was washed and re-suspended with yeast lysis buffer (50 mM Tris-HCl, pH7.5, 5 mM $MgCl_2$, 10 mM KCl, 0.1 mM EDTA, 1mM DTT) at a concentration of ~3 x 10^6 cells/µl, and the protease inhibitor cocktail for yeast (Sigma) was added at a concentration suggested by the supplier. An equal volume of 425- to 600-µm acid-washed glass beads was added to the cell re-suspension, and then cells were lysed by vigorous agitation. After the lysate was centrifuged at 2000 x *g* for 10 min at 4 °C to pellet unbroken cells and cellular debris, the supernatant was carefully transferred to a new tube as the total protein sample. The protein concentration was determined using the Bio-Rad protein assay kit (Bio-Rad) with BSA as a standard. Subsequently, protein was centrifuged at 16000 x *g* for 10 minutes at 4 °C to separate the soluble and pellet portions of the whole-cell protein obtained at 2000 x *g*. The soluble portion was removed to a new tube, and the pellet was re-suspended in an equal volume of yeast lysis buffer. Proteins of different portions were prepared with the addition of SDS loading buffer and loaded on SDS-PAGE gel for the separation by electrophoresis. The target proteins were visualized by Western blot with the anti-Sup35 antibody (santa cruz biotechnology, inc.) or the anti-GFP antibody (BD Bioscience).

Cytoduction, mating, sporulation and tetrad dissection

S. cerevisiae strain GT388k [p149.NAC], the recipient for cytoduction, was constructed through the plasmid shuffling between p316SpSup35 and p149.NAC using the procedure described above. For cytoduction experiments, *S. cerevisiae* strain GT17k

was employed as the donor strain in which different epigenetic variants of the NAC phenotype has been established in the p149.NAC plasmid background. *S. cerevisiae strains* OT56k [pJET149.WT] and OT60k [pJET149.WT] were mated to the recipient strain described above, which is karyogamy-defective *kar1* (Conde and Fink, 1976). The recipient lacks functional mitochondrial DNA and bears the recessive cyh^R mutation. After 1 day mating on YPD medium, cell mixtures were replica plated to synthetic medium containing glycerol and ethanol as the sole carbon sources and cycloheximide (5 mg/L) to select for cytoductants receiving the mitochondria from the donor strain. Cells were then characterized on 1/4YPD medium plates.

GT650 (*MAT*a *ade1-14 his3 leu2 lys2 ura3 trp1 sup35::HIS3*) [p315.NAC] and GT810 *(MATα ade1-14 his3 leu2 lys2 ura3 trp1 sup35::HIS3*) [p314.NAC] cells were mixed on YPD medium for the mating experiment. Following the overnight growth on YPD, cells were spun down and washed in 0.9% NaCl twice, after that, the cells were plated to synthetic medium lacking leucine and tryptophan (SD-Leu-Trp) to select the cells that contained both plasmids of different markers, which possibly formed as a result of mating. The cells were verified as diploids by their ability to sporulate under defined conditions. Additionally, genomic DNA was extracted and used as the template in subsequent PCR reactions that were designed to identify the mating type of cells. The forward primers MATa and MATalpha bind specifically to the genomic DNA from cells that are *MAT***a** and *MATα*, respectively, and the reverse primer MATdistal binds to the genomic DNA from both mating types. Therefore, the mating types can be reflected by the appearance of amplification products using different primers, and cells were verified as **a**/*α* diploids when products from both reactions were present (Aylon et al., 2004).

Sporulation was induced within the diploid cells generated from the mating experiments following incubation in the sporulation medium for 3 to 5 days. Tetrads were observed and dissected on YPD medium under a dissecting microscope. Following tetrad dissection, viable spores were grown on YPD for the color assay, and the genomic DNA was extracted and used in mating-type PCR to determine their mating types. Once the spores were grown on YPD plate, they were suspended in 0.9% NaCl and sequentially dropped on –Ade, -Trp and –Leu plates to check the phenotype.

Fluorescence microscopy

GFP images were scanned using a Zeiss LSM510 UV confocal laser scanning microscope (Carl Zeiss Inc., New York, NY) at wavelength 488 nm, and image analysis was conducted using a Zeiss LSM image browser (Carl Zeiss). The culture was incubated to stationary phase and subject to the observation under the fluorescence microscopy. Culture in minimal medium was spotted on slides directly.

Fluorescence recovery after photobleaching

A two-well chamber slide (Lab-Tek) was pretreated with 2 mg/mL solution of concanavalin A (Sigma-Aldrich) for 10 min. The chamber was dried by airflow and washing with distilled water and then re-dried sequentially. Cells were prepared in synthetic medium and diluted to OD_{600} = 0.1. An aliquot of cells is then added to the chamber and after 10 min, and then the unbound cells are removed by gentle washing with medium. A Zeiss LSM510 confocal microscope was use for image. Before photo bleach, the gain was adjusted to maximum of the linear range and a defined region was repeatedly photo-bleached around 15 times at 100% power strength with the 488 nm line of argon laser, which is also used for scanning but the fluorescence recovery was monitor

with 2% laser power scanning. For each strain, cells were photobleached at a given area to 20-50% of original fluorescence intensity, and 8-12 cells were examined. After normalizing each bleach experiment data by setting the maximum fluorescence to 100% and minimum to 0%, the data were compiled and used to calculate average and standard deviations for each stain as described in reference (Wu et al., 2006).

The expression and characterization of NAC-substituted Sup35NM proteins

The Sup35NM (NQ-NR-M), NAC-NR-M and SYN1-NR-M were over-expressed in *E. coli* strain BL21 [DE3] upon the addition of the inducer IPTG to the final concentration of 0.8 mM when OD_{600} reached 0.8. Following 3 h of induction, cells were harvested and lysed in 8 M urea buffer (8 M urea, 20 mM Tris-HCl, pH 8.0). Proteins were purified first on Ni-NTA column (Qiagen Inc.), using 8 M urea buffer containing 20 mM imidazole for washing and 400 mM imidazole for elution. The protein solution eluted from the Ni-NTA column was loaded on DEAE Sephacel column (Pharmacia Biotech Inc.) for another round of purification. The column loaded with the proteins was washed with 8 M urea buffer and eluted with 8 M urea buffer containing 300 mM NaCl. Proteins were methanol precipitated (at least 5 volumes to the protein solution) and air dried before they were re-suspended in 6 M GuHCl buffer and shaken overnight to disaggregate preformed polymers. The purity of the proteins was verified by SDS-PAGE electrophoresis and MALDI mass spectrometry. Proteins were quantified either by monitoring the UV absorbance at 276 nm or by using the BCA protein assay kit (Pierce Biotechnology).

To initiate fiber formation, protein was diluted by at least 100 times to the Congo red binding buffer (CRBB: 5 mM potassium phosphate, pH 7.4, 150 mM NaCl) to a final

concentration of 0.5 μM. The reactions were generally set up in a volume of 1 mL in 2-ml microcentrifuge tubes, which were rotated end-to-end on a roller. The formation the fibers was monitored by the Congo red binding assay (Chernoff et al., 2002b). At different time points, 0.5 μM peptide and 10 μM Congo red were mixed at room temperature for 30 min, and then the solution was scanned between 400 and 600 nm on a UV-Vis spectrometer. The fiber exhibited a spectra shift in absorbance, with a new peak at 540 nm, compared to unpolymerized protein or Congo red alone. The amount of Congo red bound to amyloid fiber was calculated using equation (1):

$$\text{mole Congo red bound/liter solution} = (A_{540}/25295)-(A_{477}/46306) \quad (1)$$

where A_{540} and A_{477} refer to the absorbance at 540 and 477 nm, respectively. In the seeding experiments, preformed fibers were sonicated and added to the unpolymerized protein solution at a final ratio of 5%.

Fibers were negatively stained for electron microscopy (EM) analysis (Chernoff et al., 2002). To clear the undissolved crystals, 2% uranyl acetate solution was centrifuged at 16000 x g for 5 mins, and the supernatant was transferred into a new vial. Proteins were applied to a glow-discharged 400 mesh carbon-coated copper grid and then stained with 2% (w/v) aqueous uranyl acetate. Excess liquid was removed from the grid, and the grid was then allowed to dry in a vacuum desiccator. Samples were viewed in a Philips LS-410 transmission electron microscope (TEM), at an accelerating voltage of 80 kV, at the integrated microscopy and micro-analytical facility (IM&MF) at Emory University.

The SDS solubility analysis of NAC-M amyloid

The SDS buffer (2%, W/V) was used to test the solubility of NAC-NR-M fiber. To initiate NAC-NR-M assembly, purified protein was diluted to 1 μM in saline buffer (PH7.4) and rotated end-to-end on a roller. After 48 hours, the formed NAC-NR-M fiber was aliquot equally to SDS sample buffer. One sample is boiled in a water-bath for 10 min, whereas the other sample is incubated at room temperature. The samples were then separated on a 10% SDS-PAGE gel that is subsequently stained with Coomassie Brilliant Blue.

Results

Cells containing the SYN1-G-Sup35 *and* NAC-G-Sup35 fusion proteins both showed two phenotypes

In *S. cerevisiae* strain 74D-694, the genomic *SUP35* gene has been replaced by *SUP35* on plasmid p316-Sp-Sup35. The SYN1 and NAC fragments from α-syn were substituted for the NQ domain of Sup35 individually and fused to the *N*-terminus of the wild-type NR-M-C domains of Sup35p. For the visual assays, GFP was inserted between the *N*- and *M*-domains of Sup35p to construct the chimeric aggregation domain fusion protein (ADFP) SYN1-G-SUP35and NAC-G-SUP35 (Figure 4-2), which are mentioned as SYN1p and NACp for the remainder of our studies unless otherwise noted. It was reported that the insertion of GFP in the Sup35p, which associated the fluorescence morphology in cytoplasm directly with the prion phenotypes, did not perceptibly effect either prion formation or propagation (Song et al., 2005). Sequentially, the *syn1-g-sup35* or *nac-g-sup35* fusion gene was introduced to cells with counter selection of 5-FOA and confirmed by the growth on SD–Ura plate. Since the expression of chimeric *syn1-g-*

*sup35 and nac-g-sup35*were under the control of the native *S. cerevisiae sup35* promoter, proteins would be expressed constitutively at a moderate level that was similar to that of wild-type Sup35p. As a prion protein, Sup35p can adapt two conformations: a soluble, non-prion form, ([*psi*$^-$]) and an aggregated, prion form, ([*PSI*$^+$]). The non-prion state exhibits wild-type translation termination activity, while the prion state displays attenuated activity and a heightened level of non-sense suppression. If the corresponding NACp and SYN1p proteins are aggregated in the *S. cerevisiae* host strains, the prion phenotype can be monitored through assay of the *ade1-14* marker that contains a suppressible nonsense mutation (Chernoff et al., 1995b). Once stop codon in *ade1-14* is suppressed, functional Ade1p is produced, resulting white colonies on YPD and growth on adenine-free medium. In contrast, soluble ADFPs will produce premature truncated Ade1p, so that cells appear red on YPD and cannot survive on adenine-free medium.

Since SYN1-NR-M and NAC-NR-M fusions form amyloid fiber *in vitro* (Figure 4-3), it was expected that SYN1p and NACp might display two phenotypes in yeast strains corresponding to the prion and non-prion states of Sup35p. Transformants of *S. cerevisiae* strain (GT17K) with the *nac* or *syn1* constructs were analyzed directly on YPD media using the color assay for phenotype. The cells harboring NACp exhibited identical white or red color on YPD directly. However, the SYN1p showed higher propensity to aggregate because almost all cells with SYN1p were white or light pink on YPD plate, thus the isolation of red colonies is difficult and only the darker pink colonies were found. The presence of white/light-pink and red/ dark-pink colonies suggested that the NACp and SYN1p might be at the prion and non-prion phenotypes, which were designated as [*NAC*$^+$]/[*nac*$^-$] and [*SYN1*$^+$]/[*syn1*$^-$] respectively. The [*NAC*$^+$] phenotype appears relatively stable to spontaneous conversion with a frequency lower than one in 10^6.

Meanwhile, the conversion rate of [*nac*$^-$] cells to [*NAC*$^+$] is less than one in 10^5, which is similar to the spontaneous conversion of [*psi*$^-$] to [*PSI*$^+$]. However, unlike [*nac*$^-$] cells, [*syn1*$^-$] cells converted to [*SYN1*$^+$] at an uncommonly high frequency, which was estimated to be around 1 in 10. Accordingly, although [*syn1*$^-$] cells grew poorly on SD-Ade compared to the [*SYN1*$^+$] counterparts, they grew much better than [*nac*$^-$] cells, presumably as a consequence of conversion to the prion state. (Figure 4-9)

Since SYN1p and NACp display two phenotypes upon expression in *S. cerevisiae* cells, we hypothesized that these two states arose from differential self-association of the *N*-terminal domain, in analogy to the parent Sup35p system. Further biochemical and genetic analysis were performed on the suppositious prion and non-prion states of *S. cerevisiae* strains. First, we checked the aggregation states of ADFPs *in vivo*. Second, the influence of the Hsp104 chaperone on the *de novo* prion states was assessed. Third, we investigated the inducibility of the prion state from the non-prion state through overproduction of aggregation domains. Finally, we analyzed the non-Mendelian heritance of the synthetic yeast prions.

Aggregation state of NACp and SYN1p are different

Fusion proteins with GFP is widely used to monitor Sup35p aggregation, because the fluorescence of fusion proteins is centralized in [*PSI*$^+$] cells but diffuse in [*psi*$^-$] cells, which is visualized by confocal fluorescent microscopy. (Borchsenius et al., 2001; DePace et al., 1998; Song et al., 2005; Zhou et al., 2001) In stationary phase culture, [*NAC*$^+$] cells exhibited visible foci while the [*nac*$^-$] cells displayed diffuse fluorescence. (Figure 4-8) The fluorescence data for the NACp system is similar to the control system derived from Sup35p construct NGMC. In [*PSI*$^+$] cells, punctuate foci appeared but [*psi*$^-$]

cells showed diffused cytoplasmic fluorescence. (Song et al., 2005) Since the SYN1 peptide showed a higher tendency for aggregation, based on its high conversion rate from [*syn*$^-$] to [*SYN*$^+$], SYN1p should be aggregated in the cells in the stationary state as well. Unexpectedly, cells containing the Syn1p displayed diffuse fluorescence in [*SYN1*$^+$] cells. (Figure 4-8)

The results from differential centrifugation analysis were consistent with those from the GFP visual assay. Large prion aggregates sedimented from the yeast cell lysates, while smaller aggregates and unassociated protein remain in the supernatant fraction. Equal volumes of cell lysate can be separated by SDS-PAGE and analyzed using immuno-blotting. It was found that the ratio of the protein in the pellet portion to that in the supernatant portion was higher in [*NAC*$^+$] cells than in [*nac*$^-$] cells. After treating [*NAC*$^+$] cells with 5 mM GuHCl, most of the NACp appeared in the supernatant, as was observed similarly in the [*nac*$^-$] cells. (Figure 4-6) Unexpectedly, in both [*SYN1*$^+$] and [*syn1*$^-$] cells, the SYN1p appeared almost completely in the supernatant fraction (Figure 4-6).

The method of fluorescence recovery after photo-bleaching (FRAP) is useful for determining the mobility of fluorescently tagged proteins, which can be related to their relative size through the rate of diffusion-based recovery of the fluorescent signal (Lippincott-Schwartz et al., 2003; Lippincott-Schwartz et al., 2001; Reits and Neefjes, 2001; Wu et al., 2006). This method has been employed to estimate the relative size of prion aggregates *in vivo*, and to distinguish between prion and non-prion states on the basis of protein self-association. In the stationary phase, both [*NAC*$^+$] and [*PSI*$^+$] cells displayed visible fluorescence foci in cells. As shown in Figure 4-12, once the Sup35p granule was photo-bleached, the fluorescence of the irradiated region recovered

approximately 10%, which suggested that the Sup35p granule is almost immobilized. Whereas, the recovery rate of the SYN1p fluorescence in [*SYN1*$^{+}$] cells was significantly faster than that of the NACp fluorescence in [*NAC*$^{+}$] cells. After 20 seconds, the fraction fluorescence recovery of [*NAC*$^{+}$] continued to recover, while that of [*SYN1*$^{+}$] was in balance within 5 seconds, similar to the response of [*nac*$^{-}$] cells. These results were consistent with the fluorescent microscopy image and suggested that the sizes of the SYN1p aggregates are smaller than the NACp aggregates. In the mid-log phase of cells, previous study showed that both [*PRION*$^{+}$] and [*prion*$^{-}$] cells exhibited diffuse fluorescence (Song et al., 2005), but the FRAP rate of NACp aggregates are still slower than that of SYN1p aggregates with increasing recovery time of 5 seconds. (Figure 4-13) Even though the FRAP result of [*SYN1*$^{+}$] cells closely corresponds to that observed for non-prion [*nac*$^{-}$] state, phenotypic tests performed on [*SYN1*$^{+}$] cells, such as growth on SD–Ade plate, are consistent with protein aggregation. However, the similarity between monomeric NACp and oligomeric SYN1p might due to other factors, including the association between monomeric Sup35p variant and the ribosome or other cytosolic proteins (Wu et al., 2005).

[NAC^{+}] was reversibly curable

Molecular chaperones, such as ring-shaped Hsp104p complex, appears able to mediate ATP-dependent fragmentation of "soft" aggregates of up to 600 kDa; the action of additional chaperones of the Hsp70 family allows the components of such aggregates, once released, to refold and attain the native conformation (Glover and Lindquist, 1998). When chaperones are not sufficient to disaggregate large protein assemblies, diseases may occur. Even though the mechanism for the prion propagation in mammals with

chaperones has not yet been clear, previous experiments showed that in yeast, prions required intermediate levels of Hsp104p to disaggregate mis-folded aggregates and facilitate prion propagation (Chernoff et al., 1995b; Moriyama et al., 2000; Wu et al., 2005). Guanidine hydrochloride (GuHCl) is the most frequently used prion-curing agent in yeast because it acts by inhibiting Hsp104 activity. By passing [*NAC*$^{+}$] cells on YPD containing 5 mM GuHCl 3-5 times, almost all white cells became red on subsequent culture in YPD medium. Although previous restoration experiment of [*SYN1*$^{+}$] suggested that the white phenotype was caused by aggregation of SYN1p, apparently no cure was observed on [*SYN1*$^{+}$] after passing 5 consecutive times on YPD containing 5 mM GuHCl.

Deletion of *hsp104* gene has also been employed as a non-reversible process to cure yeast prions. An *hsp104* gene knockout was introduced in *S. cerevisiae* strains displaying the prion-like phenotypes [*NAC*$^{+}$] and [*SYN1*$^{+}$]. As expected, upon transformation with the *hsp104* knockout, the [*NAC*+] strain displayed a phenotype consistent with loss of the prion, i.e., red colony formation on YPD medium and absence of growth on Ade- medium. (Figure4-10) In contrast, [*SYN1*$^{+}$] were maintained its prion-like phenotype even after *hsp104* deletion. These results were similar to those obtained from the GuHCl treatment.

Overproduction of Hsp104p cures yeast prion [*PSI*$^{+}$] but not other yeast prion. After the P_{CUP1}-promoted *hsp104* gene was induced in [*SYN1*$^{+}$] and [*NAC*$^{+}$] cells for 1 day, cells were streaked on YPD plate for color assay; however, neither [*SYN1*$^{+}$] nor [*NAC*$^{+}$] cells were converted to the corresponding non-prion state by the over-expression of Hsp104p.

Overproduction of the NAC-NR-M-GFP induced [NAC+] phenotype

It's known that overproduction of the protein should increase the frequency of the prion conversion, presumably through mass action effects that shift the process toward the aggregate. The prions [*PSI*$^{+}$] and [*URE3*$^{+}$] can be induced by increasing the intra-cellular concentration of the Sup35p or Ure2p, respectively. Moreover, overproduction the prion-forming domain is a more efficient method to induce the conformational conversion than over-expression of the full-length protein (Chernoff et al., 2002b; Derkatch et al., 1996b; Masison and Wickner, 1995). Thus, prion induction through over-expression has been employed as one of the criteria for determination of the presence of a prion phenomenon. In [*NAC*$^{+}$], the aggregation prone domain is NAC fragment, so that, in order to test whether [*NAC*$^{+}$] can fulfill this criterion, fusion protein NAC-M-GFP was constructed and cloned under inducible promoter *Pcup1*. When the fusion protein NAC-M-GFP was induced in [*nac*$^{-}$] cells and over-expressed, we found that the conversion to prion-like phenotype increased significantly to more than 50 % of the cells. After passing the white colonies on GuHCl, it was confirmed that the white phenotype was caused by formation of [*NAC*$^{+}$] prion.

[*NAC*$^{+}$] was induced by cytoduction

Cytoduction is a convenient method for monitoring non-Mendelian inheritance of yeast prions. Cytoduction occurs at a low frequency during mating through mixing of the cytoplasm of two cells without fusion of their nuclei. The daughter cells have the nuclei from one parent but contain cytoplasm from both parents. The donor strain, usually the strain to be tested for the presence of a prion, delivers its cytoplasmic materials including prion particles to the recipient strain, which contains the protein in a non-prion state (Chernoff et al., 2002b). Prions are usually transferred with high efficiency so essentially

all cytoductants should contain prion, which should result in conversion of the cytoductant strain to the corresponding prion.

Yeast strain GT388k was used to construct the recipient strain. The strain possesses a *kar1* mutation, which inhibits karyogamy, i.e., the fusion of nuclei between the two cells, and significantly increases the frequency of cytoduction. The recipient strain displays the *rho-* phenotype and cannot grow on minimal media containing glycerol and ethanol as the only carbon sources. Transfer of mitochondria from the *rho+* donor occurs during cytoplasmic mixing, which allows the cytoductants to grow on the selective media. Furthermore, the recipient strain has a recessive cyh_R marker that allows the selection against the donor and rarely occurring diploid cells.

Wild-type Sup35p was replaced with NACp to make the recipient *S. cerevisiae* strain [*GT388k-nac*], which were treated with GuHCl to generate the [nac^-] phenotype. The cytoduction was conducted using various donor strains with the non-prion recipient strain. As shown in Table 4-15, thirty-five of thirty-six cytoductants from the [NAC^+] donor grew on SD–Ade plates and exhibited [NAC^+] phenotype, while the ones from the [nac^-] donor retained the non-prion phenotype. This indicates that [NAC^+] is cytoplasmically heritable and one may plausibly suggest that the protein behaves as a prion *in vivo*. An interesting event we observed is the similar cytoduction effect caused by [$SYN1^+$], in which 21 out of 22 screened colonies exhibited Ade+ phenotypes. This observation suggests that the NACp can be efficiently incorporated into the SYN1p aggregates through a templating effect. In a control experiment, an *S. cerevisiae* strain with [PSI^+] phenotype was used as a donor strain. Although the [PSI^+] donor caused a less effective conversion of the [*nac*-] state to the [*NAC*+] state, the nucleation effect caused by [PSI^+] is observably stronger than those from [*nac*-] and [*psi*-] strains.

Previous research has shown that two nucleation regions are present in Sup35p, the major one within the NQ domain and the other on existing in NR domain. (Tessier and Lindquist, 2007) Generally, NACp mainly uses NAC domain as nucleation region, but the presence of nucleation region in the NR domain of Sup35 may provide a secondary cluster that can nucleate the soluble NACp and produce the moderate cytoduction effect.

To further test the hypothesis, we examined the self-assembly kinetic of the NAC-NR-M fiber formation that adopted an amyloid conformation by using a well-characterized *in vitro* method, based on Congo red binding assay. (Glover et al., 1997; King et al., 1997) The denatured NAC-NR-M protein in 8 M urea solution was diluted 100-fold into a final concentration of 1.0 μM and the time course of self-assembly was monitored using this assay. A lag phase of 100 minutes was observed for initiation of amyloid assembly. Adding 5% sonicated NAC-NR-M fiber as a seed into solution of un-polymerized protein solution, the lag time was shortened to 30 minutes, which suggested that the preformed fibers could acts as a template to accelerate nucleation. A similar degree of acceleration was observed when the same amount fiber solution of NAC-PEG_{3000}, a protein without the NR domain, was added to the un-polymerized NAC-NR-M solution. (Figure 4-16) Thus, the similar seeding effects of the NAC-NR-M fiber and the NAC-PEG_{3000} revealed the fact that NAC-NR-M monomer was primarily incorporated into the template nucleus formed with the NAC domain. When the pre-formed SYN1-NR-M fibers were used as a template, the seeding effect was obvious stronger than the NAC-NR-M fiber with a negligible lag phase as the Congo red binding reached the highest value at 150 minutes. All the seeding fibers we used were subjected to sonication before adding to the NAC-NR-M monomer solution. To eliminate the possibility that the seeding effect of SYN1-NR-M fibers might primarily come from the NR domain, the

SYN1-PEG3000 fibers were used as seeding nucleus and similar seeding time course was observed, which indicated that the SYN1 domain is the primary seeding template. (Figure 4-16) After sonication, the electron microscopy analysis demonstrated that the SYN1-NR-M fibers exhibited a shorter and more fragmented morphology compared to the elongated NAC-NR-M fibers. (Figure 4-17) Thus the shorter SYN1-NR-M fibers may be accountable for acceleration of the self-assembly process through presentation of a complementary interface that can template fiber formation. These results are in agreement with the cytoduction experiments in which [$SYN1^{+}$] induce the conversion from [nac^{-}] to [NAC^{+}]. Interestingly, the polymerization of NAC-NR-M monomer initiated by NQ-NR-M fiber or NR-M fiber showed an even more accelerated effect on Congo red binding, without a detectable lag phase. By contrast, with the NQ domain only, the NQ-PEG3000 fibers display no detectable influence on the rate of the polymerization of NAC-NR-M monomer. (Figure 4-16) Consequently, we suggested that the NR domain may also provide a nucleus to seed the polymerization of NAC-NR-M monomer and the acceleration effect on the polymerization of NAC-NR-M may be due to the shorter fibers derived from self-assembly of the NQ-NR-M protein. (Figure 4-17) These kinetic experiments also explained that the observation that about 25% of the [nac^{-}] cells were converted to [NAC^{+}] upon cytoduction the recipient strain with the [PSI^{+}] donor strain.

[NAC+] represents an extra-chromosomal genetic element that can be confirmed with mating and sporulation experiment

The inheritance pattern of [NAC^{+}] was also assessed using mating and sporulation experiments. The gene encoding NACp was introduced into the *S. cerevisiae* mating strains GT650 [*MATa*] and GT810 [*MATα*] as plasmid bound copies that were

maintained using *LEU2* and *TRP1* nutritional selection, respectively. The haploid [NAC^{+}] and [nac^{-}] cells were separated from each mating type and then subjected to the cross mating in YPD media. The diploid cells, which were selected from double dropout minimum medium SD-Trp-Leu were sporulated for 3-5 days. We examined 10 dissected tetrads and all spores exhibited white color on YPD plate and growth on SD-Ade plates, showing 4:0 non-Mendelian segregation behaviors. (Figure 4-18)

Discussion

A number of amyloidogenic proteins can facilitate the formation of β-sheet-rich misfolded protein aggregates through templating the conformational conversion of the soluble, unassociated form of the protein. This process is often associated with the occurrence of neurodegenerative diseases in humans and, in the case of the prion phenotype, may be transmissible and heritable. In *S. cerevisiae* yeast, the amyloidogenic features of prion proteins are associated with N/Q-rich sequences that can undergo conformational conversion from a soluble form of the protein to an extended, β-sheet assembly. The assemblies can template subsequent conformational conversion of soluble protein, which provides a mechanism for propagation and transmission of the aggregates (Maddelein and Wickner, 1999; Santoso et al., 2000; Serio et al., 1999). Previous research illuminated the switchable nature of the NQ domain of Sup35p that underlies the two phenotypic states. Substitution of the native NQ domain with other N/Q-rich regions afforded chimaeric proteins that could support similar conformational behavior, which resulted in the formation of novel, artificial prions (Nemecek et al., 2009; Santoso et al., 2000). However, the presence of an N/Q-rich sequence is not a necessary for prion

behavior as it is absent from the first identified mammalian prion, PrP^{Sc}, indicating that non-N/Q-rich domain structure component can drive the formation of infectious aggregates as well (Safar and Prusiner, 1998). It remains uncertain as to whether other non-N/Q-rich, aggregation-prone sequences could substitute for the NQ domain in Sup35p to generate novel infectious prions. In this study, we used the non-NQ-rich aggregation-prone peptides from α-synuclein, NAC and SYN1, to replace the $(NQ)_{1-39}$ domain of Sup35p for generation of novel chimeric yeast prions. *S. cerevisiae* strains harboring NACp or SYN1p fusion protein can exhibit prion-like phenotypes on the basis of conventional assays. While the *in vitro* and *in vivo* data indicate that self-association of NACp and SYN1p occurs, only the former protein fits the criteria for prion behavior. Since prions are infectious proteins and are distinguished from mutations in nucleic acid replicons, the [*PRION*+] phenotype should be reversibly curable. Previous studies have shown that most yeast prions were cured via manipulation of the functional level of the Hsp104 chaperone (Chernoff et al., 1995b; Paushkin et al., 1996). Similarly, the [NAC^+] phenotype is alleviated with inhibition of the chaperone Hsp104 by 5 mM guanidine hydrochloride or deletion of the copy of genomic *hsp104*. In addition to curability, prion phenotypes should also be inducible under appropriate conditions. The most common approach to induce *de novo* prion transformation involves transient over-expression of the prion domain. Our results indicate that increasing the intracellular NACp concentration increases the frequency of *de novo* generation of the [NAC^+] state, which is consistent with the presence of a novel [NAC^+] prion. Moreover, both cytoduction and mating/sporulation experiments indicated that [NAC^+] is a dominant, epigenetic element that is stably heritable. These results indicated that [NAC^+] is a novel chimeric prion in yeast.

In contrast, the SYN1p chimaera displays significant differences in behavior even though its sequence is contained within the NAC peptide. *S. cerevisiae* strains containing the [*SYN1*$^{+}$] factor have been isolated that display cellular characteristics similar to those of the [*NAC*$^{+}$] prion, i.e., white colonies that arise from red precursors and growth on SD-Ade media, but the formation of these species is apparently irreversible. Inhibition, deletion or overexpression of Hsp104, a chaperone specifically required for the maintenance of all known yeast prions, cannot cure the suppositious [*SYN1*$^{+}$] prion. Thus, [*SYN1*$^{+}$] does not qualify as a prion based on the criterion of curability.

Many previous studies have suggested that the prion inheritance derives from a mechanism of self-propagation in which new amyloid particles grow from the existing β-sheet templates by fragmentation of larger aggregates. (Kushnirov et al., 2007; Osherovich et al., 2004) Similarly, for NACp, when the aggregates are produced, Hsp104 should mediate the breaking of polymer into smaller oligomeric 'seeds'. Thus, the absence of functional Hsp104 impedes the disaggregation of the [*NAC*$^{+}$] prion, hindering its propagation through enhancing the formation of large aggregates that are not inherited by the cellular progeny. (Chernoff et al., 1995a; Ripaud et al., 2003; Wegrzyn et al., 2001) We found that increasing the expression level of Hsp104 did not alleviate the [*NAC*$^{+}$], arguing that Hsp104 cannot restore the NACp aggregates back to original soluble monomer. In the light of chaperone-mediated yeast prion propagation mode, the inheritance of [*SYN1*$^{+}$] phenotype may be explained by the SYN1p aggregation state. If the SYN1p generates oligomer seeds but not larger polymers, [*SYN1*$^{+}$] is inherited independently of Hsp104 and is not curable using common prion curing strategies that rely on manipulation of Hsp104 activity. The observed differences between the [NAC] and [SYN1] systems can also be detected in the level of *in vivo* self-association of the

GFP-tagged fusions. The FRAP results suggest that the cellular aggregates of NACp display lower mobility than those of SYN1p, which may be due to the different granule sizes, polymeric, in the case of NACp, and oligomeric, in the case of SYN1p. Furthermore, visible fluorescent foci were observed in [NAC^+] cells, while more diffuse fluorescence was observed in [SYN^+] cells, which suggested that [NAC^+] cells supported the formation of larger, aggregated granules. The observed differences *in vivo* aggregation behavior between the two prion candidates presumably arises from the differences in amino acid sequence of the aggregation-prone peptide. Previous research on the *in vitro* assembly of the α-synuclein fragments reported that the size and the charge properties of residue E83 influenced aggregation propensity.(Chou and Fasman, 1974; Rivers et al., 2008; Waxman et al., 2010; Waxman et al., 2009) Due to the absence of E83 in the SYN1 peptide, the SYN1p may exhibits a stronger propensity toward aggregation, which may underlie the volatility of the [$syn1^-$] state with respect to the high rate of conversion to [$SYN1^+$]. A possible explanation is that the SYN1p proteins tend to aggregate *in vivo*, which results in the formation of numerous oligomeric seeds that cannot easily merge into larger prion polymers.

Using the structural context of the yeast prion Sup35, we have substituted the $NQ_{1\text{-}39}$ sequence within the prion domain of native Sup35p with the NAC peptide sequence of α-synuclein, a non-N/Q-rich aggregation prone sequence, and successfully constructed a novel chimeric yeast prion [NAC^+]. The [NAC^+] prion propagates stably and can be easily monitored using the conventional phenotypic assays for the [PSI^+] prion. Future mutagenesis study on NAC might be useful to further investigate the individual residues that are important for the aggregation. Moreover, the [NAC^+] yeast prion might be useful for screening drugs that can specifically inhibit α-synuclein

aggregation or for identifying the potential for transmission of the aggregation phenotype to full-length α-synuclein. Our findings illustrate that, in addition to the N/Q-rich tracts of the native Sup35p protein, non-N/Q-rich aggregation prone peptides can support prion formation in the appropriate protein architecture. Furthermore, the differences observed in aggregation behavior between the [NAC^{+}] and [$SYN1^{+}$] assemblies give useful clues in designing novel prion switches and controlling self-association *in vivo*. We hypothesize that other aggregation-prone sequences might be employed as switchable sequences in yeast, potentially including *de novo* designed peptides, and that these sequences in the appropriate cellular context may be exploited to control cellular biochemistry at the phenotypic level through modulation of self-association behavior of the prion domain under controllable conditions.

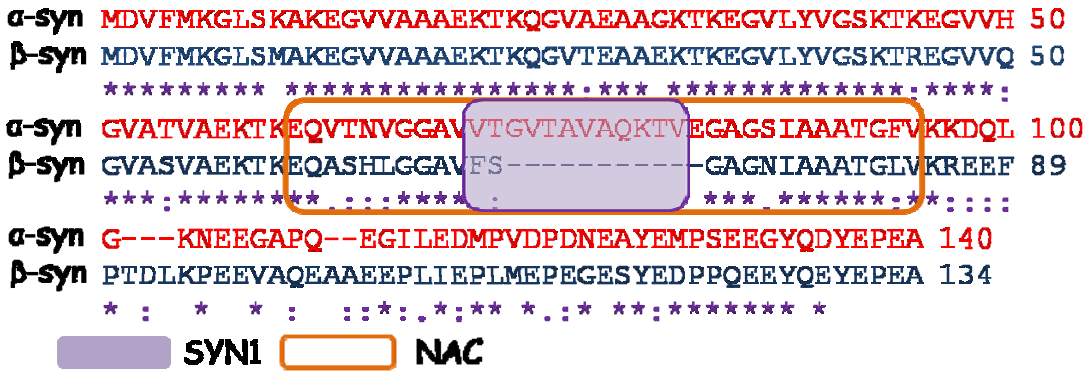


Figure 4-1. Comparison of the amino acid sequences of α-synuclein and β-synuclein. The SYN1 and NAC peptide are highlighted in the orange and purple rectangles, respectively.

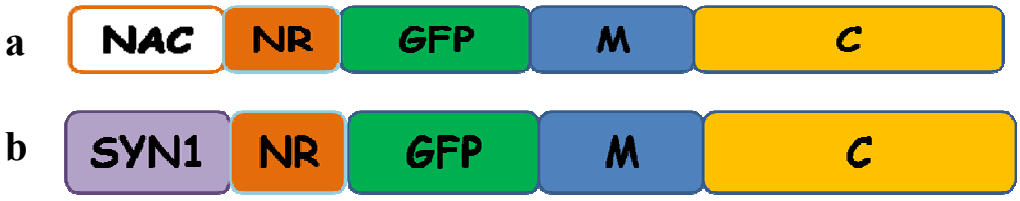


Figure 4-2. Schematic representations of the modified proteins NAC-G-Sup35 (a) and SYN1-G-Sup35 (b), respectively.

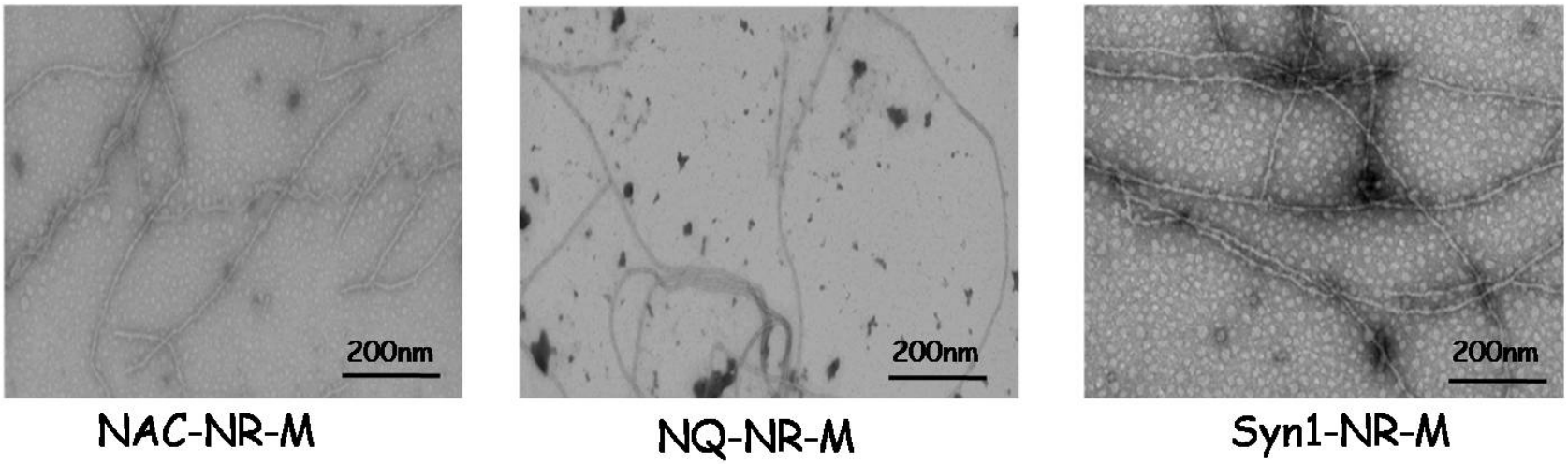


Figure 4-3. Electron microscopy image of fibers formed by NAC-NR-M, NQ-NR-M and Syn1-NR-M.

.

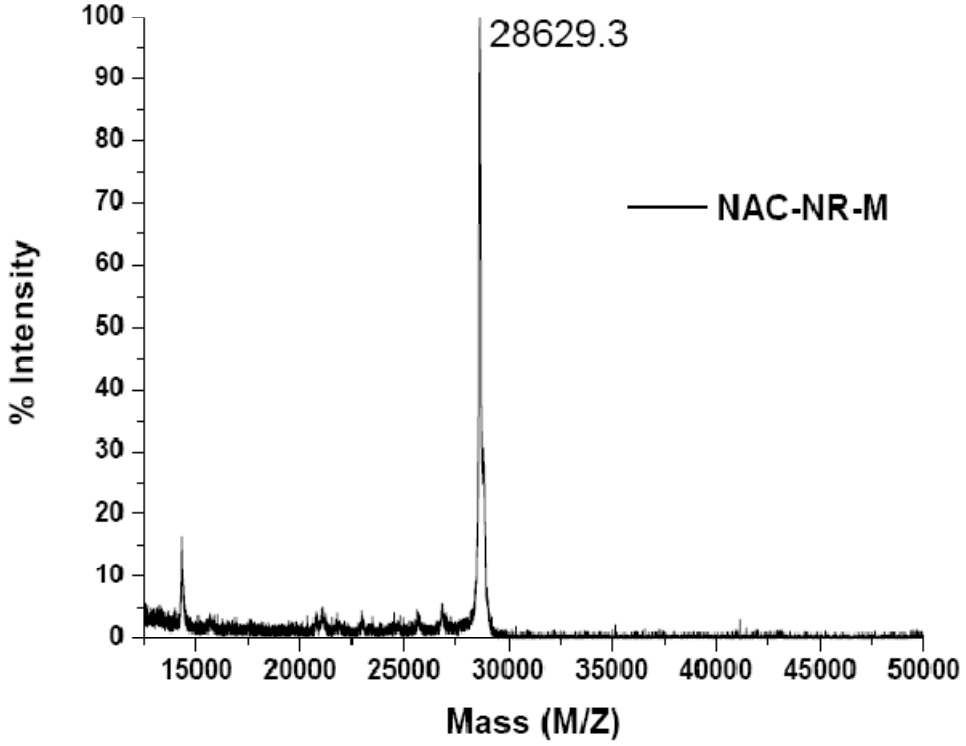


Figure 4-4 (a). MALDI-TOF mass spectrum of purified NAC-NR-M protein. Calculated molecular weight of NAC-NR-M is 28638.0.

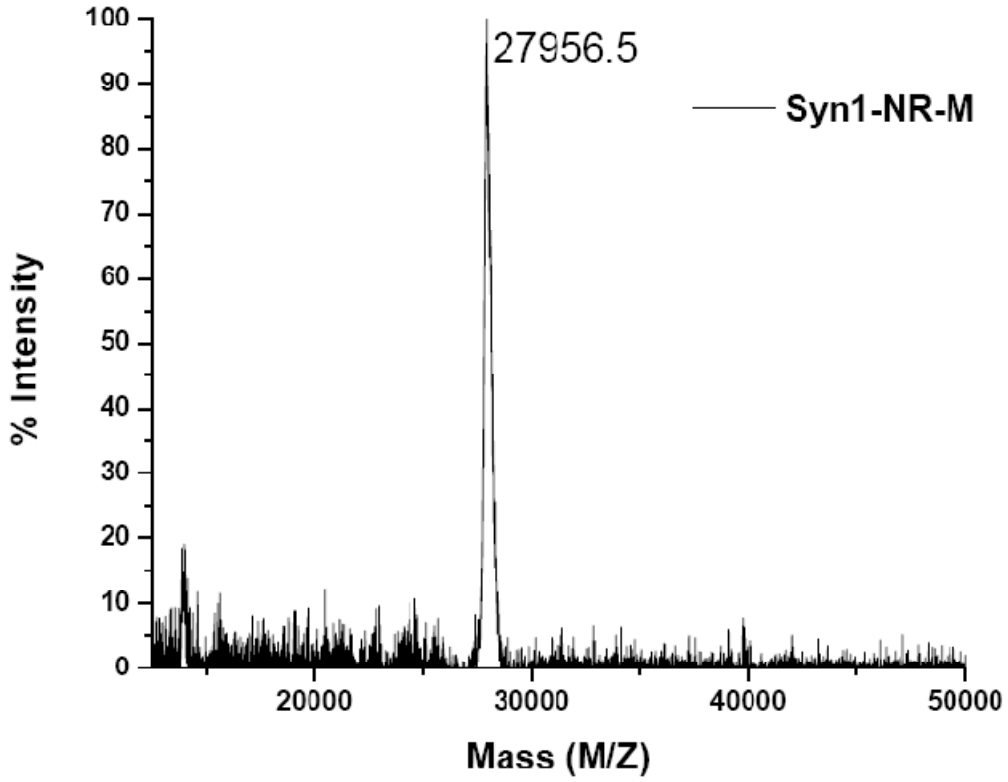


Figure 4- 4 (b). MALDI-TOF mass spectrum of purified Syn1-NR-M protein. Calculated molecular weight of Syn1-NR-M is 28214.0.

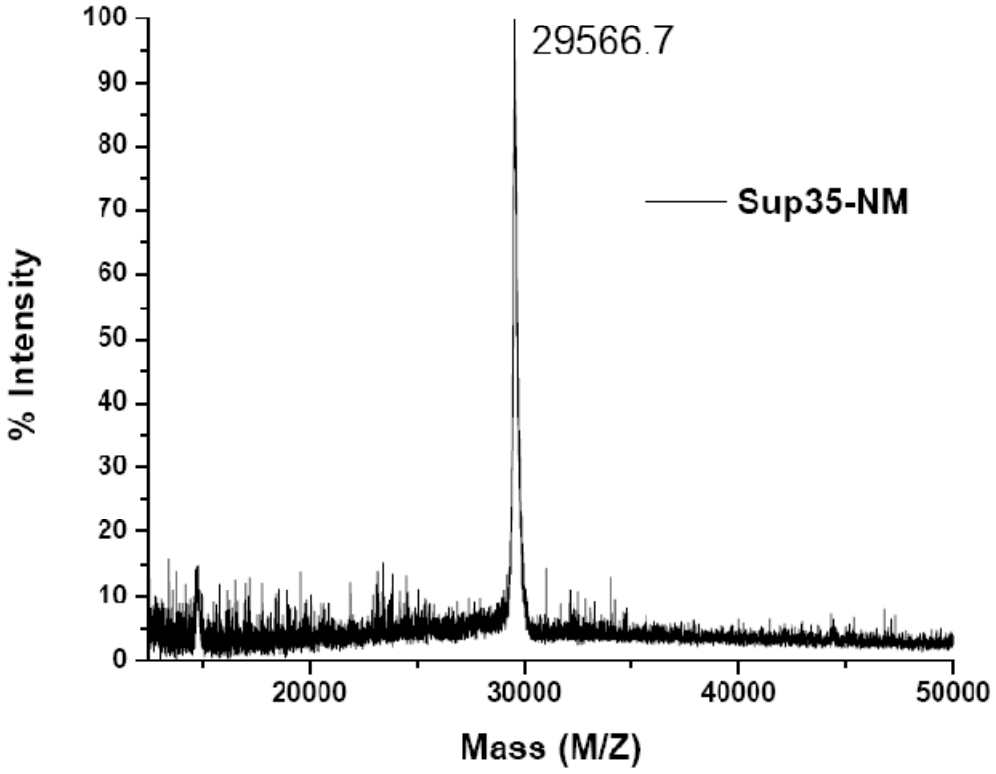
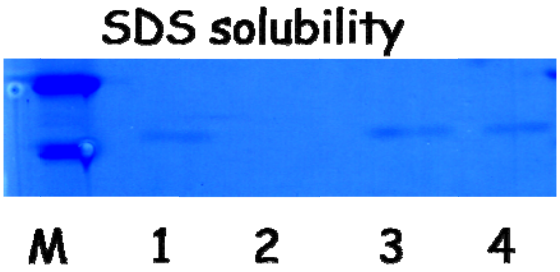


Figure 4- 4 (c). MALDI-TOF mass spectrum of purified Sup35-NM (NQ-NR-M) protein. Calculated molecular weight of Sup35-NM is 30012.



1. Fibrillar NAC-NR-M, 100^{o} C, 20mins
2. Fibrillar NAC-NR-M, RT, 20mins
3. Monomer NAC-NR-M, 100^{o} C, 20mins
4. Monomer NAC-NR-M, RT, 20mins

Figure 4-5. SDS solubility analysis of His-tagged NAC-NR-M fibers.

The amyloid fiber NAC-NR-M (Lane 1 and 2) or un-polymerized NAC-NR-M (Lane 3 and 4) was incubated in 2% (w/v) SDS with (Lane 1 and3) or without (Lane 2 and 4) boiling are shown.

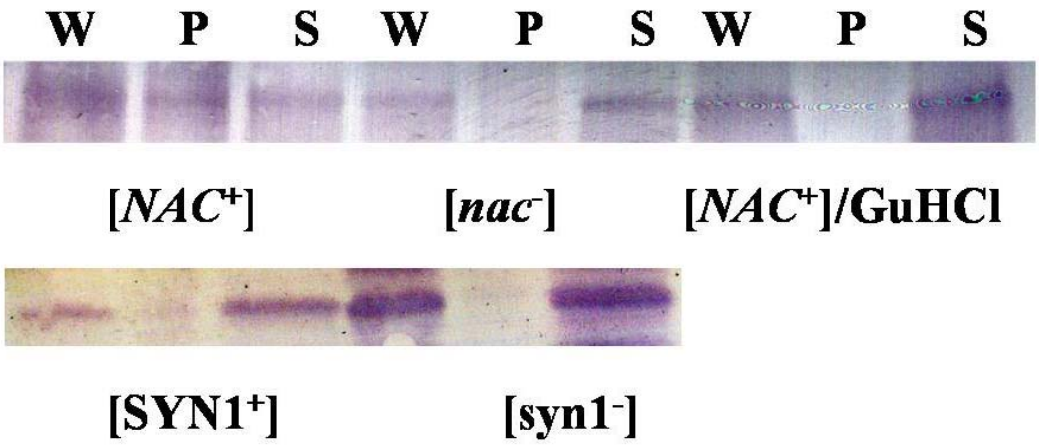


Figure 4-6. Protein sedimentation assay showing the protein distribution in the supernatant (s) and pellet (p) portions.

Strains subjected to the assay were: [*NAC*$^{+}$] (white), [*nac*$^{-}$] (red), [*NAC*$^{+}$] after GuHCl treatment (w-GuHCl), [*SYN1*$^{+}$] (white), and [*syn1*$^{-}$] (red).

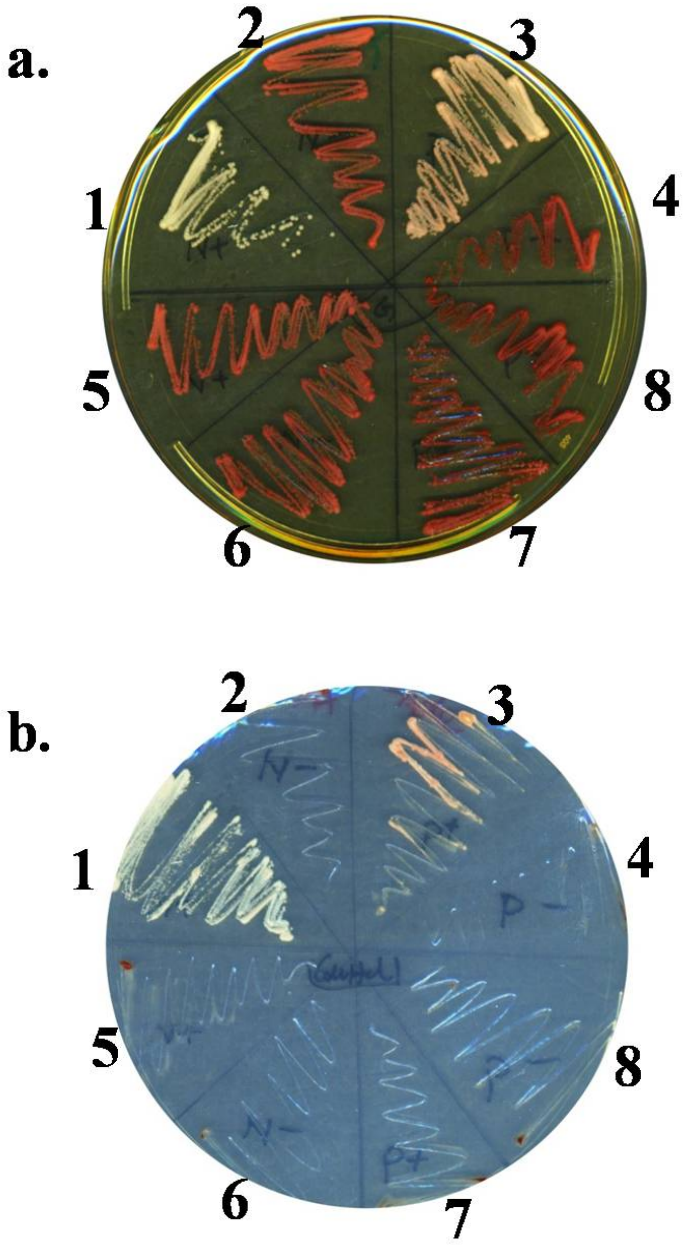


Figure 4-7. The cells expressing NACp grow on YPD (a) and SD-Ade (b).

(1) [NAC^{+}] cells; (2) [nac^{-}]cells; (3) [PSI^{+}] cells; (4) [psi^{-}] cells; (5) [NAC^{+}] cells after GuHCl treatment; (6) [nac^{-}]cells after GuHCl treatment; (7) [PSI^{+}] cells after GuHCl treatment; (8) [psi^{-}] cells after GuHCl treatment;

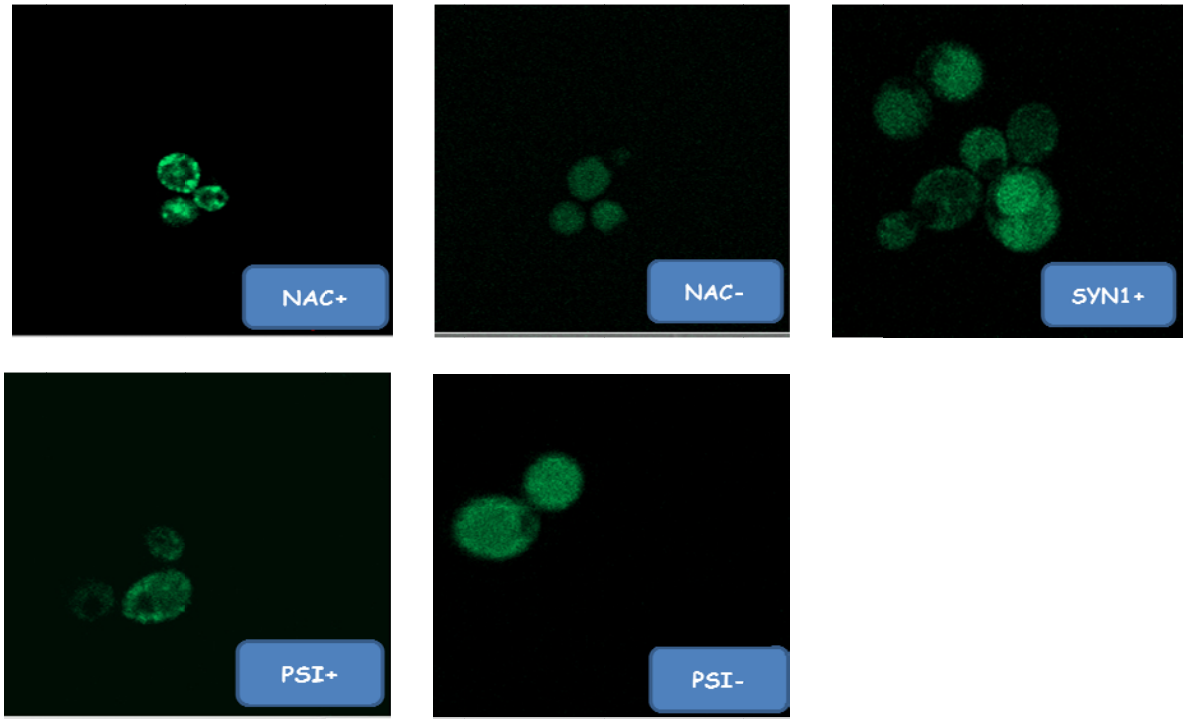


Figure 4-8. The fluorescent microscopy image of [NAC^{+}], [nac^{-}], [$SYN1^{+}$], [PSI^{+}] and [psi^{-}] cells.

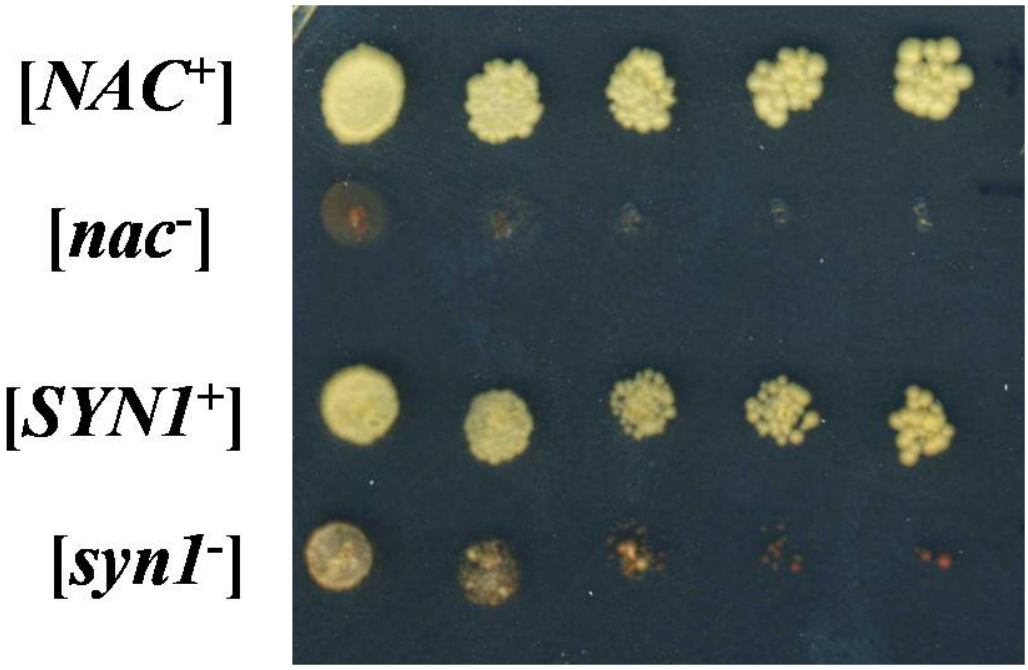


Figure 4-9. Spontaneous conversion rates estimated on SD-Ade plate.

The [*syn1*$^{-}$] cells exhibited high conversion rate to [*SYN1*$^{+}$] phenotype because approximately 1:10 of cells exhibited Ade+ phenotype; while [*nac*$^{-}$] cells were more stable and no growth was detected on SD-Ade plate.

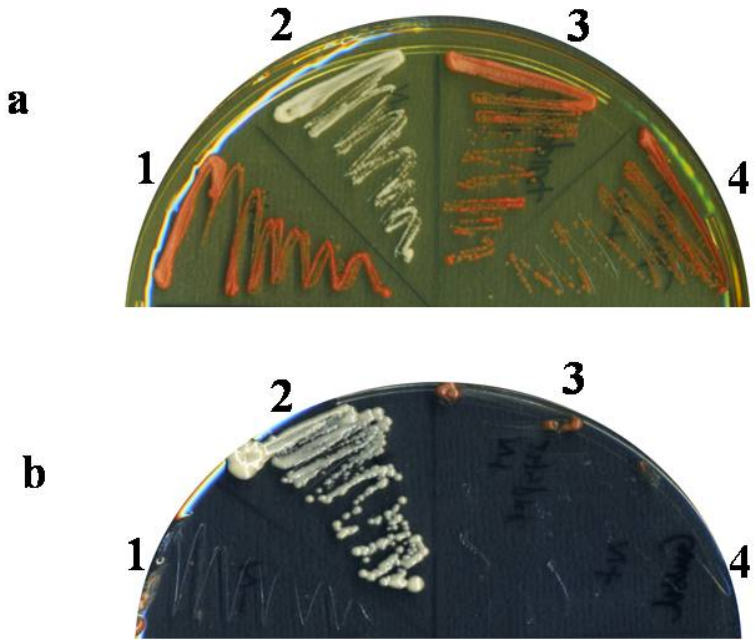


Figure 4-10. Deletion or inhibition of Hsp104 cures [NAC] prion.

On the YPD plate (a), the [*nac*$^{-}$] cells (1) are read and the [*NAC*$^{+}$] cells (2) are white. To cure the prion phenotype of [*NAC*$^{+}$] cells, either deletion of Hsp104 (3) or inhibition of Hsp104 with GuHCl successfully switch the cells to [*nac*$^{-}$] phenotype, in which the cells exhibited red on YPD plate. Accordingly, on the SD–His-Ade plate (b), only the [*NAC*$^{+}$] cells (2) displayed growth but the cells cured with Hsp104 deletion (3) or GuHCl (4) unable to survive on –Ade medium.

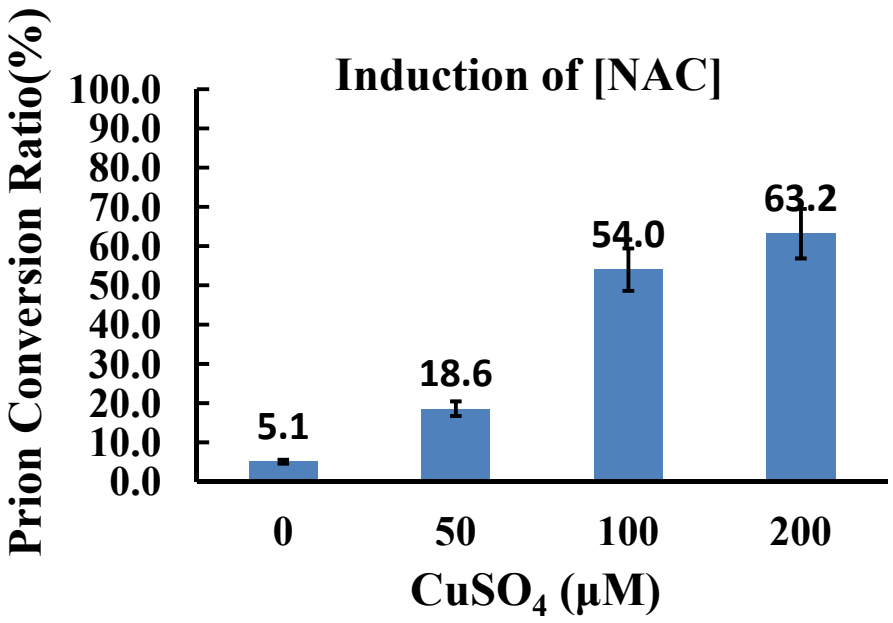


Figure 4-11. Overexpression of NAC-M-GFP fusion protein induced the formation of [NAC] prion. NAC-M-GFP fusion protein, under the *P*$_{CUP1}$ promoter, was induced by various concentration of $CuSO_4$. Over-expression caused the phenotype conversion, scored on YPD medium plates. White colonies were randomly picked and confirmed for the prion status.

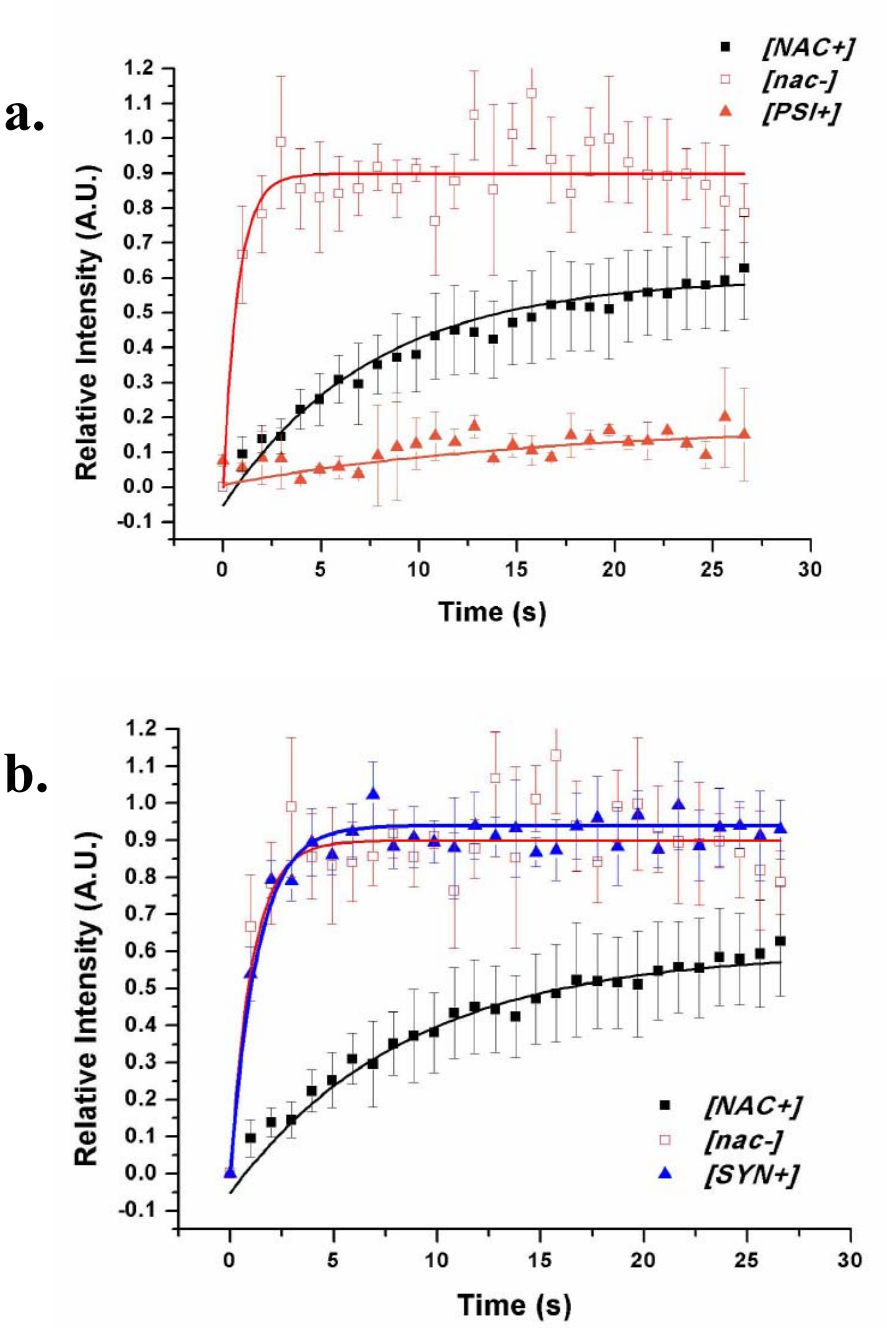


Figure 4-12 Diffusion and dynamics of GFP-fused NACp, SYN1p and Sup35p in the stationary phase cells.a) FRAP of NACp in the [NAC^{+}] and [nac^{-}] cells compared with Sup35 in [PSI^{+}] cells. b) FRAP of NACp in the [NAC^{+}] and [nac^{-}] cells compared with SYN1p in [$SYN1^{+}$] cells

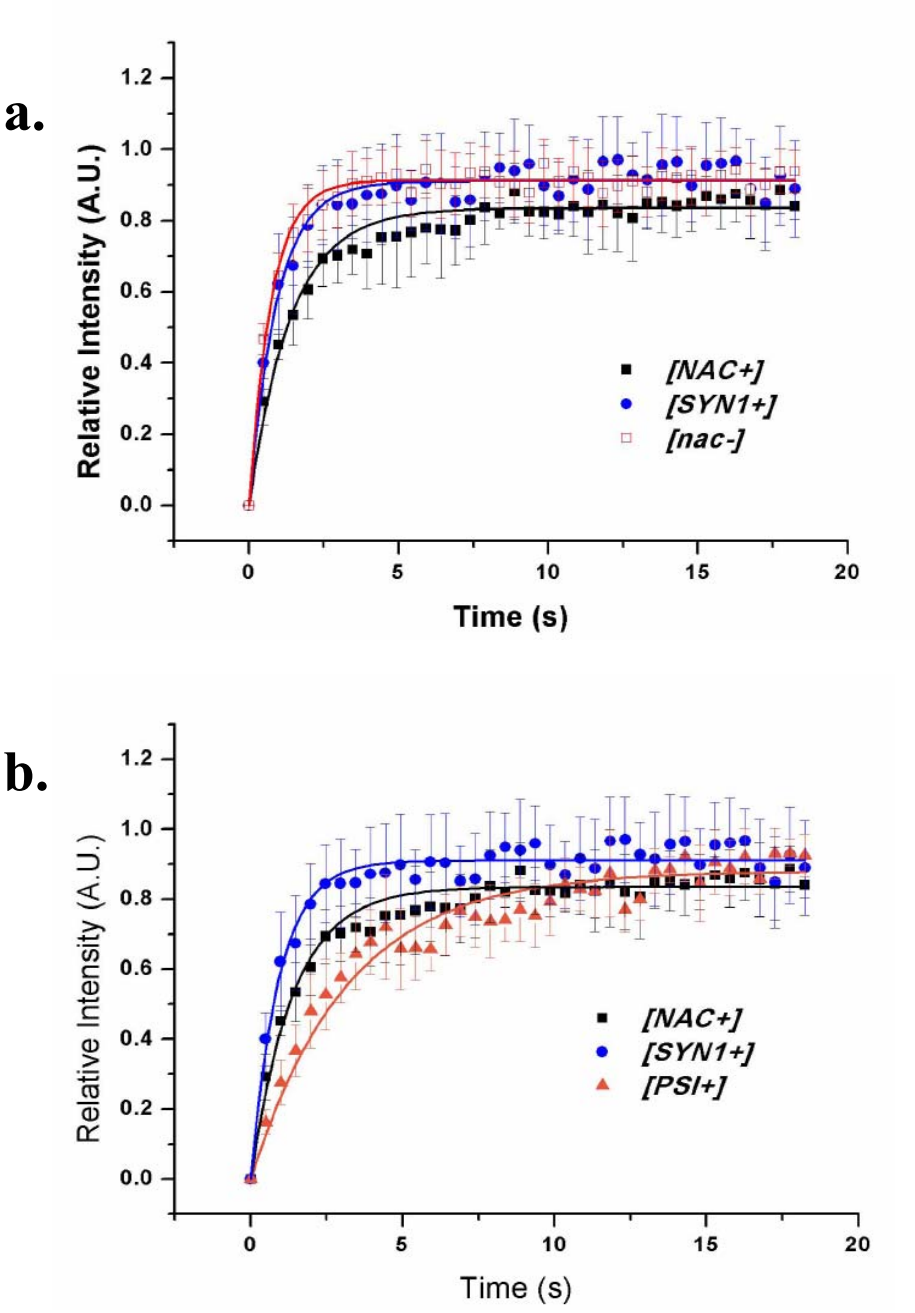
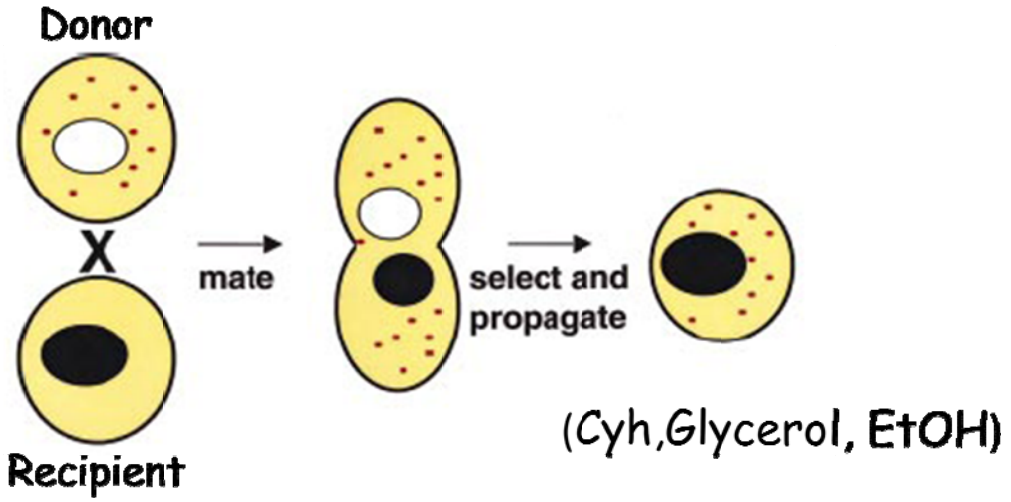


Figure 4-13. Diffusion and dynamics of GFP-fused NACp, SYN1p and Sup35p in the mid-log phase cells. a) FRAP of NACp in the [NAC^+] and [nac^-] cells compared with SYN1p in [$SYN1^+$] cells. b) FRAP of NACp in the [NAC^+] cells compared with SYN1p in [$SYN1^+$] cells and Sup35 in [PSI^+] cells.

a



b

Donor	Phenotype
$[NAC^{+}]$	**35 Ade+/36 Total**
$[nac^{-}]$	**0 Ade+/ 42 Total**
$[SYN1^{+}]$	**21 Ade+/ 22 Total**
$[PSI^{+}]$	**6 Ade+/ 24 Total**
$[psi^{-}]$	**0 Ade+/ 29 Total**

Figure 4- 14. Cytoduction experiment of [NAC] prion.

a. The scheme that shows the transfer of cytoplasmic materials during cytoduction(Sondheimer and Lindquist, 2000). In cytoduction, a strain containing the desired phenotypic element (the donor) is co-incubated with a strain lacking the element (the recipient). The two cell types fuse cytoplasmic content, but not their nuclei, preventing gene exchange. Haploid buds form from the heterokaryon. Desired offspring buds can be selected using nuclear markers from the recipient and propagated into colonies. b. The recipient strain was [*GT388k-nac*] and the cytoductant phenotype results are listed in the table.

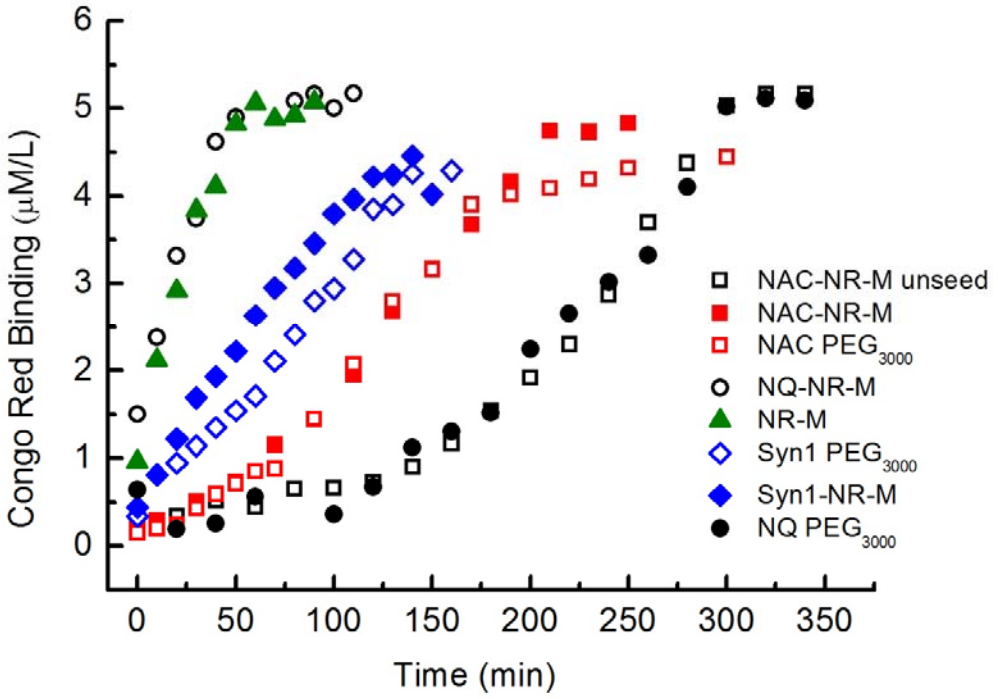


Figure 4- 15. Kinetics of aggregation by NAC peptide.

The kinetics of aggregation by NAC peptide unseeded ( ), with homologous seeding with NAC-NR-M ( ) and NAC-PEG_{3000} ( ) or heterologous seeding with NQ-NR-M (Sup35NM) ( ), NQ-PEG_{3000} (Sup35-PEG) ( ), NR-M (ΔNQ) ( ), Syn1-NR-M (Syn1-NM) ( ) and Syn1-PEG_{3000} ( ).

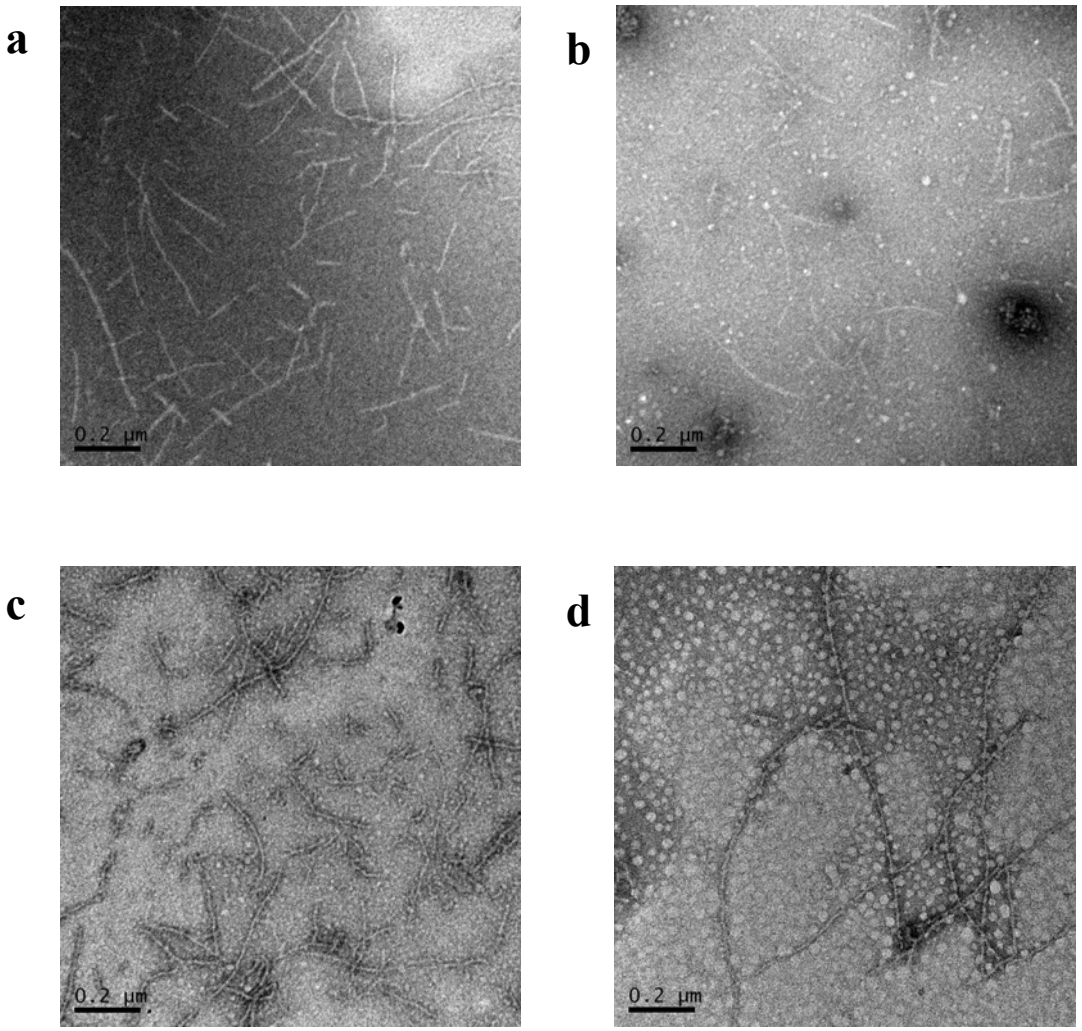


Figure 4- 16. The electron microscopy image of ulta-soniced fibers.

The fibers were formed by peptide NQ-NR-M (a), ΔNQ (b), Syn1-NR-M (c) and NAC-NR-M (d), which are used as seeds for NAC-NR-M self assembly.

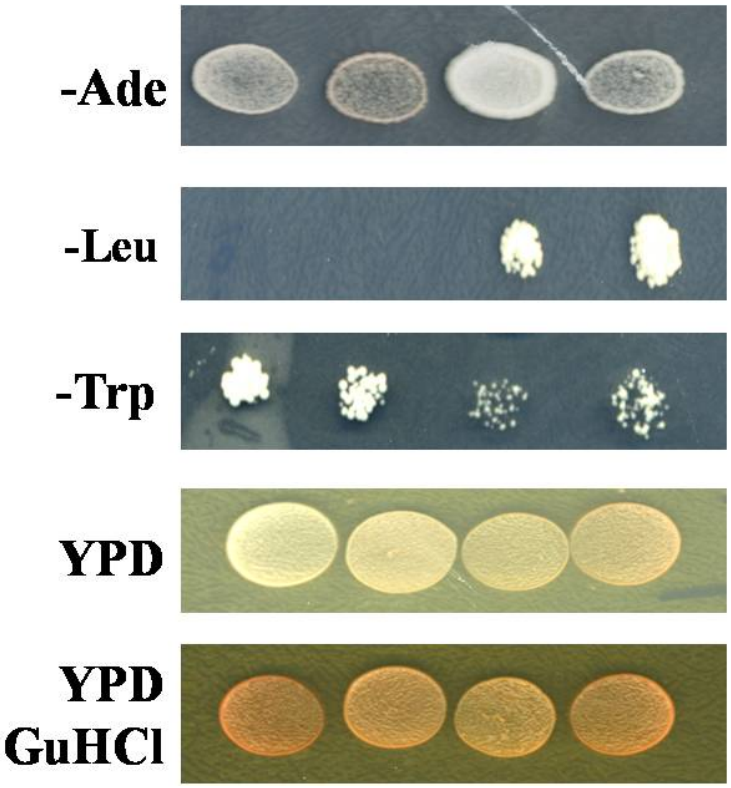


Figure 4-17. The mating and sporulation experiment of [NAC] prion.

The Ade+ phenotype is due to a non-chromosomal genetic element [NAC]. Cross mating [NAC^{+}] and [nac^{-}] cells leaded to 4:0 Ade+ in spores, while other markers segregated as 2+:2- in a typical tetrad. All spores exhibited white or light-pink color on YPD plate and were cured to red after GuHCl treatment.

Reference

Aylon, Y., Liefshitz, B., and Kupiec, M. (2004). The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. The EMBO Journal *23*, 4868-4875.

Borchsenius, A.S., Wegrzyn, R.D., Newnam, G.P., Inge-Vechtomov, S.G., and Chernoff, Y.O. (2001). Yeast prion protein derivative defective in aggregate shearing and production of new 'seeds'. EMBO J *20*, 6683-6691.

Chernoff, Y., Lindquist, S., Ono, B., Inge-Vechtomov, S., and Liebman, S. (1995a). Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]

10.1126/science.7754373. Science *268*, 880-884.

Chernoff, Y., Uptain, S., and Lindquist, S. (2002a). Analysis of prion factors in yeast. Methods Enzymol *351*, 499-538.

Chernoff, Y.O., Lindquist, S.L., Ono, B., Inge-Vechtomov, S.G., and Liebman, S.W. (1995b). Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. Science *268*, 880-884.

Chernoff, Y.O., Uptain, S.M., and Lindquist, S.L. (2002b). Analysis of prion factors in yeast. Methods Enzymol *351*, 499-538.

Chou, P.Y., and Fasman, G.D. (1974). Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. Biochemistry *13*, 211-222.

Conde, J., and Fink, G. (1976). A Mutant of Saccharomyces cerevisiae Defective for Nuclear Fusion. Proceedings of the National Academy of Sciences *73*, 3651-3655.

Cookson, M.R. (2005). The biochemistry of Parkinson's disease. Annu Rev Biochem *74*, 29-52.

Davidson, W.S., Jonas, A., Clayton, D.F., and George, J.M. (1998). Stabilization of alpha-synuclein secondary structure upon binding to synthetic membranes. J Biol Chem *273*, 9443-9449.

DePace, A.H., Santoso, A., Hillner, P., and Weissman, J.S. (1998). A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. Cell *93*, 1241-1252.

Derkatch, I.L., Bradley, M.E., Zhou, P., Chernoff, Y.O., and Liebman, S.W. (1997). Genetic and Environmental Factors Affecting the de novo Appearance of the [PSI(+)] Prion in Saccharomyces cerevisiae. Genetics *147*, 507-519.

Derkatch, I.L., Chernoff, Y.O., Kushnirov, V.V., Inge-Vechtomov, S.G., and Liebman, S.W. (1996a). Genesis and Variability of [PS1] Prion Factors in Saccharomyces cerevisiae. Genetics *144*, 1375-1386.

Derkatch, I.L., Chernoff, Y.O., Kushnirov, V.V., Inge-Vechtomov, S.G., and Liebman, S.W. (1996b). Genesis and variability of [PSI] prion factors in Saccharomyces cerevisiae. Genetics *144*, 1375-1386.

Dobson, C.M. (2001). Protein folding and its links with human disease. Biochem Soc Symp, 1-26.

Dobson, C.M. (2003). Protein folding and misfolding. Nature *426*, 884-890.

El-Agnaf, O.M., and Irvine, G.B. (2000). Review: formation and properties of amyloid-like fibrils derived from alpha-synuclein and related proteins. J Struct Biol *130*, 300-309.

El-Agnaf, O.M., Jakes, R., Curran, M.D., Middleton, D., Ingenito, R., Bianchi, E., Pessi, A., Neill, D., and Wallace, A. (1998). Aggregates from mutant and wild-type alpha-synuclein proteins and NAC peptide induce apoptotic cell death in human neuroblastoma cells by formation of beta-sheet and amyloid-like filaments. FEBS Lett *440*, 71-75.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B.*, et al.* (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387-391.

Giasson, B.I., Murray, I.V., Trojanowski, J.Q., and Lee, V.M. (2001). A hydrophobic stretch of 12 amino acid residues in the middle of alpha-synuclein is essential for filament assembly. J Biol Chem *276*, 2380-2386.

Glover, J.R., Kowal, A.S., Schirmer, E.C., Patino, M.M., Liu, J.J., and Lindquist, S. (1997). Self-seeded fibers formed by Sup35, the protein determinant of [PSI+], a heritable prion-like factor of S. cerevisiae. Cell *89*, 811-819.

Glover, J.R., and Lindquist, S. (1998). Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. Cell *94*, 73-82.

Hara, H., Nakayashiki, T., Crist, C.G., and Nakamura, Y. (2003). Prion domain interaction responsible for species discrimination in yeast [PSI+] transmission

doi:10.1111/j.1365-2443.2003.00694.x. Genes to Cells *8*, 925-939.

Kelly, J.W. (1998). The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Curr Opin Struct Biol *8*, 101-106.

Khurana, V., and Lindquist, S. (2010). Modelling neurodegeneration in Saccharomyces cerevisiae: why cook with baker's yeast? Nat Rev Neurosci *11*, 436-449.

King, C.Y., Tittmann, P., Gross, H., Gebert, R., Aebi, M., and Wuthrich, K. (1997). Prion-inducing domain 2-114 of yeast Sup35 protein transforms in vitro into amyloid-like filaments. Proc Natl Acad Sci U S A *94*, 6618-6622.

Kushnirov, V.V., Vishnevskaya, A.B., Alexandrov, I.M., and Ter-Avanesyan, M.D. (2007). Prion and nonprion amyloids: a comparison inspired by the yeast Sup35 protein. Prion *1*, 179-184.

Lavedan, C. (1998). The synuclein family. Genome Res *8*, 871-880.

Lippincott-Schwartz, J., Altan-Bonnet, N., and Patterson, G.H. (2003). Photobleaching and photoactivation: following protein dynamics in living cells. Nat Cell Biol *Suppl*, S7-14.

Lippincott-Schwartz, J., Snapp, E., and Kenworthy, A. (2001). Studying protein dynamics in living cells. Nat Rev Mol Cell Biol *2*, 444-456.

Maddelein, M.L., and Wickner, R.B. (1999). Two prion-inducing regions of Ure2p are nonoverlapping. Mol Cell Biol *19*, 4516-4524.

Maries, E., Dass, B., Collier, T.J., Kordower, J.H., and Steece-Collier, K. (2003). The role of alpha-synuclein in Parkinson's disease: insights from animal models. Nat Rev Neurosci *4*, 727-738.

Masison, D.C., and Wickner, R.B. (1995). Prion-inducing domain of yeast Ure2p and protease resistance of Ure2p in prion-containing cells. Science *270*, 93-95.

Moriyama, H., Edskes, H.K., and Wickner, R.B. (2000). [URE3] prion propagation in Saccharomyces cerevisiae: requirement for chaperone Hsp104 and curing by overexpressed chaperone Ydj1p. Mol Cell Biol *20*, 8916-8922.

Nemecek, J., Nakayashiki, T., and Wickner, R.B. (2009). A prion of yeast metacaspase homolog (Mca1p) detected by a genetic screen. Proc Natl Acad Sci U S A *106*, 1892-1896.

Osherovich, L.Z., Cox, B.S., Tuite, M.F., and Weissman, J.S. (2004). Dissection and design of yeast prions. PLoS Biol *2*, E86.

Paushkin, S.V., Kushnirov, V.V., Smirnov, V.N., and Ter-Avanesyan, M.D. (1996). Propagation of the yeast prion-like [psi+] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor. EMBO J *15*, 3127-3134.

Polymeropoulos, M.H., Lavedan, C., Leroy, E., Ide, S.E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R.*, et al.* (1997). Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science *276*, 2045-2047.

Qi, X. (2007). Generation of Novel Yeast Prions through Domain Substitutions of Saccharomyces cerevisiae Sup35p. In Chemistry (Atlanta, Emory).

Reits, E.A., and Neefjes, J.J. (2001). From fixed to FRAP: measuring protein mobility and activity in living cells. Nat Cell Biol *3*, E145-147.

Ripaud, L., Maillet, L., and Cullin, C. (2003). The mechanisms of [URE3] prion elimination demonstrate that large aggregates of Ure2p are dead-end products. EMBO J *22*, 5251-5259.

Rivers, R.C., Kumita, J.R., Tartaglia, G.G., Dedmon, M.M., Pawar, A., Vendruscolo, M., Dobson, C.M., and Christodoulou, J. (2008). Molecular determinants of the aggregation behavior of alpha- and beta-synuclein. Protein Sci *17*, 887-898.

Safar, J., and Prusiner, S.B. (1998). Molecular studies of prion diseases. Prog Brain Res *117*, 421-434.

Santoso, A., Chien, P., Osherovich, L.Z., and Weissman, J.S. (2000). Molecular basis of a yeast prion species barrier. Cell *100*, 277-288.

Serio, T.R., Cashikar, A.G., Moslehi, J.J., Kowal, A.S., and Lindquist, S.L. (1999). Yeast prion [psi +] and its determinant, Sup35p. Methods Enzymol *309*, 649-673.

Sherman, M.Y. (2004). Yeast prions: protein aggregation is not enough. PLoS Biol *2*, E125.

Sikorski, R., and Hieter, P. (1989). A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19-27.

Sondheimer, N., and Lindquist, S. (2000). Rnq1: an epigenetic modifier of protein function in yeast. Mol Cell *5*, 163-172.

Song, Y., Wu, Y.X., Jung, G., Tutar, Y., Eisenberg, E., Greene, L.E., and Masison, D.C. (2005). Role for Hsp70 chaperone in Saccharomyces cerevisiae prion seed replication. Eukaryot Cell *4*, 289-297.

Spillantini, M.G., Crowther, R.A., Jakes, R., Hasegawa, M., and Goedert, M. (1998). alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies. Proc Natl Acad Sci U S A *95*, 6469-6473.

Tessier, P.M., and Lindquist, S. (2007). Prion recognition elements govern nucleation, strain specificity and species barriers. Nature *447*, 556-561.

Ueda, K., Fukushima, H., Masliah, E., Xia, Y., Iwai, A., Yoshimoto, M., Otero, D.A., Kondo, J., Ihara, Y., and Saitoh, T. (1993). Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. Proc Natl Acad Sci U S A *90*, 11282-11286.

Uptain, S.M., and Lindquist, S. (2002). Prions as protein-based genetic elements. Annu Rev Microbiol *56*, 703-741.

Waxman, E.A., Emmer, K.L., and Giasson, B.I. (2010). Residue Glu83 plays a major role in negatively regulating alpha-synuclein amyloid formation. Biochem Biophys Res Commun *391*, 1415-1420.

Waxman, E.A., Mazzulli, J.R., and Giasson, B.I. (2009). Characterization of hydrophobic residue requirements for alpha-synuclein fibrillization. Biochemistry *48*, 9427-9436.

Wegrzyn, R.D., Bapat, K., Newnam, G.P., Zink, A.D., and Chernoff, Y.O. (2001). Mechanism of prion loss after Hsp104 inactivation in yeast. Mol Cell Biol *21*, 4656-4669.

Weinreb, P.H., Zhen, W., Poon, A.W., Conway, K.A., and Lansbury, P.T., Jr. (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. Biochemistry *35*, 13709-13715.

Wu, Y.X., Greene, L.E., Masison, D.C., and Eisenberg, E. (2005). Curing of yeast [PSI+] prion by guanidine inactivation of Hsp104 does not require cell division. Proc Natl Acad Sci U S A *102*, 12789-12794.

Wu, Y.X., Masison, D.C., Eisenberg, E., and Greene, L.E. (2006). Application of photobleaching for measuring diffusion of prion proteins in cytosol of yeast cells. Methods *39*, 43-49.

Zhou, P., Derkatch, I.L., and Liebman, S.W. (2001). The relationship between visible intracellular aggregates that appear after overexpression of Sup35 and the yeast prion-like elements [PSI(+)] and [PIN(+)]. Mol Microbiol *39*, 37-46.