DISSERTATION

MOLECULAR BASIS OF [PSI$^+$] YEAST PRION NUCLEATION AND PROPAGATION

Submitted by

Zobaida A. Ben Musa

Graduate Degree Program in Cell and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2013

Doctoral Committee:

Advisor: Eric Ross

Debbie Crans
Mark Zabel
Santiago Di Petro

ABSTRACT

MOLECULAR BASIS OF [PSI$^+$] YEAST PRION NUCLEATION

AND PROPAGATION

Many fatal diseases arise from the conversion of soluble, functional proteins to insoluble misfolded amyloid aggregates. Amyloid fibers are characterized by filamentous morphology, protease resistance and cross-beta structure. Prions (infectious amyloids) are a specific subset of amyloid fibers, differing from other classes of amyloids by their infectivity. Prions are found in both mammals and yeasts, but there are differences between these two groups. Most yeast prions are characterized by the presence of large numbers of glutamine and asparagine (Q/N) residues, and some other common characteristics have been noted, including the presence of few hydrophobic and charged residues. Although, several attempts have been made with limited success to develop valuable systems to predict prion activity, there is no accurate algorithm that has the ability to predict the prion-forming proteins among the Q/N-rich protein group. In the yeast, it has been shown that amino acid composition, not primary sequence, drives prion activity. Recently, preliminary efforts to define the role of amino acid composition in prion formation have been examined. The fundamental question of this project is how, in yeast Q/N-rich prions, the sequence requirements for nucleation versus propagation differ, and how this information can be used in order to develop a precise prion prediction system. By answering this question we will be able to more accurately identify additional prions in both yeast and other organisms. Our long-term goal in the comprehensive studies of prion formation and propagation mechanisms is to apply this information to mammalian prion diseases. Consequently, we will be able to identify targets for therapeutic intervention to avoid, slow-down, or reverse the

development of related diseases. The study determined that the amino acids required for prion formation differ from those required for prion propagation. Identifying the sequence feature for both activities is the first step towards mechanistic studies to examine how these sequences perform their function.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

I-Amyloid fibril protein

A wide range of distinct proteins have the ability to structurally convert from the normal soluble conformation to an insoluble amyloid form. Amyloid fibrils are highly organized protein aggregates that are associated with the pathology of various fatal diseases. The accumulation of amyloid fibers in various tissues of the body leads to pathological condition called amyloidosis. Amyloidosis frequently affects the heart, kidneys, liver, spleen, gastrointestinal tract, and pancreas. Insulin-producing Islets of Langerhans of the pancreas become clogged with amyloid deposits, likely contributing to type II diabetes due to impact on the normal function of pancreas. Amyloid may also localize in nervous system and cause multiple neurodegenerative diseases including Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's (HD), and prion related diseases.

The number of proteins known to have the ability to enter amyloid state and form elongated fibers has been growing. The International Society of Amyloidosis (ISA) defined an amyloid-forming protein as a naturally occurring protein that forms un-branched fibers, usually extracellular, and exhibits affinity for Congo red dye and then shows green birefringence upon viewing by polarization microscopy (Sipe *et al*., 2012). There is an additional molecular structure-related definition, in which amyloid fiber must present a characteristic cross-β fiber diffraction pattern (Eisenberg, 2012, Figure1.1). The exposure of protein backbone amide can lead to amyloid based aggregation due to hydrogen bond formation with other protein molecules. β strands run near perpendicular to the fiber axis and stack to form cross β sheet. This can

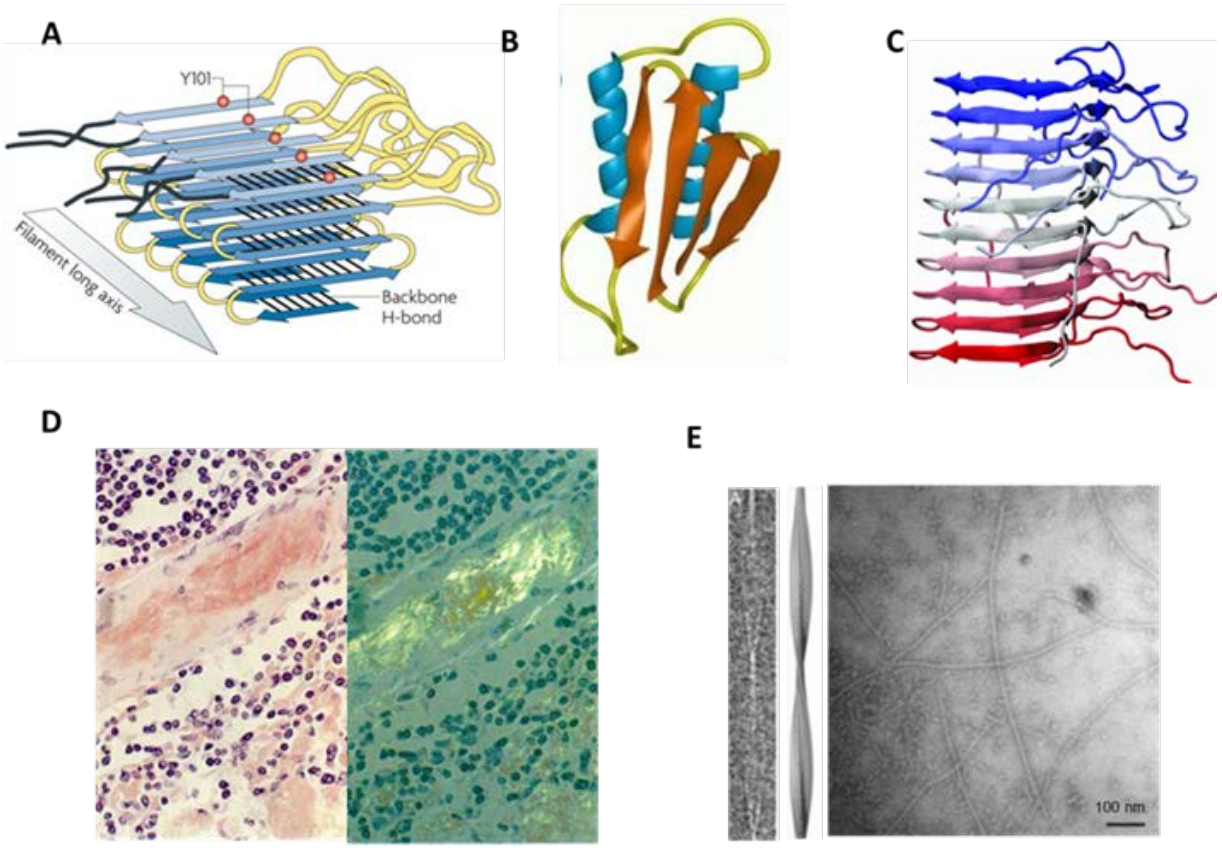


Figure1. 1: Molecular structure of various amyloid forming proteins.

A. Solid state NMR structure of the Sup35NM. Strand in the parallel in-register β-sheet structure of the prion domain (blue arrows) run perpendicular to the long axis of the filaments and are connected by loops (yellow) (Wickner and Edskes, 2007). **B**. Cartoon model of PrPc protein (http://www.bioquest.org/index.php). **C**. Solid-state NMR structure of a fibril of the HET-s fungal amyloid. The fibril is made up of five monomers of the amyloidogenic core of HET-s (Wasmer *et al*, 2008). D. A tissue section showing the pink-red staining of amyoid with the Congo red technique, nuclei-blue. Right: Same section viewed under polarized light to demonstrate the characteristic apple green birefringence of amyloid (Woods and Ellis, 2000). **E**. Electron micrograph of amyloid fibrils formed in water from Aβ1 stained with 0.1% phosphotungstic acid (Sachse, Fandrich, and Grigorieff, 2008).

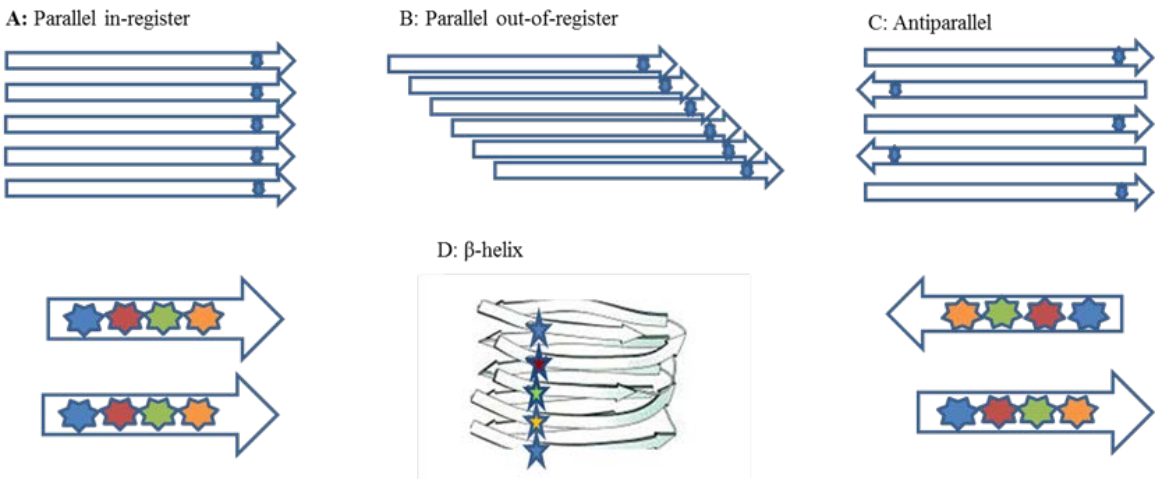


Figure1. 2: The four different arrangements of β sheet architecture.**A.** Parallel in-register type results in identical residues of adjacent strands separated by distance of 4.7 Å , the same colored stars represented identical residues. **B**. Parallel out-of-register. **C.** Antiparallel β-sheet. **D.** β-helix. In both the antiparallel β-sheet and β-helix different amino acid residues are opposed and might be complementary. In case of antiparallel this arrangement results in an identical residuesbeing separated by 9.4 Å.

produce four different arrangements of β sheet architecture (Wickner *et al.* 2010, Figure1.2). First, β strands may run in the same direction, forming parallel sheets. This form might be in-register, where the identical residues of adjacent strands align on top of each other along the fiber axis separated by distance of 4.7 Angstrom. Second, the parallel form might be out-of-register. Third, the strands can run in opposite directions and form antiparallel sheets. While the strands still separated by 4.7 Angstrom, identical residues are generally separated by 9.4 Angstrom. Fourth, the parallel beta strands may associate in a helical pattern to form β-helix.

Amyloid fibrils are composed of several filamentous subunits of 25-30 Angstrom in diameter arranged to form 75-100 Angstrom-wide amyloid fibrils. These subunits twist on each other to form the amyloid fibrils *in vitro*. *In situ* these fibrils stick together to form large deposit (Sipe and Cohen *et al*., 2000). Although the one-dimensional nature of the amyloid fibrils makes it challenging to study fibers composed of full-length proteins by three-dimensional crystallization, there have been some successes with segments of proteins. Recently, the Eisenberg lab has attempted to apply the three-dimensional crystallization approach to short fragments of fibril-forming peptide of amyloid protein including Sup35, Ab, insulin, amylin and tau (Ivanova *et al*., 2009; Nelson *et al*., 2005; Sawaya *et al.*, 2007; Wiltzius *et al*., 2009 ; Wiltzius *et al*., 2008). The peptides grew three-dimensional microcrystals and share a common structure in both fibrils and microcrystal under the same conditions. Furthermore, the microcrystals can seed amyloid fibrils growth and the calculated fiber diffraction of the microcrystals and the diffraction from their fibril equivalents are very similar. In reality, multiple peptide segments can build more than one cross β-sheet spine. Also, sometimes a single protein contains multiple segments that are sufficient to drive amyloid formation. These segment can be either overlapped or be distinct within the sequence. However, ssNMR study found the peptide

formed different conformation in the microcrystals and fibrils (Greenwald and Roland 2010), raising questions about the applicability of three-dimensional crystallization results to amyloid structure. Although the microcrystal approach was able to study segments of peptide out of the fibrils context, it is not applicable to be used for representing the conformations of the segments within the amyloids (Van der Wel, 2010).

Even though Solid State NMR is the most successful molecular approach for determining the structure of amyloid fibers (Jaroniec *et al*., 2002; Petkova *et al*., 2002 ; Ritter *et al.,* 2005), it does not provide a high resolution molecular structure of amyloid. Currently there is no technique that has the capability to determine the atomic resolution structure of amyloid fiber, making it very hard to understand the basis for fibers formation and the physical forces that stabilize these fibers. Therefore, it is been very difficult to predict whether a given protein will be able to form amyloid. Genetic study may not only provide information that helps in understanding the basis of amyloid formation, but also that allows for prediction of amyloid propensity of peptide segments.

II-Mammalian prions disease

Despite the fact that most amyloid fibers are not infectious, prions, a special form of amyloid, are infectious pathogens. Prions are associated with a variety of incurable chronic mammalian neurodegenerative diseases including: Creutzfelt-Jacob Disease (CJD), Fatal Familial Insomnia, Kuru, and animal Transmissible Spongiform Encephalopathies (TSE's), which include Bovine Spongiform Encephalopathy (BSE) in cattle, Scrapie in sheep and Chronic Wasting Disease in cervids. These diseases may present in spontaneous, infectious or inherited forms. Hence, prions obviously are a source of interest in biology and medicine. Transmission and inheritance of prion aggregates results from conversion of protein from its natural form to an

amyloid form, rather than from changes in nucleotide sequence (Chien and Weissman *et al.*, 2004).

So far PrP is the only prion protein known in mammalian kingdom. PrP was first discovered as a causative agent of Scrapie (Prusiner *et al.*, 1982). PrP is associated with TSEs in both, human and animals. TSEs are incurable neurodegenerative diseases characterized by loss of motor control, dementia, and paralysis ended by death (Prusiner *et al.*, 2004). PrP is a glycoprotein, anchored on the cell surface through glycosylphosphatidyl inositol (GPI) (Jones *et al*., 2004). Although, the precise function of the protein PrP still remains elusive, a number of functions have been proposed, including a role in copper ion homeostasis, in that way protecting neurons from oxidative damage (Roucou *et al*., 2005). PrP is also found in the cytoplasm, indicating that it could also have unknown intracellular functions (Roucou *et al*., 2005).

The ‘‘protein only’’ hypothesis (Laurent, *et al*., 1996) proposed that mammalian prion diseases result from conformational changes of naturally occurring, protease-susceptible, soluble isoform of prion protein (PrP^{C}) to a protease-resistant, aggregated pathogenic, isoform (PrP^{SC}). These two different conformations have the same primary structure, or amino acid sequence, but differ in their secondary and tertiary structures. Although the molecular mechanisms leading to the conformation conversion are still debated, the presence of PrP^{C} is required for disease. PrP^{C}-/- knockout mice are resistant to scrapie infection (Brandner, Sebastian, *et al.* 1996) and the restoration of the PrP gene brings back the susceptibility to the infection (Bueler *et al* 1993 and Fischer *et al*., 1996). PrP^{SC} is characterized by high β-sheet content, whereas PrP^{C} has higher α-helix content (Zahn *et al.*, 2000). Nuclear magnetic resonance (NMR) approaches have described

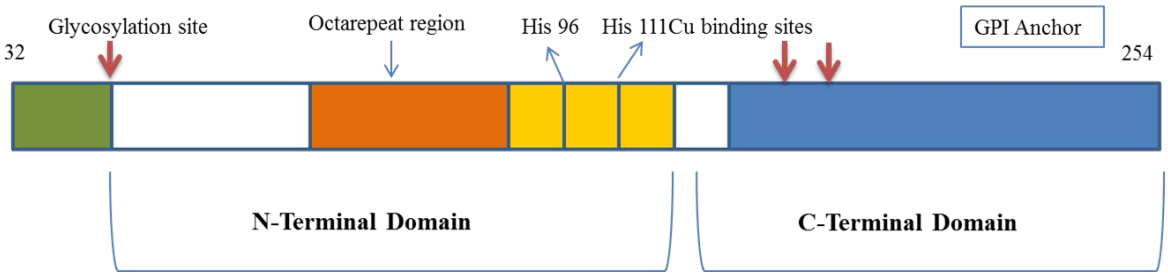


Figure1. 3: Schematic of PrP protein. The N terminal domain (unstructured) made of an Octarepeat region (PHGGGWGQ) and encodes His 96 and 111 copper binding sites. The C terminal domain (structured) contains the sequence for the GPI anchor and the regions connected by cystine bonds. Both regions contain post-translational glycosylation sites. The figure adapted from Prion Protein PrP - Biology, Davidson College.

the soluble forms of PrP as a consisting of two domains, a globular domain that extends approximately from residues 125–228 and contains the sequence for the GPI anchor and an N-terminal disordered domain extends approximately from residues 23-124 consisting of an octapeptide repeated with four copies of the sequence (PHGGGWGQ) (Riek *et al*., 1997; Donne *et al*., 1997; Jones *et al.,* 2004), Figure 1.3.

There are a number of substantial challenges in studying mammalian prions. The most important reason is that spontaneous prion formation is a very rare event. Furthermore, the incubation period of diseases is very long. Therefore, there is significant value in finding model systems that can be used to serve this branch of science.

III-Yeast prions

Whereas PrP is the only prion protein known in mammals, at least nine prion-forming proteins have been identified in the yeast *Saccharomyces cerevisiae*: Sup35p forms the [PSI^+] prion, Ure2p forms [URE3], Cyc8p forms [OCT^+], Mot3p forms [MOT3], Swi1p forms [SWI^+], Sfp1 forms [ISP^+], Mod5 forms [MOD^+] and Nup100 forms [$NUP100^+$]. Each is transmitted from mother to daughter cells as self-propagating protein aggregates. Most of prion-forming proteins participate in significant cellular processes such as translation (Sup35), transcription (Mot3, Cyc8, and Ure2), chromatin remodeling (Swi1). Yeast prions have provided a powerful platform to study the mechanism of prion formation and propagation.

Prion conformers produce new phenotypes by altering processes as diverse as nitrogen metabolism and translation termination. These phenotypes are heritable because prion aggregates attract soluble protein monomers and convert them into the prion conformation. The [$PRION^+$] phenotype is dominant, meaning if a [$PRION^+$] cell mates with a [$prion^-$] cell, the result would be a diploid cell having the [$PRION^+$] phenotype. Moreover, in yeast, prion aggregates can be

transmitted horizontally; prion fibers are found in the cytosols allowing for transmission by cytoduction. Cytoduction is a biological mating of yeast cell's cytosols without nuclear fusion (Zakharov and Yarovoy *et al*., 1977). Even though prions differ from other classes of amyloids by their infectivity, both share many common features, including similar general amyloid structure.

a-The discovery of yeast prions

[PSI^+] and [URE3], were the first prions discovered in yeast. Both were serendipitously discovered. [PSI^+] was identified in a screen for nonsense suppressors (Cox, 1965) and [URE3] in a screen for nitrogen uptake mutations (Lacroute, 1971). While the molecular basis for [PSI^+] and [URE3] phenotypes were initially not understood, they share a number of unique characteristics. Both, [PSI^+] and [URE3] show non-mendelian inheritance and dominance of prion phenotypes in genetic crosses. In both cases the infectious amyloid aggregates can be transmitted by cytoduction, as mentioned earlier. This confirmed that the responsible genetic material is located somewhere in the cell cytosol (it was originally thought to be in the mitochondria, but this was disproved).

Many years later, in 1994, yeast geneticist Reed Wickner proposed that [PSI^+] and [URE3] were prion forms of normal functional cellular proteins Sup35 and Ure2, respectively. Wickner was able to confirm prion feature of Sup35 and Ure2 using four principles (Wickner *et al.,* 1994). First, a transient increase in the level of both proteins substantially stimulates the formation of prions (Tuite *et al*., 2011; Chernoff *et al.,* 1993). It has been proposed that the reasons for *de novo* prion formation caused by protein overexpression may due to increase the protein affinity to misfold. Increasing the local concentration of proteins could also increase the chance for monomers to find each other and form the nuclei. By contrast, other infectious agents

(Such as viruses and bacteria) should be insensitive to the protein concentration. Second, [PSI^+] and [URE3] can be reversibly eradicated by temporary growth in the presence of guanidine hydrochloride (GuHCl), which cures cells of prions by inhibiting Hsp104p, a molecular chaperone needed for prion maintenance (Chernoff *et al*., 2004; Lund & Cox *et al.*, 1981; Tuite *et al*., 1981). This phenomenon can be used to differentiate between prions and DNA mutation where both have the same phenotype. Third, both proteins are needed for the cells in order to form and propagate the prion phenotype. Fourth, the prion phenotypes mimic loss of function mutations in both proteins (Tuite *et al*., 2011; Lacroute *et al*., 1971; Cox *et al*., 1965). Subsequently, it was shown that transformation of [$prion^-$] cells with either *in vitro* formed aggregates or prions extracted from [$PRION^+$] cells was sufficient to introduce prion infection (Brachmann *et al*., 2005; Tanaka *et al*., 2004). While the discovery of [PSI^+] and [URE3] was serendipitous, since then intensive searches to identify additional prions have been performed (discussed later).

b- Yeast prions characterized by loss of function

For all known yeast prions (except Rnq1, which has no known function), prion formation results in the loss of function of the corresponding proteins. Ure2, Cyc8, Mot3p, Sfp1, or Swi1 are each transcriptional factor, so conversion of these proteins to their prion form may cause dramatic change on genes expression. As an example, the normal function of globular domain of Ure2 is in nitrogen regulation, blocking the uptake of poor nitrogen sources in the presence of a good nitrogen source (Coschigano & Magasanik *et al*., 1991; Cooper *et al*., 2006). Ure2 accomplishes this by binding to the transcriptional factor Gln3, thereby preventing Gln3 from entering the nucleus and activating genes involved in nitrogen uptake. Upon prion formation, Ure2 loses the ability to bind to Gln3, leading to uptake poor nitrogen sources, even in the

presence of good sources. Overall, in all cases, [PRION^{+}] cells exhibit phenotype dissimilar to [prion^{-}] cells, allowing for the development of simple assays to distinguish between the two conformations.

c- Are prions beneficial or deleterious?

Although it is obvious that mammalian prions are disease-causing, determining whether yeast prions are helpful or harmful is still very controversial. Prions segregate 4:0; consequently they should quickly dominate the population, unless there is selection against them. In reality the frequency of prion appearance in wild population is very low, indicating a harmful effect on the cells. Wickner's lab has screened 70 wild yeast strains obtained from diverse location and under different environmental conditions looking for [PSI^{+}] and [URE3^{+}]; none had either, indicating their harmful effect (Nakayashiki and Toru, *et al*., 2005).

Nevertheless, the lines of evidence supporting the beneficial theory of yeast prions have been growing. It has been proposed that [PSI^{+}] cells showed more stress resistance than corresponding [psi^{-}] under selective culture condition (when cells are not well-suited to their environments) (Eagle-stone *et al*., 1999). It is clear that yeast prions act as genetic machinery that replicate the information (True, Berlin *et al*. 2004 and Tuite and Cox 2009), but the question is whether it is helpful or harmful genetic information. This concept may also applicable for [Het-s], a prion that facilitates heterokaryon incompatibility in *Psodospora anserine (*coustou *et al*., 1997). [Het-s] protects cells from being infected with fungal viruses by blocking vegetative fusion of strains with different genetic background (Wickner, 1999).

Recently, Susan Lindquist's lab published a paper offering additional evidence supporting the beneficial hypothesis of yeast prions. Approximately 700 wild strains of *Saccharomyces* were biochemically screened for [PSI^{+}] or [MOT3^{+}], and prions were found in

many isolates (Halfmann *et al*., 2012). Halfmann *et al* presented a hypothesis to elucidate how prion could be beneficial despite their relative rarity. The rate of *de novo* prion formation is approximately $1:10^6$ for most yeast prions. When yeast colony approaches appropriate size, it is likely to have a slight amount of [PRION$^+$] cells. If prion is detrimental, prion containing cells will be selected against. Because of their rarity, this will only have a slight effect on the colony. However, if the prion confers a substantial benefit, the prion positive cells will rapidly take over the population. Moreover, [prion$^-$] may reemerge again in [PRION$^+$] population balancing the survival scales.

Overall, I believe there are two possibilities consistent with evidence. First, prions are modestly harmful. The rare natural *de novo* formation of yeast prion indicates that the prion phenotype is not generally helpful. Second, prions are generally detrimental, but under certain conditions confer a substantial benefit; for example, certain stress conditions may induce [PSI$^+$] formation to help the cell overcome the stress by altering translation process (Tyedmers *et al*., 2008).

d- Molecular basis of yeast prion proteins

Most yeast prion proteins share a common overall architecture (with exception of Mod5), with a prion forming domain (PFD) that is structurally distinct from the functional domain of the protein. The PFDs are unstructured and generally located at the N-terminus of the proteins. These PFDs are required for prion formation, but generally dispensable for the normal cellular function of the protein (the Rnq1 is only exception, as it has no known function other than prion formation). Each PFD region is glutamine/asparagine (Q/N) rich, with few hydrophobic and charged residues. Scrambling of the amino acid sequence of the PFDs of both Sup35 and Ure2 does not block prion formation, indicating that amino acid composition, not primary sequence,

drives prion formation (Ross *et al*, 2005). Yeast prion forming domains are generally modular and portable, when appended to other proteins, the protein will exhibit prion activity (Lindquist *et al.,* 2000; Baxa *et al.,* 2002). This feature of yeast prion domains does not exist in mammalian prions (Colby and Prusiner, 2011). This modular set-up provides a powerful *in vivo* system for defining the sequence requirements for prion activities.

The prion field initially used the yeast prions to model PrP, but increasingly it appears that PrP is substantially different from the yeast prion proteins. Although the exact sequence features that drive PrP aggregation are still obscure, it is clear that the yeast prions and PrP have very different amino acids compositions. Moreover, all algorithms designed to predict aggregation by non-Q/N-rich proteins are not sufficient to distinguish between Q/N rich proteins with and without prion activity. Therefore, it is likely that yeast prions will be a not ideal model for PrP, but they will probably be a respectable model for human Q/N-rich aggregating proteins. The amino acid composition of known PFDs potentially provides a method to distinguish yeast prions from other high Q/N content amyloid-forming proteins. Such characteristic of PFDs has been used to identify more prion-forming proteins by looking for proteins with similar compositions to existing prions. Such methods are obviously effective at picking out potential prion candidates from genomes. Yet, these methods exhibited no capability to distinguish among these candidates. Unfortunately, these attempts have got inadequate success (discussed later). Additionally, a shared feature of yeast prion-forming proteins that distinguishes them from many other amyloids is that they are disordered in the native state of the protein. This makes the domains accessible for amyloid formation (Tuite and Cox, 2003). By contrast, in some other amyloid forming proteins, the amyloidegenic region that is responsible for amyloid formation is

buried in the native state of the proteins, therefore, the stability of the native state can prevent the protein from adopting the amyloid conformation.

e- Prion amyloid polymerization and aggregation

In order for a protein to act as a prion in yeast it must be able to not only form aggregates, but also propagate these aggregates over many generations of cell division. Based on current knowledge, the mechanism of prion fibers formation consists of two major steps: fiber nucleation also referred to as the *de novo* aggregation, and the fiber elongation. These two polymerization steps are followed by fiber cleavage. Although the exact nucleation mechanism is unclear, it is thought that nucleation involves an unknown number of monomers assembling to form a stable nucleus. This nucleation is the rate limiting step and is dependent on protein concentration (Serio and Cashikar, 2000). Therefore, if the proteins concentration is elevated, this lag phase is diminished, and if seeds (preassembled fibers) are added to the solution, the lag phase will be bypassed (Serio and Cashikar, 2000; Scheibel and Lindquist, 2001).

Nucleus formation is followed by the temperature-dependent step of fiber elongation. Fibers only grow from the ends, although it is yet unknown whether addition of the units to both ends of the fiber takes place at the same rate (Scheibel, Kowal *et al*. 2001; DePace and Weissman 2002). It is clear that this growth phase consists of two-stage processes: the interaction of soluble units (monomer, oligomer or both) with nuclei and the conformational conversion of functional protein molecule to misfolded form; however, which one occurs first is still vagu, Figure 1.4.

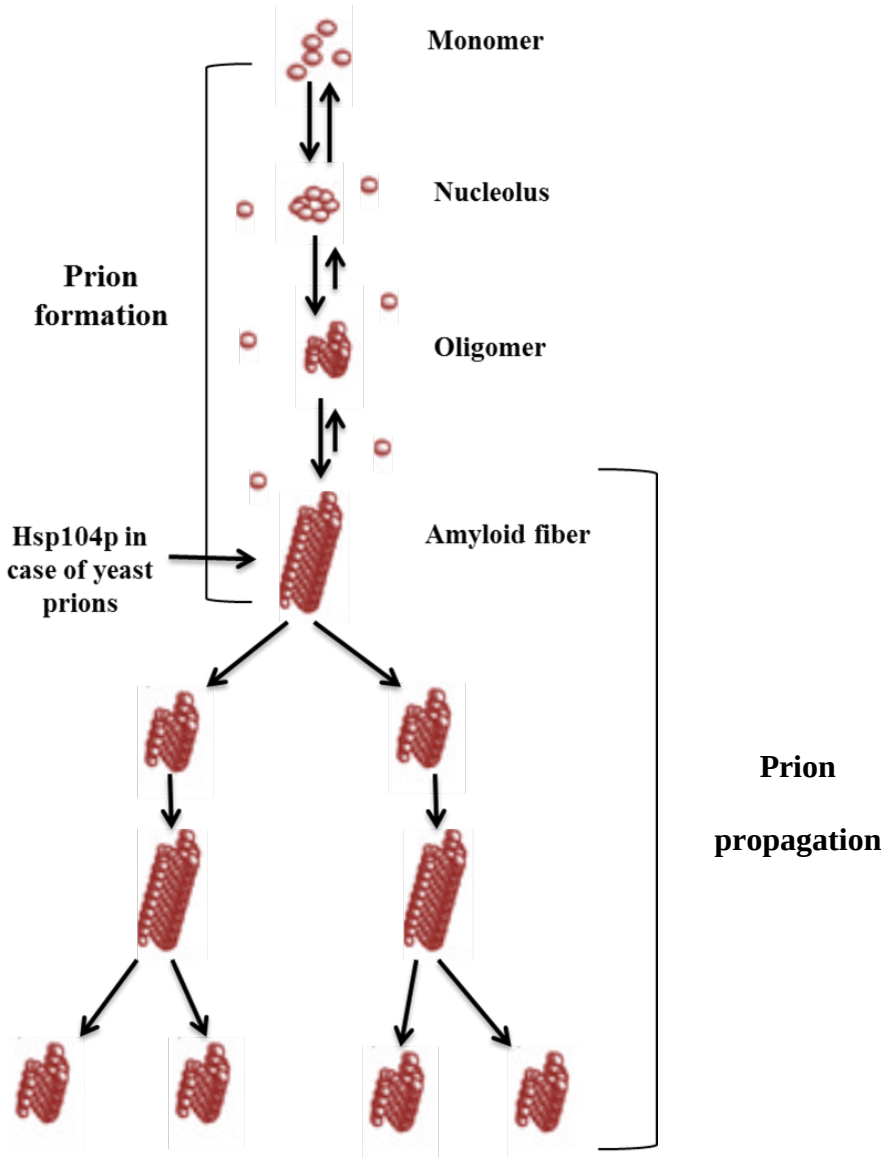


Figure1. 4: Schematic representation of prion formation and aggregation *De novo* prion formation starts with a nucleated polymerization reaction with a lag phase required to form an aggregation nucleus. This is follow by an elongation phase in which monomers are rapidly recruited into growing aggregates. Oligomers are common folding intermediates of amyloid fibrils. Efficient prion propagation followings the nucleation step. In yeast, the chaperone protein Hsp104 is important to sever long amyloid fiber into smaller propagons that are passed to daughter cells during cell division or mating in yeast. Essential mammalian prion propagation co-factors remain vague (Krammer, Schätzl and Vorberg, 2009).

The final step that required for prion activities is fiber fragmentation. Fragmentation is required to generate new independently segregating prion aggregates (seeds) to offset dilution by cell division. In the yeast, Hsp104 is responsible for prion fibers fragmentation.

Therefore, amyloid inheritance hinges on the cellular concentration Hsp104. Hsp104 is a member of the Adenosine triphosphotases Associated with diverse Activities family (Chernoff *et al*. 1995). In order for prion amyloids to be disseminated to the next generation, cells should have an appropriate amount of Hsp104, to fragment single large fibers into small propagons (Shorter and Lindquist, 2004) and initiate new rounds of prion propagation by providing free ends (Kushnirov and Ter-Avanesyan, 1998). Indeed, a decline in the cellular level of Hsp104 by gene down regulation leads to buildup of long fragments of amyloid fiber within the cell and failure to be transferred to daughter cells (Satpute-Krishnan *et al.*, 2003). In theory, excessive level of Hsp104 could also result in prion loss by fully fragmenting fibers back to their monomer subunits. In reality, such curing by over expression is seen for some yeast prions, including [PSI^+], but not for others. The prion splitting model proposes that in order for the amyloid fiber to be fragmented into small pieces, Hsp104 attacks the fiber by drawing out a single monomer from the middle of the fiber through the central pore of hexametric structure of Hsp104 in an ATP-dependent mechanism (Grimminger-Marquardt & Lashuel *et al*., 2010) Figure 1.5. However, if Hsp104 attacks the end of polymer, rather than fragmenting the fiber, it will simply release a soluble monomer from the fiber end. Therefore, for prions to be propagated, the on-rate must be faster than the off-rate.

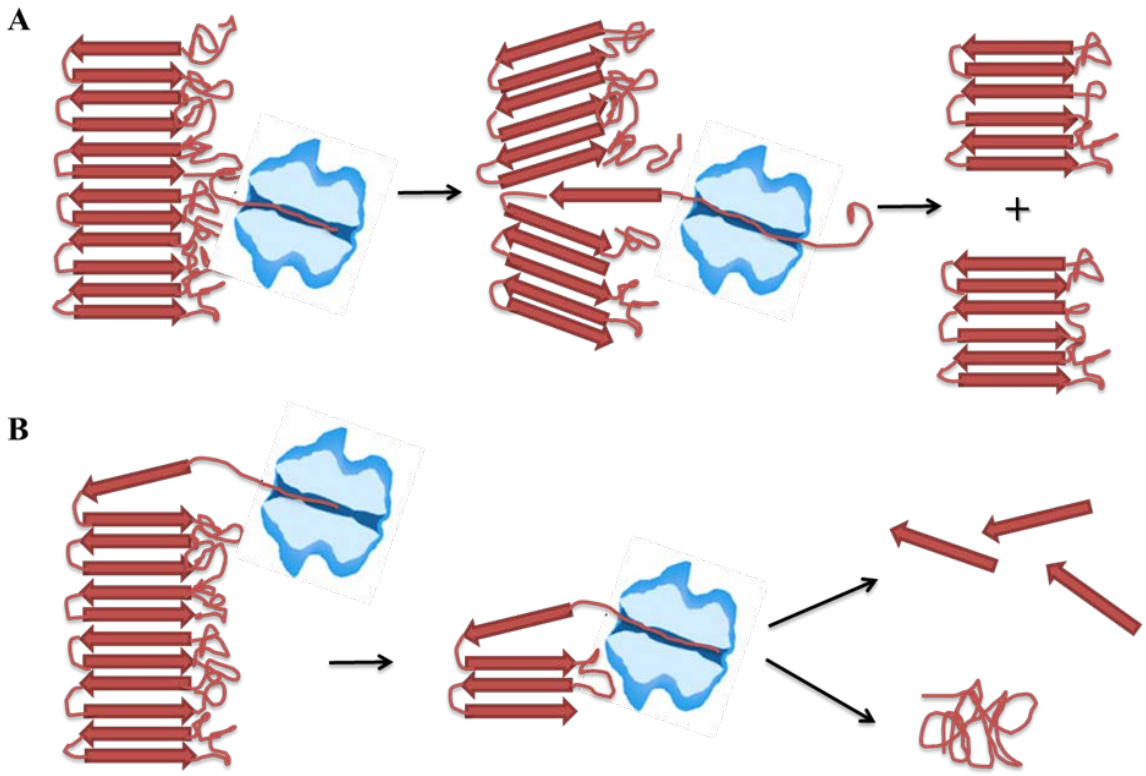


Figure1. 5: Mechanism of chaperone Hsp104 in prion propagation. **A**. Hsp104 mechanism for amyloid fiber severing activity. Hsp104 effectively yields amyloid seeds out of long fibrils by attacking the middle of long fiber and pulling one monomer result in splitting the fiber to prion propagons. This type of interaction is ATP dependent and requires optimum concentrations of active Hsp104. **B**. Hsp104 may also approach the end of amyloid fibrils, leading to fiber dissociation and monomer release. Complete dissociation only occurs if the on-rate of amyloid formation is lower than the off-rate or if released monomers are removed from the amyloid formation reaction by achieving their native, normal folding state. Figure modified from Grimminger-Marquardt and Lashuel, 2009.

Yet, our knowledge on the role of amino acid composition in driving amyloid fiber formation and propagation separately is still poorly understood, leading to poor prediction of prion propensity. Although it is clear that prion activity requires distinct prion formation and propagation steps, the sequence requirement for each step are poorly understood. Separately defining the sequence requirements for prion formation versus prion propagation will provide insight into the mechanisms underlying these two activities, and will improve prion prediction.

IV- Research model: the prion-forming protein Sup35

Yeast prion proteins particularly Sup35, have been extensively used in the prion field due to their simplicity, ease of genetic manipulation, high growth rate, and ease of movement between *in vivo* to *in vitro* assays. These features make the yeast prions influential models for studying the amino acid sequence requirement for prion formation and propagation. The soluble cytosolic yeast *Saccharomyces cerevisiae* protein Sup35, referred to as eRF3, allies with Sup45, eRF1, in a heterodimer that is responsible for translation termination (Stansfield, Jones *et al*. 1995; Zhouravleva, Frolova *et al*. 1995). The presence of [PSI^+], the prion conformation of Sup35, decreases the level of soluble Sup35. This leads to read through of stop codons, causing increase non-sense suppression (Patino, Liu *et al*. 1996; Paushkin, Kushnirov *et al*. 1996). The genetic and physical interactions between Sup35 and Sup45 have been demonstrated (Ebihara & Nakamura *et al*., 1999). Overexpression of Sup45 prevents de novo nucleation of [PSI^+], but has no effect on the propagation of [PSI^+] (Derkatch *et al*., 1998).

[PSI^+] formation can be detected by observation of non-sense suppression of the mutant ade2-1 allele (Patino, Liu *et al*. 1996; Paushkin, Kushnirov *et al*. 1996). Ade2 is required for adenine biosynthesis. Therefore, ade2 cells are unable to grow without adenine and grow red on limiting adenine due to accumulation of the substrate of the Ade2 enzyme. In [PSI^+] cells, the

premature stop codon is read through and full length functional Ade2 protein is synthesized, giving the cells the ability to maintain their normal white color on limiting adenine and grow in absence of adenine, Figure 1.6.

The Sup35 protein is segmented into three functionally and structurally distinct domains: a C-terminal domain (amino acids 254-685) that is important for the normal function of Sup35 as a translation termination factor; the prion-forming domain (PFD; amino acids1-114), which drives prion formation and propagation; and the highly charged middle domain, which has no known function other than its ability to stabilize [PSI^+] fibers (Ter-Avanesyan, Kushnirov *et al*. 1993), Figure 1.6. Like other yeast prion domains, the Sup35 PFD is rich of glutamine and asparagine (Q/N) and has low hydrophobic content. A wealth of data demonstrated that transient overexpression of PFD alone is able to induce the *de novo* appearance of [PSI^+], while deletion the same region leads to loss the nonsense suppression phenotype of Sup35, meaning that the PFD is required for prion activity (Derkatch *et al.*,1996; Ter-Avanesyan *et al.*, 1994).

Within the PFD, the first 39 residues, which are highly glutamine/asparagine-rich (Q/N-rich), appear to play an important role in the prion nucleation (DePace, Santoso *et al.* 1998; Osherovich, Cox *et al.* 2004). The following 57 residues, the oligopeptide repeat subdomain (ORD) facilitates Hsp104-dependent cleavage to maintain the propagons (Liu and Lindquist 1999; Osherovich, Cox *et al*. 2004). The ORD consists of 5 and half repeats of the consensus sequence (P/Q) QGGYQ (Q/S) YN (Crist, Nakayashiki *et al*. 2003). In fact, replacing the ORD repeat with random sequence or poly Qs, adding more Sup35 repeats, or deletion of one or two repeats, each has a dramatic effect on prion propagation either by aiding or preventing the prion propagation (Osherovich, Cox *et al.* 2004; Shkundina, Kushnirov *et al.* 2006; Crist, Nakayashiki *et al.* 2003; Ter-Avanesyan, Kushnirov *et al.* 1993; Liu and Lindquist 1999; Parham, Resende *et*

al. 2001; Osherovich, Cox *et al.* 2004). This indicates an important role of the ORD in facilitating the chaperone-dependent fiber cleavage (Osherovich, Cox *et al.* 2004), but the exact mechanism is still unclear. However, the yeast protein Ure2p is capable of forming and propagating prions efficiently without an ORD. Additionally, randomizing the order of amino acid in the Sup35 PFD does not block [PSI^+] formation and propagation (Ross, Edskes *et al.* 2005), suggesting that amino acid composition, not primary sequence, is the main determining factor of the prion phenotype. Interestingly, the nucleation domain and ORD have distinct compositional requirements for prion formation and propagation (Alexandrov *et al.*, 2008; Toombs, Ross *et al.*, 2011)

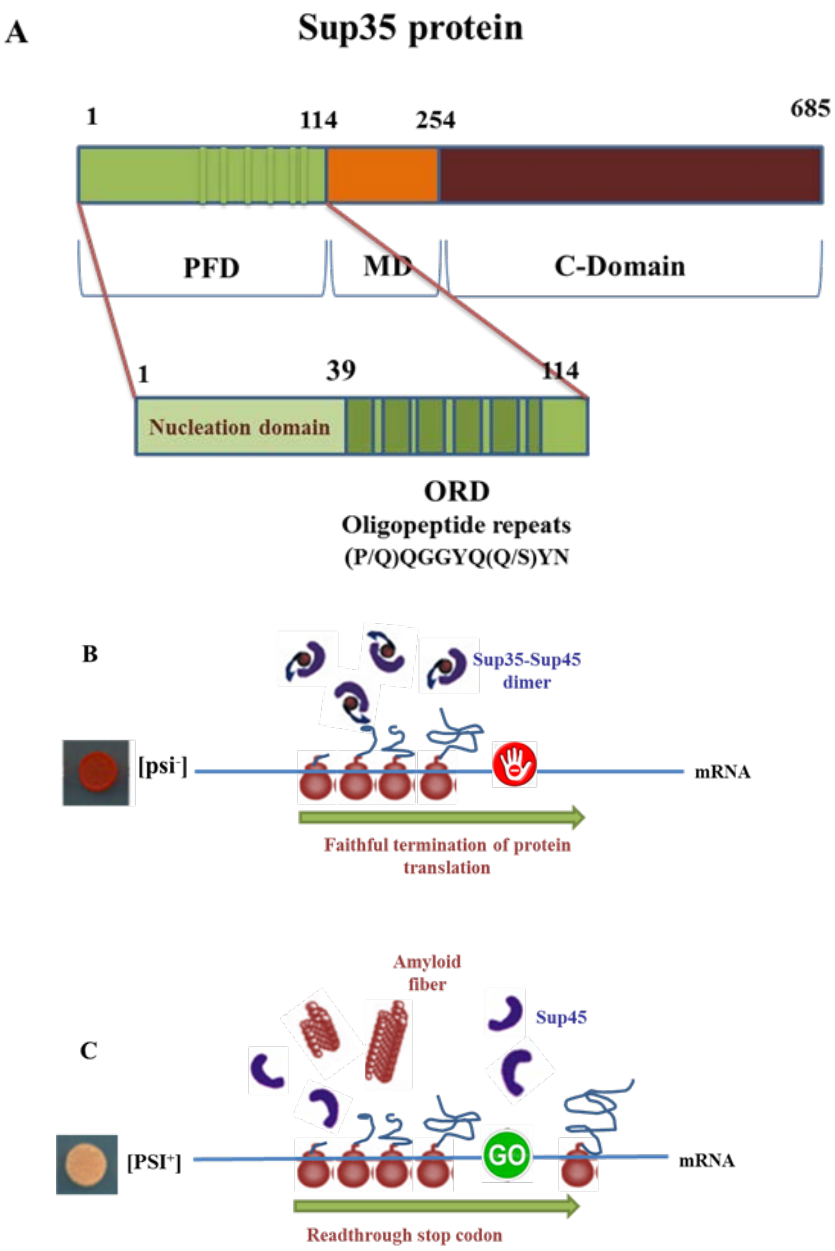


Figure1. 6: Molecular Structure of Sup35 protein. **A**. Molecular Structure of Sup35 protein. The PFD, the highly charged middle domain (MD) and the C-terminal translation termination domain (C-domain). The PFD is magnified to demonstrate the Q/N rich nucleation sub-domain and the oligopeptide repeat sub-domain (ORD), which consists of 5½ repeats of the consensus sequence (P/Q)QGGYQ(Q/S)YN. **B.** Function of the Sup35 in functional cell. The translation termination factor Sup35-Sup45 dimer is required for reliable termination of translation. The dimer stimulates ribosome to release a nascent polypeptide when approaches the one of stop codons UAG, UGA or UAA. **C**. In prion cells, Sup35 is sequestered in amyloid fibers. This leads to read-through of stop codons.

V- Attempts to predict prion activity

The finding of yeast prion forming proteins Sup35 and Ure2 stimulated many labs to direct their efforts towards identifying more prions. The unique compositional feature of these two prion proteins, particularly on their high Q/N content and low representation of hydrophobic residues, were widely used as platforms for these searches. Therefore, early bioinformatics studies looked for new prions based on compositional similarity to known prion forming domains. Three prion proteins have been discovered based on this method: New1 (Santoso *et al*, 2000*)*, Rnq1 (Sondheimer and Lindquist, 2000), and Mot3 (Alberti *et al*, 2009). Obviously, this method is effective at identifying potential prion candidates. Yet, it is poor at distinguish among these candidates. One eecent study identified 100 proteins that have domains that are compositionally similar to those of Sup35, Ure2, Rnq1, and New1 by using a Hidden Markov Model. These 100 proteins were tested for prion activity by using four different assays: An *in vitro* aggregation assay (ThT fluorescence) to screen fiber formation; SDD-AGE gel to monitor the formation of SDS-resistant aggregates *in vivo*; insertion of the domains in the place of the Sup35 PFD to test the ability to support prion formation and propagation; and *in vivo* aggregation of GFP fusion (Alberi *et al*, 2009). The study found 18 potential new prion candidates that passed all 4 assays. However, there was very poor correlation between the compositional similarity to the known yeast prions and prion propensity of these 100 prion-like domains. The most likely explanation for this failure is that the ability of compositional similarity algorithm is established on the assumption that all compositional deviations from the known yeast PFDs will reduce prion activity. In reality, this is not the case. Known PFDs are likely not optimized for maximal prion propensity. In fact, amino acids may be rare in prion domains either because they strongly inhibit prion activity (proline and charged residues) or

because too strongly promote prion formation (hydrophobic residues); therefore some compositional changes will increase (or decrease) prion propensity, and some will decrease prion propensity.

Since compositional changes can affect prion propensity, it is not surprising that simply counting the number of compositional changes from known prion domains would not be an accurate method for scoring prion activity. To more accurately predict prion activity, Toombs *et al* developed a method to score the prion propensity of each amino acid. The PFD of scrambled version Sup35 was targeted for random mutagenesis. An 8 codon segment was replaced with random oligopeptide sequence, thereby generating a library of mutants, each with different sequence in the random region. A prion-propensity score for each amino acid was measured by comparing the amino acid composition of the naïve library with the library of isolates that were able to form prions. A sliding window approach using a window size of 41 amino acids could then use to score the prion propensity of a region by summing the prion propensity scores for the individual amino acids across the window. Combining the experimentally determined prion propensity values with FoldIndex provided the most faithful prion propensity prediction algorithm. Toombs was able to develop an algorithm capable of distinguishing between Q/N-rich proteins with and without prion activity with 90% accuracy (Toombs *et al.*, 2010). This algorithm was later named Prion Aggregation Prediction Algorithm (PAPA).

Although PAPA is by far the most accurate prediction algorithm for Q/N-rich domains, it is still far from perfect. Recent study has support the hypothesis that the nucleation domain (residues 1-39) and ORD domain (residues 40-114) of Sup35 have distinct compositional requirements (Toombs *et al.*, 2011). Scrambling the order of Sup35 ORD has no effect on [PSI^+] propagation, while replacing the Sup35 ORD with scrambling version of Sup35 nucleation

domain blocks [PSI$^+$] propagation. Therefore, one obvious caveat with PAPA is that it does not separately score these two activities. Instead it simply scores overall prion activities. We hypothesized that by separately scoring prion formation and propagation propensity, we could improve the accuracy of PAPA to predict prion propensity of other proteins in yeast and beyond. Specifically, it will help to distinguish which Q/N rich domain proteins have ability to form prions and which do not.

VI- Prion like domains and diseases

With increased life expectancy has come an increasing number of neurodegenerative diseases, which is a serious danger to public health worldwide. Many neurodegenerative diseases, including Alzheimer disease, Huntington disease, Parkinson disease, Amyotrophic lateral sclerosis (ALS also known as Lou Gehrig's disease), frontemporal lobar degeneration (FTLD), and inclusion body myopathy with frontemporal dementia (IBMPFD) are associated with buildup of insoluble proteins mainly in the brain and characterized by unifying pathway of pathogenesis (Gitler and Shorter *et al.,* 2011). Recently, bioinformatics approaches have predicted the presence of prion-like domains (PrLDs) in more than two hundred human proteins (King *et al.,* 2012). Many of these are RNA-binding proteins and are characterized by having RNA recognition motif (RRM). In yeast, a PAPA score above 0.05 was generally associated with prion activity. A few of the human RRM-containing proteins scored above this threshold. Some of remaining are very close, such that a single point mutation in PrLDs could drive them above the threshold (King *et al.,* 2012). A handful of these proteins have recently been associated with several neurodegenerative diseases. TDP-43, FUS, TAF-15, hnRNPA1 and hnRNPA2B1 are involved in amyloid aggregation disorders of some cases of ALS and FTLD (Gitler *et al.,* 2011; Couthouis *et al.,* 2011).

Moreover, single point mutations in the RNA-binding proteins hnRNPA1 and hnRNPA2B1 have been reported in patients with IBMPFD coupled with ALS. These proteins have been examined by using both Alberti *et al* and PAPA to predict prion activity. PAPA scores the wild type proteins just below threshold for prion formation, while Alberti *et al* algorithm predicts that these proteins should form prions. Interestingly, PAPA correctly predicts that the disease-associated mutations should drive the aggregation propensity past the threshold for aggregation. By contrast, the Alberti *et al* algorithm predicts that these mutations should have little effect on aggregation. The disease-causing mutations have been tested in *Drosophila* and were sufficient to cause muscle degeneration (King *et al.*, 2012). Further examination of these two proteins using modular feature of yeast protein Sup35 has been made (Kim, Hong Joo, *et al*, 2013). It has been recommended that further investigation must be done as soon as possible on RNA-binding proteins that are not yet reported in association with neurodegenerative disease (King *et al*., 2012). Consequently, we have begun to investigate other RNA-binding proteins that have high predicted prion propensity, but have not yet been associated with diseases.

The principal aims of my research are to develop a better understanding of the fundamental features of prion formation and propagation and use this information to improve the ability of PAPA on identifying more prion forming candidates throughout genomic searches. The research addresses the questions of how sequence requirements differs from prion formation versus prion propagation and how amino acid composition, particularly hydrophobic amino acids within the ORD affects prion formation.

References

1. Alberti S, Halfmann R, King O, Kapila A, Lindquist S. (2009). A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. Cell 137:146-58.
2. Alexandrov IM, Vishnevskaya AB, Ter-Avanesyan MD, Kushnirov VV. (2008). Appearance and propagation of polyglutamine-based amyloids in yeast: tyrosine residues enable polymer fragmentation. J Biol Chem; 283:15185-92.
3. Baxa, U., Speransky, V., Steven, A. C., & Wickner, R. B. (2002). Mechanism of inactivation on prion conversion of the Saccharomyces cerevisiae Ure2 protein. Proceedings of the National Academy of Sciences, 99(8), 5253-5260.
4. Brachmann, A., Baxa, U., & Wickner, R. B. (2005). Prion generation in vitro: amyloid of Ure2p is infectious. The EMBO journal, 24(17), 3082-3092.
5. Brandner, S., Raeber, A., Sailer, A., Blättler, T., Fischer, M., Weissmann, C., & Aguzzi, A. (1996). Normal host prion protein (PrPC) is required for scrapie spread within the central nervous system. Proceedings of the National Academy of Sciences, 93(23), 13148-13151.
6. Büeler, H., Aguzzi, A., Sailer, A., Greiner, R. A., Autenried, P., Aguet, M., & Weissmann, C. (1993). Mice devoid of PrP are resistant to scrapie. Cell, 73(7), 1339.
7. Caughey, B., Baron, G. S., Chesebro, B., & Jeffrey, M. (2009). Getting a grip on prions: oligomers, amyloids and pathological membrane interactions. Annual review of biochemistry, 78, 177.

8. Chernoff, Y. O., Derkach, I. L., & Inge-Vechtomov, S. G. (1993). Multicopy SUP35 gene induces de-novo appearance of psi-like factors in the yeast *Saccharomyces cerevisiae*. Current genetics, 24(3), 268-270.
9. Chernoff, Y. O., Inge-Vechtomov, S. G., Derkach, I. L., Ptyushkina, M. V., Tarunina, O. V., Dagkesamanskaya, A. R., & Ter-Avanesyan, M. D. (2004). Dosage-dependent translational suppression in *yeast Saccharomyces cerevisiae*. Yeast, 8(7), 489-499.
10. Chernoff, Y. O., Lindquist, S. L., Ono, B. I., Inge-Vechtomov, S. G., & Liebman, S. W. (1995). Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. Science (New York, NY), 268(5212), 880.
11. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., & Dobson, C. M. (1999). Designing conditions for in vitro formation of amyloid protofilaments and fibrils. Proceedings of the National Academy of Sciences, 96(7), 3590-3594.
12. Colby, D. W., & Prusiner, S. B. (2011). Prions. Cold Spring Harbor Perspectives in Biology, 3(1).
13. Cooper, T. G. (2006). Transmitting the signal of excess nitrogen in Saccharomyces cerevisiae from the Tor proteins to the GATA factors: connecting the dots. FEMS microbiology reviews, 26(3), 223-238.
14. Coschigano, P. W., & Magasanik, B. O. R. I. S. (1991). The URE2 gene product of Saccharomyces cerevisiae plays an important role in the cellular response to the nitrogen source and has homology to glutathione s-transferases. Molecular and cellular biology, 11(2), 822-832.

15. Coustou, V., C. Deleu, et al. (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. USA 94(18): 9773-8.
16. Couthouis, J., Hart, M. P., Shorter, J., DeJesus-Hernandez, M., Erion, R., Oristano, R., & Gitler, A. D. (2011). A yeast functional screen predicts new candidate ALS disease genes. Proceedings of the National Academy of Sciences, 108(52), 20881-20890.
17. Cox BS (1965). [PSI], a cytoplasmic suppressor of super-suppressors in yeast. Heredity 20:505–521
18. 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(1), 23-33.
19. Crist, C. G., T. Nakayashiki, et al. (2003). "[PHI+], a novel Sup35-prion variant propagated with non-Gln/Asn oligopeptide repeats in the absence of the chaperone protein Hsp104." Genes Cells 8(7): 603-18.
20. DePace, A. H. and J. S. Weissman (2002). "Origins and kinetic consequences of diversity in Sup35 yeast prion fibers." Nat. Struct. Biol. 9(5): 389-96.
21. DePace, A. H., A. Santoso, et al. (1998). "A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion." Cell 93(7): 1241-52.
22. Derkatch, I. L., Bradley, M. E., & Liebman, S. W. (1998). Overexpression of the SUP45 gene encoding a Sup35p-binding protein inhibits the induction of the de novo appearance of the [PSI+] prion. Proceedings of the National Academy of Sciences, 95(5), 2400-2405.

23. Derkatch, I. L., Chernoff, Y. O., Kushnirov, V. V., Inge-Vechtomov, S. G., & Liebman, S. W. (1996). Genesis and variability of [PSI] prion factors in Saccharomyces cerevisiae. Genetics, 144(4), 1375-1386.

24. Donne, D. G., Viles, J. H., Groth, D., Mehlhorn, I., James, T. L., Cohen, F. E., ... & Dyson, H. J. (1997). Structure of the recombinant full-length hamster prion protein PrP (29–231): the N terminus is highly flexible. Proceedings of the National Academy of Sciences, 94(25), 13452-13457.

25. Eaglestone, S. S., Cox, B. S., & Tuite, M. F. (1999). Translation termination efficiency can be regulated in Saccharomyces cerevisiae by environmental stress through a prion-mediated mechanism. The EMBO journal, 18(7), 1974-1981.

26. Ebihara, K., & Nakamura, Y. O. S. H. I. K. A. Z. U. (1999). C-terminal interaction of translational release factors eRF1 and eRF3 of fission yeast: G-domain uncoupled binding and the role of conserved amino acids. RNA, 5(6), 739-750.

27. Eisenberg, David, and Mathias Jucker. (2012). "The amyloid state of proteins in human diseases." Cell 148.6, 1188-1203.

28. Fischer M, Rulicke T, Raeber A, Sailer A, Moser M, et al. (1996) Prion protein (PrP) with amino-proximal deletions restoring susceptibility of PrP knockout mice to scrapie. Embo J 15: 1255–1264.

29. Gitler, A. D., & Shorter, J. (2011). RNA-binding proteins with prion-like domains in ALS and FTLD-U. Prion, 5(3), 179-187.

30. Greenwald, J., & Riek, R. (2010). Biology of amyloid: structure, function, and regulation. Structure, 18(10), 1244-1260.

31. Grimminger-Marquardt, V., & Lashuel, H. A. (2010). Structure and function of the molecular chaperone Hsp104 from yeast. Biopolymers, 93(3), 252-276.

32. Halfmann, R., Jarosz, D. F., Jones, S. K., Chang, A., Lancaster, A. K., & Lindquist, S. (2012). Prions are a common mechanism for phenotypic inheritance in wild yeasts. Nature, 482(7385), 363-368.

33. Jaroniec, C. P., MacPhee, C. E., Astrof, N. S., Dobson, C. M., & Griffin, R. G. (2002). Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril. Proceedings of the National Academy of Sciences, 99(26), 16748-16753.

34. Kim, H. J., Kim, N. C., Wang, Y. D., Scarborough, E. A., Moore, J., Diaz, Z., ... & Taylor, J. P. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature.

35. King, O. D., Gitler, A. D., & Shorter, J. (2012). The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain research, 1462, 61-80.

36. Kushnirov, V. V. & Ter-Avanesyan, M. D. (1998). “Structure and Replication of Yeast Prions."Cell 94, 13–16.

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

38. Laurent, M. (1996). Prion diseases and the'protein only'hypothesis: a theoretical dynamic study. Biochemical Journal, 318(Pt 1), 35.

39. Liu, J. J. and S. Lindquist (1999). "Oligopeptide-repeat expansions modulate 'proteinonly' inheritance in yeast." Nature 400(6744): 573-6.

40. Lund, P. M., & Cox, B. S. (1981). Reversion analysis of [psi⁻] mutations in *Saccharotnyces cerevisiae*. Genet. Res., Camb, 37, 173-182.

41. Markowitz, V. M., Mavromatis, K., Ivanova, N. N., Chen, I., Min, A., Chu, K., & Kyrpides, N. C. (2009). IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics, 25(17), 2271-2278.

42. Osherovich, L. Z., B. S. Cox, et al. (2004). "Dissection and design of yeast prions." PLoS Biol. 2(4): E86.

43. Parham, S. N., C. G. Resende, et al. (2001). "Oligopeptide repeats in the yeast protein Sup35p stabilize intermolecular prion interactions." EMBO J. 20(9): 2111-9.

44. Patino, M. M., Liu, J. J., Glover, J. R., & Lindquist, S. (1996). Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. SCIENCE-NEW YORK THEN WASHINGTON-, 622-626.

45. Paushkin, S. V., Kushnirov, V. V., Smirnov, V. N., & Ter-Avanesyan, M. D. (1996). Propagation of the yeast prion-like [psi+] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor. The EMBO Journal, 15(12), 3127.

46. Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., & Tycko, R. (2002). A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. Proceedings of the National Academy of Sciences, 99(26), 16742-16747.

47. Prusiner, S. B. (1982). Novel proteinaceous infectious particles cause scrapie. Science (New York, NY), 216(4542), 136.

48. Prusiner, S. B. (1998). Prions. Proceedings of the National Academy of Sciences, 95(23), 13363-13383.

49. Prusiner, S. B. (2004). An introduction to prion biology and diseases. Cold Spring Harbor Monograph Series, 41, 1-88.

50. Riek, R., Hornemann, S., Wider, G., Glockshuber, R., & Wüthrich, K. (1997). NMR characterization of the full-length recombinant murine prion protein, mPrP (23-231). FEBS letters, 413(2), 282-288.

51. Ritson, G. P., Custer, S. K., Freibaum, B. D., Guinto, J. B., Geffel, D., Moore, J., ... & Taylor, J. P. (2010). TDP-43 mediates degeneration in a novel Drosophila model of disease caused by mutations in VCP/p97. The Journal of Neuroscience, 30(22), 7729-7739.

52. Ritter, C., Maddelein, M. L., Siemer, A. B., Lührs, T., Ernst, M., Meier, B. H., ... & Riek, R. (2005). Correlation of structural elements and infectivity of the HET-s prion. Nature, 435(7043), 844-848.

53. Ross, E. D., H. K. Edskes, et al. (2005). "Primary sequence independence for prion formation." Proc. Natl. Acad. Sci. USA 102(36): 12825-12830.

54. Roucou, X., & LeBlanc, A. C. (2005). Cellular prion protein neuroprotective function: implications in prion diseases. Journal of molecular medicine, 83(1), 3-11.

55. Santoso A, Chien P, Osherovich LZ, Weissman JS (2000). "Molecular basis of a yeast prion species barrier." Cell 100:277-88.

56. Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., ... & Eisenberg, D. (2007). Atomic structures of amyloid cross-&bgr; spines reveal varied steric zippers. Nature, 447(7143), 453-457.

57. Scheibel, T. and S. L. Lindquist (2001). "The role of conformational flexibility in prion propagation and maintenance for Sup35p." Nat. Struct. Biol. 8(11): 958-62.

58. Scheibel, T., A. S. Kowal, et al. (2001). "Bidirectional amyloid fiber growth for a yeast prion determinant." Curr Biol 11(5): 366-9.

59. Serio, T. R., A. G. Cashikar, et al. (2000). "Nucleated conformational conversion and the replication of conformational information by a prion determinant." Science 289(5483): 1317-21.

60. Shkundina, I. S., V. V. Kushnirov, et al. (2006). "The role of the N-terminal oligopeptide repeats of the yeast Sup35 prion protein in propagation and transmission of prion variants." Genetics 172(2): 827-35.

61. Shorter, J. and S. Lindquist (2004). "Hsp104 catalyzes formation and elimination of self-replicating Sup35 prion conformers." Science 304(5678): 1793-7.

62. Sipe JD, Benson MD, Buxbaum JN, et al (2012). Amyloid fibril protein nomenclature: 2010 recommendations from the nomenclature committee of the International Society of Amyloidosis. *Amyloid*.19:167–170.

63. Sipe, Jean D., and Alan S. Cohen. (2000). "Review: history of the amyloid fibril." Journal of structural biology 130.2, 88-98.

64. Sondheimer N, Lindquist S. (2000) "Rnq1: an epigenetic modifier of protein function in yeast." Mol Cell 5:163-72.

65. Stansfield, I., Jones, K. M., Kushnirov, V. V., Dagkesamanskaya, A. R., Poznyakovski, A. I., Paushkin, S. V., ... & Tuite, M. F. (1995). The products of the SUP45 (eRF1) and SUP35 genes interact to mediate translation termination in *Saccharomyces cerevisiae.* The EMBO journal, 14(17), 4365.

66. Tanaka, M., Chien, P., Naber, N., Cooke, R., & Weissman, J. S. (2004). Conformational variations in an infectious protein determine prion strain differences. Nature, 428(6980), 323-328.

67. Ter-Avanesyan, M. D., Dagkesamanskaya, A. R., Kushnirov, V. V., & Smirnov, V. N. (1994). The Sup35 Omnipotent Suppressor Gene Is Involved in the Maintenance of the Non-Mendelian Determinant [PSI$^+$] in the Yeast Saccharomyces Cerevisiae. Genetics, 137(3), 671-676.

68. Ter-Avanesyan, M. D., V. V. Kushnirov, et al. (1993). "Deletion analysis of the SUP35 gene of the yeast Saccharomyces cerevisiae reveals two non-overlapping functional regions in the encoded protein." Mol. Microbiol. 7(5): 683-92.

69. Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., ... & Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial “pan-genome”. Proceedings of the National Academy of Sciences of the United States of America, 102(39), 13950-13955.

70. Toombs, J. A., Liss, N. M., Cobble, K. R., Ben-Musa, Z., & Ross, E. D. (2011). [PSI+] Maintenance Is Dependent on the Composition, Not Primary Sequence, of the Oligopeptide Repeat Domain. PloS one, 6(7), e21953.

71. True, H. L., Berlin, I., & Lindquist, S. L. (2004). Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. Nature, 431(7005), 184-187.

72. Tuite, M. F., & Cox, B. S. (2003). Propagation of yeast prions. Nature Reviews Molecular Cell Biology, 4(11), 878-890.

73. Tuite, M. F., & Cox, B. S. (2009). Prions remodel gene expression in yeast. Nature cell biology, 11(3), 241-243.

74. Tuite, M. F., Mundy, C. R., & Cox, B. S. (1981). Agents that cause a high frequency of genetic change from [psi$^+$] to [psi$^-$] in *Saccharomyces cerevisiae*. Genetics, 98(4), 691-711.

75. Tuite, M., Marchante, R., & Kushnirov, V. (2011). Fungal prions: structure, function and propagation. Prion Proteins, 257-298.

76. Tyedmers, J., Madariaga, M. L., & Lindquist, S. (2008). Prion switching in response to environmental stress. PLoS biology, 6(11), e294.

77. Van der Wel, P. C., Lewandowski, J. R., & Griffin, R. G. (2010). Structural characterization of GNNQQNY amyloid fibrils by magic angle spinning NMR. Biochemistry, 49(44), 9457.

78. Wickner RB (1994). [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. Science 264:566–569

79. Wickner, R. B., & Chernoff, Y. O. (1999). 6 Prions of Fungi:[URE3],[PSI], and [Het-s] Discovered as Heritable Traits. Cold Spring Harbor Monograph Archive, 38, 229-272.

80. Wickner, Reed B., et al. (2010)"Prion amyloid structure explains templating: how proteins can be genes." FEMS yeast research 10.8, 980-991.

81. Wiltzius, J. J., Landau, M., Nelson, R., Sawaya, M. R., Apostol, M. I., Goldschmidt, L., ... & Eisenberg, D. (2009). Molecular mechanisms for protein-encoded inheritance. Nature structural & molecular biology, 16(9), 973-978.

82. Zahn, R., Liu, A., Lührs, T., Riek, R., Von Schroetter, C., García, F. L., & Wüthrich, K. (2000). NMR solution structure of the human prion protein. Proceedings of the National Academy of Sciences, 97(1), 145-150.

83. Zakharov, I. A., & Yarovoy, B. P. (1977). Cytoduction as a new tool in studying the cytoplasmic heredity in yeast. Molecular and Cellular Biochemistry, 14(1), 15-18.

84. Zhouravleva, G., Frolova, L., Le Goff, X., Le Guellec, R., Inge-Vechtomov, S., Kisselev, L., & Philippe, M. (1995). Termination of translation in eukaryotes is governed by two

interacting polypeptide chain release factors, eRF1 and eRF3. The EMBO journal, 14(16), 4065.

Chapter 2: Different Amino Acid Composition Requirements for Prion Formation and Propagation in the [PSI$^+$] Yeast Prion[1]

Summary

Misfolding of a wide range of proteins leads to formation of amyloid fibrils: ordered, β-sheet-rich protein aggregates. Most of the known yeast prion proteins contain glutamine and/or asparagine (Q/N) rich domains that drive amyloid formation. Q/N-rich domains are found in 1-4% of the proteins in most eukaryotic proteomes, but very few of these proteins have been shown to undergo amyloid structural conversion. Use of bioinformatic screens for prions in yeast on the basis of Q/N-richness has had some notable successes. Despite advances in predicting which Q/N-rich domains may turn out to be *bona fide* prions in yeast and *in vitro*, predictions remain imperfect. To make better predictions of prion propensity requires a better understanding of the distinct and poorly understood factors that separately affect amyloid formation and propagation. Yeast prion (infectious proteins) formation is driven primarily by amino acid composition. Previously, we have developed a novel *in vivo* assay to quantitatively assess how composition affects prion formation. Unfor[1]tunately, the down-side of this work was unable to separately assess how composition affects prion formation versus prion propagation. Here we extended the previous method by performing an inclusive investigation to define the compositional features that promote prion formation versus prion propagation.

[1] "A modified version of this work is being prepared for publication as: Kyle S. MacLea, Zobaida Ben-Musa, Kacy R. Paul, Margaret Gruca, and Eric D. Ross. "Different Amino Acid Composition Requirements for Prion Formation and Propagation in the [PSI+] Yeast Prion."

Introduction

Amyloid fibers formation has been progressively implicated with human and animal diseases including Creutzfeldt–Jacob, Transmissible encephalopathy , Scrapie, Chronic wasting diseases, noninfectious neurodegenerative diseases such as Alzheimer, Parkinson and Huntington diseases. The causative agent that accountable for these diseases rises from proteins that have the ability to convert from soluble, functional conformation into insoluble amyloid aggregates.

Amyloid fibrils are organized, aggregated protein with high β-sheet content, protease resistance, filamentous morphology and yellow-green birefringence upon staining with Congo red (Kisilevsky and Fraser 1997; Sipe and Cohen 2000). Most amyloid fibers are not infectious, however, only a small subset of amyloids (called prions), can be transferred from cell to another vertically and horizontally including the causative agents of TSEs in mammals and [URE3], [PSI^+], [PIN^+], and others in *Saccharomyces cerevisiae*.

A well-studied model prion protein from *S. cerevisiae* is [PSI^+], the prion form of the translational terminator protein Sup35 (Wickner *et al.* 1994). Like other yeast prion proteins, Sup35 is modular, containing a distinct prion-forming domain (PFD), middle domain (M), and C-terminal domain (C) (Ter-Avanesyan *et al.* 1993; Ter-Avanesyan *et al.* 1994; DePace *et al.* 1998 ; Liu J-J *et al*. 2002) . The PFD (amino acids 1-114) drives the conversion of Sup35 into its amyloid form (Ter-Avanesyan *et al.* 1994), while the charged M domain has no known function other than its ability to stabilize [PSI^+] fibers (Ter-Avanesyan, Kushnirov *et al*. 1993) and the C domain is an essential component responsible for translational termination (Ter-Avanesyan *et al*. 1993; Liu J-J *et al*. 2002). Mutational studies of the PFDs of Sup35 and Ure2 have shown that amino acid composition not primary sequence is the predominant driving force behind prion

formation (Ross, Baxa *et al.* 2004; Ross, Edskes *et al.* 2005). We previously used a quantitative mutagenesis method to determine the prion propensity of each amino acid in the context of Q/N-rich PFDs (Toombs 2010). These prion propensity values were used to develop PAPA (Prion Aggregation Prediction Algorithm), the first prediction algorithm capable of accurately distinguishing between Q/N-rich domains with and without prion activity (Toombs *et al.* 2010, 2012; Ross *et al.* 2013).

Although PAPA represents a significant advance in prion prediction, it is far from perfect. One likely problem is that there are multiple distinct steps required for prion activity. Specifically, prion formation requires that a protein be able to both form prion aggregates and add onto these aggregates; additionally, prion propagation over multiple generations requires that these aggregates must be fragmented to create new independently segregating prion seeds to offset dilution by cell division (Chernoff *et al.* 1995; Derdowski *et al.* 2010). Each of these steps may have distinct amino acid sequence requirements, yet PAPA uses only a single prion propensity score for each amino acid. To make better predictions of prion propensity requires a better understanding of how amino acid sequence separately affects prion formation and propagation.

Sup35 is an ideal substrate for examining these sequence requirements. Within the Sup35 PFD are two sub-domains with overlapping but separate functions. Amino acids 1-39 form a Q/N-rich tract primarily responsible for nucleation and growth of prion fibers (DePace *et al.* 1998; Osherovich *et al.* 2004). The remaining portion (amino acids 40-114) has been previously implicated in prion propagation and contains the oligopeptide repeat domain (ORD), a region containing 5 and a half imperfect repeats with the consensus sequence (P/Q)QGGYQ(Q/S)YN (DePace *et al.* 1998 ;Liu J-J *et al.* 2002; Liu JJ *et al.* 1999 ; Shkundina

et al. 2006; Parham *et al.* 2001). Importantly, the Q/N-rich tract and the ORD appear to have distinct compositional requirements for their respective functions (Toombs *et al.* 2011)

Although the boundaries of the nucleation and propagation domains are not absolute (Osherovich *et al*. 2004; Toombs *et al*. 2010; Ohhashi *et al.* 2010), there is significant evidence to support the role of the ORD in promoting prion propagation, particularly the last 3 and a half repeats (Osherovich *et al*. 2004; Toombs *et al*. 2011). Removal of all or part of the ORD (Ter-Avanesyan *et al*. 1993; Osherovich *et al*. 2004; Liu JJ *et al*. 1999; Parham *et al.* 2001) or replacement of the ORD with a random sequence (Crist *et al.* 2003) destabilizes [PSI^+]. Such mutations appear to reduce prion aggregate fragmentation, resulting in larger aggregate size. Larger sizes are frequently lost as a result of imperfect segregation of aggregates into daughter cells (Derdowski *et al.* 2010). Hsp104, a yeast chaperone protein, has been shown to be essential for prion propagation (Chernoff *et al.* 1995) and does so by cleaving prion fibers into smaller fragments better suited to segregation into daughter cells (Paushkin *et al.* 1996; Wegrzyn *et al*. 2001; Ness *et al.* 2002) . Therefore, the ORD repeats have been hypothesized to facilitate for Hsp104-dependent aggregate cleavage; the repeats could act as Hsp104 binding-sites (although a site also exists in the M domain (Helsen *et al*. 2012), or by conformationally modify the amyloid core to allow chaperone access, or modulate fiber fragility (Shkundina *et al*. 2006; Alexandrov *et al*. 2012). Interestingly, the mammalian protein PrP also contains an ORD, with five repeats of consensus sequence (PHGGGWGQ) (Brown and Qin *et al.* 1997). PrP repeat expansion is associated with dominant inherited prion disease (Wadsworth *et al.* 2003; Prusiner *et al*. 1998) and removal of the repeats in a mouse model of disease slows progression (Flechsig *et al*. 2000). These results, combined with the presence of repeat elements in other yeast prion domains, Rnq1 and New1 (Osherovich *et al*. 2000; Vitrenko *et al*. 2007) suggests a role for repeats in prion

formation or propagation; however, other yeast prion proteins, such as Ure2, do not contain repeats, so repeats cannot be an absolutely necessary feature for prion. Furthermore, scrambling the Sup35 ORD does not prevent prion formation or maintenance (Toombs *et al*. 2011), indicating that the activity of the repeats is largely primary-sequence independent.

The sequence determinants of ORD function in prion propagation have been explored to a much lesser degree than the elements of Sup35 that promote prion nucleation. To date, the effects of mutagenesis on prion propagation in yeast have been explored through single- or double-residue targeted mutations (Alexandrov *et al*. 2012; 2008). It was thought that the elevated number and regular spacing of aromatic tyrosine residues in the ORD region and in the repeats of New1 might indicate recognition sites for chaperones such as Hsp104, since some chaperones are known to use exposed aromatic or hydrophobic residues as binding sites (Rüdiger *et al*. 1997; 2001). Using artificial polyglutamine-amyloids, targeted replacement of G with Y residues (Alexandrov *et al*. 2008) or other aromatic residues (Alexandrov *et al*. 2012) increased fiber fragmentation by an unknown mechanism that was Hsp104-independent. A recent study has also shown that targeted single or double mutations at a few sites in the natural ORD of Sup35 could affect the degree of fiber fragmentation (Marchante *et al.* 2013). In that study, certain single substitutions of position G58 in the second repeat of the Sup35 ORD (with, for example, aromatic residues) were associated with no loss of propagation ability, while substitution with non-aromatic hydrophobic residues, charged residues, or proline, created instability in prion maintenance (Marchante *et al.* 2013). Similar changes in the 1st, 4th, or 5th repeats of the ORD had no effect on propagation. While consistent with the polyglutamine protein mutation results in terms of the importance of aromatic residues in prion propagation, these results rely upon substitutions in the 2nd repeat of the ORD for their main conclusions, a

region of the protein known to also contribute to nucleation of the [PSI$^+$] prion (Osherovich *et al*. 2004). As a result, these mutations could be impacting both formation and propagation of prions and another assay is necessary to confirm if this biochemistry is important in the pure propagation of prions as well.

To perform a more comprehensive analysis of the compositional determinants for prion formation and propagation, we took advantage of the distinct functional roles of Sup35's Q/N tract and ORD. Using our previously developed method, we measured the prion propensities of each amino acid *in vivo* in the context of the Q/N-tract and ORD. We observed distinct compositional biases in these two domains. To confirm that these differences were due to distinct compositional biases for prion formation and propagation, we developed a new method to specifically isolate the effects of amino acid composition on prion propagation. These studies confirmed that nucleation and propagation of prions have overlapping but non-identical compositional requirements. Further, these differences allow us to make predictions about prions that should propagate well, design *de novo* mutant forms that are well maintained in yeast, and extend previously described datasets with *in silico* analysis confirming expected propagation behaviors.

Materials and methods

Strains and media.

Standard yeast media and methods were used, as described previously (Sherman, 1991), except that yeast extract-peptone-dextrose (YPD) media contained 0.5% yeast extract in place of the standard amount (1%). In all experiments, growth of yeast was at 30°C using *Saccharomyces cerevisiae.* Strain YER648 was used as the transformation host for 3R propagation mutant library formation (see below). YER648 was created from a strain (YER282

that was previously selected for loss of a *URA3* plasmid) with this starting genotype: a *kar1-1 SWQ5 ade2-1 his3 leu2 trp1 ura3 arg1*::*HIS3 sup35*::KanMx. To create strain YER648, a scrambled version of *SUP35* driven by the native *SUP35* promoter, pER188, was amplified with EDR302 (GGCAGAATATCTGTCAACCACAC) and EDR304 (GTTTCGTACTCACCCTTTCTGG) and this fragment was transformed into YER282 in the presence of *Aat*II/*Hin*dIII-cut pJ533 plasmid (Song *et al*. 2005). The mating strain used in a subsequent step of the creation of the mutant library (see below) was 780-1D/pJ533 (Song *et al*. 2005). The genotype of this strain is α *kar1-1 SUQ5 ade2-1 his3 leu2 trp1 ura3 sup35*::KanMx [PSI^+] [PIN^+]; pJ533 expresses, from a *URA3* plasmid, *SUP35* as the sole copy of SUP35 in the cell. Strain YER709 was used as the transformation host for nucleation and 3R prion formation mutant libraries. YER709 is mutant strain that we created fromYER635 (strong [PSI^+] strain that we also created by knockout Ppq1 gene) expresses SUP35MC on a URA3 from the SUP35 promoter (pER589 contains no SUP35 PFD with this starting genotype*: a kar1-1 SUQ5 ade2-1 his3 leu2 trp1 ura3 ppq1::HIS3 sup35::KanMx.*

Mutant library creation of the 3R libraries (formation or propagation) by PCR.

Both the 3R formation (Tables 3-4) and propagation experiments (Tables 5-7) used the same PCR strategy for generation of the mutant libraries. Degenerate oligonucleotides were used to randomly mutate the wild-type 3rd repeat of the ORD (wild-type amino acid sequence PDAGYQQQYN) in SUP35. Primer EDR1377 (CTGGGTACCAACAAGGTGGCTATCAACAGTACAAT$(NNB)_{10}$CCTCAAGGAGGCTACC AGCAATACAAC) was a sense primer, made by Invitrogen, containing a degenerate segment with a 25% mix of each nucleotide at positions 1 and 2 of each mutated codon, and a 33% mix of C, G, and T at the third position. This primer was paired with EDR262

(GCATCAGCACTGGTAACATTGG) to amplify the C-terminal region of the SUP35 PFD and middle domain and then reamplified with EDR1394 (CTGGGTACCAACAAGGTGGCTATC) and EDR262. In a separate PCR, a primer complementary to the nondegenerate 5' region of EDR1377, EDR1385 (GATAGCCACCTTGTTGGTACCCAG), was paired with EDR257 (GACCATGATTACGCCAAGCTC) to amplify the N-terminal region of the SUP35 PFD.

Mutant library creation of the ND library by PCR.

The region of the SUP35 prion forming domains, amino acids 21-28 (ND- wild-type amino acid sequence NQQQGNNR) was replaced with the random sequence (NNB). EDR1388 (CAAGGCTACCAGGCTTACAATGCTCAAGCCCAACCTGCAG) was sense primer that paired with EDR304 (GTTTCGTACTCACCCTTTCTGG) to pre-amplify C-terminal region of the SUP35 PFD and middle domain for mutagenesis of the nucleation domain (ND). And then reamplified with EDR1380 GCAAAACTACCAGCAATACAGCCAGAACGGT $(NNB)_8$ TACCAAGGCTACCAGGCTTACAATGC, sense primers, made by Invitrogen (as above). This primer was paired with EDR304 to mutate 8 residues in ND of SUP35. EDR1389 (GTTCTGGCTGTATTGCTGGTAGTTTTGCTGATTGTTGCCTTGGTTTGAATCC) was primer complementary to the nondegenerate 5' region of allowing for 2-piece PCR. In a separate PCR, this primer was paired with EDR302 (GGCAGAATATCTGTCAACCACAC) to amplify the N-terminal region of the SUP35 PFD.

Transformation conditions for generation of the ND prion formation and 3R prion formation and propagation libraries

SUP35 mutant versions were generated using a two-step PCR procedure in which regions 5' and 3' to the site of mutagenesis were amplified in separate reactions. The individual products of these reactions were combined and reamplified with EDR302 and EDR262. To

insert the combined PCR products into yeast under the control of the native SUP*35* promoter, transformation was undertaken in the presence of *Bam*HI/*Hin*dIII-cut pJ526 (*cen LEU2*; from Dan Masison, National Institutes of Health) into yeast strain YER709(ND and 3R prion formation) and YER648 (3R prion propagation) and selected on synthetic complete medium lacking leucine (Sc-Leu). Transformants were spotted onto medium containing 5-fluoroorotic acid (5-FOA) to select for loss of pJ533.

Screening for prion forming transformers.

Libraries mutants that grew on 5-FOA were then spotted onto medium lacking adenine (Sc-ade), yeast complete medium (YPAD), and prion color based medium (YPD). To eliminate any colonies with SUP35 function, manually select colonies that grew as red on YPD and failed to grow on Sc-ade, those colonies should have a functional SUP35 mutant. These cells were then pooled into different mini-libraries (~50 colonies each). This library size prevented strong prion to overpower. And then OD_{600nm} readings were used to plate equal numbers of the libraries strains (derived from YER709) onto Sc-ade at concentrations of 10^5 and 10^6 cell per plate and grown for 5 days at 30°C. Then individual clones were tested to see whether they are legitimate prions by spreading them on YPD media to allow color selection. Ade+ colonies were streaked on YPD and YPD plus 4 mM GdHCl to test for curability. Clones in which the Ade+ phenotype was stable and curable were sequenced (selected clones). The *SUP35* from single library clones proceeding to prion selection was sequenced as well to generate a naïve library data base.

Mating and screening for prion propagators.

Library mutants that grew on 5-FOA were then collected and OD_{600nm} readings were used to place equal numbers of the library strain (derived from YER648 above, which is mating type a) and 780-1D/pJ533 (α mating type) together on YPD supplemented with adenine for a 24 hour

mating reaction. After mating, only diploid cells were allowed to grow by replica plating on synthetic defined media supplemented with adenine, tryptophan, and uracil (SD+Ade,Trp,Ura) to remove the haploid parents. Subsequent to selection, cells were again spotted on 5-FOA-containing medium to select for loss of pJ533. Cells were then spread on YPD media to allow color selection. Ade^+ colonies were streaked on YPD and YPD plus 4 mM GdHCl to test for curability. Clones in which the Ade^+ phenotype was stable and curable were sequenced (propagators). Clones with a strong Ade^- phenotype were likewise sequenced (non-propagators).

Assessment of 3R propagation mutant prion isolates' ability to pass the species barrier.

Individual non-propagating clones were tested for the ability to pass the species barrier (that is, to add on to wild-type aggregates when co-expressed in a cell). To accomplish this, the plasmids expressing their mutant versions of *SUP35* were isolated from cells grown in liquid culture and these plasmids were then used to transform the 780-1D/pJ533 strain (see above), which expresses the wild-type *SUP35* from a *URA3* plasmid. Cells were then spread on YPD media to allow color selection and also were spotted on 5-FOA-containing medium to select for loss of pJ533 after which they were spread on YPD. Clones that were not red on the initial plating on YPD that became red after 5-FOA treatment were considered to have initially passed the species barrier, but to be unable to propagate the prion as the sole copy within a cell.

Library screen selecting for 3R propagation mutants' able to add on to existing aggregates.

A more comprehensive screen for the species barrier was also undertaken. In this experiment, the PCR mutagenesis and transformation of YER648 and mating with 780-1D/pJ533 were as above. However, the selection step to remove haploids after mating was performed by replica plating on synthetic defined media supplemented with tryptophan and uracil only (SD+Trp,Ura). The lack of adenine in this medium allowed only cells that were Ade+ at this

stage (those that could form Sup35 aggregates) to remain in the library. All subsequent steps of handling were as in the previous experiment, and analysis was undertaken as below.

Calculating prion propensity scores and amino acid composition of non-propagation libraries.

Libraries designed for analysis of propagation were analyzed as described below. Other libraries where prion formation was assessed were analyzed. A comparison between each mutant library and its naïve library was done by calculating the odds ratio to represent each amino acid (over/under representation), and then we scaled these values from 1 to -1. ORaa was defined as below.

Composition of yeast prion propagators.

For each yeast mutant PFD, the odds ratio for each amino acid (OR_{aa}) was calculated as

$$OR_{aa} = [f_p/(1-f_p)]/ [f_{np}/(1-f_{np})] \quad (1)$$

where f_p is the fraction of residues present in the mutated region of propagating prions that are the indicated amino acid and f_{np} is the fraction of residues present in the mutated region of non-propagating prions that are the indicated amino acid.

Calculating prion propagation propensity scores.

Prion propagation propensity (PPP) was calculated as follows. For each mutant sequence, the amino acids found in the 3rd repeat of each mutant were counted. For each amino acid, the number of residues present were multiplied by the $\ln(OR_{aa})$ for that amino acid, and the sum of these values for all amino acids present was called the PPP score for that sequence. Any residue present at $\leq 0.7\%$ in either the propagating or non-propagating library was excluded from

consideration. As a result, E, K, M, Q, and W residues were not counted in the PPP score for sequences that contained these residues.

Mutant library creation of the 5R library by PCR.

All steps from mutant library creation through calculation of prion propagation propensity scores were as above, except, the mutant primers used in library creation were as follows. In place of EDR1377, EDR1378 was used: GGAGGCTACCAGCAATACAAC$(NNB)_9$AATCCACAAGGTGGACGTGGAAAC. In place of EDR1385, EDR1387 was used: GTTGTATTGCTGGTAGCCTCCTTGAGGATTATACTGTTGCTGGTAACCGGCGTC. In place of EDR1395, EDR1396 was used: GGATTCAAACCAAGGCAACAATCAGCAAAACTACCAGCAATACAGCC.

Creation of *de novo* mutants in the ORD.

A random proteome of 65386 residues was generated using the random number function of the Microsoft Excel software program and an equal chance of selecting any of the 20 natural amino acids. 10 amino acid windows were scored using the calculated PPP values for each window. 3628 sequences did not contain any of the low-abundance residues (E, K, M, Q, and W) and were chosen for further evaluation. Sequences at the 95^{th} and 5^{th} percentile in terms of PPP score were chosen as mutants with no selection other than being compositionally non-identical and non-overlapping. Mutants at the 95^{th} percentile were predicted to be prion propagators based on the PPP scoring method and those at the 5^{th} percentile were predicted to be non-propagators. PCR amplification and mutagenesis was undertaken as described above, but specific synthetic oligonucleotides (Invitrogen) were synthesized to generate these clones. The primers used were:

DN1:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATCCAGTTTATGATCATATTTATT

ATGCTTCTCCTCAAGGAGGCTACCAGCAATACAAC

DN2:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATTATAATGGTATTAATGCTTATT

TGTTTAATCCTCAAGGAGGCTACCAGCAATACAAC

DN3:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATACTGCTAGATATGGTGCTGCTT

TTGGTTATCCTCAAGGAGGCTACCAGCAATACAAC

DN4:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATGGTTTGTGTCATTTGTATTTTA

ATTGTACTCCTCAAGGAGGCTACCAGCAATACAAC

DN5:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATGCTCCATGTGCTTATCCACCAA

CTAGACATCCTCAAGGAGGCTACCAGCAATACAAC

DN6:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATTATTCTGTTGGTTTTCATAATG

ATACTACTCCTCAAGGAGGCTACCAGCAATACAAC

DN7:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATTATACTAATAGAGGTGATCCA

ACTATTAGACCTCAAGGAGGCTACCAGCAATACAAC

DN8:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATCATGTTTCTCATGGTAGATTTT

GTGATATTCCTCAAGGAGGCTACCAGCAATACAAC

DN9:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATGATATTTGTTTGTCTGATGTTG

GTGATGCTCCTCAAGGAGGCTACCAGCAATACAAC

DN10:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATTTGATTGTTGTTATTGGTTTTAT

TCCAACTCCTCAAGGAGGCTACCAGCAATACAAC

DN11:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATGATGATACTAATTTTAGATGTT

TGATTAGACCTCAAGGAGGCTACCAGCAATACAAC

DN12:

CTGGGTACCAACAAGGTGGCTATCAACAGTACAATGGTAGAGTTTTGTTTAGAAATG

GTATTGGTCCTCAAGGAGGCTACCAGCAATACAAC

Corresponding oligonucleotide primers targeting the 4^{th} and 5^{th} repeats of the ORD (that differ only in their flanking regions) were also synthesized and used for generation of yeast strains carrying the same mutations in those regions.

Creation of targeted insertion mutants in the ORD.

Mutant oligonucleotides were designed to insert specific non-aromatic hydrophobic residues (valine and isolucine) into or delete tyrosines from the 3^{rd} and 4^{th} repeats of the ORD. PCR amplification and mutagenesis was undertaken as described above, but specific synthetic oligonucleotides (Invitrogen) were synthesized to generate these clones. The primers used were:

EDR1419 (GGCGTCTGGGTTGTACTGCTGGTATCCACCTTGTTGGTACCCAG), antisense primer to make -4tyr, -2A, -2B, +2A, +2B constructs. EDR1425 (GATAGGGTTGTACTGAACCTGCTGGTAAATACCGATGGCGTCTGGGTTGTACTGCTGGTATCCACCTTGTTGGTACCCAG), antisense primer to make +6A. EDR1427 (GAGGGATGTTGTACTGTTGGACCTGGTAACCAATGGCGTCTGGGTTGTACTGCTGGTATCCACCTTGTTGGTACCCAG), antisense primer to make +6B. EDR1420 (GATACCAGCAGTACAACCCAGACGCCGGTCAGCAACAGAATCCTCAAGGAGGCCAACAGAATCCTCAAGGCGGTTATCAGC), sense primer to make-4tyr. EDR1421 (GATACCAGCAGTACAACCCAGACGCCGGTCAGCAACAGTATAATCCTCAAGGAGGCCAACAGTACAATCCTCAAGGCGGTTATCAGCAGCAATTCAATCC), sense primer to make -2A. EDR1422 (GATACCAGCAGTACAACCCAGACGCCGGTTACCAGCAACAGAATCCTCAAGGAGGCTATCAACAGAATCCTCAAGGCGGTTATCAGCAGCAATTCAATCC), sense primer to make -2B. EDR1423 (GATACCAGCAGTACAACCCAGACGCCGGTTACCAGCAAATTCAGTATAATCCTCAAGGAGGCTATGTTCAACAGTACAATCCTCAAGGCGGTTATC), sense primer to make +2A. EDR1424 (GATACCAGCAGTACAACCCAGACGCCGGTTACCAGCAAGTTCAGTATAATCCTCAAGGAGGCTATCAACAGATTTACAATCCTCAAGGCGGTTATC), sense primer to make +2B. EDR1426 (GCAGGTTCAGTACAACCCTATCCAAGGAGGCTATCAACAGGTATACGTTAATCCTCAAGGCGGTTATCAGCAGCAATTCAATCC), sense primer to make +6A. ER1428 (GTCCAACAGTACAACATCCCTCAAGGAGGCGTATATCAACAGATTTACGTTAATCC

TCAAGGCGGTTATCAGCAGCAATTCAATCC), sense primer to make +6B. Seven different constructs were made with hydrophobic residues insertion and deletion. Two constructs have 6 extra hydrophobic residues (+6A and +6B) and two constructs have two extra hydrophobic residues (+2A and +2B). An additional 3 constructs have deletions of hydrophobic residues. Two constructs have two tyrosines deletions (-2A and -2B), and one construct lacks 4 tyrosines (-4Tyr). Then each construct was shuffled in to wild type SUP35 as described above (except the host strain wasYER632).

Screening for prion formation by induction.

To build induction plasmids, the N and M domains of the 7 mutants SUP35 were amplified by PCR with primers EDR301 (CGTCACAGTGTTCGAGTCTG) and EDR304. And then reamplified with EDR1008 (GAGCTACTGGATCCACAATGTCGGATTCAAACCAAGGCAAC), and EDR1084 (CGATGCTACTCGAGTTTACATATCGTTAACAACTTCGTCATCCAC) to insert a *BamHI* site before the start codon and a stop codon and an *XhoI* site after the M domain. PCR products (inserts) were digested with *BamHI-HF / XhoI* and ligated to *BamHI/ XhoI/CIP*–cut pKT24. Ligation products were then transformed into NEB DH5α cells and examined by DNA sequencing. The mutant copies of SUP35 were transformed with either plasmid containing the *GAL1* promoter (un-induced) or with a derivative plasmid (induction plasmid) expressing the SUP35 PFD domain from the *GAL1* promoter (induced). Strains were then grown in galactose/raffinose -trp +ade dropout mix medium (made the same as SC media, but with 2% Galactose and 1% raffinose instead of glucose) for 3 days and 5 serial dilutions were plated onto medium lacking adenine (Sc-ade) and grown for 5 days to select for [PSI^+].

***In silico* reanalysis of Alberti *et al.* data set.**

Amino acid compositions were compared by calculating the percentage of each amino acid out of the total number of amino acids in the previously-identified prion-like domain of each protein (Alberti *et al*. 2009). The 18 proteins that passed all four tests in the assays of Alberti *et al.* were as follows: Ure2, Sup35, Rnq1, New1, Puf2, Nrp1, Swi1, Ybr016w, Cbk1, Lsm1, Ybl081w, Pub1, Ksp1, Asm4, Nsp1, Gln3, Ypr022c, and Rlm1. The 12 proteins that failed only in the Sup35 fusion protein expression assay (demonstrating a potential propagation defect) were as follows: Snf5, Gts1, Scd5, Sgf73, Sok2, Mot3, Ngr1, Jsn1, Pdr1, Cyc8, Pan1, and Ybr108w.

Statistics.

Two-sided Student's *t*-tests were undertaken using JMP 6.0 (SAS, Inc., Raleigh, N.C.). Fisher's exact tests and *P*-value calculation from Z-scores used GraphPad (GraphPad Software, Inc., La Jolla, CA).

Results

Random mutagenesis of the *SUP*35 nucleation and oligopeptide repeat domains.

Given the separate compositional requirements and functions of the nucleation domain (ND) and oligopeptide repeat domain (ORD, involved in prion propagation) (DePace *et al*. 1998; Liu *et al*.,1999; Parham *et al.* 2001), libraries of yeast clones were created in which segments of the ND (amino acids 21-28) or the entire 3rd repeat of the *SUP35* ORD were replaced with a segment of random amino acid residues. Briefly, the oligonucleotides used to create the two libraries were designed to bind in the regions flanking the target regions for mutagenesis and replace only the target codons with $(NNB)_n$, where N is any of the four nucleotides, B is any of the nucleotides except adenine, and *n* is requisite number of amino acids to replace (for the ND library, n=8 and for the 3rd repeat library, n=10). Although disallowing adenine at the final

position does eliminate two of the three stop codons and thereby make screening easier, leaving out adenine does not otherwise eliminate any amino acids from the mutant region. PCR with these chemically-synthesized "random" oligonucleotides was used to create libraries of mutated versions of *SUP35* which were then transformed into yeast cells in which the sole copy of the *SUP35* gene was expressed from a plasmid. Through plasmid shuffling, the plasmid-encoded version was replaced with the random library.

Yeast cells were transformed with the PCR-generated mutants in the ND and ORD regions (Figure 2.1A). Mutant regions were sequenced for two groups of clones within each library: those cells selected for [PSI^+] (the "selected" library) and other cells that contained mutant SUP35 genes but were not selected for [PSI^+] (the "unselected" library) (Table 1). [PSI^+] selected isolates were tested to remove any clones that turned white due to DNA mutation rather than presence of [PSI^+]. To distinguish clones that displayed an Ade^+ phenotype due to DNA mutation from genuine prions, isolates were grown on guanidine, which "cures" prions, resulting in a conversion to a red phenotype. All [PSI^+] clones that were curable when grown on guanidine-containing media were included in the study; all non-curable clones were excluded. *SUP35* was sequenced from the 37 [PSI^+] clones and the 61 unselected library clones analyzed in this study and each sequence is reported in Table 1.

Compositional biases among the ND and ORD mutant isolates.

For each of the amino acids, an observed odds ratio (OR_{aa}) was determined, which represents the degree of over- or underrepresentation of that amino acid among the [PSI^+] isolates, as previously described (Toombs *et al.*, 2010). A statistically significant overrepresentation ($\alpha = 0.05$) among the [PSI^+] isolates among mutants in the ND region of Sup35 was observed for Phe, and a significant underrepresentation was noted for Gly and Arg

(Table 2). Due to small sample sizes, some propensities did not show statistical significance, but grouping these amino acids together allowed detection of these subtle biases by increasing the effective sample sizes for those groups. Through the use of groups of chemically-similar amino acids, we observed an overrepresentation of aromatic amino acids and non-aromatic hydrophobic residues (except Leu) along with an underrepresentation of charged amino acids in [PSI^{+}] ND-mutant isolates (Table 2).

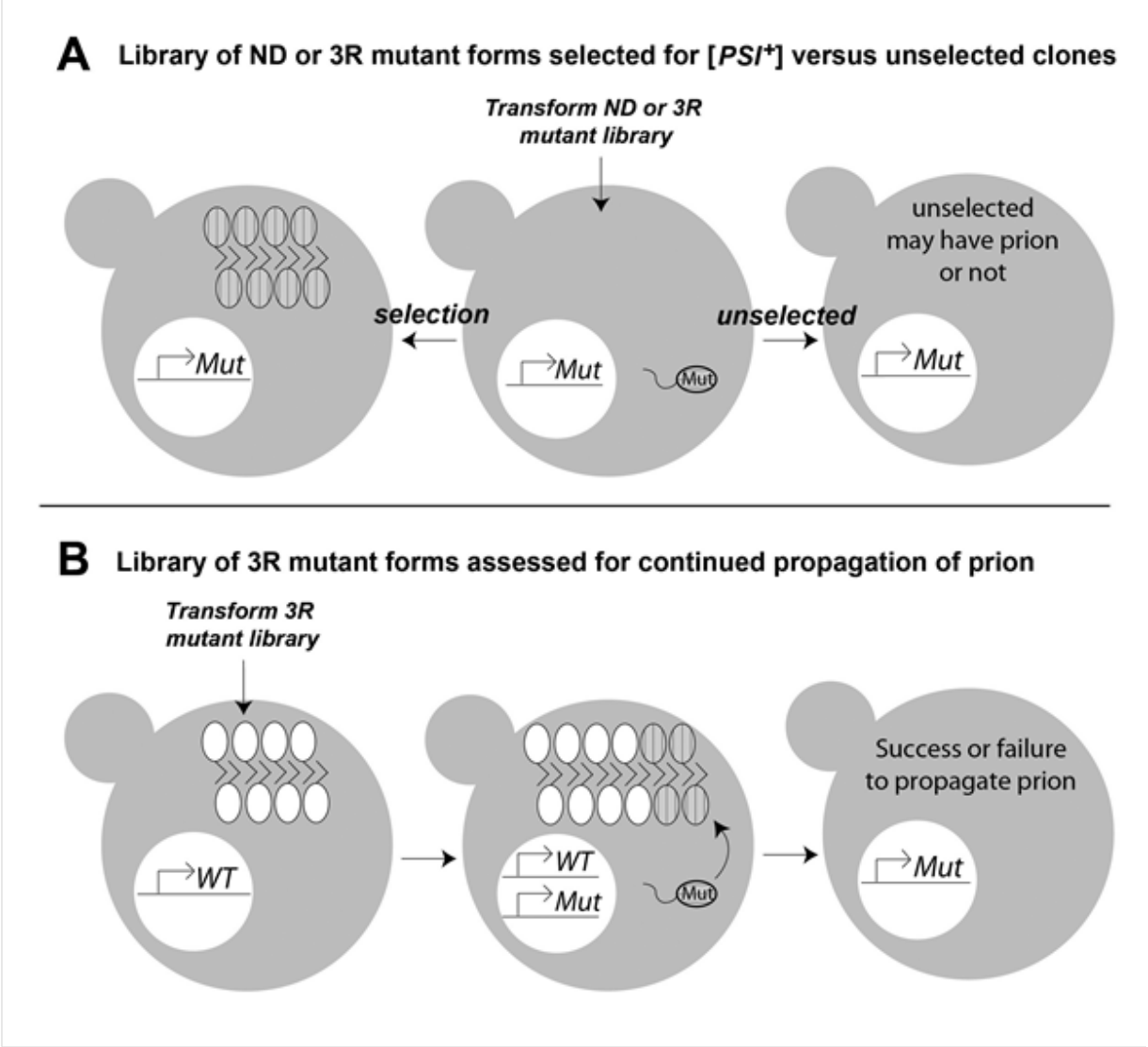


Figure2. 1: Aggregate formation and maintenance assays used for library analyses. (A) For the experiments described in Tables 1-4, [psi$^-$] cells were transformed with libraries of mutant forms in the ND or ORD regions (middle cell) and either selected for presence of [PSI$^+$] aggregates (left cell) or not selected (right cell). (B) For the experiments described in Tables 5-7, [PSI$^+$] cells expressing a wild-type Sup35 gene from a plasmid (WT, left cell) were mated to a cell containing a library of mutated versions (Mut) of Sup35 (middle cell), generating a library of diploid cells coexpressing WT and Mut proteins. Cells were selected for loss of WT and were then assayed for the continued presence of aggregates (right cell).

Table2.1: Nucleation domain (ND) mutated sequences (amino acids 21-28). Nucleation domain (ND) mutated sequences (amino acids 21-28) from the pool of ND mutant clones selected for prion formation and the unselected library.

Unselected sequences		[*PSI*$^{+}$] sequences	
LIARSNHC	SNYPSCFI	RNVFLVGL	HWLFCFG
LRGGGGHI	VVFGLGQG	YHMIGVVW	SSGSICVS
LRIILDGA	YICSISMD	GFFTSSVF	GCSWQPS
RRAYFSLP	LYVITNFI	GYLSSCVF	IHIHNSGR
CGAHDGG	RCGCGRGP	GGTFIFGC	SALGHNV
LTWLWTL	GSNDLDTS	VNRVTCNV	TYFDQYG
CVCCGLD	WGCPSSGH	AITICGGV	SLTTCFCA
SCGSNVM	SNSSCANL	FVGAFGIF	TVFFIGDL
RYGSALW	DVTSSLLM	YISVYVAG	IFTTHFSR
AYNVSQR	CDLECVGR	CGYFNAYC	IAISYVNA
PIRGNVLM	YTLRTWRL	RDAIYYSR	SATYSMV
SLHYLGW	YGRHTDAC	TGNGGPGL	PTHAFFNS
TRIRRICV	SSAFFRCV	GACNGLFV	VVFLYSD
DHALGDW	QSADFSAF	TSYVFGPC	GTGSIYFL
RVVRASPS	GAYQRGPV	GLICHIYN	LVIAGDIS
QCASTGIS	SSVIRYFA	NVMWLTSG	DIFRNCFY
IGSQSPCA	FGTRGTYG	FHVEIYLN	PNVVSNV
SRGDRSSG	REDAGCNC	DFCASYVI	GFVLAFD
FVLARRTG	GGTFIFGC	LIFGISTV	
GIRRDCGC	HGSGHVAI		
GCSWQPS	HFHIGSYG		
GNPYDGG	VRGRIICV		
GRVYPVC	VGDVVGVA		
SGVTGSD	SCGEPGLT		
PNSCRLTG	RIGSPAVG		
TIGDGNVR	NGVSGAAF		
YGGLMHR	GSGLVRWG		
QASGGYV	PGGDWLSM		
CVSGGCIS	VFRTGYMH		
CSPHSGSG	GSSIGRIT		
NVERFYGF	FGARGVGL		

Table2. 2: Library amino acid representation of ND mutants mutating amino acids 21-28 of Sup35. [a]Odds ratios (OR_{aa}) were calculated using equation 1 and then the natural logarithm of each value was taken. Values are provided for groups of amino acids, but only the values for individual amino acids were used in subsequent calculations. [b]P value is derived from the two-tailed Fisher's exact test. [c]The natural logarithm of the odds ratio could not be computed for lysine because zero lysines were observed in either library. [d]See Table 6 for prion propagation propensity results. [e]As described in Toombs et al., 2010, when multiple consecutive prolines were separated by no more than one residue, a value of zero was used for the natural logarithm of the odds ratio for each proline after the first in that cluster (Alberti *et al*. 2009).

Amino acid(s)	**ND Mutant Prion Propensity Results**				**3R Propagation Propensity Results[d]**	
	Frequency					
	[PSI+] selected library	**Unselected library**	**ln (OR_{aa})[a]**	***P* value[b]**	**ln (OR_{aa})**	***P* value**
Tryptophan (W)	0.014	0.021	–0.41	0.59	1.4	0.011
Tyrosine (Y)	0.069	0.042	0.49	0.17	1.03	4.2×10^{-5}
Glutamic acid (E)	0.003	0.008	–0.87	0.66	0.99	0.07
Cysteine (C)	0.065	0.071	–0.10	0.88	0.45	0.045
Phenylalanine (F)	0.121	0.042	1.06	3.05×10^{-4}	0.33	0.15
Alanine (A)	0.053	0.055	–0.04	1.00	0.28	0.15
Asparagine (N)	0.061	0.031	0.67	0.06	0.18	0.57
Serine (S)	0.109	0.125	–0.14	0.63	0.18	0.28
Glutamine (Q)	0.007	0.016	–0.88	0.34	0.11	1.00
Proline (P)[e]	0.017	0.035	–0.73	0.18	0.06	0.81

Glycine (G)	0.130	0.213	–0.49	0.025	0.05	0.81
Threonine (T)	0.061	0.042	0.37	0.30	–0.06	0.89
Histidine (H)	0.031	0.029	0.08	0.40	–0.06	0.90
Aspartic acid (D)	0.028	0.046	–0.51	0.25	–0.19	0.44
Valine (V)	0.125	0.078	0.47	0.068	–0.35	0.049
Leucine (L)	0.065	0.071	–0.10	0.88	–0.48	0.05
Isoleucine (I)	0.088	0.055	0.47	0.13	–0.57	0.066
Arginine (R)	0.024	0.095	–1.37	2.42×10^{-4}	–0.88	6.14×10^{-6}
Methionine (M)	0.014	0.014	–0.04	1.00	–1.80	0.087
Lysine (K)[c]	0.000	0.000	NA	NA	NA	0.27
Groups						
Aromatic (FWY)	0.228	0.112	0.71	1.07×10^{-3}	0.76	3.1×10^{-6}
Non-aromatic hydrophobic (ILMV)	0.364	0.259	0.34	0.054	–0.53	9.0×10^{-5}
Without leu (IVM)	0.260	0.162	0.47	0.017	–0.48	0.30
Charged (DEKR)	0.037	0.162	-1.46	1.28×10^{-4}	–0.54	2.0×10^{-4}
All polar (HNQST)	0.333	0.292	0.13	0.44	0.33	0.0046

When prion formation was examined in mutant isolates of the 3rd repeat (3R) of the ORD (Table 3), however, some differences were noted (Table 4). No statistically significant overrepresentations among the [PSI$^+$] isolates in the 3rd repeat were observed, but significant underrepresentation was noted for Arg (Table 4). Groupings of amino acids revealed that there was a near-significant overrepresentation of aromatic amino acids and a significant underrepresentation of charged amino acids in [PSI$^+$] 3R mutant isolates. However, the absence of a positive effect of hydrophobic residues (as seen in the ND experiment in Table 2 and in previous work (Toombos *et al.*, 2010), suggested that there were different compositional requirements in the ND and ORD regions.

Random mutagenesis of the *SUP*35 oligopeptide repeat domain for assessment of propagation by 3R mutant clones.

We know that the functions and overall compositional requirements for the ND and ORD regions are different (DePace , 1998; Liu, 1999; Shkundina ,2006; Parham, 2001). The results contained in this work (Tables 2 and 4) further suggested that aromatic residues may drive prion formation in both regions, while hydrophobic residues are less important, or even negative, within the ORD. These preceding experiments were imperfect, however, because both the ND and 3R formation libraries required proteins to both form and propagate prions. Given that the ORD differences may reflect unique requirements for propagation in that site, a new method was required that would isolate the effects of composition on propagation only.

To do this, using PCR mutagenesis, a separate library of 3rd repeat mutants of the Sup35 ORD was created as above, but the ability of those clones to add on to existing wild-type [PSI$^+$] fibers and then continue propagation in the absence of the wild-type protein was assessed. To undertake this the library of yeast transformant clones were then mated with a strain expressing

the wild-type *SUP35* gene from a plasmid (Figure 2.1B, left cell), creating diploid cells that expressed both versions of *SUP35* (Figure 2.1B, middle cell). After selecting for loss of the wild-type *SUP35* plasmid (Figure 2.1B, right cell), clones were screened for Sup35 activity using the ade2-1 allele. To do this, all clones were spotted onto medium containing limiting adenine (YPD). Cells that grew on YPD but expressed a strongly red phenotype (due to build-up of the pigmented adenine precursor) were unable to continue propagation after loss of the wild-type plasmid and were considered to be non-propagators.

Cells that expressed a similar (white) phenotype to the wild-type Sup35 phenotype were able to continue propagation after loss of the wild-type plasmid and were considered to be propagators. In order to be considered successful propagators, these white isolates were tested to remove any clones that turned white due to DNA mutation rather than presence of [PSI^+] by curing with guanidine hydrochloride. All propagating clones that were curable when grown on guanidine-containing media (representative example in Figure 2.2A) were included in the study; all non-curable clones were excluded. Non-propagating clones were also tested on guanidine (representative example on Figure 2.2B) but demonstrated no phenotypic change (because the prion had already been lost). Clones of intermediate phenotype (which were either pink on YPD or showed any loss of the prion upon streaking) were also excluded from the study. *SUP35* was then sequenced from the 65 clearly propagating and 87 clearly non-propagating library clones analyzed in this study. The sequences are reported in Table 1.

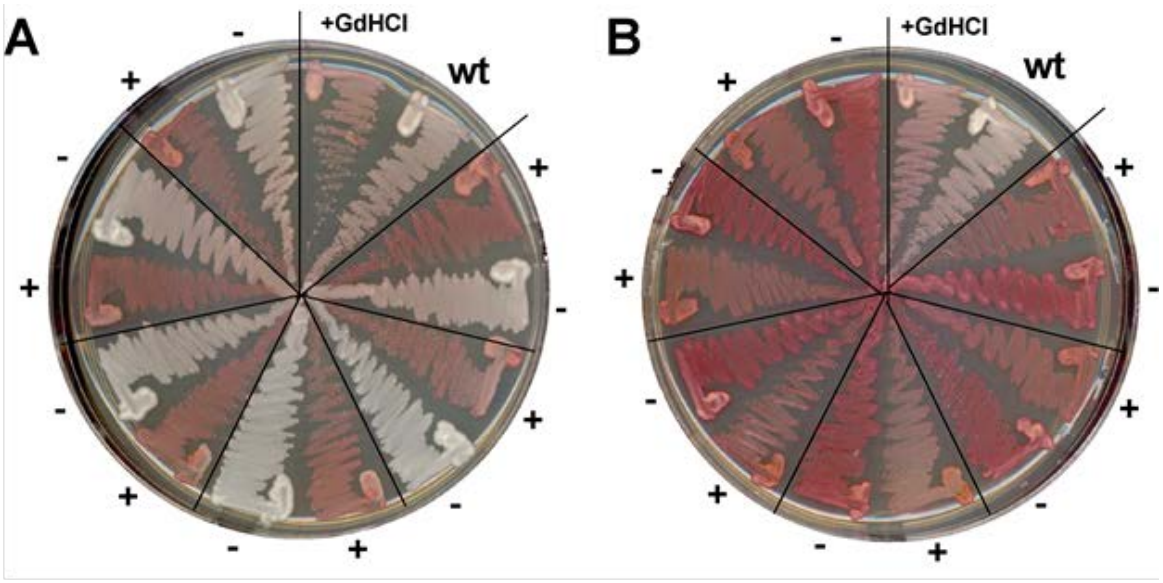


Figure2. 2 : Representative examples of propagating and non-propagating clones of the 3rd repeat library. After being plated on YPD with and without treatment with 4 mM guanidine hydrochloride (GdHCl), wild-type Sup35 and mutant clones – and + GdHCl treatment were streaked side-by-side on the same YPD plate. (A) Propagating clones demonstrate curability of the Sup35 prion phenotype. Only clones that showed a similar phenotype to wild-type Sup35 and were curable were sequenced and included in the propagating library. (B) Non-propagating clones demonstrate no difference in phenotype before and after GdHCl treatment. 65 clearly propagating and 87 clearly non-propagating library clones analyzed in this study. The sequences are reported in Table 5.

Table2. 3: 3rd repeat mutated sequences from the pool of 3R mutant. 3rd repeat mutated sequences from the pool of 3R mutant clones selected for prion formation (not specifically for propagation as in Tables 5-7) and the unselected library.

Unselected sequences		[*PSI*$^+$] sequences	
PTDVCA	KLTGGCTD	FYGVLVCH	YCLCTG
RVFGVS	GSVVPIYGD	YYPIVVTGT	LYTWTS
SGGIHRT	SHTFGAVAC	FHGAFSAVA	DLVWPF
RGLCRW	VVPCCRALF	STVYMCAS	SPYTIFR
LIGAGA	IDFSLNRAL	VGGVFVCN	GCCND
AFCRAG	TIGVLSVRG	TFNIYPNGV	VNVTV
PFNHLA	SERCFFSVS	GIIFGFHSYY	ASCNCR
SGAALV	NHGITGSVM	ADFSHVSGY	GGCLN
IIRFIERD	HSAPLRPAL	NFSSYPGRV	NAGRSP
TLSRYC	DNSHPFDFA	ADSCFLGAF	GDGCG
YSIGCG	RRVCIVGLF	TAGGLSMDI	IGTQFD
DMAQG	LVDSHSWCS	IAALYVPMS	SNARRA
SSIRPNP	GPDHGGVV	CAENLIGWF	SSGIYLS
YDYRYS	KRCHRGSR	LLSCGIGSFA	VDHEC
RAFRGV	IRVSYHSPA	SFLYYFDVC	AGFSHY
VIRGFC	LKISFNRPFS	HLCFDSNRC	
VYRLSG	RTASGGELP		
CCRIRG	FSLWISYCR		
AGCAIVI	RLPGYDDYI		
IVCRVF	GPDSCTGRF		
DTVIFY	NSYCSHLVR		
WSCDGN	FSGTRGDNG		
CGSFGD	SLVGSSFISR		
VVLPCIS	YGLTGYPLC		
GYRHGP	PCVCIHFRR		
ASGSGG	PVYAHDGG		
WATVD	AFDSYNSVA		
FRGAYF	IDGMGRSSV		
VTHGGH	VRGAWPHC		

Table2. 4: Library amino acid representation of 3R mutants in the prion formation experiment (Table 6). [a]See Table 2 for an explanation of terms. [b]See Table 6 for 3R propagation library results.

Amino acid(s)	3R Mutant Prion Propensity Results				3R Propagation Propensity Results[b]	
	Frequency[a] [PSI+] selected library	Unselected library	ln (OR_{aa})	*P* value	ln (OR_{aa})	*P* value
Tryptophan (W)	0.016	0.016	0.04	1.00	1.4	0.011
Tyrosine (Y)	0.076	0.055	0.34	0.29	1.03	4.2×10^{-5}
Glutamic acid (E)	0.006	0.012	–0.63	0.51	0.99	0.07
Cysteine (C)	0.084	0.062	0.30	0.32	0.45	0.045
Phenylalanine (F)	0.088	0.060	0.37	0.20	0.33	0.15
Alanine (A)	0.095	0.068	0.34	0.22	0.28	0.15
Asparagine (N)	0.047	0.021	0.81	0.058	0.18	0.57
Serine (S)	0.123	0.113	0.08	0.73	0.18	0.28
Glutamine (Q)	0.003	0.003	–0.07	1.00	0.11	1.00
Proline (P)	0.037	0.053	–0.36	0.40	0.06	0.81
Glycine (G)	0.144	0.151	–0.05	0.92	0.05	0.81
Threonine (T)	0.040	0.038	0.07	0.85	–0.06	0.89
Histidine (H)	0.047	0.038	0.23	0.59	–0.06	0.90
Aspartic acid (D)	0.033	0.058	–0.56	0.14	–0.19	0.44

Valine (V)	0.076	0.086	–0.12	0.69	–0.35	0.049
Leucine (L)	0.058	0.060	–0.04	1.00	–0.48	0.05
Isoleucine (I)	0.040	0.058	–0.37	0.33	–0.57	0.066
Arginine (R)	0.044	0.107	–0.89	3.60×10^{-3}	–0.88	6.14×10^{-6}
Methionine (M)	0.016	009	63	33	1.80	087
Lysine (K)	0.003	0.010	–1.17	0.43	NA	0.27
Groups						
Aromatic (FWY)	0.202	0.142	0.35	0.084	0.76	3.1×10^{-6}
Non-aromatichydrophobic (ILMV)	0.220	0.250	–0.13	0.53	–0.53	9.0×10^{-5}
Without leu (IVM)	0.144	0.167	–0.15	0.54	–0.48	0.30
Charged (DEKR)	0.054	0.211	–1.36	1.85×10^{-4}	–0.54	2.0×10^{-4}
All polar (HNQST)	0.319	0.247	0.26	0.14	0.33	0.0046

Compositional biases among the propagating prion isolates.

For each of the amino acids, odds ratios (OR_{aa}) were calculated as shown in equation 1 (see Materials and Methods). Among the propagating isolates, a statistically significant overrepresentation was observed for Trp, Tyr, and Cys, and a significant underrepresentation was noted for Val, Leu, and Arg (Table 6). Because of small sample sizes, some biases did not rise to the level of statistical significance. However, groupings of amino acids allowed the detection of these subtle biases by increasing effective sample sizes for those groups. Through chemically-similar amino acids, we observed an overrepresentation of aromatic and polar amino acids along with an underrepresentation of hydrophobic and charged amino acids in propagating isolates (Table 6). Through comparison with earlier published results on prion propensity (Toombs *et al*., 2009) and our initial results in this study (Table 2 and 4), similarities and differences in composition effects on prion propagation and formation were noted (Table 6, compare $\ln(OR_{aa})$ for propagation with prion propensity score from (Toombs *et al*., 2009) on the right of the table). The 3R propagation experiment noted compositional differences between propagating and non-propagating clones that were more significant than the differences observed in the 3R prion formation and ND libraries seen in Tables 4 and 2, respectively. For example, non-aromatic hydrophobics were near beneficial for prion formation in previous work(Toombs *et al*., 2009), slightly beneficial in the ND library (Table 2), modestly worse but only at near-statistical significance in the 3R formation assay (Table 4), and significantly worse in the 3R experiment that assessed propagation ability only (Table 6). Given these results, several confirmatory experiments and additional experiments were undertaken.

Table2. 5: 3rd repeat (3R) mutated sequences from prion-propagating and non-propagating libraries. Asterisks are used to indicate non-propagating clones that were analyzed for presence of a species barrier.

Propagating sequences		Non-propagating sequences	
PYSYVRCSCD	DHAHGIPRSC	RGVYCAVIGA	VIVDRGQICG*
YTGGISAAGT	PNGSHIIFSF	NFRPLHAVAR	IDYSGRSLLL*
CPCAPLGCW	VLESSSFERF	LYTYDASMAF	IAIVGFGGTG
IVAPVHGWSC	GNCGSLHDER	SRIRRRTSFH	PGDRGLSRAR
HYAAGYPCA	ACSAEHRWGL	AHCSGFSGRV	HSVVRRCTR
VYDRRGVGE	TAAVRNFGAS	VWCRGSVDAL	DSRLLRPGCS
VRSGYYLYN	VAAAIRTAQT	HARTSMDHLA	HDTLCNGVRF
SGSFCSFGRY	NSIRAGSCNL	RHDITVNFGD	RVSGVVSLHL
LPGYGNHEFV	GITFFGDASG	IDHVISDSFR	FGRSLNSFRY
QFCQASFSAG	SSTNYDYHRA	YAVVGSANHC	PRIGHSDLVN
HHVPWVSTC	AANTHREWCF	SRCSHVSISR	PAGDDRNFVS
SPTPGLYAAS	LYPLFVVDTV	RFGGTDDVDV	HCYCHRAVAR
DGYHFCQWC	CIDNVCRSGW	FAISRSVLRG	ANLVVHCVTQ
PLRGFYYHAF	CIPDRYDCSA	HYVVASGDRR	RPVTRGHRYH
GFYYSGIRDT	YGPPSVSVGD	RRSTAVMHVA	RGPTSGPSGD
GSNGYPGNC	SVGVAEFDRA	RRTRTRVRCT	YPRISHGGQC
DVFASIYRAC	FDYFFNGDNT	RAGVDHSSRG	EITVMIAMSR
SNRCPAYRSS	HVNFVAVAVH	GALVPDTSLL	PRVVVVILNL
LFTLRFPFSV	FVVLASRCAY	WDFSVDAALG	DFRRHRHLGF
SHYMSGVAD	VSVGSCVPGC	GCRSRINGFN	PQCDTSSSCG
SFEGNVLRHP	DYVSGANHNS	SPGIFRGSAL	FGVPCVNVPV
TFYTNSSAPV		RSGWRSRGSV	PDNVVGNPPS
WDLCGDVAG		FYSVSILDRR*	VVRPGLSDRS
FLGIGSLCAG		GCPRVVIHVD*	CSGFLDDRCI
HGDTSYADS		ADACLSLVSV	FMYGRDTRVS
ACSGGHAGT		AVRSRRRENR	PNHYRFGAVC
RDLSGYFDGG		NRYSIHGNGV	IIIVPSDRAL
VVSPSLGATT		ATCIDSSTNS	LVRYFNVGDC
SWNFYPCHG		TLDACYVCRR	RKFAFSDSAG
VGNVGWGVA		ASRIKRPLGE	LGASVYLVSG
YVCGHIDVFD		RVFGGGRAC*	ATSLTVLGNG
YANYDPSHCT		PRPNVISFGA	DTVGALHTRV
CGFSVFVGRC		LVRVFSPPHS*	CHARGGCLVR
CSCLVTGFEP		YDGADVFPPG	FAYFSVGSIT
DCFRYCLSGV		VVSVSDGIWC*	RRKGYCGGFL
GACVAACFA		PVGGNFAAV*	GYGGTVGFAH
DCVSRQTFGG		VRVTFLNRNG	IGCTSISLAP
YSLYGSFYPS		NSDCVACFLS*	RGSVTHGVGQ
VYASCVWSRI		VPSYRRCAVV	SLSARVYCIS
CDHGSYFDC		DPHPSTGDCF*	VIGRSARYLL
CAFGSLRSSV		GGCNGEVFFH*	DLELDVRTDM
PYPDCAYGSV		LSAGVFGCAV*	IDDCMPPGDL
IRVSYHSPAP		RPAIARDNVG*	QSAFDDVPL
LYFNRGSSRA		PHFALVHSTH*	

Table2. 6: Amino acid representation among the library of Sup35 3R mutants. [a]Propagating values represent the frequency of occurrence of the amino acid(s) among the prion-forming clones subsequent to the loss of the wild-type-expressing plasmid; non-propagating values represent the frequency of occurrence of the amino acid among clones that do not form prions after loss of the wild-type-expressing plasmid from the cell. [c]Propensity values are from Toombs *et al*., 2010. The prion propensity score is natural logarithm of the odds ratios calculated in those experiments. [d]The natural logarithm of the odds ratio could not be computed for lysine because zero lysines were observed in the propagating library. [e]As described in Toombs *et al*., 2010, when multiple consecutive prolines were separated by no more than one residue, a value of zero was used for the natural logarithm of the odds ratio for each proline after the first in that cluster (Alberti *et al*. 2009).

Amino acid(s)	**3rd Repeat Prion Propagation Results**				**Previous Prion Propensity Results[c]**	
	Frequency[a]					
	Propagating library	**Non-propagating library**	**ln (OR_{aa})[b]**	***P* value[b]**	**Prion propensity score**	***P* value**
Tryptophan (W)	0.018	0.005	1.40	0.011	0.67	0.32
Tyrosine (Y)	0.074	0.028	1.03	4.2×10^{-5}	0.78	0.099
Glutamic acid (E)	0.015	0.006	0.99	0.070	–0.61	0.51
Cysteine (C)	0.072	0.047	0.45	0.045	0.42	0.43
Phenylalanine (F)	0.068	0.049	0.33	0.15	0.84	0.040

Alanine (A)	0.089	0.069	0.28	0.15	–0.40	0.38
Asparagine (N)	0.038	0.032	0.18	0.57	0.080	0.88
Serine (S)	0.120	0.102	0.18	0.28	0.13	0.68
Glutamine (Q)	0.008	0.007	0.11	1.00	0.069	1.00
Proline (P)	0.048	0.045	0.06	0.81	–1.20[e]	0.0020
Glycine (G)	0.114	0.109	0.05	0.81	–0.039	1.00
Threonine (T)	0.037	0.039	–0.06	0.89	–0.12	0.75
Histidine (H)	0.040	0.043	–0.06	0.90	–0.28	0.50
Aspartic acid (D)	0.054	0.064	–0.19	0.44	–1.28	0.041
Valine (V)	0.085	0.116	–0.35	0.049	0.81	0.015
Leucine (L)	0.040	0063	0.48	0.050	0.040	1.00
Isoleucine (I)	0.025	0.043	–0.57	0.066	0.81	0.015

Arginine (R)	0.054	0.121	–0.88	6.14×10^{-6}	–0.41	0.31
Methionine (M)	0.002	0.009	–1.80	0.0871	0.67	0.19
Lysine (K)[d]	0.000	0.003	NA[d]	0.2652	–1.58	0.028
Groups						
Aromatic (FWY)	0.160	0.082	0.76	3.1×10^{-6}	0.84	0.002
Hydrophobic(FILMV)	0.218	0.280	–0.33	0.0062	0.79	3.0×10^{-5}
Non-aromatic hydrophobic (ILMV)	0.151	0.231	–0.53	9.0×10^{-5}	0.68	8.4×10^{-4}
Charged (DEKR)	0.123	0.194	–0.54	2.0×10^{-4}	–0.89	8.8×10^{-4}
Positive (KR)	0.054	0.124	–0.91	2.1×10^{-6}	–0.73	0.024
Negative (DE)	0.069	0.070	–0.01	1.00	–1.08	0.034
All polar (HNQST)	0.317	0.251	0.33	0.0046	–0.020	0.92
Q/N	0.046	0.039	0.17	0.5203	0.077	0.79

Propagating mutants have a biased amino acid composition.

Given the per-amino acid differences in prion propagation propensity represented by the $\ln(OR_{aa})$ values (Table 6), we examined the sequences found in both the propagating and non-propagating libraries (Table 5) for overall biases in the 10 amino acid mutant region. Because using the same data set to analyze the sequences from which the odds ratio scores were derived was not appropriate, we analyzed the data by randomly splitting the sequence sets into two halves. The first half was used to determine the odds ratio values for each amino acid (that is, to "train" the algorithm) and the second half of the sequences was analyzed using these values (the "test" phase). Ten independent randomizations were undertaken with odds ratio values calculated for each training set and corresponding scores generated for all the sequences in each test set. Prion propagation propensity (PPP) scores were generated as the sum of the numbers of each type of amino acid multiplied by the $\ln(OR_{aa})$ for each amino acid (see Materials and Methods). Each of the ten iterations of the train-test paradigm was plotted in a box-and-whisker plot (Figure 2.3A-J). In each case, propagating clones were found to have a higher PPP score (Figure 2.3, white boxes) than the non-propagating clones (Figure 2.3, grey boxes). Although the separation between each set was highly statistically significant (see Figure 2.3 legend for *P* values), it was not possible to draw a simple threshold line that would separate the propagating and non-propagating clones on the basis of PPP score for any of the iterations.

ORD mutants that failed to propagate were not due to failure to add on to existing wild-type aggregates.

In measuring the ability of 3rd repeat mutants to continue propagation after loss of the wild-type Sup35 plasmid from cells (Figure 2.1), the possibility existed that some of the non-propagating mutants (Table 5) may have failed due to inability to add onto wild-type aggregates

(Figure 2.1B, middle cell) instead. To discount this possibility, plasmids expressing mutant *SUP35* from individual non-propagating clones were isolated and re-transformed in cells containing wild-type [PSI^+] and their phenotype before the loss of wild-type plasmid was examined. Of the 14 clones examined (indicated with an asterisk in Table 5), none completely lacked the ability to propagate in the presence of wild-type aggregates. Only three (FYSVSILDRR, GCPRVVIHVD, PHFALVHSTH) showed slightly reduced efficiency of adding on to wild-type aggregates (pink phenotype in the *ade2-1* assay) (data not shown).

To more comprehensively examine the question of a species barrier influence on the results of the library experiment, the experiment was repeated with an additional selection step. After mating the mutant library strains with wild-type Sup35-containing cells, the selection to remove haploid cells was undertaken in the absence of adenine. Without adenine, only cells that remained [PSI^+] in the presence of both mutant and wild-type plasmids could remain Ade^+ and grow on the sc-ade media. Then, after selecting with 5-FOA to allow for loss of the wild-type plasmid, each clone was examined for its ability to propagate as the sole copy in the cell. DNA sequencing of a smaller set of 19 successful propagators and 26 non-propagators demonstrated that the broad trends shown in Table 1 held in this altered experimental system (Table 7). In brief, overrepresentation of aromatic and polar amino acids and underrepresentation of hydrophobic and charged amino acids in propagating isolates were again observed. However, the effect of the presence of non-aromatic hydrophobic residues was reduced and lacked statistical significance (Table 7).

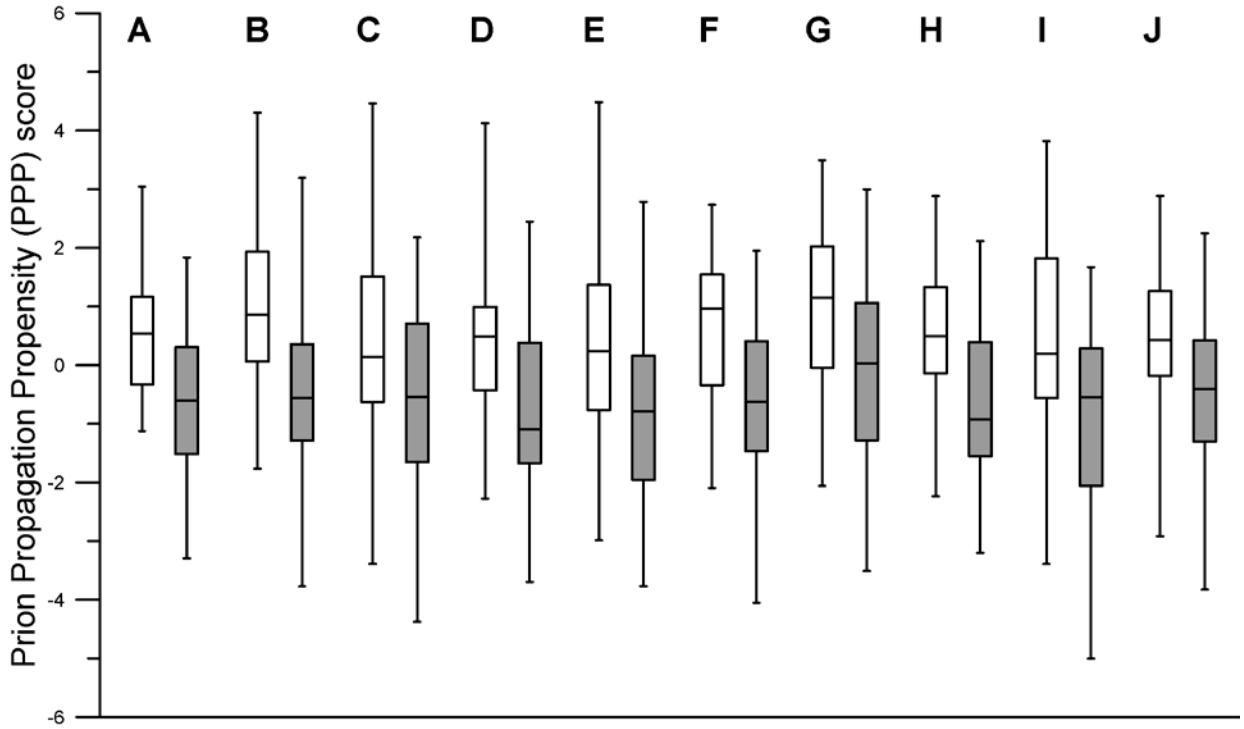


Figure2. 3: Propagation ability of clones with mutant sequences in the 3rd repeat propagation library is based on amino acid composition. To assess propagation ability of library clones, the propagating and non-propagating data sets shown in Table 5 were randomly divided using Excel in a jackknife experiment into two half-sets. The first half-set, called the training set, was used to derive prion propagation propensities ($\ln(OR)_{aa}$, generated the same way as in Table 6 see Materials and Methods for more detail) for each amino acid. The second half-set, called the test set, used the propensities determined in the training set to generate scores for each of the propagating and non-propagating clones. The random division was iteratively repeated 10 times and each of the sets is shown graphically in a box-and-whisker plot. In each case, statistical analysis using a two-sided *t*-test demonstrated clear statistical differences (at α=0.05) between the means of the propagating clones (shown with white boxes) and the non-propagating clones (shown with red boxes). For propagating clones, *N*=32 in the training set and *N*=33 in the test set. For non-propagating clones, *N*=43 in the training set and *N*=44 in the test set. *P*-values for each of the ten iterations are as follows: A, $p<0.0001$; B, $p<0.0001$; C, $p=0.0401$; D, $p=0.0013$; E, $p=0.0015$; F, $p<0.0001$; G, $p=0.0027$; H, $p=0.0004$; I, $p=0.0023$; J, $p=0.0029$.

Table2. 7: Library amino acid representation after initial selection for 3R mutants able to add on to wild-type fibers. Briefly, the experiment reported in Tables 1 and 2 was repeated, with the additional requirement that mutant aggregates be able to add on to the wild-type fibers (shown in Figure 2.1B, middle cell). The ability to add on to wild-type fibers was selected for by an additional plating step on medium lacking adenine to ensure all cells were Ade^+ prior to the loss of the wild-type plasmid (Figure 2.1B, right cell). [a]See Table 2 for an explanation of terms.

Groups of Amino acid(s)	**3rd Repeat Prion Propagation Results** (initial requirement to add on to wild-type Sup35 fibers)				**Previous Prion Propensity Results[a]**	
	Frequency[a]					
	Propagating library	**Non-propagating library**	**ln (OR_{aa})[a]**	***P* value[a]**	**Prion propensity score**	***P* value[a]**
Aromatic (FWY)	0.253	0.073	1.46	1.4×10^{-7}	0.84	0.002
Non-aromatic hydrophobic (ILMV)	0.179	0.215	–0.23	0.40	0.68	8.4×10^{-4}
Charged (DEKR)	0.047	0.188	–1.54	1.4×10^{-5}	–0.89	8.8×10^{-4}
All polar (HNQST)	0.347	0.258	0.43	0.047	–0.020	0.92

Inserting hydrophobic residues into the ORD was not able to abolish propagation.

To examine the effect of hydrophobic amino acids on prion activity, we made targeted mutations in the 3rd and 4th repeat of *SUP35* ORD (regions that have been suggested to have a role in prion propagation (Osherovich, 2004 and Toombs, *et al*. 2011). Libraries of *SUP35* clones with variable number of non-aromatic hydrophobic amino acids inserted or deleted were created (Figure 2.4A). Toombs *et al*. has suggested that hydrophobic residues encourage prion formation. This suggestion was supported by an additional study in our lab where we found an almost linear correlation between the numbers of hydrophobic residues randomly inserted in different positions of Sup35 nucleation domain and the rate of prion formation. Furthermore, other research has suggested that prion formation may be promoted by aromatic rather than hydrophobic amino acids (Chernoff *et al*. 2010). In order to focus on hydrophobicity rather than aromaticity, the effect of valine and isoleucine (non-aromatic hydrophobic residues) were tested. Methionine and leucine were excluded because M has low frequency and L has neutral effects on prion activity, likely due to its low β-sheet propensity. Unexpectedly, the prion formation activity assay of these mutants showed there was a slightly positive effect of the one of +2 and one of +6 constructs, but none of the others (Figure 2.4B). These results suggested the sensitivity to the position of insertion.

A

	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	89	90	91	92	93	94	94	95	96	97	98	99	100	101	102	103	104	105
WT	Q	Q	Y	N	P	D	A	G	Y	Q	Q	Q	Y	N	P	Q	G	G	Y	Q	Q	Y	N	P	Q	G	G	Y	Q	Q	Q	F	N	P	Q	G	G	R	G	N						
(-)4Y	Q	Q	Y	N	P	D	A	G	X	Q	Q	Q	X	N	P	Q	G	G	X	Q	Q	X	N	P	Q	G	G	Y	Q	Q	Q	F	N	P	Q	G	G	R	G	N						
(-)2A	Q	Q	Y	N	P	D	A	G	X	Q	Q	Q	Y	N	P	Q	G	G	X	Q	Q	Y	N	P	Q	G	G	Y	Q	Q	Q	F	N	P	Q	G	G	R	G	N						
(-)2B	Q	Q	Y	N	P	D	A	G	Y	Q	Q	Q	X	N	P	Q	G	G	Y	Q	Q	X	N	P	Q	G	G	Y	Q	Q	Q	F	N	P	Q	G	G	R	G	N						
(+)2A	Q	Q	Y	N	P	D	A	G	Y	Q	Q	Q	I	Y	N	P	Q	G	G	Y	V	Q	Q	Y	N	P	Q	G	G	Y	Q	Q	Q	F	N	P	Q	G	G	R	G	N				
(+)2B	Q	Q	Y	N	P	D	A	G	Y	Q	Q	Q	V	Y	N	P	Q	G	G	Y	Q	Q	I	Y	N	P	Q	G	G	Y	Q	Q	Q	F	N	P	Q	G	G	R	G	N				
(+)6A	Q	Q	Y	N	P	D	A	I	G	I	Y	Q	V	Q	Q	Y	N	I	P	Q	G	G	Y	V	Q	V	Q	Y	N	P	Q	G	G	Y	Q	Q	Q	F	N	P	Q	G	G	R	G	N
(+)6B	Q	Q	Y	N	P	D	A	G	I	Y	Q	Q	V	Q	Y	N	P	I	Q	G	G	Y	V	Q	Q	Y	I	N	V	P	Q	G	G	Y	Q	Q	Q	F	N	P	Q	G	G	R	G	N

B

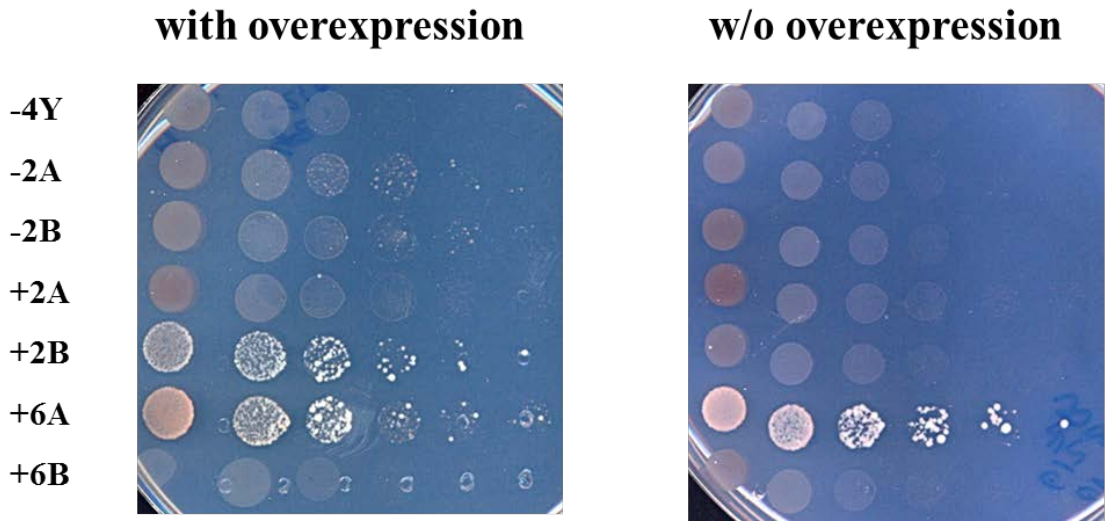


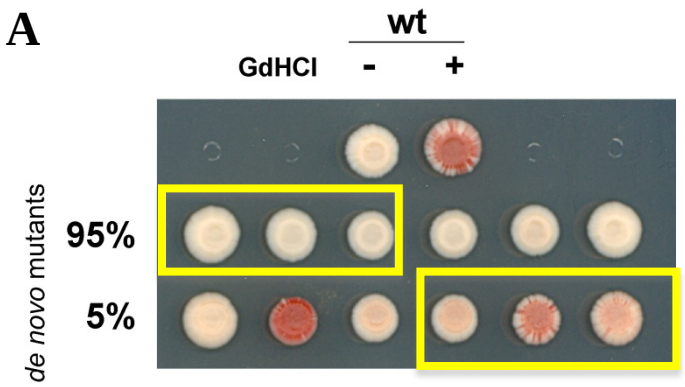
Figure2. 4: Mutagenesis in the ORD. **(A)** Targeted mutants that insert/delete non-aromatic hydrophobic residues in the 3rd and 4th repeats of the ORD were introduced to wild-type [PSI$^+$]-containing cells. Seven clones were made; -4Y, -2A, -2B, +2A, +2B, +6A, and +6B. **(B)** Induction experiment. Strains expressing the mutant SUP35 were tested for *de novo* prion formation. Strains expressing the mutant were transformed with either plasmid pKT24 containing the *GAL1* promoter (uninduced) or with a derivative of pKT24 expressing the matching prion forming domain from the *GAL1* promoter (induced). Strains were then grown in galactose/raffinose –trp + ade dropout medium and serial dilutions plated onto medium lacking adenine to select for [PSI$^+$].

New mutants in the 3rd repeat of the ORD can be generated that successfully propagate prions.

To assess if novel mutations could be randomly generated solely on the basis of prion propagation propensity values, six randomly-designed mutant versions that were predicted to be very good propagators (95th percentile) and six versions predicted to propagate poorly (5th percentile) were cloned by recombination and expressed under the control of the SUP35 promoter in yeast cells. After 5-FOA treatment and loss of wild-type plasmid, cells were assessed for ability to propagate the mutant versions (Figure 2.5A). While all six predicted propagators were uniformly white when plated on YPD in the *ade2-1* assay, the predicted non-propagators were more variable. Two clones showed greater prion loss (DDTNFRCLIR and RDICPRIFPD), but the others showed lower levels of prion loss (Figure 2.5A).

Mutants in the 4th and 5th repeats of the ORD can also successfully propagate prions.

Six of the same *de novo* mutants that were created for the 3rd repeat were also generated in the context of the 4th and 5th repeats. Three 95% clones and three 5% clones for the 4th and 5th repeats were compared with their 3rd repeat counterpart generated in Figure 2.5A. The 4R and 5R versions behaved very similarly to the 3R clones (Figure 2.5B).



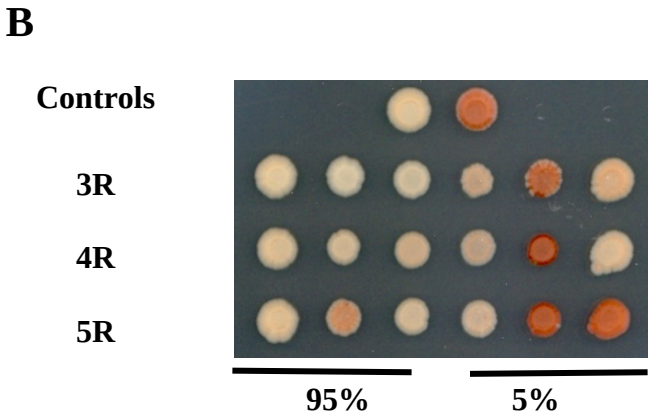


Figure2. 5: *De novo* repeat mutants. *De novo* repeat mutants randomly selected to score at the top (95th percentile—good propagators) or bottom (5th percentile—poor propagators) using the prion propagation propensity (PPP) score behave largely as predicted in yeast cells. Mutants were introduced to wild-type [*PSI*$^{+}$]-containing cells; subsequent to 5-FOA selection for loss of the wild-type *URA*-containing plasmid, mutant cells were plated on YPD medium and examined for color phenotype. The sequences were generated randomly (see Methods) using an Excel spreadsheet and all amino acids were eligible for inclusion, except E, K, M, Q, and W, because of their small sample size in the original library experiment (see Table 2). **(A)** *De novo* 3rd repeat mutants. The 95th percentile sequences, from left to right, were as follows: **PVYDHIYYAS**, **YNGINAYLFN**, **TARYGAAFGY**, GAINYTYCVA, YNITYDVTYN, and FIDYAPPNAY. The 5th percentile sequences were: LIVVIGFIPT, DDTNFRCLIR, GRVLFRNGIG, **TRPDPICNLR**, **RDICPRIFPD**, and **LILALRNNIN**. The clones in bold above are indicated in a yellow box and represent the 95% and 5% clones analyzed in the context of the three different repeats in the (B) portion of this figure. **(B)** *De novo* 3rd, 4th, and 5th repeat mutants. Wild-type [*PSI*$^{+}$]-containing cells and wild-type cells treated with GdHCl are in the top row, left, and top row, right, respectively. 2nd row: 3rd repeat mutants (same as in (A). 3rd row: Same mutants placed within the 4th repeat. 4th row: Same mutants placed within the 5th repeat. For rows 2-4, the clones observed are three 95% clones on the left and three 5% clones on the right (bold above and yellow-boxed in (A)

Yeast PFDs that successfully propagate show similar compositional biases.

The very large data set of Alberti *et al.* (Alberti *et al.* 2009), in which the 100 most compositionally similar yeast ORFs to the PFDs of the prions Sup35, Ure2, Rnq1, and New1 were tested in four distinct assays of prion-like activity, provides a useful tool in assessing the validity of predictions from these experiments and others. The four assays used in this landmark study included three distinct measures of aggregation (formation of fluorescent foci when expressed as an EYFP fusion, formation of SDS-resistant aggregates in an SDD-AGE assay, and *in vitro* aggregation by thioflavin-T fluorescence after bacterial expression) and one assay (replacement of the PFD of Sup35 with a portion of each ORF) that requires that the fusion protein be successfully propagated in *S. cerevisiae* (Alberti *et al.* 2009).

For corroborating our predictions of effects of mutation on prion propagation, the Alberti *et al.* data set contains two useful subsets (Alberti *et al.* 2009), 18 proteins identified from among the 100 tested actually passed all four assays and represent likely new prions, forming one important subset. Among the remaining 82 proteins that did not pass all four tests, many failed in multiple tests. However, 12 of the proteins passed all three of the aggregation assays, but failed to propagate as Sup35 fusion proteins, indicating a possible defect in propagation and making this a useful subset for comparison (Alberti *et al.* 2009).

Using the two protein subsets, we looked for differences in amino acid content (Figure 2.6). In terms of Q/N content, both subsets were identical (Figure 2.6A). However, in both content of aromatic residues (FWY, Figure 2.6 B) and non-aromatic hydrophobic residues (ILMV, Figure 2.6C) there appeared to be differences in the distributions that were in keeping with the prion propagation propensities identified in Table 6. Only the difference in ILMV residues proved to be significant ($P < 0.05$).

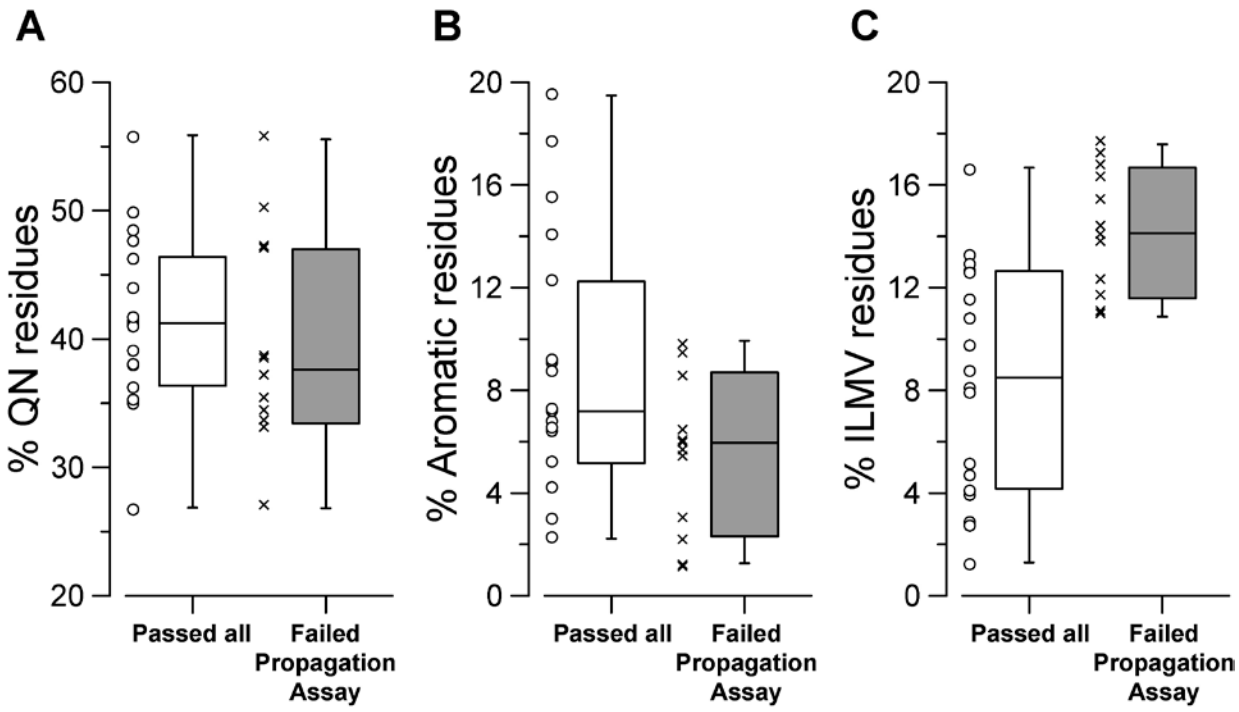


Figure2. 6: Re-analysis of the prion domains. Re-analysis of the prion domains for the subset of proteins examined by Alberti, *et al*. is suggestive of the importance of aromatic and non-aromatic hydrophobic amino acid residues in prion propagation. The proteins in the Alberti, *et al*., data set that passed all tests (left side of each panel) or those in the subset that passed all tests except the Sup35-fusion protein assay (the only assay that requires prion propagation, right side of each panel), were scored on the percentage of the following amino acids in the identified "prion-like domains" (see Alberti, *et al*., 2009) and graphed using box-and-whisker and scatter plots: (A) total percentage of Q and N residues; (B) total percentage of F, W, and Y residues; and (C) total percentage of I, L, M, and V residues.

Discussion

The oligopeptide repeat domain of Sup35 has long been known to be important for maintenance of [PSI$^+$] in yeast cells (DePace *et al.* 1998; Osherovich *et al*. 2004; Liu J-J *et al.* 2002; Ohhashi *et al.* 2010; Toombs *et al*. 2011; Crist *et al*. 2003; Tank *et al.* 2007; Kalastavadi *et al.* 2008). In marked contrast to experiments that deleted repeats (Shkundina *et al*. 2006), experiments that scrambled the ORD region demonstrated no reduction in efficiency of prion nucleation, propagation, or sensitivity to overexpression of the chaperone Hsp104 (Toombs *et al*. 2011). Therefore, reduced importance of the primary sequence of the ORD region is clear. Furthermore, the inability to replace the ORD with scrambled versions of the nucleation domain (Toombs *et al*. 2011) indicates that amino acid composition requirements are distinct between the ND and ORD regions of the Sup35 PFD.

Given the divergent amino acid compositions of these two regions, dissection of the residues that are required in the ORD for [PSI$^+$] maintenance is important for understanding the mechanism(s) of yeast prion propagation. The only previous study to examine residue requirements in the ORD proposed the importance of tyrosine residues in fiber fragmentation (Alexandrov *et al*. 2008). However, subsequent experiments in which scrambled versions of the ND replaced the ORD (Toombs *et al*. 2011), showed nearly identical tyrosine content that was still insufficient to promote propagation.

Earlier work to analyze the overall amino acid composition requirements of the Sup35 prion (Toombs *et al*.,2010; Ross *et al.* 2010) provided evidence of the importance of many residues in promoting and opposing prion propensity. However, because this experiment required both prion formation and propagation, it was not possible to separate the distinct requirements of the two processes. Here, we have undertaken an experiment that explicitly

separates those two functions (Figure 2.1B) and defines a set of amino acid composition requirements for propagation (Table 6) that are overlapping but distinct from experiments in which both prion formation and propagation are required (Table 2 and 4). We can conclude that while prions generally favor the presence of hydrophobic and aromatic residues in prion domains, while disfavoring charged residues (Parham *et al*. 2001), non-aromatic hydrophobic residues (ILMV) are disfavored specifically in terms of their effects on the propagation function (Table 6).

Analysis of the different amino acids present in the ND and ORD previously suggested that given the overrepresentation of polar residues and the underrepresentation of glycine and proline in the ND relative to the ORD (Toombs *et al*. 2011), we might expect to see some effect on the presence of these residues in the ORD. However, the original experiments demonstrated that, of these, only proline had a negative effect on prion formation (Toombs *et al*., 2009). In these experiments, polar residues in fact had a slightly positive effect on prion propagation propensity, and proline and glycine were neutral (Table 6). These results underscore that there is still much to be learned about prion amino acid composition requirements and that a simple analysis of what is already present in the ND and ORD is not sufficient to gain a mechanistic understanding of the forces at work. Indeed, many of the amino acid residues examined in this work (Table 2, 4, 6) and earlier work (Toombs *et al*.,2009) are not normally present in the Sup35 PFD at all.

Since amino acid composition appeared to be the dominant factor that affected success or failure of propagation, we examined the amino acid content of the proteins identified in the Alberti *et al.* prion screen (Alberti *et al* 2009). Eighteen putative prions that passed all of their screening assays were quite similar in overall Gln/Asn content to the 12 proteins that failed in the

one assay requiring propagation (Figure 2.5A). However, non-aromatic hydrophobic residues were significantly enriched in the failed propagators compared to those that propagated well (Figure 2.5C). The observed differences in aromaticity, while not statistically significant, also suggested broad agreement with our data (Figure 2.5B).

References

1. Alberti S, Halfmann R, King O, Kapila A, Lindquist S. 2009. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell 137:146–58.
2. Alexandrov AI, Polyanskaya AB, Serpionov G V, Ter-Avanesyan MD, Kushnirov V V. 2012. The effects of amino acid composition of glutamine-rich domains on amyloid formation and fragmentation. PloS one 7:e46458.
3. Alexandrov IM, Vishnevskaya AB, Ter-Avanesyan MD, Kushnirov V V. 2008. Appearance and propagation of polyglutamine-based amyloids in yeast: tyrosine residues enable polymer fragmentation. The Journal of biological chemistry 283:15185–92.
4. Brown DR, Qin K, Herms JW, Madlung A, Manson J, Strome R, Fraser PE, Kruck T, Von Bohlen A, Schulz-Schaeffer W, Giese A, Westaway D, Kretzschmar H. 1997. The cellular prion protein binds copper in vivo. Nature 390:684–7.
5. 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 (New York, N.Y.) 268:880–4.
6. Chiti F, Dobson CM. 2006. Protein misfolding, functional amyloid, and human disease. Annual review of biochemistry 75:333–66.
7. Crist CG, Nakayashiki T, Kurahashi H, Nakamura Y. 2003. [PHI+], a novel Sup35-prion variant propagated with non-Gln/Asn oligopeptide repeats in the absence of the chaperone protein Hsp104. Genes to cells : devoted to molecular & cellular mechanisms 8:603–18.
8. DePace AH, Santoso A, Hillner P, Weissman JS. 1998. A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. Cell 93:1241–52.

9. Derdowski A, Sindi SS, Klaips CL, DiSalvo S, Serio TR. 2010. A size threshold limits prion transmission and establishes phenotypic diversity. Science (New York, N.Y.) 330:680–3.
10. Derkatch IL, Bradley ME, Hong JY, Liebman SW. 2001. Prions affect the appearance of other prions: the story of [PIN(+)]. Cell 106:171–82.
11. Flechsig E, Shmerling D, Hegyi I, Raeber AJ, Fischer M, Cozzio A, Von Mering C, Aguzzi A, Weissmann C. 2000. Prion protein devoid of the octapeptide repeat region restores susceptibility to scrapie in PrP knockout mice. Neuron 27:399–408.
12. Helsen CW, Glover JR. 2012. Insight into molecular basis of curing of [PSI+] prion by overexpression of 104-kDa heat shock protein (Hsp104). The Journal of biological chemistry 287:542–56.
13. Kalastavadi T, True HL. 2008. Prion protein insertional mutations increase aggregation propensity but not fiber stability. BMC biochemistry 9:7.
14. Lindquist S. 1997. Mad cows meet psi-chotic yeast: the expansion of the prion hypothesis. Cell 89:495–8.
15. Liu JJ, Lindquist S. 1999. Oligopeptide-repeat expansions modulate “protein-only” inheritance in yeast. Nature 400:573–6.
16. Liu J-J, Sondheimer N, Lindquist SL. 2002. Changes in the middle region of Sup35 profoundly alter the nature of epigenetic inheritance for the yeast prion [PSI+]. Proceedings of the National Academy of Sciences of the United States of America 99 Suppl 4:16446–53.
17. MacLea KS, Ross ED. 2011. Strategies for identifying new prions in yeast. Prion 5:1–6.
18. Marchante R, Rowe M, Zenthon J, Howard MJ, Tuite MF. 2013. Structural Definition Is Important for the Propagation of the Yeast [PSI+] Prion. Molecular Cell 50:675–685.

19. Michelitsch MD, Weissman JS. 2000. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. Proceedings of the National Academy of Sciences of the United States of America 97:11910–5.
20. Ness F, Ferreira P, Cox BS, Tuite MF. 2002. Guanidine hydrochloride inhibits the generation of prion “seeds” but not prion protein aggregation in yeast. Molecular and cellular biology 22:5593–605.
21. Ohhashi Y, Ito K, Toyama BH, Weissman JS, Tanaka M. 2010. Differences in prion strain conformations result from non-native interactions in a nucleus. Nature chemical biology 6:225–230.
22. Osherovich LZ, Cox BS, Tuite MF, Weissman JS. 2004. Dissection and design of yeast prions. PLoS biology 2:E86.
23. Parham SN, Resende CG, Tuite MF. 2001. Oligopeptide repeats in the yeast protein Sup35p stabilize intermolecular prion interactions. The EMBO journal 20:2111–9.
24. Paushkin S V, Kushnirov V V, Smirnov VN, Ter-Avanesyan MD. 1996. Propagation of the yeast prion-like [psi+] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor. The EMBO journal 15:3127–34.
25. Prusiner SB, Scott MR, DeArmond SJ, Cohen FE. 1998. Prion protein biology. Cell 93:337–48.
26. Prusiner SB. 1982. Novel proteinaceous infectious particles cause scrapie. Science (New York, N.Y.) 216:136–44.
27. Prusiner SB. 1991. Molecular biology of prion diseases. Science (New York, N.Y.) 252:1515–22.

28. Ross ED, Edskes HK, Terry MJ, Wickner RB. 2005. Primary sequence independence for prion formation. Proceedings of the National Academy of Sciences of the United States of America 102:12825–30.
29. Ross ED, MacLea KS, Anderson C, Ben-Hur A. 2013. A bioinformatics method for identifying Q/N-rich prion-like domains in proteins. Methods in molecular biology (Clifton, N.J.) 1017:219–28.
30. Ross ED, Toombs JA. 2010. The effects of amino acid composition on yeast prion formation and prion domain interactions. Prion 4:60–65.
31. Rüdiger S, Germeroth L, Schneider-Mergener J, Bukau B. 1997. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. The EMBO journal 16:1501–7.
32. Rüdiger S, Schneider-Mergener J, Bukau B. 2001. Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. The EMBO journal 20:1042–50.
33. Shkundina IS, Kushnirov V V, Tuite MF, Ter-Avanesyan MD. 2006. The role of the N-terminal oligopeptide repeats of the yeast Sup35 prion protein in propagation and transmission of prion variants. Genetics 172:827–35.
34. Sondheimer N, Lindquist S. 2000. Rnq1: an epigenetic modifier of protein function in yeast. Molecular cell 5:163–72.
35. Song Y, Wu Y-X, Jung G, Tutar Y, Eisenberg E, Greene LE, Masison DC. 2005. Role for Hsp70 chaperone in Saccharomyces cerevisiae prion seed replication. Eukaryotic cell 4:289–97.

36. Tanaka M, Collins SR, Toyama BH, Weissman JS. 2006. The physical basis of how prion conformations determine strain phenotypes. Nature 442:585–9.

37. Tank EMH, Harris DA, Desai AA, True HL. 2007. Prion protein repeat expansion results in increased aggregation and reveals phenotypic variability. Molecular and cellular biology 27:5445–55.

38. Ter-Avanesyan MD, Dagkesamanskaya AR, Kushnirov V V, Smirnov VN. 1994. The SUP35 omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant [psi+] in the yeast Saccharomyces cerevisiae. Genetics 137:671–6.

39. Ter-Avanesyan MD, Kushnirov V V, Dagkesamanskaya AR, Didichenko SA, Chernoff YO, Inge-Vechtomov SG, Smirnov VN. 1993. Deletion analysis of the SUP35 gene of the yeast Saccharomyces cerevisiae reveals two non-overlapping functional regions in the encoded protein. Molecular microbiology 7:683–92.

40. Toombs JA, Liss NM, Cobble KR, Ben-Musa Z, Ross ED. 2011. [PSI+] maintenance is dependent on the composition, not primary sequence, of the oligopeptide repeat domain. PLoS ONE 6:e21953.

41. Toombs JA, McCarty BR, Ross ED. 2010. Compositional determinants of prion formation in yeast. Molecular and Cellular Biology 30:319–332.

42. Toombs JA, Petri M, Paul KR, Kan GY, Ben-Hur A, Ross ED. 2012. De novo design of synthetic prion domains. Proceedings of the National Academy of Sciences of the United States of America 109:6519–24.

43. Vitrenko YA, Pavon ME, Stone SI, Liebman SW. 2007. Propagation of the [PIN+] prion by fragments of Rnq1 fused to GFP. Current genetics 51:309–19.

44. Wadsworth JDF, Hill AF, Beck JA, Collinge J. 2003. Molecular and clinical classification of human prion disease. British medical bulletin 66:241–54.

45. Wegrzyn RD, Bapat K, Newnam GP, Zink AD, Chernoff YO. 2001. Mechanism of prion loss after Hsp104 inactivation in yeast. Molecular and cellular biology 21:4656–69.

46. Weissmann C. 1991. Spongiform encephalopathies. The prion's progress. Nature 349:569–71.

47. Wickner RB, Masison DC, Edskes HK. 1995. [PSI] and [URE3] as yeast prions. Yeast (Chichester, England) 11:1671–85.

48. Wickner RB. 1994. [URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae. Science (New York, N.Y.) 264:566–9.

Chapter 3: Prion-like domains and neurodegenerative diseases[2]

Summary

Faultless protein function is crucial for all cells within organisms. Diseases may be caused by proteins, especially essential proteins, which are functioning inappropriately. The conversion of pr[2]otein to misfolded-amyloid aggregates is one of the manners that proteins can lose function or misbehave. This form of aggregated proteins has been associated with several neurodegenerative diseases, including: Alzheimer Disease, Huntington's Disease, Parkinson's Disease, Amyotrophic Lateral Sclerosis (ALS), and Front Temporal lobar Degeneration (FTLD). Additional proteins may be associated with neurodegenerative disease. This study examined three prion-like domain-containing proteins to see if they have prion activity.

Introduction

The common definition of prion is infectious protein that results from the conversion of proteins from its natural form into a self-templating aggregated form. Thus, the self-templating is the basis of infection; however, many human proteins (such as Aβ and Tau) also form self-templating aggregates, but do not act as prions. Although these types of proteins can transfer the infection from one cell to another cell by an unidentified mechanism, and spread the infectious phenotype to surrounding tissues, these diseases have not been shown to be infectious between individuals. Furthermore, attempts to create infectious material from pure protein have been unsuccessful for these proteins. Therefore, they are called prionoid, or prion-like, instead.

[2] Zobaida Ben Musa and Eric Ross

Prion diseases are a dissimilar cluster of incurable human and animal neurodegenerative disorders, including: Creutzfelt-Jacob Disease (CJD), Fatal Familial Insomnia (FFI), Kuru, and animal Transmissible Spongiform Encephalopathies (TSE's), which include Bovine Spongiform Encephalopathy (BSE) in cattle, Scrapie in sheep and goats, and Chronic Wasting Disease in cervids. Presently, these diseases are all incurable. They can be acquired by infection, inherited, or arise spontaneously (Prusiner *et al*., 1982; 1998). Although this group of diseases has different clinical manifestations depending on the involved tissues, all have the same causative agent, an abnormal isoform of mammalian prion protein PrP. PrP is a cell-surface glycoprotein, the infection is spread by conversion of normal cellular protein isoform (PrP^{C}) that is rich in α-helices into insoluble aggregates consisting of β-sheets in PrP^{SC}. The self-templating ability of PrP^{SC} forms the basis of prion infectivity by recruiting PrP^{C} and templating its structural conversion, leading to depletion of the normal and functional isoform from the population (Prusiner *et al.,* 1982; Golby and Prusiner *et al.,* 2011). Unlike other kinds of infectious etiology (bacteria and viruses), prions are transferred, not by nucleic acid, but through merely protein (Laurent *et al*., 1996).

While PrP is the only prion-forming protein known in mammals, several prion-forming proteins have been identified in the yeast, *Saccharomyces cerevisiae*: Sup35p forms the [PSI^{+}] prion, Ure2p forms [URE3], Nup100 forms [NUP100 $^{+}$], Cyc8p forms [OCT^{+}], Mot3p forms [MOT3], Swi1p forms [SWI^{+}], Sfp1 forms [ISP^{+}], Mod5 forms [MOD^{+}], Nup100 forms [NUP100 $^{+}$]. With exception of Mod5, all others are characterized by having prion domains rich in glutamine and asparagine amino acid (Q/N).

Recently, some progress has been made in predicting the aggregation propensity of Q/N-rich domains. Previous studies showed that prion propensity of Q/N-rich proteins are determined

predominately by amino acid composition, not primary sequence (Ross *et al.*, 2005). Therefore, Alberti *et al*. (2009) developed an algorithm to identify domains with high compositional similarity to yeast prion-forming domains (PFDs). However, while this algorithm proved very effective at identifying prion candidates, it was completely ineffective at predicting which of the top ranked candidates could act as prions. Our lab developed a quantitative in vivo method for ranking the prion propensity of each amino acid. Using this data, PAPA was developed, the first algorithm capable of predicting whether a given Q/N-rich protein would show prion activity.

In recent times, prion-like domains (PrLDs), defined as domains that compositionally resemble yeast prion forming domains (PFD). PrLDs have been recognized in more than two hundred human proteins, (see Table 1). Almost all of them are RNA-binding proteins containing RNA recognition motifs (RRM) (King *et al*., 2010). Some of these, including FUS, TDP-43, TAF15 and EWSR1, have been associated with neurodegenerative diseases, including: Amyotrophic lateral sclerosis (ALS also known as Lou Gehrig's disease), Frontemporal Lobar Degeneration (FTLD), and inclusion body myopathy with Frontemporal Dementia (IBMPFD) (King *et al*., 2010; Da Cruz, and Cleveland *et al.*, 2011; Weihl *et al.*, 2008; Couthouis *et al*., 2012; Neumann *et al*., 2011). All four proteins contain regions identified as prion-like by the Alberti algorithm. All except TDP-43 are predicted to be highly aggregation-prone by PAPA, while TDP-43 is predicted to be right at the threshold for aggregation.

Recently, two new prion-like domain-containing proteins were added to the list of diseases associated with PrLDs. In examining two families with an inherited form of IBMPFD, two mutations were linked to the disease. In one family, an aspartic acid to valine substitution at position 290 of hnRNPA2 was linked to disease. In a second family, a similar aspartic acid to valine substitution at the position 262 in a related protein, hnRNPA1, was linked to disease.

These proteins have been tested by using both Alberti *et al*. (2009) and PAPA to predict prion activity, Figure 3.1. PAPA scored the wild-type proteins right below threshold for prion formation (cutoff 0.05). While the Alberti *et al*. algorithm predicts that these proteins should form prions. Remarkably, PAPA properly predicts that incorporation of the disease-associated mutations should drive the aggregation propensity past the threshold for aggregation. By contrast, the Alberti *et al* algorithm predicts that these mutations should have slight effect on aggregation. Both mutations resulted in mislocalization and inclusion body formation by respective proteins. When expressed in *Drosophila*, the mutant proteins resulted in muscle degeneration. Likewise, when the wild-type or mutant PrLDs were inserted into Sup35 in the place of Sup35 PFD and expressed in yeast, the fusion proteins showed mutation-dependent prion formation. These results support the value of yeast proteins as a model system to detect prion activities among human proteins.

The rapid expansion of the number of human proteins that are involved in neurodegenerative diseases suggests that there are likely more disease-associated proteins yet to be identified. Based on the observation that fragments from hnRNPA2 and hnRNPA1 showed mutation-dependent prion formation when inserted into the Sup35 prion forming domain (Kim *et. al*., 2013), other human protein fragments that were predicted by PAPA to have high prion propensity were tested to see if they might support prion activity in yeast.

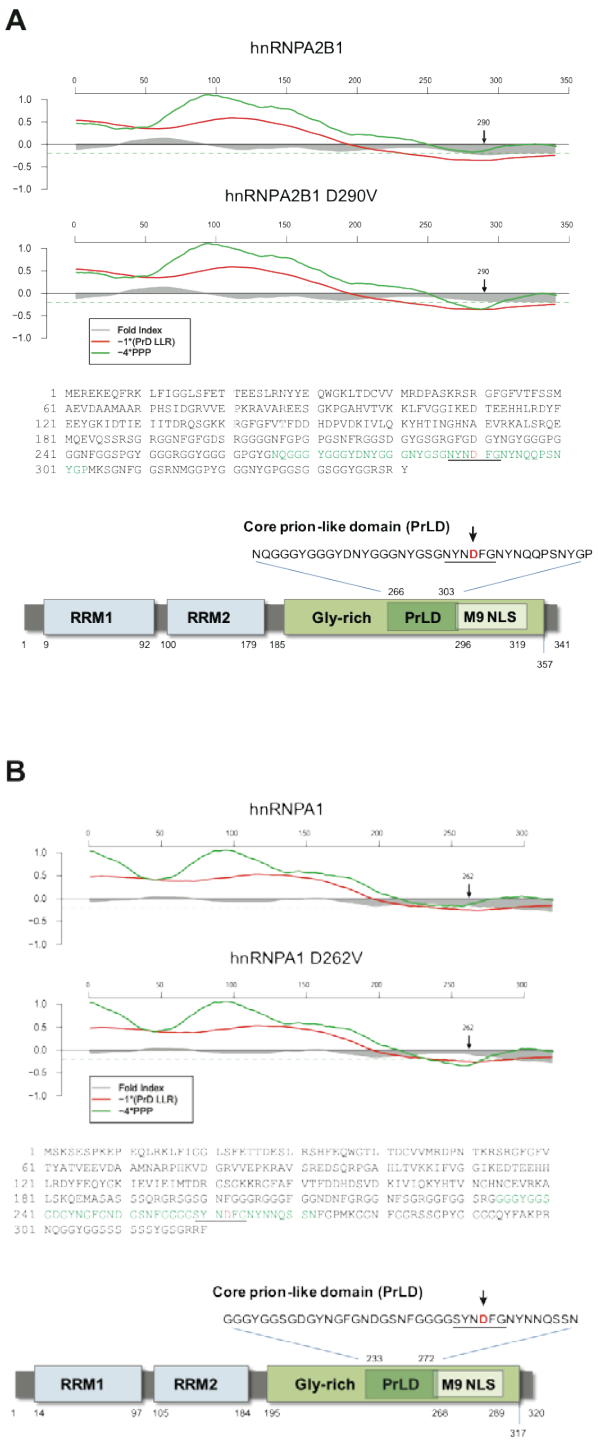


Figure3. 1: The disease causing mutations impact a PrLD In hnRNPA2B1 and hnRNPA. **(A and B)** Prion domain prediction for hnRNPA2B1 and hnRNPA1. The top panel shows prion propensity predictions for wild types. The lower panel shows the prediction after introducing the mutations, the D290V mutation in hnRNPA2B and the D262V mutation in hnRNPA1. FoldIndex predictions (grey line) with values less than zero indicate a region that is predicted to

be intrinsically disordered. A prion prediction algorithm developed by Alberti et al. (red line) identifies potential prion domains (indicated by values below the dotted green line) in both the wild-type and mutant proteins. By contrast, PAPA (green line) predicts that the wild-type proteins narrowly missed the cutoff for prion-like aggregation (indicated by values below the dotted green line), while each of the disease-associated mutants were predicted to make this region more prionogenic and enough to shift the proteins below this threshold. (**D** and **C)** Domain architecture of hnRNPA2 and hnRNPA1, showing the positions of the disease-associated mutation (red).

Table**3. 1 :** Human proteins from Ensembl release GRCh37.59 were scanned for prion-like domains (PrLD's) using the Alberti or PAPA.Proteins with RRM domains (PFAM ID PF00076.15) were identified using BioMart. 29 of 210 RRM-bearing proteins were identified to have the PrLD's according to the Alberti algorithm and are ranked in the entire proteome and among RRM proteins. The position of the PrLD's and a core region of highest score are provided. In PAPA, yeast proteins that are predicted to have an extended intrinsically unfolded region (minus value of FoldIndex score) and have prion propensity scores past PAPA threshold (greater than 0.05) are considered prion-forming. The sequence regions that almost satisfy both criteria are given in the table, along with the corresponding score. Scores that passed both thresholds are indicated in red. The toxicity and aggregation phenotype upon overexpression in yeast are provided. ND = not determined. The three proteins that have been chosen to be tested are checked with red. Table adapted from (King, Gitler, & Shorter, 2012). NB: the table does not show all 29 proteins.

Protein	Prion domain rank (whole genome) (Alberti et al., 2009)	Prion domain rank (RRM proteins) (Alberti et al., 2009)	Prion domain (core) residues (Alberti et al., 2009)	Prion domain central residues (Toombs et al., 2010)	Prion propensity score (FoldIndex) (Toombs et al., 2010)	Yeast overexpression phenotype (toxicity & localization) (Couthouis et al., 2011)
FUS	12	1	1–237 (118–177)	40–80	0.101 (−0.211)	Highly toxic, cytoplasmic aggregates
TAF15	22	2	1–152 (33–92)	33–73	0.126 (−0.268)	Mildly toxic, cytoplasmic aggregates
EWSR1	25	3	1–280 (205–264)	209–249	0.057 (−0.277)	Mildly toxic, cytoplasmic aggregates
HNRPDL	27	4	316–420 (341–400)	353–393	0.117 (−0.29)	Not toxic, cytoplasmic aggregates
HNRNPD	29.5	5	262–355 (281–340)	292–332	0.164 (−0.291)	Mildly toxic, diffuse nuclear
HNRNPA2B1	32	6	197–353 (276–335)	274–314	0.043 (−0.208)	Highly toxic, cytoplasmic aggregates
HNRNPA1	38	7	186–372 (266–325)	278–318	0.093 (−0.092)	Highly toxic, cytoplasmic aggregates
HNRNPAB	39	8	235–327 (235–294)	253–293	0.123 (−0.327)	ND
HNRNPA3	41	9	207–378 (287–346)	302–342	0.057 (−0.194)	No expression
TDP -43	43	10	277–414 (301–360)	361–401	0.043 (0.001)	Highly toxic, cytoplasmic aggregates
TIA1	53	11	292–386 (292–351)	307–347	0.115 (−0.079)	Highly toxic, cytoplasmic aggregates
HNRNPA1L2	57	12	198–320 (243–302)	227–267	0.052 (−0.091)	ND
HNRNPH1	63	13	382–472 (388–447)	407–447	0.137 (0.039)	ND
SFPQ	79	14	41–104 (41–100)	638–678	−0.077 (0.054)	ND

Materials and methods

Strains and media.

Strains and media: Standard yeast media and methods were used, except that YPD contained 0.5% yeast extract instead of the standard 1%. Yeast were grown at 30°C in all experiments. All experiments were performed in YER 632 (*α kar1-1 SUQ5 ade2-1 his3 leu2 trp1 ura3 sup35::KanMx. pJ533),* which expresses SUP35 from a URA3 plasmid as the sole copy of SUP35 in the cell. YER632 is a derivative of 7801D/pJ533 (Song *et al*., 2005). Additionally, YER635 *(α kar1-1 SUQ5 ade2-1 his3 leu2 trp1 ura3 Ppq::His3)* carries the maintainer SUP35 plasmid pJ533. YER635 is a derivative of 780-1D/pJ533 (Song *et al*., 2005), in which Ppq has been knocked out to enhance prion formation (Lindquist, 2008).

Constructing PrLD's of human proteins:

Three RNA-binding proteins, hnRNPA3, hnRPDL, and hnRNPD, were chosen based on their PAPA prion propensity scores. In each case, the 41 amino acids segment [hnRNPA3 (FEKWGTLTDCVVMRDPQTKRSRGFGFVTYSCVEEVDAAMCA), hnRPDL (GFNNYYDQGYGNYNSAYGGDQNYSGYGGYDYTGYNYGNYGY), and hnRNPD (ATAAVGGSAGEQEGAMVAATQGAAAAAGSGAGTGGGTASGG)] was identified with a prion score (0.5-0.1) and reasonable disorders score. This segment was then cloned into SUP35 in the place of the SUP35 nucleation domain (amino acid 3-40, QGNINQNVYQQYSQNGNQIIQQGNNRYQGYQA) using two separate PCR reactions. The SUP35 promoter was amplified from pJ533 using primer EDR302 (GGCAGAATATCTGTCAACCACAC) paired with EDR261 (GTTCTTCGACCTTTGTTGGCTC). In other reaction, C-terminus of SUP35 was amplified from pJ533 using EDR304 (GTTTCGTACTCACCCTTTCTGG) paired with EDR259

(CCAAAGCTCCCATTGCTTCTG). In order to insert the PrLD's of each protein, the sense primer of hnRNPA3, EDR1782 (GGAAACTACGGAGGTTGTGGTAACTATAATGATTTTGGTAATTACTCTGGTCAACAGCAATCTAACTATCCTGCAGGTGGGTACTACC) and hnRPDL, EDR1788 (GTGCATATGGAGGTGACCAGAACTATTCTGGTTACGGTGGCTATGATTATACTGGTTATAACTACGGTAATTACGGTTATCCTGCAGGTGGGTACTACC) and hnRNPD, EDR1784 (CTATGGATACAACTCCCAGGGATACGGTGGTTATGGTGGTTATGATTATACTGGTTATAACAATTACTATGGTTATCCTGCAGGTGGGTACTACC) were paired with EDR262 (GCATCAGCACTGGTAACATTGG) to reamplify the C-terminus. In separate reaction antisense primers of each protein EDR1781, (GTTACCACAACCTCCGTAGTTTCCACCACCAAAATTACCACCTTCATTATAACCATCATATCCGCCACCGCCACCATACGACATTGTTGCTAGTGGGCAG), EDR1787 (GTTCTGGTCACCTCCATATGCACTATTATAGTTTCCGTAACCTTGATCATAATAATTGTTAAAGCCCGACATTGTTGCTAGTGGGCAG), and EDR1783 (GTATCCCTGGGAGTTGTATCCATAGTTACCATAGCCTTGATTCCAATAATTAGAGTAACCTTGGTTCCAATTCGACATTGTTGCTAGTGGGCAG) were paired with EDR301 (CGTCACAGTGTTCGAGTCTG) to reamplify N-terminus. Then, for each construct, the two corresponding PCR products were combined and re-amplified with EDR261 and EDR301. The final PCR products were co-transformed with AatII/HindIII-cut vector pJ526 (Lue2) into yeast strain YER635 (wild type Sup35 with ppq1 gene deletion to boost prion activity). Transformants were plated on Sc-leu to select for successful recombinants and then selected for loss of the wild-type SUP35 plasmid by stamping onto 5-fluoroorotic acid (5-FOA) containing medium. Then plasmids were screened by DNA sequencing.

Creating Induction plasmids.

Plasmids were built for prion domain overexpression. Each fusion construct was amplified with primers EDR301 and EDR304. These products were re-amplified with EDR1084 (CGATGCTACTCGAGTTTACATATCGTTAACAACTTCGTCATCCAC) and specific primers for each protein to insert a stop codon and an *XhoI* digestive site. PCR products were digested with *BamHI* and *XhoI* and ligated into *BamHI/XhoI*-cut vector pKT24 (TRP1) which contains the GAL1 promoter. Ligation products were transformed into DH5α bacterial cells and tested by DNA sequencing.

Testing for [PSI$^+$] maintenance and curing.

Strains expressing the PrLD-SUP35 were transformed with either plasmid pKT24 (un-induced) or with a derivative pKT24 expressing the matching PFD from the *GAL1* promoter (induced). Strains were then grown in galactose/raffinose –trp + ade dropout mix medium (made the same as SC media, but with 2% Galactose and 1% raffinose instead of glucose) for 3 days with shaking at 30C$^{0.}$ Serial dilutions were then plated onto medium lacking adenine and grown for 5 days to select for [PSI$^+$].

To confirm that the Ade+ phenotype was a result of [PSI$^+$] formation rather than DNA mutation, Ade+ cells were grown on YPAD or YPAD plus 4mM guanidine HCl (GdHCl). Guanidine HCl is a protein that blocks prion propagation and leading the cell to return back to [spi$^-$]. Single colonies were then stamped side by side onto YPD to test for loss of [PSI$^+$].

Results

Assess prion-like activities of three human proteins.

Sup35 is portable and modular, meaning that it can maintain prion activity when fused to other proteins (Baxa *et al.,* 2002). If the *SUP35* PFD appends to other protein, then that protein will exhibit prion activities. Therefore, this modular set-up was used to evaluate prion activities of PrLD's in three human RNA-binding proteins: hnRNPA3, hnRPDL, and hnRNPD. These proteins were picked based on their PAPA scores, which all were above or just below the threshold (0.05), (see Figure 3.2). It was hypothesized that each of these proteins that had a high score on the PAPA must have prion activities. By proving this, it was expected to identify more proteins that support prion activity, as well as test the PAPA accuracy by applying it to human proteins. To test this hypothesis, three chimeric proteins were generated in which PrLD's of hnRNPA3,

(FEKWGTLTDCVVMRDPQTKRSRGFGFVTYSCVEEVDAAMCA),hnRPDL

(GFNNYYDQGYGNYNSAYGGDQNYSGYGGYDYTGYNYGNYGY),and hnRNPD

(ATAAVGGSAGEQEGAMVAATQGAAAAAGSGAGTGGGTASGG) were substituted for the SUP35 nucleation domain (3-40 amino acids) (see Figure 3.3).

Fusion constructs were created by PCR and cloned into a plasmid under the control of the SUP35 promoter. These plasmids were introduced into yeast cells that lack an endogenous copy of SUP35, but that carry a maintainer copy expressed from a URA3 plasmid. The strains were assessed for their tendency to form [PSI^+] after loss of maintainer plasmids.

[PSI^+] cells were analyzed by monitoring nonsense suppression of the ade2 allele (loss of function of *SUP35*). [psi^-] cells (nonsense mutation) are unable to grow without adenine, but in the presence of limiting adenine, the colonies grow red due to the accumulation of adenine

precursor. In contrast, [PSI^{+}] cells cause read-through of the nonsense mutation (stop codon), allowing for the ability to grow without adenine, and white colony formation in the existence of limiting adenine. Ade+ selection assay was done with and without transient overexpression of the corresponding prion-forming domains. Increasing protein or prion-forming domain concentration can boost the rate of *de novo* prion formation. Hence, if a strain forms prions, the rate of prion formation would increase with increasing protein overexpression (Wickner *et al*., 1994).

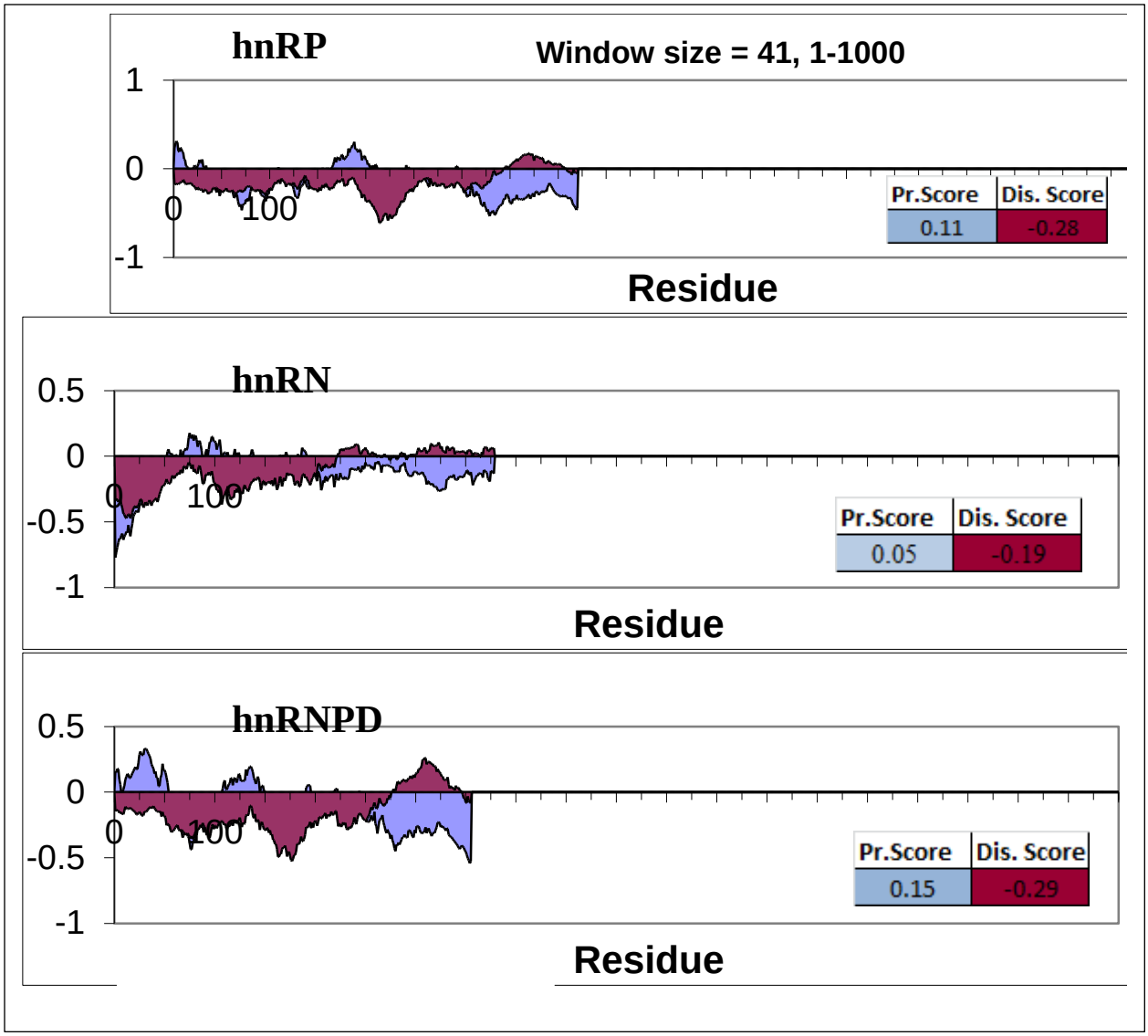


Figure3. 2 : Identification of prion-prone regions (PrLD's) within human RNA-binding proteins: hnRPDL, hnRNPA3, and hnRNPD.
The proteins were scanned using a window size of 41 amino acids, calculating the average order propensity for each window, FoldIndex (red) and prion propensity (blue). All 3 proteins show a prion-like activity region.

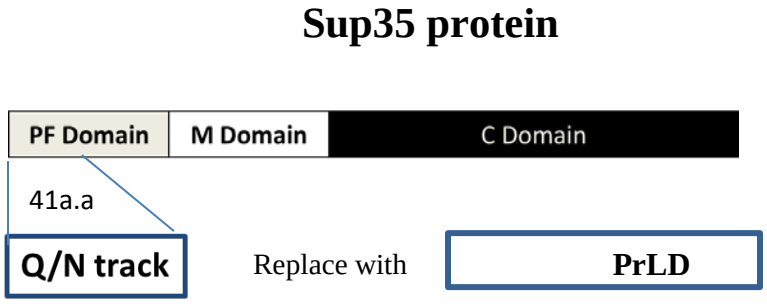


Figure3. 3: Schematic of Sup35 protein, indicating the region that was replaced with PrLDs. Replacement of the 41 amino acids segment of human RNA-binding proteins (hnRPDL, hnRNPA3, and hnRNPD) with the SUP35 nucleation domain (amino acid 3-40) using PCR.

Remarkably, all three chimeric proteins presented the ability to form and maintain the Ade+ phenotype. Transient overexpression of the corresponding prion forming domain increased the rate of Ade+ cells formation for each of the fusion constructs, signifying that the loss of function phenotype is the result of a prion (see Figure 3.4-A).

Confirming the maintenance of prion

The Ade+ phenotype could also result from a genomic mutation. To confirm that the maintenance of the Ade+ phenotype was a result of [PSI^+] maintenance, a test was conducted to see whether the Ade+ phenotype could be cured by treatment with low concentrations of guanidine HCl. Guanidine HCl cures [PSI^+] (Aigle and Lacroute 1975; Tuite, Mundy *et al*. 1981) by interfering with Hsp104p activity (Ferreira and Ness *et al*. 2001; Jung and Masison 2001; Jung, Jones *et al*. 2002; Grimminger-Marquardt & Lashuel *et al*., 2010), a chaperone protein that facilitates prion propagation. If the loss of Sup35p function is due to prion aggregation, then the Ade+ phenotype should be curable. The Ade+ phenotype was efficiently cured by treatment with 4 mM guanidine HCl in all cases, (see Figure 3.4-B), indicating that these fusion proteins were indeed maintaining [PSI^+] rather than DNA mutation.

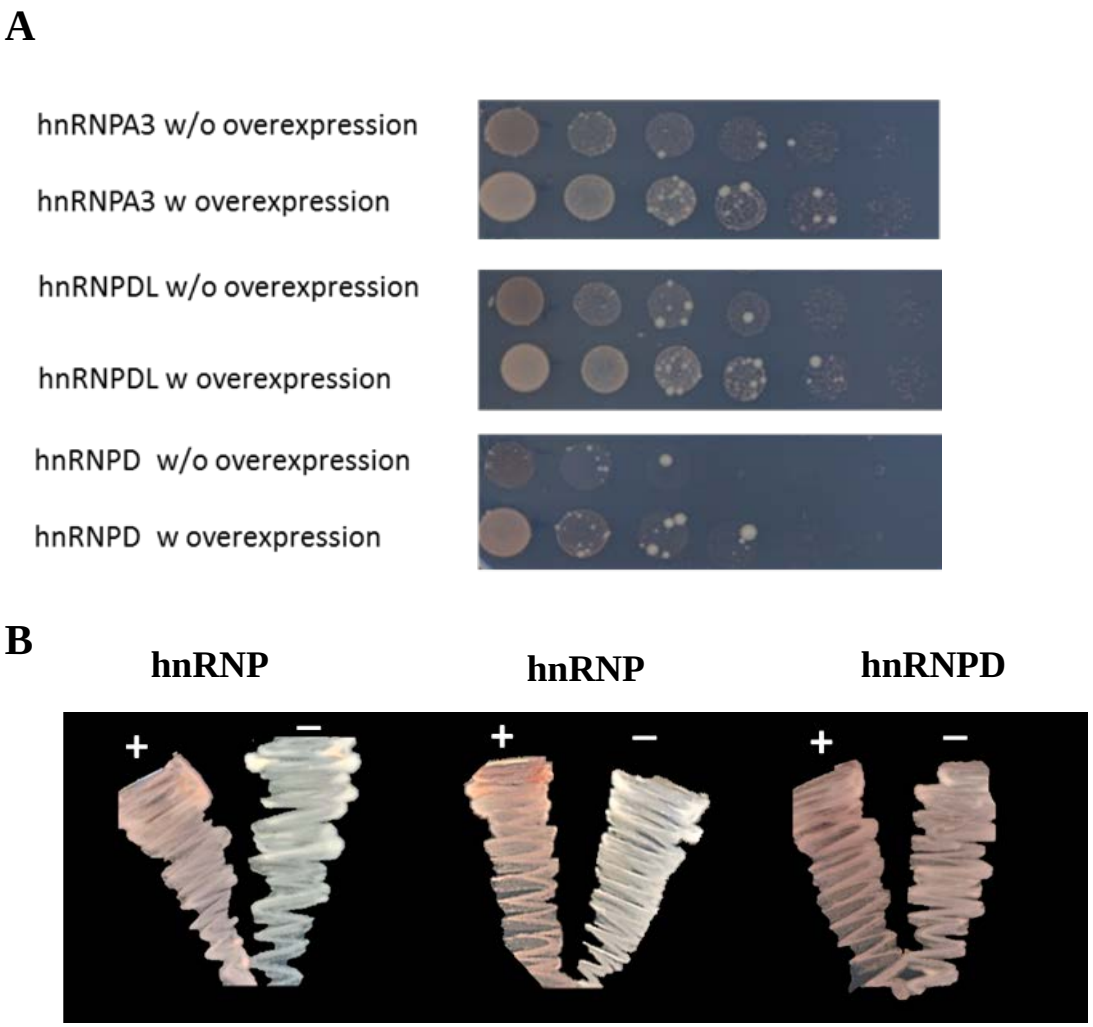


Figure3. 4: Induction and curability assays for PrLD-SUP35**. (A)** Strains expressing the PrLD-SUP35 fusions were tested for *de novo* prion formation. Strains expressing the fusion proteins were transformed with either plasmid pKT24 containing the *GAL1* promoter (uninduced) or with a derivative of pKT24 expressing the matching prion forming domain from the *GAL1* promoter (induced). Strains were grown in galactose/raffinose –trp + ade dropout medium and serial dilutions plated onto medium lacking adenine to select for [PSI^{+}]. (**B)** Curability assay to test for Ade+ maintenance. Ade+ colonies were spread on YPD media to allow color selection. Ade+ colonies were streaked on YPD and YPD plus 4 mM GdHCl to test for curability. All three proteins were successfully cured.

Discussion

Considering what facilities some human proteins to aggregate within a cell, and in what manner this aggregation can occasionally affect neighboring cells and tissues and lead to disease, are extremely important. Due to the deficiency of specific knowledge and an optimal approach to predict proteins that have an ability to aggregate and develop diseases, the identification of further proteins engaged in human aggregation-based disorders has been inadequate. Even though the essential features driving prion-like aggregation and propagation are still poorly understood, very recently it has been proposed that many aggregation-based diseases may develop by means of a prion-like mechanism and spread from cell to cell within adjacent regions of the tissue by of converting the protein to self-templating protein conformers (Brundin *et al*., 2010; Cushman *et al.*, 2010; Dunning *et al.*, 2011; Goedert *et al*., 2010; Polymenidou and Cleveland, 2011; Prusiner, 1984; Walker *et al*., 2006). This phenomena, referred to as a prion-like mechanism because it is not a *bona fide* prion, meaning the protein aggregates propagate by a self-templating mechanism, but natural transmission to another individual has not yet been reported (Clavaguera *et al*., 2009; Desplats *et al*., 2009; Eisele *et al.*, 2010; Meyer-Luehmann *et al*., 2006).

These experiments have worked to develop a better understanding of PAPA accuracy on predicting prion propensity of PrLD-harboring human proteins. PAPA was mainly developed to predict prion propensity of yeast prions. The majority of yeast proteins are characterized by having prion forming domains (PFD's) rich in uncharged polar amino acids, which are predominantly glutamine and asparagine (Q/N-rich) (Alberti *et al*., 2009; Toombs *et al*., 2010). With the new observations of Q/N-rich human proteins hiding prion-like activity, the unifying feature of Q/N-rich yeast prions could be a useful guide to investigate the prion-like machinery

accountable for several neurodegenerative disorders. Although PAPA was not optimally developed for proteins that compositionally differ from yeast proteins and its accuracy decreases as the deviation from the amino acids composition of known prions increases. Human proteins harboring a RNA recognition motif (RRM) have been scanned for prion-like domains using the hidden Markov model (Alberti *et al.*, 2009; Couthouis *et al*., 2011; Cushman *et al*., 2010), and were ranked by prion domain score according to the Alberti algorithm and PAPA (see Table 1). Together, taking the proteins list of this study and the lab compiling list, three RNA-binding proteins having PrLD's were chosen: hnRPDL, hnRNPD, and hnRNPA3. These proteins were consecutively ranked 4, 5, and 9 according to the Alberti algorithm, had a prion propensity score of 0.11, 0.15, and 0.05, and the FoldIndex was -0.28, -0.29, and -0.19, respectively, according to PAPA, (see Figure 3.2).

This study took advantage of the fact that the Sup35 yeast prion domain is modular and transferable, meaning it can maintain prion activity when attached to other proteins (Wickner *et al*. 2000 and Baxa *et al*., 2002). The 40 amino acids of PrLD's of these proteins were inserted in the place of the *SUP*35 nucleation domain (the amino acid 3-40), in order to assay whether they were able to form prions and confirm the PAPA score. These assays were performed under controlled conditions in Sup35 yeast protein. Remarkably, in vitro results of all 3 protein's PrLDs showed an agreement with their PAPA score. Even though PAPA provided authentic prion propensity prediction for human proteins, it was not specifically created to detect human protein candidates. Therefore, it is still an important need for more efficient ways to detect more candidates among human protein populations.

This experiments show that these fragments can support prion activity, but that does not mean that they would have this same activity in their nature context. Therefore, more experiments will be needed to explore these domains in their nature context.

Chapter 4: Conclusion

Prions are infectious proteins because they are capable of transferring their trait to neighboring or daughter cells. This method of transferring is distinctive as no nucleic acid component is involved for prion maintenance or transmission. A well-studied model prion from Saccharomyces cerevisiae is [PSI^+], the prion form of the translational terminator protein Sup35. The power and properties of this model has provided platform to understand prion biology as well as its molecular aspect. One of the most surprising discoveries however was that the prions formation ability is mainly independent of primary amino acid sequence; instead, that amino acid composition is the main principal determinant of prion formation (Ross, Baxa *et al*. 2004; Ross, Edskes *et al*. 2005).

The demonstrated importance of ORD region of Sup35 would seems to argue against the idea that prion activity is primary sequence independent. Specifically, deletion of one or more repeats reduces the efficiency of prion propagation. However, the Sup35 ORD can be scrambled without any effect on propagation, indicating that the ORD is important for its length and composition not primary sequence. Taking advantage of some of the unique features of prion proteins (Q/N reach, low hydrophobics and high charged residues) and the discovery that composition not primary sequence drives prion formation has provided researchers with platforms to create methods of identifying new prion candidates. Early bioinformatics studies screened the genome looking for new prions based on compositional similarity to known prion forming domains. Unfortunately, no study provides broad understanding of which amino acids stimulate/inhibit prion formation and propagation.

Recent study showed that replacing the Sup35 ORD with scrambling version of Sup35 nucleation domain blocks prion propagations (Toombs *et al*. 2011), proving that amino acid composition required for prion formation differs than amino acids required for prion propagation. Achieving an understanding of which amino acids are required for each process will specifically help to distinguish which Q/N rich domain proteins have ability to form prions and which do not.

Consequently, the principal aims of my research are to develop a better understanding of the essential features of prion formation and propagation independently and use this information to improve our ability to predict prion activity. The research addresses the questions of how sequence requirements differs from prion formation versus prion propagation and how amino acid composition, particularly hydrophobic amino acids within the ORD, affects prion formation.

Different Amino Acid Composition Requirements for Prion Formation and Propagation in the [PSI^{+}] Yeast Prion

We previously determined prion propensity of each amino acid *in vivo*, and used these values to develop a prion prediction algorithm, PAPA. Although PAPA so far is the most accurate prediction algorithm for Q/N rich domain, it is still far from perfect. One likely caveat is that there are two separate steps required for prion activity. Precisely, in order for a protein to act as a prion in yeast it must be able to not only form aggregates, but also propagate these aggregates over multiple generations by fragmenting the aggregates to create new independently segregating prion seeds to offset dilution by cell division (Chernoff *et al*. 1995; Derdowski *et al.* 2010). Each of these steps may have distinct amino acid sequence requirements, yet PAPA uses only a single prion propensity score for each amino acid and does not separately score these

activities. To make better predictions of prion propensity requires a better understanding of how amino acid sequence separately affects prion formation and propagation.

In Chapter 2, we commenced mutagenesis experiments that clearly separate those two functions and outlined a set of amino acid composition requirements for prion formation and propagation that are overlapping but distinct. We found that while prions generally favor the presence of hydrophobic and aromatic residues in prion domains, while disfavoring charged residues (Parham *et al.* 2001), non-aromatic hydrophobic residues (ILMV) are disfavored specifically in terms of their effects on the propagation function. Additionally, we also observed that polar residues in fact had a slightly positive effect on prion propagation propensity, and proline and glycine were neutral. Therefore, this information would be highly valuable if incorporated into a second generation of PAPA and provide a first step to understand the mechanistic basis of prion formation.

Prion-like domains and neurodegenerative diseases

Prions are self-templating aggregated forms of natural and functional proteins. These aggregates are "infectious" because they have the ability to transfer vertically to daughter cells or horizontally to neighboring cells and prompt aggregation of normal soluble prion-forming proteins. Prions are associated with a variety of incurable chronic mammalian diseases including Creutzfelt-Jacob Disease (CJD), Fatal Familial Insomnia (FFI), Kuru, and animal Transmissible Spongiform Encephalopathies (TSE's), which include Bovine Spongiform Encephalopathy (BSE) in cattle, Scrapie in sheep and goats, and Chronic Wasting Disease in cervids. Most yeast prions are characterized by the presence of large numbers of glutamine and asparagine (Q/N) residues, and some other common characteristics have been noted, including the presence of few

hydrophobic and charged residues (Alberti *et al*., 2009; Toombs *et al*., 2010). These features are used to identify more prion forming protein candidates.

Recently, around 250 of human proteins have been identified with regions similar to the yeast prion-forming domains (PFD's) regarding to amino acid composition. Consequently, these regions have been named prion-like domains (PrLD's). Many of them are RNA-binding proteins containing RNA recognition motifs (RRM) (King *et al*., 2010). Some of these, including FUS, TDP-43, TAF15 and EWSR1, have been associated with neurodegenerative diseases, including: Amyotrophic lateral sclerosis (ALS also known as Lou Gehrig's disease), Frontemporal Lobar Degeneration (FTLD), and inclusion body myopathy with Frontemporal Dementia (IBMPFD) (King *et al*., 2010; Da Cruz, and Cleveland *et al.,* 2011; Weihl *et al.,* 2008; Couthouis *et al*., 2012; Neumann *et al*., 2011). The rapid expansion of the number of human proteins that are involved in neurodegenerative diseases suggests that there are likely more disease-associated proteins yet to be identified.

In Chapter 3 we have worked to develop a better understanding of PAPA accuracy on predicting prion propensity of PrLD-harboring human proteins. PAPA was able to accurately predict the disease-causing effect of the single amino acid mutation observed in human proteins hnRNPA2 and hnRNPA1. These two RNA-binding proteins aggregate into amyloid fibrils in patients with IBMPFD coupled with ALS (Kim *et al.,* 2013). These observations were encouragement to test other human proteins that have reasonably high prion propensity scores. We have tested three RNA-binding proteins harboring PrLD's: hnRPDL, hnRNPD, and hnRNPA3. Taking advantage of the fact that Sup35 is modular, we replaced Sup35 nucleation domain (amino acids 3-40) with 40 amino acids of PrLD of these proteins in order to examine

whether they were able to act as prions. In this chapter we concluded that all three proteins have prion activities and suggesting PAPA can identify aggregation segments in human PrLD's.

An important caveat is that while these experiments show that fragments from these three proteins can support prion activity when fused to Sup35 that does not necessarily mean that they will have similar aggregation activity in their nature context. Therefore, future experiments will be needed to examine the aggregation activity of the full length proteins. Additionally, the relationship between aggregation and diseases is not fully understood, so even if full length proteins aggregate, that does not mean that they will cause disease.

References

1. Aguzzi A, Rajendran L (2009) The transcellular spread of cytosolic amyloids, prions, and prionoids. Neuron 64 (6):783-790. doi:10.1016/j.neuron.2009.12.016
2. 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 (1):146-158
3. Alexandrov IM, Vishnevskaya AB, Ter-Avanesyan MD, Kushnirov VV (2008) Appearance and propagation of polyglutamine-based amyloids in yeast: tyrosine residues enable polymer fragmentation. J Biol Chem 283 (22):15185-15192.
4. Baxa, U., Speransky, V., Steven, A. C., & Wickner, R. B. (2002). Mechanism of inactivation on prion conversion of the Saccharomyces cerevisiae Ure2 protein. Proceedings of the National Academy of Sciences, 99(8), 5253-5260.
5. Brundin, P., Melki, R., Kopito, R. (2010). Prion-like transmission of protein aggregates in neurodegenerative diseases. Nat. Rev. Mol. Cell Biol. 11, 301–307.
6. 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 (5212):880-884
7. Clavaguera, F., Bolmont, T., Crowther, R.A., Abramowski, D., Frank, S., Probst, A., Fraser, G., Stalder, A.K., 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.
8. Couthouis J, Hart MP, Shorter J, Dejesus-Hernandez M, Erion R, Oristano R, Liu AX, Ramos D, Jethava N, Hosangadi D, Epstein J, Chiang A, Diaz Z, Nakaya T, Ibrahim F, Kim HJ, Solski JA, Williams KL, Mojsilovic-Petrovic J, Ingre C, Boylan K, Graff-Radford NR,

Dickson DW, Clay-Falcone D, Elman L, McCluskey L, Greene R, Kalb RG, Lee VM, Trojanowski JQ, Ludolph A, Robberecht W, Andersen PM, Nicholson GA, Blair IP, King OD, Bonini NM, Van Deerlin V, Rademakers R, Mourelatos Z, Gitler AD (2011) A yeast functional screen predicts new candidate ALS disease genes. Proceedings of the National Academy of Sciences of the United States of America 108:20881-20890. doi:10.1073/pnas.1109434108

9. Couthouis, J., Hart, M. P., Erion, R., King, O. D., Diaz, Z., Nakaya, T., ... & Gitler, A. D. (2012). Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. Human molecular genetics, 21(13), 2899-2911.
10. Cox BS (1965) PSI, a cytoplasmic suppressor of super-suppressor in yeast. Heredity 26:211-232
11. Cushman, M., Johnson, B. S., King, O. D., Gitler, A. D., & Shorter, J. (2010). Prion-like disorders: blurring the divide between transmissibility and infectivity. Journal of Cell Science, 123(8), 1191-1201.
12. Da Cruz, S., & Cleveland, D. W. (2011). Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. Current opinion in neurobiology, 21(6), 904-919.
13. Derkatch IL, Bradley ME, Hong JY, Liebman SW (2001) Prions affect the appearance of other prions: the story of [PIN(+)]. Cell 106 (2):171-182.
14. Du Z (2011). The complexity and implications of yeast prion domains. Prion 5 (4):311-316. doi:10.4161/pri.18304
15. Du Z, Park KW, Yu H, Fan Q, Li L (2008) Newly identified prion linked to the chromatin-remodeling factor Swi1 in Saccharomyces cerevisiae. Nat Genet 40 (4):460-465

16. Dunning, C.J., Reyes, J.F., Steiner, J.A., Brundin, P. (2011). Can Parkinson's disease pathology be propagated from one neuron to another? Prog Neurobiol. doi:10.1016/j.pneurobio.2011.11.003.
17. Eisele, Y.S., Obermuller, U., Heilbronner, G., Baumann, F., Kaeser, S.A., Wolburg, H., Walker, L.C., Staufenbiel, M., Heikenwalder, M., Jucker, M. (2010). Peripherally applied Abeta-containing inoculates induce cerebral beta-amyloidosis. Science 330, 980–982.
18. Esteras-Chopo A, Serrano L, de la Paz ML (2005) The amyloid stretch hypothesis: Recruiting proteins toward the dark side. Proc Natl Acad Sci USA 102 (46):16672-16677
19. 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 (6):1357-1369.
20. Gitler AD, Shorter J (2011) RNA-binding proteins with prion-like domains in ALS and FTLD-U. Prion 5 (3):179-187. doi:10.4161/pri.5.3.17230
21. Goedert, M., Clavaguera, F., Tolnay, M. (2010). The propagation of prion-like protein inclusions in neurodegenerative diseases. Trends Neurosci. 33, 317–325.
22. Halfmann R, Alberti S, Krishnan R, Lyle N, O'Donnell CW, King OD, Berger B, Pappu RV, Lindquist S (2011) Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins. Mol Cell 43 (1):72-84.
23. Halfmann R, Wright J, Alberti S, Lindquist S, Rexach M (2012) Prion formation by a yeast GLFG nucleoporin. Prion 6 (4)
24. Harrison PM, Gerstein M (2003) A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes. Genome Biol 4 (6):R40.

25. Kim, H. J., Kim, N. C., Wang, Y. D., Scarborough, E. A., Moore, J., Diaz, Z., ... & Taylor, J. P. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature. 11922

26. King, O. D., Gitler, A. D., & Shorter, J. (2012). The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain research, 1462, 61-80.

27. Laurent, M. (1996). Prion diseases and the'protein only'hypothesis: a theoretical dynamic study. Biochemical Journal, 318(Pt 1), 35.

28. Li L, Lindquist S (2000) Creating a protein-based element of inheritance. Science 287 (5453):661-664.

29. Maclea KS, Ross ED (2011) Strategies for identifying new prions in yeast. Prion 5 (4):263-268

30. Meyer-Luehmann, M., Coomaraswamy, J., Bolmont, T., Kaeser, S., Schaefer, C., Kilger, E., Neuenschwander, A., Abramowski, D., Frey, P., Jaton, A.L., Vigouret, J.M., Paganetti, P., Walsh, D.M., Mathews, P.M., Ghiso, J., Staufenbiel, M., Walker, L.C., Jucker, M. (2006). Exogenous induction of cerebral beta-amyloidogenesis is governed by agent and host. Science 313, 1781–1784

31. 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 USA 97 (22):11910-11915.

32. Neumann, M., Bentmann, E., Dormann, D., Jawaid, A., DeJesus-Hernandez, M., Ansorge, O., ... & Mackenzie, I. R. (2011). FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. Brain, 134(9), 2595-2609.

33. Patel BK, Gavin-Smyth J, Liebman SW (2009) The yeast global transcriptional co-repressor protein Cyc8 can propagate as a prion. Nature cell biology 11 (3):344-349

34. Polymenidou, M., Lagier-Tourenne, C., Hutt, K.R., Huelga, S.C., Moran, J., Liang, T.Y., Ling, S.C., Sun, E., Wancewicz, E., Mazur, C., Kordasiewicz, H., Sedaghat, Y., Donohue, J.P., Shiue, L., Bennett, C.F., Yeo, G.W., Cleveland, D.W. (2011). Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat. Neurosci. 14, 459–468.

35. 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 (16):3435-3438.

36. Prusiner SB (1982) Novel proteinaceous infectious particles cause scrapie. Science 216 (4542):136-144.

37. Prusiner SB (1998) Prions. Proc Natl Acad Sci USA 95:13363-13383

38. Prusiner, S. B. (1982). Novel proteinaceous infectious particles cause scrapie. Science (New York, NY), 216(4542), 136.

39. Prusiner, S. B. (1998). Prions. Proceedings of the National Academy of Sciences, 95(23), 13363-13383.

40. Prusiner, S.B.(1984). Some speculations about prions, amyloid, and Alzheimer's disease. N. Engl. J. Med. 310, 661–663.

41. Rambaran RN, Serpell LC (2008) Amyloid fibrils: abnormal protein assembly. Prion 2 (3):112-117

42. Rogoza T, Goginashvili A, Rodionova S, Ivanov M, Viktorovskaya O, Rubel A, Volkov K, Mironova L (2010) Non-Mendelian determinant [ISP+] in yeast is a nuclear-residing prion

form of the global transcriptional regulator Sfp1. Proc Natl Acad Sci USA 107 (23):10573-10577.

43. Ross CD, McCarty BM, Hamilton M, Ben-Hur A, Ross ED (2009) A promiscuous prion: Efficient induction of [URE3] prion formation by heterologous prion domains. Genetics 183 (3):929-940.
44. Ross ED, Baxa U, Wickner RB (2004) Scrambled Prion Domains Form Prions and Amyloid. Mol Cell Biol 24 (16):7206-7213
45. Ross ED, Edskes HK, Terry MJ, Wickner RB (2005) Primary sequence independence for prion formation. Proc Natl Acad Sci USA 102 (36):12825-12830
46. Sheikh S, Safia, Haque E, SS M (2013) Neurodegenerative diseases: multifactorial conformational diseases and their therapeutic interventions. Journal of Neurodegenerative Diseases 2013
47. Sondheimer N, Lindquist S (2000) Rnq1: an epigenetic modifier of protein function in yeast. Mol Cell 5 (1):163-172.
48. Suzuki G, Shimazu N, Tanaka M (2012) A Yeast Prion, Mod5, Promotes Acquired Drug Resistance and Cell Survival Under Environmental Stress. Science 336 (6079):355-359. doi:10.1126/science.1219491
49. Toombs JA, McCarty BR, Ross ED (2010) Compositional determinants of prion formation in yeast. Mol Cell Biol 30 (1):319-332
50. Toombs JA, Petri M, Paul KR, Kan GY, Ben-Hur A, Ross ED (2012) De novo design of synthetic prion domains. Proceedings of the National Academy of Sciences of the United States of America 109 (17):6519-6524

51. Walker, L.C., Levine III, H., Mattson, M.P., Jucker, M. (2006). Inducible proteopathies. Trends Neurosci. 29, 438–443.

52. Weihl, C. C., Temiz, P., Miller, S. E., Watts, G., Smith, C., Forman, M., ... & Pestronk, A. (2008). TDP-43 accumulation in inclusion body myopathy muscle suggests a common pathogenic mechanism with frontotemporal dementia. Journal of Neurology, Neurosurgery & Psychiatry, 79(10), 1186-1189.

53. Wickner RB (1994) [URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae. Science 264 (5158):566-569.

54. Wickner, R.B., Taylor, K.L., Edskes, H.K., Maddelein, M.L. (2000). Prions: portable prion domains. Curr. Biol. 10, R335–R337.