NORTHWESTERN UNIVERSITY

Examining prionogenic domains in the Swi1 protein, yeast evolution, and novel disease-related aggregators

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Neuroscience

By

Dustin K. Goncharoff

EVANSTON, ILLINOIS

June 2021

ABSTRACT

Examining prionogenic domains in

the Swi1 protein, yeast evolution, and novel disease-related aggregators

Dustin K. Goncharoff

Thesis Advisor: Liming Li

Prions are self-perpetuating, alternative protein conformations associated with neurological diseases and normal cellular functions. *Saccharomyces cerevisiae* contains many endogenous prions – providing a powerful system to study prionization. Previously, the Li Lab demonstrated that Swi1, a component of the SWI/SNF chromatin-remodeling complex, can form the prion [SWI^+]. A small region, $Swi1_{1\text{-}38}$, with a unique amino-acid composition of low complexity, acts as a prion domain and supports [SWI^+] propagation. Here, we further examine $Swi1_{1\text{-}38}$ through site-directed mutagenesis. We found that mutations of the two phenylalanine residues at positions 3 and 4 or the threonine tract at positions 15 to 22 inhibit $Swi1_{1\text{-}38}$ aggregation and prion propagation. In addition, mutating both phenylalanines can abolish *de novo* prion formation by $Swi1_{1\text{-}38}$ whereas mutating only one phenylalanine does not. Replacement of half or the entire eight-threonine tract with alanines has the same effect, possibly disrupting a core region of $Swi1_{1\text{-}38}$ aggregates. We also show that $Swi1_{1\text{-}38}$ and its prion-fold-maintaining mutants form high-molecular-weight, SDS-resistant aggregates whereas the double phenylalanine mutants eliminate these protein species. These results indicate the necessity of the large hydrophobic residues and threonine tract in $Swi1_{1\text{-}38}$ in prionogenesis – possibly acting as important aggregatable regions. Our findings thus highlight the importance of specific amino-acid residues in the Swi1 prion domain in prion formation and propagation.

Given the multitude of endogenous yeast prions identified in *S. cerevisiae*, including the aforementioned Swi1, there remains an open question as to what is the evolutionary importance of these proteins and the prion domains contained within. A recent study sequenced over a thousand *S. cerevisiae* isolates from diverse geographic and ecologic origins. This data set has provided us the opportunity to probe the selection pressures that yeast prion proteins have undergone. Through comparison to *S. paradoxus* – the closest extant relative of *S. cerevisiae*, we have examined the rate of nonsynonymous (d_N) and synonymous (d_S) substitutions in yeast prion genes as well as several yeast prion-related genes including the *FLO* genes, which regulate yeast multicellularity and are regulated by a yeast prion network. In doing so, we have found an absence of evidence for correlation between the ecological origin of a yeast isolate and the observed selection pressures on both yeast prion genes and the *FLO* genes. However, there exists a divergent group of isolates undergoing strong selection pressure for *FLO11*. Most intriguingly, we find that yeast prion genes experience greater variation in their computed d_N / d_S values for their PrD-encoding regions as compared to their non-PrD-encoding regions. Additionally, the PrD portions of the genes are under purifying selection though the selection pressure is weaker than the non-PrD regions as viewed by the d_N / d_S ratio. Lastly, we examined predicted prionogenicity – as evaluated by the PLAAC and PAPA algorithms – as a stand-in for the functioning of the proteins that would be produced by the diverse alleles of yeast prion genes. General ecological origin did not correlate with changes in prionogenicity, and no strong correlations were seen in prionogenicity of different yeast prion proteins. In summary, although *S. cerevisiae* exists in a variety of ecological niches, the multiple yeast prion genes contained in the species are relatively maintained in their prionogenic potential but still undergoing relatively greater nonsynonymous changes in their PrD-encoding regions.

Prion proteins also play an important role throughout other organisms including humans. Unfortunately, these proteins and related prion-like proteins underpin numerous incurable neurological diseases. Recent research has suggested a role for the prion phenomenon in cancers as p53 – which is widely mutated in cancers – can aggregate in a prion-like fashion that is also pathologically relevant. As such, we set out to use a bioinformatically-informed approach to identify novel protein aggregators involved in cancer. By using a combination of established prediction algorithms for prionogenicity, amyloidogenicity, disorderedness, and general aggregation propensity, we selected a group of top-ranked cancer-related proteins. From these proteins, we have found that several BAF complex proteins readily aggregate in a yeast model system. The BAF complex – also known as the human SWI/SNF complex – is highly mutated across virtually all cancer types and contains proteins functionally homologous to the yeast prion protein Swi1. Thus, the aggregation of these proteins in a yeast model may indicate possible disease-relevant aggregation is possible in humans as well. Furthermore, we identified that PrLDs contained in the clathrin interacting protein CLINT1 and in the circadian clock protein CLOCK can act as bona fide prion domains in a yeast assay. CLINT1 PrLD and CLOCK PrLD A – so termed due to the presence of two predicted PrLDs in CLOCK – both form monomeric, diffuse conformations in yeast as well as alternative conformations. These alternative conformations are stable and heritable and form high-molecular weight, SDS-resistant protein species. Given the important role that CLINT1 plays in vesicular trafficking and CLOCK plays in circadian rhythms and healthy homeostasis, aggregation driven by these PrLDs may possibly act as a foundation to human diseases including cancer, schizophrenia, and neurodegeneration.

ACKNOWLEDGMENTS

This thesis was made possible by Dr. Liming Li and her continued support of me in her lab and in my career as a scientist. As my thesis advisor, she has always provided wise and candid advice and a sense of scientific excitement – even when I was at my most jaded. I often would describe Liming as the kindest person that I have ever met to others and her caring attitude made my life easier throughout my graduate career. She was steadfast in noting that the main thing that I needed to worry about was doing good science. I believe that the final product is worthy of being considered good science and Liming's support made everything possible.

The experiments and data within these pages contain not just the work of myself but also of multiple undergraduates and post-baccalaureate students. These talented young scientists – Sarah Applebey, Raudel Cabral, Brandon Cho, Rowan Hussein, Manasa Pagadala, and Jake Rosenthal – were indispensable to completing my thesis work. Most have since moved on to other scientific pursuits at the University of Pennsylvania, Northwestern University, or elsewhere. I have confidence that all will be successful doctors, scientists, or whatever else they may find fulfilling. Mentoring them in the lab was the highlight of my doctoral work.

The aforementioned individuals and the other members of the Li Lab were a second family to me and supported me in a plethora of ways during my time at Northwestern. Chief amongst them was Dr. Zhiqiang Du, whose constant expertise in the lab – particularly with anything yeast – made my experiments possible. He would always remark "should be fine" when we would try something out of the ordinary or jury-rig lab equipment. His scientific input was invaluable. Another was Dr. Stephanie Valtierra, a fellow graduate student who was always a great sounding board for any number of problems in lab or elsewhere. She worked on some projects that were

related to one of my own and always had pertinent advice on troubleshooting experiments. Her ability to keep things light-hearted in lab made the problematic days much more tolerable. Other close colleagues over the years include Luzivette Robles and James Matsumura.

Outside of the Li Lab, the input provided by my thesis committee members allowed for me to wholly improve the thesis presented here now. Drs. Jeffrey Savas, Marc Mendillo, and Robert Vassar always had another experimental idea or another astute critique or another piece of career advice – while pushing me to complete this phase of my studies and move onto other things. To a naïve graduate student, the thesis committee could seem to be more of a barrier to completion than a vital and helpful group of guides, but my committee was most definitely the latter.

Collaborators played a key role in furthering the projects in my thesis. In regard to the second chapter focusing on the evolution of prion domains in yeast, the input and guidance of the laboratory of Dr. Manyuan Long at the University of Chicago proved valuable. Critically, Dr. Shengjun Tan conducted the bioinformatic programming and calculations needed to compare the DNA sequences of the various yeast isolates. Regarding the third chapter focusing on the identification of novel protein aggregators, the laboratory of my committee member, Dr. Mendillo played a vital role. In particular, Dr. Milad Alasady and Jasen Jackson worked on evaluating the candidates in cell culture systems in tandem with the yeast work conducted under my purview. Furthermore, Drs. Randal Halfmann and Shriram Venkatesan of Stowers Institute for Medical Research provided the opportunity to evaluate proteins in their high-throughput system.

In addition to the support of everyone in the Li Lab, my committee members, and my collaborators, I had a similarly large team of people behind me at home. Chief amongst them was my girlfriend and best friend, Elizaveta Klementieva who would always take my mind off any troubles with experiments, looming deadlines, or other stressors. And when experiments were

moving well, she would roll right with me while I explained why a picture of a yeast plate or the black bands of a western blot were great. I am a lucky guy to have someone that shares in my joy when I stick a print-out of figure-quality data to the fridge. Moreover, her sharing of her own work as an attorney with me is part of the reason that I have grown an interest in patent law – leading me to choose to pursue a career in the field. We enrich each other's lives and are best together. I have asked much of Liza throughout the past four-plus years that we have spent together thus far, and she has shown understanding, patience, and love. I cannot wait to fulfill her dreams of living between the mountains and the sea, and this dissertation is one step on the way to building a family with each other. Her adventurous spirit has managed to balance out my cautious self and her pride and belief in me makes me a happy man.

My family has always been extremely supportive of my doctoral work and my scientific work before even that begun. Sharing my love of science with my niece, Savannah has enriched my life. My older sister, Patricia and my younger brother, Derek were always there – especially for some good laughs. My dad, Ken kept me supported by helping me with whatever I needed and via our regular weekday lunches downtown. He would ask me to send him any papers that I authored or contributed to, and he would look at them, even though it was likely all gibberish to him. My stepmom, Celine Goncharoff and stepsister, Margaux Di Iorio have been wonderful and kept me fueled with home-cooked meals and other goodies. A couple of furry friends – Stormy and Leo – have also been there to offer some warm, unquestioning companionship when needed (although bribery with a couple of cat treats was sometimes involved).

Lastly, nothing here would have been accomplished without my biggest supporter and mother, Robin. She dedicated herself to raising my sister and I, and she was always immensely proud of whatever I accomplished. I would get annoyed with her when, in typical mom fashion,

she would force me to tell people about what I am studying or how well I was doing in school. Now I recognize that this was not just for her to brag about her kids, but that she wanted to see me be proud and confident in myself.

When I moved to Chicago for graduate school, I would talk to her weekly by phone and these calls were always a great way to hash things out about anything stressing me. She went to the hospital during the winter holidays. I talked to her from the hospital a couple of times. The last phone call broke me. I went from lab one day to supporting my mother at home in hospice the next. She passed away from heart failure. Just like that, my main support was gone. This was not how my time in graduate school was supposed to go. It took many struggles to get back to feeling like a person again, let alone capable of focusing on any of my experiments.

Losing my mom forever changed my life and will prevent me from having her present at my thesis defense as well as other significant life moments such as my wedding, the birth of my first child, and so on. Her limitless love and support for me paved my way in life as I knew that I could do whatever I needed to do or wanted to do. During one of my final conversations with her, she made me promise that I would stay in graduate school and finish my PhD. This thesis is ultimately dedicated to her. I love you mom.

ABBREVIATIONS

4xTA	mutation of 4 threonines to 4 alanines
4xTN	mutation of 4 threonines to 4 asparagines
5-FOA	5-fluoroorotic acid
8xTA	mutation of 8 threonines to 8 alanines
8xTN	mutation of 8 threonines to 8 asparagines
A	alanine
Aβ	amyloid-beta
AD	Alzheimer's disease
Ade	adenine
-Ade	synthetic complete media lacking adenine
ALS	amyotrophic lateral sclerosis
AP-1	adaptor protein-1
ASC	apoptosis-associated speck-like protein containing a CARD
ATCC	American Type Culture Collection
BSE	bovine spongiform encephalopathy
CARD	caspase activation and recruitment domain
CBD	corticobasal degeneration
CDS	coding sequence
CJD	Creutzfeldt-Jakob disease
COSMIC	Catalogue of Somatic Mutations in Cancer
CTE	chronic traumatic encephalopathy

CWD	chronic wasting disease
DAmFRET	distributed amphifluoric FRET
d_N	rate of nonsynonymous base substitutions
d_S	rate of synonymous base substitutions
ENTH	epsin-N-terminal-homology
F	phenylalanine
FF-AA	mutation of adjacent phenylalanines to alanines
FF-NN	mutation of adjacent phenylalanines to asparagines
FRET	Förster resonance energy transfer
FTD	frontotemporal dementia
FUS	fused in sarcoma
GdnHCl	guanidine hydrochloride
GFP	green fluorescent protein
GLFG	glycine-leucine-phenylalanine-glycine
HNRNPA1/2B1	heterogeneous nuclear ribonucleoprotein A1 and A2B1
IDR	intrinsically disordered region
kDa	kilodalton
LD	Luminidependens
Leu	leucine
-Leu	synthetic complete media lacking leucine
-Leu-Ura	synthetic complete media lacking leucine and uracil
LiAc	lithium acetate
MAVS	mitochondrial antiviral-signaling protein

N	asparagine
NaOH	sodium hydroxide
NEF	nucleotide exchange factor
NES	nuclear export signal
NFT	neurofibrillary tangle
ORF	open reading frame
PAPA	prion aggregation prediction algorithm
PASTA	prediction of amyloid structural aggregation
PCR	polymerase chain reaction
PD	Parkinson's disease
PDB	Protein Data Bank
PEG	polyethylene glycol
PEU	PASTA Energy Unit
PLAAC	Prion-Like Amino Acid Composition
PrD	prion domain
PrLD	prion-like domain
PrP	prion protein
PrP^{C}	prion protein cellular; the non-prion form of PrP
PrP^{Sc}	prion protein scrapie; the prion form of PrP
PSP	progressive supranuclear palsy
PVDF	polyvinylidene fluoride
PYD	pyrin domain
Q	glutamine

RIG-I	retinoic acid-inducible gene I
rpm	revolutions per minute
RT-PCR	reverse-transcription polymerase chain reaction
SDD-AGE	semi-denaturing detergent agarose gel electrophoresis
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SOD1	superoxide dismutase 1
ssDNA	single-stranded DNA
$Sup35_{FL}$	full-length Sup35
$Sup35_{MC}$	Sup35 construct consisting of the M and C domains
$Swi1_{1-32}$	Swi1 construct consisting of the first 32 amino-acid residues
$Swi1_{1-38}$	Swi1 construct consisting of the first 38 amino-acid residues
$Swi1_{1-42}$	Swi1 construct consisting of the first 42 amino-acid residues
$Swi1_{1-46}$	Swi1 construct consisting of the first 46 amino-acid residues
$Swi1_{1-65}$	Swi1 construct consisting of the first 65 amino-acid residues
$Swi1_{1-74}$	Swi1 construct consisting of the first 74 amino-acid residues
$Swi1_{1-133}$	Swi1 construct consisting of the first 133 amino-acid residues
$Swi1_{1-176}$	Swi1 construct consisting of the first 176 amino-acid residues
$Swi1_{1-224}$	Swi1 construct consisting of the first 224 amino-acid residues
$Swi1_{1-38}$Mut-MC	$Swi1_{1-38}$ mutant fused to $Sup35_{MC}$
$Swi1_{1-38}$Mut-YFP	$Swi1_{1-38}$ mutant fused to YFP
$Swi1_{239-259}$	Swi1 construct consisting of only amino-acid residues 239 through 259
$Swi1_{FL}$	full-length Swi1
$Swi1_{N}$	Swi1 construct consisting of only the N domain (residues 1-327)

$Swi1_{NQ}$	Swi1 construct consisting of the N and Q domains (residues 1-536)
$Swi1_{NQ\Delta 239-259}$	Swi1 construct consisting of $Swi1_{NQ}$ with $Swi1_{239-259}$ deleted
T	threonine
TCGA	The Cancer Genome Atlas
TDP-43	TAR (transactive response) DNA-binding protein 43
TEM	transmission electron microscopy
TGN	trans-Golgi network
ThS	thioflavin S
ThT	thioflavin T
TSE	transmissible spongiform encephalopathy
Ura	uracil
-Ura	synthetic complete media lacking uracil
WT	wild-type
YFP	yellow fluorescent protein
YPD	yeast extract peptone dextrose

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

Introduction

Proteinaceous infectious particles, or prions were first described by Prusiner in 1986 (1). Initially, this concept was applied to a singular protein, which is now known as the namesake prion protein (PrP). PrP can exist in multiple conformations – the normal, cellular form (PrP^{C}) that consists of a primarily alpha-helical structure and numerous disease-associated, prion forms (PrP^{Sc}) that principally consist of beta-sheets. PrP and other prion proteins have several defining characteristics: 1. A protein-only existence; 2. Capability to transmit a prion fold from molecule to molecule; 3. Infectivity from cell-to-cell or organism-to-organism. These early descriptions of prions were done in conjunction with the investigation of a group of deadly neurological diseases known as transmissible spongiform encephalopathies (TSEs), which include scrapie and bovine spongiform encephalopathy (BSE; i.e., mad cow disease) in farm animals and Creutzfeldt-Jakob disease (CJD) in humans. The prion concept has since grown to encompass proteins from across diverse species and involved in myriad diseases and even in a variety of functional roles.

Prions and Neurodegeneration

Currently, diseases with prion-like pathology – other than the aforementioned TSEs – include neurodegenerative diseases as well as some cancers, which is a topic to be covered in the next section. Chief amongst these neurodegenerative diseases are Alzheimer's disease (AD), Parkinson's diseases (PD), and amyotrophic lateral sclerosis (ALS). Common pathologies can be found amongst these diseases – aging, proteasomal disfunction, mitochondrial disfunction, widespread cell death, among others. Protein aggregation exists as a hallmark of

neurodegeneration. In AD, amyloid-beta (Aβ) and tau form the characteristic plaques and tangles (2). In PD, alpha-synuclein (α-syn) forms various deposits including the well-known Lewy bodies (3). In ALS, a variety of proteins including HNRNPA1/2B1 (heterogeneous nuclear ribonucleoprotein A1 and A2B1), FUS (fused in sarcoma), SOD1 (superoxide dismutase 1), and TDP-43 (TAR DNA-binding protein 43) can form aggregates (4). Curiously, many of these proteins have shown characteristics of a prion. While proteins generally must have demonstrated infectivity to be termed prions, another classification – prion-like – has been created for proteins that display prion behavior without proving to be infectious from one organism to another. Many neurodegenerative proteins involved in the formation of aggregates have been found to act in a prion-like manner. These prion-like characteristics include intercellular transmission, templating of normally folded proteins, formation of amyloids, and the existence of different conformational strains.

The Aβ peptide drives the neuronal cell death found in AD. In this disease, Aβ forms toxic oligomers and deposits known as plaques. The appearance of these plaques in mice can be seeded through injection of AD brain lysates (5). This seeding can be prevented by denaturation or treatment with Aβ-specific antibodies, indicating that the Aβ protein is necessary for successful seeding (6). Additionally, Aβ isolated from patients with various diseases exists as amyloids with varying structures – supporting the notion of prion-like strains of Aβ conformations (7). Stages of AD can be determined in post-mortem brains through examination of which areas of the brain have developed plaques and neurofibrillary tangles (NFTs) formed by the tau protein (8). This progressive appearance of plaques in various brain areas serves to underscore the prion-like spreading of Aβ pathology in AD. In these experiments and many others, Aβ behaves in a prion-like manner and its associated pathology reflects this behavior (9).

The tau protein normally functions as a microtubule-associated protein and its structure is highly disordered (10, 11). In AD pathology, tau can be found aggregated – typically as amyloids and hyperphosphorylated – in NFTs. Moreover, tau pathology underpins a multitude of other neurodegenerative diseases such as corticobasal degeneration (CBD), frontotemporal dementia (FTD), chronic traumatic encephalopathy (CTE), and progressive supranuclear palsy (PSP), which collectively are known as tauopathies. As such, the development of treatment strategies for tau aggregation and its associated toxicity and cell death has great importance. Recent studies have revealed strong evidence for the prion-like behavior of tau in disease. Just as with Aβ, brain homogenates from diseased patients or transgenic mice can propagate tau inclusions when injected into mice brains and homogenates from these mice can also be secondarily used for injection-based propagation (12, 13). This seeding capability of tau has been confirmed with other *in vivo* assays (14). Furthermore, the different progressions and disease phenotypes of various tauopathies have been hypothesized to occur due to differences in initial formation of pathological tau aggregates and the route through which they spread (15, 16).

Perhaps the most convincing evidence for the prion-like nature of tau came in the form of identification of disease-specific strains of tau aggregates (17). Through usage of brain homogenates from patients with different tauopathies, the aggregatable tau RD (repeat domain; previously identified as an aggregation core) was seeded to form a multitude of strains in a cell culture model (18). Aggregate morphology and patterns of high-molecular weight aggregates observed via SDD-AGE (semi-denaturing detergent agarose gel electrophoresis) could clearly delineate between these different strains (17). In summation, the prion-like tau protein is amyloidogenic, forms disease-specific strains, spreads intercellularly, and can be artificially propagated between organisms – indicating the prion-like nature of tauopathies.

The pathology of PD is marked by inclusions of α-syn in the form of Lewy bodies as well as other aggregates (3). Aggregation of α-syn often takes place in the form of amyloids and different fibril strains with variable properties have been identified (19, 20). Moreover, these amyloid aggregates are capable of seeding α-syn aggregation when injected in mice brains (21, 22). Intercellular transmission of α-syn misfolding and aggregation has been well-documented and underpins modern models of PD and other synucleinopathies (i.e., diseases with α-syn pathology) (23). The structure and histology-based post-mortem Braak staging of PD and the observable progression of brain structures affected in PD reflect the likelihood that prion-like neuron-to-neuron transmission drives α-syn pathology (24). Significantly, recent studies have also found that relatively less invasive introduction of α-syn fibrils (e.g., through the peripheral nervous system or even orally or intravenously) can induce neurological disease in mice (25, 26). In summary, α-syn recapitulates most of the characteristics of a bona fide prion protein and the prion-like spreading of α-syn aggregation underlies the neurodegeneration observed in PD and other synucleinopathies.

As mentioned earlier, ALS has a number of prion-like proteins – including HNRNPA1/2B1, FUS, SOD1, and TDP-43 – involved in its pathology (27). RNA-binding proteins account for many of these proteins including all of those mentioned other than SOD1. The presence of regions with prion-like sequences in many RNA-binding proteins – particularly of those involved in neurodegeneration – has been well-documented (28, 29). In fact, HNRNPA1 and HNRNPA2B1 present with disease-specific mutations that increase their predicted amyloidogenicity and enhances the aggregation of both proteins (30). A predicted prion-like domain (PrLD) from HNRNPA2B1 can even act as a prion domain as part of a fusion protein in a yeast assay. Another protein, FUS has been shown to phase separate as well as form amyloid

aggregates (31–34). Moreover, mutant FUS can form aggregates that recruit wild-type (WT) FUS to also aggregate – a behavior reflective of confirmed prion proteins (33). A fourth RNA-binding protein, TDP-43 shows similarities to the amino-acid composition of yeast prion proteins (28). Not only can TDP-43 phase separate and form amyloid aggregates, but these aggregates are intercellularly transmissible in cultured cells (35–39). Similarly to Aβ and tau, significant evidence for the prion-like disposition of TDP-43 comes in the form of injection of diseased human brain homogenates into the brains of transgenic mice, where the homogenates seeded further TDP-43 aggregation (40). Lastly, as the first protein linked to inherited forms of ALS, SOD1 has been the focus of much study (41). Inclusions of SOD1 can be found in both familial and sporadic forms of the disease, and the protein is capable of forming amyloid fibrils *in vitro* (42–44). The intercellular prion-like transmission of SOD1 aggregation has been demonstrated in cell culture and the ability for mutant SOD1 to recruit WT SOD1 – as with other prion-like proteins – was shown (45, 46). Once again, SOD1 is also a protein whose aggregation can be seeded in mice through injection of homogenates from diseased spinal cords (47). Finally, as discussed with tau, evidence of distinct propagatable strains of SOD1 aggregates exists – further supporting the prion-like nature of SOD1 (48).

The range of prion-like proteins involved in neurodegenerative diseases displays the important and significant role of this mechanism in human disease. Furthermore, the understanding of prion formation and propagation holds the promise of development of new efficacious therapeutics for these currently incurable diseases. As such, examination of any and all prion and prion-like proteins – whether in humans, model organisms, or elsewhere – is bound to elucidate critical information.

Prions and Cancer

As described in the previous section, prion and prion-like proteins have primarily been implicated in neurodegenerative diseases. However, recent studies have implicated prion-like aggregation as an important pathological process in cancers. Cancer, the unregulated growth and division of aberrant cells results from mutations in a plethora of genes. There exists a perspective of neurodegeneration and cancer as two ends of a spectrum of disease – with uncontrolled cell death at one end and unregulated cellular proliferation at the other, respectively. This view has been supported by longitudinal studies that the risk of particular neurodegenerative diseases (e.g., AD) are negatively correlated with some cancers (49, 50). Additionally, there appears to be a transcriptional basis for this divergence between the two sets of diseases (51, 52). As such, protein aggregation, let alone prion-like protein behavior, has typically been perceived as the realm of degenerative diseases – not proliferative diseases.

On the other hand, there exists several similarities between cancer and neurodegenerative diseases. Foremost among these likenesses is the age-related risk of the diseases. Additionally, disruption of homeostasis is the cornerstone of any disease, though traditionally, neurodegeneration has been seen to revolve around aberrant proteostasis and cancer ultimately revolving around genetic disorder (53, 54). However, both neurodegeneration and cancer do have notable related mechanisms across both proteostasis, DNA damage, and cell cycle dysregulation among others (55). Indeed, recent research has revealed that even aberrant protein aggregation – typically tied to cytotoxicity and cell death – has an important role in the formation of cancers (56).

The most commonly mutated gene across all cancer types is the gene encoding p53 (57). As a tumor suppressor protein, p53 has been commonly termed the ‘guardian of the genome’ due

to its functioning in the cell cycle, DNA damage repair, and apoptosis. Intriguingly, p53 and other genes commonly mutated in cancers sometimes present with hot-spot residues – i.e., amino-acid residues that have a high frequency of mutations detected in cancer (58). In the case of p53, some of these hot-spot mutations have been found to associate with a toxic gain-of-function as well as changes in physical interactions with other proteins (59). This observed toxic gain-of-function exhibits dominant-negative behavior through conformational changes of wild-type p53 expressed alongside the mutant (60, 61). This dominant-negative, gain-of-function behavior mirrors that of prion-like proteins in neurodegenerative diseases.

Mutant p53 has the capability to aggregate, form amyloids, and recruit WT copies of the protein to co-aggregate (62–64). Initially, examination of breast cancer patient samples displayed a co-localization of p53 and amyloids (as recognized by the amyloid oligomer-specific A11 antibody) (65). Further investigation yielded that p53 could aggregate *in vitro* and form amyloid fibrils as reported by thioflavin T (ThT) fluorescence (66). Intriguingly, aggregated p53 was found to cause other proteins including p63 and p73 to co-aggregate, providing a mechanism through which mutant p53 can achieve a toxic gain-of-function (67). Additional study has revealed that p53 can act as a prion in a yeast model system and in doing so, forms amyloid aggregates and disrupts normal p53 functioning (68). Also, in yeast cells, p53 can phase separate upon cellular stress – indicating there may be a range of p53 proteostatic states important to its function and disfunction. Given these various traits of mutant p53 and its aggregation abilities, some have termed this behavior as prion-like; however, this application of terminology and its significance for the pathology and treatment of cancers has not been without controversy (56, 69, 70).

Perhaps the most convincing evidence of the prion-like aggregation of p53 having a significant role in cancer development rests in successful experiments to target tumorigenesis

through prevention of p53 aggregation. High-grade serious ovarian carcinomas overwhelmingly have mutations in p53 (71). Researchers found that the resistance of these cancers to platinum treatment correlates with the presence of p53 aggregates – due to likely gain-of-function changes (72). Intriguingly, the usage of a peptide designed against a predicted amyloid core sequence of p53 (residues 252-258) not only blocks aggregation of the protein in a cell line model but also causes cancer cell death and decreases the size of tumor xenografts in a mouse model (73). A similar study found that a mutant p53 containing a charged residue that interrupts its predicted amyloid core prevented mutant p53 from recruiting and aggregating with wild-type p53 (74). These results highlight the opportunity that exists for new cancer therapeutic strategies focusing on preventing or disrupting p53 aggregates or any other protein that may aggregate in a relevant manner in cancer. As such, the expansion of the prion concept beyond the realm of neurodegeneration highlights the importance of understanding the biological phenomenon.

Prions across Life

Prions and prion-like proteins can be found in organisms of all types. As already touched upon, PrP in both non-prion and prion forms can be found in humans as well as myriad mammals including sheep, deer, and cows. In these species, prion disease manifests as scrapie, chronic wasting disease (CWD), and BSE. Although all these organisms can succumb to the neurodegeneration brought upon them by PrP^{Sc}, there does exist species barriers preventing infectivity between some of them (75). However, this barrier may be less extant than previously thought as the barrier has been broken in some experimental models (76, 77). This barrier – regardless of actual permeability – demonstrates the existence of PrP^{Sc} variants as well as the range of species that can be affected by prion disease. Although we have largely discussed prions in the

context of incurable neurological diseases, the prion concept also encompasses a range of prion-like proteins involved not in illness but in normal cellular functioning.

Humans suffer from a multitude of degenerative diseases underpinned by prion-like proteins; however, a vital cellular infection defense system relies on the performance of a prion-like protein (78). The mitochondrial antiviral-signaling protein (MAVS) acts in the signaling cascade triggered upon detection of viral RNA (79). During this process, the protein RIG-I (retinoic acid-inducible gene I) binds to MAVS and induces the formation of MAVS aggregates (80). Furthermore, *in vitro* experiments showed that MAVS can form amyloid fibrils. When investigated through usage of the yeast-based Sup35 *de novo* prion formation assay, the CARD (caspase activation and recruitment domain) contained within MAVS could act as a prion domain (PrD) (81). The prion formed by the $MAVS_{CARD}$-$Sup35_{C}$ (the functional domain of Sup35) fusion protein could be stably inherited, transmitted by cytoplasm only, and form high-molecular weight (MW) SDS-resistant aggregates. Binding of RIG-I can also induce this fusion protein to prionize – reflecting the model of the interaction of RIG-I with the full-length MAVS protein. Taken together, these results indicate that MAVS is indeed a prion-like protein. Another protein involved in the immune response, ASC (apoptosis-associated speck-like protein containing a CARD) also contains a prion-like domain (81). The pyrin domain (PYD) of ASC acted as a capable PrD when fused with $Sup35_{C}$ and could also form infectious amyloid fibrils *in vitro*. The combination of the PrLDs in MAVS and ASC suggest that prion-like aggregation of these proteins importantly functions in the viral response signaling cascade (78).

Functional prion-like proteins can also be found in lower eukaryotic organisms. The sea slug *Aplysia* acts as an important model for investigations of learning and memory. In *Aplysia*, the cytoplasmic polyadenylation element binding (CPEB) protein plays an important role in regulating

protein synthesis in neurons as well as supporting the strengthening of synaptic connections via long-term facilitation (82). Intriguingly, this protein contains a Q-rich domain capable of forming at least two conformational states – one diffuse and one aggregated – in a yeast model (83). The aggregated, prion-like state can also be inherited in this yeast model. Further investigation demonstrated that additional aggregated amyloid states can be established with the CPEB Q-rich domain, mimicking the variants seen with prion proteins (84). These amyloid states of Aplysia CPEB have since been confirmed to occur in neuronal cells as well and correlate with long-term facilitation at synapses (85). The neuronal CPEB protein is conserved into the fruit fly, *Drosophila melanogaster*. And as may be predicted, the *Drosophila* CPEB homolog, Orb2 contains a PrLD, can form amyloid fibrils, and the oligomerization of Orb2 plays an important role in synaptic strengthening (86, 87). Carrying further up the ladder of eukaryotes, the *Aplysia* CPEB and *Drosophila* Orb2 homolog in mice, CPEB3 has also been implicated as a prion-like protein. Indeed, CPEB3 contains a disordered PrLD, can form amyloid fibrils, and aggregates in neurons (88). Functionally, the prion-like aggregation of CPEB3 appears to play a significant role in the learning and memory of mice (89). These data establish a lineage of a prion-like protein in critical neuronal functioning throughout higher eukaryotes.

The prion concept extends even to plants. The common model organism, *Arabidopsis thaliana* has been shown to contain a prion protein. This protein, Luminidependens (LD) was first identified as a candidate based on evaluation with an algorithm which computes the likelihood of a given protein being prionogenic. The predicted PrD of LD can form foci when expressed as a GFP fusion in yeast. Furthermore, the LD PrD demonstrated its ability to adopt a prion form ([LD^+]) when fused with $Sup35_C$ and assayed using the widely used Sup35 assay. [LD^+] could be cytoplasmically inherited and cured through inhibition of the molecular chaperone Hsp70 but not

Hsp104. Although the LD PrD has been shown to capably prionize in an experimental context, whether the full-length protein acts in a prion-like fashion in *Arabidopsis* has not yet been elucidated. However, the fact that this single species of plant contains at least one capable PrD supports the possibility that prion and prion-like proteins may exist across other plant life.

Prions have been identified in a variety of fungal species. Most of these prion proteins were discovered in the budding yeast, *Saccharomyces cerevisiae* and will be reviewed in the next section. One well-studied fungal prion from a species other than *S. cerevisiae* is the [Het-s] prion from the filamentous fungus *Podospora anserina* (90). The het-s protein acts in determining heterokaryon (i.e., existence of two different nuclei in one cytoplasm) incompatibility (91, 92). In the non-prion state, strains expressing het-s have incompatibility with strains expressing het-S (another highly related protein). However, with the [Het-s] prion form, strains expressing het-s are neutral towards strains expressing het-S. The [Het-s] phenotype relies on the expression of the het-s protein, can convert non-prionized het-s, and can be found more frequently after overexpression of het-s – indicating the prion nature of [Het-s] (90). Furthermore, het-s was found to aggregate in cells harboring [Het-s] and a proteinase K-resistant region of het-s forms amyloid in vitro (93, 94). As such, het-s has been thoroughly shown to form an infectious, amyloid-based prion, [Het-s] in a filamentous fungus.

While there has been much investigation of prions in eukaryotes, the concept has also been expanded to include proteins in bacteria. One such protein is the transcription terminator Rho from a strain of *Clostridium botulinum* (Cb-Rho). Cb-Rho was initially identified as a possible prion domain-containing protein alongside other bacterial proteins through the usage of a prion prediction algorithm (95). Indeed, the candidate PrD from Cb-Rho demonstrated amyloid formation and deletion of the PrD from full-length Cb-Rho prevented amyloidogenesis (95, 96).

Usage of the Sup35 *de novo* prion formation assay (which is discussed in more detail in the next section) confirmed the prionogenic potential of the Cb-Rho PrD. Lastly, this prion forming ability extended to the full-length Cb-Rho protein when expressed in a non-*Clostridium botulinum* bacterium (*Escherichia coli*) and the prion state could be successfully propagated for many generations, displaying its heritability (96). Thus, the prionogenicity of Cb-Rho demonstrates the extension of the prion phenomena into prokaryotes – and many additional bacterial prions may exist.

Recent research has shown that a prion exists in a virus (97). The LEF-10 protein is a late expression factor of the autographa californica multiple nucleopolyhedrovirus, which is a virus that infects insects (98). In insect cells, LEF-10 can form stable aggregates as observed via fluorescent microscopy and SDD-AGE. These aggregates prevent LEF-10 from performing its normal function and activating the late genes of the virus – hindering the production of new viral particles (97). This mechanism allows the LEF-10 prion to act as a functional element. Furthermore, LEF-10 displayed the ability to act as a prion in the Sup35 *de novo* prion formation assay in yeast – where LEF-10 was fused to the middle supporting domain and the functional domain of the Sup35 protein, together known as $Sup3_{MC}$. By testing various truncations of LEF-10, a small region of either 41 or 23 amino-acids was found that could act as a bona fide prion domain in this assay (97). Interestingly, the prion formed by these LEF-10-$Sup35_{MC}$ fusion proteins formed SDS-resistant, high molecular weight aggregates and could be cured by inhibition of Hsp104 – much like endogenous yeast prions. Prior work has detailed thousands of potential viral prion proteins as determined through computational prediction – indicating that LEF-10 may be the first of many identified viral prion proteins (99).

Yeast Prions

Saccharomyces cerevisiae, also known as the budding yeast or baker's yeast, harbors many endogenous prions that act as epigenetic elements. Currently identified yeast prions include [*PSI*$^+$], [URE3], [*RNQ*$^+$], [*SWI*$^+$], [*OCT*$^+$], [*GAR*$^+$], [*MOD*$^+$], [*NUP100*$^+$], [*MOT3*$^+$], [*SMAUG*$^+$], and [*ESI*$^+$] (100–107). (Yeast prions are typically denoted with brackets to signify their non-Mendelian inheritance and capitalization and italics indicate their dominance when inherited.) The aforementioned prions have protein determinants of Sup35, Ure2, Rnq1, Swi1, Cyc8, Pma1/Std1, Mod5, Nup100, Mot3, Vts1, and Snt1, respectively.

Table 1.1. Currently identified yeast prions.

Prion	Protein	Q/N Rich?	Protein Function	Source
[*ESI*$^+$]	Snt1	No	Histone deacetylase scaffold	(105)
[*GAR*$^+$]	Pma1/Std1	No	Proton pump / glucose signaler	(108)
[*MOD*$^+$]	Mod5	No	tRNA isopentenyl transferase	(102)
[*MOT3*$^+$]	Mot3	Yes	Transcription factor	(109)
[*NUP100*$^+$]	Nup100	Yes	Nucleoporin component	(103)
[*OCT*$^+$]	Cyc8	Yes	Transcription factor	(101)
[*PSI*$^+$]	Sup35	Yes	Translation termination factor	(110)
[*RNQ*$^+$]	Rnq1	Yes	Unknown	(107)
[*SMAUG*$^+$]	Vts1	No	Post-transcriptional regulator	(104)
[*SWI*$^+$]	Swi1	Yes	Chromatin remodeling complex subunit	(100)
[URE3]	Ure2	Yes	Transcription factor	(111)

Yeast prions can be divided into a few different groups based on the amino-acid composition of their protein determinants and the structural characteristics of their prion aggregates. The first grouping of yeast prions – [*PSI*$^+$], [URE3], [*RNQ*$^+$], [*SWI*$^+$], [*MOT3*$^+$], [*NUP100*$^+$] – consists of those that have been experimentally demonstrated to form amyloids and with protein determinants rich in glutamine (Q) and asparagine (N). Additionally, the [*OCT*$^+$] prion has a Q-rich protein determinant and is predicted to form amyloids; however, this amyloidogenicity has not been experimentally tested yet. The second grouping of yeast prions –

[*MOD*$^{+}$] – includes those prions which form amyloids but whose protein determinants are not enriched in Q/N residues. Lastly, the third grouping of yeast prions – [*GAR*$^{+}$], [*SMAUG*$^{+}$], [*ESI*$^{+}$] – contains those prions that do not form amyloids.

The first identified yeast prions were [URE3], formed by the Ure2 protein and [*PSI*$^{+}$], whose protein determinant was later determined to be the translation termination factor, Sup35 (106). Originally characterized as simply a non-Mendelian cytoplasmic element, [*PSI*$^{+}$] was associated with a phenotype of inefficient termination of translation (110, 112). Further research theorized that [*PSI*$^{+}$] was an alternative form of Sup35, and indeed, cells harboring [*PSI*$^{+}$] were found to exhibit aggregates of the Sup35 protein (113). Moreover, the overexpression of Sup35 leads to an increase in the appearance of [*PSI*$^{+}$] – a characteristic of prionogenesis now widely accepted and experimentally used (114). The important positioning of the [*PSI*$^{+}$] protein determinant, Sup35 as a translation termination factor allows for powerful phenotypic assays for prion formation that are unavailable in any other model system.

The Sup35 assay for *de novo* prion formation takes advantage of the stop codon read-through that occurs when Sup35 prionizes (107). In this assay, a protein of interest is fused with the region of Sup35 providing the translation termination – Sup35$_{C}$. This results in a chimera where the Sup35 PrD, Sup35$_{N}$ has been replaced with a candidate PrD. If the fusion protein can prionize, some yeast isolates will display impeded translation termination as prion aggregates sequester the fusion. This reduction in activity can be ascertained via a premature stop codon in the *ADE1* gene. Proper translation termination, as in non-prion isolates, leads to the Ade1 enzyme not being fully translated – preventing the yeast from synthesizing adenine and causing a build-up of a precursor that makes the yeast red in color. Meanwhile, prionized fusion protein allows for read-through to occur, leading to translation of the Ade1 enzyme – permitting the yeast to synthesize adenine,

decreasing accumulation of the adenine precursor, and leading to colonies of white or pink coloration. This assay plays a critical role in two of the projects presented in this thesis.

As a yeast prion, [*PSI*$^{+}$] relies on the molecular chaperone machinery for propagation. The most important protein in this system, Hsp104 acts to break-up larger protein aggregates – generating seeds that can be readily distributed to daughter cells (115–117). "Curing" or elimination of yeast prions like [*PSI*$^{+}$] can occur through the disruption of the normal equilibrium of Hsp104 – either by deleting its gene, overexpressing it, or inhibiting its functioning through treatment with guanidine hydrochloride (GdnHCl) (115, 118). Other molecular chaperones including Hsp70s (e.g., Ssa1 and Ssa2), Hsp40s (e.g., Ydj1 and Sis1), and nucleotide exchange factors (NEFs; e.g., Sse1) also have roles in [*PSI*$^{+}$] propagation and lead to loss or destabilization of the prion (119–123). Perhaps most interestingly, a range of overexpressed proteins can act as [*PSI*$^{+}$] inducible (Pin^{+}) factors (i.e., the rate of [*PSI*$^{+}$] formation increases in their presence) (124–126). Other yeast prion proteins such as Rnq1, Cyc8, and Swi1 were both first noted as potentially prion-like due to these experiments; in fact, Rnq1, Cyc8, and Swi1 all form their own prions, [*RNQ*$^{+}$], [*CYC*$^{+}$], and [*SWI*$^{+}$] respectively.

As the other yeast prion discovered early on, [URE3] was first identified as a cytoplasmically inherited epigenetic element in yeast (111). Yeast with [URE3] were able to grow on media containing ureidosuccinate with ammonium, which normally prevented uptake of ureidosuccinate. This phenotype could be recapitulated by deletion of the URE2 gene; however, such mutants could not present with the [URE3] element – indicating that Ure2 was necessary for [URE3] (127). Seminal work posited that [URE3] was the result of prionization of the Ure2 protein (106). The evidence greatly supports the prion nature of [URE3] – a protein-only basis, curable

through disruption with GdnHCl, non-Mendelian dominant inheritance, transmitted via cytoplasmic mixing, and overexpression of the Ure2 increases the rate at which [URE3] appears.

Further research has detailed the [URE3] prion. Just like [*PSI*$^+$] and many other yeast prions yet to be discussed, the protein determinant of [URE3], Ure2 forms aggregates in prionized cells (128). The [URE3] prion also possesses sensitivities to the disruption of molecular chaperones other than Hsp104. Overexpression of the Hsp70 Ssa1 can cure cells of [URE3]; however, overexpression of another Hsp70, Ssa2 does not (129). In contrast, mutation of Ssa2 can cure the prion whereas mutation of Ssa1 cannot (130). [URE3] is also sensitive to overexpression of the Hsp40, Ydj1 or NEF, Sse1 (131, 132). These differences in susceptibility highlight the complexity and variation of the interactions between molecular chaperones and yeast prions such as [URE3]. In another layer of complexity, heterologous prion interactions were observed between [URE3] and [*PSI*$^+$] with [URE3] acting as a Pin$^+$ factor and [URE3] can also induce [*RNQ*$^+$] (126, 129). Intriguingly, Ure2 normally acts as a transcription factor and regulates a variety of genes targeted by other identified yeast prion proteins (e.g., Cyc8, Swi1, Mot3) (133, 134). These numerous direct and indirect interactions between Ure2 and [URE3] and other prion proteins suggest a possible regulatory network of yeast prions. On the other hand, the growth-hampering and/or toxic phenotypes that can occur with [URE3] and [*PSI*$^+$] suggest that these prions may simply be diseases in yeast – reflecting the nature of many prion and prion-like proteins in humans (135).

Unlike the protein determinants of [*PSI*$^+$] and [URE3], Sup35 and Ure2, the yeast prion protein Rnq1 has an unknown function outside of forming the [*RNQ*$^+$] prion. As the name may suggest, Rnq1 has an amino-acid sequence rich in N and Q residues. The [*RNQ*$^+$] prion was first classified as a [*PIN*$^+$] prion – alongside other potential prion proteins – due to its discovery as a [*PSI*$^+$] inducer (124–126, 136). Furthermore, testing of a fusion protein that consisted of the Rnq1

PrD in place of the Sup35 PrD confirmed that Rnq1 was indeed a prion (107). This study also established the now widely used Sup35 assay for *de novo* prion formation. In terms of curability, [*RNQ*$^+$] can be eliminated by Hsp104 disruption via GdnHCl treatment or deletion of the Hsp104 gene. Deletions of an Hsp70, Sis1 also lead to a loss of [*RNQ*$^+$] (122, 137). Like other well-characterized prion proteins, Rnq1 produces a range of prion variants as observed by both phenotypic assays as well as biochemical (e.g., SDD-AGE) (138). Interestingly, further research has revealed that [*RNQ*$^+$] can induce not only [*PSI*$^+$] but also [URE3] and [*SWI*$^+$] (139, 140). This broad induction of other prions by [*RNQ*$^+$] and the lack of a known function for the Rnq1 protein suggests that Rnq1 may primarily act as an epigenetic regulator, working through the induction of yeast prions (141).

The [*MOT3*$^+$] prion was initially identified in a wide-ranging screen that examined candidate yeast proteins chosen by a prediction algorithm (109). This algorithm checked for proteins whose amino-acid compositions were similar to already characterized prion proteins. (Additional technical details of the algorithm are detailed in a later section.) Mot3, a transcription factor was one of many yeast proteins determined to contain a candidate PrD. The Mot3 PrD displayed formation of aggregate foci in yeast (as visualized by fusion with YFP), formation of high-molecular weight, SDS-resistant species (as assessed by SDD-AGE), and formation of amyloids (as determined by ThT binding) (109). Furthermore, a Mot3 PrD-Sup35$_C$ fusion protein exhibited formation of prion and non-prion states as assayed by the Sup35 prion formation assay. Additional experiments confirmed the existence of [*MOT3*$^+$] by usage of a custom Mot3 reporter system reliant on the *DAN1* promoter – which Mot3 represses under non-prion conditions.

Added investigation into Mot3 and the [*MOT3*$^+$] prion using the *DAN1* reporter system revealed that adoption of the prion led to yeast gaining multicellular features (133). This driving

of multicellularity by [*MOT3*$^{+}$] was determined to be a result of Mot3 regulation of flocculin (*FLO*) genes. Interestingly, multiple yeast prion proteins involved in transcription target the *FLO* genes – indicating a likely complicated prion-based regulation of these genes (133, 134). Deletion of Mot3 was found to not fully recapitulate [*MOT3*$^{+}$] phenotypes. This discrepancy likely results from [*MOT3*$^{+}$] not only abolishing the normal functioning of Mot3 through its aggregation but also recruiting other transcriptional proteins to aggregate alongside Mot3 (133). In fact, a similar situation occurs with the [*SWI*$^{+}$] prion as discussed in a later section (134). Interestingly, the [*MOT3*$^{+}$] prion exhibits changes in response to environmental cues such as ethanol concentration and hypoxia. As such, not surprisingly, [*MOT3*$^{+}$] has been documented as naturally occurring in yeasts involved in fermentative conditions (e.g., wine production) (133, 142). The behavior of [*MOT3*$^{+}$] demonstrates that yeast prions can act as functional elements that provide beneficial adaptations.

Also identified as a possible prion protein in the screen conducted by Alberti et al., Nup100 normally functions as a component of the nuclear pore complex (NPC) (109). Multiple NPC proteins contain intrinsically disordered domains enriched in phenylalanine and glycine (143, 144). Furthermore, some of these proteins specifically contain repeats of glycine-leucine-phenylalanine-glycine (GLFG). Nup100 is one such protein. The sequence of Nup100 also contains an enrichment of asparagine and glutamine residues. Additional investigation following the more widely focused Alberti et al. screen determined that Nup100 could form aggregates that sequester other NPC proteins and that a 100 residue-long region of Nup100 could form amyloid fibrils *in vitro* (103). Moreover, this region could also act as a bona fide prion domain when fused with Sup35$_{C}$ to allow for a phenotypic assay for *de novo* prion formation. Lastly, full-length Nup100

was shown to form inheritable prion aggregates *in vivo* that were curable by treatment with GdnHCl. This prion formed by Nup100 is known as [*NUP100*$^{+}$].

The Q-rich protein Cyc8 can form the yeast prion [*OCT*$^{+}$]. First identified as a prion protein candidate, Cyc8 could be overexpressed to increase the rate of de novo [*PSI*$^{+}$] formation (126). Through overexpression of a predicted Q-rich PrD contained in Cyc8, cells containing potential prionized Cyc8 ([*OCT*$^{+}$]) were obtained (101). The prion-containing cells demonstrated enhanced growth on lactate media as all cells were *cyc1Δ*, eliminating the only source other than Cyc8 for cytochrome *c* – a necessity for growth on non-fermentable carbon sources. Furthermore, [*OCT*$^{+}$] yeast demonstrated enhanced flocculation (i.e., the clumping of cells in liquid media). These phenotypes were found to be dominant (101). As with many other yeast prions, [*OCT*$^{+}$] can be cured by disruption of Hsp104 – whether by treatment with GdnHCl or expression of a dominant-negative Hsp104 mutant. Although Cyc8 has been shown to aggregate in [*OCT*$^{+}$] yeast, it has not been determined if Cyc8 forms amyloid – making this protein unique amongst the group of Q/N-rich yeast prion proteins.

Although a distinctive feature of many yeast prions is a sequence rich in Q/N residues, there are a growing number of identified yeast prions that do not have such enrichment. First amongst these prions is [*MOD*$^{+}$], which is formed by the Mod5 protein. Initially identified through a yeast screen, Mod5 demonstrated the ability to promote the formation of [*PSI*$^{+}$] and prionization of a chimeric protein (Q_{62}-Sup35, a 62-residue-long polyglutamine peptide fused with $Sup35_{\Delta 1\text{-}40}$). The Mod5 protein can form amyloid fibrils as assayed via ThT binding and transmission electron microscopy (TEM). Furthermore, transformation of naïve yeast with Mod5 amyloids could generate the [*MOD*$^{+}$] prion and aggregation of Mod5 in [*MOD*$^{+}$] cells could be visualized with a Mod5-GFP fusion protein. Cytoduction confirmed that [*MOD*$^{+}$] was inheritable in yeast by

only cytoplasmic mixing. And as with the yeast prions discussed thus far, [*MOD*$^{+}$] could be cured by inhibition of Hsp104 via GdnHCl treatment. In terms of functional yeast phenotypes, [*MOD*$^{+}$] cells display resistance to various drugs including nocodazole (a microtubule disruptor) and fluconazole (an antifungal medication). These acquired resistances on conversion from [*mod*$^{-}$] to [*MOD*$^{+}$] exemplifies the potential for yeast prions to provide beneficial adaptations on a quicker timescale than traditional genetic mutations.

A peculiar yeast prion known as [*GAR*$^{+}$] was initially identified through its phenotype of cells being able to use non-glucose carbon sources in the presence of glucose (108). The prion was determined to not have one but two protein determinants, Pma1 and Std1 and be curable through disruption of Hsp70. The overexpression of Std1 could induce [*GAR*$^{+}$] but deletion of Std1 did not eliminate [*GAR*$^{+}$]. Meanwhile, overexpression of Pma1, a proton pump was found to increase the rate at which [*GAR*$^{+}$] appeared. If the disordered cytosolic N-terminus of Pma1 was deleted, no impact on the prion was observed; however, if both this N-terminus and Std1 were both deleted, [*GAR*$^{+}$] would be eliminated. Thus, it was concluded that the [*GAR*$^{+}$] prion relies on both Pma1 and Std1 as protein determinants and is driven by changes in conformations and oligomerization of the two proteins.

Perhaps just as interesting as the dual protein basis of [*GAR*$^{+}$] is that this prion can be induced by chemical signals secreted by bacteria. Growth of bacteria (e.g., *Staphylococcus hominis*) adjacent to yeast on solid media leads to [*GAR*$^{+}$] induction. In fact, usage of filter-sterilized liquid media that had previously hosted bacteria can also lead to induction of [*GAR*$^{+}$] – indicating that a bacterial chemical secretion remaining in the media drives the induction. This ability for bacteria to induce [*GAR*$^{+}$] extends to a wide range of bacterial species. The ability to induce [*GAR*$^{+}$] may benefit bacteria by reducing the amount of ethanol produced by the yeast.

Moreover, [*GAR*$^{+}$] yeast demonstrate a growth advantage in multiple experimental setups compared to [*gar*$^{-}$] cells. Once again, a yeast prion exhibits the potential to provide measurable benefits to the cells harboring the prion. In fact, numerous wild yeast isolates harbor the [*GAR*$^{+}$] prion and a similar epigenetic trait can be found in other fungal species.

More recently, a large screening experiment has identified numerous potential yeast prions that do not form amyloids and are largely not Q/N-rich (145). By performing transient overexpression of most open reading frames (ORFs) in the yeast genome, Chakrabortee et al. found that 46 proteins expressed in this manner generated epigenetically inherited phenotypes. These proteins were enriched in intrinsically disordered regions (IDRs) based on evaluation with the DISOPRED3 prediction algorithm (see next section for more information regarding DISOPRED3) (146). One protein identified was Vts1. The transient overexpression of Vts1 led to yeast with greater sensitivity to ultraviolet (UV) irradiation and this epigenetic trait was termed [*VTS1*$^{+}$]. Further experimentation found that [*VTS1*$^{+}$] was not dependent on Hsp104 for propagation but was dependent on Hsp70 (145).

Additional, targeted investigation of Vts1 revealed that instead of forming amyloids as other yeast prion proteins, Vts1 forms gel-like condensates (104). The IDR of Vts1 proved to be sufficient and necessary for formation of these condensates. These higher-order condensates were determined to be inheritable and reversible – defining characteristics of prions. Thus, this prion formed by Vts1 was termed [*SMAUG*$^{+}$], named after the protein's name in higher eukaryotic organisms (e.g., humans, Drosophila). Examination of [*SMAUG*$^{+}$] in yeast revealed a prion-based regulation of numerous gene transcripts through modulation of Vts1 binding to loops in mRNAs. This functional change results in phenotypic outcomes including on carbohydrate metabolism and

meiosis and sporulation (147). The [*SMAUG*$^+$] prion has been identified in wild yeast isolates from a multitude of different environments as well.

Another yeast prion first identified by the transient overexpression screen conducted by Chakrabortee et al. was that formed by Snt1 (145). Normal functioning of Snt1 sees it acting as a scaffold protein in the Set3C histone deacetylase complex (148). In the initial screening experiments, yeast that had experienced transient overexpression of Snt1 had greater zinc tolerance. This acquired trait was shown to be curable by inhibition of Hsp90. Once again, further experimentation demonstrated that Snt1 aggregates underlie the prion and can be transmitted from cell-to-cell (105). This prion, [*ESI*$^+$] impacts the transcriptional profile of yeast through the expression of genes in sub-telomeric domains. In doing so, the cells gain greater resistance to stresses such as the aforementioned zinc. Interestingly, the [*ESI*$^+$] prion state seems to be driven in part by the phosphorylation of IDRs within Snt1 (105). The existence of [*SMAUG*$^+$] and [*ESI*$^+$] point towards the likely existence of additional yeast prions that are not Q/N-rich and contain IDRs and have yet to be fully investigated.

Lastly, the yeast prion at the heart of the work presented here, [*SWI*$^+$] was initially discovered by our lab (100). It was the fourth prion identified in *S. cerevisiae*. The initial identification of [*SWI*$^+$] was made possible through its induction of [*PSI*$^+$] as a Pin$^+$ factor and the knowledge that *swi1Δ* cells exhibited poor growth on non-glucose carbon sources (e.g., raffinose, galactose, glycerol) (126, 149). Thus, yeast containing both [*PSI*$^+$] and the candidate [*SWI*$^+$] were produced through Swi1 overexpression. By exploiting the unique susceptibility of [*PSI*$^+$] to curing via Hsp104 overexpression, [*SWI*$^+$]-only cells were isolated (115). From here, further characterization confirmed the existence of [*SWI*$^+$] – Swi1 aggregates in [*SWI*$^+$] cells, the prion cannot be maintained in the absence of Swi1, and the prion can be cytoplasmically inherited (100).

The protein determinant of [*SWI*$^+$], Swi1 functions as a component of the SWI/SNF chromatin remodeling complex, which modulates the expression of more than 15% of yeast genes (150, 151). This modulation results in a range of phenotypes including the previously mentioned poor growth on non-glucose carbon sources and abolishment of multicellular features (e.g. flocculation, pseudohyphae) (Figure 1.1) (100, 134). This regulation of multicellularity occurs through targeting of the *FLO* genes by Swi1. As such, aggregation of Swi1 and co-aggregation of Swi1-interacting proteins plays an important role in determining these phenotypes (134). Intriguingly, other yeast prion proteins including Cyc8 and Mot3 exhibit transcriptional controls on some of the same genes – suggesting a yeast prion regulatory network (152). If a network of functional prions regulates multicellular features in yeast, this situation would suggest that there exist benefits to forming these prions. In fact, some of the features observed in yeast harboring [*SWI*$^+$] may play a role in increasing perceived fitness by allowing quick adaptation to environmental cues (153). Finally, a recent study revealed that the transcriptome of [*SWI*$^+$] yeast significantly differs from *swi1Δ* yeast (151). This result indicates that [*SWI*$^+$] changes the transcriptional profile not just by removing normal Swi1 function but also by interactions with other proteins – potentially through co-aggregation as noted above.

In addition to relationships at the transcriptional level, [*SWI*$^+$] can engage in heterologous interactions with other yeast prions. As already mentioned, [*SWI*$^+$] can act as a Pin$^+$ factor and induce [*PSI*$^+$] (126). Conversely, the presence of [*PSI*$^+$] or [*RNQ*$^+$] can enhance [*SWI*$^+$] formation (140). While data on the interaction of [*SWI*$^+$] with [*MOT3*$^+$] has not yet been published, there stands to be potentially informative changes in propagation and formation of the two prions as they work largely opposite of each other in modulating multicellular features (133, 134). In terms of dependencies on molecular chaperones, [*SWI*$^+$] displays many sensitivities as already discussed in

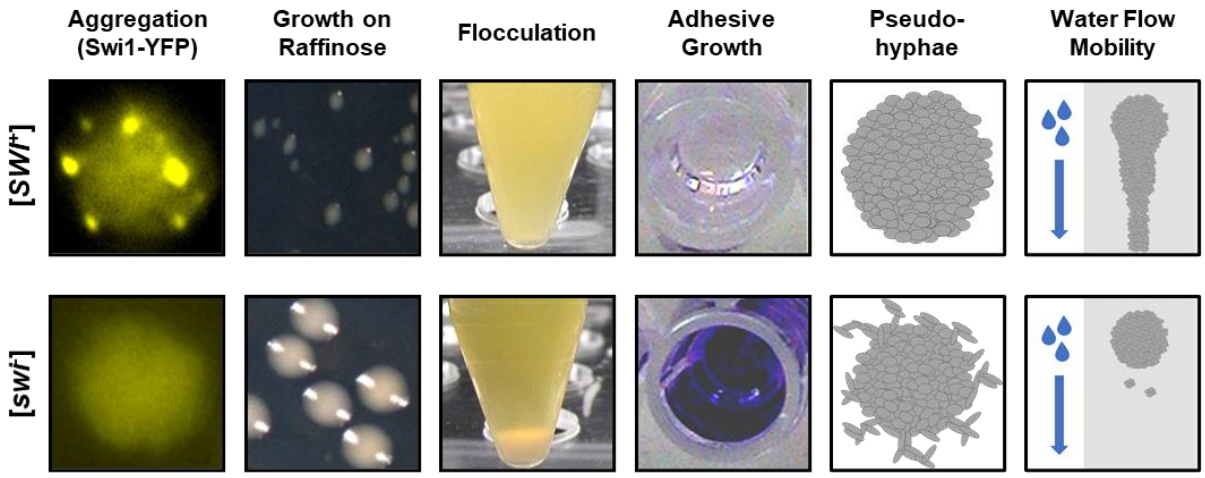


Figure 1.1. Phenotypes of yeast containing [*SWI*$^+$].

Yeast with the [*SWI*$^+$] prion display several phenotypes. Aggregation of Swi1 in prion cells can be visualized with a fluorescently tagged copy of the protein or Swi1-PrD-containing fragment. A partial loss-of-function phenotype can be observed on non-glucose (e.g., raffinose, glycerol) media where [*SWI*$^+$] cells grow poorly. Abolishment of multicellular features in yeast from prionization of Swi1 leads to loss of flocculation (clumping together of yeast cells), adhesive growth (as visualized by crystal blue staining after washing), and pseudohyphae formation by diploids (cartoon depiction). The loss of these multicellular features also leads to [*SWI*$^+$]-containing cells having greater mobility in water flow scenarios as depicted in the final panel. Figure adapted from Goncharoff et al., 2018.

relation to other yeast prions. The disruption of Hsp104 through either GdnHCl treatment or expression of a dominant-negative mutant cures [*SWI*$^+$] (100). Additionally, [*SWI*$^+$] shows remarkable sensitivity to perturbations of the Hsp70 system (154). Disruption of Ssa1 through mutation and deletion of Sse1, a NEF for the Hsp70s both robustly cure the prion. However, overexpression of Sse1 can cure [*SWI*$^+$] in some yeast strains but not others – indicating different sensitivities may exist depending on genetic background (140, 154). Finally, the J-proteins of Sis1, Ydj1, and Apj1 were found to destabilize [*SWI*$^+$] upon overexpression and, interestingly, Sis1 repression can have the same curing effect.

Structurally, the Swi1 protein can be broken down into three domains (Figure 1.2) (155). The N domain ($Swi1_N$; residues 1-327) consists of the amino-terminal region enriched in asparagine residues and contains the Swi1 PrD. This region has the capability to form amyloid fibers *in vitro* (155). [*SWI*$^+$] can be introduced into naïve cells via transformation with these fibrils. The Q domain (residues 328-536) comprises the middle of the protein enriched in glutamine residues and acts in supporting the prionization of the N domain as well as the functioning of the Swi1 C domain. This C domain (residues 537-1314), at the carboxyl-terminal, was found to be indispensable for the normal functioning of Swi1. Expression of this C domain can rescue the poor growth on non-glucose carbon sources – though multicellularity cannot be rescued in this manner due to co-aggregation of other non-Swi1 proteins that help produce those phenotypes (134, 155).

Further study revealed that the Swi1 PrD, $Swi1_N$ can be further truncated and retain its aggregation capabilities and the ability to propagate the [*SWI*$^+$] prion fold (156). This smaller truncation, $Swi1_{1\text{-}38}$ consists of the first 38 residues and aggregates alongside full-length Swi1 ($Swi1_{FL}$). Furthermore, $Swi1_{1\text{-}38}$ can maintain this aggregation upon removal of $Swi1_{FL}$ and then transmit the prion conformation back to $Swi1_{FL}$ upon reintroduction. Additional experimentation

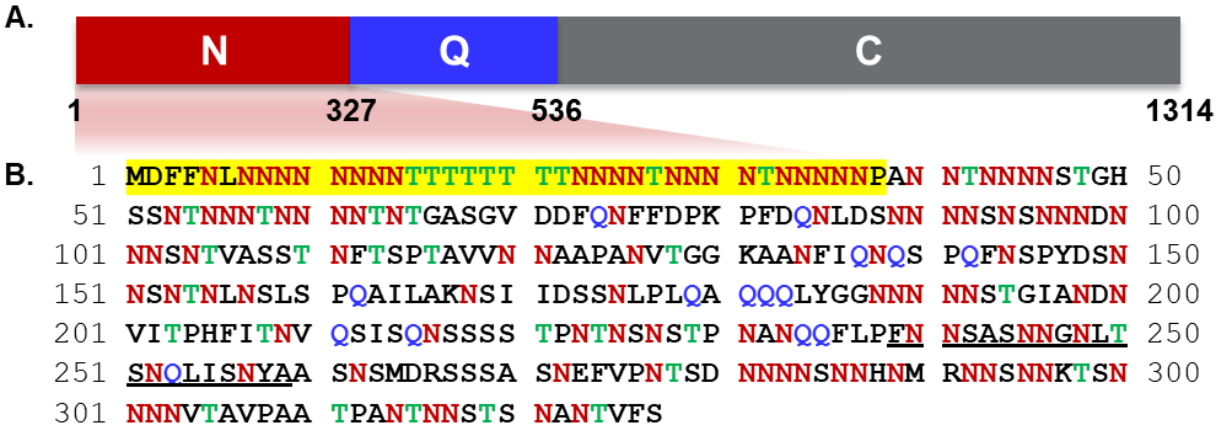


Figure 1.2. Schematic of the domains of Swi1.

A. The Swi1 protein consists of three domains. The N domain (residues 1-327) is highly enriched in asparagine (N) residues, acts as a prion domain (PrD), and can form fibrils to seed [*SWI*$^{+}$] formation. The middle Q domain (residues 328-536) is rich in glutamine (Q) residues and modulates the aggregation pattern of the N region and Swi1 functions. The carboxyl-terminal C domain acts as the functional domain of Swi1.

B. The sequence of the N domain is shown. The extreme N-terminal region, Swi1$_{1\text{-}38}$ – shown to be able to stably transmit the prion confirmation in the absence of full-length Swi1 – is highlighted. A reported amyloidogenic region, Swi1$_{239\text{-}259}$ whose role in prionogenesis has not been tested is also underlined. Asparagine, glutamine, and threonine residues are highlighted in red, blue, and green, respectively.

Figure adapted from Goncharoff et al., 2018.

revealed that $Swi1_{1-38}$ can *de novo* form a prion when fused with $Sup35_{MC}$ (157). Even smaller truncations such as $Swi1_{1-32}$ can also perform these tasks; however, stability was decidedly decreased. Taken together, these results demonstrate that this small, asparagine-rich region of Swi1 can act as a bona fide prion domain.

Prion Prediction

Due to the prevalence of sequencing technologies, the prediction of prionogenic proteins from amino-acid sequences remains a significant area of research. These efforts to predict protein properties from sequence data extends to related characteristics including disorderedness, amyloidogenicity, and general aggregation propensity. Existing protein-characteristic prediction algorithms include DISOPRED3, FoldIndex, PAPA (Prion Aggregation Prediction Algorithm), PLAAC (Prion-Like Amino Acid Composition), PrionW, TANGO, Waltz, ZipperDB, and Zyggregator (146, 158–165). Each of these tools utilizes different approaches and initial data sets – leading to varying biases and restrictions in prediction.

Table 1.2. A sampling of prediction algorithms.

Name	Property Predicted	Source
DISOPRED3	Disorderedness	(146)
FoldIndex	Disorderedness	(158)
PAPA	Prionogenicity	(159)
PASTA	Amyloidogenicity, disorderedness	(166)
PLAAC	Prionogenicity	(160)
PrionW	Prionogenicity	(161)
Waltz	Amyloidogenicity	(163)
ZipperDB	Amyloidogenicity	(164)
Zyggregator	Aggregation propensity	(165)

Yeast prion work has directly contributed to creation of some prediction algorithms. The precursor to PLAAC was developed to screen the *Saccharomyces cerevisiae* proteome for possible

novel PrDs that share amino-acid compositions similar to the already-identified PrDs (in Sup35, Ure2, and Rnq1) at the time (109). Indeed, this initial screening identified Mot3 as a yeast prion protein and its associated prion, [*MOT3*$^{+}$]. The algorithm uses a hidden Markov model – a statistical model that assumes an observable phenomenon depends on another variable which can exist in hidden, unobservable states. The further development of the algorithm resulted in PLAAC by adding the amino-acid frequencies of over 20 additional PrDs or PrLDs (160). PLAAC also expands upon the original algorithm by allowing for modification of the amino-acid background frequencies used as well as the minimum length of a predicted prion core. Not surprisingly, an important constraint of PLAAC remains the limited number of identified bona fide PrDs and PrLDs. This bias may lead to not identifying novel prion proteins lacking sequence similarities to the largely N/Q-rich proteins that form a significant portion of the training set for PLAAC (145). Nonetheless, the algorithm has displayed usefulness in examining disease-associated mutations in hnRNPA1 and hnRNPA2/B1 (30). An example output for one protein of interest, SS18 in Figure 1.3 displays the easy-to-interpret graphics produced by the web implementation of PLAAC (http://plaac.wi.mit.edu/) – as well as the inclusion of two other algorithms, FoldIndex and PAPA, to be discussed in the following paragraphs.

Another yeast-based algorithm, PAPA has also been developed for predicting prion propensity (159, 167). Random mutagenesis of a small critical region of the Sup35 PrD revealed which amino acids allowed for or disrupted prion formation by Sup35. This data combined with the hydrophobicity, α-helix propensity, and β-sheet propensity of each amino acid allowed for the calculation of an odds ratio of each amino acid to be included in a prion domain. The sum of these odds ratios over a window (i.e., a consecutive stretch of residues in a peptide sequence) provides the prion propensity of that protein region. Testing known prion domains revealed a bimodal

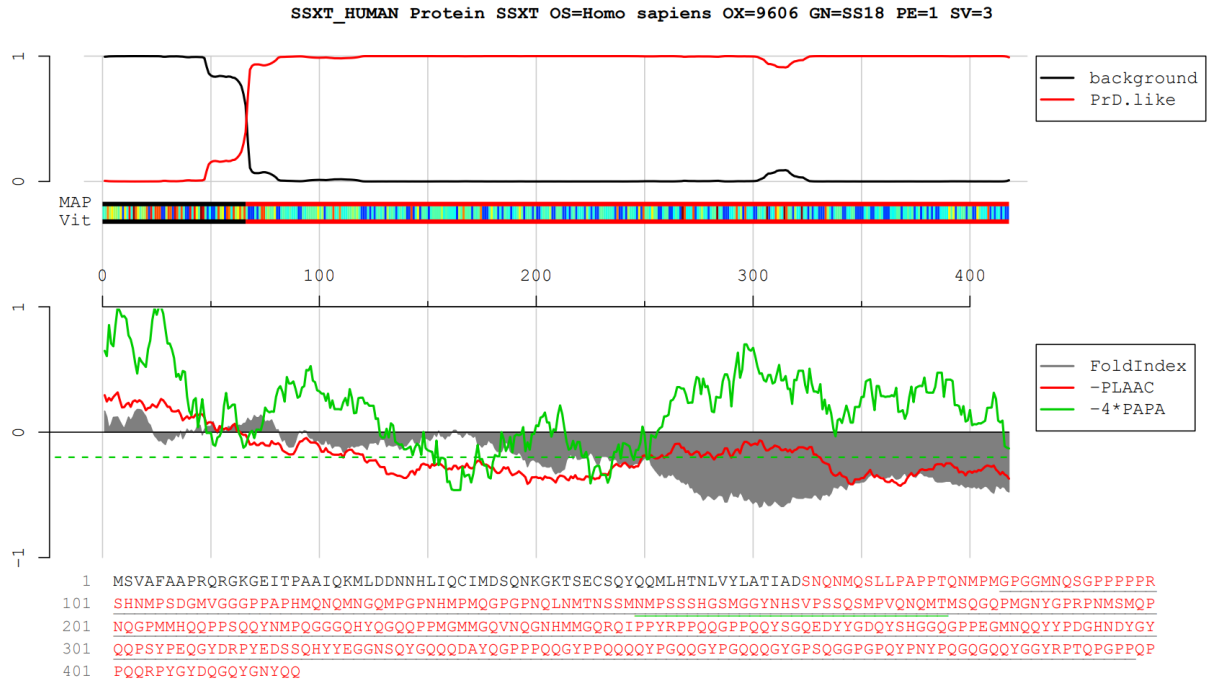


Figure 1.3. Visual output generated by evaluating the human SS18 protein with the PLAAC web server.

The top graph displays the likelihood calculated by the PLAAC algorithm that the corresponding amino acid position exists in a region predicted to be PrD-like (red line) or not (black line). The x-axis is the protein sequence of SS18 from N-terminus to C-terminus as diagramed with colors corresponding to each amino acid. The bottom graph displays the FoldIndex score (gray), negative PLAAC score (red line), and PAPA score scaled by a factor of -4 (green line). The entire sequence of SS18 written at the bottom with residues in black for predicted non-PrD-like regions or in red for predicted PrD-like regions. Default core size of 60 used. Background amino acid frequencies of *Homo sapiens* used.

distribution where a cutoff could be established that would allow for >90% accuracy in differentiating prion from non-prion (159). Using this cutoff, possible prionogenic regions can be predicted and therefore, possible prionogenic proteins. Due to the easily modifiable nature of the cutoff value used, the sensitivity of PAPA can be adjusted as necessary. Though, such changes may introduce reductions in accuracy if one assumes those proteins not meeting a more stringent cutoff value as being non-prion instead of merely not passing this particular predictive test. PAPA exists via a publicly available web server (https://combi.cs.colostate.edu/supplements/papa/) as well as in code form for local implementation.

The prion prediction algorithm of PrionW (http://bioinf.uab.cat/prionw/) exists an amalgamation of other methods – FoldIndex for disorderedness prediction, pWALTZ for amyloidogenicity, and a relatively simple calculation of the proportion of residues that are N or Q (161). By using FoldIndex, PrionW first determines whether a given protein sequence contains any disordered regions of at least 60 amino-acid residues in length. These regions are then evaluated for whether they are made up of at least 25% N or Q residues before being passed to pWALTZ for final analysis. As an iteration of the WALTZ algorithm, pWALTZ relies on the position-specific scoring matrix initially used to define the scoring conducted by WALTZ. Using a library of 116 amyloid-forming and 103 non-amyloid-forming hexapeptides, the probability of finding each amino-acid at a particular residue position was determined (163). Combining this position-specific scoring with a physiochemical scoring and structural scoring, the WALTZ algorithm provides an overall score for query sequences. Through usage of a lower scoring cutoff and requiring that a window of 21 consecutive residues scored above this cutoff value, pWALTZ was derived (168). Testing of pWALTZ on known yeast prions demonstrated a promising level of accuracy and comparable balancing of sensitivity and specificity to PAPA. If a particular protein

sequence examined by PrionW through this combinatorial method results in the identification of multiple possible PrDs, the algorithm returns the region with the highest scoring amyloid core (by pWALTZ) as the putative PrD for that protein (161).

As already mentioned, the FoldIndex algorithm functions to predict disorderedness of a given protein sequence (158). Once again, this method also uses a window (with a default of 51 residues) through which scores are calculated on a per residue basis. The calculation relies on combining previous work mathematically defining mean net charge and mean hydrophobicity of a peptide sequence (169, 170). In doing so, FoldIndex provides scoring that is either positive or negative – corresponding to a folded region or an unfolded region. The web implementation of the FoldIndex algorithm provides access to modify the size of the window used – leading to potentially varied results on a domain-scale, though calculations for entire proteins would remain the same. As can be gathered from the discussion of PLAAC and PrionW, FoldIndex has been used as a quick and efficient method for predicting disorderedness of proteins when attempting to predict potential PrDs. FoldIndex can be used either through the built-in implementation with the aforementioned algorithms or directly (https://fold.weizmann.ac.il/fldbin/findex).

An additional algorithm for prediction of protein disorderedness is DISOPRED. Now in its third iteration, this algorithm was initially developed through the usage of a neural network using the residues in the peptide sequence as the known states to predict either disordered or ordered on a per-residue basis (171). Training of this original version of DISOPRED utilized solved protein structures from the Protein Data Bank (PDB) by identifying residues that were not able to be solved via X-ray-based methods (i.e., residues that were not in a consistent position). The first revision of this algorithm, DISOPRED2 improved upon the initial approach by using a different form of learning algorithm known as a support vector machine, which provides a greater tolerance to

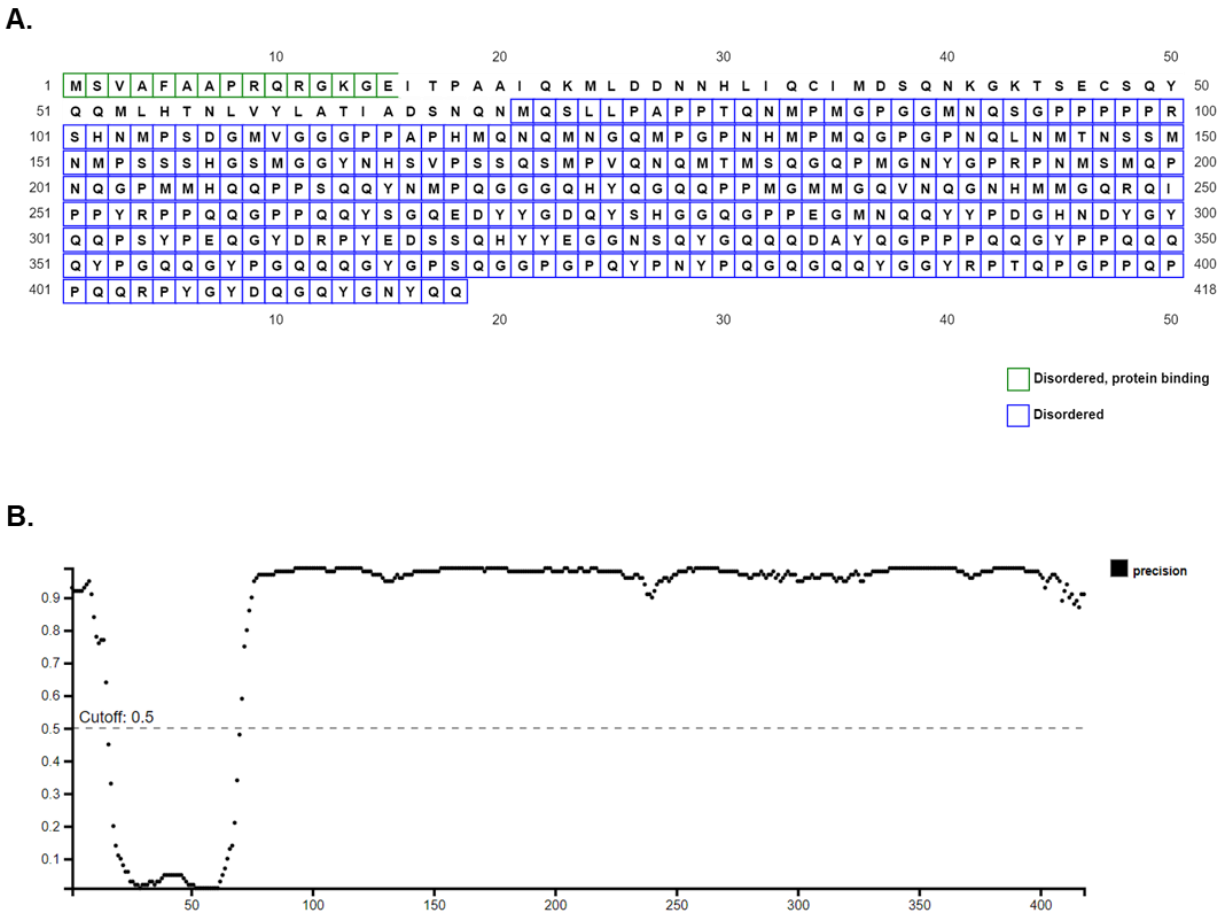


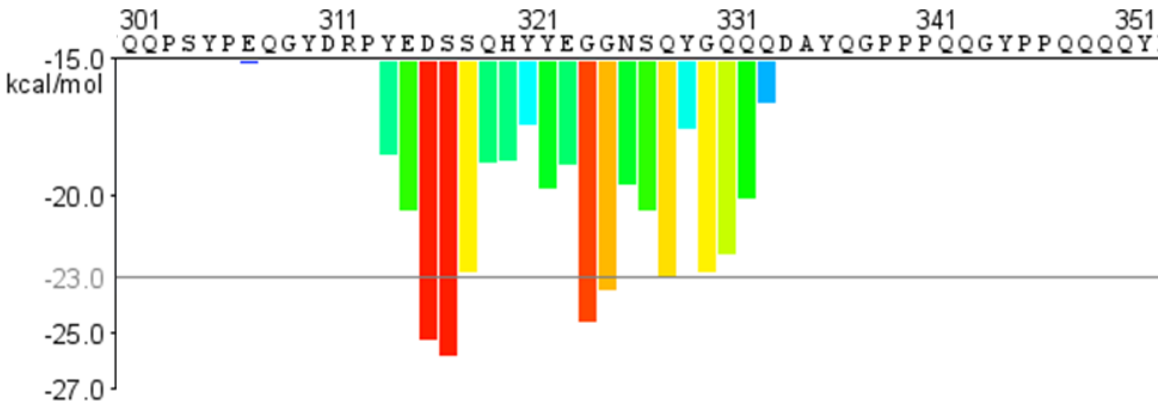
Figure 1.4. Visual output generated by evaluating the human SS18 protein with the DISOPRED3 web server.

A. Classification of each residue of SS18 into disordered (blue boxes), disordered with protein binding (green boxes), or neither (no box).
B. Graph of disordered (1.0) versus ordered (0.0) prediction and the corresponding precision for the prediction per residue. The x-axis is the residue number from N-terminus to C-terminus.

overfitting the data (172, 173). However, the most recent revision, DISOPRED3 combined the approaches of the first two versions by having both a support vector machine-based algorithm and a neural network-based algorithm (146). This recreated neural network trained on a new set of proteins that have known long stretches of disordered residues (to combat poor handling of such areas by DISOPRED2). These analyses were then combined with a few other predictors to form the final algorithmic approach. A web server (http://bioinf.cs.ucl.ac.uk/psipred/) allows for general availability of DISOPRED3 and associated algorithms and generates a plot of the predictions as well as an easily readable, per-residue prediction (Figure 1.4). DISOPRED3 can predict intrinsically disordered regions (IDRs) that have been indicated as having important roles in allowing for the prion or prion-like behavior of some proteins.

As opposed to the prediction of disorderedness by FoldIndex and DISOPRED3, a range of algorithms attempt to predict the ability of proteins to form the highly ordered structures of amyloid fibrils. One such algorithm is that of ZipperDB (https://services.mbi.ucla.edu/zipperdb/). This database uses a three-dimensional fitting of sequences of six residues in length from a given input protein sequence to formulate its prediction of amyloid fibril formation (164). The yeast prion protein, Sup35 contains the sequence NNQQNY which has been crystallized. This NNQQNY peptide also can form the cross-beta structure of an amyloid fibril (174). To produce its predictions, this algorithm samples every peptide of six residues that can be produced from the input sequence. The one restriction being that the six residues cannot contain a proline – such peptides are ignored and not scored. The predicted energy of the peptide being fit into the three-dimensional structure of NNQQNY is calculated by another program, Rosetta Design. If this energetic model is past the threshold of -23 kJ/mol (as determined from experimentally tested amyloid-forming peptides), then the given window of six residues is predicted to be amyloid fibril-

A.



B.

Position	Sequence	Rosetta Energy (kcal/mol)
307	EQGYDR	-12.3
314	YEDSSQ	-18.5
315	EDSSQH	-20.5
316	DSSQHY	-25.2
317	SSQHYY	-25.8
318	SQHYYE	-22.8
319	QHYYEG	-18.8
320	HYYEGG	-18.7
321	YYEGGN	-17.4
322	YEGGNS	-19.7
323	EGGNSQ	-18.9

Position	Sequence	Rosetta Energy (kcal/mol)
324	GGNSQY	-24.6
325	GNSQYG	-23.4
326	NSQYGQ	-19.6
327	SQYGQQ	-20.5
328	QYGQQQ	-23
329	YGQQQD	-17.6
330	GQQQDA	-22.8
331	QQQDAY	-22.1
332	QQDAYQ	-20.1
333	QDAYQG	-16.6

Figure 1.5. Visual output generated by evaluating the human SS18 protein with the 3D profile method of ZipperDB.

A. Graphical representation of the energies calculated by the 3D profile method using Rosetta Energy. The six residue-long peptide starting at each position is fit into the NNQQNY fibril crystal structure and the corresponding energetic fit is calculated using Rosetta Energy. Shown is the region of 301-352 of SS18. The default threshold of -23.0 kcal/mol (gray line) is indicated. The bars of the energies are displayed in colors with colors towards blue being higher energies (and thus less favorable structural fits) and colors towards being lower energies (and thus more favorable structural fits). Peptides that would contain a proline (P) are not evaluated and calculating Rosetta Energy at corresponding positions is not conducted.

B. Table of evaluated peptide sequences and Rosetta Energy scores for the region of SS18 represented in the graph in A.

forming (Figure 1.5). Given that many known human prion-like proteins and yeast prions can form amyloids, the presence of predicted fibril-forming regions may hold indications of the ability of a larger protein being able to do so under the correct conditions.

Another prediction algorithm that focuses on amyloid is PASTA, prediction of amyloid structural aggregation. This algorithm relies on multiple functions that attempt to mathematically describe the energies of each amino acid in a β-strand based on the hydrogen bonding of β-strands in known globular protein structures (175). Possible energies of both parallel and antiparallel arrangements are calculated in this manner. Through examination of segments of different lengths across a given sequence, the best pairing segments can be determined on the most favorable (i.e., lowest) energy. These calculated energies are translated on residue position basis to amyloid aggregation propensities. An iteration of PASTA, so-called PASTA 2.0 improved upon the initial algorithm by using a new, larger dataset of protein structures for the energy function as well as by incorporating predictions of secondary structure and disorderedness (166). More specifically, disorderedness prediction was added to PASTA through incorporation of another algorithm, ESpritz (176). A web server (http://protein.bio.unipd.it/pasta2/) provides bulk processing and visual output of results from PASTA (Figure 1.6). Benchmarking of PASTA indicates its usefulness in identifying implicated amyloid-forming regions and thus may allow for identification of core regions of some prion or prion-like proteins.

While prediction algorithms such as ZipperDB and PASTA focus specifically on the formation of amyloid aggregates, other approaches attempt to address general aggregation propensity. One such prediction system is that of Zyggregator (165). This prediction algorithm attempts to predict aggregation propensity from the physical and chemical properties – charge, hydrophobicity, and secondary structure – of the amino acids present in the protein being

A.

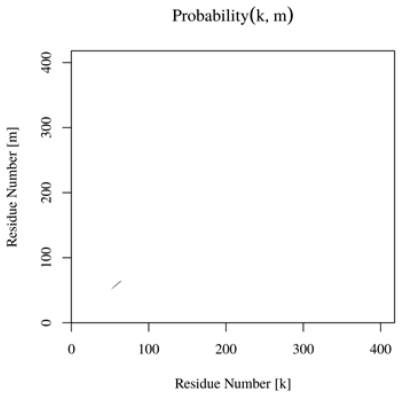
```
1        10        20        30        40
MSVAFAAPRQRGKGEITPAAIQKMLDDNNHLIQCIMDSQNKGKTSECSQY
DDDDDDDDDDDDOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOO
CCBBBBCCCCCCCCCCCHHHHHHHHHCCCCBBBBBBCCCCCCCCCHHHHH
----------------------------------------------------

51       60        70        80        90
QQMLHTNLVYLATIADSNQNMQSLLPAPPTQNMPMGPGGMNQSGPPPPPR
OOOOOOOOOOOOOOOOOOOOOOOOOOOOODDDDDDDDDDDDDDDDDDDDD
HHHHHHCHHHHHHHHCCCCHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCC
--PPPPPPPPPPPPP-----------------------------------

101      110       120       130       140
SHNMPSDGMVGGGPPAPHMQNQMNGQMPGPNHMPMQGPGPNQLNMTNSSM
DDDDDDDDDDDDDDDDDDOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOO
CCCCCCCCCCCCCCCCCCHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
----------------------------------------------------
```

D: Disordered residue
O: Ordered residue

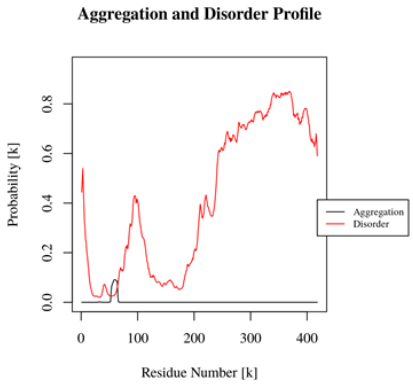
H: α-Helix residue
B: β-strand residue
C: Coiled residue

P: Parallel aggregation
A: Anti-parallel aggregation
-: Non aggregating residue

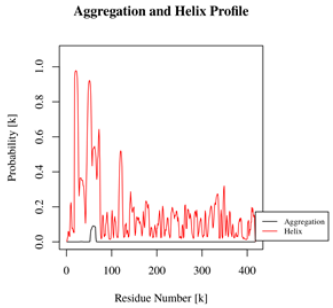
B.

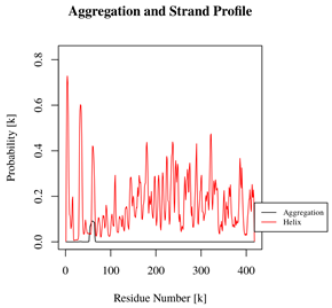


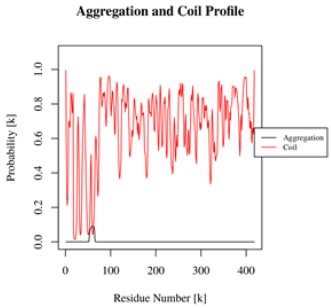
C.



D.







E.

Length (aa)	# Amyloids	Best Energy	% Disorder	% α-Helix	% β-Strand	% Coil
418	10	-6.41565	68.18	8.85	3.83	87.32

Figure 1.6. Visual output generated by evaluating the human SS18 protein with the PASTA web server.

A. Residue assignment of the SS18 protein. The first line notes the sequence position. The second line notes the amino acid present. The third line notes the prediction of the residue being ordered (O) or disordered (D). The third line notes the prediction of the secondary structure present – α-helix (H), β-strand (B), or coil (C). The fourth line notes the prediction of any amyloid aggregation – parallel (P), anti-parallel (A), or non-aggregating (-). Shown are residues 1-150 of the SS18 protein.
B. Probability plot of amyloid aggregation for the SS18 protein. Pairings of individual amino-acid residues (k, m) are plotted if predicted to form amyloid by the energy function of PASTA.
C. Disorder probability plot for the SS18 protein. The predicted probability of disorderedness on a per residue basis is plotted in red. The predicted probability of amyloid aggregation on a per residue basis is plotted in black.
D. Secondary structure probability plots for the SS18 protein. The predicted probability of α-helix, β-strand, or coil on a per residue basis is plotted in red. The predicted probability of amyloid aggregation on a per residue basis is plotted in black.
E. Table of summary data for the SS18 protein. Number of amyloids refers to the number of unique peptide sequences of 10 residues in length predicted to form amyloid. Best energy refers to the most favorable (i.e., lowest) energy scoring of any predicted amyloid aggregate-forming region. Units for best energy are PASTA Energy Units (PEU) and 1 PEU = 1.192 Kcal/mol. The SS18 protein was evaluated using an energy threshold of -5 PEU.

evaluated. Initial development of Zyggregator depended on experiments with point mutations in various disease-related proteins including Aβ42, α-syn, and tau and the associated calculations surrounding their physiochemical properties (177). For each residue position, Zyggregator calculates the intrinsic aggregation propensity using each amino-acids previously determined charge, hydrophobicity, α-helix formation ability, and β-sheet formation ability (165). Furthermore, an overall aggregation propensity for the entire sequence being examined is calculated by comparison to scoring for randomly generated sequences. The Zyggregator method has been validated through high correlation between predicted aggregation propensities and experimentally observed aggregation (178–180). Importantly for studies focused on the involvement of mutated proteins in human disease, Zyggregator can distinguish changes in aggregation propensity of even single point mutations (179).

The prediction algorithms covered here – DISOPRED3, FoldIndex, PLAAC, PAPA, PASTA, PrionW, ZipperDB, and Zyggregator – are a limited selection of myriad approaches. Whether predicting prionogenicity, disorderedness, amyloidogenicity, or general aggregation propensity, these algorithms can and do inform research trajectories. However, the biases and limitations – including specificity, sensitivity, training sets, and more – of the methods used in each algorithm must be considered. Additionally, algorithms exist to predict other aspects of protein structure and behavior such as phase separation, a highly related topic to those discussed here. Given the significance of protein aggregation and misfolding to neurodegeneration as well as other disease, the prediction of these various protein attributes plays an important role in understanding disease and development of potential therapeutics.

CHAPTER TWO

Defining key residues of the Swi1 prion domain in prionogenesis and propagation

Introduction

Prions were initially identified as infectious abnormal protein conformations that underpin incurable neurological diseases (181). While the prion concept originally applied to the namesake protein, the idea has grown to encompass additional proteins in a multitude of organisms (17, 81, 96, 182, 183). The budding yeast, *Saccharomyces cerevisiae* harbors a number of endogenous proteins that can adopt alternative, heritable protein conformations (100–106, 109, 110). These proteins – termed yeast prions – have greatly contributed to our understanding of prions.

One such yeast prion, [*SWI*$^+$] was identified by our laboratory (100). The protein determinant of [*SWI*$^+$], Swi1 normally functions as part of the SWI/SNF chromatin-remodeling complex, which modulates the expression of more than 15% of yeast genes (150, 151). Due in part to this role, the prionization of Swi1 leads to multiple phenotypes in yeast including poor growth on non-glucose carbon sources (e.g. raffinose, glycerol), aggregation of the Swi1 protein, and loss of multicellular features (e.g. flocculation, invasive growth) (100, 134). Swi1 can be divided into three domains (155). The N-terminal, asparagine-rich domain contains the Swi1 prion domain (PrD) – the region necessary and sufficient for prionization. The N region has previously been shown to capably form amyloid fibrils *in vitro* (155). A middle glutamine-rich domain follows and a C-terminal, functional domain completes the protein. Expression of this functional domain rescues the poor growth on raffinose phenotype and restores multicellular features (155).

Further research into the Swi1 PrD revealed that the protein's first 38 amino acids could act to maintain and propagate the [*SWI*$^+$] prion fold (156). Moreover, this region, Swi1$_{1\text{-}38}$ could

be further truncated – down to $Swi1_{1-32}$ – and retain the ability to aggregate and propagate (157). $Swi1_{1-38}$ was also shown to act as a transferable PrD. When fused with $Sup35_{MC}$ – the Sup35 protein without its N-terminal prion domain – for assaying purposes, $Swi1_{1-38}$ can *de novo* form a prion that has been termed [SPS^+] (Swi1 conferred [PSI^+]). This prion, formed by $Swi1_{1-38}$-MC, exhibits aggregation when visualized by $Swi1_{1-38}$-YFP, displays impaired translation termination due to the primary function of $Sup35_{MC}$, and is curable by treatment with guanidine hydrochloride. Once again, a shorter truncation, $Swi1_{1-31}$ was found to also be capable of prionization. In all, this small N-terminal region of Swi1 stands as the smallest currently identified PrD.

$Swi1_{1-38}$ is highly enriched in polar residues, particularly asparagine. Of the 38 residues, 22 are asparagine and 10 are threonine. Inclusion of asparagine and/or glutamine residues is common among currently characterized yeast prions (109, 167). Meanwhile, the six remaining residues comprise a methionine necessary as a start codon, an adjacent aspartate likely playing a role in the protein half-life, an ending proline with probable unimportance for prion capabilities, and three hydrophobic residues – a leucine and two phenylalanines. This largely uncomplicated primary sequence of $Swi1_{1-38}$ gives rise to a protein domain capable of aggregating, maintaining and propagating an alternative fold, and initializing a prion (156, 157). To better understand the prionogenicity of $Swi1_{1-38}$, we performed a series of mutagenesis experiments to characterize the contributions of residues to the prionogenic characteristics of this small PrD.

Materials and Methods

Yeast strains and media

Yeast strains used in this study are listed in Table 2.1. The W303 *sup35Δ / SUP35::TRP1 / p316SUP35$_{FL}$* strain was provided by the Weissman Lab (UCSF, San Francisco, CA, USA).

Yeast were grown according to established protocols at 30° C in either yeast extract-peptone-dextrose (YPD) or synthetic complete (SC) media minus appropriate amino acids (e.g. leucine) (184). When indicated, media was supplemented with 1 g/l 5-FOA for counter-selection against a *URA3*-carrying plasmid or with 5 mM GdnHCl for inactivation of Hsp104 to disrupt prion propagation.

Table 2.1. Yeast strains used in chapter two.

Identifier	Background	Relevant Genotype	Prion State	Source
742	BY4741		[*SWI*$^+$]	(100)
756	BY4741	*swi1Δ / p416TEF-SWI1*	[*SWI*$^+$]	(156)
YJW561	W303	*sup35Δ / SUP35::TRP1 / p316SUP35*$_{FL}$	[*PSI*$^+$][*PIN*$^+$]	Weissman Lab

Plasmid construction

Plasmids used in this study are listed in Table 2.2. Briefly, the *p415TEF-SWI1*$_{1\text{-}38}$*-YFP* plasmid (156) was used as template to produce the various mutant *SWI1*$_{1\text{-}38}$ plasmids via PCR. See Table 2.3 for primer information. For mutations in the first portion of *SWI1*$_{1\text{-}38}$, the mutant PCR product was digested with *Spe*I/*Xho*I for cloning into similarly digested *p415TEF-SWI1*$_{1\text{-}38}$*-YFP* to produce *p415TEF-SWI1*$_{1\text{-}38}$*F3N-YFP*, *p415TEF-SWI1*$_{1\text{-}38}$*F3A-YFP*, *p415TEF-SWI1*$_{1\text{-}38}$*F4N-YFP*, *p415TEF-SWI1*$_{1\text{-}38}$*FFNN-YFP*, *p415TEF-SWI1*$_{1\text{-}38}$*FFAA-YFP*, *p415TEF-SWI1*$_{1\text{-}38}$*L6N-YFP*, and *p415TEF-SWI1*$_{1\text{-}38}$*L6A-YFP*. This approach was also used with *p415TEF-SWI1*$_N$*-YFP* and *p415TEF-SWI1*$_{NQ}$*-YFP* to make *p415TEF-SWI1*$_N$*F3N-YFP*, *p415TEF-SWI1*$_N$*F3A-YFP*, *p415TEF-SWI1*$_N$*F4N-YFP*, *p415TEF-SWI1*$_N$*F4A-YFP*, *p415TEF-SWI1*$_N$*FFNN-YFP*, *p415TEF-SWI1*$_N$*FFAA-YFP*, *p415TEF-SWI1*$_{NQ}$*F3N-YFP*, *p415TEF-SWI1*$_{NQ}$*F3A-YFP*, *p415TEF-SWI1*$_{NQ}$*F4N-YFP*, *p415TEF-SWI1*$_{NQ}$*F4A-YFP*, *p415TEF-SWI1*$_{NQ}$*FFNN-YFP*, and *p415TEF-SWI1*$_{NQ}$*FFAA-YFP*. For mutations in the middle of *SWI1*$_{1\text{-}38}$, the mutant PCR product was cloned

into *p415TEF-SWI1$_{1-38}$-YFP* via *Sac*I/*Xho*I sites to make *p415TEF-SWI1$_{1-38}$4xTN-YFP*, *p415TEF-SWI1$_{1-38}$4xTA-YFP*, *p415TEF-SWI1$_{1-38}$8xTN-YFP*, and *p415TEF-SWI1$_{1-38}$8xTA-YFP*. For mutations in the back portion of *SWI1$_{1-38}$*, the mutant PCR product was cloned into *p415TEF-SWI1$_{1-38}$-YFP* via *Sac*I/*Bam*HI sites to produce *p415TEF-SWI1$_{1-38}$T27N-YFP*, *p415TEF-SWI1$_{1-38}$T27A-YFP*, *p415TEF-SWI1$_{1-38}$T32N-YFP*, and *p415TEF-SWI1$_{1-38}$T32A-YFP*. To produce the collection of mutant *p416TEF-SWI1$_{1-38}$-YFP* plasmids, the mutant SWI1$_{1-38}$-YFP was cloned from the respective *p415TEF-SWI1$_{1-38}$-YFP* plasmid and into *p416TEF* via *Spe*I/*Xho*I sites.

The *p415TEF-SWI1$_{NQ\Delta239-259}$-YFP* plasmid was produced by PCR-amplifying *SWI1$_{1-238}$* with the SWI1$_{\Delta239-259}$ Front For and SWI1$_{\Delta239-259}$ Front Rev primers and PCR-amplifying *SWI1$_{260-517}$-YFP* with SWI1$_{\Delta239-259}$ Back For and SWI1$_{\Delta239-259}$ Back Rev primers. These two PCR products were then linked by using a mixture of both as template and the SWI1$_{\Delta239-259}$ Front For and SWI1$_{\Delta239-259}$ Back Rev primers – producing the full-length *SWI1$_{NQ\Delta239-259}$-YFP* product. Was subsequently cloned into *p415TEF* via *SpeI*/*XhoI* sites. The *p415TEF-SWI1$_{NQ\Delta239-259}$ FFNN-YFP* and *p415TEF-SWI1$_{NQ\Delta239-259}$FFAA-YFP* constructs were produced in the same manner as the other FF-NN and FF-AA mutants above.

The *p415TEF-SWI1$_{1-38}$-MC* plasmid was produced by PCR-amplifying *SWI1$_{1-38}$* from *p415TEF-SWI1$_{1-38}$-YFP* with the SpeI-SWI1$_{1-38}$ For and SWI1$_{1-38}$-BamHI-Linker Rev primers and PCR-amplifying MC from *p316SUP35$_{FL}$* with the Linker-SUP35$_{MC}$ For and SUP35$_{MC}$-XhoI Rev primers. These two PCR products were then linked by using a mixture of both as template and the SpeI-SWI1$_{1-38}$ For Short and SUP35$_{MC}$-XhoI Rev Short primers – producing the full-length *SWI1$_{1-38}$-Linker-MC* product where the DPGGPGGG linker contains a *Bam*HI site. *SWI1$_{1-38}$-Linker-MC* was subsequently cloned into *p415TEF* via *Spe*I/*Xho*I sites. The collection of mutant *p415TEF-*

SWI1$_{1-38}$-MC plasmids was generated by cloning the mutant *SWI1$_{1-38}$* from the respective *p415TEF-SWI1$_{1-38}$-YFP* plasmids and into *p415TEF-SWI1$_{1-38}$-MC* via *Sac*I/*Bam*HI sites.

Table 2.2. Plasmids used in chapter two.

Name	Marker	Replicon	Promoter	Use	Source
p415TEF-SWI1$_{1-38}$-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$-YFP	(156)
p415TEF-SWI1$_{1-38}$ F3N-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F3N-YFP	This study
p415TEF-SWI1$_{1-38}$ F3A-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F3A-YFP	This study
p415TEF-SWI1$_{1-38}$ F4N-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F4N-YFP	This study
p415TEF-SWI1$_{1-38}$ F4A-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F4A-YFP	This study
p415TEF-SWI1$_{1-38}$ FFNN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ FFNN-YFP	This study
p415TEF-SWI1$_{1-38}$ FFAA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ FFAA-YFP	This study
p415TEF-SWI1$_{1-38}$ L6N-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ L6N-YFP	This study
p415TEF-SWI1$_{1-38}$ L6A-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ L6A-YFP	This study
p415TEF-SWI1$_{1-38}$ 4xTN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 4xTN-YFP	This study
p415TEF-SWI1$_{1-38}$ 4xTA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 4xTA-YFP	This study
p415TEF-SWI1$_{1-38}$ 8xTN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 8xTN-YFP	This study
p415TEF-SWI1$_{1-38}$ 8xTA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 8xTA-YFP	This study
p415TEF-SWI1$_{1-38}$ T27N-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T27N-YFP	This study
p415TEF-SWI1$_{1-38}$ T27A-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T27A-YFP	This study
p415TEF-SWI1$_{1-38}$ T32N-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T32N-YFP	This study
p415TEF-SWI1$_{1-38}$ T32A-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T32A-YFP	This study
p415TEF-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of YFP	(134)
p416TEF-SWI1$_{FL}$	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{FL}$	(100)
p416TEF-SWI1$_{FL}$-mCherry	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{FL}$-mCherry	(156)
p416TEF-SWI1$_{1-38}$-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$-YFP	This study
p416TEF-SWI1$_{1-38}$ F3N-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F3N-YFP	This study
p416TEF-SWI1$_{1-38}$ F3A-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F3A-YFP	This study
p416TEF-SWI1$_{1-38}$ F4N-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F4N-YFP	This study
p416TEF-SWI1$_{1-38}$ F4A-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F4A-YFP	This study
p416TEF-SWI1$_{1-38}$ FFNN-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ FFNN-YFP	This study
p416TEF-SWI1$_{1-38}$ FFAA-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ FFAA-YFP	This study
p416TEF-SWI1$_{1-38}$ L6N-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ L6N-YFP	This study
p416TEF-SWI1$_{1-38}$ L6A-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ L6A-YFP	This study
p416TEF-SWI1$_{1-38}$ 4xTN-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 4xTN-YFP	This study
p416TEF-SWI1$_{1-38}$ 4xTA-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 4xTA-YFP	This study
p416TEF-SWI1$_{1-38}$ 8xTN-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 8xTN-YFP	This study
p416TEF-SWI1$_{1-38}$ 8xTA-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 8xTA-YFP	This study
p416TEF-SWI1$_{1-38}$ T27N-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T27N-YFP	This study
p416TEF-SWI1$_{1-38}$ T27A-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T27A-YFP	This study
p416TEF-SWI1$_{1-38}$ T32N-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T32N-YFP	This study
p416TEF-SWI1$_{1-38}$ T32A-YFP	URA3	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T32A-YFP	This study
p415TEF-SWI1$_{1-38}$-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$-MC	This study
p415TEF-SWI1$_{1-38}$ F3N-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F3N-MC	This study
p415TEF-SWI1$_{1-38}$ F3A-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F3A-MC	This study
p415TEF-SWI1$_{1-38}$ F4N-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F4N-MC	This study
p415TEF-SWI1$_{1-38}$ F4A-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ F4A-MC	This study
p415TEF-SWI1$_{1-38}$ FFNN-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ FFNN-MC	This study
p415TEF-SWI1$_{1-38}$ FFAA-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ FFAA-MC	This study
p415TEF-SWI1$_{1-38}$ L6N-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ L6N-MC	This study
p415TEF-SWI1$_{1-38}$ L6A-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ L6A-MC	This study
p415TEF-SWI1$_{1-38}$ 4xTN-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 4xTN-MC	This study
p415TEF-SWI1$_{1-38}$ 4xTA-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 4xTA-MC	This study
p415TEF-SWI1$_{1-38}$ 8xTN-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 8xTN-MC	This study
p415TEF-SWI1$_{1-38}$ 8xTA-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ 8xTA-MC	This study
p415TEF-SWI1$_{1-38}$ T27N-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T27N-MC	This study

p415TEF-SWI1$_{1-38}$ T27A-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T27A-MC	This study
p415TEF-SWI1$_{1-38}$ T32N-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T32N-MC	This study
p415TEF-SWI1$_{1-38}$ T32A-MC	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-38}$ T32A-MC	This study
p415TEF-SWI1$_{N}$-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{N}$-YFP	(155)
p415TEF-SWI1$_{NQ}$-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{NQ}$-YFP	(155)
p415TEF-SWI1$_{NQ}$ F3N-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{NQ}$ F3N-YFP	This study
p415TEF-SWI1$_{NQ}$ F3A-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{NQ}$ F3A-YFP	This study
p415TEF-SWI1$_{NQ}$ F4N-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{NQ}$ F4N-YFP	This study
p415TEF-SWI1$_{NQ}$ F4A-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{NQ}$ F4A-YFP	This study
p415TEF-SWI1$_{NQ}$ FFNN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{NQ}$ FFNN-YFP	This study
p415TEF-SWI1$_{NQ}$ FFAA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{NQ}$ FFAA-YFP	This study
p415TEF-SWI1$_{N}$ F3N-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{N}$ F3N-YFP	This study
p415TEF-SWI1$_{N}$ F3A-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{N}$ F3A-YFP	This study
p415TEF-SWI1$_{N}$ F4N-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{N}$ F4N-YFP	This study
p415TEF-SWI1$_{N}$ F4A-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{N}$ F4A-YFP	This study
p415TEF-SWI1$_{N}$ FFNN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{N}$ FFNN-YFP	This study
p415TEF-SWI1$_{N}$ FFAA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{N}$ FFAA-YFP	This study
p415TEF-SWI1$_{NQ\Delta239-259}$-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{NQ\Delta239-259}$-YFP	This study
p415TEF-SWI1$_{NQ\Delta239-259}$ FFNN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{NQ\Delta239-259}$ FFNN-YFP	This study
p415TEF-SWI1$_{NQ\Delta239-259}$ FFAA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{NQ\Delta239-259}$ FFAA-YFP	This study
p415TEF-SWI1$_{1-224}$-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-224}$-YFP	This study
p415TEF-SWI1$_{1-224}$ FFNN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-224}$ FFNN-YFP	This study
p415TEF-SWI1$_{1-224}$ FFAA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-224}$ FFAA-YFP	This study
p415TEF-SWI1$_{1-176}$-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-176}$-YFP	This study
p415TEF-SWI1$_{1-176}$ FFNN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-176}$ FFNN-YFP	This study
p415TEF-SWI1$_{1-176}$ FFAA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-176}$ FFAA-YFP	This study
p415TEF-SWI1$_{1-133}$-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-133}$-YFP	This study
p415TEF-SWI1$_{1-133}$ FFNN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-133}$ FFNN-YFP	This study
p415TEF-SWI1$_{1-133}$ FFAA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-133}$ FFAA-YFP	This study
p415TEF-SWI1$_{1-74}$-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-74}$-YFP	This study
p415TEF-SWI1$_{1-74}$ FFNN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-74}$ FFNN-YFP	This study
p415TEF-SWI1$_{1-74}$ FFAA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-74}$ FFAA-YFP	This study
p415TEF-SWI1$_{1-65}$-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-65}$-YFP	This study
p415TEF-SWI1$_{1-65}$ FFNN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-65}$ FFNN-YFP	This study
p415TEF-SWI1$_{1-65}$ FFAA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-65}$ FFAA-YFP	This study
p415TEF-SWI1$_{1-46}$-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-46}$-YFP	This study
p415TEF-SWI1$_{1-46}$ FFNN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-46}$ FFNN-YFP	This study
p415TEF-SWI1$_{1-46}$ FFAA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-46}$ FFAA-YFP	This study
p415TEF-SWI1$_{1-42}$-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-42}$-YFP	This study
p415TEF-SWI1$_{1-42}$ FFNN-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-42}$ FFNN-YFP	This study
p415TEF-SWI1$_{1-42}$ FFAA-YFP	LEU2	CEN6/ARSH4	TEF1	Expression of Swi1$_{1-42}$ FFAA-YFP	This study
p316Sup35$_{FL}$	URA3	CEN6/ARSH4	SUP35	Expression of Sup35	Weissman Lab

Site-directed mutagenesis

The suite of *SWI1$_{1-38}$* mutants was produced via incorporation of base substitutions in PCR primers (Table 4) and usage of *p415TEF-SWI1$_{1-38}$-YFP* as the template. All PCRs were conducted using PrimeSTAR HS DNA polymerase (Takara Bio, Mountain View, CA, USA) and the manufacturer's recommended protocols. Custom primers were ordered from Integrated DNA

Technologies (Coralville, IA, USA) and annealing temperatures were estimated via the Integrated DNA Technologies' OligoAnalyzer Tool.

Table 2.3. Primers used in chapter two.

Name	Sequence (5' - 3')	Resulting Plasmid
Swi1 F3N For	AGAACTAGTATGGATAACTTTAATTTGAAT	*p415TEF-SWI1$_{1-38}$ F3N-YFP*
Swi1 F3A For	AGAACTAGTATGGATGCCTTTAATTTGAAT	*p415TEF-SWI1$_{1-38}$ F3A-YFP*
Swi1 F4N For	AGAACTAGTATGGATTTCAACAATTTGAAT	*p415TEF-SWI1$_{1-38}$ F4N-YFP*
Swi1 F4A For	AGAACTAGTATGGATTTCGCCAATTTGAAT	*p415TEF-SWI1$_{1-38}$ F4A-YFP*
Swi1 FF-NN For	AGAACTAGTATGGATAACAACAATTTGAAT	*p415TEF-SWI1$_{1-38}$ FF-NN-YFP*
Swi1 FF-AA For	AGAACTAGTATGGATGCCGCCAATTTGAAT	*p415TEF-SWI1$_{1-38}$ FF-AA-YFP*
Swi1 L6N For	ACTAGTATGGATTTCTTTAATAACAATAATAATAATAATAATAATAATACTACTACT	*p415TEF-SWI1$_{1-38}$ L6N-YFP*
Swi1 L6A For	ACTAGTATGGATTTCTTTAATGCGAATAATAATAATAATAATAATAATACTACTACT	*p415TEF-SWI1$_{1-38}$ L6A-YFP*
Swi1 4xTN For	ACTACTAACAACAACAACAATAACAATAATACTAATAATAATAATACT	*p415TEF-SWI1$_{1-38}$ 4xTN-YFP*
Swi1 4xTN Rev	GTTATTGTTGTTGTTGTTAGTAGTAGTAGTATTATTATTATTATTATTATTCAA	*p415TEF-SWI1$_{1-38}$ 4xTN-YFP*
Swi1 4xTA For	ACTACTGCAGCAGCAGCAAATAACAATAATACTAATAATAATAATACT	*p415TEF-SWI1$_{1-38}$ 4xTA-YFP*
Swi1 4xTA Rev	GTTATTTGCTGCTGCTGCAGTAGTAGTAGTATTATTATTATTATTATTATTCAA	*p415TEF-SWI1$_{1-38}$ 4xTA-YFP*
Swi1 8xTN For	AATAATAACAACAACAACAACAACAACAACAATAACAATAATACT	*p415TEF-SWI1$_{1-38}$ 8xTN-YFP*
Swi1 8xTN Rev	GTTGTTGTTGTTGTTGTTATTATTATTATTATTATTATTATTCAA	*p415TEF-SWI1$_{1-38}$ 8xTN-YFP*
Swi1 8xTA For	AATAATGCAGCAGCAGCAGCAGCAGCAAATAACAATAATACT	*p415TEF-SWI1$_{1-38}$ 8xTA-YFP*
Swi1 8xTA Rev	TGCTGCTGCTGCTGCTGCATTATTATTATTATTATTATTATTCAA	*p415TEF-SWI1$_{1-38}$ 8xTA-YFP*
Swi1 T27N Rev	GGTGGATCCGGATTATTATTATTATTAGTATTATTATTATTATTATTATTGTTATTGGT	*p415TEF-SWI1$_{1-38}$ T27N-YFP*
Swi1 T27A Rev	GGTGGATCCGGATTATTATTATTATTAGTATTATTATTATTAGCATTATTGTTATTGGT	*p415TEF-SWI1$_{1-38}$ T27A-YFP*
Swi1 T32N Rev	GGTGGATCCGGATTATTATTATTATTGTTATTATTATTATTAGT	*p415TEF-SWI1$_{1-38}$ T32N-YFP*
Swi1 T32A Rev	GGTGGATCCGGATTATTATTATTATTAGCATTATTATTATTAGT	*p415TEF-SWI1$_{1-38}$ T32A-YFP*
p415TEF-SWI1$_{1-38}$-YFP For	TTATCTACACGACGGGGAGTCA	*Multiple SWI1$_{1-38}$ mutants*

p415TEF-SWI1$_{1-38}$-YFP Rev	AATGTAAGCGTGACATAACTAATTACATGA	*Multiple SWI1$_{1-38}$ mutants*
p415TEF-SWI1$_{1-38}$-YFP For 4x	CAAGACGATAGTTACCGGATAAGG	*Multiple SWI1$_{1-38}$ mutants*
p415TEF-SWI1$_{1-38}$-YFP Rev 4x	TGGATTTTGATGTAATTGTTGGGATTC	*Multiple SWI1$_{1-38}$ mutants*
p415TEF-SWI1$_{1-38}$-YFP For 4x HT	AAGACGATAGTTACCGGATAAGGCGCA	*Multiple SWI1$_{1-38}$ mutants*
p415TEF-SWI1$_{1-38}$-YFP Rev 4x HT	AGAATAGACCGAGATAGGGTTGAGTGTTGT	*Multiple SWI1$_{1-38}$ mutants*
SpeI-SWI$_{1-38}$ For	GGTTCAAGCTATGCGTCAGACCCCGTAGAAAAGATCAAAGG	*p415TEF-SWI1$_{1-38}$-MC*
SWI$_{1-38}$-BamHI-Linker Rev	ACCACCACCAGGACCACCTGGATCCGGATTATTATTATTATTAGTATTATTATTATTAGT	*p415TEF-SWI1$_{1-38}$-MC*
Linker-SUP35$_{MC}$ For	GGTGGTCCTGGTGGTGGTATGTCTTTGAACGACTTTCAAAAGC	*p415TEF-SWI1$_{1-38}$-MC*
SUP35$_{MC}$-XhoI Rev	CTGCGAGCCCTCGAGTTACTCGGCAATTTTAACAATTTTACCAATTGCT	*p415TEF-SWI1$_{1-38}$-MC*
SpeI-SWI$_{1-38}$ For Short	TCAGACCCCGTAGAAAAGATCAAAGG	*p415TEF-SWI1$_{1-38}$-MC*
SUP35$_{MC}$-XhoI Rev Short	CTGCGAGCCCTCGAGTTACTC	*p415TEF-SWI1$_{1-38}$-MC*
SWI1 SRT For	TCTAACTCTACTCCGAATGCAAATC	N/A
SWI1 SRT Rev	ACGTTGATATTAATATTGCTATTCAAGCT	N/A
ACT1 RT For	TTGGTTATTGATAACGGTTCTGGTATG	N/A
ACT1 RT Rev	GGTGAACGATAGATGGACCACTT	N/A
SWI1$_{\Delta 239-259}$ Front For	CACAGGCTACACGGTTAATTACAAATCTAGAACTAGTATGGATTTC	*p415TEF-SWI1$_{NQ\Delta 239-259}$-YFP*
SWI1$_{\Delta 239-259}$ Front Rev	TCTGTCCATTGAATTTGAAGCTGGCAAAAATTGTTGATTTGC	*p415TEF-SWI1$_{NQ\Delta 239-259}$-YFP*
SWI1$_{\Delta 239-259}$ Back For	CAACAATTTTTGCCAGCTTCAAATTCAATGGACAGATCATCC	*p415TEF-SWI1$_{NQ\Delta 239-259}$-YFP*
SWI1$_{\Delta 239-259}$ Back Rev	AATGGCAGACTAATTACATGACTCGAGGCCG	*p415TEF-SWI1$_{NQ\Delta 239-259}$-YFP*

Yeast transformation

Yeast were transformed as previously described (184). In brief, cells were spun down at 2500 rpm for 3 minutes, supernatant removed, and cells resuspended in 1 ml H_2O. Cells were then spun down again at 2500 rpm for 3 minutes, supernatant removed, and cells resuspended in 1 ml of 0.1M lithium acetate. After 10 minutes, the cells were pelleted again, supernatant removed, and cells resuspended in 100 µl of Li-PEG (0.1M lithium acetate, 30% polyethylene-glycol 3350 in

H_2O). From this mixture, 94.5 µl of resuspended cells were combined with 3.5 µl of ssDNA and 2.0 µl of the appropriate plasmid. The transformation mixture was then incubated at 42 °C for 30 minutes. Following, the transformation mixture was moved to ice for 5 minutes before spreading onto the appropriate selective media.

Microscopy

Images were captured using a Zeiss Axiovert 200 epifluorescence microscope with an attached camera and the Axiovision AC software (Zeiss, Oberkochen, Germany). Cell samples were visualized with a 100x objective and the appropriate filters for differential interference contrast (DIC), mCherry, or yellow fluorescent protein (YFP). Images were analyzed using Fiji software (185, 186).

RT-PCR

Yeast samples for RT-PCR were grown overnight in selective media (3 ml). The next day, the cultures were spun down at 2500 rpm for 5 min and the media was removed. The cell pellet was resuspended in 1 ml of H_2O before spinning down again at 2500 rpm for 5 min. The supernatant was once again removed, and the pellet resuspended in 600 µl of RLT buffer from the Qiagen RNeasy Mini Kit (Hilden, Germany). The resuspended cells were transferred to a screw-cap tube with silica beads and additional RLT buffer was added to fill the tube to maximum. A Mini-Beadbeater 16 (Bio Spec Products, Bartlesville, OK, USA) was used to lyse the suspended cells by beating 5x in 1-minute intervals with resting on ice for 1 minute in between. Tubes were spun down at 8000 g for 15 seconds.

The clarified lysates were transferred to microcentrifuge tubes and thereafter, the Qiagen RNeasy Mini Kit protocol was followed. RNA concentration was quantified using a Take3 Micro-Volume Plate with a Synergy HT plate reader and Gen5 software (BioTek, Winooski, VT, USA).

Corresponding cDNA was synthesized using the SuperScript III First-Strand Synthesis System (Invitrogen, Carlsbad, CA, USA). The resulting cDNA was immediately used for PCR using the SWI1 SRT For and Rev primers and the ACT1 RT For and Rev primers.

SDS-PAGE

Yeast samples for SDS-PAGE were grown overnight in selective media (3 ml) and prepared via alkaline lysis similarly to previously described (187). The following day, the OD_{600} of the cultures were measured and a volume of culture equal to an OD_{600} of 2.0 was transferred to a microcentrifuge tube. Cells were pelleted at 13000 rpm for 1 minute, media removed, and washed with 500 ul of ice-cold water before being pelleted again. The washed cell pellet was resuspended in 200 µl of 0.1 M NaOH and incubated at room temperature for 10 minutes. After another centrifugation at 13000 rpm for 1 minute, the pellet was resuspended in 50 ul of 2x Laemmli buffer (Bio-Rad, Hercules, CA, USA). Samples were boiled for 10 minutes prior to loading on to a 4-20% Mini-PROTEAN TGX precast protein gel (Bio-Rad, Hercules, CA, USA). After completion of electrophoresis, samples were transferred to a PVDF membrane using an iBlot Dry Blotting System (Invitrogen, Carlsbad, CA, USA).

SDD-AGE

Yeast samples for SDD-AGE were grown overnight in selective media (3 ml) and prepared similarly to previously described (121). The next day, the culture was diluted into a larger volume of selective media (30 ml total) and grown over approximately 4 hours at 30 °C with shaking at 225 rpm. The yeast was harvested afterwards by spinning down at 2500 rpm for 5 minutes. The media was removed and the resulting cell pellet was washed with 10 ml of H_2O. After another spin down, the H_2O was removed and 800 µl of cell lysis buffer (50 mM Tris-HCL pH 7.5, 50 mM KCl, 10 mM $MgCl_2$, 5% glycerol, 10 mM PMSF, cOmplete Mini Protease Inhibitor Cocktail

(Roche, Basel, Switzerland)) was added. The cell suspension was transferred to a 2.0 ml screw-cap tube filled halfway with silica beads and additional cell lysis buffer was added to fill the tube to maximum. A Mini-Beadbeater 16 was used to lyse the suspended cells by beating 5x in 1-minute intervals with resting on ice for 1 minute in between. The resulting samples were then used for SDD-AGE.

SDD-AGE was conducted as described previously (188). Briefly, yeast lysates were first mixed with 4x Laemmli sample buffer (2x TAE, 20% glycerol, 8% SDS, 0.1% bromophenol blue). Samples were either incubated at room temperature for 7 minutes or boiled for 10 minutes. Samples were loaded onto 1.5% agarose, 0.1% SDS gels. After completion of electrophoresis, samples were transferred to PVDF membrane using capillary action and 1x TBS.

Immunoblotting

Membranes were blocked via incubation in 5% milk in PBS at either 4 C overnight or room temperature for 2 hours. Blots were washed 3x for 5 minutes with PBS + 0.01% Tween-20 before probing with primary antibody for 2 hours at room temperature. The following primary antibodies were used for detection: JL-8 anti-GFP antibody (Clontech, Mountain View, CA, USA), anti-Sup35 antibody (gift from the Liebman Lab, University of Nevada, Reno, NV, USA), or anti-actin antibody, clone C4 (Chemicon, Temecula, CA, USA). All primary antibodies were used at a 1:2500 dilution. Blots were washed 3x for 5 minutes with PBS + 0.01% Tween-20 before probing with horseradish peroxidase-conjugated rat anti-mouse secondary antibody (Cell Signaling Technology, Danvers, MA, USA) for 1 hour at room temperature. Blots were washed 3x for 5 minutes with PBS + 0.01% Tween-20 before incubation with Clarity Western ECL Substrate (Bio-Rad, Hercules, CA, USA). Blots were imaged using a ChemiDoc Imaging System (Bio-Rad, Hercules, CA, USA).

De novo prion formation assay

W303 *sup35Δ / p316SUP35$_{FL}$* [*PSI*$^{+}$] cells were independently transformed with each of the *p415TEF-SWI1$_{1-38}$-MC* wild-type and mutant constructs. Transformants were grown on -LU media and a color change to red was observed – indicating that the fusion proteins were functional in translational termination. Three red colonies were selected for each construct and were streaked onto -L+5-FOA media to select against *p316SUP35$_{FL}$*. Resulting colonies were selected and re-streaked onto both -L and -LU media to confirm loss of *p316SUP35$_{FL}$*. Afterwards, three different colonies from each of the three *sup35Δ / p415TEF-SWI1$_{1-38}$Mut-MC* isolates were transformed with the corresponding *p416TEF-SWI1$_{1-38}$Mut-YFP* plasmid. The resulting plates of *sup35Δ / p415TEF-SWI1$_{1-38}$Mut-MC / p416TEF-SWI1$_{1-38}$Mut-MC* colonies were then checked for coloration and aggregation via fluorescence microscopy.

Results

Multiple Swi1$_{1-38}$ mutants cannot co-aggregate with Swi1$_{FL}$ in [*SWI*$^{+}$] cells

To dissect the contributions of various residues to the prionogenicity of Swi1$_{1-38}$, we targeted the minority of residues which are non-asparagine amino acids for mutagenesis (Figure 2.1A). The first two amino-acid residues – methionine and aspartate – were not mutated due to the need for the start codon and the N-end rule, respectively (189). The last amino-acid residue – proline – was also not mutated due to our lab's previous work displaying that this residue is not necessary for aggregation, maintenance of [*SWI*$^{+}$], or prionogenesis (157). Additionally, proline is not known to be particularly aggregation- or prion-promoting.

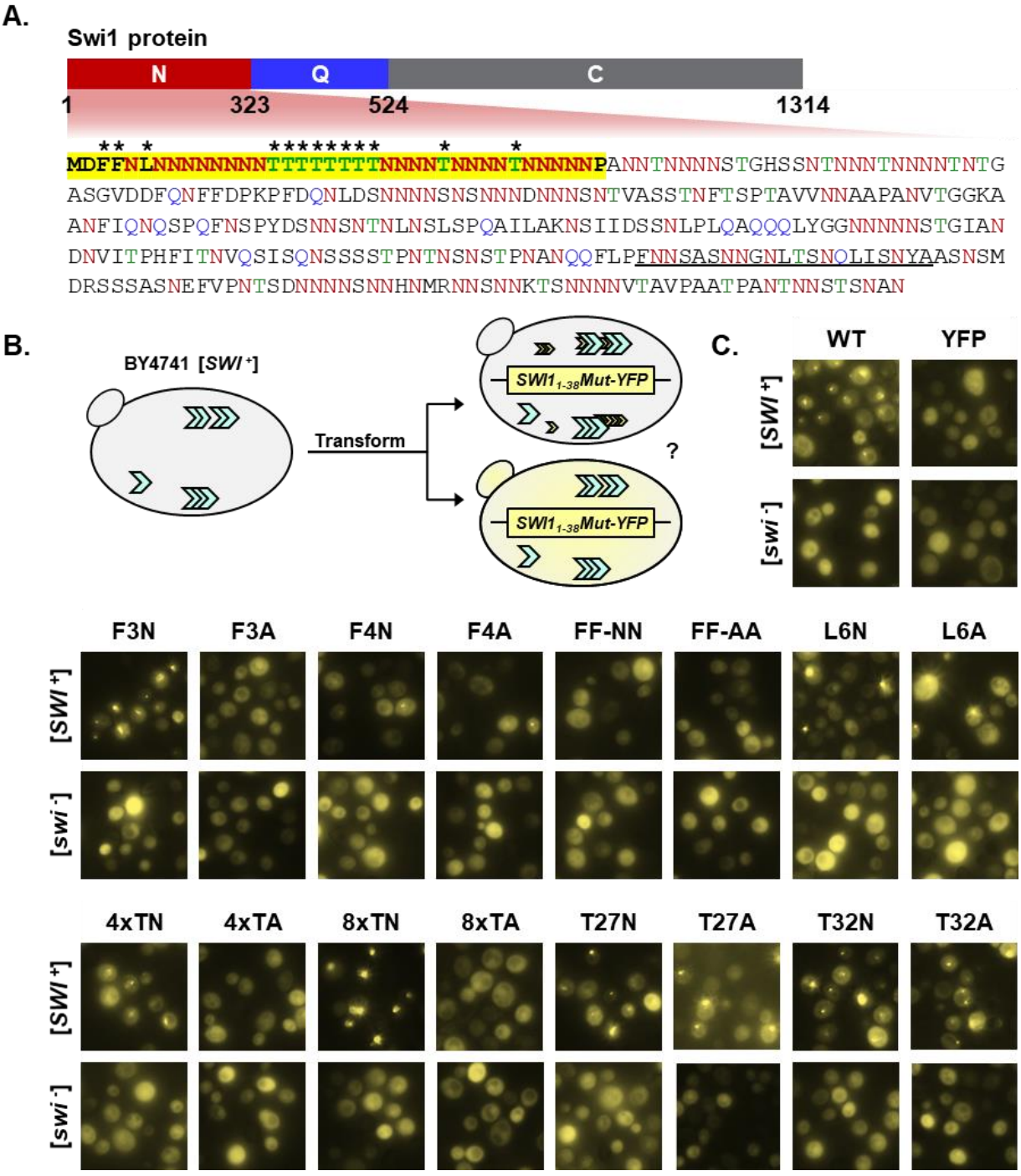

A.
Swi1 protein
N
Q
C
1
323
524
1314
MDFFNLNNNNNNNNTTTTTTTTNNNNTNNNNTNNNNNPANNTNNNNSTGHSSNTNNNTNNNNTNTG
ASGVDDFQNFFDPKPFDQNLDSNNNNSNSNNNDNNNSNTVASSTNFTSPTAVVNNAAPANVTGGKA
ANFIQNQSPQFNSPYDSNNSNTNLNSLSPQAILAKNSIIDSSNLPLQAQQQLYGGNNNNNSTGIAN
DNVITPHFITNVQSISQNSSSSTPNTNSNSTPNANQQFLPFNNSASNNGNLTSNQLISNYAASNSM
DRSSSASNEFVPNTSDNNNNSNNHNMRNNSNNKTSNNNNVTAVPAATPANTNNSTSNAN
B.
BY4741 [SWI+]
Transform
SWI1 1-38 Mut-YFP
SWI1 1-38 Mut-YFP
?
C.
WT
YFP
[SWI+]
[swi-]
F3N
F3A
F4N
F4A
FF-NN
FF-AA
L6N
L6A
[SWI+]
[swi-]
4xTN
4xTA
8xTN
8xTA
T27N
T27A
T32N
T32A
[SWI+]
[swi-]

Figure 2.1. Mutation of the phenylalanine residues or threonine tract disrupts $Swi1_{1-38}$ co-aggregation with $Swi1_{FL}$.

A. Diagram of Swi1 protein domains. The N region ($Swi1_{1-323}$) includes the Swi1 prion domain and the sequence of said region is presented. Amino-acid residues of asparagine (N), glutamine (Q), and threonine (T) are in red, blue, and green, respectively. The first 38 amino acid residues ($Swi1_{1-38}$) are **bolded** and highlighted in yellow. A previously predicted amyloid core region ($Swi1_{239-259}$) is underlined. Asterisks (*) indicate those residues that were targeted for mutagenesis.
B. Diagram of experiment. BY4741 [*SWI*$^{+}$] or [*swi*$^{-}$] cells were transformed with *p415TEF-SWI1*$_{1-38}$*-YFP* (WT), *p415TEF-SWI1*$_{1-38}$*Mut-YFP*, or *p415TEF-YFP*. Transformants were observed using fluorescence microscopy for aggregate foci or diffuse signal.
C. Fluorescence images of BY4741 [*SWI*$^{+}$] or [*swi*$^{-}$] cells transformed with *p415TEF-SWI1*$_{1-38}$*-YFP* (WT), *p415TEF-SWI1*$_{1-38}$*Mut-YFP*, or *p415TEF-YFP*. See Table 1 for amino-acid sequences of mutants. For each construct, 3 independent transformations of both [*SWI*$^{+}$] and [*swi*$^{-}$] yeast were conducted. On average, approximately 900 cells (from across 3 colonies) were observed per transformation. Shown are representative views and quantitative results are shown in Figure 2C.

The remaining non-asparagine residues including phenylalanine (F), threonine (T), and leucine (L) in $Swi1_{1\text{-}38}$ were mutagenized. Noticeably, these amino acids are not overrepresented in the asparagine/glutamine-rich PrDs of identified yeast prions though some of them have been reported as amyloidogenic (167, 190). Codons for individual amino acids were swapped via PCR mutagenesis to codons for either asparagine (N) or alanine (A) (Table 2.4). Asparagine was selected due to its importance in prionogenicity and the fact that $Swi1_{1\text{-}38}$ is already very N-rich – small changes to the number of N residues are unlikely to have a sizeable effect (167). Replacement of the phenylalanine(s) in $Swi1_{1\text{-}38}$ with the polar, uncharged asparagine allows us to determine the value of the singular and/or duplicated phenylalanine(s) and the prionogenic hydrophobicity provided by them. Furthermore, replacing threonine residues allows us to evaluate whether $Swi1_{1\text{-}38}$ requires unique contributions of threonine or its tandem tract. On the other hand, alanine was selected due to its lack of prionogenicity and its simple and small structure – particularly compared to amino acids such as phenylalanine with a large aromatic side chain. The phenylalanine residues at positions 3 and 4 were mutated singularly or in tandem – producing the mutants F3N, F3A, F4N, F4A, and FF-NN and FF-AA. The leucine residue at position 6 was mutated singularly to construct the mutants L6N and L6A. For the threonine tract in the center of $Swi1_{1\text{-}38}$, the last four threonine residues (positions 19-22) were replaced with either all asparagine or all alanine to produce 4xTN or 4xTA, respectively. The entire threonine tract (positions 15-22) was mutated to be either entirely asparagine (8xTN) or entirely alanine (8xTA). The interspersed threonine residues in the back portion of $Swi1_{1\text{-}38}$ were singularly mutated to create T27N, T27A, T32N, and T32A. Together with the wild-type (WT) $Swi1_{1\text{-}38}$ construct, these mutants were initially assayed for their ability to co-aggregate with $Swi1_{FL}$.

Table 2.4. Swi1$_{1-38}$ mutants.

Name	DNA Mutation	Amino-Acid Sequence
WT	-	MDFFNLNNNN NNNNTTTTTT TTNNNNTNNN NTNNNNNP
F3N	TTC → AAC	MD**N**FNLNNNN NNNNTTTTTT TTNNNNTNNN NTNNNNNP
F3A	TTC → GCC	MD**A**FNLNNNN NNNNTTTTTT TTNNNNTNNN NTNNNNNP
F4N	TTT → AAC	MDF**N**NLNNNN NNNNTTTTTT TTNNNNTNNN NTNNNNNP
F4A	TTT → GCC	MDF**A**NLNNNN NNNNTTTTTT TTNNNNTNNN NTNNNNNP
FF-NN	TTCTTT → AACAAC	MD**NN**NLNNNN NNNNTTTTTT TTNNNNTNNN NTNNNNNP
FF-AA	TTCTTT → GCCGCC	MD**AA**NLNNNN NNNNTTTTTT TTNNNNTNNN NTNNNNNP
L6N	TTG → AAC	MDFFN**N**NNNN NNNNTTTTTT TTNNNNTNNN NTNNNNNP
L6A	TTG → GCG	MDFFN**A**NNNN NNNNTTTTTT TTNNNNTNNN NTNNNNNP
4xTN	ACTACTACTACC → AACAACAACAAC	MDFFNLNNNN NNNNTTTT**NN** **NN**NNNNTNNN NTNNNNNP
4xTA	ACTACTACTACC → GCAGCAGCAGCA	MDFFNLNNNN NNNNTTTT**AA** **AA**NNNNTNNN NTNNNNNP
8xTN	ACTACTACTACTACTACTACTAAC → AACAACAACAACAACAACAACAAC	MDFFNLNNNN NNNN**NNNNNN** **NN**NNNNTNNN NTNNNNNP
8xTA	ACTACTACTACTACTACTACTAAC → GCAGCAGCAGCAGCAGCAGCAGCA	MDFFNLNNNN NNNN**AAAAAA** **AA**NNNNTNNN NTNNNNNP
T27N	ACT → AAT	MDFFNLNNNN NNNNTTTTTT TTNNNN**N**NNN NTNNNNNP
T27A	ACT → GCT	MDFFNLNNNN NNNNTTTTTT TTNNNN**A**NNN NTNNNNNP
T32N	ACT → AAC	MDFFNLNNNN NNNNTTTTTT TTNNNNTNNN N**N**NNNNNP
T32A	ACT → GCT	MDFFNLNNNN NNNNTTTTTT TTNNNNTNNN N**A**NNNNNP

Each mutant was tagged with yellow fluorescent protein (YFP) and individually transformed into BY4741 [*SWI*$^{+}$] and [*swi*$^{-}$] yeast (Figure 2.1B). This process was repeated for three biological replicates. Wild-type Swi1$_{1-38}$-YFP served as a positive control – specifically aggregating in [*SWI*$^{+}$] cells – while YFP alone served as a negative control. While F3N displayed aggregation similar to WT, the other phenylalanine mutants displayed greatly hampered aggregation formation (Figure 2.1C). Indeed, the FF-NN and FF-AA constructs were not observed to have any puncta visible. The other mutant constructs that resulted in deficient aggregation were 4xTA and 8xTA; however, 4xTN and 8xTN displayed aggregation akin to WT. Thus, the replacement of these threonine residues with alanine removed the polar side-groups that are aggregation-prone and greatly disrupted aggregation. On the other hand, maintaining that polarity via mutation to the similarly polar asparagine allowed for aggregation. The remaining mutations (L6N, L6A, T27N, T27A, T32N, T32A) had aggregation similar to WT. All constructs did not produce observable aggregates in [*swi*$^{-}$] cells – indicating that the observed aggregation was

specific to $Swi1_{1-38}$Mut-YFP adopting the prion fold of the existing [*SWI*$^+$] and not amorphous aggregates forming solely due to overexpression.

Overexpression of $Swi1_{FL}$ allows co-aggregation of additional $Swi1_{1-38}$ mutants

To further examine the aggregation capabilities of the $Swi1_{1-38}$ mutant constructs, we performed the same co-aggregation assay in the presence of higher $Swi1_{FL}$ expression (Figure 2.2A). We started with BY4741 *swi1Δ / p416TEF-SWI1$_{FL}$* [*SWI*$^+$] and [*swi*$^-$] yeast for this experiment. In these cells, instead of expressing *SWI1$_{FL}$* from its chromosomal locus under its endogenous promoter, *SWI1$_{FL}$* was expressed from a plasmid under the significantly stronger *TEF* promoter. These conditions should provide an overexpression context for $Swi1_{FL}$ and additional opportunity for the $Swi1_{1-38}$ mutants to decorate the existing [*SWI*$^+$] aggregates.

Once again, based on three biological replicates, several mutants exhibited aggregation akin to WT (Figure 2.2B). These mutants included F3N, L6N, L6A, 4xTN 8xTN, T27N, T27A, T32N, and T32A. These results once again highlighted that most of the singular mutants were capable of co-aggregating and adopting the prion conformation of $Swi1_{FL}$. Interestingly, the other singular phenylalanine mutants (F3A, F4N, and F4A) that had low (<20%) aggregation rates in the initial assay displayed increased aggregation frequency (~60%) under $Swi1_{FL}$-overexpression conditions (Figure 2.2C). On the other hand, the double-phenylalanine mutants (FF-NN and FF-AA) displayed greatly impaired ability to decorate $Swi1_{FL}$ aggregates; however, there were low levels of observable puncta. Another mutant, 4xTA also displayed low aggregation frequency (~20%) in the $Swi1_{FL}$-overexpression context whereas aggregates were not seen under the non-$Swi1_{FL}$-overexpression conditions (~0%). The 8xTA mutant was unable to form notable aggregates even under the $Swi1_{FL}$-overexpression conditions – suggesting that replacing the polar threonine tract with small, hydrophobic, nonpolar alanine residues in the center of $Swi1_{1-38}$ likely

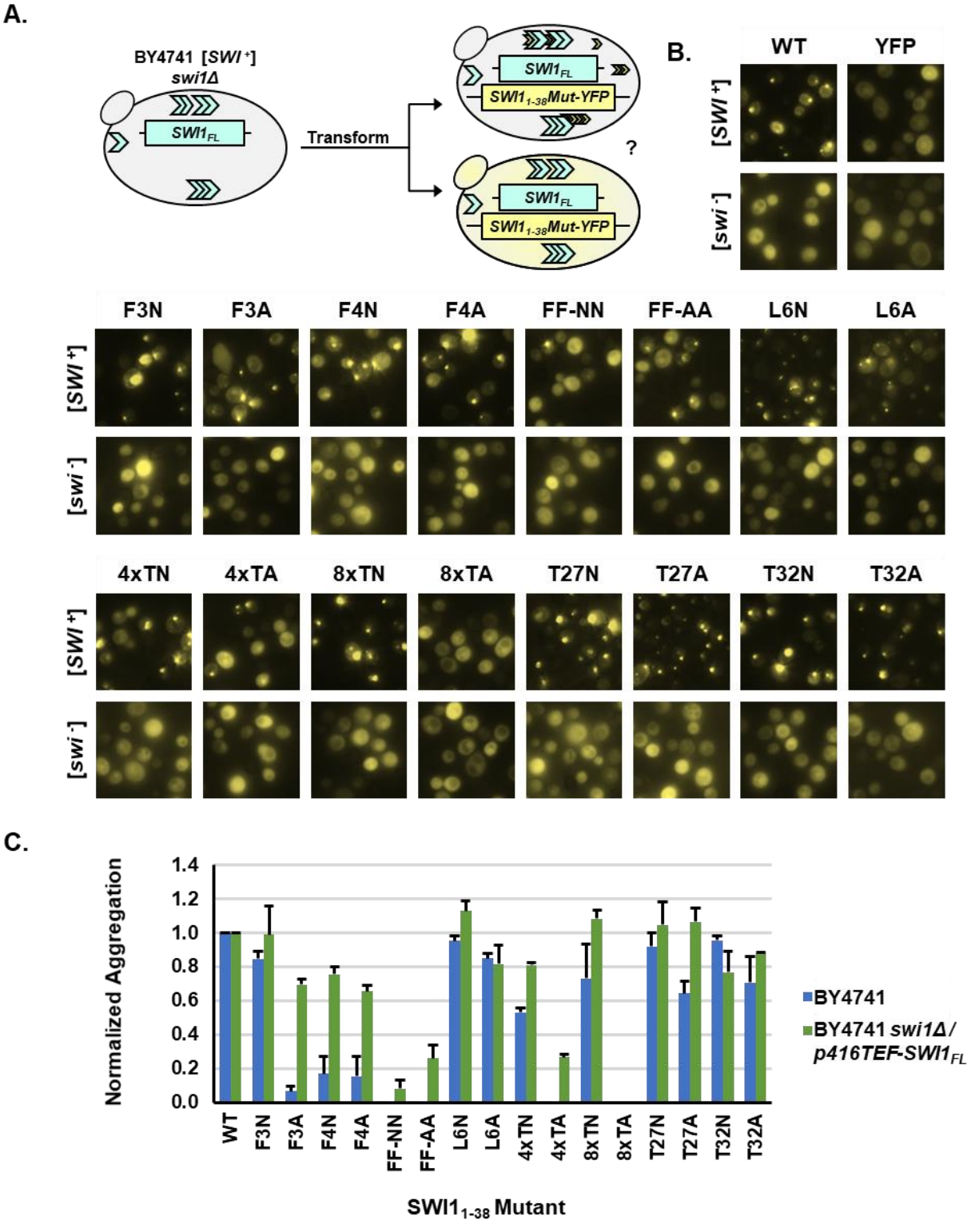

A.
BY4741 [SWI+]
swi1Δ
SWI1FL
Transform
SWI1FL
SWI1 1-38 Mut-YFP
?
SWI1FL
SWI1 1-38 Mut-YFP
B.
WT
YFP
[SWI+]
[swi-]
F3N
F3A
F4N
F4A
FF-NN
FF-AA
L6N
L6A
[SWI+]
[swi-]
4xTN
4xTA
8xTN
8xTA
T27N
T27A
T32N
T32A
[SWI+]
[swi-]
C.
Normalized Aggregation
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
WT
F3N
F3A
F4N
F4A
FF-NN
FF-AA
L6N
L6A
4xTN
4xTA
8xTN
8xTA
T27N
T27A
T32N
T32A
SWI1 1-38 Mutant
BY4741
BY4741 swi1Δ/
p416TEF-SWI1FL

Figure 2.2. Substantially higher $Swi1_{FL}$ expression promotes aggregation of $Swi1_{1-38}$ mutants in [*SWI*$^+$] cells.

A. Diagram of experiment. BY4741 *swi1Δ / p416TEF-SWI1*$_{FL}$ [*SWI*$^+$] or [*swi*$^-$] cells were transformed with *p415TEF-SWI1*$_{1-38}$*Mut-YFP*. Transformants were observed using fluorescence microscopy for aggregate foci or diffuse signal.
B. Fluorescence images of BY4741 *swi1Δ / p416TEF-SWI1*$_{FL}$ [*SWI*$^+$] or [*swi*$^-$] cells transformed with *p415TEF-SWI1*$_{1-38}$*-YFP* (WT), *p415TEF-SWI1*$_{1-38}$*Mut-YFP*, or *p415TEF-YFP*. See Table 1 for amino-acid sequences of mutants. Similar to that described for Figure 1C, for each construct, 3 independent transformations of both [*SWI*$^+$] and [*swi*$^-$] yeast were conducted. On average, approximately 900 cells (from across 3 colonies) were observed per transformation. Shown are representative views and quantitative results are shown in Figure 2C.
C. Quantification of $Swi1_{1-38}$Mut-YFP aggregation observed in Figure 1C and Figure 2B. Cells were manually counted using Fiji software. Normalized aggregation is defined as the number of aggregate-containing cells divided by the total number of cells with YFP fluorescence and then normalized to WT (which had a raw aggregation percentage of ~50-80%) per biological replicate. The mean number of cells with YFP fluorescence observed per mutant per replicate was approximately 900. Error bars represent standard error of the mean.

was disruptive to the aggregation core of the protein. No construct resulted in consistent aggregation in [*swi*$^-$] cells; although, single cells with puncta were observed in a minimal (<5%) number of colonies. These rare instances in the originally [*swi*$^-$] cells likely reflect randomly generated Swi1$_{1\text{-}38}$ or Swi1$_{FL}$ aggregates from the very favorable overexpression conditions.

Although there was an increase in Swi1$_{1\text{-}38}$ mutant co-aggregation when Swi1$_{FL}$ was overexpressed, the most deleterious mutants still had significant effects. The removal of aromatic groups via the replacement of the phenylalanine residues displayed a stepwise effect on aggregation frequency with removal of both leading to a greater decrease than just one. A peculiar site-specific effect was observed as well with the lack of an effect on aggregation with the F3N mutant versus F4N. Moreover, whether the aggregates of these and other Swi1$_{1\text{-}38}$ mutants were stable without the presence of Swi1$_{FL}$ aggregates remained an open question.

Swi1$_{1\text{-}38}$ requires a phenylalanine to maintain [*SWI*$^+$]

We proceeded to use the BY4741 *swi1Δ / p416TEF-SWI1$_{FL}$ / p415TEF-SWI1$_{1\text{-}38}$Mut-YFP* [*SWI*$^+$] transformants to investigate the maintenance of the prion fold by the various Swi1$_{1\text{-}38}$ mutants in the absence of Swi1$_{FL}$ (Figure 2.3A). To do so, isolates containing aggregates were transferred to media containing 5-fluoroorotic acid (5-FOA). Cells containing the URA3 marker – in this case on the plasmid *p416TEF-SWI1$_{FL}$* – would process 5-FOA into a toxic chemical, killing the cells. Thus, using this selection system, we generated cells with the *swi1Δ / p415TEF-SWI1$_{1\text{-}38}$Mut-YFP* genotype that have no full-length Swi1 present.

Among these newly generated yeast isolates, we examined whether individual colonies contained aggregates – indicating maintenance of an adopted prion fold (Figure 2.3B, 2.3C). The single phenylalanine mutants (F3N, F3A, F4N, F4A) displayed observable puncta in cells in ~50% of colonies. However, once both phenylalanine residues were replaced with either asparagine or

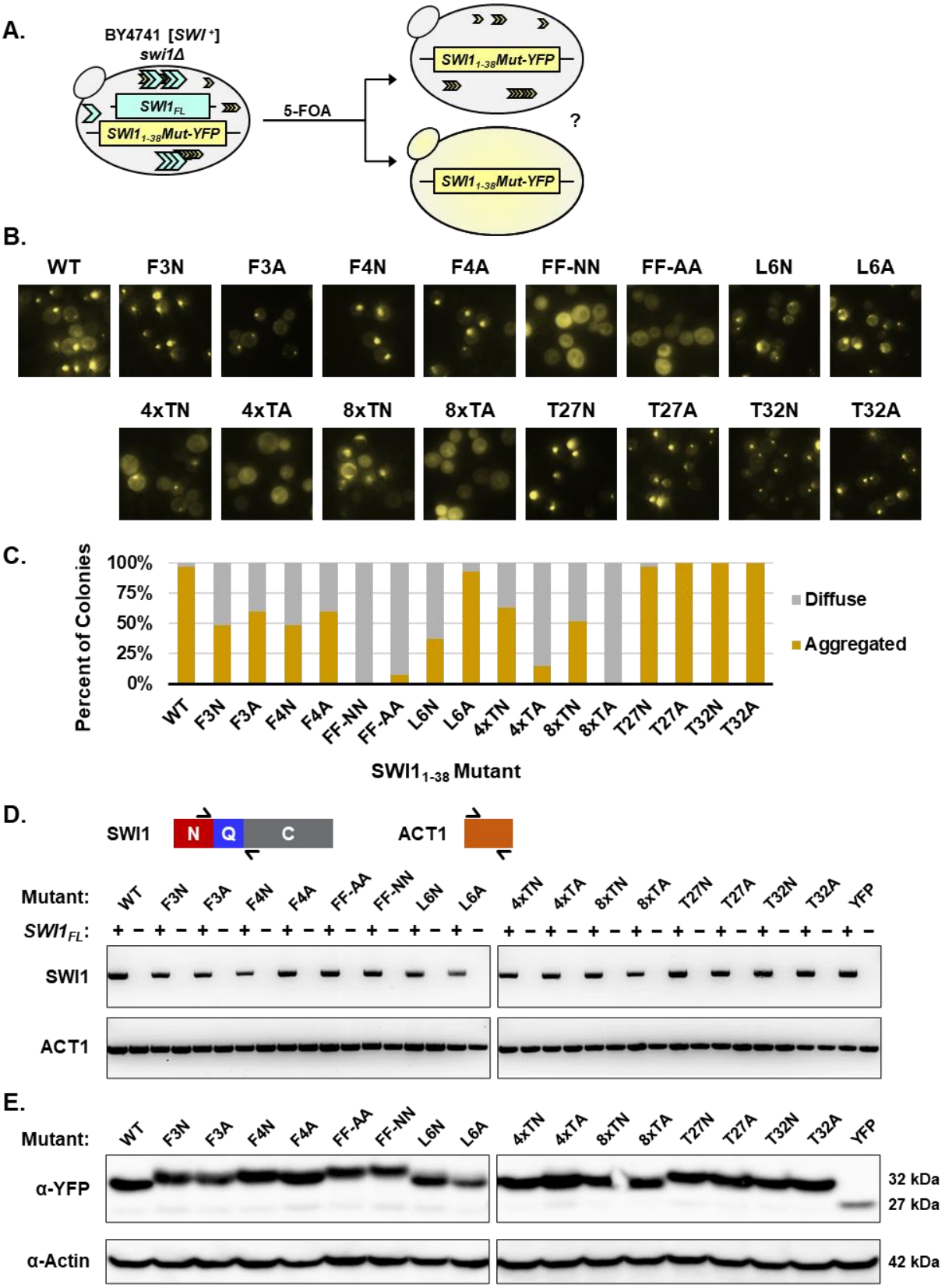

A.
BY4741 [SWI+]
swi1Δ
SWI1FL
SWI11-38Mut-YFP
5-FOA
SWI11-38Mut-YFP
?
SWI11-38Mut-YFP
B.
WT
F3N
F3A
F4N
F4A
FF-NN
FF-AA
L6N
L6A
4xTN
4xTA
8xTN
8xTA
T27N
T27A
T32N
T32A
C.
Percent of Colonies
100%
75%
50%
25%
0%
Diffuse
Aggregated
SWI11-38 Mutant
D.
SWI1
N
Q
C
ACT1
Mutant:
SWI1FL:
SWI1
ACT1
YFP
E.
Mutant:
α-YFP
α-Actin
32 kDa
27 kDa
42 kDa

Figure 2.3. Mutation of both phenylalanine residues leads to Swi1$_{1-38}$ being unable to maintain the prion fold in the absence of Swi1$_{FL}$.

A. Diagram of experiment. BY4741 *swi1Δ* / *p416TEF-SWI1$_{FL}$* / *p415TEF-SWI1$_{1-38}$Mut-YFP* [*SWI^{+}*] cells containing aggregates were treated with 5-FOA to select against cells containing the *p416TEF-SWI1$_{FL}$* plasmid. Resulting BY4741 *swi1Δ* / *p415TEF-SWI1$_{1-38}$Mut-YFP* cells were observed using fluorescence microscopy for aggregate foci or diffuse signal.
B. Representative fluorescence images of the resulting BY4741 *swi1Δ* / *p415TEF-SWI1$_{1-38}$Mut-YFP* cells. See Table 1 for amino-acid sequences of mutants. For each construct, 3 aggregate-containing BY4741 *swi1Δ* / *p416TEF-SWI1$_{FL}$* / *p415TEF-SWI1$_{1-38}$Mut-YFP* isolates were treated with 5-FOA to drop out the full-length Swi1-expression plasmid. From there, 9 colonies for each isolate (for a total of 27 colonies) were examined for each construct. Shown are representative views and quantitative results are shown in Figure 3C.
C. Quantification of yeast colonies retaining Swi1$_{1-38}$Mut-YFP aggregates after removal of Swi1$_{FL}$ via 5-FOA treatment. In experiments related to Figure 3, aggregated colonies had >25% of cells containing aggregates and diffuse colonies showed aggregation in <5% of cells. Percent of colonies was calculated as the number of each of the two types of colonies (aggregated, diffuse) divided by the total number of colonies examined for each construct.
D. Diagram of RT-PCR primer targets and resulting RT-PCR amplification visualized by agarose gel. Primers flanking the Q region of *SWI1* were used to confirm the loss of *SWI1$_{FL}$* in the BY4741 *swi1Δ* / *p415TEF-SWI1$_{1-38}$Mut-YFP* cells. Primers covering *ACT1* were used as a positive control. Samples labeled as + *p416TEF-SWI1$_{FL}$* correspond to the pre-5-FOA BY4741 *swi1Δ* / *p415TEF-SWI1$_{1-38}$Mut-YFP* cells. Samples labeled as - *p416TEF-SWI1$_{FL}$* correspond to the post-5-FOA BY4741 *swi1Δ* / *p415TEF-SWI1$_{1-38}$Mut-YFP* cells.
E. Western blot of BY4741 *swi1Δ* / *p415TEF-SWI1$_{1-38}$Mut-YFP* cells. Membrane was probed with either α-GFP or α-actin. Estimated molecular weights based on sequence are listed at right.

alanine as in FF-NN and FF-AA, aggregation was almost completely abolished – only one colony was found to contain any aggregation. This result indicates that the phenylalanine residues play a pivotal role in maintaining $Swi1_{1\text{-}38}$ aggregation.

Other mutants also displayed deficiencies in maintaining aggregation once $Swi1_{FL}$ was removed. The 4xTA mutant had aggregates in ~10% of colonies while no observed 8xTA colonies contained aggregates (Figure 2.3C). The substitution of the polar threonine residues with the nonpolar alanine residues likely disrupted stability of any prion fold adopted by $Swi1_{1\text{-}38}$. Meanwhile, the 4xTN and 8xTN mutants both presented with a reduced maintenance ability (~60% and ~50% of colonies, respectively) – suggesting that the presence of the threonine tract remains important but not required for $Swi1_{1\text{-}38}$ aggregation. Once again, the T27N, T27A, T32N, and T32A mutants did not present with any meaningful deviation from the WT control. In addition, the L6A mutant did not demonstrate impairment in maintaining aggregation. The L6N mutant did display a substantial decrease in the number of colonies with cells containing aggregates. This difference between the L6A and L6N mutants may be due to the similarities between alanine and leucine – both nonpolar, hydrophobic amino acids – as opposed to the polarity introduced by an asparagine residue.

Cells containing $Swi1_{1\text{-}38}$Mut-YFP aggregates were examined for the curability of this aggregation. For $Swi1_{1\text{-}38}$ WT and the various mutants, multiple isolates were streaked onto selective media containing 5 mM guanidine hydrochloride (GdnHCl). GdnHCl cures – or rids cells of many endogenous yeast prions including [*SWI*$^{+}$] – through inactivation of the molecular chaperone Hsp104 (118). Treatment with GdnHCl resulted in loss of $Swi1_{1\text{-}38}$Mut-YFP aggregation for WT and all aggregate-maintaining mutants (data not shown). This result indicates that the aggregation was prion in nature. Additionally, we tested whether the aggregated $Swi1_{1\text{-}}$

$_{38}$Mut-YFP could transmit its prion fold back to Swi1$_{FL}$ by transforming the BY4741 *swi1Δ* / *p415TEF-SWI1$_{1-38}$Mut-YFP* cells with *p416TEF-SWI1$_{FL}$-mCherry*, which allows for expression of a fluorescently tagged version of Swi1$_{FL}$ (Figure 2.4A). Colonies containing cells with Swi1$_{1-38}$Mut-YFP aggregates also displayed Swi1$_{FL}$-mCherry aggregates when visualized via fluorescence microscopy and the puncta of Swi1$_{1-38}$Mut-YFP and Swi1$_{FL}$-mCherry were largely co-localized (Figure 2.4B). Mutants unable to maintain aggregation without Swi1$_{FL}$, and thus, not having aggregates to support transmission of a prion fold back to Swi1$_{FL}$-mCherry, did not display any mCherry foci. Thus, these aggregates present in the BY4741 *swi1Δ* / *p415TEF-SWI1$_{1-38}$Mut-YFP* cells could transmit a prion fold back to Swi1$_{FL}$ – indicating that the observed Swi1$_{1-38}$Mut-YFP aggregates are prion in nature.

To confirm the validity of these results, the yeast isolates were checked for *SWI1$_{FL}$* mRNA via RT-PCR using a pair of primers in the *SWI1* coding region but downstream of *SWI1$_{1-38}$* to verify the absence of *SWI1$_{FL}$* expression (Figure 2.3D). We confirmed that none of the examined isolates after 5-FOA treatment contained *SWI1$_{FL}$* (Figure 2.3D). The expression of the mutant Swi1$_{1-38}$ constructs was also examined at the protein level to address the possibility that the results could be influenced by variations in expression level rather than the mutations. No notable differences were observed when assessed via Western blot (Figure 2.3E). Slight variations in band location were seen on the blot, but these differences were likely due to the changes in molecular weight and electrophoretic mobility due to the mutations combined with a high acrylamide percentage on a gradient gel.

Loss of both phenylalanine residues disrupts *de novo* prion formation by Swi1$_{1-38}$

We next examined if these Swi1$_{1-38}$ mutants were able to *de novo* form a prion. Our lab's previous research established that Swi1$_{1-38}$ can act as a bona fide prion domain and *de novo* form

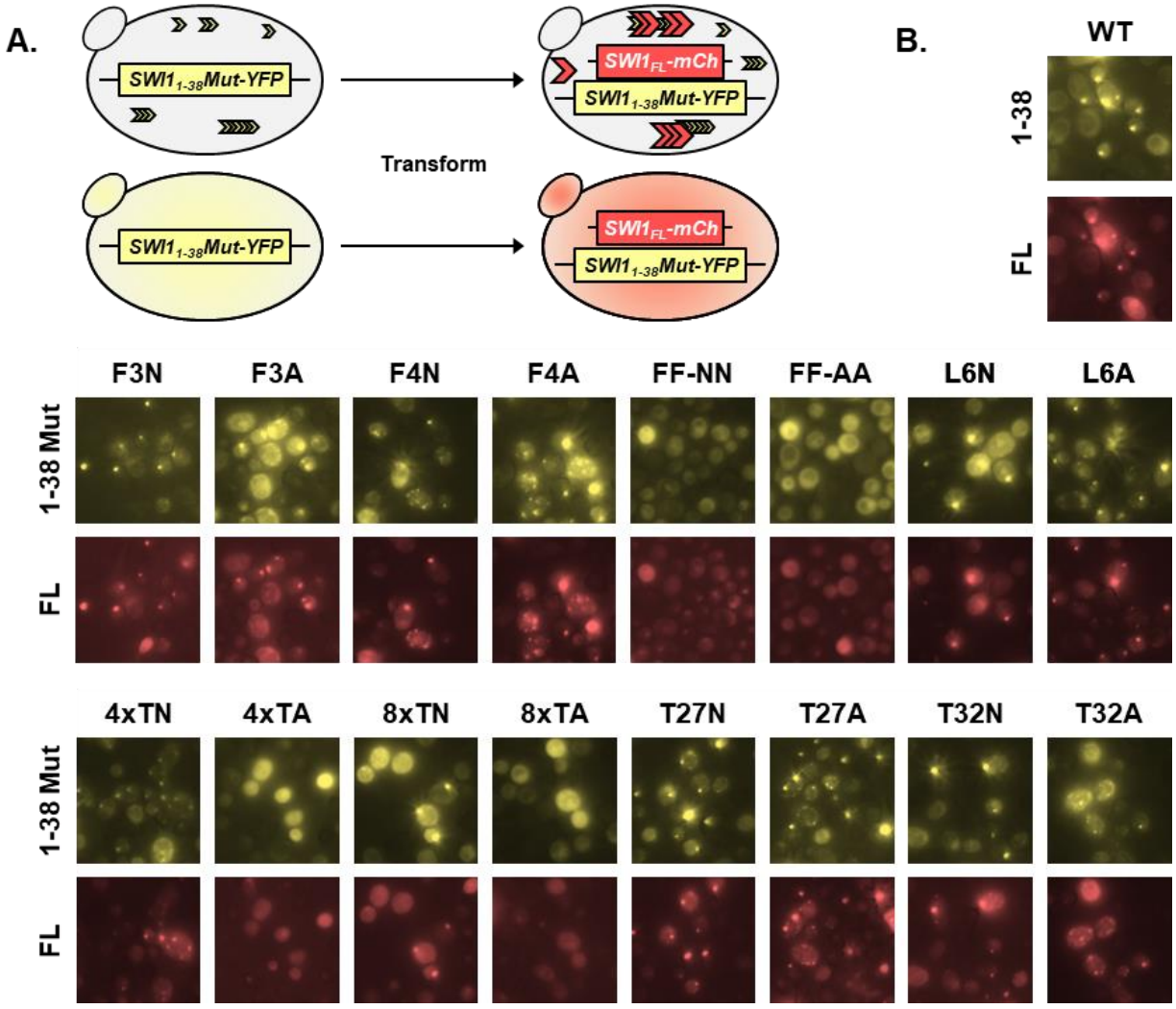


Figure 2.4. Mutant Swi1$_{1\text{-}38}$ aggregates can transmit the prion fold back to Swi1$_{FL}$.

A. Diagram of experiment. BY4741 *swi1Δ* / *p415TEF-SWI1$_{1\text{-}38}$Mut-YFP* cells were transformed with *p416TEF-SWI1$_{FL}$-mCherry*. Transformants were observed using fluorescence microscopy for aggregate foci or diffuse signal.

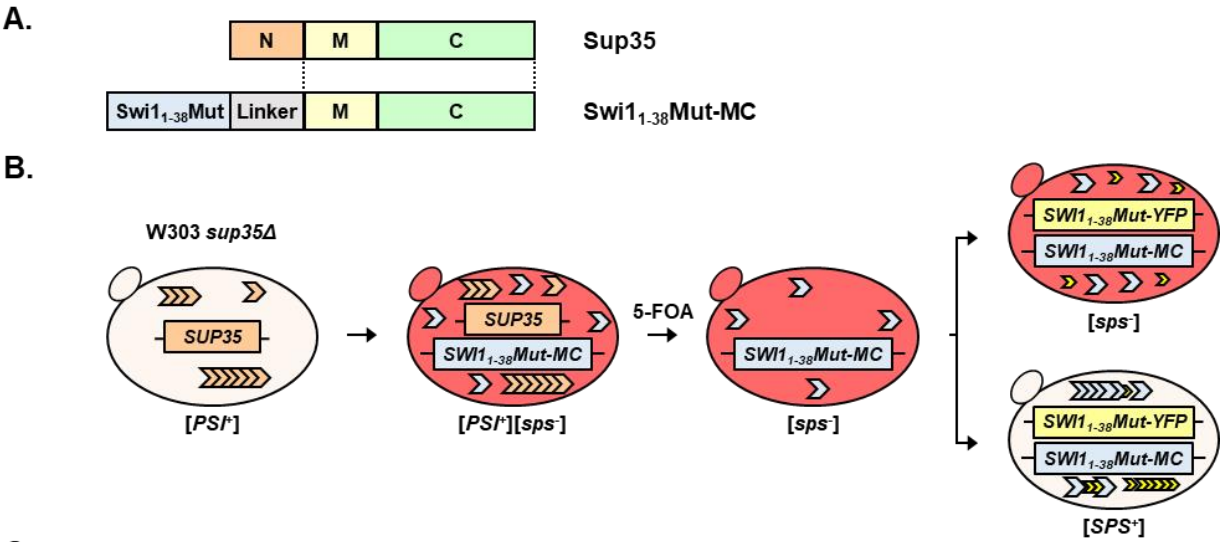
B. Fluorescence images of the resulting BY4741 *swi1Δ* / *p415TEF-SWI1$_{1\text{-}38}$Mut-YFP* / *p16TEF-SWI1$_{FL}$-mCherry* cells. See Table 1 for amino-acid sequences of mutants. Samples were imaged with the appropriate filters for Swi1$_{1\text{-}38}$-YFP / Swi1$_{1\text{-}38}$Mut-YFP (1-38 / 1-38 Mut) and Swi1$_{FL}$-mCherry (FL). For each construct, 3 different transformants were examined. Shown are representative views.

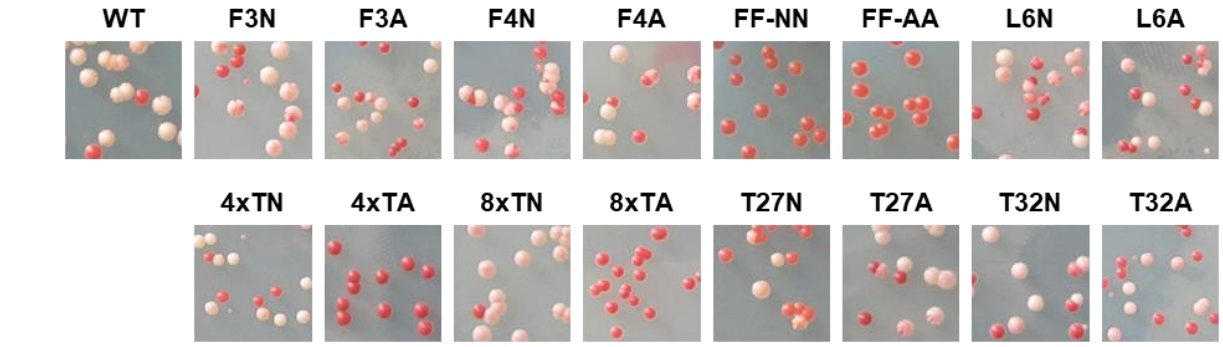
a prion termed [*SPS*$^{+}$] (157). To examine the ability of the Swi1$_{1\text{-}38}$ mutants to do so, we employed the widely used Sup35 assay – in which a prion or prion-like domain of interest is attached to the MC regions of Sup35 in place of its own prion-domain-containing N region (107, 109).

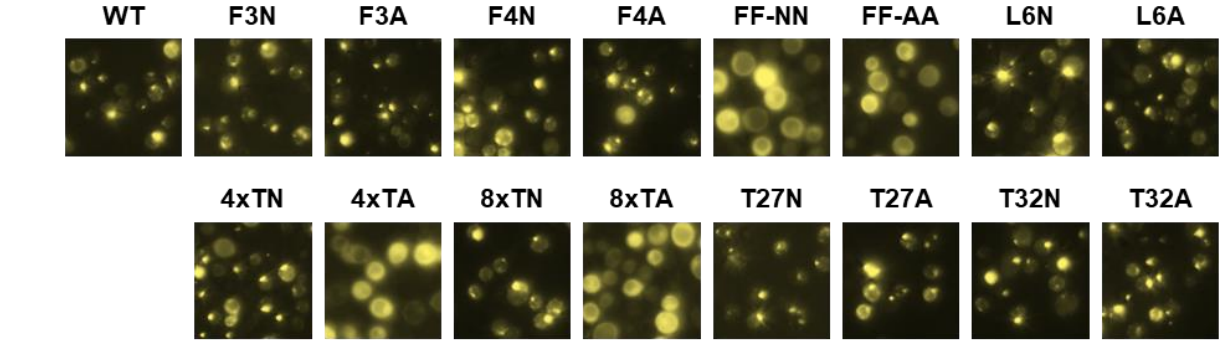
Sup35 functions normally as a translation terminator in yeast, and this function combined with a genetic alteration to the *ADE1* gene (*ade1-14*) provides a useful tool for evaluating prionogenesis (115). Under non-prion conditions, Sup35 acts an efficient translation terminator, recognizes the premature stop codon introduced with *ade1-14*, and prevents creation of a necessary enzyme in the adenine synthesis pathway. This prevention results in the build-up of an adenine precursor that provides the yeast cells with a red hue. When prionized, Sup35 can no longer efficiently function as a translation terminator and read-through of the premature stop codon results in the production of the requisite enzyme. This situation results in the synthesis of adenine and little build-up of the red-colored adenine precursor – leading to the yeast colonies having a white or light-pink coloration.

To initialize the assay, Swi1$_{1\text{-}38}$ mutants were linked to Sup35$_{MC}$ and transformed into W303 *sup35Δ / p316SUP35* [*PSI*$^{+}$] yeast provided by the Weissman laboratory (Figure 2.5A, 2.5B). After confirmation of a white-to-red color change indicating that the Swi1$_{1\text{-}38}$Mut-MC fusions were functional, the *SUP35* plasmid was removed via treatment with 5-FOA. From there, three red isolates for each Swi1$_{1\text{-}38}$Mut-MC fusion were transformed with p415TEF-SWI1$_{1\text{-}38}$Mut-YFP in order to provide an overexpression of Swi1$_{1\text{-}38}$Mut to induce *de novo* prion formation at a high rate. Additionally, we confirmed via Western blot that the expression of Swi1$_{1\text{-}38}$Mut-MC was consistent among the different mutants (Figure 2.5E).

Majority of the Swi1$_{1\text{-}38}$ constructs produced colonies with colors indicative of prionization – e.g., white, light pink, sectored with multiple hues (Figure 2.5C). In addition to WT, these


A.
N
M
C
Sup35
Swi1 1-38 Mut
Linker
M
C
Swi1 1-38 Mut-MC
B.
W303 sup35Δ
SUP35
[PSI+]
SUP35
Swi1 1-38 Mut-MC
[PSI+][sps-]
5-FOA
Swi1 1-38 Mut-MC
[sps-]
Swi1 1-38 Mut-YFP
Swi1 1-38 Mut-MC
[sps-]
Swi1 1-38 Mut-YFP
Swi1 1-38 Mut-MC
[SPS+]


C.
WT
F3N
F3A
F4N
F4A
FF-NN
FF-AA
L6N
L6A
4xTN
4xTA
8xTN
8xTA
T27N
T27A
T32N
T32A


D.
WT
F3N
F3A
F4N
F4A
FF-NN
FF-AA
L6N
L6A
4xTN
4xTA
8xTN
8xTA
T27N
T27A
T32N
T32A

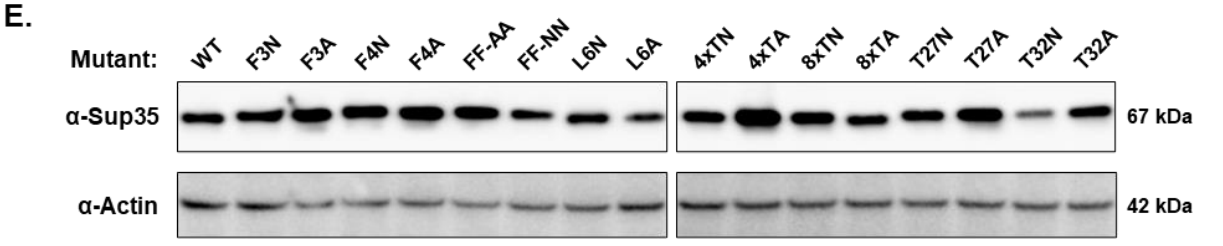

E.
Mutant:
WT
F3N
F3A
F4N
F4A
FF-AA
FF-NN
L6N
L6A
4xTN
4xTA
8xTN
8xTA
T27N
T27A
T32N
T32A
α-Sup35
67 kDa
α-Actin
42 kDa

Figure 2.5. Swi1$_{1-38}$ requires at least one of its phenylalanine residues for *de novo* prion formation.

A. Diagram of fusion proteins created. Swi1$_{1-38}$Mut was linked to Sup35$_{MC}$ via a DPGGPGGG linker to allow for usage of the Sup35 assay for *de novo* prion formation.
B. Diagram of experiment. W303 *sup35Δ* / *p316SUP35$_{FL}$* [*PSI*$^{+}$] cells were transformed with *p415TEF-SWI1$_{1-38}$-MC* (WT) or *p415TEF-SWI1$_{1-38}$Mut-MC*. The transformants were treated with 5-FOA to select against cells containing the *p316SUP35$_{FL}$* plasmid. Resulting W303 *sup35Δ* / *p415TEF-SWI1$_{1-38}$Mut-MC* cells were then transformed with *p416TEF-SWI1$_{1-38}$-YFP* (WT) or the corresponding *p416TEF-SWI1$_{1-38}$Mut-YFP*. Transformants were grown and spread onto -LU plates and checked for color change corresponding to prionization of the Swi1$_{1-38}$Mut-MC protein.
C. Representative images of W303 *sup35Δ* / *p415TEF-SWI1$_{1-38}$Mut-MC* / *p416TEF-SWI1$_{1-38}$Mut-YFP* colonies on -LU plates. Images are representative of full-plate images captured of 3 biological replicates.
D. Representative fluorescence images of W303 *sup35Δ / p415TEF-SWI1$_{1-38}$Mut-MC / p416TEF-SWI1$_{1-38}$Mut-YFP* yeast. Prion-forming constructs display aggregation visualized from cells from white or light-pink colonies. Constructs unable to form [*SPS*$^{+}$] (FF-NN, FF-AA, 4xTA, 8xTA) display diffuse signal as seen in cells from red colonies. Images are representative of multiple examined colonies for each mutant.
E. Western blot of W303 *sup35Δ* / *p415TEF-SWI1$_{1-38}$Mut-MC* / *p416TEF-SWI1$_{1-38}$Mut-YFP* cell lysates. Membrane was probed with either α-Sup35 or α-actin. Estimated molecular weights based on sequence are listed at right.

constructs included the single phenylalanine mutants (F3N, F3A, F4N, F4A), the other single residue mutants (L6N, L6A, T27N, T27A, T32N, T32A), and the threonine tract asparagine mutants (4xTN, 8xTN). When treated with GdnHCl, the majority of non-red colonies could be reverted to red – indicating curability (data not shown). While replacing one phenylalanine with an asparagine or alanine did not disrupt prion formation, replacing both phenylalanine residues (FF-NN, FF-AA) completely abolished *de novo* prion formation by $Swi1_{1\text{-}38}$. The aromaticity of particular amino-acid sidechains may play a crucial role in nucleating the prion fold – explaining the lack of prion formation of the FF-NN and FF-AA mutants. The 4xTA and 8xTA mutants also led to $Swi1_{1\text{-}38}$ losing its prion forming ability. In this case, the addition of multiple alanine residues with their small methyl side chains in what otherwise would be a long stretch of polar residues proved deleterious – as swapping one polar amino acid for another polar amino acid (as in 4xTN, 8xTN) did not result in a similar prionization impairment.

We examined the generated *sup35Δ* / *p415TEF-SWI1*$_{1\text{-}38}$*Mut-MC* / *p416TEF-SWI1*$_{1\text{-}38}$*Mut-YFP* colonies for aggregation using fluorescence microscopy. There were a small number of wholly white colonies that presented on the FF-NN, FF-AA, 4xTA, and 8xTA plates. All such colonies were checked for aggregate formation and none contained puncta of any sort – indicating they were non-prion cells containing mutations in the adenine synthetic pathway (Figure 2.5D, data not shown). On the other hand, randomly selected white, light pink, or sectored colonies from among all other mutants and WT displayed aggregates (Figure 2.5D).

We treated the [*SPS*$^{+}$] colonies with 5-FOA to select against the *p416TEF-SWI1*$_{1\text{-}38}$*Mut-YFP* plasmid. This process removes a portion of the overexpression condition by theoretically halving the overall expression of *SWI1*$_{1\text{-}38}$*Mut*. After treatment with 5-FOA, some colonies – regardless of mutation – stably maintained the [*SPS*$^{+}$] phenotypes while others did not (data not

shown). This result was not surprising as the higher overexpression condition with the *p416TEF-SWI1$_{1-38}$Mut-YFP* plasmid present likely supported the maintenance of weaker variants. Thus, the inability of the double phenylalanine mutants to *de novo* form [*SPS*$^{+}$] even in the highly favorable two plasmid overexpression condition indicates that these mutations indeed abolish the prion-forming capability of Swi1$_{1-38}$-MC.

Double phenylalanine mutants do not form high MW, SDS-resistant aggregates

With aggregation, maintenance of the [*SWI*$^{+}$] prion fold, and *de novo* prion formation by Swi1$_{1-38}$ being deleteriously affected by replacement of its phenylalanine residues, we next examined these mutants via semi-denaturing detergent agarose gel electrophoresis (SDD-AGE). This technique allows identification of high-molecular-weight, detergent-resistant protein aggregates. We cultivated both BY4741 *swi1Δ / p415TEF-SWI1$_{1-38}$Mut-YFP* and W303 *sup35Δ / p415TEF-SWI1$_{1-38}$Mut-MC / p416TEF-SWI1$_{1-38}$Mut-YFP* yeast to check for such aggregates of Swi1$_{1-38}$Mut-YFP and Swi1$_{1-38}$Mut-MC, respectively.

In the presence of 2% SDS, we found that the WT Swi1$_{1-38}$-YFP protein formed a noticeable smear of high-MW, SDS-resistant species when probed with anti-YFP (Figure 2.6A). Upon boiling, the high-MW species dissembled to become low-MW monomers (Figure 2.6A). The single phenylalanine mutants (F3N, F3A, F4N, F4A) all form a similar smear. On the other hand, both FF-NN and FF-AA only display a faint lower banding that corresponds to where monomeric species are found (as seen in the boiled WT sample). These results correlate with the minimal aggregation observed in the FF-NN and FF-AA samples and the inability of these mutants to maintain the [*SWI*$^{+}$] fold in the absence of Swi1$_{FL}$.

Similarly, high-MW, SDS-resistant forms of WT Swi1$_{1-38}$-MC were seen by SDD-AGE (Figure 2.6B). The single phenylalanine mutants displayed much the same pattern as WT and the

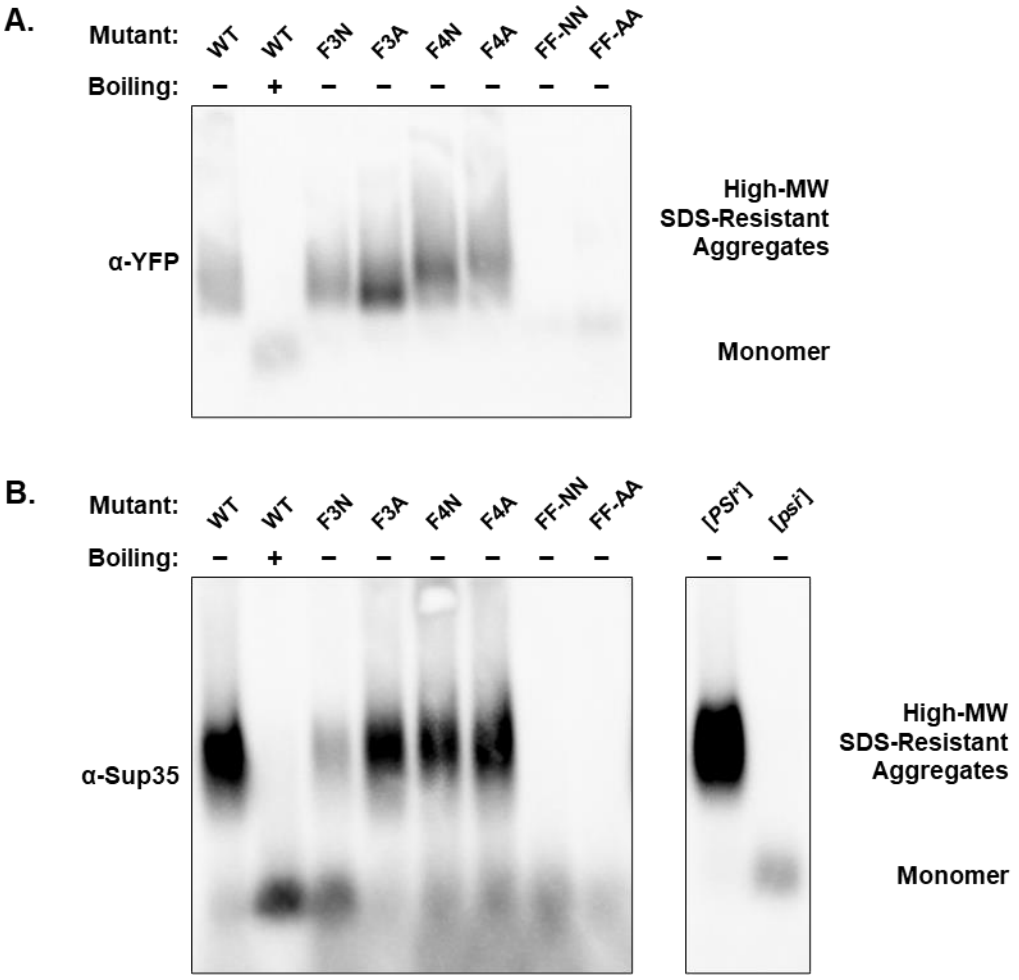


Figure 2.6. Swi1$_{1\text{-}38}$ no longer forms high-molecular-weight, SDS-resistant aggregates when both phenylalanine residues are replaced.

A. Blot of SDD-AGE of BY4741 *swi1Δ* / *p415TEF-SWI1$_{1\text{-}38}$Mut-YFP* cells with (+) or without (-) boiling. Membrane was probed with α-GFP to detect Swi1$_{1\text{-}38}$Mut-YFP.
B. Blot of SDD-AGE of W303 *sup35Δ* / *p415TEF-SWI1$_{1\text{-}38}$Mut-MC* / *p416TEF-SWI1$_{1\text{-}38}$Mut-YFP* cells with (+) or without (-) boiling and W303 *sup35Δ / p316SUP35$_{FL}$* cells. Membrane was probed with α-Sup35 to detect Swi1$_{1\text{-}38}$Mut-MC or Sup35$_{FL}$.

double phenylalanine mutants showed only signal in the monomeric region. This loss of high MW, SDS-resistant species mirrored the loss of *de novo* prion formation by FF-NN and FF-AA – indicating the inability of these mutants to adopt stable prion aggregates.

ThT staining of yeast expressing Swi1$_{1\text{-}38}$ remains inconclusive

To further investigate the structural characteristics of Swi1$_{1\text{-}38}$ and its mutants, we attempted to employ thioflavin T (ThT) for staining purposes. As the SDD-AGE experiments can only implicate the formation of high MW, SDS-resistant species and not whether these protein species are amyloid in nature, the specificity of ThT staining for amyloids would provide crucial additional information as whether the short PrD of Swi1$_{1\text{-}38}$ can form amyloid is not yet known. Previous research in yeast has highlighted difficulties with ThT staining due to exceedingly high background fluorescence necessitating very high expression of the protein of interest (191, 192). We implemented the most widely used protocol for staining yeast with ThT (193).

We examined BY4741 *swi1Δ* / *p415TEF-SWI1$_{1\text{-}38}$-YFP*, BY4741 *swi1Δ* / *p415TEF-SWI1$_{N}$-YFP*, BY4741 *swi1Δ* / *p415TEF-SWI1$_{NQ}$-YFP*, and BY4741 *swi1Δ* / *p415TEF-YFP* cells – Swi1$_{N}$ has previously been shown to form amyloid *in vitro*. All isolates were derived from a BY4741 *swi1Δ* / *p416TEF-SWI1$_{FL}$* [*SWI*$^{+}$] strain, and thus, they contained widespread Swi1$_{NQ/N/1\text{-}38}$-YFP aggregates. Initial attempts at staining with ThT (at a concentration of 0.001%) resulted in extremely high background fluorescence (data not shown). We attempted to optimize the staining protocol through usage of both ThT and the similar thioflavin S (ThS) as well as lower concentrations of both. At a concentration of 0.0001% ThT, the staining still led to whole cell fluorescence (Figure 2.7). Furthermore, an indeterminant foci was observed across all samples – possibly corresponding to a structure such as the insoluble protein deposit (IPOD) – including the negative control expressing YFP alone (194, 195). Staining with ThS did not result in clearer

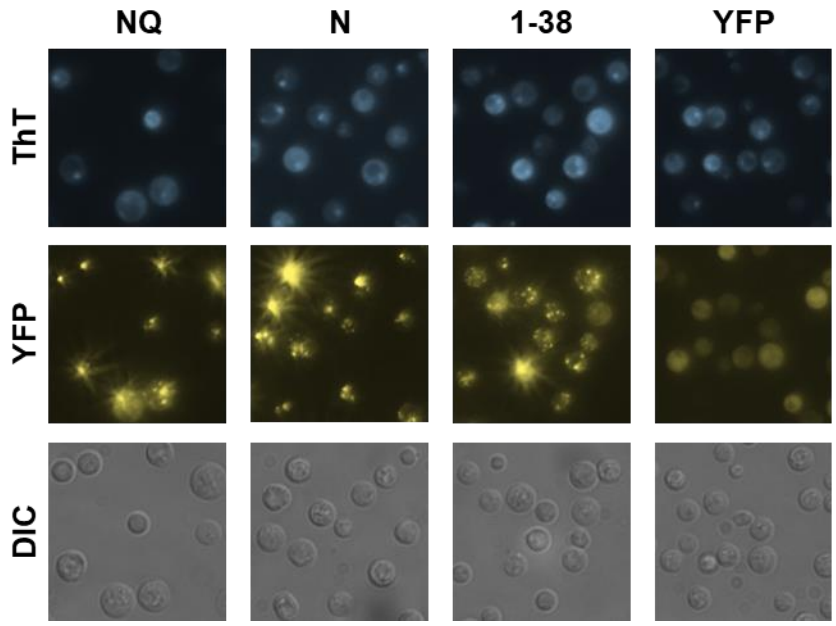


Figure 2.7. ThT staining of BY4741 *swi1Δ* / *p415TEF-SWI1*$_{NQ/N/1-38}$-YFP yeast.

Yeast expressing Swi1$_{NQ}$-YFP, Swi1$_{N}$-YFP, Swi1$_{1-38}$-YFP, or YFP alone were stained with 0.0001% thioflavin T (ThT). Universally high background of ThT fluorescence was observed in all samples including the YFP-only control. Foci observed from the ThT fluorescence did not correspond to the aggregates observed from the YFP fluorescence. Representative images of two examined samples shown.

results as background fluorescence was significantly higher than with ThT (data not shown). Further study will require optimizations in ThT staining and deployment of a functioning positive control (as the expected positives of $Swi1_{N}$ and $Swi1_{NQ}$ did not exhibit notable colocalized ThT staining).

Double phenylalanine mutants do not impact aggregation of large Swi1 constructs

As the double phenylalanine mutants resulted in the most dramatic disruptions of prion propagation and formation by $Swi1_{1\text{-}38}$, we next tested whether these mutations would affect larger Swi1 constructs such as $Swi1_{N}$ and $Swi1_{NQ}$ (Figure 2.8A). The full range of phenylalanine mutants – F3N, F3A, F4N, F4A, FFNN, FFAA – were all reproduced in $Swi1_{N}$ and $Swi1_{NQ}$. In both BY4741 and BY4741 *swi1Δ* / *p416TEF-SWI1*$_{FL}$ backgrounds, all of the phenylalanine mutants did not result in any significant differences in aggregation of $Swi1_{N}$ or $Swi1_{NQ}$ versus WT (Figure 2.8B, data not shown). Maintenance of the [*SWI*$^{+}$] prion conformation by $Swi1_{N}$ or $Swi1_{NQ}$ in the absence of $Swi1_{FL}$ also was not affected by the phenylalanine mutants (data not shown). These results point to regions outside $Swi1_{1\text{-}38}$ supporting the ability of $Swi1_{N}$ and $Swi1_{NQ}$ to adopt and propagate the [*SWI*$^{+}$] prion conformation.

Given that the double phenylalanine mutants abolish aggregation of $Swi1_{1\text{-}38}$ but do not impact the aggregation of $Swi1_{N/NQ}$, we next tested whether these mutations could impact aggregation of intermediate Swi1 PrD constructs. A range of previously constructed $Swi1_{N}$ truncations were used – $Swi1_{1\text{-}224}$, $Swi1_{1\text{-}176}$, $Swi1_{1\text{-}133}$, $Swi1_{1\text{-}74}$, $Swi1_{1\text{-}65}$, $Swi1_{1\text{-}46}$, and $Swi1_{1\text{-}42}$ (156). For each Swi1 truncation, we created and transformed WT, FF-NN, and FF-AA constructs into BY4741 *swi1Δ* / *p416TEF-SWI1*$_{FL}$ yeast to provide the greatest opportunity for aggregation. Aggregation was observed for all constructs in a proportion of transformed yeast – mirroring the potential for even the $Swi1_{1\text{-}38}$ FF-NN and FF-AA mutants to aggregate in some yeast (Figure

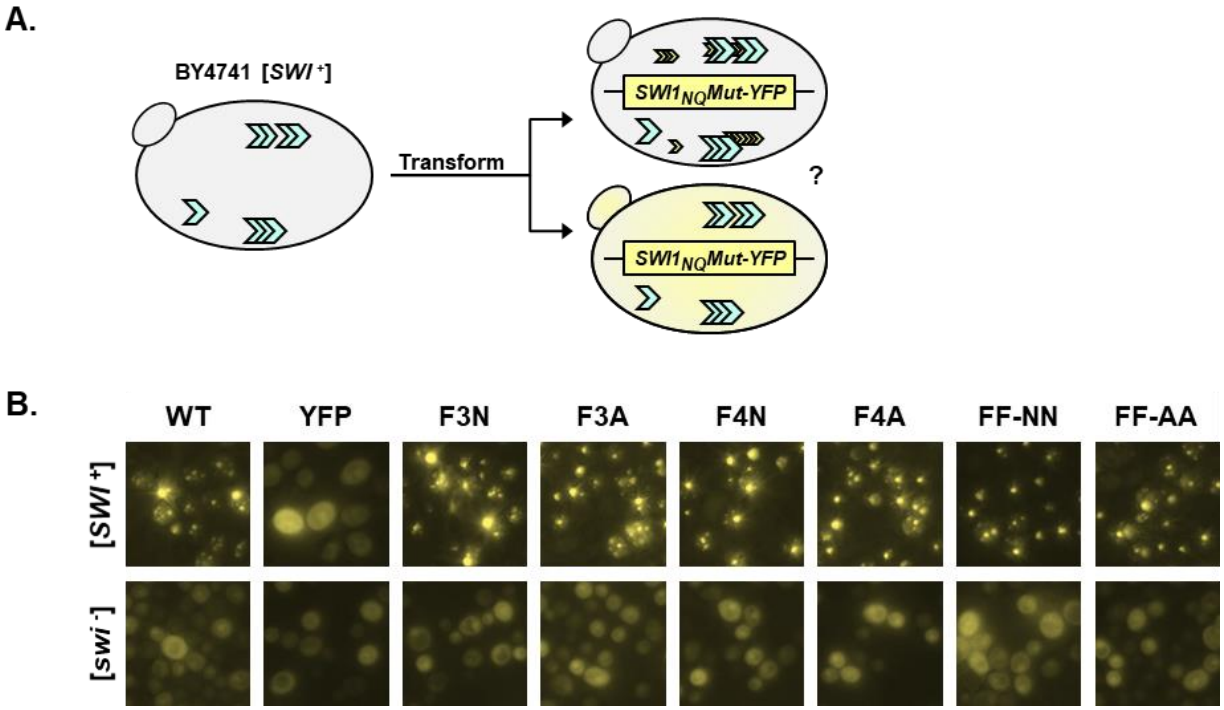


Figure 2.8. Phenylalanine-targeted mutations fail to disrupt Swi1$_{NQ}$ aggregation.

A. Diagram of experiment. BY4741 [*SWI*$^{+}$] or [*swi*$^{-}$] cells were transformed with *p415TEF-SWI1*$_{NQ}$*-YFP* (WT), *p415TEF-SWI1*$_{NQ}$*Mut-YFP*, or *p415TEF-YFP*. Transformants were observed using fluorescence microscopy for aggregate foci or diffuse signal.

B. Representative fluorescence images of BY4741 [*SWI*$^{+}$] or [*swi*$^{-}$] cells transformed with *p415TEF-SWI1*$_{NQ}$*-YFP* (WT), *p415TEF-SWI1*$_{NQ}$*Mut-YFP*, or *p415TEF-YFP*. Shown are representative views.

2.2C). When $Swi1_{FL}$ was dropped out from aggregate-containing yeast to produce BY4741 *swi1Δ* / *p415TEF-SWI1$_{Trunc-Mut}$-YFP* cells, an intriguing pattern of aggregation maintenance was observed. The largest truncations, $Swi1_{1-224}$ and $Swi1_{1-176}$ kept aggregates in the absence of $Swi1_{FL}$ in all examined isolates – WT and FF-NN and FF-AA (Table 2.5). Smaller truncations – $Swi1_{1-133}$, $Swi1_{1-74}$, $Swi1_{1-46}$, $Swi1_{1-42}$ – all lost the ability to maintain aggregation with the FF-NN or FF-AA mutations and without the presence of $Swi1_{FL}$. The WT versions of these truncations readily maintained aggregates in BY4741 *swi1Δ* cells. Curiously, $Swi1_{1-65}$ was the lone outlier as FF-NN and FF-AA mutants aggregated just as WT did in the absence of $Swi1_{FL}$.

Table 2.5. Aggregation of $Swi1_N$ truncations with FF-NN and FF-AA mutations.

Truncation	Mutation	Colonies with Aggregation
$Swi1_{1-42}$	WT	100%
	FF-NN	0%
	FF-AA	0%
$Swi1_{1-46}$	WT	100%
	FF-NN	13%
	FF-AA	13%
$Swi1_{1-65}$	WT	100%
	FF-NN	100%
	FF-AA	100%
$Swi1_{1-74}$	WT	88%
	FF-NN	25%
	FF-AA	6%
$Swi1_{1-133}$	WT	100%
	FF-NN	0%
	FF-AA	0%
$Swi1_{1-176}$	WT	100%
	FF-NN	100%
	FF-AA	100%
$Swi1_{1-224}$	WT	100%
	FF-NN	100%
	FF-AA	100%

We had hypothesized that the reason for the FF-NN and FF-AA mutants not impacting $Swi1_N$ or $Swi1_{NQ}$ was due to the availability of other large hydrophobic amino-acid residues that could replace the functioning of the phenylalanines in $Swi1_{1-38}$. This reasoning does not explain the maintenance of aggregation by $Swi1_{1-65}$ with the FF-NN and FF-AA mutants – as no additional

large, hydrophobic amino acids are present in the 39-65 region. Additionally, the inability for $Swi1_{1\text{-}133}$ to maintain aggregation while having multiple phenylalanines and valines in the 66-133 region does not concur with the hypothesis. As such, there may be greater complications in structure requirements to maintain the Swi1 prion fold present in these BY4741 *swi1Δ* / *p416TEF-SWI1$_{FL}$* folds. These experiments ask for further confirmation and testing of this discrepancy. Moreover, these results only regard maintenance of the [*SWI*$^+$] prion fold and whether the FF-NN and FF-AA mutants impact de novo prion formation for larger Swi1 constructs has yet to be examined.

A predicted amyloid core region, $Swi1_{239\text{-}259}$ is unnecessary for aggregation of Swi1

Another region of Swi1 implicated in aggregation and therefore prion activities is $Swi1_{239\text{-}259}$. Previously predicted to act as an amyloid core region by Sant'Anna et al., $Swi1_{239\text{-}259}$ has been shown to form amyloid *in vitro* (196). We synthesized the $Swi1_{239\text{-}259}$ peptide with the assistance of the Northwestern Peptide Synthesis Core. With this newly synthesized $Swi1_{239\text{-}259}$ peptide, we confirmed its ability to form amyloid *in vitro* through a ThT-binding assay (Figure 2.9A). Using the $Swi1_{239\text{-}259}$ amyloid fibrils, naïve yeast was transformed with a marker plasmid and either sonicated $Swi1_{239\text{-}259}$ fibrils or $Swi1_N$ fibrils. Although this fibril transformation experiment was repeated multiple times, no [*SWI*$^+$] (as assayed via either a Swi1-MC reporter system or *FLO1-URA3* reporter system) colonies were generated. However, the transformation of the positive control of $Swi1_N$ also did not result in [*SWI*$^+$] isolates. Thus, we cannot conclude whether $Swi1_{239\text{-}259}$ fibrils may confer [*SWI*$^+$] when transformed into yeast.

As a predicted amyloid core for Swi1, we next examined whether deletion of $Swi1_{239\text{-}259}$ would impede aggregation of $Swi1_{NQ}$. In both BY4741 and BY4741 *swi1Δ* / *p416TEF-SWI1$_{FL}$* yeast, aggregation by $Swi1_{NQ\Delta 239\text{-}259}$ was readily observed in [*SWI*$^+$] cells and comparable to WT

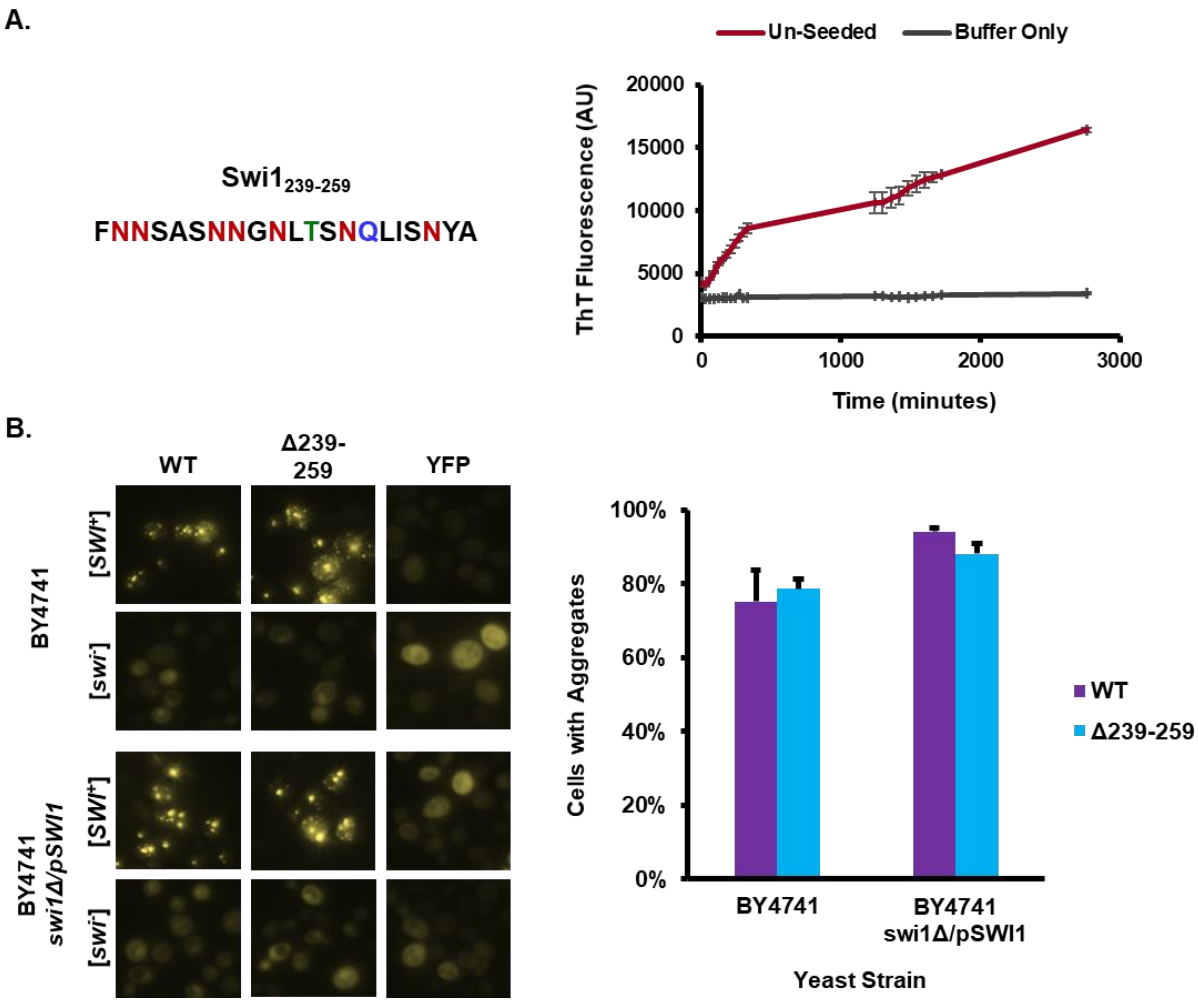


Figure 2.9. Swi1$_{239-259}$ forms amyloid fibrils but is unnecessary for aggregation of Swi1$_{NQ}$.

A. Synthesized Swi1$_{239-259}$ peptide demonstrates increasing thioflavin T (ThT) fluorescence over time – indicating formation of amyloid fibrils and confirmation of the results of Sant'Anna et al., 2016. Unseeded Swi1$_{239-259}$ monomer and buffer only were shaken and followed over 48 hours (2880 minutes). Error bars represent standard error of 3 biological replicates.

B. Deletion of Swi1$_{239-259}$ from Swi1$_{NQ}$ does not prevent aggregation of Swi1$_{NQ}$ in BY4741 [*SWI*$^+$] or BY4741 *swi1Δ / p416TEF-SWI1* [*SWI*$^+$] cells. Shown are representative images. Error bars represent standard error of 3 biological replicates.

(Figure 2.9B, data not shown). No aggregation was observed for any construct in [*swi*$^-$] yeast. This aggregation was also stable upon removal of Swi1$_{FL}$ from the BY4741 *swi1Δ* / *p416TEF-SWI1$_{FL}$* / *p415TEF-SWI1$_{NQΔ239-259}$* yeast. We next asked whether the combination of the deletion of 239-259 and either the FF-NN or FF-AA mutations (as Swi1$_{NQΔ239-259\ FF-NN/FF-AA}$) would impact the aggregation of Swi1$_{NQ}$. Either of these modifications alone had not affected Swi1$_{NQ}$ aggregation in either BY4741 or BY4741 *swi1Δ* / *p416TEF-SWI1$_{FL}$* cells (Figure 2.8B, 2.9B). We found that Swi1$_{NQΔ239-259\ FF-NN/FF-AA}$ aggregated just as the WT Swi1$_{NQ}$ in [*SWI*$^+$] cells – indicating that regions beyond Swi1$_{1-38}$ and Swi1$_{239-259}$ support Swi1 prion aggregation (data not shown). In summary, these results indicate that this region is not specifically required for Swi1 aggregation.

Discussion

Our lab initially discovered the [*SWI*$^+$] prion and documented the existence of the Swi1 prion domain within the protein's N-region (100, 155). Further study revealed that Swi1$_{1-38}$ could recapitulate the aggregation phenotype of [*SWI*$^+$] and function as a bona fide prion domain (156, 157). In this study, we further dissected Swi1$_{1-38}$ and its ability to aggregate, maintain the [*SWI*$^+$] prion fold, and *de novo* form a prion.

Multiple mutants of Swi1$_{1-38}$ that we created via replacement of singular non-asparagine residues with either asparagine or alanine had no significant effect. Previous work by our lab showed that Swi1$_{1-31}$, a truncation of Swi1$_{1-38}$ which did not include T32, could still transmit the [*SWI*$^+$] prion fold and form a prion when fused with Sup35$_{MC}$. The lack of necessity for this end portion of Swi1$_{1-38}$ suggests that mutations at this location would likely be more easily tolerated than at other locations. Indeed, we observed that the T32N and T32A mutations maintained similarity to WT throughout our various assays. Interestingly, the T27N and T27A also did not

demonstrate significant deviations from WT in aggregation, maintenance, or prionization. Meanwhile, the L6A mutation did not generate meaningful differences in the functioning of $Swi1_{1-38}$ as a prion domain. However, L6N did display decreases in maintaining the [*SWI*$^{+}$] prion fold in the absence of $Swi1_{FL}$. This difference between the mutation to alanine versus mutation to asparagine may indicate that the decrease in hydrophobicity interrupts a buried region of aggregated $Swi1_{1-38}$.

The threonine tract mutants (4xTN, 4xTA, 8xTN, 8xTA) demonstrated a dichotomy based on maintaining the polarity of the residues versus losing said polarity. The 4xTN and 8xTN mutants where the tract was partially or wholly replaced with the similarly polar, uncharged asparagine saw little variance from the WT $Swi1_{1-38}$. Conversely, the 4xTA and 8xTA mutants exhibited severely reduced aggregation in both genetic backgrounds as well as completely abolished *de novo* prion formation. Threonine tracts of similar length to the one found in $Swi1_{1-38}$ can be found in some adhesins or flocculins in various yeasts (197–199). In those contexts, such polythreonine stretches are thought to be important for the formation of β-sheet structures and for the surface-binding properties of the proteins. The threonine tract of $Swi1_{1-38}$ may also play a similar role for its aggregation – though the exact structure of this prion domain has yet to be determined. However, threonine and other uncharged polar residues such as asparagine and serine have a noted role in promotion of aggregation and amyloidogenesis. In following, the threonine tract of $Swi1_{1-38}$ may provide a stable core for formation of the high-MW, SDS-resistant species observed in this study.

Mutating the two phenylalanine residues at the beginning of $Swi1_{1-38}$ resulted in a fairly direct relationship between the number of phenylalanines and maintenance of the [*SWI*$^{+}$] prion fold as well as prionization. Replacing one phenylalanine residue (F3N, F3A, F4N, F4A) led to

~50% of colonies maintaining aggregates in the BY4741 *swi1Δ / p415TEF-SWI1$_{1-38}$Mut-YFP* cells (Figure 2.3C) and replacing the second phenylalanine residue (FF-NN, FF-AA) led to almost all complete loss of the prion as observed via aggregation. If the capability of Swi1$_{1-38}$ to maintain a prion fold is dependent on the aromaticity present in these phenylalanine residues, then perhaps the replacement of the phenylalanine residues with other aromatic amino acids (i.e., tryptophan or tyrosine) may have no effect versus WT. *De novo* prion formation by Swi1$_{1-38}$-MC also relied on the presence of at least one phenylalanine residue with neither FF-NN or FF-AA capable of prionization. This reliance of the prionogenicity of Swi1$_{1-38}$ on a single amino-acid residue being present belies the fact that the jump from aggregatable to prionogenic can be extremely small. Indeed, previous research found that just a small number of mutations could lead an existing asparagine/glutamine-rich domain to gain prion capabilities (200). The mutations presented in that research primarily relied on replacing non-prionogenic residues (e.g., charged amino acids) and introduction of hydrophobic and/or aromatic residues (e.g., phenylalanine) – much the opposite of some of the deleterious mutants produced in Swi1$_{1-38}$.

Given the impact of the removal of the aromatic side groups on Swi1$_{1-38}$, we also examined whether either the FF-NN or FF-AA mutation affected aggregation in context of longer regions, such as Swi1$_{N}$ or Swi1$_{NQ}$. However, no change in aggregation was observed in BY4741 [*SWI*$^{+}$] cells (data not shown). This result indicates that other residues or regions of Swi1$_{N}$ can stand-in for the loss of the two phenylalanine residues at positions 3 and 4. Indeed, multiple aromatic amino acids can be found downstream of Swi1$_{1-38}$ (*i.e.*, positions 73, 76, 77, 82). These other aromatic-containing residues may indeed provide the necessary underpinning of the region's prion forming capacity when the first two phenylalanine residues are replaced. Additionally, Sant'Anna et al. showed and we confirmed that a predicted amyloidogenic region (Swi1$_{239-259}$) can in fact form

amyloid *in vitro* (Figure 2.9A) (196). Regions such as $Swi1_{239-259}$ likely provide any required stabilization needed to offset the destabilization of $Swi1_{1-38}$ – allowing maintenance and propagation of the prion fold. Taken together, the presence of multiple aromatic residues and the amyloidogenic region located downstream of the $Swi1_{1-38}$ PrD suggests that [*SWI*$^+$] formation is likely a favorable event in *S. cerevisiae*. In this regard, it has been shown that the [*SWI*$^+$] can confer fungicide resistance, tolerance of certain alcohols, and aid yeast to adapt environmental changes (109, 151, 153).

Intriguingly, many of the phenylalanine residues in $Swi1_N$ region – those that were mutated at positions 3 and 4 as well as those closely downstream at positions 73, 76, 77, and 82 – are conserved in other *Saccharomyces* species (data not shown). For example, the Swi1 gene in *S. boulardii*, *S. paradoxus*, and *S. pastorianus* all contain the aforementioned phenylalanine residues (201–203). The asparagine content of the corresponding $Swi1_N$ regions across these species is highly similar (~31-34%) although *S. cerevisiae* $Swi1_N$ contains a greater number of asparagine by raw count. Moreover, in the case of *S. boulardii*, the threonine tract can also be found within the corresponding $Swi1_{1-38}$ region. It should be noted that additional charged amino-acid residues present in *S. pastorianus* may prevent the extreme N-terminal of Swi1 from acting similarly to the $Swi1_{1-38}$ examined in this study. In all, we do not currently know if Swi1 exhibits prionogenicity in these other species; however, the gene appears to retain the components that likely provide the basis for prionogenicity in *S. cerevisiae*. Further research may elucidate the possibility of [*SWI*$^+$] existence in other species.

Although the structure of aggregated or prionized $Swi1_{1-38}$ (or its various fusions) is yet unknown, our lab has previously demonstrated that $Swi1_N$ can form amyloid. In this study, we have demonstrated that $Swi1_{1-38}$ forms high molecular weight, SDS-resistant aggregates in the case

of either Swi1$_{1\text{-}38}$-YFP initially aggregated alongside Swi1$_{FL}$ in the [*SWI*$^{+}$] prion form or the [*SPS*$^{+}$] prion *de novo* formed by Swi1$_{1\text{-}38}$-MC. It is likely that these protein species visualized by SDD-AGE are of an amyloid variety as such patterning mirrors that of larger Swi1 constructs, Sup35 (Figure 2.6B), and other amyloid-forming proteins.

In all, select amino acids in Swi1$_{1\text{-}38}$ are crucial to this prion domain's ability to aggregate, maintain the [*SWI*$^{+}$] prion fold, and *de novo* form a prion. While the overall asparagine-rich composition of Swi1$_{1\text{-}38}$ provides a basis for prionization, this region depends on the presence of its two phenylalanine residues for the ability to prionize – though these two specific residues are not vital for [*SWI*$^{+}$] in the context of Swi1$_{N}$ or Swi1$_{NQ}$. The other non-asparagine residues, which are mainly threonine, likely maintain the favorable uncharged, polar sidechains that favor disorder but also aggregation. As such, it remains likely that like other prionogenic proteins, Swi1$_{1\text{-}38}$ and its larger iterations, Swi1$_{N}$, Swi1$_{NQ}$, and full-length Swi1 achieve their prionogenicity largely via overall composition.

CHAPTER THREE

Examining the evolution of yeast prion domains

Introduction

Prion proteins are proteins capable of adopting transmissible alternative conformations. Initially identified in relation to neurodegenerative diseases such as transmissible spongiform encephalopathy (TSE), the prion concept has grown to encompass proteins involved in the pathology of other diseases such as Alzheimer's (AD), Parkinson's (PD), and amyotrophic lateral sclerosis (ALS). Moreover, the discovery of heritable, protein conformation-based epigenetic elements in the budding yeast, *Saccharomyces cerevisiae* has further expanded our understanding of prions. There has been eleven such yeast prions discovered – [*ESI*$^{+}$], [*GAR*$^{+}$], [*MOD*$^{+}$], [*MOT3*$^{+}$], [*NUP100*$^{+}$], [*OCT*$^{+}$], [*PSI*$^{+}$], [*RNQ*$^{+}$], [*SMAUG*$^{+}$], [*SWI*$^{+}$], and [URE3] (100, 101, 110, 102–109). The protein determinants of these prions are Snt1, Pma1/Std1, Mod5, Mot3, Nup100, Cyc8, Sup35, Rnq1, Vts1, Swi1, and Ure2, respectively.

Each prion protein typically has a region responsible for its prion-forming capabilities. This region is termed its prion domain (PrD). Many yeast prion proteins have PrDs enriched in polar, uncharged glutamine (Q) and asparagine (N) residues (109). These amino acids are overrepresented in yeast PrDs likely due to their lack of unfavorability to prionization and high penchant for disorderedness (167). Meanwhile, charged amino acids are rare and hydrophobic residues are underrepresented – as they are needed for aggregation and formation of amyloids but too many would likely prevent the establishment of a non-prion state. Yeast prion proteins that are not enriched in Q/N residues tend to still be largely disordered – a quality observed in many human prion-like proteins as well. The composition-based prionogenicity of yeast PrDs also makes it

possible to randomly scramble a PrD – thus altering sequence but not composition – and have the protein retain its prion-forming capabilities (204). In all, an overall disordered-prone composition with interspersed large, hydrophobic residues is more important than the sequence of the residues in a PrD (205).

Transcriptional regulators are overrepresented among currently yeast prion proteins. Furthermore, prionization of Mot3, Cyc8, and Swi1 can all effect the transcription, and therefore expression, of a group of genes involved in yeast multicellularity, the flocculin (*FLO*) genes (133, 134). This setup provides for existence of a possible prion-based gene regulatory network and supports the case for yeast prions acting not as diseases but instead functional epigenetic elements. Additional evidence has been found for this perspective in the form of other beneficial phenotypes – including resistance to antifungal drugs, mobility enhancements, and fitness increases over non-prion cells (102, 133, 145, 147, 151, 153, 206). Given such possible phenotypic benefits, environmental pressures could impact the evolution of yeast PrDs. Mutations of limited scope may affect the formation rate of prions by protein determinants. This effect has been experimentally demonstrated as prionogenicity was increased through the addition of large prionogenic, hydrophobic residues (e.g., isoleucine (I), valine (V)) or replacement of charged amino-acids (e.g., lysine (K), arginine (R)) with the disorder-prone Q/N (200, 207, 208). These data create questions regarding the evolution of yeast prion proteins and the PrDs residing in them.

Previous observations have found that a limited number of eukaryotic species harbor high proportions of Q/N-rich proteins (209). As already discussed, yeast contain an enrichment of these proteins as well as many demonstrated endogenous prions. However, the sequences of these prion proteins are not highly conserved into higher organisms including humans. Moreover, prokaryotes also do not exhibit a high frequency of Q/N-rich regions among their proteins (210). These data

implicate that yeast prion proteins are uniquely evolved epigenetic elements functioning in the quick environmental adaptation of yeast.

Recently, an extensive effort to evaluate the evolution of *S. cerevisiae* generated a trove of sequencing data on yeast isolates from across the globe (211). This collection of whole genome sequencing of 1,011 *S. cerevisiae* isolates covers a range of ecological origins including soil samples, wine samples, and human clinical samples. The resource generated through these sequences allows for us, in partnership with Dr. Shengjun Tan, to investigate the diversity and evolution of yeast prion genes as well as genes (e.g., *FLO*) regulated by yeast prion proteins. We also analyzed whether the variation in ecological origin – and thus differences in evolutionary pressures – impacted prionogenicity of known yeast prion proteins.

Materials and Methods

Yeast sequence data

Saccharomyces cerevisiae genome sequences were obtained from data produced by Peter et al., 2018 (211). FASTA sequences of isolates were reconstructed from the mapped sequencing reads available on the 1002 Yeast Genome website (http://1002genomes.u-strasbg.fr/files/). Single nucleotide polymorphism (SNP) and indel (insertion or deletion of bases) information was downloaded (1011Matrix.gvcf). The S288C reference genome (from Ensembl, ensemble.org; version R64-1-1.92) was used for mapping the FASTA sequences.

Gene orthologs for *Saccharomyces paradoxus* were determined based on the Broad Institute's fungal orthogroups. Gene orthologs for other *Saccharomyces* species were obtained by examining genomes from NCBI (National Center for Biotechnology Information) – *S. arboricola* (ALIE00000000), *S. boulardii* (LIOO00000000), *S. jurei* (GCA_900290405.1), *S. kudriavzevii*

(AJHS00000000), *S. mikatae* (AABZ00000000), *S. pastorianus* (AZCJ00000000), and *S. uvarum* (AACG00000000). The software tool Exonerate was used to compute orthologs from the given genomes. Exonerate uses algorithms detailed by Slater and Birney, 2005 (212). If multiple orthologs were detected for a given gene, the ortholog with the highest identity and length was retained. Calculated orthologs for *S. arboricola*, *S. mikatae*, and *S. uvarum* were confirmed by comparison to public data. All orthologs corresponded to the same genes – confirming the reliability of the Exonerate-based method.

Coding sequence (CDS) reconstruction

Data was handled in Perl. The SNP and indel information from 1011Matrix.gvcf included position and nucleotide substitution type. From this data, any heterozygous sites (i.e., sites where there were substantial sequencing reads corresponding to two different bases) were handled by selecting the substitution with the highest proportion of sequencing reads (>5 reads more than the substitution with the next highest proportion) as the actual nucleotide at the site. If the exact nucleotide could not be determined, the site was masked as 'N'. All indels were also masked with 'N'. Missing data in some sequences were likewise masked with 'N'. All sites masked with 'N' – regardless of reason – were omitted from involvement with calculation of d_N and d_S values.

Clade information

Clade information for all isolates was obtained from supplementary table 1 of Peter et al., 2018 (211). For ecological groupings used in peptide sequence analysis, ecological origins were obtained from supplementary table 1 as well. Broader ecological groupings (e.g., domestic, human, nature, unknown) were ascribed to isolates based on the original, more specific ecological groupings. Domestic included bakery, beer, bioethanol, cider, dairy, distillery, industrial, palm

wine, probiotic, sake, wine, and general fermentation isolates. Human included clinical and general human isolates. Nature included flower, fruit, insect, soil, water, and general nature isolates.

d_N / d_S

For clade-based calculations, the rate of nonsynonymous (d_N) and synonymous (d_S) base substitutions for each gene CDS was calculated using the yn00 module of PAML (phylogenetic analysis by maximum likelihood) (213, 214). The yn00 package implements the d_N, d_S, and d_N / d_S calculation method detailed by Yang and Nielsen, 2000 (215). Pairwise d_N and d_S were calculated for each gene in each clade. The median value of these pairwise calculations were selected. The d_N / d_S ratio represents overall constraint on the genes within each clade.

In some cases, d_N / d_S could not be calculated. If a premature stop codon exists in the gene of interest, d_N / d_S could not be calculated due to limitations of the yn00 package. If a clade only had one isolate, d_N / d_S could not be calculated due to requiring at least two isolates. For these situations, NA (not applicable) was recorded instead of a calculated value.

For species-based calculations, the rate of nonsynonymous (d_N) and synonymous (d_S) base substitutions for each gene CDS was calculated using the codeml module of PAML (214). The codeml package implements a multitude of codon substitution models for analysis of nonsynonymous and synonymous mutations. *S. arboricola*, *S. bayanus*, *S. boulardii*, *S. carlsbergensis*, *S. eubayanus*, *S. jurei*, *S. kudriavzevii*, *S. mikatae*, or *S. paradoxus* was used as the outgroup in one branch. The other branch contained all *S. cerevisiae* isolates. The ratio represents the overall constraint on the genes as compared between the outgroup and all *S. cerevisiae* isolates.

In some cases, d_N / d_S could not be calculated. If the outgroup species did not have an ortholog for the gene of interest, d_N / d_S could not be calculated. If an ortholog was <50% of the

length of the gene in *S. cerevisiae*, the d_N / d_s is not reliable. For these situations, no value was calculated or an unreliable value was omitted for further analysis.

Protein sequence construction

The FASTA sequences for the CDS of genes of interest were used to translate to protein amino acid sequences. In some cases, codons could not be translated to amino acid due to the sequence quality issues detailed in the CDS reconstruction methods. For these codons, the amino acid was masked with '?'. All sites masked with '?' were omitted from involvement in calculations of prionogenicity.

Prionogenicity analysis

Constructed protein sequences for genes of interest were analyzed for prionogenicity using PAPA (prion aggregation prediction algorithm) and PLAAC (Prion-Like Amino Acid Composition) (109, 159, 160, 167). Default parameters were used for analysis by PAPA. The default core length (60) and background amino acid frequencies from *S. cerevisiae* were used for analysis by PLAAC.

Statistical analysis

Data produced by d_N / d_S calculations, PAPA analysis, or PLAAC analysis were handled and further statistically analyzed using Prism software (GraphPad Software, San Diego, CA, USA).

Results

FLO genes display high levels of selection pressure

In partnership with Dr. Shengjun Tan, we first performed a *de novo* reconstruction of all sequenced genes by reprocessing the reported single nucleotide polymorphisms (SNPs) and indels

reported by Peter et al., 2008 (211). The rate of nonsynonymous nucleotide substitutions (d_N), rate of synonymous nucleotide substitutions (d_S), and the ratio of these rates (d_N / d_S) values for the genes of interest were calculated per clade using the previously defined clades. These values allow for evaluation of the selection pressures that the genes are under (Table 3.1). However, given the small size of some clades and the fact that one clade contained ~1/3 of all *S. cerevisiae* isolates, we quickly moved to evaluating d_N / d_S of selected prion-related genes (Table 3.2) as compared to the closest extant relative of *S. cerevisiae, S. paradoxus*.

Table 3.1. Interpretation of d_N / d_S values.

d_N / d_S	Interpretation
<1	Purifying selection
=1	Neutral evolution
>1	Positive selection

Table 3.2. Yeast genes of interest in chapter three.

Gene	Function	Gene	Function
ASM4	Nucleoporin	NRP1	RNA binding protein
CUR1	Sorting factor	NUP100	Nucleoporin
CYC8	Transcriptional regulator	PGD1	RNA polymerase II mediator
FLO1	Flocculin	PIN3	Actin factor regulator
FLO5	Flocculin	PIN4	Cell cycle regulator
FLO8	Transcription factor	PMA1	Proton pump
FLO9	Flocculin	POL32	DNA polymerase complex
FLO10	Flocculin	PSP1	Unknown
FLO11	Flocculin	PUB1	RNA binding protein
GLN3	Transcriptional regulator	PUF1	RNA binding protein
HSP104	Disaggregase	RLM1	Transcription factor
KAP120	Nuclear importer	RNQ1	Unknown
KSP1	Protein kinase	SAP30	Histone deacetylase complex
MCA1	Cysteine protease	SFP1	Transcription regulator
MOD5	tRNA transferase	STD1	Glucose signaler
MOT3	Transcriptional regulator	SUP35	Translation terminator
MRN1	Transcriptional regulator	SWI1	Chromatin remodeler
MSN1	Transcriptional regulator	TUP1	Transcription regulator
MSS11	Transcription factor	URE2	Transcription regulator
NEW1	Translation terminator	YBL081W	Unknown

Upon doing so, we found that of the genes we chose to examine, the group of *FLO* genes exhibited the greatest positive selection pressures and overall substitution rates (Figure 3.1, data

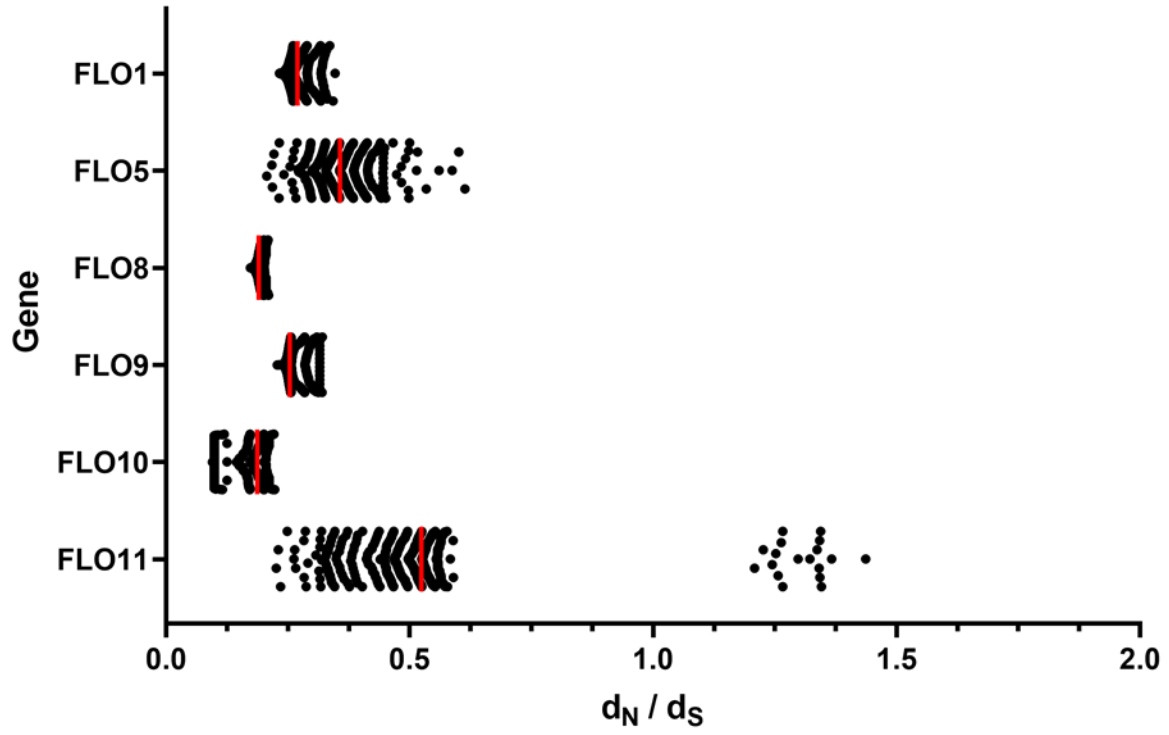


Figure 3.1 d_N / d_S distributions for *FLO* genes in 1,011 *S. cerevisiae* isolates.

Scatter plot of d_N / d_S values for FLO genes in 1,011 *S. cerevisiae* isolates. d_N / d_S values were calculated using orthologs from *S. paradoxus* as the outgroup. Red line represents median.

not shown). As already mentioned, these genes are of particular interest due to being regulatory targets for multiple yeast prion proteins and therefore, regulated by a network of yeast prions (133, 134). Moreover, this group of genes encode a range of flocculin proteins (Flo1, Flo5, Flo9, Flo10, and Flo11) and a *FLO* gene-targeting transcription factor (Flo8) that play critical roles in yeast multicellularity – making them important proteins in responding to environmental cues. In particular, an outgroup of isolates displayed d_N / d_S values that were >1.0 for *FLO11*, indicating that this gene is likely undergoing positive selection in those isolates (Figure 3.1). Additionally, the *FLO* genes were amongst those examined – alongside the prion gene, RNQ1 – with the most frequent premature stop codons (Table 3.3), however, whether these changes affect functionality of these genes is not known.

Table 3.3. Genes containing premature stop codons.

Gene	Number of Isolates with Stops	Locations in Protein (aa)	Length of Protein
FLO1	16	34, 103, 228, 1347	1537
FLO8	22	142	799
FLO9	43	29, 44	1322
FLO10	15	45, 715, 976	1169
FLO11	21	30, 94, 111, 113	1367
RNQ1	25	65, 193, 235, 249	405

Moving forward, we checked whether the isolates displaying large *FLO11* d_N / d_S values shared a common ecological origin that could explain the positive selection pressure on the *FLO11* gene. However, there was not a shared ecological origin amongst these isolates. Some isolates displayed domestic origins (e.g., beer, dairy, sake), some isolates displayed human origins, and some isolates displayed nature origins (e.g., fruit, tree). Furthermore, there was not a geographic relationship present either – the isolates had been gathered from nine different countries across four different continents. We also checked whether this group of outliers correlated with particular

d_N / d_S values for other *FLO* genes, but no such correlation was observed via Pearson r correlation matrix (for linear correlation) or direct graphing of paired d_N / d_S values for the various *FLO* genes (Figure 3.2, data not shown).

Prion domains exhibit greater genetic variability than non-prion regions in *S. cerevisiae*

As previously mentioned, prion proteins contain regions known as prion domains (PrDs) that are necessary and sufficient for formation of a prion. We hypothesized that PrDs may experience different evolutionary pressures due to their functioning as epigenetic elements and the importance of overall composition rather than primary sequence in PrD determination. Thus, we examined the seven yeast prion protein genes for which PrDs have been experimentally defined – *CYC8*, *MOT3*, *NUP100*, *RNQ1*, *SUP35*, *SWI1*, and *URE2* (Table 3.3). Each protein was divided into PrD and non-PrD regions and the rate of nonsynonymous and synonymous nucleotide substitutions (d_N and d_S) were calculated independently for each portion for each of the 1,011 *S. cerevisiae* isolates versus the corresponding regions in the orthologs in *S. paradoxus*. We used *S. paradoxus* in this calculation due to it being the closest extant relative to *S. cerevisiae*.

Table 3.4. Characterized yeast prion domains (PrDs).

Prion	Gene	Protein	PrD Coordinates (aa)	Source
[*OCT*$^+$]	*CYC8*	Cyc8	1-36, 442-678	(101)
[*MOT3*$^+$]	*MOT3*	Mot3	1-295	(109)
[*NUP100*$^+$]	*NUP100*	Nup100	300-400	(103)
[*RNQ*$^+$]	*RNQ1*	Rnq1	153-405	(107)
[*PSI*$^+$]	*SUP35*	Sup35	1-137	(216)
[*SWI*$^+$]	*SWI1*	Swi1	1-327	(155)
[URE3]	*URE2*	Ure2	1-89	(217)

Intriguingly, we found that there was greater variation in the d_N / d_S ratio for all seven PrDs versus the non-PrD regions from the corresponding proteins (Figure 3.3). Indeed, the variance of

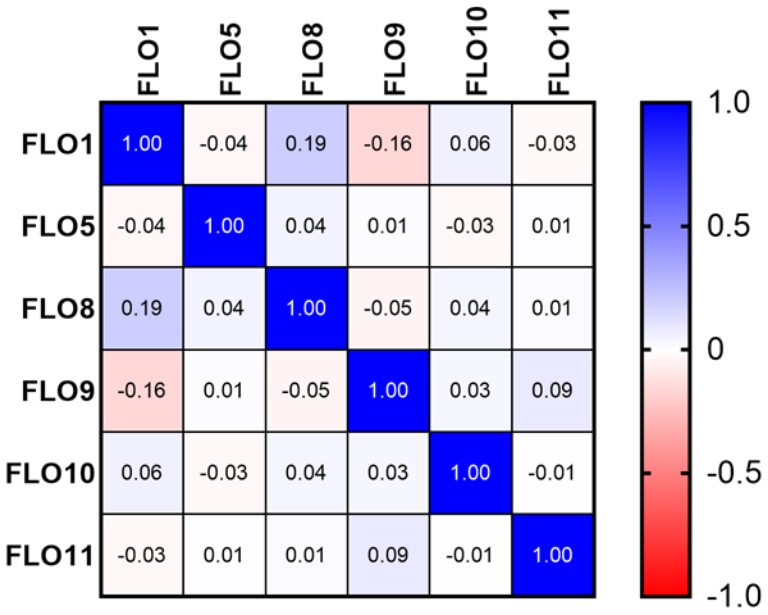


Figure 3.2 Pearson r correlation matrix for d_N / d_S values for *FLO* genes.

Heat map matrices for Pearson r correlation to test for linear correlation between d_N / d_S values for *FLO* genes among 1,011 *S. cerevisiae* isolates. Individual values for Pearson r are given in the corresponding cells. d_N / d_S values were calculated using orthologs from *S. paradoxus* as the outgroup.

the d_N / d_S values for PrDs was larger than the variance of the corresponding non-PrD regions for all seven genes examined (data not shown). Furthermore, all PrDs had significantly higher average d_N / d_S values than the non-PrD region from the same gene (paired two-tailed t-test, $P < 0.0001$). These results suggest that the PrDs are under weaker purifying selection pressure than the non-PrD regions but that this selection pressure may vary given the geographical and ecological conditions that each isolate experienced.

Given the greater average d_N / d_S values for the PrDs of these seven yeast prion genes, we backtracked and looked at the d_N values alone for the PrDs versus the non-PrD regions. Was the d_N / d_S value greater for PrDs due to differences in the overall rate of mutation or due to maintaining a larger number of nonsynonymous substitutions? For two yeast prion genes – *MOT3* and *NUP100*, d_N did not vary significantly between PrD and non-PrD region (Figure 3.4). One yeast prion gene, *CYC8* displayed the opposite relationship as expected – d_N was significantly higher for the non-PrD region than the PrD ($P<0.0001$). The remaining four yeast prion genes – *RNQ1*, *SUP35*, *SWI1*, and *URE2* – exhibited d_N values for their PrDs that were significantly greater than the d_N for their non-PrD regions (Figure 3.4). The rate of nonsynonymous substitutions was at least double for the PrD than the non-PrD component for these four genes.

Prionogenicity of yeast prion proteins across 1,011 *S. cerevisiae* isolates

Having had investigated the genomic evolution of ten yeast prion proteins, we then turned to examining the functional outcomes for these mutations. As noted previously, limited changes in amino-acid residues can result in significantly different prion formation capabilities for a protein. However, we could not feasibly investigate the de novo prion formation rates for nine different prions – [*GAR*$^+$], [*MOD*$^+$], [*MOT3*$^+$], [*NUP100*$^+$], [*OCT*$^+$], [*PSI*$^+$], [*RNQ*$^+$], [*SWI*$^+$], and [URE3] – in 1,011 *S. cerevisiae* isolates. Given that we already had the CDS for the yeast prion

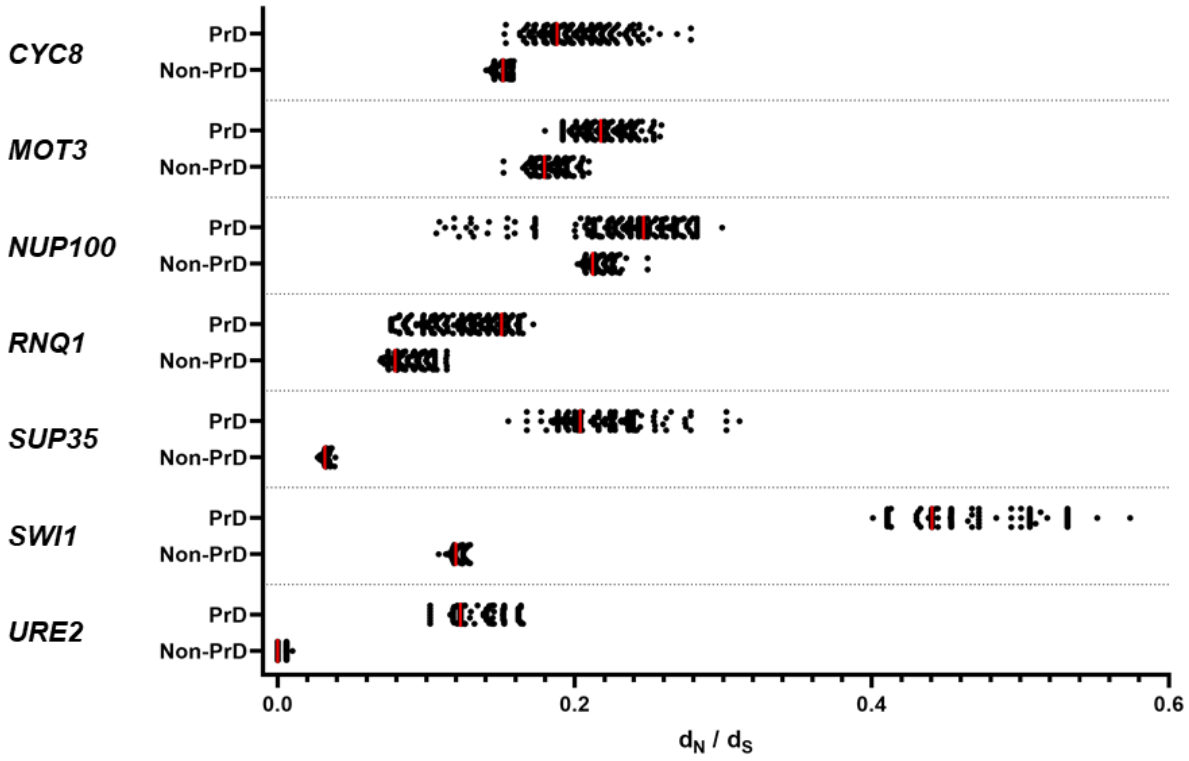


Figure 3.3 PrDs of *S. cerevisiae* prion genes display greater variation and weaker purifying selection than corresponding non-PrD regions.

Scatter plot of d_N / d_S values for prion domains (PrDs) and non-PrD regions of 7 yeast prion genes in 1,011 *S. cerevisiae* isolates. d_N / d_S values were calculated using orthologs from *S. paradoxus* as the outgroup. Red line represents median. See Table 3.3 for the amino acid boundaries of the PrDs.

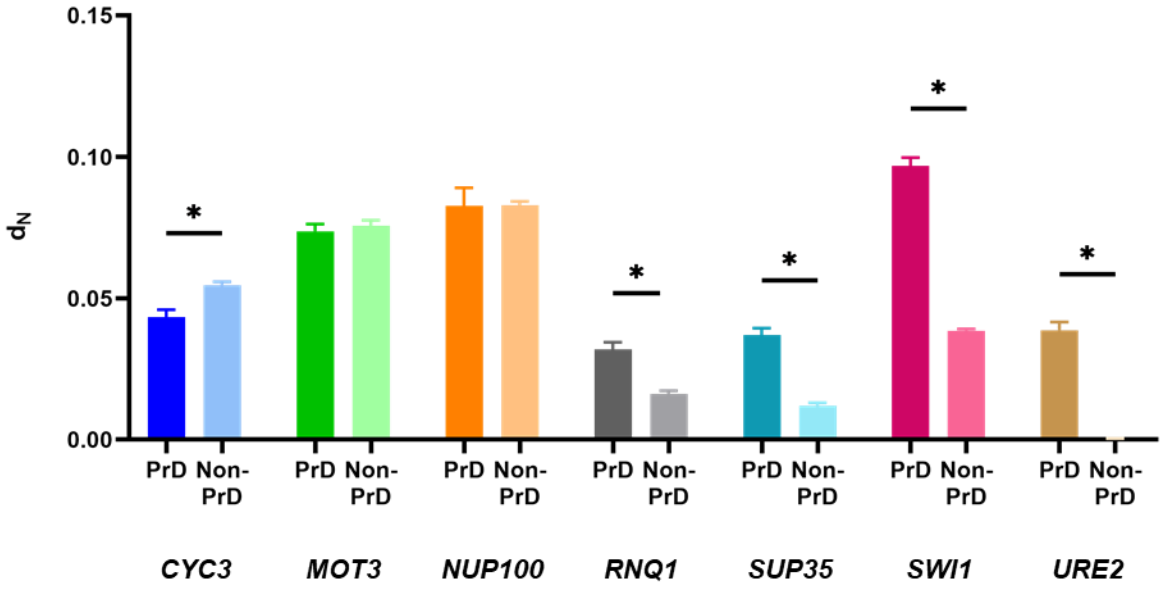


Figure 3.4 Some PrDs exhibit greater rates of nonsynonymous changes than corresponding non-PrD regions.

Bar plot of d_N values for prion domains (PrDs) and non-PrD regions of 7 yeast prion genes in 1,011 *S. cerevisiae* isolates. d_N values were calculated using orthologs from *S. paradoxus* as the outgroup. See Table 3.3 for the amino acid boundaries of the PrDs. Significance represented by * is P<0.0001.

genes, we decided to construct the corresponding protein sequences and evaluate these sequences with prion prediction algorithms.

The PLAAC and PAPA prion prediction algorithms were both developed at least partially from either yeast prion protein sequences or yeast-based prion assays (109, 159, 160, 167). Thus, these algorithms provide a useful mechanism for evaluating changes in prionogenicity resulting from sequence variation across the 1,011 isolates. However, the PLAAC scoring of protein sequences for three yeast prion proteins – Mod5, Pma1, and Std5 – were not analyzed as no predicted PrD was found using the parameters used. In evaluating the prionogenicity of the various yeast prion proteins from the isolates, we categorized the isolates into groups based on general ecological origin – domestic, human, or nature. Interestingly, we did not find any significant differences in prionogenicity as predicted by PLAAC or PAPA (Figure 3.5, 3.6). Although outliers existed across the different prion proteins and ecological groupings, the overall distribution of PLAAC and PAPA scores was fairly similar regardless of general ecological origin. We did note that there appeared to be bimodal distributions of prionogenicity for Cyc8 (based on PLAAC and PAPA), Mot3 (based on PLAAC), and Swi1 (based on PAPA) – all regulators of *FLO* and part of the hypothesized yeast prion regulatory network for those genes.

Finally, we used the prionogenicity prediction scores from PLAAC and PAPA to investigate whether there were correlations in prionogenicity of different yeast prion proteins. Our reasoning for this was multifold: 1. Multiple yeast prions can exist in the same cell; 2. Some yeast prions can heterologously induce other prions; and 3. Ecological pressures for different traits may at the same time select for or against formation of multiple yeast prions – particularly those regulating the *FLO* genes. We first examined whether the bimodal distributions of prionogenicity for Cyc8 and Mot3 (as evaluated by PLAAC) and for Cyc8 and Swi1 (as evaluated by PAPA) in

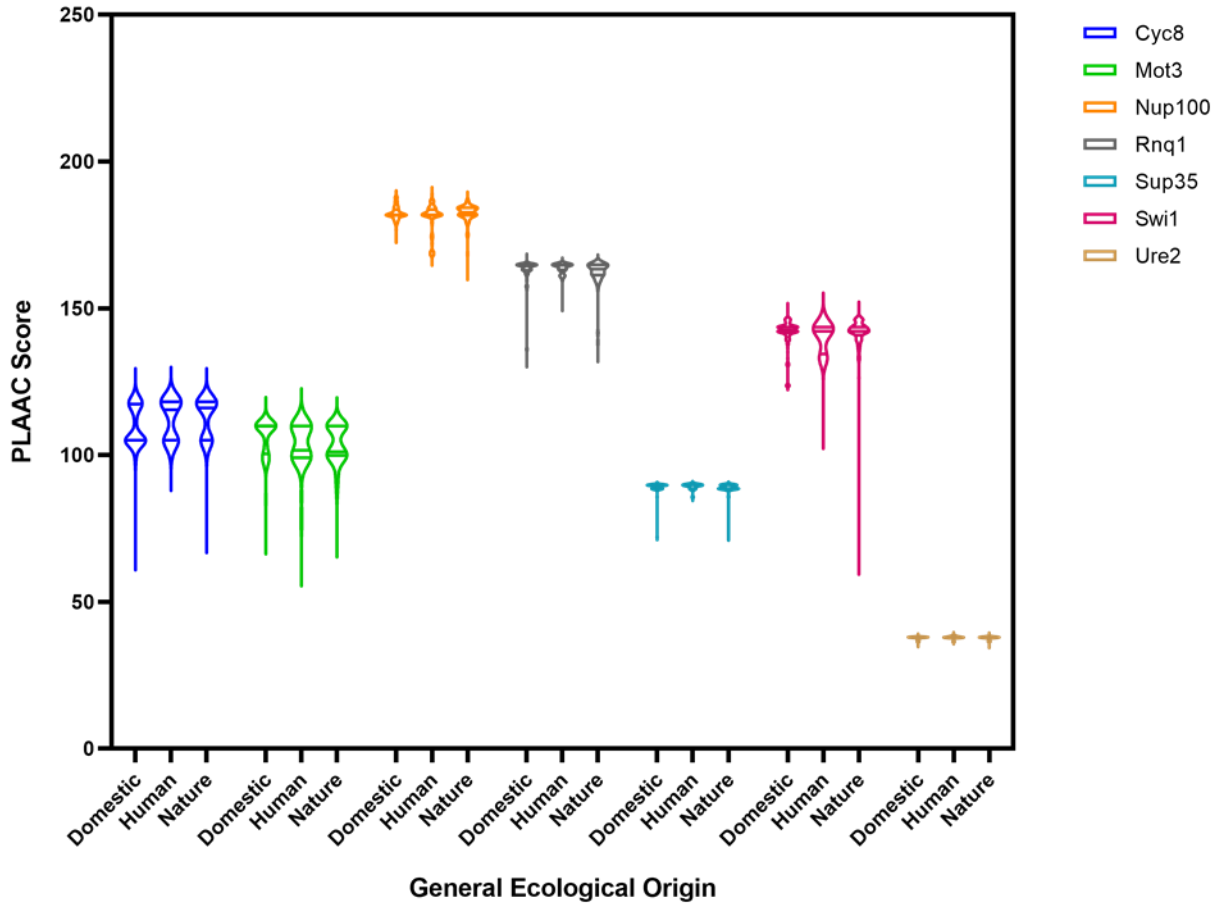


Figure 3.5 PLAAC analysis of prion proteins in *S. cerevisiae* isolates by ecological origin.

Violin plot of PLAAC propensity values for 7 yeast prion proteins in 981 *Saccharomyces cerevisiae* isolates belonging to ecological origins classified as domestic, human, or nature. General ecological origins were determined by grouping together isolates based on the specific ecological origin as defined by Peter et al., 2018. Domestic ecological origin included bakery, beer, bioethanol, cider, dairy, distillery, industrial, palm wine, probiotic, sake, wine, and general fermentation isolates. Human ecological origin included clinical and general human isolates. Nature ecological origin included flower, fruit, insect, soil, water, and general nature isolates. Default core length of 60 and background amino acid frequencies from S. cerevisiae were used for PLAAC analysis. PLAAC Score refers to the PLAAC PRDscore, which is the total log-likelihood ratios for across the predicted core region. The yeast prion proteins of Mod5, Pma1, and Std1 were not included due to no predicted PrD with the parameters used. Positive values are more prionogenic. Plots extend from minimum value to maximum value.

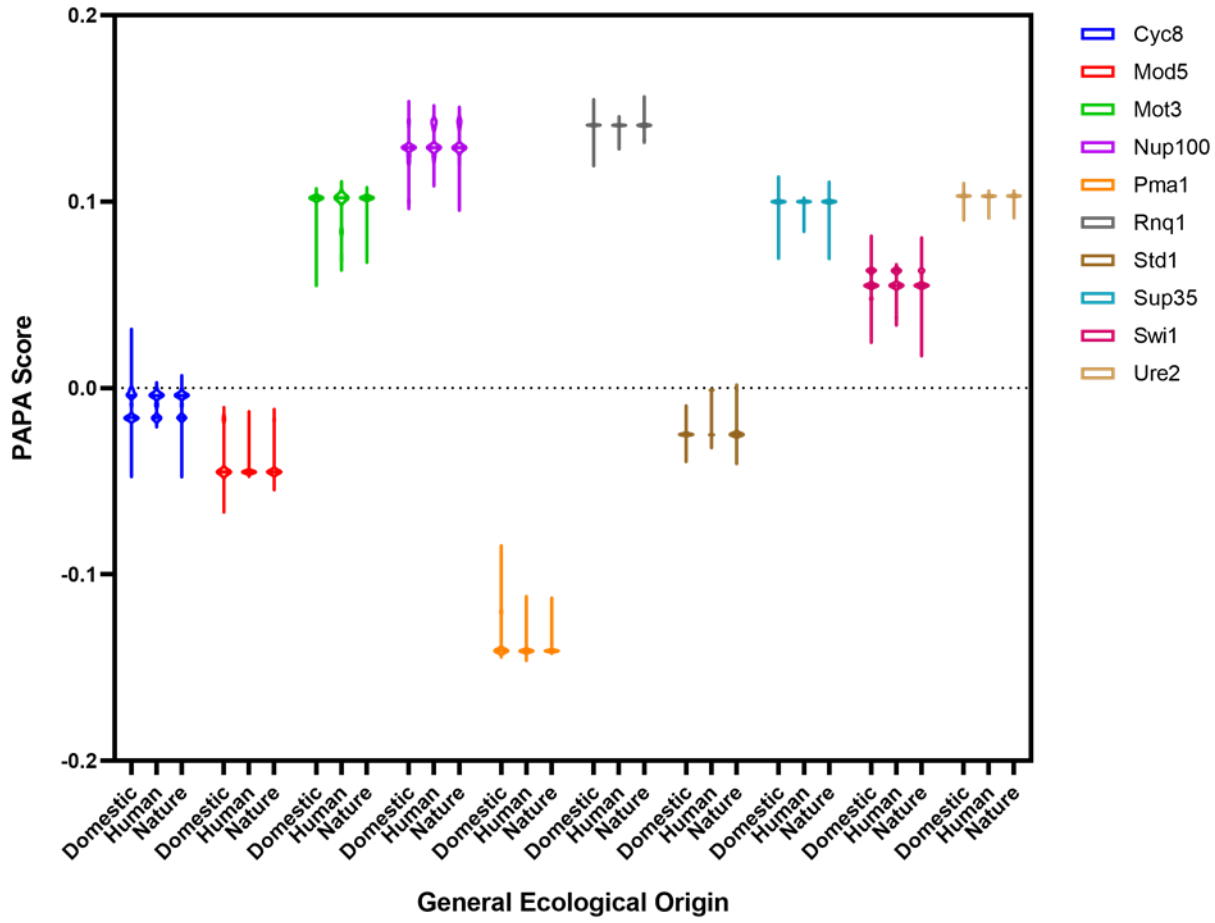


Figure 3.6 PAPA analysis of prion proteins in *S. cerevisiae* isolates by ecological origin.

Violin plot of PAPA propensity values for 10 yeast prion proteins in 981 *Saccharomyces cerevisiae* isolates belonging to ecological origins classified as domestic, human, or nature. General ecological origins were determined by grouping together isolates based on the specific ecological origin as defined by Peter et al., 2018. Domestic ecological origin included bakery, beer, bioethanol, cider, dairy, distillery, industrial, palm wine, probiotic, sake, wine, and general fermentation isolates. Human ecological origin included clinical and general human isolates. Nature ecological origin included flower, fruit, insect, soil, water, and general nature isolates. Default parameters for PAPA analysis were used. PAPA Score refers to the PAPA propensity value, which is the maximal prion propensity score in a region with a negative FoldIndex score (indicating a disordered region). Positive values are more prionogenic and negative values are less prionogenic. Plots extend from minimum value to maximum value.

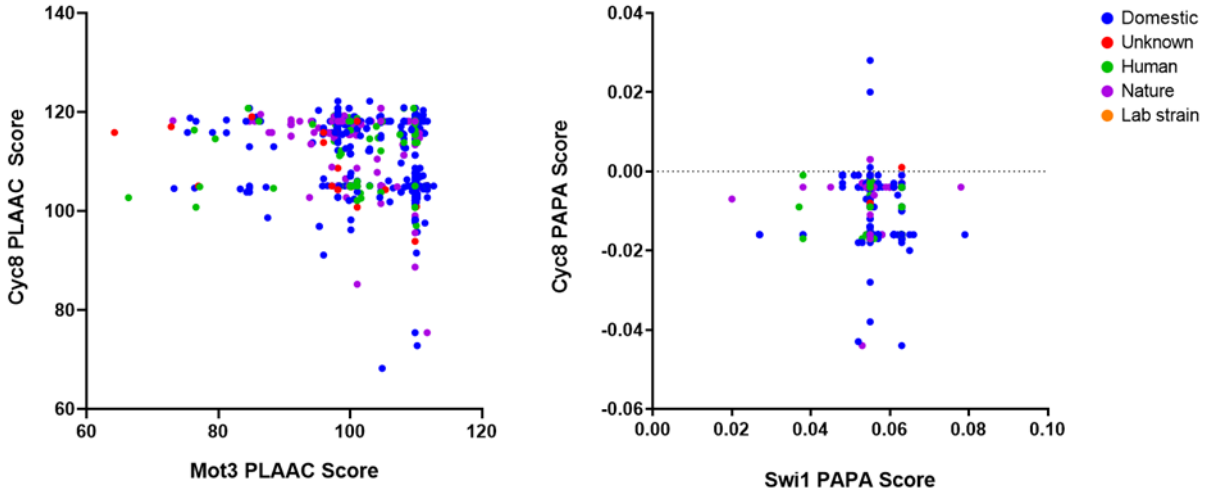


Figure 3.7 Bimodal distributions of Cyc8, Mot3, and Swi1 prionogenicity prediction scores do not strongly correlate with each other.

A. Scatter plot of Cyc8 PLAAC score versus Mot3 PLAAC score. For PLAAC, default core length of 60 and background amino acid frequencies from *S. cerevisiae* were used for analysis. PLAAC Score refers to the PLAAC PRDscore, which is the total log-likelihood ratios for across the predicted core region.
B. Scatter plot of Cyc8 PAPA score versus Swi1 PAPA score. For PAPA, default parameters for analysis were used. PAPA Score refers to the PAPA propensity value, which is the maximal prion propensity score in a region with a negative FoldIndex score (indicating a disordered region).

Plotted points are color-coded according to general ecological origin.

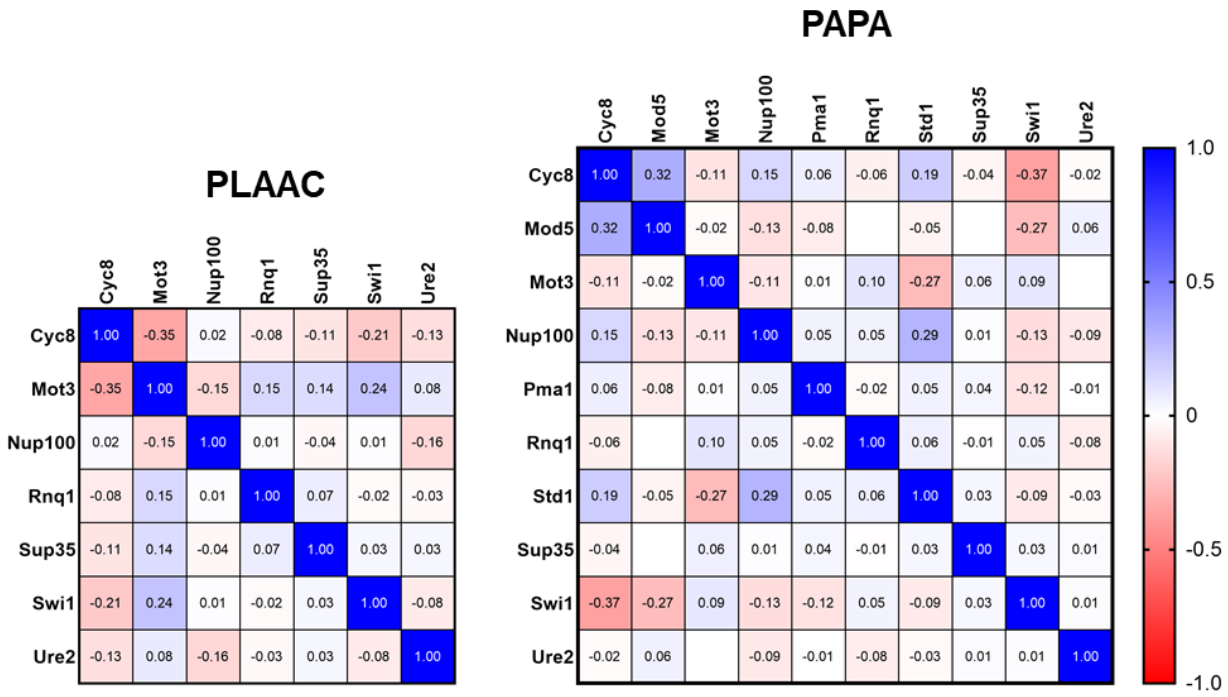


Figure 3.8 Pearson r correlation matrices for PLAAC and PAPA scores of *S. cerevisiae* prion proteins.

Heat map matrices for Pearson r correlation to test for linear correlation between scoring of prionogenicity of yeast prion proteins among 1,011 *S. cerevisiae* isolates. Individual values for Pearson r are given in the corresponding cells – a blank white cell indicates a value approximately equal to 0. For PLAAC, default core length of 60 and background amino acid frequencies from S. cerevisiae were used for analysis. PLAAC Score refers to the PLAAC PRDscore, which is the total log-likelihood ratios for across the predicted core region. The yeast prion proteins of Mod5, Pma1, and Std1 were not included in the PLAAC heatmap due to no predicted PrD with the parameters used. For PAPA, default parameters for analysis were used. PAPA Score refers to the PAPA propensity value, which is the maximal prion propensity score in a region with a negative FoldIndex score (indicating a disordered region).

fact appeared to correlate with each other. No strong correlation was observed (Figure 3.7). Given this result, we next decided to check for simple linear relationships between PLAAC scores for all yeast prion proteins and between PAPA scores for all yeast prion proteins. The generated heat map matrix for linear regression displayed some relatively weak relationships (Figure 3.8). Indeed, most yeast prion protein pairings showed virtually no linear correlation between each other. The most correlated proteins in terms of PLAAC scores were Cyc8 and Mot3. Whereas the most correlated proteins in terms of PAPA scores were Cyc8 and Mod5 and Cyc8 and Swi1.

Discussion

The new resource provided by the sequencing of over a thousand genomes from geographically and ecologically diverse isolates of the budding yeast, *Saccharomyces cerevisiae* offers opportunities to examine the diversity of yeast prion and prion-related genes. In partnership with Dr. Shengjun Tan, we took advantage of this genetic information and investigated the potential selection pressures imparted upon individual genes.

One set of our targets, the group of *FLO* genes were chosen due to its regulation by multiple yeast prion proteins and thus, yeast prions themselves. Moreover, these genes largely encode for proteins exposed to the extracellular environment and are critical to yeast multicellularity. However, these genes did not demonstrate any ecologically tied difference in selection pressures as we had hypothesized. Interestingly, an outlying group of isolates showed significantly large *FLO11* d_N / d_S values as compared to all other isolates. The geographical and ecological origins of these yeast were not uniform or related – making any assumptions of shared environmental pressures on this flocculin gene unobtainable. Furthermore, the isolates with large *FLO11* d_N / d_S values did not have correlated changes in d_N / d_S values in other *FLO* genes. As such, the reason

behind and the importance of this outlying group of isolates with large d_N / d_S values eludes our investigation.

Our investigation of possible differential selection pressures experienced by the PrD-encoding portions of yeast prion genes was more successful. By breaking up seven yeast prion genes into PrD and non-PrD portions (based on experimental evidence), we could separately evaluate the selection pressure imparted upon each portion of the gene. For all seven of the yeast prion genes examined, the PrD portion exhibited higher variation in its d_N / d_S value and a higher overall d_N / d_S value. This finding is particularly interesting because it suggests that selection is occurring at different rates in the PrDs than in the non-PrD regions. Moreover, this difference may be a result of the dissimilar constraints placed upon the sequence requirements for the PrD, where overall composition matters more than strict primary sequence (167, 204). Additionally, the presence of purifying selection of the PrDs suggests that yeast prions have an important biological function, potentially as a method to quickly adapt to environmental changes as other research has suggested (102, 133, 134, 151, 153).

We also examined just the rate of nonsynonymous changes in the PrDs versus the non-PrD regions. Interestingly, there were differing results of this examination among just the seven yeast prion proteins examined. *CYC8* displayed a greater d_N value in the non-PrD regions of the gene. However, very little is currently known about the exact Cyc8 PrD boundaries and this may impact our results (101). The yeast prion genes of *MOT3* and *NUP100* had d_N values that were similar between the PrD and non-PrD regions of the gene. Although, there may be limitations to our study here as well. The Mot3 PrD has had limited targeted experimentation to determine its boundaries – relying mostly on the PLAAC algorithm to delineate the PrD from the rest of the protein (109, 133). This uncertainty may mask effects in this study. In regard to Nup100, this protein has a

unique amino-acid composition among the seven yeast prion proteins examined here and once again has limited data regarding the boundaries of its PrD. The four yeast prion genes that demonstrated significantly higher rates of nonsynonymous substitutions in their PrDs versus their non-PrD regions – *RNQ1*, *SUP35*, *SWI1*, and *URE2* all have extensive research regarding their PrD boundaries (182). This specificity may have impacted our results – particularly for the small defined PrD used for *URE2*. However, these results suggest that the PrD of yeast prions may be relatively tolerant of amino acid alterations in wild strains and thus, perhaps indicating the usefulness of the PrDs in quickly responding to environmental stressors.

Lastly, our investigation of the diversity of the yeast prion proteins in these 1,011 *S. cerevisiae* isolates brought us to examining the prionization outcomes of the cataloged mutations. Although we could not perform *in vivo* work to characterize these different isolates and their prion capabilities, the *in silico* examination of their prionogenicity was possible. As such, we found that regardless of ecological origin, the yeast prion proteins had consistent distribution of predicted prionogenicity. This result would seem to argue against the importance of the role of yeast prions in environmental adaptation. On the other hand, as most isolates retain the prionogenicity of these proteins, this outcome would suggest that prionization – at the appropriate rates – provides a beneficial boost to fitness. Given the limitations of the prediction algorithms we used, the differences in prion formation rates in the various strains likely alluded us. The complicated biology underpinning yeast prions, their formation rates, and their regulation may indeed make discernment of evolutionary relationships between the various identified yeast prions particularly difficult. Further *in vivo* study is required.

Moving forward, we have multiple paths to further investigate PrD evolution in yeast via sequencing data as well as *in vivo* experiments. In regard to additional bioinformatic examinations,

there have been at least another 220 budding yeast isolates sequenced across many species – allowing for further expansion of base data set and opening for more comparisons (218). Also, new yeast prions including the experimentally confirmed [*ESI*$^{+}$] and [*SMAUG*$^{+}$] have been identified (104, 105). The genes for the protein determinants of these two yeast prions – SNT1 and VTS1 – and other recently recognized yeast prion-related genes could be examined. Coinciding with these efforts, *in vivo* experiments will be critical. The *S. cerevisiae* isolates from the Peter et al., 2018 study are largely available in culturable forms (211). As such, we may be able to test these isolates for the presence of various prions – via fluorescence reporters or the like – or the capability for particular prions to form in them. This data would provide an experimental readout for prionogenicity as opposed to the calculated scores produced by predictive algorithms as we used here. Alternatively, we could replace yeast prion genes in a generic lab strain with different versions found in the many sequenced isolates – perhaps simplifying experimental processes and better controlling for variability. Any *in vivo* method used to further analyze the diversity of yeast prion proteins and genes will require dutiful efforts but may lend important insights into the significance of these epigenetic elements.

CHAPTER FOUR

Identifying novel protein aggregators linked to cancer

Introduction

In humans, prion proteins have long been associated with neurological diseases. These connections include the namesake prion protein (PrP) and Creutzfeldt-Jakob disease (CJD) or the prion-like proteins of tau in Alzheimer's disease (AD) and alpha-synuclein (α-syn) in Parkinson's disease (PD). As such, much research has focused on the role that prion and prion-like proteins play in neurodegeneration; however, recent investigations have suggested that similar phenomena may underlie another serious category of human disease – cancers.

One mode of scientific thinking has been to view neurodegenerative disease and cancer as opposing ends of a spectrum. At the one end, there exists escalating cell death and at the other end, uncontrolled cellular proliferation reigns supreme. Inverse correlations have been commonly observed between certain neurodegenerative diseases and certain types of cancer. For example, longitudinal studies have found that individuals that have had cancer have a decreased risk of AD and vice-versa (49, 50). Transcriptomic analyses have also underscored numerous divergences between cells involved in neurodegeneration or cancer (51, 52). On the other hand, both disease states clearly relate to a disruption of homeostasis and the aging organism. Research into the mechanistic bases of these two classes of disease have found notable similarities – particularly involving DNA damage and cell dysregulation (55). Through research into the behavior of mutant p53, the prion concept has been found to play a mechanistic role in both neurodegeneration and carcinogenesis.

As the proclaimed guardian of the genome, p53 plays an important role as a tumor suppressor. In fact, the p53 gene is the most mutated gene in cancers (42% by one study) and these mutations occur most frequently at hotspots in the DNA binding domain (57, 58). Recent studies have revealed that multiple of these mutations destabilize normally folded p53, allow for it to adopt an alternative conformation, and form oligomers or aggregates that can pull in other proteins including wild-type p53, p63, and p73 (62, 66, 67, 69). These characteristics mirror those of the prion-like proteins most commonly associated with neurodegenerative disease. Perhaps most convincingly, p53 function can be rescued and tumors shrunken by preventing the aggregation of mutant p53 with a designed peptide that binds to a core aggregation region (73). Thus, the finding of prion-like behavior by p53 has opened new avenues to cancer treatment.

As p53 demonstrates the ability for prion-like protein aggregation to drive oncogenesis and tumorigenesis, there may exist additional proteins that contribute to cancer via prion-like behavior. Here we attempt to identify novel protein aggregators in cancer. In doing so, we have chosen to survey cancer-related proteins for aggregator candidates using an algorithmic approach and to test them using model systems – yeast and mammalian cell lines. A range of algorithms targeting various protein characteristics such as disorderedness, amyloidogenicity, and prionogenicity have been implemented. An additional focus was placed upon proteins belonging to the BAF complex due to the presence of multiple proteins with predicted prion-like domains and the high incidence rate (~20%) of mutations in the complex in cancers (219–221). Interestingly, the BAF complex, also known as the human SWI/SNF complex, contains functional homologs to Swi1, the protein determinant of the yeast prion [*SWI*$^{+}$]. Through these investigations, we have identified multiple human proteins that may indeed be driven to aggregate or even contain a bona fide prion domain.

Materials and Methods

Bioinformatic analysis

Protein amino-acid sequences were acquired from the UniProt database (222). The canonical sequence for each protein was used. Sequences were evaluated using a range of predictive algorithms. DISOPRED3, FoldIndex, PAPA, ZipperDB, and Zyggregator analysis was conducted using the default parameters (146, 158, 159, 164, 165). PASTA analysis was conducted using an energy threshold of -5 PASTA Energy Units (PEU) (166). PLAAC analysis was conducted using the default core length of 60 and background amino-acid probabilities for *Homo sapiens* (160). For combinatorial ranking, scores from FoldIndex, PAPA, PASTA disorder, and PASTA energy were adjusted via transformation to scale from 0 to 1 and multiplied by a weighting factor (W_{FI}, W_{PAPA}, W_{PE}, W_{PD}). Positive PLAAC hits were granted a full weighting factor (W_{PLAAC}) as a score whereas negative hits received no additional score. Candidates were ranked according to the sum of all weighted scores.

Yeast strains and media

Yeast strains used in this study are listed in Table 2.1. The W303 *sup35Δ / SUP35::TRP1 / p316SUP35$_{FL}$* strain was provided by the Weissman Lab (University of California, San Francisco).

Table 4.1. Yeast strains used in chapter four.

Identifier	Background	Relevant Genotype	Prion State	Source
746	BY4741		[*RNQ*$^+$]	ATCC
YJW561	W303	*sup35Δ / SUP35::TRP1 / p316SUP35$_{FL}$*	[*PSI*$^+$][*PIN*$^+$]	Weissman Lab

Yeast were grown according to established protocols at 30° C in either yeast extract-peptone-dextrose (YPD) or synthetic complete (SC) media minus appropriate amino acids (e.g. leucine) (184). When indicated, media was supplemented with 1 g/l 5-FOA for counter-selection against a *URA3*-carrying plasmid.

Plasmid construction

Plasmids used in this study are listed in Table 4.2. Briefly, the collection of *pENTR221-Candidate PrLD* plasmids were created by Gateway cloning with BP clonase (Invitrogen, Carlsbad, CA, USA) of purchased synthesized attB-flanked gene fragments (ordered from Integrated DNA Technologies, Coralville, IA, USA) into *pDONR221* (Invitrogen, Carlsbad, CA, USA). Additional entry clone plasmids (*pENTR221*) for BAF complex proteins (ARID2, BAF53B, CREST, SS18) and their mutants were created by Gateway cloning of either purchased synthesized attB-flanked gene fragments or PCR-produced attB-flanked gene fragments into *pDONR221*. See Table 4.3 for primer information. Initial cDNA for these genes were purchased from various repositories (Dharmacon, Lafayette, CO, USA; DNASU, Tempe, AZ, USA; PlasmID, Boston, MA, USA). These *pENTR221-BAF Complex Protein* plasmids were then used for Gateway cloning with LR clonase (Invitrogen, Carlsbad, CA, USA) into a yeast expression vector (*pAG425GAL-ccdB-EGFP*) to create the final inducible expression clones (e.g., *pAG425GAL-CREST-EGFP*).

Plasmids of the *p415TEF-Candidate PrLD-MC* series were created through PCR of the Candidate PrLD DNA from the corresponding *pENTR221-Candidate PrLD*. The forward primer introduced a *SpeI* site, and the reverse primer introduced a *BamHI* site. Through these sites, Candidate PrLD DNAs were cloned into *p415TEF- -MC* to create the final Candidate PrLD-

$Sup35_{MC}$ fusion constructs (e.g., *p415TEF-CREST PrLD-MC*). All cloned plasmids were confirmed through sequencing (ACGT, Inc., Wheeling, IL, USA).

Table 4.2. Plasmids used in chapter four.

Name	Marker	Replicon	Promoter	Use	Source
pDONR221	-	-	-	Gateway cloning	Invitrogen
pENTR221-ABLIM3 PrLD	-	-	-	Gateway cloning	This study
pENTR221-BRD4 PrLD	-	-	-	Gateway cloning	This study
pENTR221-CLINT1 PrLD	-	-	-	Gateway cloning	This study
pENTR221-CLOCK PrLD A	-	-	-	Gateway cloning	This study
pENTR221-CLOCK PrLD B	-	-	-	Gateway cloning	This study
pENTR221-CLOCK PrLD AB	-	-	-	Gateway cloning	This study
pENTR221-CNOT2 PrLD	-	-	-	Gateway cloning	This study
pENTR221-CREST PrLD	-	-	-	Gateway cloning	This study
pENTR221-HNRNPA1 PrLD	-	-	-	Gateway cloning	This study
pENTR221-HNRNPDL PrLD	-	-	-	Gateway cloning	This study
pENTR221-HNRNPU PrLD	-	-	-	Gateway cloning	This study
pENTR221-ILF3 PrLD	-	-	-	Gateway cloning	This study
pENTR221-JMJD6 PrLD	-	-	-	Gateway cloning	This study
pENTR221-PDCD6IP PrLD	-	-	-	Gateway cloning	This study
pENTR221-QRICH1 PrLD	-	-	-	Gateway cloning	This study
pENTR221-TIAL1 PrLD	-	-	-	Gateway cloning	This study
pENTR221-ARID2	-	-	-	Gateway cloning	This study
pENTR221-ARID2 R285Q	-	-	-	Gateway cloning	This study
pENTR221-BAF53B	-	-	-	Gateway cloning	This study
pENTR221-BAF53B P308L	-	-	-	Gateway cloning	This study
pENTR221-BAF53B P308S	-	-	-	Gateway cloning	This study
pENTR221-BAF53B R338L	-	-	-	Gateway cloning	This study
pENTR221-BAF53B P371S	-	-	-	Gateway cloning	This study
pENTR221-BAF53B R386L	-	-	-	Gateway cloning	This study
pENTR221-BAF53B YO	-	-	-	Gateway cloning	This study
pENTR221-BAF53B YO P308L	-	-	-	Gateway cloning	This study
pENTR221-BAF53B YO P308S	-	-	-	Gateway cloning	This study
pENTR221-BAF53B YO R338L	-	-	-	Gateway cloning	This study
pENTR221-BAF53B YO P371S	-	-	-	Gateway cloning	This study
pENTR221-BAF53B YO R386L	-	-	-	Gateway cloning	This study
pENTR221-CREST	-	-	-	Gateway cloning	This study
pENTR221-CREST A263T	-	-	-	Gateway cloning	This study
pENTR221-CREST A348S	-	-	-	Gateway cloning	This study
pENTR221-CREST P372L	-	-	-	Gateway cloning	This study
pENTR221-SS18	-	-	-	Gateway cloning	This study
pENTR221-SS18 P117L	-	-	-	Gateway cloning	This study
pENTR221-SS18 P168S	-	-	-	Gateway cloning	This study
pENTR221-SS18 P252L	-	-	-	Gateway cloning	This study
pENTR221-SS18 P341S	-	-	-	Gateway cloning	This study
pAG425GAL-ccdB-EGFP	LEU2	2μ	GAL	Gateway cloning	(223)
pAG425GAL-ARID2-EGFP	LEU2	2μ	GAL	Expression of ARID2-EGFP	This study
pAG425GAL-ARID2 R285Q-EGFP	LEU2	2μ	GAL	Expression of ARID2 R285Q-EGFP	This study
pAG425GAL-BAF53B-EGFP	LEU2	2μ	GAL	Expression of BAF53B-EGFP	This study
pAG425GAL-BAF53B P308L-EGFP	LEU2	2μ	GAL	Expression of BAF53B P308L-EGFP	This study
pAG425GAL-BAF53B P308S-EGFP	LEU2	2μ	GAL	Expression of BAF53B P308S-EGFP	This study
pAG425GAL-BAF53B R338L-EGFP	LEU2	2μ	GAL	Expression of BAF53B R338L-EGFP	This study
pAG425GAL-BAF53B P371S-EGFP	LEU2	2μ	GAL	Expression of BAF53B P371S-EGFP	This study
pAG425GAL-BAF53B R386L-EGFP	LEU2	2μ	GAL	Expression of BAF53B R386L-EGFP	This study
pAG425GAL-CREST-EGFP	LEU2	2μ	GAL	Expression of CREST-EGFP	This study
pAG425GAL-CREST A263T-EGFP	LEU2	2μ	GAL	Expression of CREST A263T-EGFP	This study
pAG425GAL-CREST A348S-EGFP	LEU2	2μ	GAL	Expression of CREST A348S-EGFP	This study
pAG425GAL-CREST P372L-EGFP	LEU2	2μ	GAL	Expression of CREST P372L-EGFP	This study
pAG425GAL-SS18-EGFP	LEU2	2μ	GAL	Expression of SS18-EGFP	This study

pAG425GAL-SS18 P117L-EGFP	LEU2	2μ	GAL	Expression of SS18 P117L-EGFP	This study
pAG425GAL-SS18 P168S-EGFP	LEU2	2μ	GAL	Expression of SS18 P168S-EGFP	This study
pAG425GAL-SS18 P252L-EGFP	LEU2	2μ	GAL	Expression of SS18 P252L-EGFP	This study
pAG425GAL-SS18 P341S-EGFP	LEU2	2μ	GAL	Expression of SS18 P341S-EGFP	This study
p415TEF-ABLIM3 PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of ABLIM3 PrLD-MC	This study
p415TEF-BRD4 PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of BRD4 PrLD-MC	This study
p415TEF-CLINT1 PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of CLINT1 PrLD-MC	This study
p415TEF-CLOCK PrLD A-MC	LEU2	CEN6/ARSH4	TEF	Expression of CLOCK PrLD A-MC	This study
p415TEF-CLOCK PrLD B-MC	LEU2	CEN6/ARSH4	TEF	Expression of CLOCK PrLD B-MC	This study
p415TEF-CLOCK PrLD AB-MC	LEU2	CEN6/ARSH4	TEF	Expression of CLOCK PrLD AB-MC	This study
p415TEF-CNOT2 PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of CNOT2 PrLD-MC	This study
p415TEF-CREST PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of CREST PrLD-MC	This study
p415TEF-HNRNPA1 PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of HNRNPA1 PrLD-MC	This study
p415TEF-HNRNPDL PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of HNRNPDL PrLD-MC	This study
p415TEF-HNRNPU PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of HNRNPU PrLD-MC	This study
p415TEF-ILF3 PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of ILF3 PrLD-MC	This study
p415TEF-JMJD6 PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of JMJD6 PrLD-MC	This study
p415TEF-PDCD6IP PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of PDCD6IP PrLD-MC	This study
p415TEF-QRICH1 PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of QRICH1 PrLD-MC	This study
p415TEF-TIAL1 PrLD-MC	LEU2	CEN6/ARSH4	TEF	Expression of TIAL1 PrLD-MC	This study
p316Sup35$_{FL}$	URA3	CEN6/ARSH4	SUP35	Expression of Sup35	Weissman Lab

Site-directed mutagenesis

The suite of *ARID2*, *BAF53B*, *CREST*, and *SS18* mutants was produced via incorporation of base substitutions in PCR primers (Table 4.3) and usage of the corresponding *pENTR221-BAF Complex Protein* as the template. All PCRs were conducted using PrimeSTAR HS DNA polymerase (Takara Bio, Mountain View, CA, USA) and the manufacturer's recommended protocols. Custom primers were ordered from Integrated DNA Technologies (Coralville, IA, USA) and annealing temperatures were estimated via the Integrated DNA Technologies' OligoAnalyzer Tool.

Table 4.3. Primers used in chapter four.

Name	Sequence (5' - 3')	Resulting Plasmid
ARID2 attB For	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGGCAAACTC GACGGGG	*pENTR221-ARID2*
ARID attB Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTCTGCAGCATTTC TGAGTCTTTTTCT	*pENTR221-ARID2*
BAF53B attB For	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGAGCGGGGG CGTCTAC	*pENTR221-BAF53B*

BAF53B attB Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTGGGGCACTTTCGCTCCAC	*pENTR221-BAF53B*
CREST attB For	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAATGTCCGTGGCCTTCGCG	*pENTR221-CREST*
CREST attB Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTCAACTGCTGGTAATTTCCATACTGGCC	*pENTR221-CREST*
SS18 attB For	GGGGACAAGTTTGTACAAAAAAAGCAAGCTTAATGTCTGTGGCTTTCGCG	*pENTR221-SS18*
SS18 attB Rev	GGGGACCACTTTGTACAAGAAAGCTGGGTTCTGCTGGTAATTTCCATACTGTC	*pENTR221-SS18*
ARID2 R285Q For	GCTGGGCATTAACGATATTGAAGGACAGCAGGTACTTCAGATTGCAGTGA	*pENTR221-ARID2 R285Q*
ARID2 R285Q Rev	TCACTGCAATCTGAAGTACCTGCTGTCCTTCAATATCGTTAATGCCCAGC	*pENTR221-ARID2 R285Q*
BAF53B P308L For	GGGCCTGTTTGATCTCTCGAACGTCAAGGGCCTGTCGGG	*pENTR221-BAF53B P308L*
BAF53B P308L Rev	CCCGACAGGCCCTTGACGTTCGAGAGATCAAACAGGCCC	*pENTR221-BAF53B P308L*
BAF53B P308S For	CCCTGAGGGCCTGTTTGATTCCTCGAACGTCAAGGGCCTG	*pENTR221-BAF53B P308S*
BAF53B P308S Rev	CAGGCCCTTGACGTTCGAGGAATCAAACAGGCCCTCAGGG	*pENTR221-BAF53B P308S*
BAF53B R338L For	CGGCATGTGTGACATTGATATTCTCCCGGGCCTGTACGGG	*pENTR221-BAF53B R338L*
BAF53B R338L Rev	CCCGTACAGGCCCGGGAGAATATCAATGTCACACATGCCG	*pENTR221-BAF53B R338L*
BAF53B P371S For	CCCAGAAGACCCCATCGAGCATGCGACTGAAACTCATTGC	*pENTR221-BAF53B P371S*
BAF53B P371S Rev	GCAATGAGTTTCAGTCGCATGCTCGATGGGGTCTTCTGGG	*pENTR221-BAF53B P371S*
BAF53B R386L For	GCAACAGCACCATGGAGCTCAAGTTCAGCCCCTGGATCG	*pENTR221-BAF53B R386L*
BAF53B R386L Rev	CGATCCAGGGGCTGAACTTGAGCTCCATGGTGCTGTTGC	*pENTR221-BAF53B R386L*
BAF53B YO P308L For	CAGAGGGCTTATTCGATCTTTCAAATGTGAAAGGTCT	*pENTR221-BAF53B YO P308L*
BAF53B YO P308L Rev	AGACCTTTCACATTTGAAAGATCGAATAAGCCCTCTG	*pENTR221-BAF53B YO P308L*
BAF53B YO P308S For	CCAGAGGGCTTATTCGATTCTTCAAATGTGAAAGGTC	*pENTR221-BAF53B YO P308S*
BAF53B YO P308S Rev	GACCTTTCACATTTGAAGAATCGAATAAGCCCTCTGG	*pENTR221-BAF53B YO P308S*
BAF53B YO R338L For	ATGTGCGATATTGACATACTGCCCGGACTTTATGGGAG	*pENTR221-BAF53B YO R338L*
BAF53B YO R338L Rev	CTCCCATAAAGTCCGGGCAGTATGTCAATATCGCACAT	*pENTR221-BAF53B YO R338L*
BAF53B YO P371S For	GTTATCACAGAAGACACCTTCTAGTATGAGATTGAAACT	*pENTR221-BAF53B YO P371S*
BAF53B YO P371S Rev	AGTTTCAATCTCATACTAGAAGGTGTCTTCTGTGATAAC	*pENTR221-BAF53B YO P371S*
BAF53B YO R386L For	GTAATTCCACGATGGAACTTAAGTTTTCTCCTTGGAT	*pENTR221-BAF53B YO R386L*
BAF53B YO R386L Rev	ATCCAAGGAGAAAACTTAAGTTCCATCGTGGAATTAC	*pENTR221-BAF53B YO R386L*
CREST A263T For	AGCCACAGCCAGGGCACCGCGGAG	*pENTR221-CREST A263T*

CREST A263T Rev	CTCCGCGGTGCCCTGGCTGTGGCT	*pENTR221-CREST A263T*
CREST A348S For	ACGGGTCTTCCCAGGGAGCCCCGTCAC	*pENTR221-CREST A348S*
CREST A348S Rev	GTGACGGGGCTCCCTGGGAAGACCCGT	*pENTR221-CREST A348S*
CREST P372L For	GGAAGCTACCGAGCACTGCAGACAGCGCCG	*pENTR221-CREST P372L*
CREST P372L Rev	CGGCGCTGTCTGCAGTGCTCGGTAGCTTCC	*pENTR221-CREST P372L*
SS18 P117L For	TGGGGGTCCTCCTGCACTGCACATGCAGAA	*pENTR221-SS18 P117L*
SS18 P117L Rev	TTCTGCATGTGCAGTGCAGGAGGACCCCCA	*pENTR221-SS18 P117L*
SS18 P168S For	ACCATTCTGTGTCATCATCACAGAGCATGCCAGTACAGAATCAGA	*pENTR221-SS18 P168S*
SS18 P168S Rev	TCTGATTCTGTACTGGCATGCTCTGTGATGATGACACAGAATGGT	*pENTR221-SS18 P168S*
SS18 P252L For	GAGACAGATTCCTCTCTATAGACCTCCTCAACAGGGCCC	*pENTR221-SS18 P252L*
SS18 P252L Rev	GGGCCCTGTTGAGGAGGTCTATAGAGAGGAATCTGTCTC	*pENTR221-SS18 P252L*
SS18 P341S For	GGACCACCTTCACAACAGGGATATCCACCCCAGCAG	*pENTR221-SS18 P341S*
SS18 P341S Rev	CTGCTGGGGTGGATATCCCTGTTGTGAAGGTGGTCC	*pENTR221-SS18 P341S*
SpeI-ABLIM3 PrLD For	ATGGTCCCGACTAGTATGAACACCAGCATCCCCTAC	*p415TEF-ABLIM3 PrLD-MC*
ABLIM3 PrLD-BamHI Rev	TCAAATGGATCCCAGCTGCTGGTAGTCCTGGGT	*p415TEF-ABLIM3 PrLD-MC*
SpeI-BRD4 PrLD For	GTCCCGACTAGTATGTCTATGCAGATGCAGCTTTACTTG	*p415TEF-BRD4 PrLD-MC*
BRD4 PrLD-BamHI Rev	TCAAATGGATCCTTGCTTGTCCTTATCCGGTGC	*p415TEF-BRD4 PrLD-MC*
SpeI-CLINT1 PrLD For	TCAAATACTAGTATGTCTCAGCCGTCAAAACCCCAGCAA	*p415TEF-CLINT1 PrLD-MC*
CLINT1 PrLD-BamHI Rev	TCAAATGGATCCGGGTTGCACCGTACCACTTGT	*p415TEF-CLINT1 PrLD-MC*
SpeI-CLOCK PrLD A For	GTCGCCACTAGTATGTCTCAGGGCCTGCAGATGTTCCTG	*p415TEF-CLOCK PrLD A-MC*
CLOCK PrLD A-BamHI Rev	TCAAATGGATCCGCTCTGGGTGCTGTTCTGGGG	*p415TEF-CLOCK PrLD A-MC*
SpeI-CLOCK PrLD B For	GTCGCCACTAGTATGTCTGCCACCCAACAACAGCAGTCT	*p415TEF-CLOCK PrLD B-MC*
CLOCK PrLD B-BamHI Rev	TCAAATGGATCCTTGCTGCTGCTGCTGGCTCTG	*p415TEF-CLOCK PrLD B-MC*
SpeI-CLOCK PrLD AB For	GTCGCCACTAGTATGTCTCAGGGTCTGCAAATGTTTTTGCAACAG	*p415TEF-CLOCK PrLD AB-MC*
CLOCK PrLD AB-BamHI Rev	TCAAATAGGTCCCTGTTGTTGCTGCTGGCTTTGATG	*p415TEF-CLOCK PrLD AB-MC*
SpeI-CNOT2 PrLD For	GTCGCCACTAGTATGTCTATGAACCCCAGAAACATGATGAAC	*p415TEF-CNOT2 PrLD-MC*
CNOT2 PrLD-BamHI Rev	TCAAATGGATCCGCCGTTGAAGATGTTGCTGCT	*p415TEF-CNOT2 PrLD-MC*
SpeI-CREST PrLD For	GTCGCCACTAGTATGTCTAGCAATCAGAATATGCAATCA	*p415TEF-CREST PrLD-MC*

CREST PrLD-BamHI Rev	GTCGCCGGATCCCTGTTGATAATTACCATACTGTCC	*p415TEF-CREST PrLD-MC*
SpeI-HNRNPA1 PrLD For	TCAAATACTAGTATGTCTAGCCAGAGGGGTCGGAGCGGG	*p415TEF-HNRNPA1 PrLD-MC*
HNRNPA1 PrLD-BamHI Rev	GTCGCCTCAAATGGATCCAAATCTCCTACCAGAGCCGTA	*p415TEF-HNRNPA1 PrLD-MC*
SpeI-HNRNPDL PrLD For	GTCGCCACTAGTATGTCTCAGCAGCAACAGCAACAAAAG	*p415TEF-HNRNPDL PrLD-MC*
HNRNPDL PrLD-BamHI Rev	GTCGCCGGATCCGTAAGGCTGGTAGTTATTTTGATGATT	*p415TEF-HNRNPDL PrLD-MC*
SpeI-HNRNPU PrLD For	GTCGCCACTAGTATGTCTAGCAACAAGAATAAGTCAGGA	*p415TEF-HNRNPU PrLD-MC*
HNRNPU PrLD-BamHI Rev	GTCGCCGGATCCATAATAGCCTTGATGGTAGTG	*p415TEF-HNRNPU PrLD-MC*
SpeI-ILF3 PrLD For	TCAAATACTAGTATGTCTGGCGGCGCCAACCATGGGGGA	*p415TEF-ILF3 PrLD-MC*
ILF3 PrLD-BamHI Rev	TCAAATGGATCCTCGGTATTGATAGTTCATGGAATGATCGGCA TTTCTCCCGTA	*p415TEF-ILF3 PrLD-MC*
SpeI-JMJD6 PrLD For	TCAAATACTAGTATGTCTACCGGAATCGCCAGCGACAGC	*p415TEF-JMJD6 PrLD-MC*
JMJD6 PrLD-BamHI Rev	TCAAATGGATCCGCGTGAAGAAGACCGCTCCTTGCTTAC	*p415TEF-JMJD6 PrLD-MC*
SpeI-PDCD6IP PrLD For	TCAAATACTAGTATGTCTAGCGCACCACCACCCCAAGCT	*p415TEF-PDCD6IP PrLD-MC*
PDCD6IP PrLD-BamHI Rev	TCAAATGGATCCCTGTTGTGGGTAATATGACTGCTGAGGAGG CTG	*p415TEF-PDCD6IP PrLD-MC*
SpeI-QRICH1 PrLD For	TCAAATACTAGTATGTCTCAGCCACAAACGCAGCAAGAA	*p415TEF-QRICH1 PrLD-MC*
QRICH1 PrLD-BamHI Rev	TCAAATGGATCCTTGTACCTGGATCTGAGCGGC	*p415TEF-QRICH1 PrLD-MC*
SpeI-SS18 PrLD For	TCAAATACTAGTATGTCTTCTAATCAGAACATGCAGTCATTG	*p415TEF-SS18 PrLD-MC*
SS18 PrLD-BamHI Rev	TCAAATGGATCCTTGTTGATAATTCCCGTATTGGCC	*p415TEF-SS18 PrLD-MC*
SpeI-TIAL1 PrLD For	TCAAATACTAGTATGTCTGTGGACTATAGCCAGTGGGGG	*p415TEF-TIAL1 PrLD-MC*
TIAL1 PrLD-BamHI Rev	TCAAATGGATCCCTGTGTTTGGTAACTTGCCATTCCATA	*p415TEF-TIAL1 PrLD-MC*
Linker-SUP_{MC} For	GGTGGTCCTGGTGGTGGTATGTCTTTGAACGACTTTCAAAAG	*p415TEF-Candidate PrLD-MC*
$SUP35_{MC}$-XhoI Rev	CTGCGAGCCCTCGAGTTACTCGGCAATTTTAACAATTTTACCA ATTGCT	*p415TEF-Candidate PrLD-MC*

Yeast transformation

Yeast were transformed as previously described (184). In brief, cells were spun down at 2500 rpm for 3 minutes, supernatant removed, and cells resuspended in 1 ml H_2O. Cells were then spun down again at 2500 rpm for 3 minutes, supernatant removed, and cells resuspended in 1 ml

of 0.1M lithium acetate. After 10 minutes, the cells were pelleted again, supernatant removed, and cells resuspended in 100 µl of Li-PEG (0.1M lithium acetate, 30% polyethylene-glycol 3350 in H_2O). From this mixture, 94.5 µl of resuspended cells were combined with 3.5 µl of ssDNA and 2.0 µl of the appropriate plasmid. The transformation mixture was then incubated at 42 °C for 30 minutes. Following, the transformation mixture was moved to ice for 5 minutes before spreading onto the appropriate selective media.

Microscopy

Images were captured using a Zeiss Axiovert 200 epifluorescence microscope with an attached camera and the Axiovision AC software (Zeiss, Oberkochen, Germany). Cell samples were visualized with a 100x objective and the appropriate filters for differential interference contrast (DIC) or green fluorescent protein (GFP).

SDD-AGE

Yeast samples for SDD-AGE were grown overnight in selective media (3 ml) and prepared similarly to previously described (121). The next day, the culture was diluted into a larger volume of selective media (30 ml total) and grown over approximately 4 hours at 30 °C with shaking at 225 rpm. The yeast was harvested afterwards by spinning down at 2500 rpm for 5 minutes. The media was removed, and the resulting cell pellet was washed with 10 ml of H_2O. After another spin down, the H_2O was removed and 800 µl of cell lysis buffer (50 mM Tris-HCL pH 7.5, 50 mM KCl, 10 mM $MgCl_2$, 5% glycerol, 10 mM PMSF, cOmplete Mini Protease Inhibitor Cocktail (Roche, Basel, Switzerland)) was added. The cell suspension was transferred to a 2.0 ml screw-cap tube filled halfway with silica beads and additional cell lysis buffer was added to fill the tube to maximum. A Mini-Beadbeater 16 was used to lyse the suspended cells by beating 5x in 1-minute

intervals with resting on ice for 1 minute in between. The resulting samples were then used for SDD-AGE.

SDD-AGE was conducted as described previously (188). Briefly, yeast lysates were first mixed with 4x Laemmli sample buffer (2x TAE, 20% glycerol, 8% SDS, 0.1% bromophenol blue). Samples were either incubated at room temperature for 7 minutes or boiled for 10 minutes. Samples were loaded onto 1.5% agarose, 0.1% SDS gels. After completion of electrophoresis, samples were transferred to PVDF membrane using capillary action and 1x TBS.

Immunoblotting

Membranes were blocked via incubation in 5% milk in PBS at either 4 C overnight or room temperature for 2 hours. Blots were washed 3x for 5 minutes with PBS + 0.01% Tween-20 before probing with primary antibody for 2 hours at room temperature. The following primary antibodies were used for detection: anti-Sup35 antibody (gift from the Liebman Lab, University of Nevada, Reno, NV, USA) or anti-FLAG antibody. All primary antibodies were used at a 1:2500 dilution. Blots were washed 3x for 5 minutes with PBS + 0.01% Tween-20 before probing with horseradish peroxidase-conjugated rat anti-mouse secondary antibody (Cell Signaling Technology, Danvers, MA, USA) for 1 hour at room temperature. Blots were washed 3x for 5 minutes with PBS + 0.01% Tween-20 before incubation with Clarity Western ECL Substrate (Bio-Rad, Hercules, CA, USA). Blots were imaged using a ChemiDoc Imaging System (Bio-Rad), Hercules, CA, USA).

***De novo* prion formation assay**

W303 *sup35Δ / p316SUP35*$_{FL}$ [*PSI*$^{+}$] cells were independently transformed with *p415TEF-Candidate PrLD-MC* constructs. Transformants were grown on -LU media and a color change to red was observed – indicating that the fusion proteins were functional in translational termination. Three red colonies were selected for each construct and were streaked onto -L+5-FOA media to

select against *p316SUP35$_{FL}$*. Resulting colonies were selected and re-streaked onto both -Leu and -LU media to confirm loss of *p316SUP35$_{FL}$*. The resulting W303 *sup35Δ* / *p415TEF-Candidate PrLD-MC* yeast isolates were grown in -Leu media overnight and then counted and spread onto YPD (10^3 cells) and -Ade (10^6, 10^7, 10^8 cells) media. Colonies on all plates were counted as possible – plates containing >1000 colonies were not marked as uncountable. Colonies on the YPD media that exhibited any white or pink coloration were selected for further investigation and patched onto -Ade plates. Colonies on the -Ade media were patched onto new -Ade plates. Patched yeast isolates were replica plated onto YPD and YPD + GdnHCl media and then both replica plated back to YPD media to check for curability of the color phenotype.

Results

Multiple algorithm approach identifies candidate novel protein aggregators

To determine the focus of our search for novel protein aggregators involved in cancer, we analyzed a wide range of protein sequences with multiple prediction algorithms (Figure 4.1). A set of reference human protein sequences was collected from the UniProt database. This sequence set included any protein with connections to the cancer-related keywords (e.g., carcinoma, leukemia, melanoma, sarcoma). In all, the initial set totaled 9260 unique proteins after removal of duplicate hits from the various keywords. This number represents approximately ~45% of all reviewed and manually annotated human proteins in the UniProt database. From here, we processed all protein sequences with a range of algorithms focused on predicting different protein characteristics (i.e., prionogenicity, amyloidogenicity, disorderedness).

The PAPA (prion aggregation prediction algorithm) and PLAAC (Prion-Like Amino Acid Composition) algorithms evaluated the prionogenicity of the cancer-related protein set (159, 160).

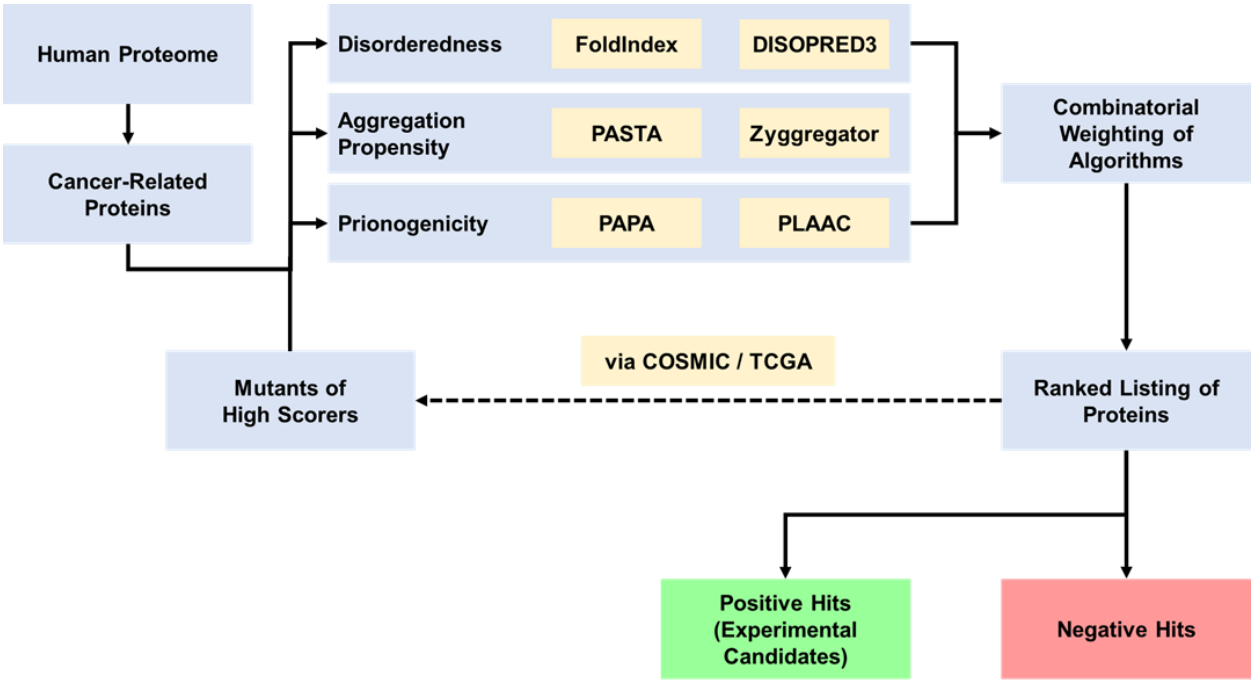


Figure 4.1. Identification and ranking of candidate proteins.

Cancer-related proteins were gleaned from the UniProt database using associated cancer-related keywords (e.g., carcinoma, leukemia). All proteins were evaluated by the listed algorithms to predict three main characteristics. FoldIndex and DISOPRED3 evaluated disorderedness. PASTA and Zyggregator evaluated aggregation propensity. PAPA and PLAAC evaluated prionogenicity. The scores from these various algorithms were normalized and combined after weighting to provide a synthesized score. Proteins were ranked by this score and approximately the top 100 were selected as experimental candidates. Mutants of high scoring candidates were fed through the same algorithms to predict whether these mutations would increase any scores.

In terms of parameters, PAPA was used with its default threshold while PLAAC used the default core length (60) and the background amino acid frequencies for *Homo sapiens* instead of *Saccharomyces cerevisiae*. Of the 9260 proteins evaluated, PAPA reported positive hits for possible prion-like domains in 723 proteins (~7.8%). PLAAC yielded a smaller list of positive hits with only 140 proteins (~1.5%) with predicted prion-like regions. A total of 87 proteins (~0.9%) were predicted to contain prionogenic regions by both PAPA and PLAAC (Figure 4.2A).

Disorderedness acts as a common trait among many prion or prion-like proteins. For evaluating the disorderedness of the cancer-related protein set, we implemented FoldIndex and the disorderedness component of PASTA (prediction of amyloid structural aggregation) (158, 166). Of the 9260 proteins evaluated, FoldIndex reported that 5759 (~62.2%) contained a disordered region of greater than 50 residues in length and 2652 (~28.6%) contained a disordered region of greater than 100 residues in length (Figure 4.2B). Meanwhile, PASTA predicted that 3558 (~38.4%) candidate proteins were at least 25% disordered and of those, 1219 (~13.1%) were at least 50% disordered (Figure 4.2C). Not surprisingly – given the disordered nature of many prion-favorable amino-acids, the group of proteins with at least 25% disorderedness contained ~85.7% (120) of PLAAC positive hits. However, PAPA positive hits were not obviously biased towards proteins predicted by PASTA to be highly disordered as only ~40.0% (289) were at least 25% disordered (similar to the proportion of the entire protein set).

For evaluation of amyloid aggregation propensity, all proteins were assigned a predicted PASTA energy score based on the best scoring predicted amyloid-forming peptide contained within the protein. Of the 9260 proteins in the initial set, the overwhelming majority (8447, ~91.2%) contained at least a short amyloidogenic sequence at the energy threshold used (5 PEU) (Figure 4.2C). Even doubling this threshold (to 10 PEU) reduces the pool of proteins to 2653

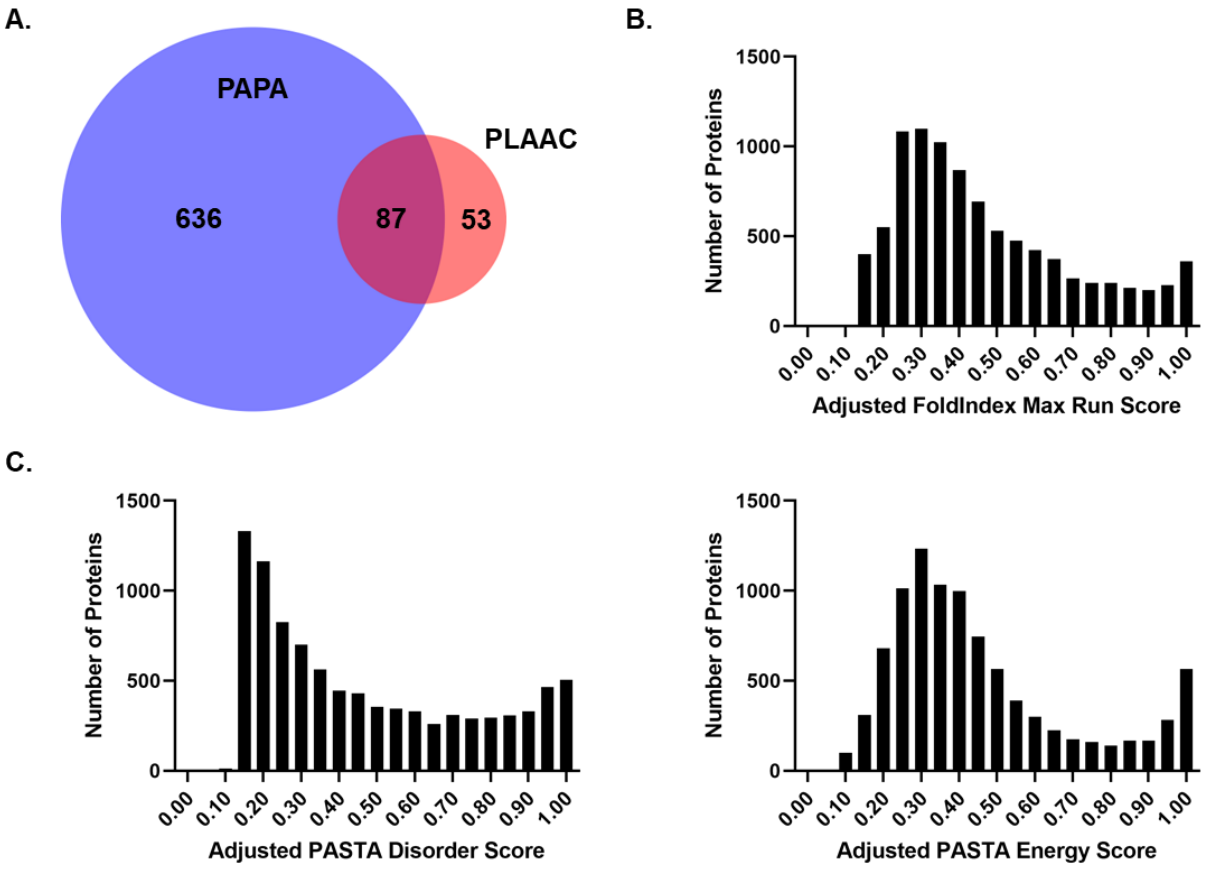


Figure 4.2. Predicted prionogenicity and disorderedness of human cancer-related proteins.

A. Venn diagram of positive hits from PAPA and PLAAC. Of the 9260 proteins evaluated, a total of 776 proteins (~8.3%) were predicted to contain a prion-like region by either PAPA (723, ~7.8%) or PLAAC (140, ~1.5%). Created using DeepVenn.
B. Distribution of adjusted FoldIndex Max Run scores of the entire protein set. Raw FoldIndex Max Run scores were normalized and transformed to scale from 0 (smaller unfolded run) to 1 (longer unfolded run).
C. Distribution of adjusted PASTA scores of the entire protein set. Raw PASTA disorder and energy scores were normalized and transformed to scale from 0 (more ordered) to 1 (more disordered). The PASTA energy score was also inverted.

(~28.7%). These results that the great sensitivity at the expense of specificity for the PASTA energy equation. Given the lack of a necessity for amyloid formation for some prions and many types of aggregation, this distribution of scoring should provide a modest boost to actual amyloidogenic proteins in our data set. Combination of the various scores (FoldIndex, PASTA, PAPA, PLAAC) with weights generated a quantitative ranking from which we selected top ranking proteins to examine further with various assays.

Table 4.4. Top ranking proteins.

Protein	General Function	Protein	General Function
ABLIM3	Actin binding	KRT10	Cytoskeleton
ACBD3	Golgi structure regulation	KRT2	Cytoskeleton
ADAM29	Spermatogenesis	KRT76	Cytoskeleton
AKAP8	Anchoring / scaffolding	KRT9	Cytoskeleton
ANXA11	Phospholipid binding	MAML2	Transcription activation
ARNT	Nuclear transport	MED12	Transcription initiation
ATRX	Chromatin remodeling	MED15	Transcription initiation
ATXN1	DNA binding	MLL5	Transcription regulation
BICRAL	Chromatin remodeling	NCOA3	Nuclear receptor coactivation
BMP2K	Bone differentiation	NFYC	Transcription regulation
BRD4	Chromatin binding	PBX3	Transcription activation
CDK19	Transcription activation	PDCD6IP	Vesicle trafficking / apoptosis
CHD7	Transcription regulation	PHC1	Transcription repression
CHD9	Transcription coactivation	PRB3	Pathogen recognition
CLINT1	Vesicle trafficking	PYGO2	Signal transduction
CLOCK	Circadian regulation	QRICH1	Unknown
CNOT2	mRNA regulation	R3HDM1	RNA binding
CREST	Chromatin remodeling	R3HDM2	RNA binding
DDX17	RNA helicase	RBM14	Nuclear receptor coactivation
DLX2	Transcriptional activation	SCAF11	RNA binding
DMKN	Skin cell differentiation	SNX16	Intracellular trafficking
EYA2	Transcription coactivation	SP1	Transcription regulation
FOXA1	Transcription activation	SS18	Chromatin remodeling
FOXO1	Transcription regulation	SUPT5H	Transcription elongation
FUS	mRNA regulation	SUPT6H	Transcription elongation
HDX	DNA binding	TAF15	Transcription regulation
HNRNPA0	RNA binding	TCERG1	Transcription elongation
HNRNPA1	RNA binding	TCF20	Transcription activation
HNRNPA2B1	RNA binding	TIAL1	RNA binding
HNRNPH1	RNA binding	TNRC6C	mRNA regulation
HNRNPH2	RNA binding	TOB1	mRNA regulation
HNRNPU	RNA binding	TRERF1	Transcription regulation
HNRNPUL1	RNA binding	WNK1	Phosphorylation
HNRNPDL	RNA binding	WWTR1	Transcription coactivation
ILF3	RNA binding	YLPM1	Telomerase regulation
JMJD6	Hydroxylation / demethylation	YTHDF1	mRNA regulation
KAT6A	Histone acetylation	ZMIZ1	Transcription coactivation
KPRP	Skin cell differentiation	ZNF384	Transcription regulation

Multiple BAF complex proteins contain predicted PrLDs and aggregate in yeast

As the BAF complex was a specific focus of our investigation into protein aggregation in cancer, we noted that five BAF complex proteins contained predicted prion-like domains (PrLDs). These five proteins were the SWI1 functional homologs of ARID1A, ARID1B, and ARID2 as well as SS18 and its homolog, CREST (SS18L1). Moreover, both CREST and SS18 were among our top ranked candidates from our combinatorial method (Table 4.4). BAF53B was also selected due the presence of a couple of predicted prion-like domains buried in folded regions that could potentially be destabilized through mutations – similar to p53. The PrLDs of CREST and SS18 both span approximately three-quarters of the full-length protein (Figure 4.3A). Those regions are also predicted to be largely disordered by DISOPRED3, FoldIndex, and PASTA (Figure 4.3A, 4.3B, data not shown). The combination of predicted disorderedness, prionogenicity, and amyloidogenicity pointed towards CREST and SS18 as prime candidates for potential disease-related aggregation.

As both ARID1A and ARID1B harbored multiple predicted PrLDs and are over 2000 residues in length, these proteins were not investigated further here due to constrained feasibility and experimental limitations. The other four proteins – ARID2, BAF53B, CREST, and SS18 – were expressed in yeast as GFP fusions under a galactose-inducible promoter to examine the potential for aggregation under overexpression conditions. ARID2-EGFP and BAF53B-EGFP had poor expression though an exceptionally low proportion of cells exhibited faint fluorescent puncta (data not shown). This hampered expression likely stemmed from a combination of a lack of codon optimization, possible toxicity, and the large size of the protein in the case of ARID2. To attempt to rectify the expression issue with BAF53B, a yeast codon-optimized version of the BAF53B cDNA was synthesized. This version of the BAF53B gene robustly expressed in yeast under the

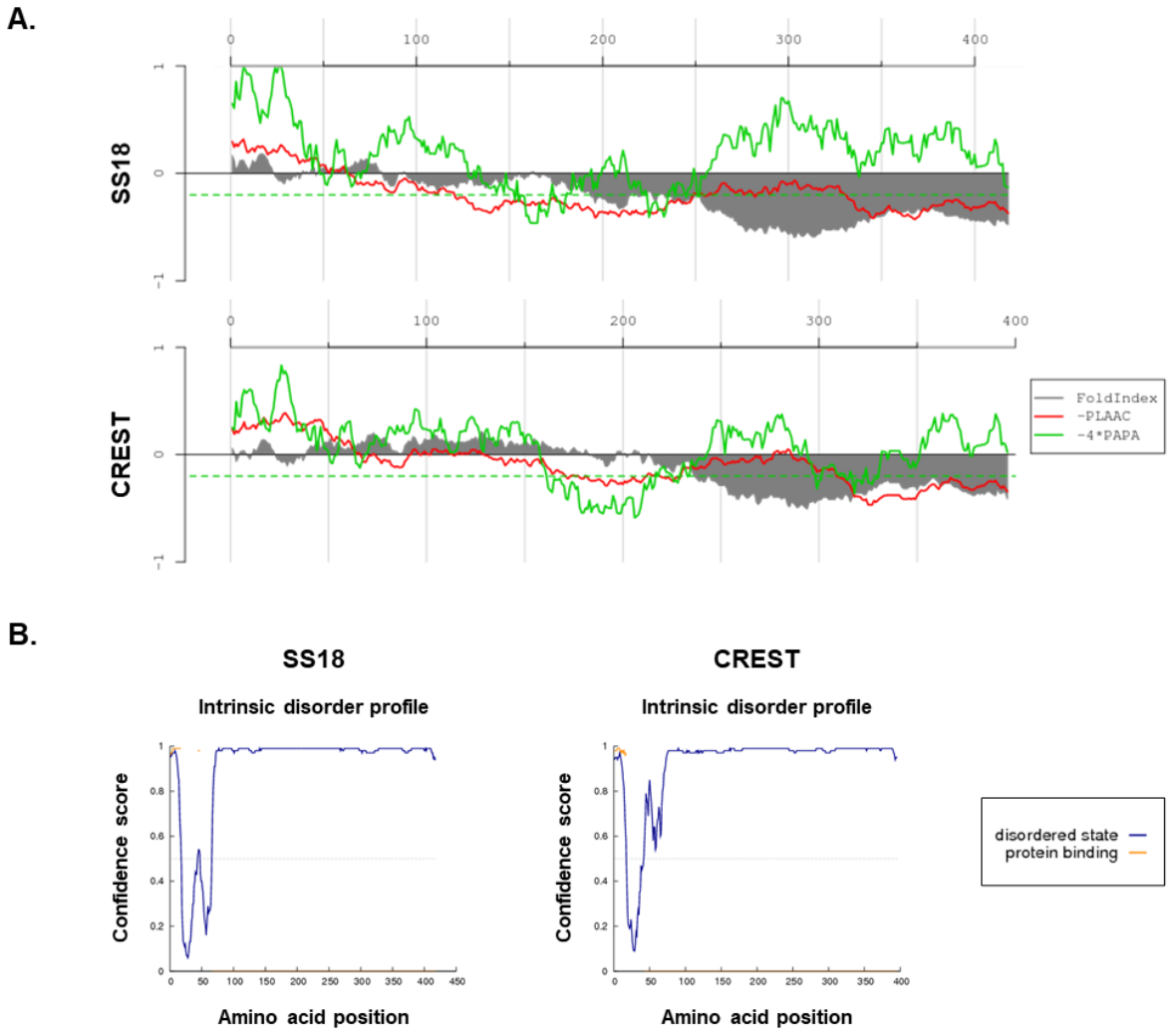


Figure 4.3. SS18 and CREST contain predicted PrLDs in their disordered regions.

A. Plots of FoldIndex, PAPA, and PLAAC analysis of the SS18 and CREST proteins. FoldIndex (gray) indicates folded and unfolded regions by positive and negative scoring per residue, respectively. PAPA (green) log-likelihood scores are plotted per residue – scores were scaled by a factor of -4 for graphing onto the shared axis and thus, lower scores are indicative of greater prionogenicity. Scaled threshold for PAPA shown in a dashed green line. PLAAC (red) log-likelihood scores are plotted per residue – scores were scaled by a factor of -1 for graphing onto the shared axis and thus, lower scores are indicative of greater prionogenicity. PLAAC analysis used a core length of 60 and the background probabilities of *Homo sapiens*. X-axis represents amino-acid position.

B. Plots of DISOPRED3 analysis of the SS18 and CREST proteins.

same, previously used galactose-inducible promoter (Figure 4.4). Most colonies exhibited fluorescent puncta in a minority of cells (<15%); however, a small proportion of examined colonies showed widespread (>50% of cells) aggregation. This distribution of observed aggregation can typically be seen when inducing the appearance of yeast prions through overexpression – indicating that BAF53B may misfold and form prion-like conformations.

We also examined a handful of BAF53B mutations that were predicted to increase scoring in predictions of either prionogenicity or aggregation propensity. These attributes were examined with PAPA and PLAAC or Zyggregator, respectively. The mutations were reported via COSMIC and/or TCGA. The five mutations consisted of P308L, P308S, R338L, P371S, and R386L – thus replacing a proline (P) or arginine (R) with the more aggregation-prone and prion-favorable serine (S) or leucine (L). We were unable to determine any significance of these mutations as no mutants behaved differently than WT BAF53B in the aggregation assays (data not shown).

Mutations of CREST and SS18 reported from samples derived from cancer patients were selected and generated for examination in the yeast aggregation model alongside wild-type (WT). For each protein, the full range of reported mutants in the COSMIC and TCGA databases were collected (57, 224). Missense mutations recorded in at least two independent cases were assessed by predictive algorithms (i.e., PAPA, PLAAC, ZipperDB, Zyggregator) to examine whether the mutations enhanced prionogenicity, amyloidogenicity, or general aggregation propensity (data not shown). For CREST, we selected mutations that existed within the predicted PrLD – A263T, A348S, and P372L. For SS18, we selected mutations where a hydrophobic leucine or polar, uncharged serine replaced a non-prionogenic, non-amyloidogenic proline residue – P117L, P168S, P252L, and P341S. All mutations were generated via PCR mutagenesis. Both WT CREST-EGFP and SS18-EGFP showed aggregates in some cells as early as 3 hours post-induction and in the

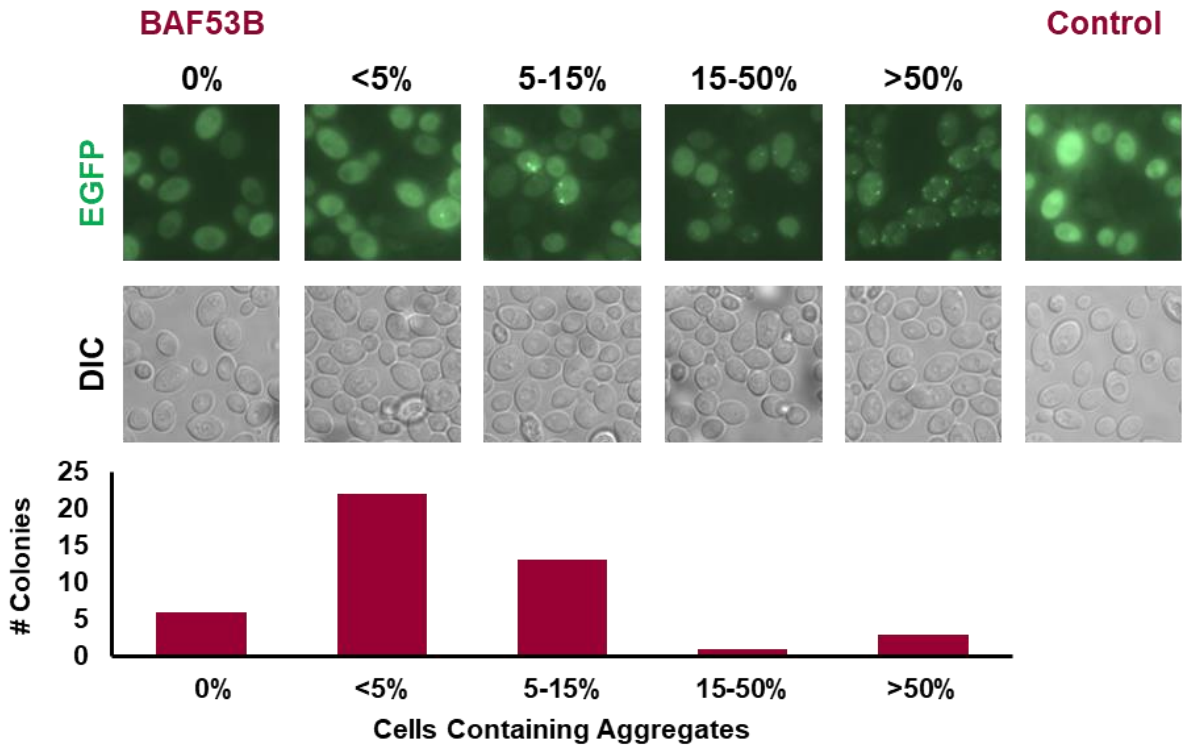


Figure 4.4. BAF53B aggregates in yeast when overexpressed.

BAF53B forms aggregates in BY4741 yeast. An EGFP-only control displays only diffuse fluorescence. Images taken at 24 hours post-induction with 2% galactose and are representative of three biological replicates.

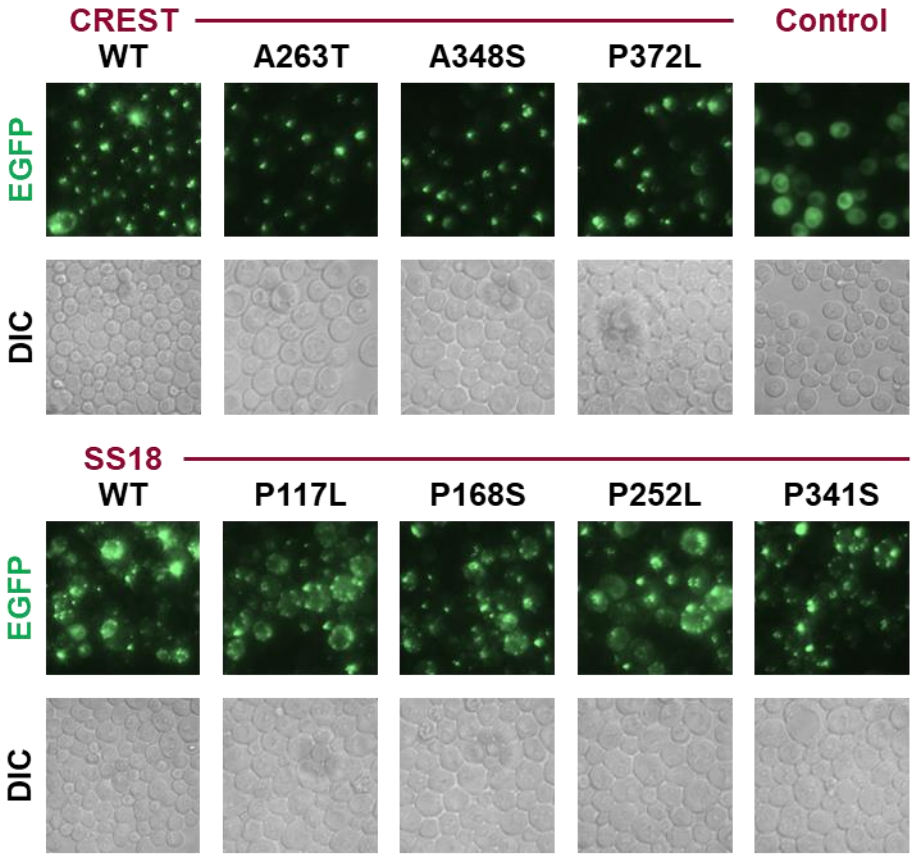


Figure 4.5. CREST and SS18 aggregate in yeast when overexpressed.

Both wild-type (WT) and mutant forms of CREST (A263T, A348S, P372L) and SS18 (P117L, P168S, P252L, P341S) form aggregates in BY4741 yeast. An EGFP-only control exhibits only diffuse fluorescence. Images taken at 24 hours post-induction with 2% galactose and are representative of three biological replicates.

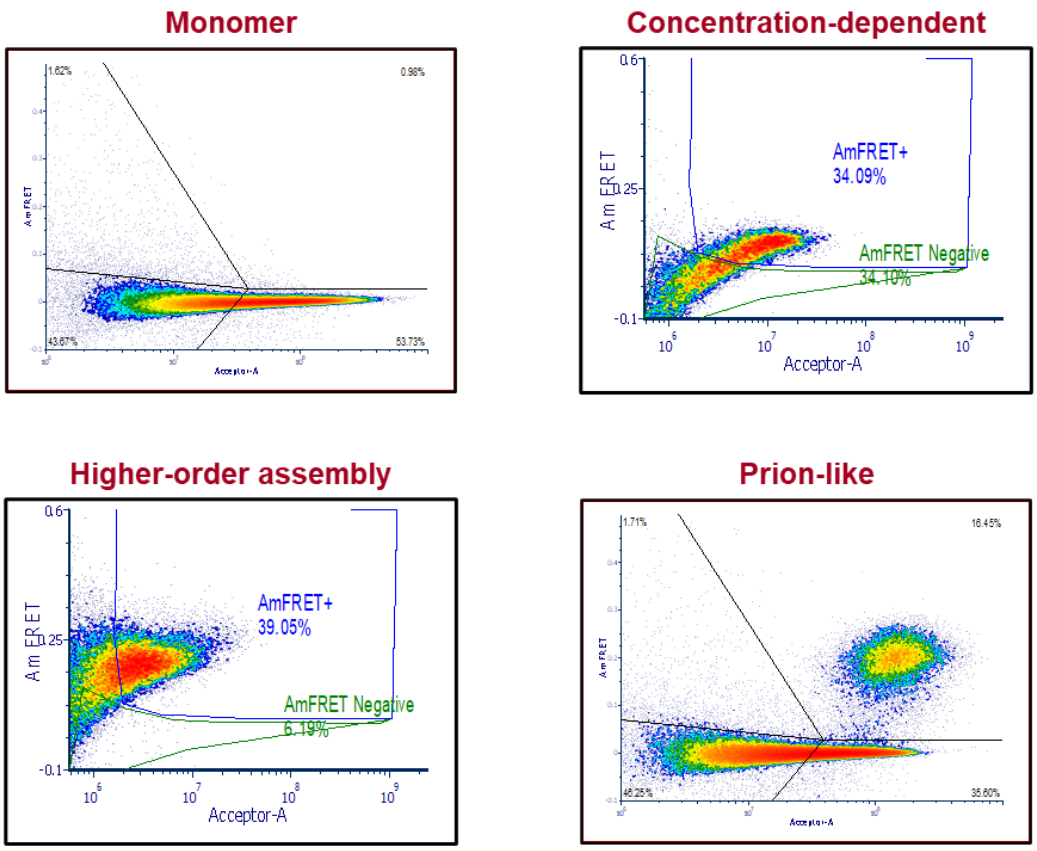
large majority of cells at 24 hours post-induction (Figure 4.5, data not shown). All mutations exhibited similar aggregation patterns to WT. As such, whether these mutations indeed increase aggregation propensity remains an open question – a more detailed kinetic study would be required. A GFP-only control showed only diffuse fluorescent signal.

Fourteen proteins display either concentration-dependent or higher-order assemblies

Moving back to the broader search for novel protein aggregators in cancer, we collaborated with the Halfmann Lab at Stowers Research Institute for initial screening of the top candidates produced by the multi-algorithmic approach. Their DAmFRET system allows for automated and high-throughput examination of phase tendencies of proteins expressed in a yeast model system (225). From the aforementioned protein ranking, the top 79 candidates were selected for experimental evaluation using DAmFRET. Some proteins that ranked high were removed and not tested due to having been previously reported to form aggregates or act in a prion-like fashion (e.g., FUS, EWS, TDP-43). Other high-ranking proteins were omitted due to their membrane-bound existence or an unavailability in the Human ORFeome collection.

Of the 79 candidate proteins examined, the majority (56, ~70.9%) showed little to no expression in the yeast model system – making them unable to be evaluated (Figure 4.6B). This issue likely resulted from a myriad of factors but chief amongst them was that the genes were not yeast-codon optimized (As will be detailed in a subsequent section, codon optimization can significantly boost expression of human proteins in yeast.) Although these proteins were not further investigated here, this experiment should not be taken as these proteins being unlikely to aggregate or act in a prion-like fashion as nothing can be concluded from the lack of data. The remaining candidates fell into three categories – proteins which expressed modestly and monomerically, proteins which demonstrated concentration-dependent aggregation, and proteins

A.



B.

Category	# of Candidates	Specific Proteins
Higher-order assembly	4	BRD4 CNOT2 JMJD6 TIAL1
Concentration-dependent	10	ABLIM3 CLINT1 CLOCK HNRNPA1 HNRNPDL HNRNPU ILF3 PDCD6IP QRICH1 SS18
Modest, monomeric expression	9	
Little expression	21	
No expression	35	

Figure 4.6. DAmFRET assay of aggregation behavior of candidate proteins.

A. Top ranked protein candidates were evaluated in collaboration with the Halfmann Lab and their DAmFRET assay. Example plots of proteins that maintain a monomeric state (top left), aggregate in a concentration-dependent manner (top right), form higher-order assemblies even at low concentrations (bottom left), or act in a prion-like manner with monomeric species in most cells and aggregates in a small subset of cells at high concentrations (bottom right). Plots for monomer and prion-like were adapted from Khan et al., 2018.
B. DAmFRET results for the top ranked protein candidates. The number of candidates falling into each category are reported. Names of proteins displaying either higher-order assembly or concentration-dependent behavior are listed in the rightmost column. No candidate proteins displayed prion-like behavior.

which exhibited the ability to form higher-order assemblies. DAmFRET plots of representative proteins from each of these categories can be seen in Figure 4.6A, which also includes an example of a plot for a protein that behaves in a prion-like fashion. The candidates that displayed only modest, monomeric expression (9, ~11.4%) were not marked for further investigation. On the other hand, we decided to continue examining those candidates that displayed either concentration-dependent aggregation or higher-order assembly. These proteins, in alphabetical order, were ABLIM3, BRD4, CLINT1, CLOCK, CNOT2, HNRNPA1, HNRNPDL, HNRNPU, ILF3, JMJD6, PDCD6IP, QRICH1, SS18, and TIAL1.

Table 4.5. Candidate PrLDs.

Protein	Region	PrLD Coordinates (aa)	PrLD Length (aa)
ABLIM3	-	1 – 77	77
BRD4	-	926 – 1180	255
CLINT1	-	494 – 612	119
CLOCK	A	561 – 689	129
CLOCK	B	742 – 828	87
CLOCK	AB	561 – 828	268
CNOT2	-	134 – 215	82
CREST	-	67 – 396	330
HNRNPA1	-	192 – 372	181
HNRNPDL	-	316 – 420	105
HNRNPU	-	690 – 825	136
ILF3	-	658 – 894	237
JMJD6	-	336 – 403	68
PDCD6IP	-	792 – 868	77
QRICH1	-	85 – 172	88
SS18	-	67 – 418	352
TIAL1	-	288 – 375	88

Candidate PrLD-Sup35$_{MC}$ fusion proteins are functional and generate Ade^{+} colonies

For each of the candidates, we identified probable prion-like domains (PrLDs) for each using the previous results of the initial bioinformatic screening (Table 4.5). The primary

determinant for the PrLDs was the results of the two algorithms that specifically predict prionogenicity, PAPA and PLAAC. If a candidate did not have a region considered prion-like by either PAPA or PLAAC, then disorderedness as predicted by DISOPRED3 was used to outline a potential PrLD. In the case of CLOCK, two regions were predicted to be prion-like thus, each region was selected separately (CLOCK PrLD A and CLOCK PrLD B) as well as a combined domain incorporating the non-prion-like region between them (CLOCK PrLD AB). In addition to the domains from the 14 candidates selected from the DAmFRET results, the CREST PrLD was also determined due to a propensity to aggregate in yeast as outlined in a later section. While CREST was tested separately using DAmFRET, low expression levels prevented the protein from being fully evaluated in that system.

For assaying of their prionogenicity, all candidate PrLDs in Table 4.4 were fused with $Sup35_{MC}$ – which consists of the middle linker region (M) and the functional region (C) of Sup35. As an endogenous yeast prion protein, Sup35 forms the [*PSI*$^{+}$] prion but normally functions as a translation terminator. In cells containing a mutation that introduces a premature stop codon in the *ADE1* gene, Sup35 recognizes said codon and prevents the translation of an enzyme essential for the synthesis of adenine. This situation results in the build-up of an adenine precursor that bestows a red color to the yeast and an inability to grow on media lacking adenine. This phenotype is reversed when Sup35 prionizes as the protein becomes aggregated, preventing recognition of the premature stop codon and leading to proper translation of *ADE1*. Thus, colonies adopt a white or pink color – depending on the prion conformation – and can grow on media lacking adenine. By replacing the Sup35 prion domain, $Sup35_{N}$ with the candidate PrLDs, we were able to use this robust phenotypic assay to test for prionogenicity of these regions.

Table 4.6. Spreading assay for Ade$^+$ colony formation.

Candidate PrLD	Number of Colonies on YPD*	Number of Colonies on -Ade†	Ade$^+$ Colonies per 10^6 cells
ABLIM3	963	18	1.87
BRD4	1054	53	5.03
CLINT1	Too many to count	Too many to count	-
CLOCK A	607	65	10.71
CLOCK B	310	182	58.71
CLOCK AB	844	248	29.38
CNOT2	911	26	2.85
CREST	96	133	138.54
HNRNPA1	984	Too many to count	-
HNRNPDL	1044	58	5.56
HNRNPU	1231	32	2.60
ILF3	802	26	3.24
JMJD6	777	Too many to count	-
PDCD6IP	663	19	2.87
QRICH1	970	99	10.21

*Total for 10^3 cells spread onto 2-3 plates. †Total for 10^7 cells spread onto an equivalent number of plates as YPD.

Upon initial transformation of W303 *sup35Δ* / *p316SUP35$_{FL}$* with the set of *p415TEF-Candidate PrLD-MC* constructs and dropout of *p316SUP35$_{FL}$*, the majority of the resulting W303 *sup35Δ* / *p415TEF-Candidate PrLD-MC* yeast displayed a robust red coloration – indicating robust functioning of the various fusion proteins. In the minority were the yeast expressing JMJD6-MC and TIAL1-MC, these isolates all displayed constant light pink coloration – indicating that the fusion protein was likely consistently aggregated (Figure 4.7A). To examine the ability of the Candidate PrLD-MC fusions to prionize and adopt an alternative conformation, we performed a spreading assay of the W303 *sup35Δ* / *p415TEF-Candidate PrLD-MC* cells onto -Ade media – as potential yeast containing prionized Candidate PrLD-MC would have the ability to grow on this media. A control spreading of yeast onto nonselective media (i.e., YPD) was also conducted. Yeast from all candidates generated colonies that could grow on the -Ade media, henceforth termed Ade$^+$

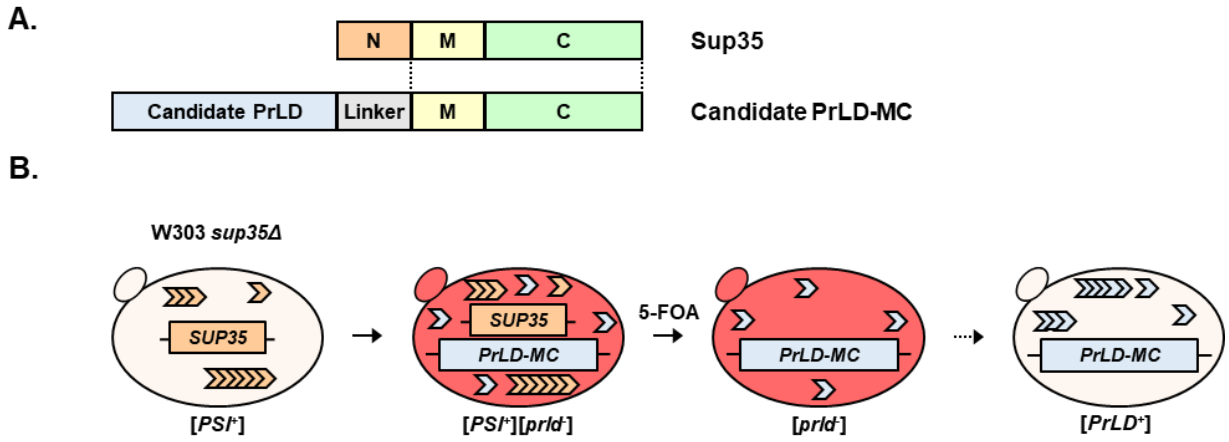


Figure 4.7. Experimental design for testing Candidate PrLDs using the Sup35$_{MC}$ *de novo* prion formation assay.

A. Diagram of fusion proteins created. Candidate PrLDs were linked to Sup35$_{MC}$ via a DPGGPGGG linker to allow for usage of the Sup35 *de novo* prion formation assay.
B. Diagram of experiment. W303 *sup35Δ* / *p316SUP35$_{FL}$* [*PSI*$^+$] cells were transformed with *p415TEF-Candidate PrLD-MC*. The transformants were treated with 5-FOA to select against cells containing the *p316SUP35$_{FL}$* plasmid. Resulting W303 *sup35Δ* / *p415PrLD-MC* cells were then checked for color change corresponding to prionization of the Candidate PrLD-MC protein.

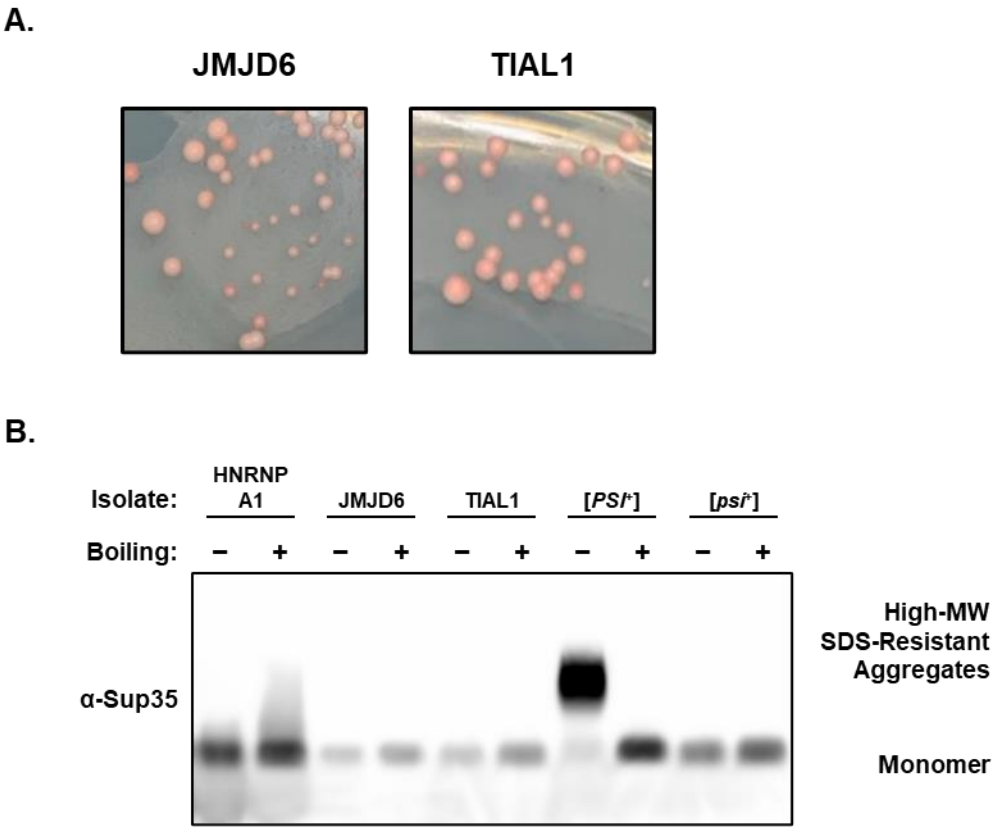


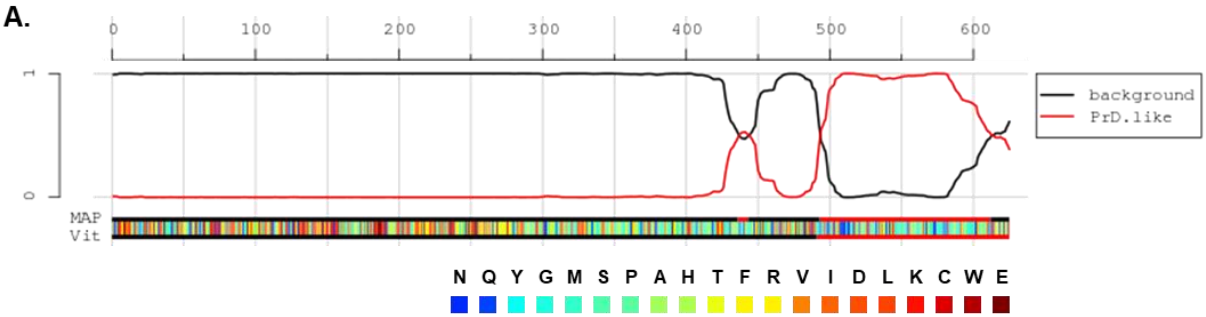
Figure 4.8. PrLDs of HNRNPA1, JMJD6, and TIAL1 do not form high molecular-weight, SDS-resistant aggregates when fused with Sup35$_{MC}$.

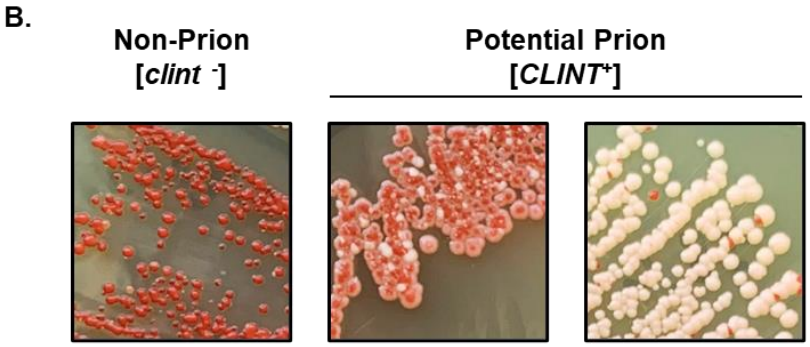
A. JMJD6-MC and TIAL1-MC do not support full translation termination function. W303 *sup35Δ* cells expressing JMJD6-MC or TIAL1-MC do not exhibit full translation termination function as assessed via the consistent light pink/white colony coloration.
B. Blot of SDD-AGE of W303 *sup35Δ* / *p415TEF-Candidate PrLD-MC* and W303 *sup35Δ* / *p316SUP35$_{FL}$* cells with (+) or without (-) boiling. For HNRNPA1-MC, JMJD6-MC, and TIAL1-MC, light-pink or white-colored yeast isolates that were also Ade$^+$ were examined. Membrane was probed with α-Sup35 to detect Candidate PrLD-MC or Sup35$_{FL}$.

colonies. (Yeast unable to grow on -Ade media are referred to as Ade$^-$.) As expected from the previously seen light pink coloration, JMJD6-MC had an uncountable (>1000) number of colonies present on the -Ade plates – further suggesting that these constructs are consistently highly aggregated. The other Candidate PrLD-MC fusions had variable numbers of Ade$^+$ colonies as seen in Table 4.6 – SS18 PrLD-MC has not yet been tested and TIAL1 PrLD-MC did not present with red colonies for usage in the assay.

CLINT1 prion-like region acts as a bona fide prion domain in yeast

The CLINT1 PrLD exists as an extreme C-terminal domain in the full-length CLINT1 protein and was predicted by PLAAC to extend from residue 494 to 612 (Figure 4.9A). Interestingly, yeast expressing CLINT1 PrLD-MC generated pink, white, or sectored (i.e., partially red, partially pink/white) colonies even on the control nonselective media. We isolated and purified these pink and white colored yeast through consecutive streaking (Figure 4.9B). In doing so, we isolated three different variants of W303 *sup35Δ* / *p415TEF-CLINT1 PrLD-MC* yeast – red (non-prion), pink (prion), and white (prion). These phenotypes were stable over many generations and curable by disruption of Hsp104 with GdnHCl – a trait common to many endogenous yeast prions. We termed the prionized form of CLINT1 PrLD-MC as [*CLINT*$^+$] and the non-prion form as [*clint1*$^-$]. To examine the aggregation status of CLINT1 PrLD-MC in these cells, we analyzed samples with SDD-AGE alongside [*PSI*$^+$] and [*psi*$^-$] controls. The pink and white [*CLINT*$^+$] isolates displayed formation of high-molecular weight (MW), SDS-resistant protein species while the red [*clint1*$^-$] isolate did not (Figure 4.9C). The [*PSI*$^+$] control displayed clear high-MW, SDS-resistant aggregates and the [*psi*$^-$] did not. Thus, the CLINT1 PrLD appears to function as PrD in this yeast model and forms SDS-resistant prion aggregates.


A.
0
100
200
300
400
500
600
1
0
background
PrD.like
MAP
Vit
N Q Y G M S P A H T F R V I D L K C W E


B.
Non-Prion
[clint -]
Potential Prion
[CLINT+]

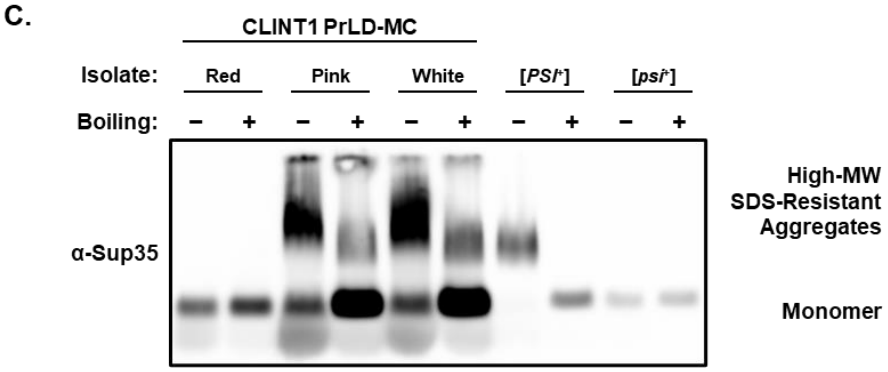
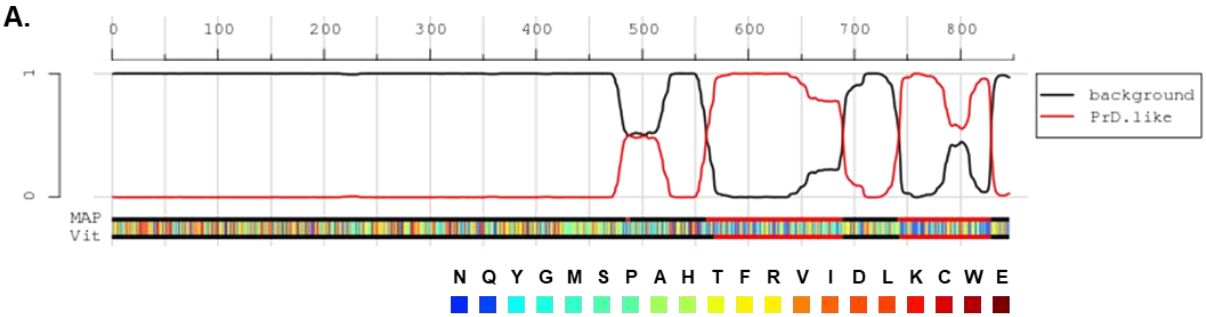

C.
CLINT1 PrLD-MC
Isolate:
Red
Pink
White
[PSI+]
[psi-]
Boiling:
– + – + – + – + – +
α-Sup35
High-MW
SDS-Resistant
Aggregates
Monomer

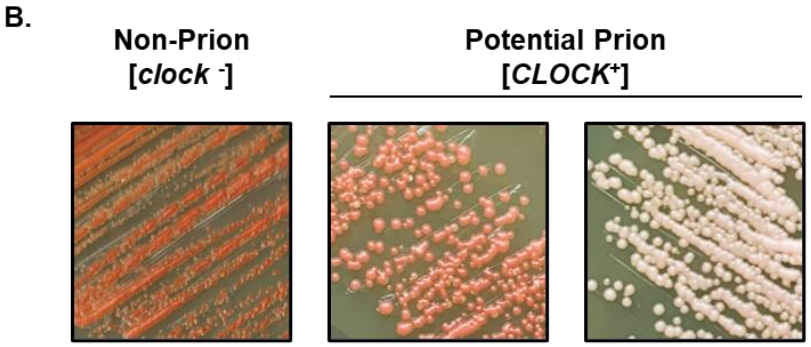
Figure 4.9. CLINT1 contains a PrLD that can *de novo* form a prion.

A. CLINT1 contains a C-terminal prion-like domain. PLAAC analysis predicts that a region stretching from 494-612 at the C-terminal of CLINT1 has amino-acid frequencies highly similar to known prion domains. PLAAC analysis used a core length of 60 and the background probabilities of *Homo sapiens*. X-axis represents amino-acid position. Amino acids are represented by colors in the bottom bar as indicated by legend.
B. CLINT PrLD-MC can *de novo* form a prion with different variants. W303 sup35 / p415TEF-CLINT1 PrLD-MC yeast can exhibit multiple phenotypes via an ade1-14 reporter. Red colonies correspond to non-prion [*clint*$^-$] cells. Pink and white colonies correspond to different prion [*CLINT*$^+$] colonies.
C. Blot of SDD-AGE of W303 *sup35Δ* / *p415TEF-CLINT1 PrLD-MC* and W303 *sup35Δ* / *p316SUP35*$_{FL}$ cells with (+) or without (-) boiling. Three isogenic isolates of yeast expressing CLINT1 PrLD-MC but with differing colony color (red, pink, white) were examined. Membrane was probed with α-Sup35 to detect CLINT1 PrLD-MC or Sup35$_{FL}$.

CLOCK prion-like region acts as a bona fide prion domain in yeast

As with the CLINT PrLD, we found that the CLOCK PrLD A could prionize in our yeast model system. Although CLOCK contains two different predicted PrLDs, we only observed prionization of the more interior PrLD stretching from residue 561 to 689 (Figure 4.10A). The aforementioned initial screening produced a number of W303 *sup35Δ* / *p415TEF-CLOCK PrLD A-MC* Ade^+ colonies. Additionally, these colonies exhibited corresponding coloration with a pink or white coloring on YPD media as compared to a red non-prion CLOCK PrLD A-MC isolate (Figure 4.10B). These phenotypes were stable across many generations. As such, we tried to cure the pink/white Ade^+ colonies by inactivating Hsp104 via treatment with GdnHCl; however, these isolates were not curable through this method (data not shown). This result could indicate that some of the CLOCK PrLD A-MC prion isolates may instead be random mutants in the reporter system used for assaying and therefore, not faithfully reporting the status of CLOCK PrLD A-MC. To this end, we evaluated the aggregation status of a pink Ade^+ isolate and a white Ade^+ using SDD-AGE. A red non-prion CLOCK PrLD A-MC isolate and control [PSI^+] and [psi^+] yeast were also included. The ensuing blot showed that the pink CLOCK PrLD A-MC isolates harbored high-MW, SDS-resistant protein species whereas the red and white isolates did not (Figure 4.10C). This result indicates that some white candidates may indeed be off-target mutants and not prionized. However, the pink Ade^+ isolates do appear to contain a prion form of CLOCK PrLD A-MC, [$CLOCK^+$] while the red Ade^- isolates are non-prion ([$clock^-$]). The control [PSI^+] and [psi^-] samples exhibited the expected pattern – high-MW, SDS-resistant band and no such band, respectively. Given this result, we find that the CLOCK PrLD A-MC fusion protein forms a prion – termed here [$CLOCK^+$] – in yeast that exhibits high-MW, SDS-resistant aggregates but cannot be cured through Hsp104 inactivation.


A.
0
100
200
300
400
500
600
700
800
1
0
background
PrD.like
MAP
Vit
N Q Y G M S P A H T F R V I D L K C W E


B.
Non-Prion
[clock -]
Potential Prion
[CLOCK+]

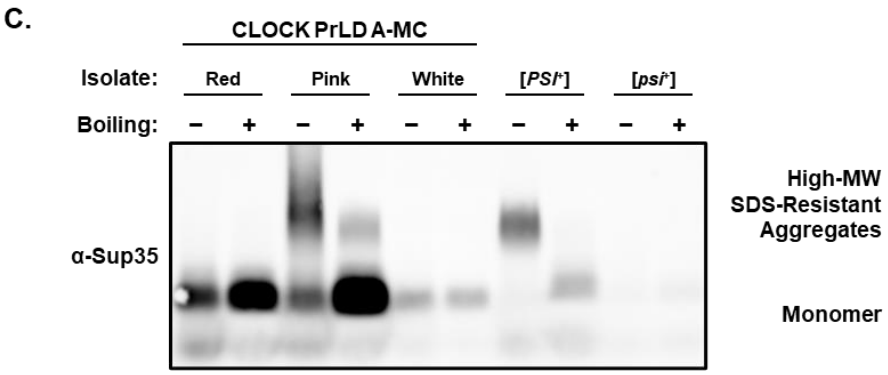

C.
CLOCK PrLD A-MC
Isolate:
Red
Pink
White
[PSI+]
[psi-]
Boiling:
– + – + – + – + – +
α-Sup35
High-MW
SDS-Resistant
Aggregates
Monomer

Figure 4.10. CLOCK contains a PrLD that can *de novo* form a prion.

A. CLOCK contains two C-terminal prion-like domains. PLAAC analysis predicts that two regions stretching from 561-689 (PrLD A) and 742-828 (PrLD B) at the C-terminal of CLOCK have amino-acid frequencies highly similar to known prion domains. PLAAC analysis used a core length of 60 and the background probabilities of *Homo sapiens*. X-axis represents amino-acid position. Amino acids are represented by colors in the bottom bar as indicated by legend.
B. CLOCK PrLD A-MC can *de novo* form a prion with different variants. W303 *sup35Δ* / *p415TEF-CLOCK PrLD A-MC* yeast can exhibit multiple phenotypes via an ade1-14 reporter. Red colonies correspond to non-prion [*clock*$^{-}$] cells. Pink and white colonies correspond to potential prion [*CLOCK*$^{+}$] colonies.
C. Blot of SDD-AGE of W303 *sup35Δ* / *p415TEF-CLOCK PrLD A-MC* and W303 *sup35Δ* / *p316SUP35*$_{FL}$ cells with (+) or without (-) boiling. Three isogenic isolates of yeast expressing CLOCK PrLD A-MC but with differing colony color (red, pink, white) were examined. Membrane was probed with α-Sup35 to detect CLOCK PrLD A-MC or Sup35$_{FL}$.

Discussion

In this study, we have used a bioinformatically guided process to test for novel protein aggregators involved in cancers. A set of cancer-related proteins were scored and ranked based on key characteristics for protein aggregation. These characteristics included disorderedness as predicted by DISOPRED3 and FoldIndex, prionogenicity as predicted by PAPA and PLAAC, and general and amyloid aggregation propensity as predicted by PASTA, ZipperDB, and Zyggregator. Due to our lab's focus on the yeast SWI/SNF complex, the proteins of the human SWI/SNF or BAF complex were of particular interest. From a set of cancer-related proteins, we collaborated with the Halfmann lab to evaluate the 79 top-scoring candidates using a DAmFRET-based assay for concentration-based aggregation patterns. Proteins displaying aggregation of any sort in that assay were then experimentally tested for prionogenicity using the widely used yeast-based Sup35 *de novo* prion formation assay. We found that a predicted PrLD in the clathrin interacting protein CLINT1 and a predicted PrLD in the circadian rhythm protein CLOCK both can act as bona fide prion domains – forming [*CLINT1*$^{+}$] and [*CLOCK*$^{+}$].

CLINT1 (also known as epsinR) was first identified through its homology to other epsin proteins (226). CLINT1 contains an epsin-N-terminal-homology (ENTH) domain which allows for binding to cargo proteins and specific lipids and appears to be highly structured (227, 228). However, the majority of the protein appears to be relatively unstructured even though it contains a middle domain for binding to clathrin and adaptor protein-1 (AP-1). It is in this unstructured region and at the C-terminal of the protein that the CLINT1 PrLD resides. Although the function of this region has not been determined, there exists limited evidence of additional clathrin binding sites residing there (227). The value of a PrLD in normal cellular functioning has been observed

with other proteins including the human MAVS (mitochondrial antiviral-signaling protein) and ASC (apoptosis-associated speck-like protein containing a caspase activation and recruitment domain) in antiviral responses and mammalian CPEB3 (cytoplasmic polyadenylation element-binding protein 3) in long-term facilitation of synapses (80, 81, 88, 89). Perhaps the CLINT1 PrLD acts in important protein-recruiting functions – whether that is recruitment of cargo, adaptors, or other proteins.

Normally, CLINT1 specifically acts in trafficking cargo between the trans-Golgi network (TGN) and endosomes (228–230). Polymorphisms in CLINT1 have been associated with disruptions in schizophrenia in humans (231, 232). Could possible mutations destabilize the normal CLINT1 conformation and lead to aberrant misfolding and prion-like aggregation (as observed with p53)? Any such aggregation activity likely would have a significant effect on neurologic functioning given the importance of vesicle trafficking in neurons. Examination of full-length CLINT1 for prion-like aggregation in a human cell line would be valuable and could assist in elucidating the function of the CLINT1 PrLD.

The driving of circadian rhythms by molecular clocks plays a critical role in maintaining organismal homeostasis. As the first identified mammalian circadian gene, CLOCK acts in a central role of this molecular clock through its interactions with and regulation of other circadian proteins and genes (e.g., BMAL1, CRY, and PER) (233). The dimerization of CLOCK and BMAL1 target the CRY and PER genes, which in turn, repress the CLOCK and BMAL1 dimer. Disruptions of circadian genes play an important role in a range of diseases – including cancer, neurodegeneration, obesity, and of course, sleep disorders (234). In cancer, disturbances of circadian rhythms can increase susceptibility to liver carcinomas and lung tumors (235, 236). Thusly, CLOCK was an intriguing candidate for our investigation.

We found that the CLOCK protein harbors two predicted PrLDs, which we termed CLOCK PrLD A and CLOCK PrLD B. In examining these two PrLDs in the Sup35 *de novo* prion formation assay, we observed the formation of yeast with pink, Ade^{+} phenotypes reflective of CLOCK PrLD A-MC prionization. We did not find phenotypically stable yeast indicating the prionization of CLOCK PrLD B-MC or the larger combination of both PrLDs and the intervening region, CLOCK PrLD AB-MC. However, the lack of susceptibility to Hsp104 inactivation by [*CLOCK*$^{+}$] formed by CLOCK PrLD A-MC may suggest that delineating between mutants that phenotypically mimic prionization and actual protein conformational changes is a limiting factor of screening prionization of these three fusion proteins. Although many endogenous yeast prions are curable by millimolar concentrations of GdnHCl due to Hsp104 inactivation, some yeast prions cannot be cured by such a treatment and instead can be cured by disruption of other molecular chaperones (e.g., Hsp70s, Hsp90s) (145). Given this knowledge, we intend to examine the possibility that [*CLOCK*$^{+}$] formed by CLOCK PrLD A-MC can be cured by expression of a dominant-negative Hsp70 (Ssa1 K69M) or treatment with the Hsp90-disrupting radicicol (206, 237, 238). The finding that CLOCK PrLD A-MC indeed exists in two different states (monomeric or high-MW, SDS-resistant species) whether in red [*clock*$^{-}$] cells or pink [*CLOCK*$^{+}$] cells, respectively as seen via SDD-AGE supports the existence of an alternative, aggregated prion state for CLOCK PrLD A-MC. Additional study will have to further confirm the transmissibility of this putative prion state to naïve cells.

The two PrLDs in CLOCK (PrLD A and PrLD B) both reside in the second half of the protein and contain an enrichment of glutamine (Q) residues. While the N-terminal region of CLOCK contains a bHLH-PAS (basic helix-loop-helix and PER-ARNT-SIM) domain that is typically found in transcriptional regulators, little is known about the C-terminal region other than

its importance for transactivation (239, 240). Given the presence of two PrLDs within this region, there may be significance in their composition for protein-protein interactions between CLOCK and BMAL1 or other circadian proteins. Disruption of these domains may then lead to aberrant circadian rhythms and contribute towards human disease.

Six proteins involved in the BAF complex were predicted to contain PrLDs. Although we were not able to effectively express the larger proteins of this group – ARID1A, ARID1B, ARID2 – in a yeast model, the other three proteins – CREST, BAF53B, SS18 – demonstrated aggregation phenotypes. Intriguingly, BAF53B exhibited a skewed distribution of aggregation frequencies across different colonies – indicating a possible prionogenic nature. CREST and SS18 demonstrated robust aggregation under overexpression conditions but no differences between WT and mutant versions. We attempted to test the prionogenicity of CREST and SS18 with the Sup35 de novo prion formation assay; however, both CREST-MC and SS18-MC were unable to provide a stable non-prion phenotype (data not shown). This issue may have been due to localization to the nucleus (in the case of CREST-MC) or consistently aggregated proteins preventing the assumption of a non-prion state.

We intend to move forward with this research into novel protein aggregators associated with cancer by moving into a human cell culture model with our collaborators in the laboratory of Dr. Marc Mendillo. Already, we have generated lentiviral expression clones for the various Candidate PrLDs fused to a 3xFLAG.V5 tag. In fact, initial experimentation has revealed that the some of the Candidate PrLDs can form high-MW, SDS-resistant aggregates in HEK293T cells (Figure 4.11A). For this experiment, Dr. Milad Alasady and Jasen Jackson of the Mendillo lab produced the stable cell lines expressing the Candidate PrLDs fused to 3xFLAG.V5 and the 0.02% SDS condition mirrored conditions previously used in investigating tau aggregate strains (17). We

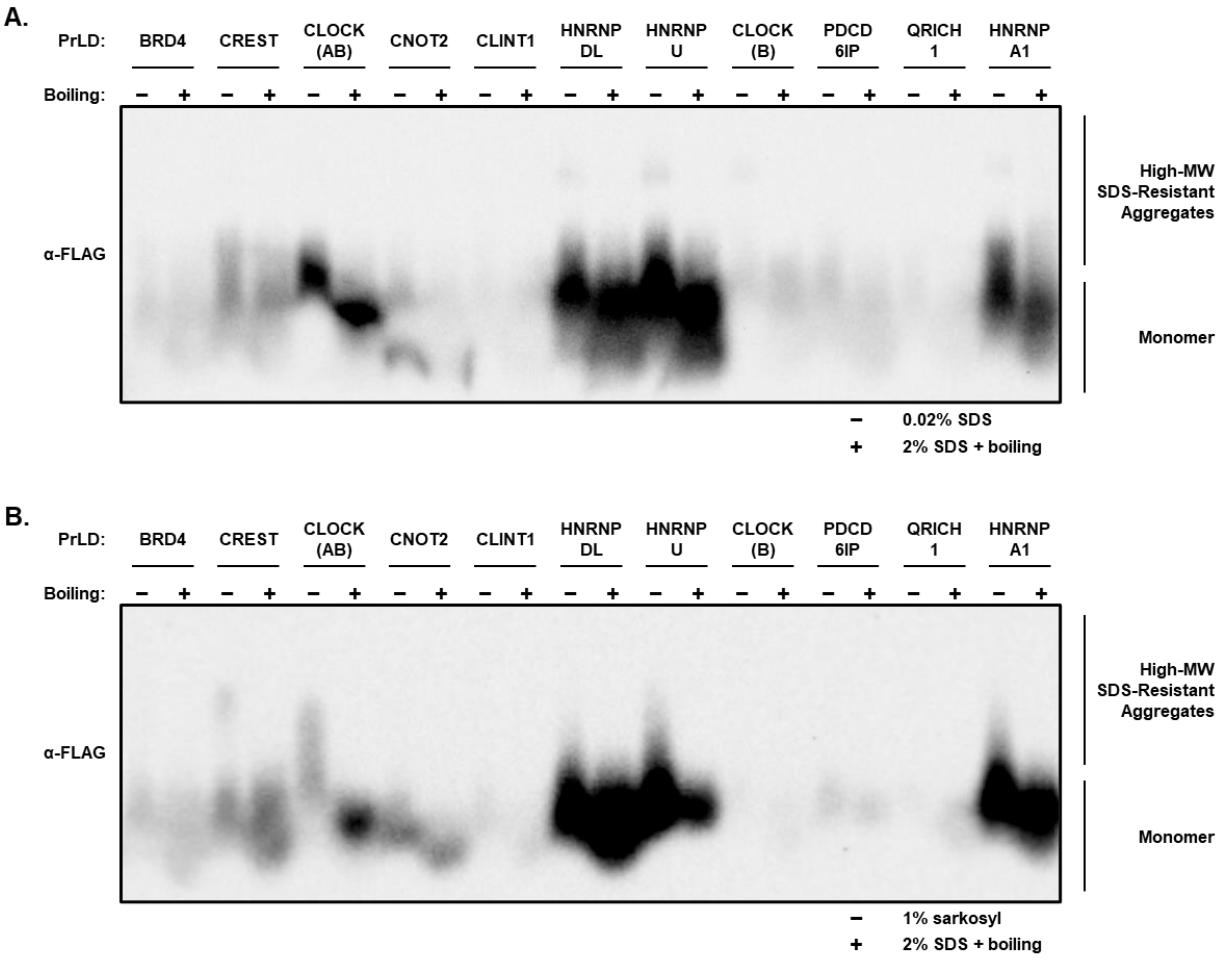


Figure 4.11. Candidate PrLDs can form high-molecular weight, SDS or sarkosyl-resistant aggregates in HEK293T cells.

A. Blot of SDD-AGE of Candidate PrLD-3xFLAG.V5 fusions expressed in HEK293T cells. Samples were incubated with 0.02% SDS (-) at room temperature or with 2% SDS and boiling (+). Membrane was probed with α-FLAG to detect Candidate PrLD-3xFLAG.V5.

B. Blot of SDD-AGE of Candidate PrLD-3xFLAG.V5 fusions expressed in HEK293T cells. Samples were incubated with 1% sarkosyl (-) at room temperature or with 2% SDS and boiling (+). Membrane was probed with α-FLAG to detect Candidate PrLD-3xFLAG.V5.

also saw similar high-MW aggregates formed by some Candidate PrLDs when 1% sarkosyl was used as the detergent condition (Figure 4.11B). We plan to further investigate the promising candidates – including both the PrLDs alone and the full-length proteins – in HEK293T and other appropriate human cell lines. In regard to the CLINT1 PrLD and CLOCK PrLD A, both proteins may be susceptible to study either via vesicular assays or established reporter systems, respectively (241–243). Additionally, the possibility of infection of cultured human cells with CLINT1 PrLD and CLOCK PrLD A seeds from [*CLINT*$^{+}$] and [*CLOCK*$^{+}$] yeast exists as a promising route of investigation.

Given a starting set of over 9000 proteins and a bioinformatically-informed selection process, we have identified at least two disease-related proteins with prionogenic regions that appear to function as PrDs in a yeast-based assay. Additionally, at least three BAF complex proteins display significant aggregation phenotypes in yeast; although, mutant forms of these proteins do not appear to differ in aggregation from the WT versions. These results indicate that our understanding of the role of protein aggregators in disease including cancers may still be limited. Future research must focus on determining the extent of the aggregation abilities of the CLINT1 PrLD and the CLOCK PrLDs and the functional importance of these domains in a human cell line model.

REFERENCES

1. Prusiner SB. 1982. Novel proteinaceous infectious particles cause scrapie. Science 216:136–144.
2. Long JM, Holtzman DM. 2019. Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. Cell 179:312–339.
3. Poewe W, Seppi K, Tanner CM, Halliday GM, Brundin P, Volkmann J, Schrag AE, Lang AE. 2017. Parkinson disease. Nat Rev Dis Prim 3:1–21.
4. Hardiman O, Al-Chalabi A, Chio A, Corr EM, Logroscino G, Robberecht W, Shaw PJ, Simmons Z, van den Berg LH. 2017. Amyotrophic lateral sclerosis. Nat Rev Dis Prim 3:17071.
5. Kane MD, Lipinski WJ, Callahan MJ, Bian F, Durham RA, Schwarz RD, Roher AE, Walker LC. 2000. Evidence for seeding of beta -amyloid by intracerebral infusion of Alzheimer brain extracts in beta -amyloid precursor protein-transgenic mice. J Neurosci 20:3606–11.
6. Meyer-Luehmann M, Coomaraswamy J, Bolmont T, Kaeser S, Schaefer C, Kilger E, Neuenschwander A, Abramowski D, Frey P, Jaton AL, Vigouret J-M, Paganetti P, Walsh DM, Mathews PM, Ghiso J, Staufenbiel M, Walker LC, Jucker M. 2006. Exogenous induction of cerebral beta-amyloidogenesis is governed by agent and host. Science 313:1781–4.
7. Eisenberg D, Jucker M. 2012. The amyloid state of proteins in human diseases. Cell 148:1188–1203.
8. Braak H, Braak E. 1995. Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol Aging 16:271–8; discussion 278-84.

9. Walker LC, Schelle J, Jucker M. 2016. The Prion-Like Properties of Amyloid-β Assemblies: Implications for Alzheimer's Disease. Cold Spring Harb Perspect Med 6.

10. Götz J, Halliday G, Nisbet RM. 2019. Molecular Pathogenesis of the Tauopathies. Annu Rev Pathol Mech Dis 14:239–261.

11. Orr ME, Sullivan AC, Frost B. 2017. A Brief Overview of Tauopathy: Causes, Consequences, and Therapeutic Strategies. Trends Pharmacol Sci 38:637–648.

12. Clavaguera F, Akatsu H, Fraser G, Crowther RA, Frank S, Hench J, Probst A, Winkler DT, Reichwald J, Staufenbiel M, Ghetti B, Goedert M, Tolnay M. 2013. Brain homogenates from human tauopathies induce tau inclusions in mouse brain. Proc Natl Acad Sci U S A 110:9535–9540.

13. Clavaguera F, Bolmont T, Crowther RA, Abramowski D, Frank S, Probst A, Fraser G, Stalder AK, Beibel M, Staufenbiel M, Jucker M, Goedert M, Tolnay M. 2009. Transmission and spreading of tauopathy in transgenic mouse brain. Nat Cell Biol 11:909–913.

14. Holmes BB, Furman JL, Mahan TE, Yamasaki TR, Mirbaha H, Eades WC, Belaygorod L, Cairns NJ, Holtzman DM, Diamond MI. 2014. Proteopathic tau seeding predicts tauopathy in vivo. Proc Natl Acad Sci U S A 111:E4376–E4385.

15. Liu L, Drouet V, Wu JW, Witter MP, Small SA, Clelland C, Duff K. 2012. Trans-synaptic spread of tau pathology in vivo. PLoS One 7:1–9.

16. Clavaguera F, Hench J, Goedert M, Tolnay M. 2015. Invited review: Prion-like transmission and spreading of tau pathology. Neuropathol Appl Neurobiol 41:47–58.

17. Sanders DW, Kaufman SK, DeVos SL, Sharma AM, Mirbaha H, Li A, Barker SJ, Foley AC, Thorpe JR, Serpell LC, Miller TM, Grinberg LT, Seeley WW, Diamond MI. 2014. Distinct tau prion strains propagate in cells and mice and define different tauopathies.

Neuron 82:1271–1288.

18. Wischik CM, Novak M, Thogersen HC, Edwards PC, Runswick MJ, Jakes R, Walker JE, Milstein C, Roth M, Klug A. 1988. Isolation of a fragment of tau derived from the core of the paired helical filament of Alzheimer disease. Proc Natl Acad Sci U S A 85:4506–4510.

19. Conway KA, Harper JD, Lansbury PT. 2000. Fibrils formed in vitro from α-synuclein and two mutant forms linked to Parkinson's disease are typical amyloid. Biochemistry 39:2552–2563.

20. Guo JL, Covell DJ, Daniels JP, Iba M, Stieber A, Zhang B, Riddle DM, Kwong LK, Xu Y, Trojanowski JQ, Lee VMY. 2013. Distinct α-synuclein strains differentially promote tau inclusions in neurons. Cell 154:103–117.

21. Luk KC, Kehm VM, Zhang B, O'Brien P, Trojanowski JQ, Lee VMY. 2012. Intracerebral inoculation of pathological α-synuclein initiates a rapidly progressive neurodegenerative α-synucleinopathy in mice. J Exp Med 209:975–988.

22. Mougenot AL, Nicot S, Bencsik A, Morignat E, Verchère J, Lakhdar L, Legastelois S, Baron T. 2012. Prion-like acceleration of a synucleinopathy in a transgenic mouse model. Neurobiol Aging 33:2225–2228.

23. Valdinocci D, Radford RAW, Siow SM, Chung RS, Pountney DL. 2017. Potential modes of intercellular α-synuclein transmission. Int J Mol Sci 18.

24. Braak H, Del Tredici K, Rüb U, De Vos RAI, Jansen Steur ENH, Braak E. 2003. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging 24:197–211.

25. Kim S, Kwon SH, Kam TI, Panicker N, Karuppagounder SS, Lee S, Lee JH, Kim WR, Kook M, Foss CA, Shen C, Lee H, Kulkarni S, Pasricha PJ, Lee G, Pomper MG, Dawson VL, Dawson TM, Ko HS. 2019. Transneuronal Propagation of Pathologic α-Synuclein from

the Gut to the Brain Models Parkinson's Disease. Neuron 103:627-641.e7.

26. Lohmann S, Bernis ME, Tachu BJ, Ziemski A, Grigoletto J, Tamgüney G. 2019. Oral and intravenous transmission of α-synuclein fibrils to mice. Acta Neuropathol 138:515–533.

27. McAlary L, Plotkin SS, Yerbury JJ, Cashman NR. 2019. Prion-Like Propagation of Protein Misfolding and Aggregation in Amyotrophic Lateral Sclerosis. Front Mol Neurosci 12:1–21.

28. King OD, Gitler AD, Shorter J. 2012. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain Res 1462:61–80.

29. Gitler AD, Shorter J. 2011. RNA-binding proteins with prion-like domains in ALS and FTLD-U. Prion 5:179–187.

30. Kim HJ, Kim NC, Wang Y-D, Scarborough EA, Moore J, Diaz Z, MacLea KS, Freibaum B, Li S, Molliex A, Kanagaraj AP, Carter R, Boylan KB, Wojtas AM, Rademakers R, Pinkus JL, Greenberg SA, Trojanowski JQ, Traynor BJ, Smith BN, Topp S, Gkazi A-S, Miller J, Shaw CE, Kottlors M, Kirschner J, Pestronk A, Li YR, Ford AF, Gitler AD, Benatar M, King OD, Kimonis VE, Ross ED, Weihl CC, Shorter J, Taylor JP. 2013. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature 495:467–73.

31. Kato M, Han TWW, Xie S, Shi K, Du X, Wu LCC, Mirzaei H, Goldsmith EJJ, Longgood J, Pei J, Grishin NV V., Frantz DEE, Schneider JWW, Chen S, Li L, Sawaya MRR, Eisenberg D, Tycko R, McKnight SLL. 2012. Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. Cell 149:753–767.

32. Fushimi K, Long C, Jayaram N, Chen X, Li L, Wu JY. 2011. Expression of human FUS/TLS in yeast leads to protein aggregation and cytotoxicity, recapitulating key features

of FUS proteinopathy. Protein Cell 2:141–149.

33. Nomura T, Watanabe S, Kaneko K, Yamanaka K, Nukina N, Furukawa Y. 2014. Intranuclear aggregation of mutant FUS/TLS as a molecular pathomechanism of amyotrophic lateral sclerosis. J Biol Chem 289:1192–1202.

34. Qamar S, Wang GZ, Randle SJ, Ruggeri FS, Varela JA, Lin JQ, Phillips EC, Miyashita A, Williams D, Ströhl F, Meadows W, Ferry R, Dardov VJ, Tartaglia GG, Farrer LA, Kaminski Schierle GS, Kaminski CF, Holt CE, Fraser PE, Schmitt-Ulms G, Klenerman D, Knowles T, Vendruscolo M, St George-Hyslop P. 2018. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. Cell 173:720-734.e15.

35. Gasset-Rosa F, Lu S, Yu H, Chen C, Melamed Z, Guo L, Shorter J, Da Cruz S, Cleveland DW. 2019. Cytoplasmic TDP-43 De-mixing Independent of Stress Granules Drives Inhibition of Nuclear Import, Loss of Nuclear TDP-43, and Cell Death. Neuron 102:339-357.e7.

36. Chen AKH, Lin RYY, Hsieh EZJ, Tu PH, Chen RPY, Liao TY, Chen W, Wang CH, Huang JJT. 2010. Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. J Am Chem Soc 132:1186–1187.

37. Guo W, Chen Y, Zhou X, Kar A, Ray P, Chen X, Rao EJ, Yang M, Ye H, Zhu L, Liu J, Xu M, Yang Y, Wang C, Zhang D, Bigio EH, Mesulam M, Shen Y, Xu Q, Fushimi K, Wu JY. 2011. An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. Nat Struct Mol Biol 18:822–831.

38. Furukawa Y, Kaneko K, Watanabe S, Yamanaka K, Nukina N. 2011. A seeding reaction recapitulates intracellular formation of sarkosyl-insoluble transactivation response element

(TAR) DNA-binding protein-43 inclusions. J Biol Chem 286:18664–18672.

39. Nonaka T, Masuda-Suzukake M, Arai T, Hasegawa Y, Akatsu H, Obi T, Yoshida M, Murayama S, Mann DMA, Akiyama H, Hasegawa M. 2013. Prion-like Properties of Pathological TDP-43 Aggregates from Diseased Brains. Cell Rep 4:124–134.

40. Porta S, Xu Y, Restrepo CR, Kwong LK, Zhang B, Brown HJ, Lee EB, Trojanowski JQ, Lee VMY. 2018. Patient-derived frontotemporal lobar degeneration brain extracts induce formation and spreading of TDP-43 pathology in vivo. Nat Commun 9.

41. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, Donaldson D, Goto J, O'Regan JP, Deng HX. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature 362:59–62.

42. Bosco DA, Morfini G, Karabacak NM, Song Y, Gros-Louis F, Pasinelli P, Goolsby H, Fontaine BA, Lemay N, McKenna-Yasek D, Frosch MP, Agar JN, Julien JP, Brady ST, Brown RH. 2010. Wild-type and mutant SOD1 share an aberrant conformation and a common pathogenic pathway in ALS. Nat Neurosci 13:1396–1403.

43. Forsberg K, Jonsson PA, Andersen PM, Bergemalm D, Graffmo KS, Hultdin M, Jacobsson J, Rosquist R, Marklund SL, Brännström T. 2010. Novel antibodies reveal inclusions containing non-native SOD1 in sporadic ALS patients. PLoS One 5:1–9.

44. Chattopadhyay M, Durazo A, Se HS, Strong CD, Gralla EB, Whitelegge JP, Valentine JS. 2008. Initiation and elongation in fibrillation of ALS-linked superoxide dismutase. Proc Natl Acad Sci U S A 105:18663–18668.

45. Münch C, O'Brien J, Bertolotti A. 2011. Prion-like propagation of mutant superoxide dismutase-1 misfolding in neuronal cells. Proc Natl Acad Sci U S A 108:3548–3553.

46. Grad LI, Guest WC, Yanai A, Pokrishevsky E, O'Neill MA, Gibbs E, Semenchenko V,

Yousefi M, Wishart DS, Plotkin SS, Cashman NR. 2011. Intermolecular transmission of superoxide dismutase 1 misfolding in living cells. Proc Natl Acad Sci U S A 108:16398–16403.

47. Ayers JI, Fromholt S, Koch M, DeBosier A, McMahon B, Xu G, Borchelt DR. 2014. Experimental transmissibility of mutant SOD1 motor neuron disease. Acta Neuropathol 128:791–803.

48. Ayers JI, Diamond J, Sari A, Fromholt S, Galaleldeen A, Ostrow LW, Glass JD, Hart PJ, Borchelt DR. 2016. Distinct conformers of transmissible misfolded SOD1 distinguish human SOD1-FALS from other forms of familial and sporadic ALS. Acta Neuropathol https://doi.org/10.1007/s00401-016-1623-4.

49. Driver JA, Beiser A, Au R, Kreger BE, Splansky GL, Kurth T, Kiel DP, Lu KP, Seshadri S, Wolf PA. 2012. Inverse association between cancer and Alzheimer's disease: Results from the Framingham Heart Study. BMJ 344:19.

50. Musicco M, Adorni F, Di Santo S, Prinelli F, Pettenati C, Caltagirone C, Palmer K, Russo A. 2013. Inverse occurrence of cancer and Alzheimer disease: A population-based incidence study. Neurology 81:322–328.

51. Aramillo Irizar P, Schäuble S, Esser D, Groth M, Frahm C, Priebe S, Baumgart M, Hartmann N, Marthandan S, Menzel U, Müller J, Schmidt S, Ast V, Caliebe A, König R, Krawczak M, Ristow M, Schuster S, Cellerino A, Diekmann S, Englert C, Hemmerich P, Sühnel J, Guthke R, Witte OW, Platzer M, Ruppin E, Kaleta C. 2018. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. Nat Commun 9:1–11.

52. Ibáñez K, Boullosa C, Tabarés-Seisdedos R, Baudot A, Valencia A. 2014. Molecular

Evidence for the Inverse Comorbidity between Central Nervous System Disorders and Cancers Detected by Transcriptomic Meta-analyses. PLoS Genet 10:1–7.

53. Kurtishi A, Rosen B, Patil KS, Alves GW, Møller SG. 2019. Cellular Proteostasis in Neurodegeneration. Mol Neurobiol 56:3676–3689.

54. Stratton MR, Campbell PJ, Futreal PA. 2009. The cancer genome. Nature 458:719–724.

55. Seo J, Park M. 2020. Molecular crosstalk between cancer and neurodegenerative diseases. Cell Mol Life Sci 77:2659–2680.

56. Costa DCF, Oliveira GAP De, Cino EA, Soares IN, Rangel LP, Silva JL, de Oliveira GAP, Cino EA, Soares IN, Rangel LP, Silva JL. 2016. Aggregation and Prion-Like Properties of Misfolded Tumor Suppressors: Is Cancer a Prion Disease ? Cold Spring Harb Perspect Biol 8:1–22.

57. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MDM, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ, Ding L. 2013. Mutational landscape and significance across 12 major cancer types. Nature 502:333–339.

58. Baugh EH, Ke H, Levine AJ, Bonneau RA, Chan CS. 2018. Why are there hotspot mutations in the TP53 gene in human cancers? Cell Death Differ 25:154–160.

59. Freed-Pastor WA, Prives C. 2012. Mutant p53: one name, many proteins. Genes Dev 26:1268–86.

60. Milner J, Medcalf EA. 1991. Cotranslation of activated mutant p53 with wild type drives the wild-type p53 protein into the mutant conformation. Cell 65:765–774.

61. Gualberto A, Aldape K, Kozakiewicz K, Tlsty TD. 1998. An oncogenic form of p53 confers a dominant, gain-of-function phenotype that disrupts spindle checkpoint control. Proc Natl

Acad Sci U S A 95:5166–5171.

62. Rangel LP, Costa DCF, Vieira TCRG, Silva JL. 2014. The aggregation of mutant p53 produces prion-like properties in cancer. Prion 8:75–84.

63. Ghosh S, Salot S, Sengupta S, Navalkar A, Ghosh D, Jacob R, Das S, Kumar R, Jha NN, Sahay S, Mehra S, Mohite GM, Ghosh SK, Kombrabail M, Krishnamoorthy G, Chaudhari P, Maji SK. 2017. p53 amyloid formation leading to its loss of function: implications in cancer pathogenesis. Cell Death Differ 24:1784–1798.

64. De Oliveira GAP, Petronilho EC, Pedrote MM, Marques MA, Vieira TCRG, Cino EA, Silva JL. 2020. The Status of p53 Oligomeric and Aggregation States in Cancer. Biomolecules 10.

65. Levy CB, Stumbo AC, Ano Bom APD, Portari EA, Carneiro Y, Silva JL, De Moura Gallo CV. 2011. Co-localization of mutant p53 and amyloid-like protein aggregates in breast tumors. Int J Biochem Cell Biol 43:60–64.

66. Ano Bom APD, Rangel LP, Costa DCF, De Oliveira GAP, Sanches D, Braga CA, Gava LM, Ramos CHI, Cepeda AOT, Stumbo AC, De Moura Gallo CV., Cordeiros Y, Silva JL. 2012. Mutant p53 aggregates into prion-like amyloid oligomers and fibrils: Implications for cancer. J Biol Chem 287:28152–28162.

67. Xu J, Reumers J, Couceiro JR, De Smet F, Gallardo R, Rudyak S, Cornelis A, Rozenski J, Zwolinska A, Marine J-C, Lambrechts D, Suh Y-A, Rousseau F, Schymkowitz J. 2011. Gain of function of mutant p53 by coaggregation with multiple tumor suppressors. Nat Chem Biol 7:285–95.

68. Park SK, Park S, Pentek C, Liebman SW. 2021. Tumor suppressor protein p53 expressed in yeast can remain diffuse, form a prion, or form unstable liquid-like droplets. iScience

24:102000.

69. Silva JL, De Moura Gallo C V, Costa DCF, Rangel LP. 2014. Prion-like aggregation of mutant p53 in cancer. Trends Biochem Sci 39:260–7.

70. Billant O, Friocourt G, Roux P, Voisset C. 2021. P53, a victim of the prion fashion. Cancers (Basel) 13:1–15.

71. Bell D, Berchuck A, Birrer M, Chien J, Cramer DW, Dao F, Dhir R, Disaia P, Gabra H, Glenn P, Godwin AK, Gross J, Hartmann L, Huang M, Huntsman DG, Iacocca M, Imielinski M, Kalloger S, Karlan BY, Levine DA, Mills GB, Morrison C, Mutch D, Olvera N, Orsulic S, Park K, Petrelli N, Rabeno B, Rader JS, Sikic BI, Smith-Mccune K, Sood AK, Bowtell D, Penny R, Testa JR, Chang K, Dinh HH, Drummond JA, Fowler G, Gunaratne P, Hawes AC, Kovar CL, Lewis LR, Morgan MB, Newsham IF, Santibanez J, Reid JG, Trevino LR, Wu YQ, Wang M, Muzny DM, Wheeler DA, Gibbs RA, Getz G, Lawrence MS, Cibulskis K, Sivachenko AY, Sougnez C, Voet D, Wilkinson J, Bloom T, Ardlie K, Fennell T, Baldwin J, Gabriel S, Lander ES, Ding L, Fulton RS, Koboldt DC, McLellan MD, Wylie T, Walker J, O'Laughlin M, Dooling DJ, Fulton L, Abbott R, Dees ND, Zhang Q, Kandoth C, Wendl M, Schierding W, Shen D, Harris CC, Schmidt H, Kalicki J, Delehaunty KD, Fronick CC, Demeter R, Cook L, Wallis JW, Lin L, Magrini VJ, Hodges JS, Eldred JM, Smith SM, Pohl CS, Vandin F, Raphael BJ, Weinstock GM, Mardis ER, Wilson RK, Meyerson M, Winckler W, Verhaak RGW, Carter SL, Mermel CH, Saksena G, Nguyen H, Onofrio RC, Hubbard D, Gupta S, Crenshaw A, Ramos AH, Chin L, Protopopov A, Zhang J, Kim TM, Perna I, Xiao Y, Zhang H, Ren G, Sathiamoorthy N, Park RW, Lee E, Park PJ, Kucherlapati R, Absher DM, Waite L, Sherlock G, Brooks JD, Li JZ, Xu J, Myers RM, Laird PW, Cope L, Herman JG, Shen H, Weisenberger DJ, Noushmehr

H, Pan F, Triche T, Berman BP, Van Den Berg DJ, Buckley J, Baylin SB, Spellman PT, Purdom E, Neuvial P, Bengtsson H, Jakkula LR, Durinck S, Han J, Dorton S, Marr H, Choi YG, Wang V, Wang NJ, Ngai J, Conboy JG, Parvin B, Feiler HS, Speed TP, Gray JW, Socci ND, Liang Y, Taylor BS, Schultz N, Borsu L, Lash AE, Brennan C, Viale A, Sander C, Ladanyi M, Hoadley KA, Meng S, Du Y, Shi Y, Li L, Turman YJ, Zang D, Helms EB, Balu S, Zhou X, Wu J, Topal MD, Hayes DN, Perou CM, Wu CJ, Shukla S, Sivachenko AY, Jing R, Liu Y, Noble M, Carter H, Kim D, Karchin R, Korkola JE, Heiser LM, Cho RJ, Hu Z, Cerami E, Olshen A, Reva B, Antipin Y, Shen R, Mankoo P, Sheridan R, Ciriello G, Chang WK, Bernanke JA, Haussler D, Benz CC, Stuart JM, Benz SC, Sanborn JZ, Vaske CJ, Zhu J, Szeto C, Scott GK, Yau C, Wilkerson MD, Zhang N, Akbani R, Baggerly KA, Yung WK, Weinstein JN, Shelton T, Grimm D, Hatfield M, Morris S, Yena P, Rhodes P, Sherman M, Paulauskis J, Millis S, Kahn A, Greene JM, Sfeir R, Jensen MA, Chen J, Whitmore J, Alonso S, Jordan J, Chu A, Barker A, Compton C, Eley G, Ferguson M, Fielding P, Gerhard DS, Myles R, Schaefer C, Mills Shaw KR, Vaught J, Vockley JB, Good PJ, Guyer MS, Ozenberger B, Peterson J, Thomson E. 2011. Integrated genomic analyses of ovarian carcinoma. Nature 474:609–615.

72. Yang-Hartwich Y, Soteras MG, Lin ZP, Holmberg J, Sumi N, Craveiro V, Liang M, Romanoff E, Bingham J, Garofalo F, Alvero A, Mor G. 2014. P53 Protein Aggregation Promotes Platinum Resistance in Ovarian Cancer. Oncogene 34:25263447.

73. Soragni A, Janzen DM, Johnson LM, Lindgren AG, Thai-Quynh Nguyen A, Tiourin E, Soriaga AB, Lu J, Jiang L, Faull KF, Pellegrini M, Memarzadeh S, Eisenberg DS. 2016. A Designed Inhibitor of p53 Aggregation Rescues p53 Tumor Suppression in Ovarian Carcinomas. Cancer Cell 29:90–103.

74. Lei J, Qi R, Wei G, Nussinov R, Ma B. 2016. Self-aggregation and coaggregation of the p53 core fragment with its aggregation gatekeeper variant. Phys Chem Chem Phys 18:8098–107.

75. Prusiner SB, Scott MR, Dearmond SJ, Cohen FE. 1998. Prion Protein Biology Review. Cell 93:337–48.

76. Hill AF, Joiner S, Linehan J, Desbruslais M, Lantos PL, Collinge J. 2000. Species-barrier-independent prion replication in apparently resistant species. Proc Natl Acad Sci U S A 97:10248–53.

77. Castilla J, Gonzalez-Romero D, Saá P, Morales R, De Castro J, Soto C. 2008. Crossing the Species Barrier by PrPSc Replication In Vitro Generates Unique Infectious Prions. Cell 134:757–768.

78. Cai X, Xu H, Chen ZJ. 2017. Prion-like polymerization in immunity and inflammation. Cold Spring Harb Perspect Biol 9:a023580.

79. Seth RB, Sun L, Ea CK, Chen ZJ. 2005. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-κB and IRF3. Cell 122:669–682.

80. Hou F, Sun L, Zheng H, Skaug B, Jiang QX, Chen ZJ. 2011. MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. Cell 146:448–461.

81. Cai X, Chen J, Xu H, Liu S, Jiang Q-X, Halfmann R, Chen ZJ. 2014. Prion-like polymerization underlies signal transduction in antiviral immune defense and inflammasome activation. Cell 156:1207–22.

82. Si K, Giustetto M, Etkin A, Hsu R, Janisiewicz AM, Miniaci MC, Kim JH, Zhu H, Kandel ER. 2003. A Neuronal Isoform of CPEB Regulates Local Protein Synthesis and Stabilizes

Synapse-Specific Long-Term Facilitation in Aplysia. Cell 115:893–904.

83. Si K, Lindquist S, Kandel ER. 2003. A Neuronal Isoform of the Aplysia CPEB Has Prion-Like Properties. Cell 115:879–891.

84. Heinrich SU, Lindquist S. 2011. Protein-only mechanism induces self-perpetuating changes in the activity of neuronal Aplysia cytoplasmic polyadenylation element binding protein (CPEB). Proc Natl Acad Sci U S A 108:2999–3004.

85. Si K, Choi YB, White-Grindley E, Majumdar A, Kandel ER. 2010. Aplysia CPEB Can Form Prion-like Multimers in Sensory Neurons that Contribute to Long-Term Facilitation. Cell 140:421–435.

86. Majumdar A, Cesario WC, White-Grindley E, Jiang H, Ren F, Khan MR, Li L, Choi EML, Kannan K, Guo F, Unruh J, Slaughter B, Si K. 2012. Critical role of amyloid-like oligomers of Drosophila Orb2 in the persistence of memory. Cell 148:515–529.

87. Khan MR, Li L, Pérez-Sánchez C, Saraf A, Florens L, Slaughter BD, Unruh JR, Si K. 2015. Amyloidogenic Oligomerization Transforms Drosophila Orb2 from a Translation Repressor to an Activator. Cell 163:1468–1483.

88. Stephan JS, Fioriti L, Lamba N, Colnaghi L, Karl K, Derkatch IL, Kandel ER. 2015. The CPEB3 Protein Is a Functional Prion that Interacts with the Actin Cytoskeleton. Cell Rep 11:1772–1785.

89. Fioriti L, Myers C, Huang Y-Y, Li X, Stephan JS, Trifilieff P, Colnaghi L, Kosmidis S, Drisaldi B, Pavlopoulos E, Kandel ER. 2015. The Persistence of Hippocampal-Based Memory Requires Protein Synthesis Mediated by the Prion-like Protein CPEB3. Neuron 86:1433–1448.

90. Coustou V, Deleu C, Saupe S, Begueret J. 1997. The protein product of the het-s

heterokaryon incompatibility gene of the fungus Podospora anserina behaves as a prion analog. Proc Natl Acad Sci U S A 94:9773–9778.

91. Turcq B, Denayrolles M, Begueret J. 1990. Isolation of the two allelic incompatibility genes s and S of the fungus Podospora anserina 297–303.

92. Bégueret J, Turcq B, Clavé C. 1994. Vegetative incompatibility in filamentous fungi: het genes begin to talk. Trends Genet 10:441–446.

93. Coustou-Linares V, Maddelein ML, Bégueret J, Saupe SJ. 2001. In vivo aggregation of the HET-s prion protein of the fungus Podospora anserina. Mol Microbiol 42:1325–1335.

94. Balguerie A, Dos Reis S, Ritter C, Chaignepain S, Coulary-Salin B, Forge V, Bathany K, Lascu I, Schmitter JM, Riek R, Saupe SJ. 2003. Domain organization and structure-function relationship of the HET-s prion protein of Podospora anserina. EMBO J 22:2071–2081.

95. Pallarès I, Iglesias V, Ventura S. 2016. The rho termination factor of Clostridium botulinum contains a prion-like domain with a highly amyloidogenic core. Front Microbiol 6:1–12.

96. Yuan AH, Hochschild A. 2017. A bacterial global regulator forms a prion. Science 355:198–201.

97. Nan H, Chen H, Tuite MF, Xu X. 2019. A viral expression factor behaves as a prion. Nat Commun 10:359.

98. Lu A, Miller LK. 1995. The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. J Virol 69:975–982.

99. Tetz G, Tetz V. 2018. Prion-like Domains in Eukaryotic Viruses. Sci Rep 8:1–10.

100. Du Z, Park KK-W, Yu H, Fan Q, Li L. 2008. Newly identified prion linked to the chromatin-remodeling factor Swi1 in Saccharomyces cerevisiae. Nat Genet 40:460–465.

101. Patel BK, Gavin-Smyth J, Liebman SW. 2009. The yeast global transcriptional co-repressor

protein Cyc8 can propagate as a prion. Nat Cell Biol 11:344–349.

102. Suzuki G, Shimazu N, Tanaka M. 2012. A yeast prion, Mod5, promotes acquired drug resistance and cell survival under environmental stress. Science 336:355–9.

103. Halfmann R, Wright JR, Alberti S, Lindquist S, Rexach M. 2012. Prion formation by a yeast GLFG nucleoporin. Prion 6:391–399.

104. Chakravarty AK, Smejkal T, Itakura AK, Garcia DM, Jarosz DF. 2020. A Non-amyloid Prion Particle that Activates a Heritable Gene Expression Program. Mol Cell 77:251-265.e9.

105. Harvey ZH, Chakravarty AK, Futia RA, Jarosz DF. 2020. A Prion Epigenetic Switch Establishes an Active Chromatin State. Cell 180:928-940.e14.

106. Wickner RB. 1994. [URE3] as an altered URE2 protein: Evidence for a prion analog in Saccharomyces cerevisiae. Science (80-) 264:566–569.

107. Sondheimer N, Lindquist S. 2000. Rnq1: an epigenetic modifier of protein function in yeast. Mol Cell 5:163–172.

108. Brown JCS, Lindquist S. 2009. A heritable switch in carbon source utilization driven by an unusual yeast prion. Genes Dev 23:2320–2332.

109. Alberti S, Halfmann R, King O, Kapila A, Lindquist S. 2009. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. Cell 137:146–158.

110. Cox BS. 1965. Ψ, A cytoplasmic suppressor of super-suppressor in yeast. Heredity 20:505–521.

111. Lacroute F. 1971. Non-Mendelian mutation allowing ureidosuccinic acid uptake in yeast. J Bacteriol 106:519–522.

112. Cox BS, Tuite MF, McLaughlin CS. 1988. The ψ factor of yeast: A problem in inheritance. Yeast 4:159–178.

113. Patino MM, Liu JJ, Glover JR, Lindquist S. 1996. Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. Science 273:622–6.

114. Chernoff YO, Derkach IL, Inge-Vechtomov SG. 1993. Multicopy SUP35 gene induces de-novo appearance of psi-like factors in the yeast Saccharomyces cerevisiae. Curr Genet 24:268–70.

115. Chernoff YO, Lindquist SL, Ono B, Inge-Vechtomov SG, Liebman SW. 1995. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. Science 268:880–884.

116. Kryndushkin DS, Alexandrov IM, Ter-Avanesyan MD, Kushnirov V V. 2003. Yeast [PSI+] prion aggregates are formed by small Sup35 polymers fragmented by Hsp104. J Biol Chem 278:49636–49643.

117. Cox B, Ness F, Tuite M. 2003. Analysis of the generation and segregation of propagons: Entities that propagate the [PSI+] prion in yeast. Genetics 165:23–33.

118. Ferreira PC, Ness F, Edwards SR, Cox BS, Tuite MF. 2001. The elimination of the yeast [PSI+] prion by guanidine hydrochloride is the result of Hsp104 inactivation. Mol Microbiol 40:1357–1369.

119. Jones G, Song Y, Chung S, Masison DC. 2004. Propagation of Saccharomyces cerevisiae [PSI+] prion is impaired by factors that regulate Hsp70 substrate binding. Mol Cell Biol 24:3928–37.

120. Allen KD, Wegrzyn RD, Chernova TA, Müller S, Newnam GP, Winslett PA, Wittich KB, Wilkinson KD, Chernoff YO. 2005. Hsp70 chaperones as modulators of prion life cycle:

Novel effects of Ssa and Ssb on the Saccharomyces cerevisiae prion [PSI+]. Genetics 169:1227–1242.

121. Fan Q, Park KW, Du Z, Morano KA, Li L. 2007. The role of Sse1 in the de novo formation and variant determination of the [PSI+] prion. Genetics 177:1583–1593.

122. Sondheimer N, Lopez N, Craig EA, Lindquist S. 2001. The role of Sis1 in the maintenance of the [RNQ+] prion. EMBO J 20:2435–2442.

123. Kushnirov V V, Kryndushkin DS, Boguta M, Smirnov VN, Ter-Avanesyan MD. 2000. Chaperones that cure yeast artificial [PSI+] and their prion-specific effects. Curr Biol 10:1443–6.

124. Derkatch IL, Bradley ME, Zhou P, Chernoff YO, Liebman SW. 1997. Genetic and environmental factors affecting the de novo appearance of the [PSI+] prion in Saccharomyces cerevisiae. Genetics 147:507–19.

125. Derkatch IL, Bradley ME, Masse S V, Zadorsky SP, Polozkov G V, Inge-Vechtomov SG, Liebman SW. 2000. Dependence and independence of [PSI(+)] and [PIN(+)]: a two-prion system in yeast? EMBO J 19:1942–52.

126. Derkatch I, Bradley M, Hong J, Liebman S. 2001. Prions affect the appearance of other prions: the story of [PIN]. Cell 106:171–182.

127. Aigle M, Lacroute F. 1975. Genetical aspects of [URE3], a non-mitochondrial, cytoplasmically inherited mutation in yeast. MGG Mol Gen Genet 136:327–335.

128. Edskes HK, Gray VT, Wickner RB. 1999. The [URE3] prion is an aggregated form of Ure2p that can be cured by overexpression of Ure2p fragments. Proc Natl Acad Sci 96:1498–1503.

129. Schwimmer C, Masison DC. 2002. Antagonistic interactions between yeast [PSI(+)] and [URE3] prions and curing of [URE3] by Hsp70 protein chaperone Ssa1p but not by Ssa2p.

Mol Cell Biol 22:3590–3598.

130. Roberts BT, Moriyama H, Wickner RB, Wiley J. 2004. [URE3] prion propagation is abolished by a mutation of the primary cytosolic Hsp70 of budding yeast. Yeast 21:107–117.

131. Moriyama H, Edskes HK, Wickner RB. 2000. [URE3] Prion Propagation in Saccharomyces cerevisiae: Requirement for Chaperone Hsp104 and Curing by Overexpressed Chaperone Ydj1p. Mol Cell Biol 20:8916–8922.

132. Kryndushkin D, Wickner RB. 2007. Nucleotide exchange factors for Hsp70s are required for [URE3] prion propagation in Saccharomyces cerevisiae. Mol Biol Cell 18:2149–54.

133. Holmes DL, Lancaster AK, Lindquist S, Halfmann R. 2013. Heritable remodeling of yeast multicellularity by an environmentally responsive prion. Cell 153:153–165.

134. Du Z, Zhang Y, Li L. 2015. The Yeast Prion [SWI+] Abolishes Multicellular Growth by Triggering Conformational Changes of Multiple Regulators Required for Flocculin Gene Expression. Cell Rep 13:2865–2878.

135. Wickner RB, Edskes HK, Bateman D, Kelly AC, Gorkovskiy A. 2011. The yeast prions [PSI+] and [URE3] are molecular degenerative diseases. Prion 5:258–262.

136. Osherovich LZ, Weissman JS. Multiple Gln/Asn-rich prion domains confer susceptibility to induction of the yeast [PSI(+)] prion. Cell 106:183–94.

137. Aron R, Higurashi T, Sahi C, Craig EA. 2007. J-protein co-chaperone Sis1 required for generation of [RNQ+] seeds necessary for prion propagation. EMBO J 26:3794–3803.

138. Huang VJ, Stein KC, True HL. 2013. Spontaneous variants of the [RNQ+] prion in yeast demonstrate the extensive conformational diversity possible with prion proteins. PLoS One 8:e79582.

139. Bradley ME, Edskes HK, Hong JY, Wickner RB, Liebman SW. 2002. Interactions among prions and prion “strains” in yeast. Proc Natl Acad Sci U S A 99:16392–16399.

140. Du Z, Goncharoff DK, Cheng X, Li L. 2017. Analysis of [SWI(+)] formation and propagation events. Mol Microbiol 104:105–124.

141. Stein KC, True HL. 2011. The [RNQ+] prion: A model of both functional and pathological amyloid. Prion 5:291–298.

142. Halfmann R, Jarosz DF, Jones SK, Chang A, Lancaster AK, Lindquist S. 2012. Prions are a common mechanism for phenotypic inheritance in wild yeasts. Nature 482:363–368.

143. Wente SR, Rout MP, Blobel G. 1992. A new family of yeast nuclear pore complex proteins. J Cell Biol 119:705–723.

144. Denning DP, Patel SS, Uversky V, Fink AL, Rexach M. 2003. Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded. Proc Natl Acad Sci U S A 100:2450–2455.

145. Chakrabortee S, Byers JS, Jones S, Garcia DM, Bhullar B, Chang A, She R, Lee L, Fremin B, Lindquist S, Jarosz DF. 2016. Intrinsically Disordered Proteins Drive Emergence and Inheritance of Biological Traits. Cell 167:369-381.e12.

146. Jones DT, Cozzetto D. 2015. DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31:857–863.

147. Itakura AK, Chakravarty AK, Jakobson CM, Jarosz DF. 2020. Widespread Prion-Based Control of Growth and Differentiation Strategies in Saccharomyces cerevisiae. Mol Cell 77:266-278.e6.

148. Pim Pijnappel WWM, Schaft D, Roguev A, Shevchenko A, Tekotte H, Wilm M, Rigaut G, Séraphin B, Aasland R, Francis Stewart A. 2001. The S. cerevisiae SET3 complex includes

two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. Genes Dev 15:2991–3004.

149. Neigeborn L, Carlson M. 1984. Genes affecting the regulation of SUC2 gene expression by glucose repression in Saccharomyces cerevisiae. Genetics 108:845–858.

150. Malovichko Y V., Antonets KS, Maslova AR, Andreeva EA, Inge-Vechtomov SG, Nizhnikov AA. 2019. RNA Sequencing Reveals Specific Transcriptomic
 Signatures Distinguishing Effects of the [SWI+] Prion and SWI1 Deletion in Yeast Saccharomyces cerevisiae. Genes (Basel) 10:212.

151. Du Z, Regan J, Bartom E, Wu W-S, Zhang L, Goncharoff DK, Li L. 2020. Elucidating the regulatory mechanism of Swi1 prion in global transcription and stress responses. Sci Rep 10:21838.

152. Goncharoff DK, Du Z, Li L. 2018. A brief overview of the Swi1 prion-[SWI+]. FEMS Yeast Res 18:1–9.

153. Newby GA, Lindquist S. 2017. Pioneer cells established by the [SWI+] prion can promote dispersal and out-crossing in yeast. PLOS Biol 15:e2003476.

154. Hines JK, Li X, Du Z, Higurashi T, Li L, Craig EA. 2011. [SWI+], the prion formed by the chromatin remodeling factor Swi1, is highly sensitive to alterations in hsp70 chaperone system activity. PLoS Genet 7:27–29.

155. Du Z, Crow ET, Kang HS, Li L. 2010. Distinct subregions of Swi1 manifest striking differences in prion transmission and SWI/SNF function. Mol Cell Biol 30:4644–4655.

156. Crow ET, Du Z, Li L. 2011. A small, glutamine-free domain propagates the [SWI(+)] prion in budding yeast. Mol Cell Biol 31:3436–44.

157. Valtierra S, Du Z, Li L. 2017. Analysis of Small Critical Regions of Swi1 Conferring Prion

Formation, Maintenance, and Transmission. Mol Cell Biol 37:1–16.

158. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL. 2005. FoldIndex©: A simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21:3435–3438.

159. Toombs J a, Petri M, Paul KR, Kan GY, Ben-Hur A, Ross ED. 2012. De novo design of synthetic prion domains. Proc Natl Acad Sci U S A 109:6519–24.

160. Lancaster AK, Nutter-Upham A, Lindquist S, King OD. 2014. PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition. Bioinformatics 30:2501–2502.

161. Zambrano R, Conchillo-Sole O, Iglesias V, Illa R, Rousseau F, Schymkowitz J, Sabate R, Daura X, Ventura S. 2015. PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores. Nucleic Acids Res 1–7.

162. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. 2004. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 22:1302–1306.

163. Maurer-Stroh S, Debulpaep M, Kuemmerer N, De La Paz ML, Martins IC, Reumers J, Morris KL, Copland A, Serpell L, Serrano L, Schymkowitz JWH, Rousseau F. 2010. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. Nat Methods 7:237–242.

164. Thompson MJ, Sievers S a, Karanicolas J, Ivanova MI, Baker D, Eisenberg D. 2006. The 3D profile method for identifying fibril-forming segments of proteins. Proc Natl Acad Sci U S A 103:4074–4078.

165. Tartaglia GG, Vendruscolo M. 2008. The Zyggregator method for predicting protein aggregation propensities. Chem Soc Rev 37:1395–1401.

166. Walsh I, Seno F, Tosatto SCE, Trovato A. 2014. PASTA 2.0: An improved server for protein aggregation prediction. Nucleic Acids Res 42:1–7.

167. Toombs JA, McCarty BR, Ross ED. 2010. Compositional determinants of prion formation in yeast. Mol Cell Biol 30:319–32.

168. Sabate R, Rousseau F, Schymkowitz J, Ventura S. 2015. What Makes a Protein Sequence a Prion? PLoS Comput Biol 11:e1004013.

169. Uversky VN, Gillespie JR, Fink AL. 2000. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins Struct Funct Genet 41:415–427.

170. Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132.

171. Jones DT, Ward JJ. 2003. Prediction of Disordered Regions in Proteins From Position Specific Score Matrices, p. 573–578. *In* Proteins: Structure, Function and Genetics.

172. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. 2004. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. J Mol Biol 337:635–645.

173. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. 2004. The DISOPRED server for the prediction of protein disorder. Bioinformatics 20:2138–2139.

174. Balbirnie M, Grothe R, Eisenberg DS. 2001. An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated β-sheet structure for amyloid. Proc Natl Acad Sci U S A 98:2375–2380.

175. Trovato A, Chiti F, Maritan A, Seno F. 2006. Insight into the structure of amyloid fibrils

from the analysis of globular proteins. PLoS Comput Biol 2:1608–1618.

176. Walsh I, Martin AJM, Di domenico T, Tosatto SCE. 2012. Espritz: Accurate and fast prediction of protein disorder. Bioinformatics 28:503–509.

177. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. 2003. Rationalization of mutational effects on protein aggregation rates. Nature 424:805–808.

178. DuBay KF, Pawar AP, Chiti F, Zurdo J, Dobson CM, Vendruscolo M. 2004. Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. J Mol Biol 341:1317–1326.

179. Pawar AP, Dubay KF, Zurdo J, Chiti F, Vendruscolo M, Dobson CM. 2005. Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. J Mol Biol 350:379–92.

180. Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F, Vendruscolo M. 2008. Prediction of Aggregation-Prone Regions in Structured Proteins. J Mol Biol 380:425–436.

181. Prusiner SB. 2013. Biology and genetics of prions causing neurodegeneration. Annu Rev Genet 47:601–23.

182. Liebman SW, Chernoff YO. 2012. Prions in Yeast. Genetics.

183. Chakrabortee S, Kayatekin C, Newby GA, Mendillo ML, Lancaster A, Lindquist S. 2016. Luminidependens (LD) is an Arabidopsis protein with prion behavior. Proc Natl Acad Sci U S A 113:6065–70.

184. Amberg DC, Burke DJ, Strathern JN. 2005. Methods in yeast genetics: A Cold Spring Harbor Laboratory course manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

185. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S,

Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: An open-source platform for biological-image analysis. Nat Methods 9:676–682.

186. Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675.
187. Kushnirov V V. 2000. Rapid and reliable protein extraction from yeast. Yeast 16:857–860.
188. Halfmann R, Lindquist S. 2008. Screening for amyloid aggregation by Semi-Denaturing Detergent-Agarose Gel Electrophoresis. J Vis Exp 20–22.
189. Bachmair A, Finley D, Varshavsky A. 1986. In vivo half-life of a protein is a function of its amino-terminal residue. Science 234:179–186.
190. Du Z. 2011. The complexity and implications of yeast prion domains. Prion 5:311–316.
191. Summers DW, Cyr DM. 2011. Use of yeast as a system to study amyloid toxicity. Methods 53:226–231.
192. Kimura Y, Koitabashi S, Fujita T. 2003. Analysis of yeast prion aggregates with amyloid-staining compound in vivo. Cell Struct Funct 28:187–193.
193. Douglas PM, Treusch S, Ren HY, Halfmann R, Duennwald ML, Lindquist S, Cyr DM. 2008. Chaperone-dependent amyloid assembly protects cells from prion toxicity. Proc Natl Acad Sci U S A 105:7206–7211.
194. Kaganovich D, Kopito R, Frydman J. 2008. Misfolded proteins partition between two distinct quality control compartments. Nature 454:1088–1095.
195. Rothe S, Prakash A, Tyedmers J. 2018. The Insoluble Protein Deposit (IPOD) in Yeast. Front Mol Neurosci 11:1–9.
196. Sant'Anna R, Fernández MR, Batlle C, Navarro S, de Groot NS, Serpell L, Ventura S. 2016.

Characterization of Amyloid Cores in Prion Domains. Sci Rep 6:34274.

197. Teunissen AW, Steensma HY. 1995. Review: the dominant flocculation genes of Saccharomyces cerevisiae constitute a new subtelomeric gene family. Yeast 11:1001–13.

198. Willaert RG. 2018. Adhesins of Yeasts: Protein Structure and Interactions. J fungi (Basel, Switzerland) 4:1–28.

199. Rauceo JM, De Armond R, Otoo H, Kahn PC, Klotz SA, Gaur NK, Lipke PN. 2006. Threonine-rich repeats increase fibronectin binding in the Candida albicans adhesin Als5p. Eukaryot Cell 5:1664–1673.

200. Paul KR, Hendrich CG, Waechter A, Harman MR, Ross ED. 2015. Generating new prions by targeted mutation or segment duplication. Proc Natl Acad Sci U S A 112:8584–8589.

201. Khatri I, Tomar R, Ganesan K, Prasad GS, Subramanian S. 2017. Complete genome sequence and comparative genomics of the probiotic yeast Saccharomyces boulardii. Sci Rep 7:371.

202. Yue JX, Li J, Aigrain L, Hallin J, Persson K, Oliver K, Bergström A, Coupland P, Warringer J, Lagomarsino MC, Fischer G, Durbin R, Liti G. 2017. Contrasting evolutionary genome dynamics between domesticated and wild yeasts. Nat Genet 49:913–924.

203. Salazar AN, Gorter de Vries AR, van den Broek M, Brouwers N, de la Torre Cortès P, Kuijpers NGA, Daran J-MG, Abeel T. 2019. Chromosome level assembly and comparative genome analysis confirm lager-brewing yeasts originated from a single hybridization. BMC Genomics 20:916.

204. Ross ED, Baxa U, Wickner RB. 2004. Scrambled Prion Domains Form Prions and Amyloid. Mol Cell Biol 24:7206–7213.

205. Ross ED, Edskes HK, Terry MJ, Wickner RB. 2005. Primary sequence independence for

prion formation. Proc Natl Acad Sci U S A 102:12825–12830.

206. Jarosz DF, Brown JCS, Walker GA, Datta MS, Ung WL, Lancaster AK, Rotem A, Chang A, Newby GA, Weitz DA, Bisson LF, Lindquist S. 2014. Cross-kingdom chemical communication drives a heritable, mutually beneficial prion-based transformation of metabolism. Cell 158:1083–1093.

207. Gonzalez Nelson AC, Paul KR, Petri M, Flores N, Rogge RA, Cascarina SM, Ross ED. 2014. Increasing prion propensity by hydrophobic insertion. PLoS One 9.

208. Paul KR, Ross ED. 2015. Controlling the prion propensity of glutamine/asparagine-rich proteins. Prion 9:347–354.

209. An L, Harrison PM. 2016. The evolutionary scope and neurological disease linkage of yeast-prion-like proteins in humans. Biol Direct 11.

210. Michelitsch MD, Weissman JS. 2000. A census of glutamine/asparagine-rich regions: Implications for their conserved function and the prediction of novel prions. Proc Natl Acad Sci U S A 97:11910–11915.

211. Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, Sigwalt A, Barre B, Freel K, Llored A, Cruaud C, Labadie K, Aury J-M, Istace B, Lebrigand K, Barbry P, Engelen S, Lemainque A, Wincker P, Liti G, Schacherer J. 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature https://doi.org/10.1038/s41586-018-0030-5.

212. Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:1–11.

213. Yang Z. 1997. Paml: A program package for phylogenetic analysis by maximum likelihood. Bioinformatics 13:555–556.

214. Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591.

215. Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43.

216. Bradley ME, Liebman SW. 2004. The Sup35 domains required for maintenance of weak, strong or undifferentiated yeast [PSI+] prions. Mol Microbiol 51:1649–1659.

217. Maddelein M-L, Wickner RB. 1999. Two Prion-Inducing Regions of Ure2p Are Nonoverlapping. Mol Cell Biol 19:4516–4524.

218. Schneider RM, Langdon QK, Libkind D, Opulente DA, Wang M, Kurtzman CP, Shen X, Haase MAB, Endoh R, Hittinger CT, DeVirgilio J, Zhou X, Kominek J, Ohkuma M, Steenwyk JL, Buh K V, Wisecaver JH, Takashima M, Groenewald M, Boudouris JT, Doering DT, Manabe R-I, Rokas A, Čadež N, Hulfachor AB, Rosa CA. 2018. Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. Cell 175:1533-1545.e20.

219. Shain AH, Pollack JR. 2013. The Spectrum of SWI/SNF Mutations, Ubiquitous in Human Cancers. PLoS One 8.

220. Kadoch C, Hargreaves DC, Hodges C, Elias L, Ho L, Ranish J, Crabtree GR. 2013. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat Genet 45:592–601.

221. Wang X, Haswell JR, Roberts CWM. 2014. Molecular pathways: SWI/SNF (BAF) complexes are frequently mutated in cancer-mechanisms and potential therapeutic insights. Clin Cancer Res 20:21–27.

222. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, Bye-A-Jee H, Coetzee R, Cukura A, Silva A Da, Denny

P, Dogan T, Ebenezer TG, Fan J, Castro LG, Garmiri P, Georghiou G, Gonzales L, Hatton-Ellis E, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Jokinen P, Joshi V, Jyothi D, Lock A, Lopez R, Luciani A, Luo J, Lussi Y, MacDougall A, Madeira F, Mahmoudy M, Menchi M, Mishra A, Moulang K, Nightingale A, Oliveira CS, Pundir S, Qi G, Raj S, Rice D, Lopez MR, Saidi R, Sampson J, Sawford T, Speretta E, Turner E, Tyagi N, Vasudev P, Volynkin V, Warner K, Watkins X, Zaru R, Zellner H, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter MC, Bolleman J, Boutet E, Breuza L, Casals-Casas C, de Castro E, Echioukh KC, Coudert E, Cuche B, Doche M, Dornevil D, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Hyka-Nouspikel N, Jungo F, Keller G, Kerhornou A, Lara V, Le Mercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto TB, Paesano S, Pedruzzi I, Pilbout S, Pourcel L, Pozzato M, Pruess M, Rivoire C, Sigrist C, Sonesson K, Stutz A, Sundaram S, Tognolli M, Verbregue L, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Yeh LS, Zhang J, Consortium TU. 2021. UniProt: The universal protein knowledgebase in 2021. Nucleic Acids Res 49:D480–D489.

223. Alberti S, Gitler A, Lindquist S. 2007. A suite of Gateway® cloning vectors for high-throughput genetic analysis in Saccharomyces cerevisiae. Yeast 913–919.

224. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR. 2008. The catalogue of somatic mutations in cancer (COSMIC). Curr Protoc Hum Genet 1–26.

225. Khan T, Kandola TS, Wu J, Venkatesan S, Ketter E, Lange JJ, Rodríguez Gama A, Box A,

Unruh JR, Cook M, Halfmann R. 2018. Quantifying Nucleation In Vivo Reveals the Physical Basis of Prion-like Phase Behavior. Mol Cell 71:155-168.e7.

226. Ford MGJ, Mills IG, Peter BJ, Vallis Y, Praefcke GJK, Evans PR, McMahon HT. 2002. Curvature of clathrin-coated pits driven by epsin. Nature 419:361–366.

227. Mills IG, Praefcke GJK, Vallis Y, Peter BJ, Olesen LE, Gallop JL, Butler PJG, Evans PR, McMahon HT. 2003. Epsinr: An AP1/clathrin interacting protein involved in vesicle trafficking. J Cell Biol 160:213–222.

228. Hirst J, Motley A, Harasaki K, Peak Chew SY, Robinson MS. 2003. EpsinR: an ENTH domain-containing protein that interacts with AP-1. Mol Biol Cell 14:625–41.

229. Huang Y, Ma T, Lau PK, Wang J, Zhao T, Du S, Loy MMT, Guo Y. 2019. Visualization of Protein Sorting at the Trans-Golgi Network and Endosomes Through Super-Resolution Imaging. Front Cell Dev Biol 7:1–13.

230. Saint-Pol A, Yélamos B, Amessou M, Mills IG, Dugast M, Tenza D, Schu P, Antony C, McMahon HT, Lamaze C, Johannes L. 2004. Clathrin adaptor epsinR is required for retrograde sorting on early endosomal membranes. Dev Cell 6:525–538.

231. Tang RQ, Zhao XZ, Shi YY, Tang W, Gu NF, Feng GY, Xing YL, Zhu SM, Sang H, Liang PJ, He L. 2006. Family-based association study of Epsin 4 and Schizophrenia. Mol Psychiatry 11:395–399.

232. Abboud Leon C, Schumacher J, Kluck N, Herold C, Schulze TG, Propping P, Rietschel M, Cichon S, Nöthen MM, Abou Jamra R. 2011. Association study of the GRIA1 and CLINT1 (Epsin 4) genes in a German schizophrenia sample. Psychiatr Genet 21:114.

233. Vitaterna MH, King DP, Chang AM, Kernhauser JM, Lowrey PL, McDonald JD, Dove WF, Pinto LH, Turek FW, Takahashi JS. 1994. Mutagenesis and mapping of a mouse gene,

clock, essential for circadian behavior. Science (80-) 264:719–725.

234. Rijo-Ferreira F, Takahashi JS. 2019. Genomics of circadian rhythms in health and disease. Genome Med 11:1–16.

235. Kettner NM, Voicu H, Finegold MJ, Coarfa C, Sreekumar A, Putluri N, Katchy CA, Lee C, Moore DD, Fu L. 2016. Circadian Homeostasis of Liver Metabolism Suppresses Hepatocarcinogenesis. Cancer Cell 30:909–924.

236. Papagiannakopoulos T, Bauer MR, Davidson SM, Heimann M, Subbaraj L, Bhutkar A, Bartlebaugh J, Vander Heiden MG, Jacks T. 2016. Circadian Rhythm Disruption Promotes Lung Tumorigenesis. Cell Metab 24:324–331.

237. Lagaudrière-Gesbert C, Newmyer SL, Gregers TF, Bakke O, Ploegh HL. 2002. Uncoating ATPase Hsc70 is recruited by invariant chain and controls the size of endocytic compartments. Proc Natl Acad Sci U S A 99:1515–1520.

238. Schulte TW, Akinaga S, Soga S, Sullivan W, Stensgard B, Toft D, Neckers LM. 1998. Antibiotic radicicol binds to the N-terminal domain of Hsp90 and shares important biologic activities with geldanamycin. Cell Stress Chaperones 3:100–108.

239. Huang N, Chelliah Y, Shan Y, Taylor CA, Yoo SH, Partch C, Green CB, Zhang H, Takahashi JS. 2012. Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. Science (80-) 337:189–194.

240. Hou Z, Su L, Pei J, Grishin N V., Zhang H. 2017. Crystal Structure of the CLOCK Transactivation Domain Exon19 in Complex with a Repressor. Structure 25:1187-1194.e3.

241. Welsh DK, Yoo S-H, Liu AC, Takahashi JS, Kay SA. 2004. Bioluminescence imaging of individual fibroblasts reveals persistent, independently phased circadian rhythms of clock gene expression. Curr Biol 14:2289–95.

242. Liu AC, Tran HG, Zhang EE, Priest AA, Welsh DK, Kay SA. 2008. Redundant function of REV-ERBα and β and non-essential role for Bmal1 cycling in transcriptional regulation of intracellular circadian rhythms. PLoS Genet 4.

243. Peek CB, Levine DC, Cedernaes J, Taguchi A, Kobayashi Y, Tsai SJ, Bonar NA, McNulty MR, Ramsey KM, Bass J. 2017. Circadian Clock Interaction with HIF1α Mediates Oxygenic Metabolism and Anaerobic Glycolysis in Skeletal Muscle. Cell Metab 25:86–92.