DISSERTATION

INVESTIGATION OF THE SEQUENCE FEATURES CONTROLLING AGGREGATION OR DEGRADATION OF PRION-LIKE PROTEINS

Submitted by

Sean Micheal Cascarina

Department of Biochemistry and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

Advisor: Eric Ross

P. Shing Ho
Santiago Di Pietro
Mark Zabel

ABSTRACT

INVESTIGATION OF THE SEQUENCE FEATURES CONTROLLING AGGREGATION OR DEGRADATION OF PRION-LIKE PROTEINS

Protein aggregates result from the conversion of soluble proteins to an insoluble form. In some cases, protein aggregates are capable of catalyzing the conversion of their soluble protein counterparts to the insoluble form, resulting in a mode of molecular self-replication. Many of these infectious proteins, or “prions”, have been identified and characterized in yeast. This has led to the development of prediction algorithms designed to identify protein domains capable of forming prions. Recently, a number human proteins with aggregation-prone prion-like domains (PrLDs) have been identified, and mutations within PrLDs have been linked to muscular and neurodegenerative disorders. However, the number and diversity of PrLD mutations linked to disease are currently limited. Therefore, the extent to which a broad assortment of PrLD mutations affect intrinsic aggregation propensity, and how well this correlates with aggregation in a cellular context, has not been systematically examined.

In Chapter 2, I present evidence suggesting that our prion aggregation prediction algorithm (PAPA) is capable of predicting the effects of a diverse range of mutations on the aggregation propensity of PrLDs in vitro and in yeast. PAPA was also able to predict the effects of many but not all PrLD mutations when the protein was expressed in Drosophila, but with slightly. Therefore, while great strides have been made in predicting intrinsic aggregation propensity, a more complete understanding of the cellular factors that influence aggregation in vivo may lead to further improvement of prion prediction methods.

Many intracellular protein quality control factors specialize in recognizing and degrading aggregation-prone proteins. Therefore, prions must evade or outcompete these quality control systems in order to form and propagate in a cellular context. However, the sequence features that promote degradation versus aggregation of prion domains and PrLDs have not been systematically defined. In Chapter 3, I present evidence that aggregation propensity and degradation propensity can be uncoupled in multiple ways. First, we find that only a subset of classically aggregation-promoting amino acids elicit a strong degradation response in PrLDs. Second, the amino acids that promoted degradation of the PrLDs did not induce degradation of a glutamine/asparagine (Q/N)-rich prion domain, and instead led to a dose-dependent increase in the frequency of spontaneous prion formation, suggesting that protein features surrounding aggregation-prone amino acids can modulate their ultimate effects. Furthermore, degradation suppression correlated with Q/N content of the surrounding prion domain, potentially indicating an underappreciated role for these residues in yeast prion domains.

The protein features that foster susceptibility or resistance to degradation are further explored in Chapter 4. We find that Q/N-rich domains resist degradation in a primary sequence-independent manner, and can even exert a dominant degradation-inhibiting effect when coupled to a degradation-prone PrLD. Furthermore, susceptibility to degradation was a relatively de-centralized feature of the PrLD, requiring a large portion of the domain surrounding degradation-promoting amino acids to permit efficient protein turnover.

Collectively, these results provide key insights into the relationship between intrinsically aggregation-prone protein features and the ability to aggregate in the context of intracellular protein quality control factors.

ACKNOWLEDGEMENTS

If the credit for the work presented here is attributed to its true sources, it would be distributed among multiple generations of incredibly supportive family members, and proper description could not be contained in a dissertation of this size.

I would like to thank my grandparents, Kenneth and Patricia Coker, and James and Maxine Cascarina, for investing so heavily in my growth and well-being throughout my life, and for fostering a sense of adventure and exploration, which has been of immense value but, most of all, deeply enjoyable. I would like to thank my parents, Richard and Alice Cascarina, for their unwavering and unrelenting support, sacrifice, and love – your lifetime accomplishments remain the greatest and most inspiring that I've ever witnessed. Your encouragement to learn, grow, and enjoy curiosity ensured my success in graduate school, and has shaped me into the person that I am today. I thank my brother, Justin Cascarina, and sister, Sara Harvey, for their support, protection, and affirmation throughout my life – I always knew I could safely venture out with you. I thank the friends that I've had throughout my life – many of whom have been like family, and have been solid rocks on ever-shifting ground. To all family and friends – your generosity and kindness is profoundly life-giving and humbling.

I would also like to thank the many people along the way who have gone beyond the call of duty and invested in my athletic, academic, emotional, relational, and spiritual growth. This includes the many coaches, teachers, mentors, and pastors who do far more than is required of them, and from whom I have learned many valuable life lessons.

Finally, I would like to thank my advisor for his support, trust, encouragement, freedom, guidance, patience, and time. You have encouraged a joy in science in its purest form – a curiosity-driven desire to ask and answer questions regarding the physical nature of life.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION[1]

Initially, the primary aim of my research was to investigate the effects of starting amino acid composition within prion-like domains (PrLDs) on the prion propensities of each amino acid, with the intent of improving current prion prediction algorithms. Using an established genetic screen, I identified systematic differences in prion-promoting and prion-inhibiting residues within glycine (G)-rich PrLDs and glutamine/asparagine (Q/N)-rich prion domains. Surprisingly, the same genetic screen also allowed me to observe and characterize composition-dependent degradation of G-rich PrLDs, which has not been observed in Q/N-rich prion domains. Therefore, in this chapter I will introduce a current understanding of prions and prion-like proteins, strategies for predicting proteins with prion activity, and cellular strategies for preventing or responding to protein aggregation.

YEAST PRIONS AND PRION PREDICTION METHODS

Prions result from the conversion of soluble proteins to an insoluble aggregated form. Typically, these aggregates are assembled into organized amyloid fibers with cross β-sheet structure and are capable of acting as self-propagating infectious agents sans nucleic acid [1]. Though this structural characterization of prion aggregates is generally accepted, the specific features that drive prion nucleation, aggregation, and propagation have proven more difficult to determine. Further complicating matters, prions can be subdivided into unique classes with fundamentally different features driving prion formation.

[1] This chapter is adapted from Cascarina, SM & Ross, ED Cell. Mol. Life Sci. (2014) 71: 2047. The contained information has been updated with recent discoveries and relevant additions.

Several naturally occurring prion-forming proteins have been identified in *Saccharomyces cerevisiae*, including Mot3, Rnq1, Swi1, Cyc8, Sfp1, Mod5, Ure2, Sup35, and Nup100 [2-10]. All of these yeast prion proteins, with the exception of Mod5, contain Q/N-rich prion forming domains (PFDs). However, many other amyloid- and prion-forming proteins are not Q/N-rich, so this feature is not required for either amyloid formation or prion activity.

The presence of such a large number of proteins that can act as prions in yeast is somewhat enigmatic. It is clear that for some proteins, amyloid or prion formation can serve beneficial functions, acting as regulatory or structural elements [11]. However, the role of prions in normal yeast physiology is less clear. Some argue that prions may be advantageous to yeast under particular conditions, allowing them to act as means of survival and adaption in fluctuating environments [12-14]. Others maintain that, since yeast prions are relatively rare in wild strains despite their ability to form spontaneously and spread, these prions likely do not confer a selective advantage [15-18]. Rather, yeast PFDs may have evolved for reasons unrelated to prion formation.

YEAST PRION DISCOVERY AND CHARACTERIZATION

The first two yeast prions discovered, [*PSI*$^+$] and [*URE3*], were initially identified in genetic screens as non-chromosomal genetic elements with non-Mendelian inheritance [19, 20]. However, the basis for this non-Mendelian inheritance was not initially known. Decades later, [*PSI*$^+$] and [*URE3*] were proposed to be the prion forms of Sup35 and Ure2, respectively, based on their unusual genetic properties [9]. Subsequent careful analyses of these and other yeast prions have revealed a series of common sequence characteristics that have allowed for more targeted searches for new prion proteins.

Ure2 and Sup35 have similar domain layouts, with an N-terminal PFD that is responsible for prion activity but dispensable for the major cellular function of the prion protein (Table 1.1), and a C-terminal functional domain [21, 23-26]. Sup35 contains an additional highly charged middle domain, termed “M”, that is not required for either prion formation or the normal cellular function of Sup35, but which helps to stabilize [*PSI*$^{+}$] [27]. Sup35 and Ure2 are modular in nature, meaning the PFDs can be transferred to unrelated proteins and still support prion formation [28, 29].

Both the Ure2 and Sup35 PFDs are Q/N-rich and intrinsically disordered [30, 31]. Scrambling the primary sequence of the Ure2 and Sup35 PFDs does not does not eliminate the ability to form prions, indicating that amino acid composition, not primary sequence, is predominantly responsible for prion activity [32, 33]. The Sup35 and Ure2 PFDs share a number of compositional features, including an under-representation of charged and highly hydrophobic residues relative to the yeast proteome, and an over-representation of polar amino acids and glycine (Table 1.2).

Many of these same general features are also found in the other Q/N-rich PFDs. All of the known Q/N-rich PFDs are predicted to be intrinsically disordered. Each has relatively few charged and highly hydrophobic residues (for a detailed review of yeast PFD composition, see Du, 2011 [34]). Consequently, a number of the subsequent prions to be discovered were identified based on compositional similarity to known prions [35]. Other compositional biases, including biases towards serine, tyrosine and glycine, are only seen in a subset of PFDs [34].

Table 1.1: Amyloid-based prions from *Saccharomyces cerevisiae*.

Prion Protein	Prion	Prion Domain
Cyc8	[*OCT*$^+$]	465-966 [5]
Mod5	[*MOD*$^+$]	194-205 [8]
Mot3	[MOT3$^+$]	1-295 [2]
Nup100	[NUP100$^+$]	201-400 [10]
Rnq1	[*PIN*$^+$] or [*RNQ*$^+$]	153-405 [7]
Sfp1	[*ISP*$^+$]	253-331 [2]
Sup35	[*PSI*$^+$]	1-114 [21]
Swi1	[*SWI*$^+$]	1-385 [2]
Ure2	[URE3]	1-89 [22]

Table 1.2: Percent amino acid composition of yeast PFDs and human disease-associated PrLDs

	Gln/Asn	Ser	Gly	Tyr	Charged[a]	Hydrophobic[b]
Yeast PFDs[c]						
Ure2	48.3	11.2	5.6	0	11.2	15.7
Rnq1	43.1	15.4	16.7	5.9	2.4	8.3
Sup35	45.6	3.5	16.7	17.5	4.4	3.5
PFD Average[d]	45.7	10.0	13.0	7.8	6.0	9.2
Yeast Genome	10.0	9.0	5.0	3.4	24.0	28.3
Human PrLDs[e]						
TDP-43	21.8	15.9	26.8	0.7	3.5	15.9
FUS	21.5	22.8	28.3	12.2	3.8	0.8
TAF15	27.7	22.4	15.1	15.1	10.6	1.3
EWSR1	18.9	15.4	9.6	13.6	3.6	4.0
hnRNPA2B1	12.7	9.6	45.2	10.8	9.5	6.4
hnRNPA1	12.3	16.0	42.2	8.0	9.7	8.1
TIA1	31.6	4.2	15.8	9.5	2.1	12.6
PrLD Average	20.9	15.2	26.1	10.0	6.1	7.0
Human Genome	8.3	8.1	6.6	2.8	22.9	26.5
PrP[f]	12.8	5.7	9.9	7.8	20.6	19.9

[a] Charged residues include D, E, K, R

[b] Hydrophibic residues include F, I, L, M, V

[c] PFDs are as defined in Table 1.1

[d] Average of the Ure2, Sup35 and Rnq1 PFDs

[e] Disease-associated PrLDs found in RRM-containing proteins, as defined by the Alberti algorithm [36]

[f] Amino acids 90-230, which constitute the protease-resistant core of prion aggregates

Minimum PFD Length Requirements Vary

Although Sup35 and Ure2 contain clearly defined PFDs, determining the exact sequence features within these PFDs that are required for prion activity has proven more challenging. One challenge is that prion activity involves a series of discrete steps that may have distinct sequence requirements (Figure 1.1). Prion proteins must be able to form prion aggregates (Figure 1.1, Steps 1 and 2). These prion aggregates must then be able to recruit additional soluble protein and convert it to the prion form (Figure 1.1, Step 3). Finally, prion aggregates must be fragmented to generate new independently-segregating aggregates (seeds) to offset dilution by cell division (Figure 1.1, Step 4).

A wide variety of *in vitro* (Figure 1.2a, b) and *in vivo* (Figure 1.2, c-f) assays have been developed to define the sequence elements required for prion activity, but many of these assays only test a subset of the steps in prion formation and propagation, and subtle differences in experimental set-up can lead to very different outcomes. Consequently, attempts to define these sequence requirements have yielded seemingly contradictory results, with some experiments suggesting that very short segments are responsible for driving prion formation, and other experiments indicating that larger regions are required for prion activity.

A widely-used method to identify key nucleating segments within PFDs is to test the ability of mutated proteins to incorporate into wild-type prion aggregates either *in vivo* (Figure 1.2d) or *in vitro* (Figure 1.2b). Various single point mutations are sufficient to substantially reduce incorporation into wild-type Sup35 aggregates, both *in vivo* and *in vitro* [23, 37]. Many of these mutations cluster in a small 19-amino-acid segment of the Sup35 PFD (amino acids 8-26), suggesting a critical role for this segment. This segment also appears critical for mediating the [*PSI*$^{+}$] prion species barrier between *S. cerevisiae* and *C. albicans*. Insertion of amino acids 8-26

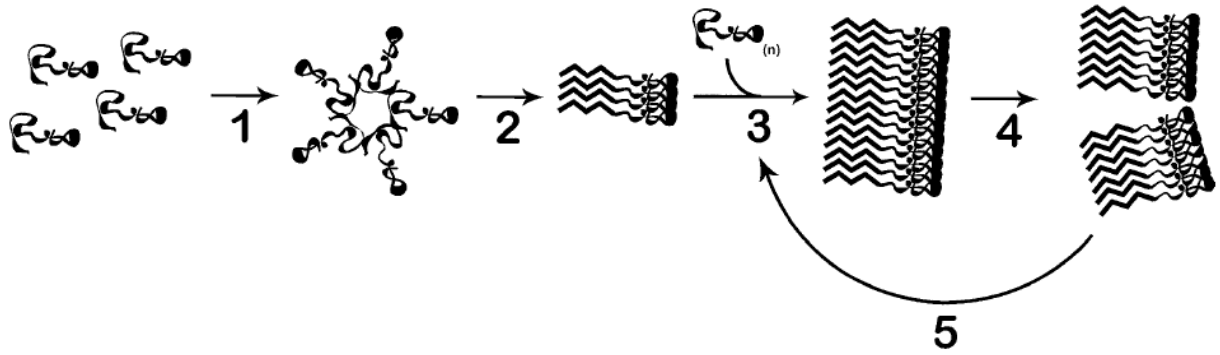


Figure 1.1: Basic steps in prion formation and propagation. Soluble proteins interact to form non-amyloid oligomers (*step 1*). These aggregates undergo a structural conversion to form amyloidogenic oligomers (*step 2*). The amyloidogenic aggregates recruit additional soluble proteins to form amyloid fibrils and to grow these fibrils (*step 3*). Fragmentation of these fibrils (*step 4*) creates new fiber ends for growth, while also creating new independently-segregating aggregates to offset dilution by cell division.

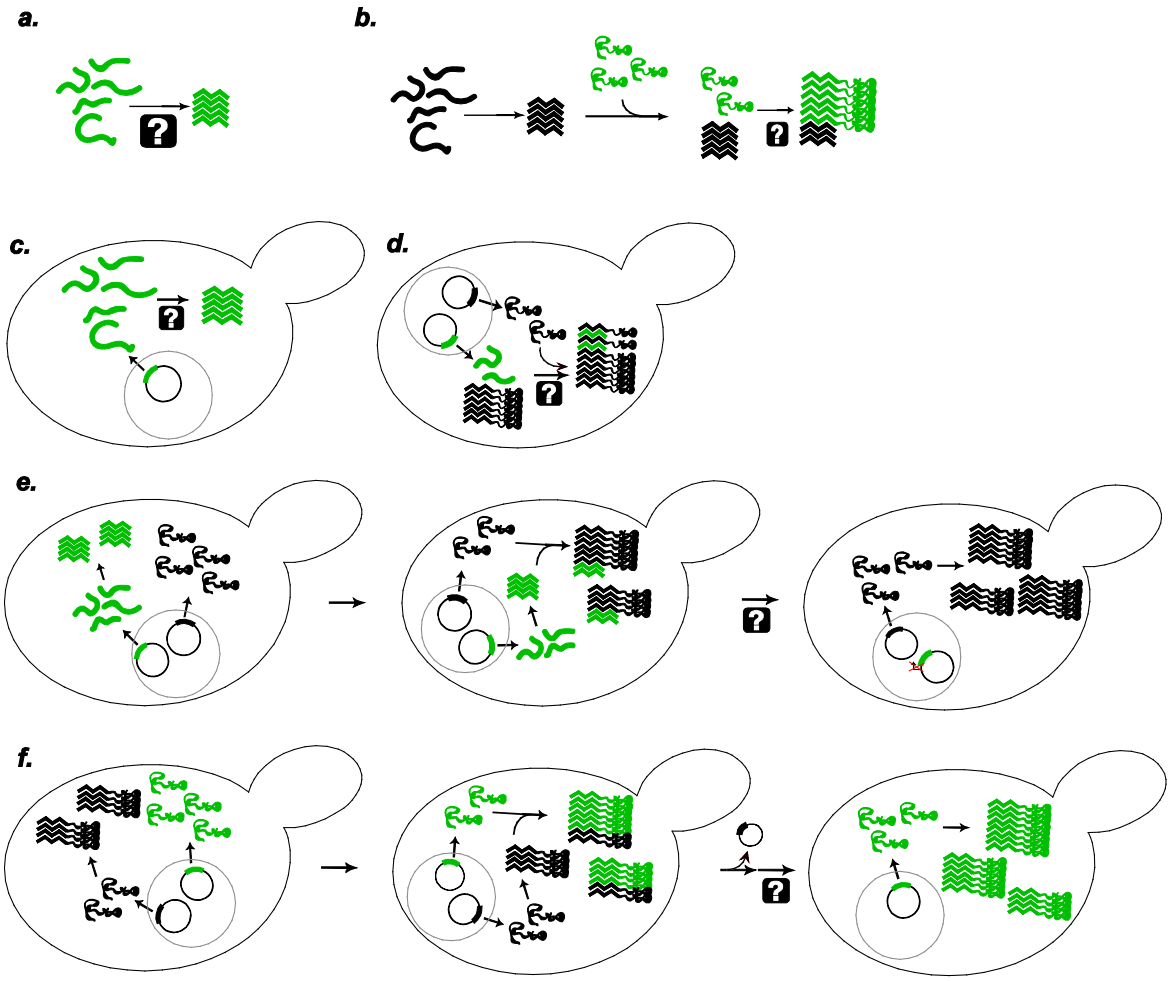


Figure 1.2: Assays to monitor prion-like activity and to define the regions of PFDs responsible for various aspects of prion activity. (A) *In vitro* aggregation. Protein fragments are incubated, generally with shaking, and aggregation is monitored using various techniques including Congo red binding, thioflavin T fluorescence, or pelleting assays. **(B)** Seeded *in vitro* aggregation. Preformed aggregates (green) are mixed with soluble protein (black) to test the ability of fragments to seed aggregation, or to test the ability of mutants to add onto preformed aggregates. **(C)** *De novo* aggregation *in vivo*. PFD fragments are transiently over-expressed. Aggregation is monitored either by fusing the fragments to GFP to observe foci formation, or through biochemical methods such as SDD-AGE (semi-denaturing detergent agarose gel electrophoresis). **(D)** Decoration of aggregates. PFD fragments are expressed in prion-positive cells to determine whether the fragments are capable of adding to preexisting aggregates. **(E)** Induction assays. PFD fragments (green) are transiently overexpressed in the presence of the full-length prion protein (black) to determine whether these fragments are sufficient to seed aggregation of the full-length protein. Aggregation is generally assayed by monitoring loss of function of the full-length protein. **(F)** Prion propagation assays. A prion-positive cell in which the chromosomal copy of the prion gene is deleted, but that carries a maintainer plasmid expressing the wild-type prion protein (black), is transformed with a plasmid expressing a prion protein mutant (green). The prion phenotype is assayed after selection for loss of the maintainer plasmid to determine if the mutant is capable of maintaining the prion.

from *S. cerevisiae* Sup35 into *C. albicans* Sup35 was sufficient to allow for efficient cross-seeding between *S. cerevisiae* and *C. albicans* Sup35 [38]. Other studies similarly indicate that short segments can play an important role in mediating the species barrier [39].

However, these short segments are not sufficient for prion activity. The Sup35 PFD contains two subdomains: an extreme N-terminal nucleation domain (amino acids 1-39) and an oligopeptide repeat domain (ORD; amino acids 40-114), which consists of five and a half copies of an imperfect nine-amino-acid sequence. The nucleation domain and the first repeat (amino acids 1-49) are required for incorporation into pre-existing aggregates. A slightly longer fragment (amino acids 1-64, which includes the first two repeats) is required for *de novo* aggregation (Figure 1.2c) or induction of prion formation by full-length Sup35 (Figure 1.2e; [40]). Furthermore, the ORD is necessary for efficient prion propagation (Figure 1.2f); deletion of some or all of the repeats destabilizes or eliminates [*PSI*$^{+}$] [40-42].

These experiments demonstrate that while short segments may act as mediators of prion aggregation, larger PFD segments are required for full prion activity, and different regions of a PFD are important for different aspects of this prion activity. Other studies further argue against the importance of short sequence motifs. The relative insensitivity of PFDs to scrambling suggests either that short sequence motifs are not important for prion activity or that the sequence requirements for any such motifs are sufficiently flexible that they are likely to be generated by random chance within scrambled PFDs. In addition, deletion analysis of one of the scrambled versions of Ure2 showed that while progressively larger truncations resulted in gradually decreasing prion-forming ability, no single segment within the PFD was absolutely required for prion activity [33]. Together, these results suggest that length and composition of PFDs are more important than any particular primary sequence element.

Curiously, much smaller segments are sufficient for *in vitro* aggregation (Figure 1.2a). Six- and seven-amino acid segments from Sup35 can form amyloid aggregates *in vitro* [43]. Likewise, eight-residue peptides from Ure2 form amyloid fibrils *in vitro* [44]. Peptide arrays of 20-amino acid fragments from Sup35 revealed multiple fragments spanning amino acids 9-39 that efficiently nucleate aggregation of the Sup35 PFD [45]. The basis for this dramatic difference in length requirements for *in vitro* versus *in vivo* aggregation is unclear.

Although some similar results have been seen for other PFDs, each has its own variations. For many of the prion proteins, the minimal prion domain has not been rigorously mapped, making it difficult to draw broad conclusions. For Ure2, amino acids 1-65 are sufficient to maintain [*URE3*] [25]. A smaller 42-amino-acid segment (amino acids 1,20-65) is capable of inducing prion formation by full-length Ure2 [33], but this fragment has not been tested for prion maintenance, so the exact minimum requirements for prion maintenance are unclear.

Intriguingly, amino acids 1-37 of Swi1 are sufficient for *in vivo* aggregation, induction, and transmission of the [SWI^+] prion [46]; this fragment is notably shorter than other minimal PFDs. By contrast, Rnq1, which forms the [PIN^+] (also known as [RNQ^+]) has a much larger and more complex PFD. The PFD spans residues 153-405, and contains four Q/N-rich segments [47]. Three of these are capable of supporting amyloid aggregation *in vitro* [48]. Deletion of any one of the Q/N-rich segments does not result in loss of [PIN^+] *in vivo*, indicating that Rnq1 contains multiple distinct prion determinants. Indeed, either the second or fourth Q/N-rich segment (spanning amino acids 218-263 and 337-405, respectively) is sufficient to maintain a very weak form of [PIN^+] when fused to the non-Q/N-rich N-terminal domain (amino acids 1-132).

Collectively, analysis of these PFDs creates a series of challenges that must be accounted for in building effective prion prediction methods. Specifically, while short stretches appear to

act as key nucleating elements, much longer segments are required for *in vivo* prion activity for each of the characterized PFDs. Furthermore, for most proteins, the exact boundaries for prion activity are not rigidly defined, as progressive PFD truncations frequently result in progressively diminishing prion activity. Finally, while PFD length seems to be a key factor in determining prion activity, the exact length requirements vary substantially between proteins.

Surrounding Regions Exert Subtle Effects on Prion Activity

Although PFDs are generally thought of as functionally independent domains, prion activity does appear to be somewhat context dependent. For example, one common assay for prion activity is carried out by replacing part or all of the PFD of Sup35 with a suspected PFD fragment from another protein and testing for loss of Sup35 activity. Although this method has helped identify new prions and candidate prion proteins in yeast [2, 7], the PFDs from two known yeast prions, Cyc8 and Mot3, show no prion activity when fused to the Sup35 C-terminus [2]. Conversely, the suspected PFD of another yeast protein, New1, shows prion activity when fused to Sup35 [49], but full-length New1 has not been shown to exhibit prion activity. Many additional candidate prion domains identified in the Alberti and Halfmann *et al.* screen show prion activity in all four assays tested, but also have not been reported to form prions in their native context [2].

Mutations outside of PFDs can also substantially affect prion activity. For Ure2, deletion of an eight amino acid segment from the middle of the functional domain increases prion induction by about 100-fold [25]. For Sup35, select mutations or deletions within the C-terminal domain of wild-type Sup35 result in minor changes in prion formation efficiency [25, 50, 51].

Likewise, mutations in the M domain can affect the efficiency of [*PSI*$^{+}$] propagation, potentially by affecting chaperone interactions [27, 52].

Regions outside core PFDs could influence prion activity by a variety of mechanisms. First, such regions could actively stabilize amyloid fibrils. For example, while the M domain of Sup35 is not required for amyloid aggregation, solid state NMR suggests that it may participate in cross-β-sheet interactions within Sup35 fibers [53]. Second, the non-prion domains could affect accessibility of the PFD; for example, the non-prion domains could directly bind to the PFD and reduce the PFD's structural flexibility. Finally, non-prion domains could affect interactions with factors, such as chaperones, that influence amyloid aggregation.

PREDICTING PRION PROPENSITY IN YEAST

Attempts at Prion Prediction

Many algorithms have been generated to predict aggregation propensity, each using a unique set of parameters. Examples include BETASCAN [54], its more recent relative STITCHER [55], Zyggregator [56], Zipper DB [57], Tango [58], SALSA [59], PASTA [60], and Waltz [61]. Although many of these algorithms have successfully predicted some amyloid proteins, none have demonstrated the ability to predict either the aggregation activity or prion activity of Q/N-rich proteins [62]. The failure to predict prion activity is not surprising, as these algorithms are specifically designed to predict aggregation, and therefore do not account for the other steps in prion activity (Figure 1.1). However, the inability to predict aggregation activity of Q/N-rich domains, as measured by both *in vivo* GFP fusion assays (Figure 1.2c) and *in vitro* amyloid aggregation assays (Figure 1.2a), suggests that there may be differences in the sequence requirements for aggregation between Q/N-rich and non-Q/N-rich proteins. Most amyloid

prediction algorithms are designed to identify short, highly amyloidogenic peptide segments that seem to characterize the majority of non-Q/N-rich amyloid domains. However, it appears that yeast PFDs are characterized by relatively long stretches of disorder-promoting, moderately aggregation-prone amino acids, rather than short stretches of high amyloid propensity [63, 64].

Therefore, while short segments may be sufficient for aggregation either in isolation or in the context of non-Q/N-rich domains, their presence is not sufficient for aggregation activity in the context of Q/N-rich domains. For example, the structure-based algorithm ZipperDB uses a 6-amino-acid window size to identify aggregation-prone segments; sequences are threaded into a known NNQQNY amyloid-forming hexapeptide crystal structure and the energetic fit is determined [57, 65]. Remarkably, insertion of a single aggregation-prone 6-amino-acid segment into an exposed loop in RNAse A is sufficient to cause amyloid formation [66]. However, the same does not seem to be true for Q/N-rich proteins. Because Q/N-rich segments tend to be intrinsically disordered, the RNAse A result would seem to suggest that Q/N-rich regions containing ZipperDB-positive segments should form amyloid aggregates. Instead, ZipperDB-positive segments are found in many Q/N-rich domains that show little or no detectable amyloid aggregation activity, and the presence of ZipperDB-positive segments shows little correlation with amyloid aggregation propensity for Q/N-rich domains [2, 62].

Similar results are seen for Waltz, another prediction algorithm that uses a 6-amino-acid window size [61]. Maurer-Stroh *et al*. analyzed over 200 hexapeptide sequences for cross β-sheet structure formation, and used these results to generate a position-specific matrix to predict amyloid propensity. Waltz-positive amyloid stretches do appear to be more common in Q/N-rich proteins that show prion activity; in one analysis of 36 Q/N-rich proteins (half that show prion-like activity, and half that are unable to support either prion or amyloid formation), Waltz-

positive segments were found in 89% of the prion-like proteins, but only 50% of the non-prion proteins [2, 62]. However, subsequent analysis suggests that this modest success is due predominantly to the compositional aspects of Waltz, not due to the position-specific components of the matrix. Amino acid composition is an inherent characteristic of any primary sequence motif. Therefore, any method focused on primary sequence inevitably runs the risk of misattributing compositional effects to primary sequence. In the Waltz scoring matrix, certain amino acids tend to be favored across most or all positions, so prions may tend to have more Waltz-positive sequences simply because they have more of these favored residues. A simple method to determine whether a primary-sequence-dependent algorithm like Waltz is truly identifying primary sequence patterns (rather than simply acting as an imperfect surrogate for assessing composition) is to make the algorithm blind to the original primary sequence of a test set of proteins by scrambling the sequences *in silico* and re-analyzing them with the algorithm. After scrambling, the prion sequences still had substantially more Waltz-positive segments than the non-prion sequences, suggesting that Waltz is detecting compositional differences between the prion and non-prion set [62].

More recently, the Waltz algorithm has been adapted to specialize in prion aggregation prediction [67]. The new Waltz algorithm, known as pWaltz (which can be accessed through the PrionW webserver [68]), utilizes the position-specific scoring matrix implemented in the original Waltz algorithm coupled with a 21-amino acid sliding window to scan protein sequences for high-scoring windows. However, as with the original Waltz algorithm, the contribution of the position-specific scoring matrix to prediction accuracy was not explored. Indeed, randomization of the position-specific scoring matrix actually appears to improve prion prediction success [69],

again suggesting that the success of Waltz/pWaltz is predominantly due to underlying composition biases rather than primary sequence effects.

Collectively, these results argue that algorithms that are built based on *in vitro* analysis of short fragments may have little ability to predict aggregation propensity of Q/N-rich proteins. However, the insensitivity of yeast PFDs to scrambling [32, 33] indicates a possible alternative prediction approach. Specifically, the dominant role of composition suggests that compositional similarity to known prions could be used to predict prion activity.

Curiously, this does not seem to be the case. Alberti *et al*. used a Hidden Markov Model to identify the 100 yeast protein domains with greatest compositional similarity to known yeast PFDs [2]. All candidates were tested in four prion-like activity assays. A remarkable number of proteins (18 out of 100) showed prion-like activity in all four assays, suggesting that compositional similarity does reasonably well at separating potential prion candidates from the bulk yeast proteome [2]. However, there was little correlation between the degree of compositional similarity to the known yeast PFDs and observed prion-like activity [64]. Other composition-based searches yield similar results: they successfully identify prion candidates but cannot identify the actual prion-forming proteins among those candidates [7, 70, 71].

It should be noted that it is difficult to evaluate exactly how good the Alberti *et al.* algorithm is at identifying prion candidates. The number of prions in yeast is not known; no one has tested what fraction of randomly selected protein fragments would show prion-like activity in these four assays, so there is no benchmark against which to judge the observation that 18 out of 100 tested fragments had clear prion-like activity. A related concern is that all four assays involve removal of the predicted PFDs from their native context, which may artificially inflate the number of fragments showing prion activity. Tartaglia *et al.* have eloquently argued that

evolutionary selection tends to reduce the aggregation propensity of proteins to just below the threshold for aggregation in their normal biological environment; consequently, even minor changes in sequence, expression level or environment may cause aggregation [72]. Therefore, while the work of Alberti *et al.* provides strong data about the intrinsic aggregation propensity of each candidate PFD (and thus provides an incredibly powerful dataset for testing any prediction algorithm), it is possible that many domains showing prion-like activity in these assays will not form prions in their native context.

However, it is unlikely that this issue fully explains the large number of proteins showing prion-like activity in the Alberti *et al.* assays. Of the 100 proteins tested, the 50 with highest compositional similarity to known prions on average had significantly higher prion-like activity than the next 50 [62]; this suggests that the algorithm has some ability to enrich for likely prion candidates. But, among the top 50 proteins, there was actually a small, statistically insignificant inverse correlation between compositional similarity to known prions and prion-like activity, suggesting that the algorithm has no ability to distinguish among the top candidates. The simplest explanation for this apparent contradiction is that the sequence features that most clearly distinguish Q/N-rich PFDs from the rest of the proteome are not necessarily the same features that would be most effective at distinguishing Q/N-rich PFDs from non-prion-forming Q/N-rich domains.

A related issue is that compositional similarity analyses implicitly assume that all deviations from the known PFDs will decrease prion-forming capacity. In reality, prion formation is an exceedingly rare event, so it is unlikely that PFDs are optimized for maximum prion propensity. Therefore, it is possible that some compositional changes may increase prion

propensity. More accurate prion prediction requires an understanding of how deviations from the compositions of known PFDs will affect prion activity.

Determining the prion propensity of each amino acid would provide a means to predict exactly how compositional changes will affect prion propensity. In a preliminary attempt to determine these prion propensities, a segment from a scrambled version of Sup35 was replaced with a random sequence, thereby generating a library of mutants [64]. By comparing the frequency of occurrence of each amino acid in the initial library to the frequency of the amino acid among the subset of mutants that maintained the ability to form prions, a prion propensity score was developed for each amino acid. In general, hydrophobic and aromatic amino acids were found to be strongly prion-promoting, polar amino acids were relatively neutral, and charged residues and prolines were strongly prion-inhibiting.

These prion propensity scores were then used to generate a prediction algorithm, called PAPA (prion aggregation prediction algorithm; http://combi.cs.colostate.edu/supplements/papa/) [62, 64]. PAPA uses a 41-amino acid sliding window, calculating the prion propensity of each window by averaging the prion propensity scores for each amino acid within the window. In addition to calculating prion propensity, PAPA uses the FoldIndex algorithm [73] to predict ordered and disordered regions within the protein. A key feature of yeast PFDs is that they are intrinsically disordered [30, 31]. Therefore, PAPA scores the overall prion propensity of each protein by identifying the 41 consecutive 41-amino acid windows that have both the highest predicted prion propensity and a negative fold index score (i.e., they are predicted to be disordered). The window size was chosen based on the observation that approximately 40 amino acids seem to be required for aggregation of most yeast PFDs, yet the flanking sequences can also affect aggregation propensity.

Strikingly, a strong correlation was seen between PAPA scores and observed prion propensity. The scores for the 100 domains tested by Alberti *et al.* ranged from approximately -0.13 to 0.15. A cutoff of 0.05 was most effective at discriminating between proteins with and without prion-like activity (Figure 1.3). Of the 18 proteins that showed no prion-like activity in any of the Alberti *et al.* assays (Figure 1.3, red diamonds), 17 scored below 0.05. By contrast, of the 18 proteins that showed prion-like activity in all four assays (Figure 1.3, green diamonds), 16 scored above 0.05. Additionally, of the 37 proteins that scored above 0.05, 36 showed prion-like activity in at least one assay. However, this cut-off is not absolute. Among proteins that scored between 0.00-0.05, many showed at least some prion-like activity. Thus, these scores may be more accurately viewed as a gradient. In general, Q/N-rich proteins scoring below 0.00 are likely to have little or no prion activity; proteins scoring between 0.00-0.05 may have some prion activity; proteins scoring from 0.05-0.10 are likely to have some prion-like activity; and proteins scoring greater than 0.10 are likely to have strong prion activity. Further supporting the utility of PAPA, the algorithm was subsequently used to design completely synthetic Q/N-rich PFDs; when these domains were substituted in place of the Sup35 PFD, they were able to support prion activity [62].

Surprisingly, there is little correlation between the frequency of occurrence of each amino acid among yeast PFDs and the amino acid's PAPA score. As expected, charged residues and prolines have low prion propensity according to PAPA, consistent with their relative rarity in yeast PFDs [64, 70]. Unexpectedly, Q/N residues scored relatively neutral despite their prevalence in yeast PFDs, while hydrophobic residues, which are rare in yeast PFDs [70], scored as highly prion promoting. The importance of intrinsic disorder likely explains this apparent contradiction, and offers a simple theory to explain the compositional make-up of yeast PFDs

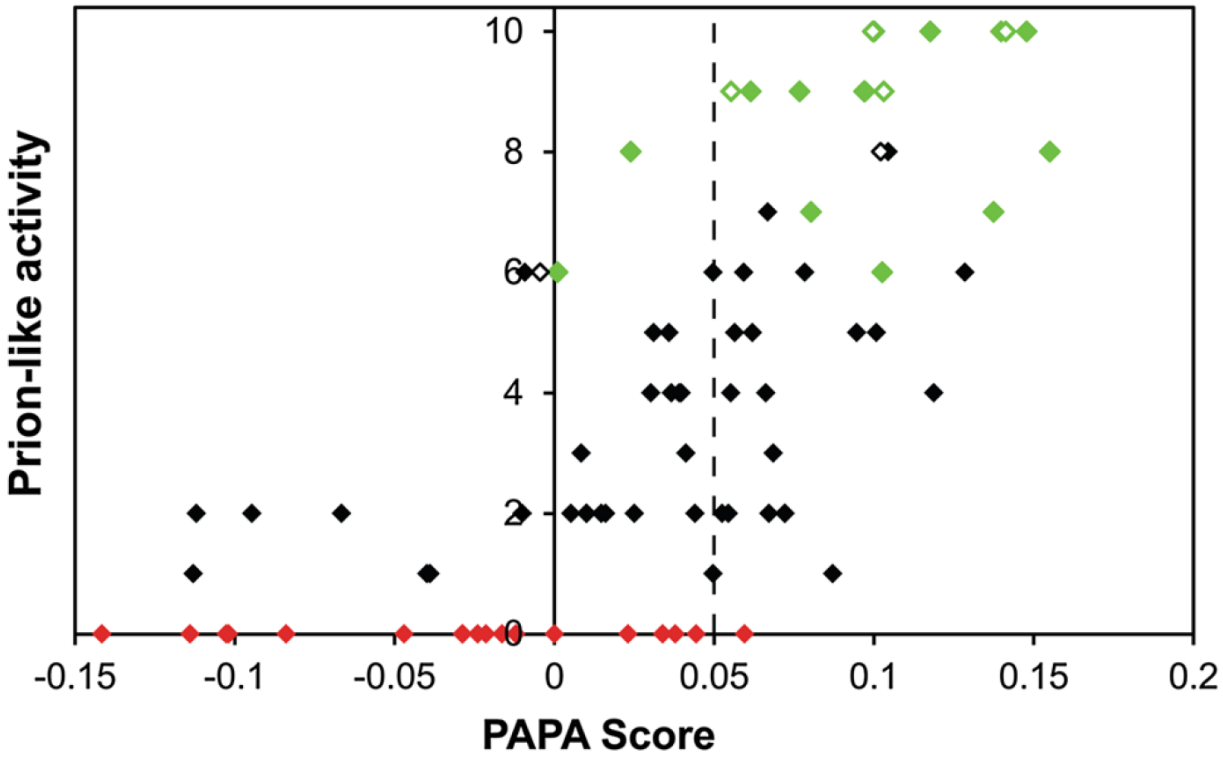


Figure 1.3: PAPA predictions for yeast prion-like proteins. Alberti *et al.* identified the 100 proteins with greatest compositional similarity to known yeast prions. Each was tested in four assays for prion-like activity, and given a prion-activity score from 0-10 based on these results. PAPA was then used to predict the prion-like activity of each protein. Domains that were not testable in one or more assays are excluded. Proteins that showed prion-like activity in all four assays are indicated in green. Proteins that failed to show prion-like activity in any assay are indicated in red. Known PFDs are indicated with open diamonds. The PAPA cutoff shown to most effectively discriminate between proteins with and without prion-like activity (0.05) is indicated with a dotted line. However, this is not an absolute cut-off; between 0-0.05, most proteins show at least some prion-like activity. Updated and adapted from [64].

[64]. The disordered nature of yeast PFDs makes the individual residues more accessible for prion formation. Q and N are likely common in yeast PFDs at least in part because they nicely balance prion propensity and disorder propensity. Most disorder-promoting residues are strongly aggregation-inhibiting. By contrast, Q/N residues promote intrinsic disorder while also providing a slight positive contribution to prion formation. In this context, very few hydrophobic residues are needed to drive aggregation. Additional hydrophobic residues would likely either make proteins excessively aggregation-prone, or create aggregates that are too stable, and thus not easily fragmented.

This theory also helps to reconcile other apparent contradictions. The proposed importance of short stretches for nucleating prion formation [37, 39, 45] seems to conflict with the insensitivity of PFDs to scrambling [32, 33]. However, if PFDs contain relatively few strongly prion-promoting amino acids, then the distribution of these amino acids will naturally create pockets of strong nucleating potential. Scrambling will simply redistribute these key amino acids, again creating nucleating sites wherever prion-promoting amino acids cluster. Indeed, the region spanning amino acids 8-26 of Sup35, which is thought to act as a critical nucleating site [37, 39, 45], contains two strongly prion-promoting amino acids (both tyrosine), and contains the longest stretch in the Sup35 PFD without any strongly prion-inhibiting amino acids. Thus, the presence of nucleating stretches can be rationalized based entirely on composition.

This may also explain why algorithms such as Waltz and ZipperDB show some ability to correctly identify key nucleating sites, but are less effective at distinguishing between proteins with and without prion-like activity. Consistent with the findings of Toombs *et al.* (2010), aromatic and hydrophobic residues tend to score high at most positions in Waltz, while charged

residues tend to score low at most positions. Therefore, although algorithms such as Waltz incorporate an additional layer of primary sequence, they may be acting predominantly as a screen for local amino acid composition.

Future Challenges in Yeast Prion Prediction

Although great strides have been made in predicting prion activity, much remains to be understood. The Alberti algorithm and PAPA have potentially complementary strengths and weaknesses (Table 1.3). The Alberti algorithm was very successful at identifying prion candidates from the *S. cerevisiae* genome, but could not accurately predict which of the candidates would demonstrate prion activity [2, 64]. Conversely, PAPA was able to accurately predict prion activity within the Alberti *et al.* candidate dataset, and could even be used to build synthetic PFDs, but it is unclear whether PAPA may itself be used to identify prion candidates from whole genomes [62]. A unified prediction method could improve prion prediction, but designing such an algorithm requires overcoming a number of current challenges.

One major challenge is the lack of good datasets on which to train and test potential algorithms. For example, the Alberti algorithm was trained on the four prion proteins that were known at the time: Sup35, Ure2, Rnq1, and the prion candidate New1 [2]. This small training set may have limited the algorithm's prediction accuracy. One of the greatest contributions of the work of Alberti *et al*. is that it provides a large, rigorously tested dataset; importantly, it includes domains that compositionally resembled yeast PFDs, but that show no prion activity. Recently, Espinosa Angarica *et al.* took advantage of this to develop a new prediction algorithm [74]. They used the full set of prion-like proteins from the Alberti *et al.* dataset to develop a probabilistic representation of Q/N-rich PFDs. This algorithm was reasonably effective both at discriminating

Table 1.3: Strengths and limitations of PAPA and the Alberti algorithm for prion prediction.

Strengths	Limitations
Alberti et al. [2]	
Built based on analysis of multiple PFDs	Ineffective at ranking the highest scoring candidates
Can identify prion candidates from proteomes	Ability to predict the effects of point mutations is unclear
High fraction of candidates show prion-like activity	Does not consider primary sequence effects
	Does not consider effects of regions outside of the PFD or interactions with heterologous proteins
PAPA	
Reasonably effective at ranking candidate PrLDs	Built based on mutagenesis of a small region of a single protein
Sufficient for de novo design of Q/N-rich PFDs	Validated only on Q/N-rich proteins
Uses experimentally derived prion propensity values for each amino acid, allowing for prediction of the effects of amino acid substitutions	Effectiveness for genomic searches is unclear
	Does not consider primary sequence effects, other than proline spacing
	Does not consider effects of regions outside of the PFD or interactions with heterologous proteins

between proteins with and without prion activity from among the Alberti *et al.* dataset and at picking known PFDs out of larger datasets. The algorithm identified 20540 predicted PFDs in 1536 organisms, but none of these new candidates have yet been tested for prion activity.

Although the growing list of known PFDs provides a broader dataset for training potential algorithms, this dataset may contain its own biases. The majority of known yeast prions or prion-like domains were identified either because of their sequence similarity to Sup35 and Ure2 or due to their ability, when overexpressed, to support [*PSI*$^{+}$] formation in a [*pin*$^{-}$] strain [35]. Therefore, they may not provide a representative sample of all yeast prion proteins, and any algorithm that uses this set of proteins as a training set runs the risk of being too narrowly focused.

In theory, one advantage of PAPA is that, because it actually scores the prion propensity of each amino acid rather than simply looking for compositional similarity, it is not constrained by any biases present in the current set of known prions. However, PAPA faces its own challenges in scoring proteins whose composition deviates from that of Sup35. PAPA's prion propensity scores for each amino acid are only estimates, based on a random sampling of prion and non-prion isolates from a library of scrambled Sup35 mutants; therefore, each prion propensity score carries large confidence intervals, which creates errors in PAPA's predictions. The further a protein's composition deviates from that of Sup35, the more these errors will likely compound. Additionally, PAPA assumes a linear relationship between the frequency of occurrence of a given amino acid and prion propensity. This assumption almost certainly is an over-simplification; some amino acids may have non-linear relationships with prion propensity or show a threshold effect. For example, within Sup35, insertion or deletion of a single Q/N residue generally has little effect on prion activity; however, it is possible that proteins with

lower concentrations of Q/N residues might be more sensitive to changes in Q/N content. The more a protein's composition deviates from that of Sup35, the higher probability that this sort of non-linear relationship will affect prediction accuracy.

While both PAPA and the Alberti algorithm focus on composition, there may be minor primary sequence elements that these algorithms do not account for. Scrambled versions of Sup35 and Ure2 form and maintain prions with different efficiencies, indicating slight primary sequence effects [32, 33]. Elucidating these subtle primary sequence features is difficult, because most experimental methods affect both primary sequence and composition. For example, deletion experiments remove specific primary sequence elements but concomitantly result in a disproportionate loss of particular amino acids, thereby altering composition. This could result in misattribution of compositional effects to primary sequence elements.

PAPA does include in its prediction method one commonly recognized primary sequence feature governing prion formation. Since proline is a known β-sheet breaker [75], the distribution of prolines can greatly affect prion formation [2, 64]. A cluster of prolines would be expected to disrupt β-strand formation at just a single location, while these same prolines dispersed across a sequence would result in multiple disruptions of the β-sheet structure. Accordingly, PAPA classifies any set of two or more prolines, separated by no more than one amino acid each, as a single proline. This is only one of potentially many subtle primary sequence features. However, analysis of the libraries used to build PAPA has not revealed any other clear primary sequence biases. There is neither clear co-variance between particular amino acids nor positional biases of individual amino acids. But, because of the limited library sizes, subtle effects could easily have been missed.

Another major challenge is that various studies suggest that the distinct steps required for prion activity (Figure 1.1) may have distinct compositional requirements. As previously discussed, the nucleation domain of Sup35 (amino acids 1-39) is thought to be primarily responsible for the initial nucleating events in prion formation and for fiber growth (Figure 1.1, steps 1-3), while the ORD is thought to be primarily responsible for chaperone-dependent prion maintenance (Figure 1.1, step 4). While the activity of the ORD is primary-sequence independent, when the composition of the ORD is changed to match that of the nucleation domain, it is no longer able to support prion maintenance [76]; indeed, recent evidence suggests that prion formation and prion maintenance have distinct compositional requirements [77]. Further complicating matters, individual amino acids can have differential effects on the discrete steps in prion formation (Figure 1.1, steps 1-3; [78]). Specifically, prion proteins are thought to first associate with each other to form soluble oligomers (Figure 1.1, step 1), and these oligomers then undergo a structural conversion to form ordered amyloid fibrils (Figure 1.1, steps 2-3). Interestingly, glutamines seem to promote the formation of soluble non-amyloid oligomers, while asparagines seem to promote the formation of mature amyloid fibrils [78]. Thus, an accurate prediction algorithm needs to consider not only the overall prion propensity of each amino acid, but also the effect of each amino acid on each step in prion formation and propagation.

PRION-LIKE DOMAINS IN DISEASE

Many protein aggregation-based diseases, like the prion diseases, involve self-templating structural conversions. But, since most protein aggregates are not infectious, the prion diseases have historically been viewed as fundamentally distinct from other aggregation-based disorders.

However, recent developments have begun to blur this distinction between prion and non-prion aggregation diseases [79-81].

There is growing evidence that various proteins implicated in many neurodegenerative disorders show prion-like behavior. Notable examples of these proteins include α-synuclein; amyloid precursor protein (APP) and tau; and Huntingtin. These proteins are implicated in Parkinson's disease, Alzheimer's disease, and Huntington's disease, respectively [80]. Although the clinical manifestations of these disorders vary, the prion-like behavior of the implicated proteins is roughly the same. Presumably, an initial misfolding event results in aggregation. The aggregates can then induce other proteins to similarly misfold and aggregate. This aggregation is thought to originate in a single epicenter and spread to neighboring tissues by an unknown mechanism, imposing aggregation commitment on nearby cells [82]. Despite this remarkable similarity to prion propagation, the hesitation in classifying these neurodegenerative proteins as bona fide prions arises from the lack of evidence of transmission between individuals [80, 83].

Additionally, a number of proteins containing domains with similar amino acid composition to yeast PFDs have recently been linked to degenerative diseases (Table 1.4), further highlighting the connection between infectious and non-infectious protein aggregation diseases. Interestingly, many of these disease-associated PrLD-containing proteins also have RNA recognition motifs (RRMs) that generally do not overlap with the predicted PrLDs. This suggests that disruptions in RNA homeostasis via prion-like aggregation may represent a common mechanism of degenerative disease. The Alberti algorithm predicts 246 out of 21,873 genes in the human genome to have sequences encoding PrLDs [36, 84]. This group contains a relatively high proportion of RNA-binding proteins – nearly 12% of predicted PrLD-containing

Table 1.4: RNA-binding proteins containing PrLDs that have been linked to degenerative disease

Protein	Alberti Algorithm		PAPA analysis		Disease
	PrLD Rank[a]	PrLD amino acids	PAPA score	Highest Scoring Segment (isoform)[b]	
TDP-43	43	277-414	0.042	339-414	ALS, FTLD
FUS	12	1-237	0.109	21-121 (2)	ALS, FTLD
TAF15	22	1-152	0.127	12-92 (2)	ALS, FTLD
EWSR1	25	1-280	0.057	194-274 (1)	ALS, FTLD
hnRNPA2B1	32	197-353	0.043[c]	241-321 (A2)	IBMPFD
hnRNPA1	38	186-372	0.093[d]	257-337 (b)	ALS, IBMPFD
TIA1	53	292-386	0.131	269-349 (1)	Welander distal myopathy

[a] PrLD rank among the human genome [36].

[b] Some of the proteins have multiple isoforms that differ either in their PAPA scores or the exact location of the highest scoring region. Shown are the amino acid positions for the highest scoring segments from the highest scoring isoform, with the isoform indicated in parentheses.

[c] The PAPA score for the disease-associated mutant is 0.88.

[d] PAPA score for Isoform B, the highest scoring isoform. hnRNPA1 has two isoforms with substantially different PAPA scores. The IBMPFD-associated mutation increases the PAPA score of Isoform B to 0.125. Isoform A scores 0.041, but the IBMPFD-associated mutation increases the PAPA score to 0.087.

proteins also contain at least one RRM. Furthermore, 20% of the top 60 proteins predicted to contain PrLDs also contained at least one RRM, raising the possibility that more of these proteins could eventually be linked to disease.

RNA-Binding Proteins with Prion-Like Domains in Human Disease

TDP-43 was the first PrLD-containing RRM protein to be associated with a degenerative disease. It was identified as a major component of aggregates in patients with either amyotrophic lateral sclerosis (ALS) or some forms of frontotemporal lobar degeneration (FTLD) [85]. TDP-43 contains two RRM's and a C-terminal PrLD. Normally, TDP-43 is primarily located in the nucleus, but in patients with ALS, it is found in cytoplasmic inclusions [85]. Interestingly, TDP-43 has since been found in inclusions in patients with a variety of other neurodegenerative diseases, including Alzheimer's and Parkinson's diseases [86]. A variety of evidence implicates the PrLD in disease. Overexpression of TDP-43 in yeast, *C. elegans*, or *Drosophila* results in TDP-43 aggregation and toxicity, and the PrLD is required for this aggregation and toxicity [87-89]. Furthermore, at least 44 mutations in TDP-43 have been identified in patients with ALS or FTLD; of these, 41 reside in the PrLD [90]. However, regions outside the PrLD also affect aggregation and toxicity. For example, the C-terminal domain is necessary, but not sufficient, for aggregation and toxicity in yeast [88]. The presence of at least one RRM is also required for toxicity in yeast. Additionally, some of the ALS-associated mutations do not accelerate aggregation *in vitro* or cause toxicity in yeast. Together, these results suggest a more complex mechanism of toxicity [91].

FUS was the second RRM-containing protein to be linked to neurodegenerative disease. It contains an N-terminal PrLD, a single RRM, and two C-terminal "RGG" domains (multiple

gly-gly motifs interspersed with arginine and aromatic residues), one of which barely misses the defined PrLD cutoff according to the Alberti algorithm [36, 92]. Mutations in FUS cause familial ALS [93, 94] and aggregation of FUS has been linked to both ALS and FTLD [90]. FUS is normally a predominantly nuclear protein, but these mutations cause FUS to form cytoplasmic aggregates. FUS has since been found in cytoplasmic inclusions in patients with other neurodegenerative diseases, including Huntington's Disease [86]. FUS is also highly aggregation-prone *in vitro* and causes toxicity when expressed in yeast [92]. The PrLD is necessary, but not sufficient for this aggregation and toxicity [92], and the RNA-binding ability of FUS seems to be critical for toxicity in yeast and *Drosophila* models [95]. The ALS-associated mutations in FUS seem to cluster in two regions: the N-terminal PrLD and a short C-terminal segment containing a predicted nuclear localization signal [90].

Two other RNA-binding proteins that ranked highly in prion prediction analyses, TAF-15 and EWSR-1, have also been linked to sporadic ALS and FTLD. Again, each protein contains a predicted N-terminal PrLD, as well as a single RRM and two RGG domains [36]. For both TAF-15 and EWSR1, mutations have been found in a small number of ALS patients that do not appear in control subjects [84, 96]. These mutations all occur outside of the PrLD, clustering in and around the RGG domains. Both proteins are inherently aggregation-prone *in vitro*, and the disease-associated mutations accelerate aggregation in each case [84, 96].

Additionally, point mutations in either hnRNPA2B1 or hnRNPA1 have been shown to cause familial IBMPFD/ALS (inclusion body myopathy with Paget's disease of bone, frontotemporal dementia, and ALS; [97]). In both cases, the causative mutation is a single aspartic acid to valine substitution within the PrLD. Additional mutations were identified in patients with both familial and sporadic forms of ALS. In normal muscle cells, hnRNPA2B1 or

hnRNPA1 are predominantly nuclear; however, in patients carrying the mutations, the proteins form large cytoplasmic inclusions. *In vitro,* the wild-type proteins are intrinsically aggregation-prone, and the disease-causing mutations accelerate this aggregation. In *Drosophila* and mouse models, expression of the mutant proteins leads to formation of large cytoplasmic inclusions and severe muscle degeneration. Furthermore, when the core PrLD from mutant hnRNPA1 and hnRNPA2 are substituted in place of the nucleation domain of Sup35, they can support prion formation in yeast, while the wild-type PrLDs cannot.

Mutations in another prion-like protein, TIA1, have recently been shown to cause Welander distal myopathy, a dominant adult-onset disorder characterized by progressive distal limb weakness [98]. Patient muscle biopsies showed TIA1 and TDP-43 staining adjacent to intracellular inclusions.

In all, of the 20 RNA-binding proteins that are scored highest by the Alberti algorithm, 10 have been linked to degenerative disease [99]. An obvious question is why so many of these RNA-binding proteins have maintained aggregation-prone PrLDs if aggregation of these domains is associated with neurodegenerative disease. A growing body of evidence suggests that the PrLDs may play a functional role, such as recruiting RNA-binding proteins to P bodies or stress granules under cellular stress [99, 100]. TIA-1 plays a critical role in stress granule formation, and the TIA-1 PrLD is required for stress granule formation [101]. Remarkably, the Sup35 PFD can substitute for the TIA-1 PrLD in supporting stress granule assembly, linking prion-like aggregation to stress granule formation [101]. Many of the disease-associated PrLD-containing proteins are recruited to stress granules, and mutations enhance this recruitment [97, 100]. A number of recent studies indicate liquid-liquid phase separation as the biophysical basis for stress granule assembly [102-106] (for review, see also [107-109]). Furthermore, disease-

associated mutations in RNA-binding proteins can alter the dynamics and characteristics of stress granule aggregates [104-106]. This suggests a model in which these RNA-binding proteins form reversible stress granule aggregates, but where mutations or changes in the cellular environment, such as prolonged stress, can lead to excessive or pathogenic stress granule formation [100, 110].

Consistent with this theory, mutations that disrupt the turnover of RNA-protein aggregates have also been linked to degenerative disease. Specifically, VCP/p97 is a well-characterized AAA ATPase that is involved in disassembling protein complexes containing ubiquitinated proteins [111]. Mutations in VCP have been shown to cause both IBMPFD [112] and ALS [113]. These mutations result in the formation of cytoplasmic inclusions containing TDP-43 and other stress granule markers [110]. Additionally, over-expression of these mutants inhibits stress granule clearance by autophagy [114]. Collectively, these results suggest that these diseases result from impairment of the normal dynamics of RNA granule assembly, disassembly, and clearance.

Predicting Disease-Associated Proteins: Successes and Future Challenges

The various disease-associated mutations in human PrLDs demonstrate both the strengths and weaknesses of existing prediction algorithms, and highlight some of the challenges in predicting human PrLDs.

hnRNPA1 and hnRNPA2 offer the best examples of proteins that are accurately predicted by current algorithms. Wild-type hnRNPA2 scores below PAPA's 0.05 threshold for high aggregation propensity, but the disease-associated mutation pushes the PAPA score past the aggregation threshold ([91]; Table 1.4). The same is true for the most highly expressed isoform of hnRNPA1 (Isoform A; [97]) – the wild-type protein scores below 0.05, but both of the

mutations linked to familial forms of IBMPFD or ALS are predicted by PAPA to push the proteins beyond the threshold of aggregation. hnRNPA1 has another isoform (Isoform B) that surpasses the PAPA threshold even in the wild-type state, but in which the mutations are predicted to further enhance the aggregation activity. ZipperDB also correctly predicts the effects of each mutation, predicting that they should create strong steric zipper motifs. Although these mutations within the hnRNP's represent important cases validating our prion predictions in human proteins, an expanded set of mutations in the hnRNP's would allow more thorough examination the efficacy of PAPA. In Chapter 2, I present evidence that PAPA is generally effective at predicting a broad range of mutations *in vitro*, in yeast, and in *Drosophila*.

While the Alberti algorithm correctly identifies the PrLDs in both proteins, it predicts that the mutations will have little effect on prion propensity. These mutations may offer a good example of the limitations of using algorithms based on compositional similarity to known PFDs to predict the effects of mutations. The two mutations linked to IBMPFD both involve substitution of an aspartic acid with a valine. Because both residues are extremely rare in yeast PFDs, algorithms based on compositional similarity will score this as a relatively neutral substitution. However, aspartic acid and valine are likely rare in yeast PFDs for opposite reasons – aspartic acid because it strongly inhibits prion formation, and valine because it too strongly promotes prion formation, creating a strong selective pressure against its inclusion in PFDs.

The Alberti algorithm also correctly predicts PrLDs in each of the other disease-associated RNA-binding proteins (Table 1.4). Likewise, PAPA scores EWSR1, FUS, TAF15, and TIA-1 above the predicted threshold for aggregation. TDP-43 scores just below the 0.05 threshold, within a range that is generally associated with some aggregation activity (Table 1.4; Figure 1.3). The fact that both the Alberti algorithm and PAPA score the wild-type proteins as

prion-like or aggregation-prone could be considered accurate, since the wild-type proteins each appear to be aggregation-prone [84]. However, this also highlights a key limitation of these algorithms – while both PAPA and the Alberti algorithm have shown success at identifying candidate disease-associated proteins, at this point none of the existing algorithms can consistently predict the exact effects of mutations on either aggregation propensity or pathogenicity.

There are a number of possible reasons for this failing. First, aggregation propensity and toxicity are not always coupled, as evidenced by the observed mutations that cause disease without a detectable change in aggregation propensity [91]. These diseases appear to broadly result from disruptions in normal RNA homeostasis, so aggregation may simply be one of many causes of such disruption.

Second, these algorithms are designed to identify aggregation-prone protein fragments; however, in a cellular context, aggregation-prone fragments may be prevented from aggregating due to factors such as protein-protein interactions (particularly, with protein chaperones or protein degradation systems – reviewed in the next section), interactions with other domains in the protein, post-translational modifications, or cellular localization. This is especially true for the RNA-binding proteins, which appear to form regulated aggregates [110]. Disruption of any of these regulatory mechanisms could ultimately affect the ability to aggregate *in vivo* regardless of the intrinsic aggregation propensity of the PrLD. Indeed, in Chapters 3 and 4, I present evidence that certain aggregation-prone features within the G-rich human PrLDs make them particularly susceptible to protein degradation systems, whereas Q/N-rich yeast PFDs are unusually resistant to degradation. These differences in intracellular regulation likely affect the formation and persistence of prion aggregates.

Third, differences between yeast PFDs and the human disease-associated PrLDs may limit the prediction accuracy of both the Alberti algorithm and PAPA. Both algorithms were designed and validated on proteins within a relatively narrow range of compositions. For example, 96 of the 100 candidate PFDs tested by Alberti *et al.* had greater than 24% Q/N-content. This is significant because, although the yeast PFDs and human PrLDs share many compositional features including an under-representation of charged residues relative to their respective proteomes (Table 1.2), they differ significantly in other ways. The eight Q/N-rich yeast PFDs range from 28.6-46.8% Q/N content, while among the human disease-associated RRM proteins, only the TAF15 and TIA1 PrLDs have greater than 22% Q/N content. Thus, while Q and N are overrepresented among both yeast PFDs and human PrLDs, they are far less overrepresented among the human PrLDs. Conversely, serine and glycine are more overrepresented among the human PrLDs than among the yeast PFDs (Table 1.2). Because the prediction accuracy of any algorithm is likely to decrease the further a protein's composition deviates from that of the algorithm's training set, these differences may limit the prediction accuracy of yeast-derived algorithms for human PrLDs. Therefore, in addition to defining protein features favoring degradation of the human PrLDs, in Chapter 3 we define the amino acids that favor or disfavor prion activity for the human PrLDs.

Finally, differences in cellular environment between yeast and human cells may impose distinct compositional requirements for prion-like activity. For example, Hsp104 is required for propagation of almost all yeast prions, and amino acid composition can affect the efficiency of Hsp104-dependent fiber fragmentation [115]. Therefore, some of the compositional biases seen in yeast prions may be due to specific requirements for Hsp104-dependent fragmentation. However, humans do not possess an Hsp104-homologue (although an Hsp110 in humans

appears to be able to perform a subset of Hsp104 activities [116]), so human proteins may have very different compositional requirements for propagation. Additionally, the reason Hsp104 is required for most yeast prions is that prion aggregates need to be fragmented to create new independently-segregating seeds to offset dilution by cell division; because mammalian neuronal cells typically do not divide rapidly, the levels of fiber fragmentation required for aggregate propagation are likely very different. The mechanism of spread of these neurodegenerative prion-like proteins to neighboring tissues may also differ from the mechanisms of propagation in yeast. Each of these differences may affect the specific compositional requirements for prion-like activity in yeast versus humans.

PROTEIN AGGREGATION AND PROTEOSTASIS

Current prediction algorithms have provided a strong foundational understanding of intrinsically prion-prone protein features. However, *in vivo*, prion aggregation occurs in the context of a variety of factors devoted to maintaining protein homeostasis, or "proteostasis". Proteostasis broadly refers to the production and maintenance of properly folded proteins, the correct assembly of protein complexes, and the clearance of damaged or misfolded proteins. The complexity and importance of proteostasis is reflected in the myriad factors devoted entirely to protein quality control, which can act co-translationally upon nascent polypeptide chains and/or post-translationally on fully synthesized chains and mature proteins. These factors fall into three main categories; the protein chaperones and chaperonins, the ubiquitin-proteasome system (UPS), and the autophagy system – collectively referred to as the "proteostasis network" (PN) [117]. The three branches of the PN are reviewed here, with a particular emphasis on components that prevent or rectify aberrant protein aggregation.

Co-translational Proteostasis

Considering the importance of protein quality control, it is no surprise that proteostasis factors begin working on nascent proteins before they are fully synthesized. Given the high local concentration of peptides at polysomes [118], the exposure of core aggregation-prone segments before the nascent protein folds [119, 120], and excluded volume effects of a crowded cytosol [121, 122], many newly synthesized peptides are at a high risk of aggregation. Eukaryotic cells possess a triad of ribosome-bound complexes that act co-translationally as nascent polypeptide quality control factors. In yeast, the ribosome associated complex (RAC) consists of the Hsp70 and Hsp40 chaperones Ssz and Zuo1 respectively [123]. Although these chaperones are ribosome-bound, they have not been shown to bind to nascent polypeptides – rather, they recruit and stimulate two closely related Hsp70 chaperones Ssb1/2 (together referred to as SSB), which bind and fold nascent chains. Interestingly, SSB substrates tend to have more short linear hydrophobic stretches, higher β-sheet content, and higher predicted aggregation propensity relative to all translated proteins [120]. Furthermore, deletion of RAC or SSB results in accumulation of protein aggregates, suggesting that SSB plays a critical role in stabilizing and folding proteins with aggregation-prone features [120, 124].

The nascent-chain associated complex (NAC) consists of a heterodimer of α and β subunits composed primarily of Egd2 and Egd1 respectively in yeast (although a functionally distinct complex can be formed by Egd2 and alternative β subunit Btt1). The NAC associates with nearly every translated polypeptide in yeast and is thought to stabilize nascent peptides until folding is initiated by downstream chaperones [125]. Although deletion of NAC does not result in growth defects in yeast, deletion of NAC results in accumulation of protein aggregates and

defects in protein sorting [124-126]. Furthermore, deletion of NAC in addition to SSB deletion enhances the widespread aggregation observed for SSB deletion alone [124].

The early-acting chaperone complexes are complemented by a ribosome-bound branch of the UPS known as the ribosome quality control complex (RQC) [127]. This complex consists of ribosome-bound scaffolding proteins Rqc1 and Rqc2, along with the ribosome-bound E3 ubiquitin ligase Ltn1 (although Ltn1 may also play a role in ribosome-independent cytosolic quality control [128]). All three proteins are necessary for efficient recruitment of the final RQC components, Cdc48 and its associated co-factors. The RQC primarily functions in degrading stalled translation products, or non-stop translation of the polyA tail [127, 129]. In addition, the ribosome-bound E3 ubiquitin ligase, Hel2, participates in the preferential ubiquitination and degradation of short proteins (<400 amino acids) with high hydrophobic content [130]. Defects in RQC can result in aggregation of stalled translation products and results in cytotoxicity via sequestration of important proteostasis chaperones [131]. Furthermore, mutation in Listerin (the mouse homolog of Ltn1) results in neurodegeneration, presumably through accumulation of protein aggregates, although similar disease-causing mutations have not yet been discovered in humans [132].

Post-translational Protein Folding and Disaggregation

Cytosolic Hsp70 chaperones (along with a suite of Hsp40 co-chaperones with various adaptor functions) act in the later stages of protein synthesis and in the early folding stages of ribosome-released substrates. Hsp70 chaperones act primarily to maintain client solubility via iterative ATP-dependent binding and release cycles, which prevents aggregation of exposed hydrophobic residues within the client peptide sequence until they are successfully buried in the

interior of the protein [133-138]. If Hsp70's are unable to complete folding, the substrates are passed to the Hsp60 chaperonin complexes GroEL/GroES (bacteria and mitochondria) and TRiC (eukaryotes), or to Hsp90 chaperones to complete protein folding (alternatively, Hsp70's have been shown to cooperate with Hsp40 chaperones in the degradation of terminally misfolded substrates [139]) [140, 141]. In addition to their roles in *de novo* protein folding, these chaperone systems likely participate in surveillance and maintenance of mature proteins to maintain proper protein folds throughout a protein's tenure in the cell. Interestingly, although the molecular mechanisms of Hsp70's, Hsp60's, and Hsp90's differ dramatically (for review, see [118, 142]), their chaperone activity fundamentally involves the stabilization of aggregation-prone features during the folding and re-folding processes.

Apart from constitutively active chaperones, a variety of small Hsp's (sHsp's), Hsp40's, Hsp70's, Hsp90's, and Hsp100's are transcriptionally induced by various cellular stress (most notably heat stress) that induce protein misfolding or denaturation. Interestingly, induction of these stress-related chaperones is coupled with transcriptional repression of the constitutively active chaperones [143], yet some of the pre-existing ribosome-associated chaperone molecules redistribute to non-native protein aggregates when present in the cell [120, 144], potentially indicating complementary and plastic roles for these chaperone networks in alleviating proteotoxic burden. In addition to the stabilization, folding, and refolding activity of most chaperones (including stress-induced chaperones), eukaryotic Hsp100's represent a specialized class of chaperones that act on pre-existing protein aggregates, dislodge their monomeric constituents, and pass them to associated Hsp70/Hsp40 co-chaperones for refolding [145, 146] or degradation [147].

The Ubiquitin-Proteasome System (UPS)

Although the protein folding and disaggregase chaperones represent a formidable arsenal of factors that counteract protein misfolding, the other two branches of the PN (the UPS and autophagy) participate in recycling terminally misfolded, damaged, aged, or aggregated proteins. Like the chaperone networks, the UPS consists of a multitude of factors whose coordinated action results in the activation and transfer of post-translational ubiquitin tags which mark proteins for degradation. This coordinated activity is carried out by a cascade of ubiquitin activating enzymes (E1's), ubiquitin conjugating enzymes (E2's), and ubiquitin ligase enzymes (E3's). Furthermore, the number of distinct proteins in each family of enzymes in the cascade increases from the E1 to the E3 families. For example, yeast possess one E1 enzyme, eleven E2 enzymes, and ~60-100 E3 enzymes [148]. The breadth of the E3 enzyme family permits ubiquitin-mediated regulation to play a role in nearly every cellular process, while the diversity of E3 enzymes helps confine ubiquitination responses to particular substrates according to cellular conditions or needs. Finally, although E3 ubiquitin ligases are adept at generating polyubiquitin chains, a small family of additional enzymes (E4's) specialize in catalyzing polyubiquitination only in the presence of pre-ubiquitinated substrates [149].

A number of E3 ubiquitin ligases play important roles in targeting misfolded and aggregation-prone proteins for degradation. As discussed above, the RQC component Ltn1 prevents aggregation of stalled translation products via ubiquitination and degradation [131]. Hul5 and Rsp5, two cytosolic E3 ligases, act in a complementary fashion to ubiquitinate the majority of misfolded cytosolic proteins upon heat stress [150, 151]. Ubiquitination accelerates the degradation of cytosolic Hul5 and Rsp5 substrates. Degradation of Hul5 substrates prevents their aggregation (the solubility status of Rsp5 substrates has not been reported). Furthermore,

recognition of misfolded substrates by Rsp5 is predominantly mediated by an Hsp40 chaperone, Yjd1, an example of cross-talk between branches of the PN.

One particularly well-characterized E3 ligase that counteracts protein aggregation is the predominantly nuclear E3 San1 [152]. San1 can directly recognize and ubiquitinate its misfolded substrates for proteasomal degradation in the nucleus [153]. Subsequent work demonstrated that as few as five exposed hydrophobic residues in San1 substrates was sufficient to recruit San1 for ubiquitination and accelerated degradation [154]. Notably, the features recognized by San1 correlate with substrate insolubility, suggesting that San1 participates specifically in the degradation of aggregation-prone proteins [155]. In the absence of San1, the expression of San1 substrates can be toxic [153] due to the formation of protein aggregates in the nucleus [154]. Although San1 is presently considered the central factor governing degradation of nuclear substrates, it has been shown to cooperate with a variety chaperones for degradation of a subset of substrates [156-158]. Interestingly, although San1 localization is thought to be restricted to the nucleus, cytosolic chaperones can deliver misfolded substrates to the nucleus for San1-mediated degradation [139, 158].

Clearance of Protein Aggregates by Autophagy

Autophagy is a process by which intracellular components are catabolically recycled in a proteasome-independent manner. This occurs through the formation of a specialized membrane-enclosed vesicle known as an autophagosome, which engulfs cellular components for delivery to the yeast vacuole (lysosomes in complex eukaryotes) for digestion by a variety of vacuolar enzymes. While autophagy is best known as a relatively non-specific bulk process in response to

starvation (reviewed in full elsewhere [159-161]), cellular components (notably, protein aggregates) can also be selectively targeted for autophagic destruction.

Selective autophagy of protein aggregates (known also as “aggrephagy”), involves the recruitment of autophagic components to protein aggregates via adaptor proteins. Currently, our understanding of aggrephagy in yeast is limited to a single adaptor protein, Cue5, which predominantly recognizes polyubiquitinated protein aggregates, including polyQ aggregates, and mediates their delivery to autophagosomes via interaction with Atg8 (LC3 in humans) [162]. Many of the Cue5 substrates were polyubiquitinated by the UPS factor Rsp5, suggesting that Rsp5 may act on aggregation-prone proteins even in the absence of heat stress. It is interesting to note that protein aggregates (particularly, amyloid aggregates, including polyQ aggregates) are sometimes delivered to or assembled into a large, perivacuolar deposit known as the “insoluble protein deposit” (IPOD) [163]. The IPOD also colocalizes with Atg8, a well-characterized marker of developing autophagosomes. However, transfer of IPOD components to the vacuole for degradation has not been reported, and it was suggested that the IPOD represents a sequestration site for terminally misfolded proteins [163]. Given the more recent observation that polyQ proteins are degraded by autophagy, it is reasonable to speculate that the IPOD may represent a centralized location by which components are delivered to the vacuole in a Cue5-dependent manner.

Currently, aggrephagy is better understood in higher eukaryotes. The human homolog of Cue5, Tollip, mediates a similar aggrephagy response in mammalian cells, by acting as an adaptor between polyubiquitinated protein aggregates and the autophagophore component LC3 [162]. Additional aggrephagy adaptor proteins include p62/SQSTM1, NBR1, Alfy and optineurin. p62, NBR1, and optineurin contain a domain capable of interacting with ubiquitin

and an LC3-interaction region (LIR), which allows them to serve as bridging factors between ubiquitinated aggregates and autophagosomes [164, 165]. p62 has been identified as a common component of ubiquitin-positive inclusions in a variety of neurodegenerative disorders [166-171]. p62 aggrephagy activity results in the clearance of Htt/polyQ aggregates and reduced cytotoxicity associated with Htt expression [172]. Additional evidence suggests that Alfy acts as a bridging factor between p62-associated polyQ aggregates and developing autophagosomes [173]. Furthermore, NBR1 was shown to cooperate with p62 in the removal of puromycin-induced inclusions [174]. Mutations in p62 and optineurin have been linked to a variety of disorders, including ALS [175-179]. Sph1 also plays a role in the selective autophagy of Lewy body aggregates [180], although the mechanism (and possibly specificity) of the Sph1 pathway appears to differ from p62-, NBR1-, and optineurin-mediated aggrephagy.

CONCLUSIONS

Significant progress has been made in defining the sequence features that drive yeast prion formation and in predicting the prion propensity of PrLDs. However, perfecting prion prediction will require overcoming a number of challenges. Translating results from yeast PFDs into methods to predict aggregation and toxicity of human PrLDs creates additional challenges due to the differences between the two systems. Collectively, these issues highlight the need for additional research to unveil the fundamental features of prion formation, propagation, and proteostatic regulation, as well as how prion-like activity relates to disease. As our knowledge of these fundamental features grows, application of this knowledge to prion prediction will lead to more accurate prediction methods and identification of new prions or prion-like proteins, potentially resulting in additional targets for treating human neurodegenerative disorders.

REFERENCES

1. Chiti, F. and C.M. Dobson, *Protein misfolding, functional amyloid, and human disease.* Annu Rev Biochem, 2006. **75**: p. 333-66.
2. Alberti, S., et al., *A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins.* Cell, 2009. **137**(1): p. 146-158.
3. Derkatch, I.L., et al., *Prions affect the appearance of other prions: the story of [PIN(+)].* Cell, 2001. **106**(2): p. 171-82.
4. Du, Z., et al., *Newly identified prion linked to the chromatin-remodeling factor Swi1 in Saccharomyces cerevisiae.* Nat Genet, 2008. **40**(4): p. 460-5.
5. Patel, B.K., J. Gavin-Smyth, and S.W. Liebman, *The yeast global transcriptional co-repressor protein Cyc8 can propagate as a prion.* Nat Cell Biol, 2009. **11**(3): p. 344-9.
6. Rogoza, T., et al., *Non-Mendelian determinant [ISP+] in yeast is a nuclear-residing prion form of the global transcriptional regulator Sfp1.* Proc. Natl. Acad. Sci. USA, 2010. **107**(23): p. 10573-7.
7. Sondheimer, N. and S. Lindquist, *Rnq1: an epigenetic modifier of protein function in yeast.* Mol. Cell, 2000. **5**(1): p. 163-72.
8. Suzuki, G., N. Shimazu, and M. Tanaka, *A Yeast Prion, Mod5, Promotes Acquired Drug Resistance and Cell Survival Under Environmental Stress.* Science, 2012. **336**(6079): p. 355-359.
9. Wickner, R.B., *[URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae.* Science, 1994. **264**(5158): p. 566-9.
10. Halfmann, R., et al., *Prion formation by a yeast GLFG nucleoporin.* Prion, 2012. **6**(4).
11. Fowler, D.M., et al., *Functional amyloid--from bacteria to humans.* Trends Biochem Sci, 2007. **32**(5): p. 217-24.
12. Eaglestone, S.S., B.S. Cox, and M.F. Tuite, *Translation termination efficiency can be regulated in Saccharomyces cerevisiae by environmental stress through a prion-mediated mechanism.* EMBO J, 1999. **18**(7): p. 1974-81.
13. True, H.L. and S.L. Lindquist, *A yeast prion provides a mechanism for genetic variation and phenotypic diversity.* Nature, 2000. **407**(6803): p. 477-83.
14. Tyedmers, J., M.L. Madariaga, and S. Lindquist, *Prion switching in response to environmental stress.* PLoS Biol, 2008. **6**(11): p. e294.
15. Bateman, D.A. and R.B. Wickner, *[PSI+] Prion transmission barriers protect Saccharomyces cerevisiae from infection: intraspecies 'species barriers'.* Genetics, 2012. **190**(2): p. 569-79.
16. Kelly, A.C. and R.B. Wickner, *Saccharomyces cerevisiae: a sexy yeast with a prion problem.* Prion, 2013. **7**(3): p. 215-20.
17. Nakayashiki, T., et al., *Yeast prions [URE3] and [PSI+] are diseases.* Proc Natl Acad Sci U S A, 2005. **102**(30): p. 10575-80.
18. Wickner, R.B., et al., *The yeast prions [PSI+] and [URE3] are molecular degenerative diseases.* Prion, 2011. **5**(4): p. 258-62.
19. Cox, B.S., *PSI, a cytoplasmic suppressor of super-suppressor in yeast.* Heredity, 1965. **26**: p. 211-232.
20. Lacroute, F., *Non-Mendelian mutation allowing ureidosuccinic acid uptake in yeast.* J. Bacteriol., 1971. **106**(2): p. 519-22.

21. Ter-Avanesyan, M.D., et al., *The SUP35 omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant [psi+] in the yeast Saccharomyces cerevisiae.* Genetics, 1994. **137**(3): p. 671-6.
22. Maddelein, M.L. and R.B. Wickner, *Two prion-inducing regions of Ure2p are nonoverlapping.* Mol. Cell. Biol., 1999. **19**(6): p. 4516-24.
23. Doel, S.M., et al., *The dominant PNM2- mutation which eliminates the psi factor of Saccharomyces cerevisiae is the result of a missense mutation in the SUP35 gene.* Genetics, 1994. **137**(3): p. 659-70.
24. Kushnirov, V.V., et al., *Nucleotide sequence of the SUP2 (SUP35) gene of Saccharomyces cerevisiae.* Gene, 1988. **66**(1): p. 45-54.
25. Masison, D.C. and R.B. Wickner, *Prion-inducing domain of yeast Ure2p and protease resistance of Ure2p in prion-containing cells.* Science, 1995. **270**(5233): p. 93-5.
26. Ter-Avanesyan, M.D., et al., *Deletion analysis of the SUP35 gene of the yeast Saccharomyces cerevisiae reveals two non-overlapping functional regions in the encoded protein.* Mol. Microbiol., 1993. **7**(5): p. 683-92.
27. Liu, J.J., N. Sondheimer, and S.L. Lindquist, *Changes in the middle region of Sup35 profoundly alter the nature of epigenetic inheritance for the yeast prion [PSI+].* Proc. Natl. Acad. Sci. USA, 2002. **99 Suppl 4**: p. 16446-53.
28. Baxa, U., et al., *Mechanism of inactivation on prion conversion of the Saccharomyces cerevisiae Ure2 protein.* Proc. Natl. Acad. Sci. USA, 2002. **99**(8): p. 5253-60.
29. Li, L. and S. Lindquist, *Creating a protein-based element of inheritance.* Science, 2000. **287**(5453): p. 661-4.
30. Pierce, M.M., et al., *Is the prion domain of soluble Ure2p unstructured?* Biochemistry, 2005. **44**(1): p. 321-8.
31. Serio, T.R., et al., *Nucleated conformational conversion and the replication of conformational information by a prion determinant.* Science, 2000. **289**(5483): p. 1317-21.
32. Ross, E.D., U. Baxa, and R.B. Wickner, *Scrambled Prion Domains Form Prions and Amyloid.* Mol. Cell. Biol., 2004. **24**(16): p. 7206-7213.
33. Ross, E.D., et al., *Primary sequence independence for prion formation.* Proc. Natl. Acad. Sci. USA, 2005. **102**(36): p. 12825-12830.
34. Du, Z., *The complexity and implications of yeast prion domains.* Prion, 2011. **5**(4): p. 311-6.
35. Maclea, K.S. and E.D. Ross, *Strategies for identifying new prions in yeast.* Prion, 2011. **5**(4): p. 263-268.
36. King, O.D., A.D. Gitler, and J. Shorter, *The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease.* Brain Res, 2012: p. Epub ahead of print.
37. DePace, A.H., et al., *A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion.* Cell, 1998. **93**(7): p. 1241-52.
38. Santoso, A., et al., *Molecular basis of a yeast prion species barrier.* Cell, 2000. **100**(2): p. 277-88.
39. Chen, B., et al., *Genetic and epigenetic control of the efficiency and fidelity of cross-species prion transmission.* Mol Microbiol, 2010. **76**(6): p. 1483-99.
40. Osherovich, L.Z., et al., *Dissection and design of yeast prions.* PLoS Biol., 2004. **2**(4): p. E86.

41. Parham, S.N., C.G. Resende, and M.F. Tuite, *Oligopeptide repeats in the yeast protein Sup35p stabilize intermolecular prion interactions.* EMBO J., 2001. **20**(9): p. 2111-9.
42. Shkundina, I.S., et al., *The role of the N-terminal oligopeptide repeats of the yeast sup35 prion protein in propagation and transmission of prion variants.* Genetics, 2006. **172**(2): p. 827-35.
43. Nelson, R., et al., *Structure of the cross-beta spine of amyloid-like fibrils.* Nature, 2005. **435**(7043): p. 773-8.
44. Fei, L. and S. Perrett, *Disulfide bond formation significantly accelerates the assembly of Ure2p fibrils because of the proximity of a potential amyloid stretch.* J Biol Chem, 2009. **284**(17): p. 11134-41.
45. Tessier, P.M. and S. Lindquist, *Prion recognition elements govern nucleation, strain specificity and species barriers.* Nature, 2007. **447**(7144): p. 556-61. Epub 2007 May 9.
46. Crow, E.T., Z. Du, and L. Li, *A small, glutamine-free domain propagates the [SWI(+)] prion in budding yeast.* Mol Cell Biol, 2011. **31**(16): p. 3436-44.
47. Vitrenko, Y.A., et al., *Propagation of the [PIN+] prion by fragments of Rnq1 fused to GFP.* Curr Genet, 2007. **51**(5): p. 309-19. Epub 2007 Apr 6.
48. Kadnar, M.L., G. Articov, and I.L. Derkatch, *Distinct type of transmission barrier revealed by study of multiple prion determinants of Rnq1.* PLoS, 2010. **6**(1): p. e1000824.
49. Osherovich, L.Z. and J.S. Weissman, *Multiple Gln/Asn-rich prion domains confer susceptibility to induction of the yeast [PSI(+)] prion.* Cell, 2001. **106**(2): p. 183-94.
50. Kabani, M., et al., *A mutation within the C-terminal domain of Sup35p that affects [PSI(+)] prion propagation.* Mol Microbiol., 2011: p. Epub ahead of print.
51. Kochneva-Pervukhova, N.V., et al., *C-terminal truncation of the Sup35 protein increases the frequency of de novo generation of a prion-based [PSI+] determinant in Saccharomyces cerevisiae.* Curr. Genet., 1998. **34**(2): p. 146-51.
52. Helsen, C.W. and J.R. Glover, *Insight into molecular basis of curing of [PSI+] prion by overexpression of 104-kDa heat shock protein (Hsp104).* J Biol Chem, 2012. **287**(1): p. 542-56.
53. Shewmaker, F., R.B. Wickner, and R. Tycko, *Amyloid of the prion domain of Sup35p has an in-register parallel beta-sheet structure.* Proc Natl Acad Sci U S A, 2006. **103**(52): p. 19754-9.
54. Bryan, A.W., Jr., et al., *BETASCAN: probable beta-amyloids identified by pairwise probabilistic analysis.* PLoS Comput Biol, 2009. **5**(3): p. e1000333.
55. Bryan, A.W., Jr., et al., *STITCHER: Dynamic assembly of likely amyloid and prion beta-structures from secondary structure predictions.* Proteins, 2011.
56. Tartaglia, G.G., et al., *Prediction of aggregation-prone regions in structured proteins.* J Mol Biol, 2008. **380**(2): p. 425-36.
57. Goldschmidt, L., et al., *Identifying the amylome, proteins capable of forming amyloid-like fibrils.* Proc. Natl. Acad. Sci. USA 2010. **107**(8): p. 3487-3492.
58. Fernandez-Escamilla, A.M., et al., *Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins.* Nat. Biotechnol., 2004. **22**(10): p. 1302-6.
59. Zibaee, S., et al., *A simple algorithm locates beta-strands in the amyloid fibril core of alpha-synuclein, Abeta, and tau using the amino acid sequence alone.* Protein Sci, 2007. **16**(5): p. 906-18.

60. Trovato, A., F. Seno, and S.C. Tosatto, *The PASTA server for protein aggregation prediction.* Protein Eng Des Sel, 2007. **20**(10): p. 521-3.
61. Maurer-Stroh, S., et al., *Exploring the sequence determinants of amyloid structure using position-specific scoring matrices.* Nature Methods, 2010. **7**(3): p. 237-42.
62. Toombs, J.A., et al., *De novo design of synthetic prion domains.* Proc Natl Acad Sci U S A, 2012. **109**(17): p. 6519-6524.
63. Esteras-Chopo, A., L. Serrano, and M.L. de la Paz, *The amyloid stretch hypothesis: Recruiting proteins toward the dark side.* Proc. Natl. Acad. Sci. USA, 2005. **102**(46): p. 16672-16677.
64. Toombs, J.A., B.R. McCarty, and E.D. Ross, *Compositional determinants of prion formation in yeast.* Mol. Cell. Biol., 2010. **30**(1): p. 319-332.
65. Thompson, M.J., et al., *The 3D profile method for identifying fibril-forming segments of proteins.* Proc Natl Acad Sci U S A, 2006. **103**(11): p. 4074-8.
66. Teng, P.K. and D. Eisenberg, *Short protein segments can drive a non-fibrillizing protein into the amyloid state.* Protein Eng Des Sel, 2009. **22**(8): p. 531-6.
67. Sabate, R., et al., *What makes a protein sequence a prion?* PLoS Comput Biol, 2015. **11**(1): p. e1004013.
68. Zambrano, R., et al., *PrionW: a server to identify proteins containing glutamine/asparagine rich prion-like domains and their amyloid cores.* Nucleic Acids Res, 2015. **43**(W1): p. W331-7.
69. Afsar Minhas F.U.A., E.D. Ross, and A. Ben-Hur, *Amino acid composition predicts prion activivty*. PLoS Comp Biol, 2017. **13**(4): p. e1005465.
70. Harrison, P.M. and M. Gerstein, *A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes.* Genome Biol, 2003. **4**(6): p. R40.
71. Michelitsch, M.D. and J.S. Weissman, *A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions.* Proc. Natl. Acad. Sci. USA, 2000. **97**(22): p. 11910-5.
72. Tartaglia, G.G., et al., *Life on the edge: a link between gene expression levels and aggregation rates of human proteins.* Trends Biochem Sci, 2007. **32**(5): p. 204-6.
73. Prilusky, J., et al., *FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded.* Bioinformatics, 2005. **21**(16): p. 3435-8.
74. Espinosa Angarica, V., S. Ventura, and J. Sancho, *Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains.* BMC Genomics, 2013. **14**: p. 316.
75. Li, S.C., et al., *Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment.* Proc Natl Acad Sci U S A, 1996. **93**(13): p. 6676-81.
76. Toombs, J.A., et al., *[PSI+] maintenance is dependent on the composition, not primary sequence, of the oligopeptide repeat domain. PLoS One,* 2011. **6**(7): p. e21953.
77. MacLea, K.S., et al., *Distinct amino acid compositional requirements for formation and maintenance of the [PSI(+)] prion in yeast.* Mol Cell Biol, 2015. **35**(5): p. 899-911.
78. Halfmann, R., et al., *Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins.* Mol Cell, 2011. **43**(1): p. 72-84.
79. Costanzo, M. and C. Zurzolo, *The cell biology of prion-like spread of protein aggregates: mechanisms and implication in neurodegeneration.* Biochem J, 2013. **452**(1): p. 1-17.

80. Cushman, M., et al., *Prion-like disorders: blurring the divide between transmissibility and infectivity.* J Cell Sci, 2010. **123**(Pt 8): p. 1191-201.
81. Hall, G.F. and B.A. Patuto, *Is tau ready for admission to the prion club?* Prion, 2012. **6**(3): p. 223-33.
82. Aguzzi, A. and L. Rajendran, *The transcellular spread of cytosolic amyloids, prions, and prionoids.* Neuron, 2009. **64**(6): p. 783-90.
83. Clavaguera, F., et al., *Transmission and spreading of tauopathy in transgenic mouse brain.* Nat Cell Biol, 2009. **11**(7): p. 909-13.
84. Couthouis, J., et al., *A yeast functional screen predicts new candidate ALS disease genes.* Proc Natl Acad Sci U S A, 2011. **108**: p. 20881-20890.
85. Neumann, M., et al., *Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis.* Science, 2006. **314**(5796): p. 130-3.
86. Lagier-Tourenne, C., M. Polymenidou, and D.W. Cleveland, *TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration.* Hum Mol Genet, 2010. **19**(R1): p. R46-64.
87. Ash, P.E., et al., *Neurotoxic effects of TDP-43 overexpression in C. elegans.* Hum Mol Genet, 2010. **19**(16): p. 3206-18.
88. Johnson, B.S., et al., *A yeast TDP-43 proteinopathy model: Exploring the molecular determinants of TDP-43 aggregation and cellular toxicity.* Proc Natl Acad Sci U S A, 2008. **105**(17): p. 6439-44.
89. Li, Y., et al., *A Drosophila model for TDP-43 proteinopathy.* Proc Natl Acad Sci U S A, 2010. **107**(7): p. 3169-74.
90. Da Cruz, S. and D.W. Cleveland, *Understanding the role of TDP-43 and FUS/TLS in ALS and beyond.* Curr Opin Neurobiol, 2011. **21**(6): p. 904-19.
91. Johnson, B.S., et al., *TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity.* J Biol Chem, 2009. **284**(30): p. 20329-39.
92. Sun, Z., et al., *Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS.* PLoS Biol, 2011. **9**(4): p. e1000614.
93. Kwiatkowski, T.J., Jr., et al., *Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis.* Science, 2009. **323**(5918): p. 1205-8.
94. Vance, C., et al., *Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6.* Science, 2009. **323**(5918): p. 1208-11.
95. Daigle, J.G., et al., *RNA-binding ability of FUS regulates neurodegeneration, cytoplasmic mislocalization and incorporation into stress granules associated with FUS carrying ALS-linked mutations.* Hum Mol Genet, 2013. **22**(6): p. 1193-205.
96. Couthouis, J., et al., *Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis.* Hum Mol Genet, 2012: p. Epub ahead of print.
97. Kim, H.J., et al., *Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS.* Nature, 2013. **495**(7442): p. 467-73.
98. Klar, J., et al., *Welander distal myopathy caused by an ancient founder mutation in TIA1 associated with perturbed splicing.* Hum Mutat, 2013. **34**(4): p. 572-7.
99. Li, Y.R., et al., *Stress granules as crucibles of ALS pathogenesis.* J Cell Biol, 2013. **201**(3): p. 361-72.
100. Wolozin, B., *Regulated protein aggregation: stress granules and neurodegeneration.* Mol Neurodegener, 2012. **7**: p. 56.

101. Gilks, N., et al., *Stress granule assembly is mediated by prion-like aggregation of TIA-1.* Mol Biol Cell, 2004. **15**(12): p. 5383-98.
102. Kroschwald, S., et al., *Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules.* Elife, 2015. **4**: p. e06807.
103. Lin, Y., et al., *Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins.* Mol Cell, 2015. **60**(2): p. 208-19.
104. Molliex, A., et al., *Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization.* Cell, 2015. **163**(1): p. 123-33.
105. Patel, A., et al., *A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation.* Cell, 2015. **162**(5): p. 1066-77.
106. Riback, J.A., et al., *Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response.* Cell, 2017. **168**(6): p. 1028-1040 e19.
107. Bergeron-Sandoval, L.P., N. Safaee, and S.W. Michnick, *Mechanisms and Consequences of Macromolecular Phase Separation.* Cell, 2016. **165**(5): p. 1067-1079.
108. Brangwynne, C.P., P. Tompa, and R.V. Pappu, *Polymer physics of intracellular phase transitions.* Nature Physics, 2015. **11**(11): p. 899-904.
109. Hyman, A.A., C.A. Weber, and F. Julicher, *Liquid-liquid phase separation in biology.* Annu Rev Cell Dev Biol, 2014. **30**: p. 39-58.
110. Ramaswami, M., J.P. Taylor, and R. Parker, *Altered Ribostasis: RNA-Protein Granules in Degenerative Disorders.* Cell, 2013. **154**(4): p. 727-36.
111. Yamanaka, K., Y. Sasagawa, and T. Ogura, *Recent advances in p97/VCP/Cdc48 cellular functions.* Biochim Biophys Acta, 2012. **1823**(1): p. 130-7.
112. Watts, G.D., et al., *Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein.* Nat Genet, 2004. **36**(4): p. 377-81.
113. Johnson, J.O., et al., *Exome sequencing reveals VCP mutations as a cause of familial ALS.* Neuron, 2010. **68**(5): p. 857-64.
114. Buchan, J.R., et al., *Eukaryotic stress granules are cleared by autophagy and Cdc48/VCP function.* Cell, 2013. **153**(7): p. 1461-74.
115. Alexandrov, A.I., et al., *The effects of amino acid composition of glutamine-rich domains on amyloid formation and fragmentation.* PLoS ONE, 2012. **7**(10): p. e46458.
116. Shorter, J., *The mammalian disaggregase machinery: Hsp110 synergizes with Hsp70 and Hsp40 to catalyze protein disaggregation and reactivation in a cell-free system.* PLoS ONE, 2011. **6**(10): p. e26319.
117. Labbadia, J. and R.I. Morimoto, *The biology of proteostasis in aging and disease.* Annu Rev Biochem, 2015. **84**: p. 435-64.
118. Hartl, F.U., A. Bracher, and M. Hayer-Hartl, *Molecular chaperones in protein folding and proteostasis.* Nature, 2011. **475**(7356): p. 324-32.
119. Ibstedt, S., et al., *Global analysis of protein aggregation in yeast during physiological conditions and arsenite stress.* Biol Open, 2014. **3**(10): p. 913-23.
120. Willmund, F., et al., *The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis.* Cell, 2013. **152**(1-2): p. 196-209.
121. Ellis, R.J. and A.P. Minton, *Protein aggregation in crowded environments.* Biol Chem, 2006. **387**(5): p. 485-97.

122. Zhou, H.X., *Influence of crowded cellular environments on protein folding, binding, and oligomerization: biological consequences and potentials of atomistic modeling.* FEBS Lett, 2013. **587**(8): p. 1053-61.
123. Gautschi, M., et al., *RAC, a stable ribosome-associated complex in yeast formed by the DnaK-DnaJ homologs Ssz1p and zuotin.* Proc Natl Acad Sci U S A, 2001. **98**(7): p. 3762-7.
124. Koplin, A., et al., *A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated complex on ribosomes.* J Cell Biol, 2010. **189**(1): p. 57-68.
125. del Alamo, M., et al., *Defining the specificity of cotranslationally acting chaperones by systematic analysis of mRNAs associated with ribosome-nascent chain complexes.* PLoS Biol, 2011. **9**(7): p. e1001100.
126. Nyathi, Y. and M.R. Pool, *Analysis of the interplay of protein biogenesis factors at the ribosome exit site reveals new role for NAC.* J Cell Biol, 2015. **210**(2): p. 287-301.
127. Brandman, O., et al., *A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress.* Cell, 2012. **151**(5): p. 1042-54.
128. Maurer, M.J., et al., *Degradation Signals for Ubiquitin-Proteasome Dependent Cytosolic Protein Quality Control (CytoQC) in Yeast.* G3 (Bethesda), 2016. **6**(7): p. 1853-66.
129. Bengtson, M.H. and C.A. Joazeiro, *Role of a ribosome-associated E3 ubiquitin ligase in protein quality control.* Nature, 2010. **467**(7314): p. 470-3.
130. Duttler, S., S. Pechmann, and J. Frydman, *Principles of cotranslational ubiquitination and quality control at the ribosome.* Mol Cell, 2013. **50**(3): p. 379-93.
131. Choe, Y.J., et al., *Failure of RQC machinery causes protein aggregation and proteotoxic stress.* Nature, 2016. **531**(7593): p. 191-5.
132. Chu, J., et al., *A mouse forward genetics screen identifies LISTERIN as an E3 ubiquitin ligase involved in neurodegeneration.* Proc Natl Acad Sci U S A, 2009. **106**(7): p. 2097-103.
133. McCarty, J.S., et al., *The role of ATP in the functional cycle of the DnaK chaperone system.* J Mol Biol, 1995. **249**(1): p. 126-37.
134. Schmid, D., et al., *Kinetics of molecular chaperone action.* Science, 1994. **263**(5149): p. 971-3.
135. Szabo, A., et al., *The ATP hydrolysis-dependent reaction cycle of the Escherichia coli Hsp70 system DnaK, DnaJ, and GrpE.* Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10345-9.
136. Flynn, G.C., et al., *Peptide-binding specificity of the molecular chaperone BiP.* Nature, 1991. **353**(6346): p. 726-30.
137. Rudiger, S., et al., *Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries.* EMBO J, 1997. **16**(7): p. 1501-7.
138. Gragerov, A., et al., *Specificity of DnaK-peptide binding.* J Mol Biol, 1994. **235**(3): p. 848-54.
139. Park, S.H., et al., *PolyQ proteins interfere with nuclear degradation of cytosolic proteins by sequestering the Sis1p chaperone.* Cell, 2013. **154**(1): p. 134-45.
140. Melville, M.W., et al., *The Hsp70 and TRiC/CCT chaperone systems cooperate in vivo to assemble the von Hippel-Lindau tumor suppressor complex.* Mol Cell Biol, 2003. **23**(9): p. 3141-51.
141. Wegele, H., et al., *Substrate transfer from the chaperone Hsp70 to Hsp90.* J Mol Biol, 2006. **356**(3): p. 802-11.

142. Balchin, D., M. Hayer-Hartl, and F.U. Hartl, *In vivo aspects of protein folding and quality control.* Science, 2016. **353**(6294): p. aac4354.
143. Albanese, V., et al., *Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells.* Cell, 2006. **124**(1): p. 75-88.
144. Olzscha, H., et al., *Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions.* Cell, 2011. **144**(1): p. 67-78.
145. Glover, J.R. and S. Lindquist, *Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins.* Cell, 1998. **94**(1): p. 73-82.
146. Mattoo, R.U., et al., *Hsp110 is a bona fide chaperone using ATP to unfold stable misfolded polypeptides and reciprocally collaborate with Hsp70 to solubilize protein aggregates.* J Biol Chem, 2013. **288**(29): p. 21399-411.
147. Lee, D.H. and A.L. Goldberg, *Hsp104 is essential for the selective degradation in yeast of polyglutamine expanded ataxin-1 but not most misfolded proteins generally.* Biochem Biophys Res Commun, 2010. **391**(1): p. 1056-61.
148. Finley, D., et al., *The ubiquitin-proteasome system of Saccharomyces cerevisiae.* Genetics, 2012. **192**(2): p. 319-60.
149. Hoppe, T., *Multiubiquitylation by E4 enzymes: 'one size' doesn't fit all.* Trends Biochem Sci, 2005. **30**(4): p. 183-7.
150. Fang, N.N., et al., *Rsp5/Nedd4 is the main ubiquitin ligase that targets cytosolic misfolded proteins following heat stress.* Nat Cell Biol, 2014. **16**(12): p. 1227-37.
151. Fang, N.N., et al., *Hul5 HECT ubiquitin ligase plays a major role in the ubiquitylation and turnover of cytosolic misfolded proteins.* Nat Cell Biol, 2011. **13**(11): p. 1344-52.
152. Gardner, R.G., Z.W. Nelson, and D.E. Gottschling, *Degradation-mediated protein quality control in the nucleus.* Cell, 2005. **120**(6): p. 803-15.
153. Rosenbaum, J.C., et al., *Disorder targets misorder in nuclear quality control degradation: a disordered ubiquitin ligase directly recognizes its misfolded substrates.* Mol Cell, 2011. **41**(1): p. 93-106.
154. Fredrickson, E.K., et al., *Exposed hydrophobicity is a key determinant of nuclear quality control degradation.* Mol Biol Cell, 2011. **22**(13): p. 2384-95.
155. Fredrickson, E.K., et al., *Substrate recognition in nuclear protein quality control degradation is governed by exposed hydrophobicity that correlates with aggregation and insolubility.* J Biol Chem, 2013. **288**(9): p. 6130-9.
156. Gallagher, P.S., S.V. Clowes Candadai, and R.G. Gardner, *The requirement for Cdc48/p97 in nuclear protein quality control degradation depends on the substrate and correlates with substrate insolubility.* J Cell Sci, 2014. **127**(Pt 9): p. 1980-91.
157. Guerriero, C.J., K.F. Weiberth, and J.L. Brodsky, *Hsp70 targets a cytoplasmic quality control substrate to the San1p ubiquitin ligase.* J Biol Chem, 2013. **288**(25): p. 18506-20.
158. Heck, J.W., S.K. Cheung, and R.Y. Hampton, *Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1.* Proc Natl Acad Sci U S A, 2010. **107**(3): p. 1106-11.
159. Bento, C.F., et al., *Mammalian Autophagy: How Does It Work?* Annu Rev Biochem, 2016. **85**: p. 685-713.
160. He, C. and D.J. Klionsky, *Regulation mechanisms and signaling pathways of autophagy.* Annu Rev Genet, 2009. **43**: p. 67-93.
161. Kaur, J. and J. Debnath, *Autophagy at the crossroads of catabolism and anabolism.* Nat Rev Mol Cell Biol, 2015. **16**(8): p. 461-72.

162. Lu, K., I. Psakhye, and S. Jentsch, *Autophagic clearance of polyQ proteins mediated by ubiquitin-Atg8 adaptors of the conserved CUET protein family.* Cell, 2014. **158**(3): p. 549-63.
163. Kaganovich, D., R. Kopito, and J. Frydman, *Misfolded proteins partition between two distinct quality control compartments.* Nature, 2008. **454**(7208): p. 1088-95.
164. Kraft, C., M. Peter, and K. Hofmann, *Selective autophagy: ubiquitin-mediated recognition and beyond.* Nat Cell Biol, 2010. **12**(9): p. 836-41.
165. Ying, H. and B.Y. Yue, *Optineurin: The autophagy connection.* Exp Eye Res, 2016. **144**: p. 73-80.
166. Gal, J., et al., *p62 accumulates and enhances aggregate formation in model systems of familial amyotrophic lateral sclerosis.* J Biol Chem, 2007. **282**(15): p. 11068-77.
167. Homma, T., et al., *Increased expression of p62/SQSTM1 in prion diseases and its association with pathogenic prion protein.* Sci Rep, 2014. **4**: p. 4504.
168. Komatsu, M., et al., *Homeostatic levels of p62 control cytoplasmic inclusion body formation in autophagy-deficient mice.* Cell, 2007. **131**(6): p. 1149-63.
169. Mann, D.M., et al., *Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72.* Acta Neuropathol Commun, 2013. **1**: p. 68.
170. Watanabe, Y., et al., *p62/SQSTM1-dependent autophagy of Lewy body-like alpha-synuclein inclusions.* PLoS One, 2012. **7**(12): p. e52868.
171. Zatloukal, K., et al., *p62 Is a common component of cytoplasmic inclusions in protein aggregation diseases.* Am J Pathol, 2002. **160**(1): p. 255-63.
172. Bjorkoy, G., et al., *p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death.* J Cell Biol, 2005. **171**(4): p. 603-14.
173. Filimonenko, M., et al., *The selective macroautophagic degradation of aggregated proteins requires the PI3P-binding protein Alfy.* Mol Cell, 2010. **38**(2): p. 265-79.
174. Kirkin, V., et al., *A role for NBR1 in autophagosomal degradation of ubiquitinated substrates.* Mol Cell, 2009. **33**(4): p. 505-16.
175. Fecto, F., et al., *SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis.* Arch Neurol, 2011. **68**(11): p. 1440-6.
176. Goode, A., et al., *Defective recognition of LC3B by mutant SQSTM1/p62 implicates impairment of autophagy as a pathogenic mechanism in ALS-FTLD.* Autophagy, 2016. **12**(7): p. 1094-104.
177. Layfield, R., et al., *p62 mutations, ubiquitin recognition and Paget's disease of bone.* Biochem Soc Trans, 2006. **34**(Pt 5): p. 735-7.
178. Teyssou, E., et al., *Mutations in SQSTM1 encoding p62 in amyotrophic lateral sclerosis: genetics and neuropathology.* Acta Neuropathol, 2013. **125**(4): p. 511-22.
179. Maruyama, H., et al., *Mutations of optineurin in amyotrophic lateral sclerosis.* Nature, 2010. **465**(7295): p. 223-6.
180. Wong, E., et al., *Molecular determinants of selective clearance of protein inclusions by autophagy.* Nat Commun, 2012. **3**: p. 1240.

CHAPTER 2: THE EFFECTS OF MUTATIONS ON THE AGGREGATION PROPENSITY OF THE HUMAN PRION-LIKE PROTEIN HNRNPA2B1[2]

INTRODUCTION

Amyloid fibrils are ordered, self-propagating, β-sheet-rich protein aggregates [1, 2]. In Saccharomyces cerevisiae, numerous prions (infectious proteins) have been identified that result from the conversion of proteins to an infectious amyloid form [3, 4]. Most of the yeast prion proteins contain low-complexity, glutamine/asparagine (Q/N) rich prion domains [5]. Hundreds of human proteins contain similar prion-like domains (PrLDs), defined as protein segments that compositionally resemble yeast prion domains [6, 7]. PrLDs are a subset of low complexity sequence domains (LCDs) that are found in about one third of the human proteome, and which are generally predicted to be intrinsically disordered [7, 8]. PrLDs are particularly enriched in RNA-binding proteins [7]. Mutations in various PrLD-containing RNA-binding proteins have been linked to degenerative disorders, including amyotrophic lateral sclerosis (ALS) and frontotemporal dementia [7, 9].

A number of these PrLD-containing RNA binding proteins are components of RNA-protein granules, such as P-bodies and stress granules [9, 10], and the PrLDs are thought to mediate interactions that are involved in the formation of these granules [11-13]. These PrLD-containing RNA binding proteins can form a range of assemblies, which differ in the degree of order in the structure, and possibly in the nature of the underlying interactions [14-17]. These range from highly dynamic liquid-liquid phase separations, in which liquid droplets are formed

[2] This chapter has been reformatted from the following publication: Paul KR, Molliex A, Cascarina S, Boncella AE, Taylor JP, Ross ED. Mol. Cell. Biol. (2017) 37:e00652-16. My contribution consisted of mutational design, building, and testing of mutations at the Y283 position within the hnRNPA2 PrLD, as well as development and implementation of all western blot assays examining protein expression levels in yeast.

in a temperature- and concentration-dependent manner [15-17]; to hydrogels, consisting of metastable amyloid-like fibers [14]; to more stable, ordered amyloid aggregates [16, 17]. Disease-associated mutations appear to specifically shift these proteins towards the amyloid state [16-18]. This raises the intriguing hypothesis that these PrLDs are evolved to mediate the weak, dynamic interactions involved in formation of dynamic RNA-protein granules, but disease-associated mutations promote conversion of the PrLDs to more stable structures [9, 10, 19].

The human heterogeneous nuclear ribonucleoprotein hnRNPA2B1 provides a useful model to examine this hypothesis. hnRNPA2B1 is a ubiquitously expressed RNA binding protein that has two alternatively spliced forms, A2 and B1, which differ by 12 amino acids at the N-terminus. The shorter hnRNPA2 is the predominant isoform in most tissues. hnRNPA2B1 contains a PrLD (Figure 2.1A), and a single point mutation (D290V in hnRNPA2) in this PrLD causes multisystem proteinopathy [18]. Interestingly, mutations at the corresponding position of a paralogous heterogeneous nuclear ribonucleoprotein, hnRNPA1, can cause either multisystem proteinopathy or familial ALS [18]. The mutations in both proteins promote incorporation into stress granules, and in *Drosophila* cause formation of cytoplasmic inclusions. In vitro, the mutations accelerate formation of amyloid fibrils. In yeast, the core prion-like domain is able to support prion formation when inserted in the place of the portion of the prion domain of Sup35 that is responsible for nucleating prion formation [18, 20]. Thus, hnRNPA2 provides a range of experimental systems to monitor the effects of mutations on protein aggregation.
Intriguingly, PAPA and ZipperDB, two algorithms designed to predict amyloid or prion propensity, both correctly predict the effects of the three known disease-associated mutations in hnRNPA2B1 and hnRNPA1 [18]. These results suggest that it might be possible to predict the effects of other mutations in these proteins, and to rationally design mutations to alter

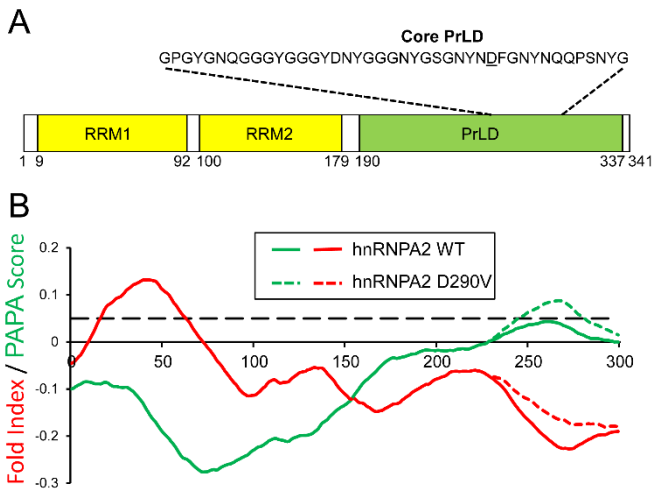


Figure 2.1: hnRNPA2 contains a predicted prion-like domain. (A) Schematic of the hnRNPA2 domain architecture. **(B)** The disease-associated D290V mutation increases predicted prion-propensity. PAPA scores (green) and FoldIndex scores (red) were calculated for hnRNPA2 wild-type (solid) and D290V (dashed). The black dotted line indicates a PAPA score of 0.05, the threshold that was most effective at separating prion-like domains with and without prion activity [21]. Regions with high PAPA scores and negative FoldIndex scores are predicted to be prion-prone. Adapted from Kim et al. [18].

aggregation propensity. However, this prediction success is currently based on a very small sample size: just one mutation in hnRNPA2, and two in hnRNPA1. Additionally, PAPA and ZipperDB use very different features to score aggregation propensity, so it is unclear which of these features is most predictive.

Specifically, PAPA was derived by replacing an 8-amino-acid segment from a scrambled version of Sup35 with a random sequence to build a library of mutants, and then screening this library for prion formation [22]. A prion propensity score was then derived for each amino acid by comparing its frequency among the prion-forming isolates relative to the starting library. PAPA predicts prion activity by first using FoldIndex to identify regions of proteins that are predicted to be intrinsically disordered, and then scanning these regions with a 41-amino acid window size, adding up the prion propensity scores of each amino acid across the window [21, 23]. By contrast ZipperDB is a structure-based algorithm designed to look for short peptide fragments with a high propensity to form steric zippers [24]. ZipperDB was developed by first solving the structure of a 6-amino-acid peptide from Sup35 in its amyloid conformation [25]. The peptide was found to form a cross-β-sheet structure, with tight steric zipper interactions between the sheets. ZipperDB predicts amyloid propensity by threading 6-amino-acid peptides into this structure in silico, and using Rosetta to determine the energetic fit.

Thus, PAPA solely considers amino acid composition, and uses a large window size, while ZipperDB uses a much smaller window, and is sensitive to primary sequence. Despite these differences, both accurately predicted the effects of the hnRNP mutations. PAPA predicts that the aggregation propensity of the wild-type hnRNPA2 PrLD falls just below the threshold for prion-like aggregation, while the mutation increases aggregation propensity well beyond this

threshold (Figure 2.1B). ZipperDB predicts that the disease-associated mutations should create a strong steric zipper [18].

Here, to define the sequence features that drive aggregation, we designed a variety of mutations in the hnRNPA2 prion-like domain. Both in yeast and *in vitro*, the effects of mutations could be predicted entirely based on amino acid composition. By contrast, while the original disease-associated mutations created predicted steric zipper motifs, such motifs were neither necessary nor sufficient for aggregation in yeast. Although composition alone accurately predicted the effects of mutations on isolated prion-like domain fragments both in yeast and *in vitro*, it was less accurate at predicting their effects in the context of the full-length protein in *Drosophila*. This highlights a critical limitation of our current prediction methods. While these methods can predict the effects of mutations on intrinsic aggregation propensity, other factors (including interactions with other parts of the protein, interacting proteins, and localization) are currently much more challenging to predict.

MATERIALS AND METHODS

Yeast Strains and Media

Standard yeast media and methods were as previously described [26]. All experiments were performed in strain YER635/pJ533 ([27]; α *kar1-1 SUQ5 ade2-1 his3 leu2 trp1 ura3 ppq1::HIS3 sup35*::KanMx). pJ533 (*URA3*) expresses *SUP35* from the *SUP35* promoter. Yeast were grown at 30°C for all experiments.

Prion Formation in Yeast

Plasmids pER599 and pER600 (cen, *LEU2*) expressing wild-type and D290V hnRNPA2-Sup35 fusions, respectively, were previously described [18]. All additional mutations were made by PCR, and confirmed by DNA sequencing. Plasmids were transformed into YER635/pJ533, selected on medium lacking leucine, and then transferred to 5-fluoroorotic-acid-containing medium to select for loss of pJ533.

To construct plasmids to transiently overexpress the PrLDs fused to GFP, the NM domain (the prion domain, plus the adjacent middle domain; see Figure 2.2A) of each hnRNPA2-Sup35 fusion was amplified with oligonucleotides EDR1624 (GAGCTACTGGATCCACAATGTCAGGACCTGGATATGGCAACCAG) and EDR1924 (GTCGATGCTACTCGAGTCGTTAACAACTTCGTCATCCACTTC). The resulting PCR products were digested with BamHI and XhoI and inserted into BamHI/XhoI cut pER760, a TRP1 plasmid that contains GFP under control of the *GAL1* promoter [28].

Prion formation assays were performed as previously described [29]. Briefly, cells expressing a given hnRNPA2-Sup35 fusion as the sole copy of Sup35 were transformed with either an empty vector (pKT24; [30]) or a plasmid expressing the matching PrLD-GFP fusion under control of the *GAL1* promoter. Cells were grown for 3 days in galactose/raffinose dropout medium lacking tryptophan. Ten-fold serial dilutions were then plated onto synthetic complete medium lacking adenine to select for [PSI^{+}] cells and onto medium with adenine to test for cell viability.

Western Blot

To probe PrLD-GFP expression levels in yeast, TRP1 plasmids expressing the PrLD-GFP fusion were transformed into the corresponding mutant strain. Low density cultures were pre-grown in raffinose dropout medium overnight, diluted to an OD of 1.0 in 10mL of 3% galactose/raffinose dropout medium, and grown for 4 h. Cells were harvested by centrifugation. Cell pellets were lysed as previously described [31], with protease inhibitor cocktail (Gold Biotechnology) included in the lysis buffer. Lysates were normalized based on total protein concentration, as determined by Bradford assay (Sigma). Proteins were separated on SDS/12% PAGE gels and transferred to a PVDF membrane, and immunoblotted, using a monoclonal anti-GPF primary antibody (Santa Cruz Biotechnology), and Alexa Fluor IR800 goat anti-mouse secondary antibody (Rockland).

To probe endogenous expression levels, log-phase cultures were harvested by centrifugation, and cells lysed as above. Proteins were separated by SDS-PAGE and analyzed by western blot, using a monoclonal antibody against the Sup35C domain (BE4 [32], from Cocalico Biologicals, kindly made available by Susan Liebman) as the primary antibody, and Alexa Fluor IR800 goat anti-mouse (Rockland) as the secondary antibody.

Fly Stocks and Culture

Mutagenesis using the QuickChange Lightning kit (Agilent) was performed on pUASTattB-wild type hnRNPA2 construct as previously described [18]. Flies carrying transgenes in pUASTattB vectors were generated by performing a standard injection and φC31 integrase-mediated transgenesis technique (BestGene Inc.). To express a transgene in muscles,

Mhc-Gal4 was used (from G. Marqués). All *Drosophila* stocks were maintained in a 25°C incubator with a 12 h day/night cycle and a standard diet.

Preparation of Adult Fly Muscle for Immunofluorescence

Adult flies were embedded in a drop of OCT compound (Sakura Finetek) on a glass slide, frozen in liquid nitrogen and bisected sagittally by using a razor blade. After being fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS), fly tissues were permeabilized in PBS containing 0.2% Triton X-100, and indiscriminant binding was blocked by adding 5% normal goat serum in PBS. The hemithoraces were stained with anti-hnRNPA2B1 (EF-67) antibody (Santa Cruz Biotechnology) followed with Alexa-488-conjugated secondary antibody (Invitrogen), Texas Red-X phalloidin (Invitrogen) and DAPI according to manufacturer's instructions. Stained hemithoraces were mounted in 80% glycerol and the muscles were imaged with a Marianas confocal microscope (Zeiss, x63).

Fly Thoraces Fractionation Protocol

Thoraces of at least 15 adult flies were dissected, homogenized in RIPA buffer, and lysed on ice for 15 min. The cell lysates were sonicated and then cleared by centrifugation at 100,000 × g for 30 min at 4 °C to generate the RIPA soluble samples. To prevent carry-overs, the resulting pellets were washed with RIPA buffer. RIPA insoluble pellets were then extracted with urea buffer (7 M urea, 2 M thiourea, 4% CHAPS, 30 mMTris, pH 8.5), sonicated, and centrifuged at 100,000 × g for 30 min at 22 °C. Protein concentration was determined by bicinchoninic acid method (Pierce), and samples were boiled for 5 min and analysed by the standard western blotting method provided by Odyssey system (LI-COR) with 4–12% NuPAGE

Bis-Tris Gel (Invitrogen) and anti-hnRNPA2B1 (DP3B3) antibody (Abcam, 1:2000) and anti-actin antibody (Santa Cruz, 1:10000).

In Vitro Aggregation Assays

A 96-well plate was treated with 5% casein solution for 5 minutes at room temperature, and then rinsed with DI water and allowed to dry. Synthetic peptides (GenScript) were dissolved at 2.5 mM in 6M guanidine HCl. Peptides were then diluted approximately100-fold to a final concentration of 25 µM in 10 mM sodium phosphate, 150 mM NaCl, 12.5 µM thioflavin T, 0.02% casein, pH 7.4 in the 96-well plate to initiate aggregation. Fluorescence was monitored in a Victor3 Perkin Elmer fluorescence plate reader, with excitation and emissions wavelengths of 460 and 490 nm, respectively. Reactions were monitored for 48 h. Between readings, reactions were incubated without agitation for 3 minutes, and then shaken for 10 sec. The fraction aggregated was calculated by normalizing relative to the final fluorescence of the well.

For electron microscopy of *in vitro* aggregation reactions, 10-20 µl of sample was incubated on carbon copper grids for 5 min, and then rinsed with distilled water. Grids were stained with 1% uranyl acetate for 30 sec, and observed on a JEOL JEM-1400 TEM, imaging with a Gatan Orius 832 Camera.

RESULTS

Hydrophobic and Aromatic Residues Promote Aggregation

We previously developed a yeast system to monitor the prion-like activity of hnRNPA2 [18]. The yeast prion [*PSI*$^{+}$] is the prion form of the translation termination factor Sup35 [33, 34]. Sup35 has three functionally distinct domains: an N-terminal prion domain that is required

for prion aggregation; a C-terminal functional domain that is necessary and sufficient for Sup35's normal function in translation termination; and a highly charged middle domain that is not required for either prion formation or Sup35's translation termination activity, but which stabilizes prion fibers (Figure 2.2A; [35-37]). Yeast prion domains are generally modular, meaning that they maintain prion activity when attached to other proteins [38]. Because simple assays are available to detect [*PSI*$^{+}$] formation, substitution of the prion domain of Sup35 with fragments from other prion-like proteins has been widely used to probe for prion activity [39-41]. The first 40 amino acids of the Sup35 prion domain are required for prion formation, while the remainder of the prion domain, which is composed of a series of imperfect oligopeptide repeats, is predominantly involved in prion maintenance (Figure 2.2A; [20, 42, 43]). Therefore, substitution of fragments in the place of the first 40 amino acids of Sup35 can be used to probe aggregation propensity of these domains, and to examine the effects of mutation on aggregation propensity [20]. The core PrLDs from mutant hnRNPA2 can support prion activity when substituted in place of the first 40 amino acids of Sup35, while the wild-type prion domain cannot [18]. Thus, these fusion proteins provide a convenient system for examining how amino acid sequence affects PrLD aggregation propensity.

The prion prediction algorithm PAPA predicts that within PrLDs, charged amino acids and proline should strongly inhibit prion formation, while aromatic and hydrophobic amino acids promote prion formation [28, 44]. The disease-associated mutations in hnRNPA2B1 and hnRNPA1 each involve substitution of a strongly prion-inhibiting amino acid (aspartic acid) with a neutral (asparagine) or prion-promoting (valine) amino acid. We therefore hypothesized that replacing the aspartic acid with any predicted prion-promoting amino acid would have a similar effect.

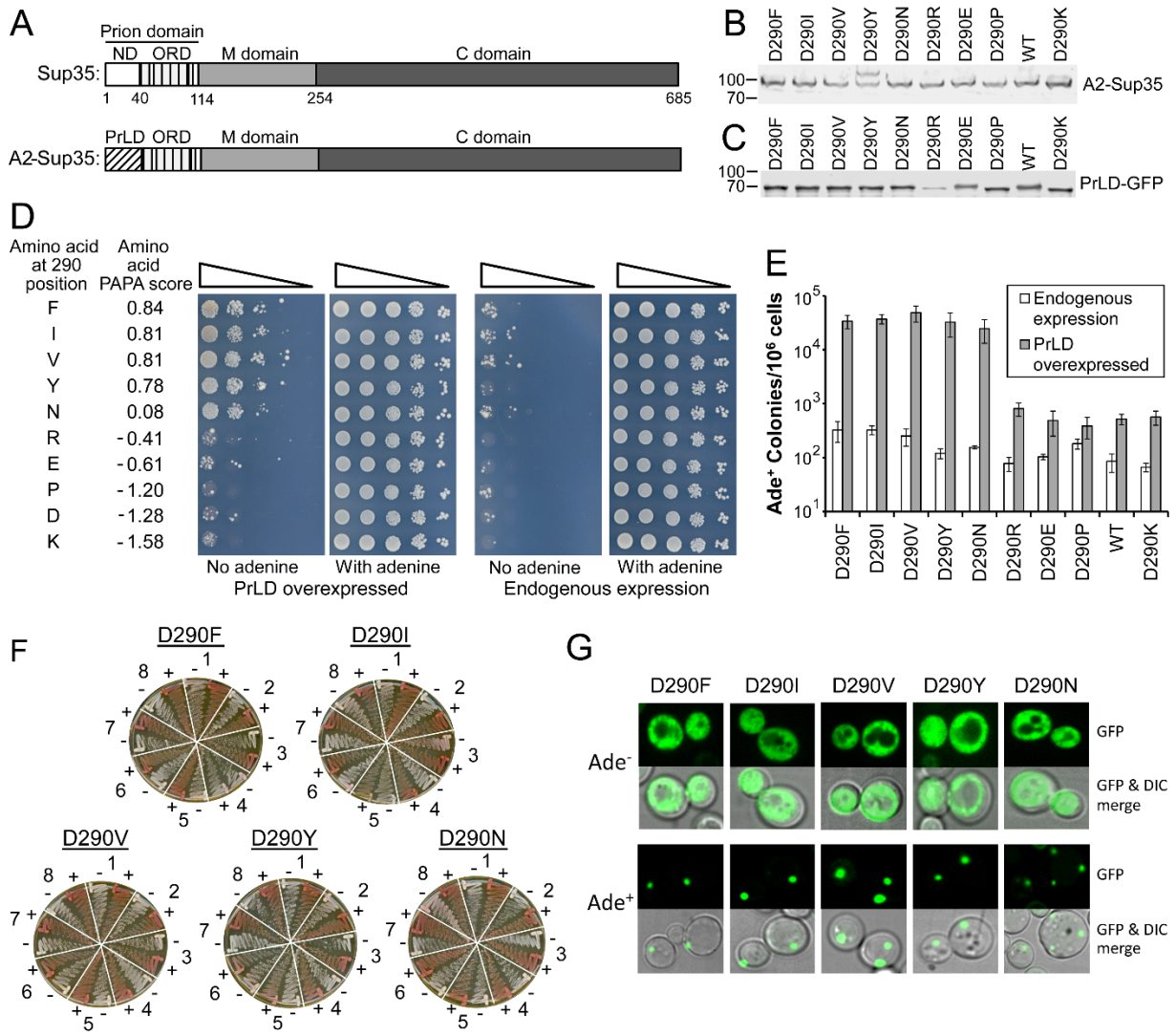


Figure 2.2: PAPA accurately predicts prion-promoting mutations at the 290 position. (A) Schematic of wild-type Sup35 and the hnRNPA2-Sup35 chimeric protein (18). Sup35 contains three domains: an N-terminal prion domain (N), a highly charged middle (M) domain, and a C-terminal domain that is responsible for Sup35's translation termination function. The prion domain contains two parts: a nucleation domain (ND) that is required for prion formation, and an oligopeptide repeat domain (ORD) that is dispensable for prion nucleation, but is required for prion propagation. In the hnRNPA2-Sup35 fusion, the ND (amino acids 3-40) of Sup35 was replaced with the core PrLD (amino acids 261-303) from hnRNPA2B1. **(B)** Western blot analysis of endogenous expression of full-length wild-type (WT) and mutant hnRNPA2-Sup35 chimeric proteins, using an antibody to the Sup35 C-terminal domain. **(C)** Western blot analysis of overexpression of PrLD-GFP fusions, using an antibody to GFP. The NM domain of each hnRNPA2-Sup35 chimera was fused to GFP and expressed from the *GAL1* promoter. **(D)** Effects of different amino acids at the D290 position. [*psi*$^-$] strains were generated that expressed hnRNPA2-Sup35 fusion proteins with the indicated substitution at the D290 position as the sole copy of Sup35 in the cell. The strains were transformed with either an empty vector (endogenous expression) or a plasmid expressing the matching PrLD-GFP mutant under control of the *GAL1*

promoter (PrLD overexpression). Cells were grown in galactose dropout medium for 3 days, and then 10-fold serial dilutions were plated onto medium lacking adenine to select for [*PSI*$^{+}$] and medium containing adenine to test for cell viability. PAPA scores for each amino acid are indicated. Wild-type (D290) and D290V were previously reported [18]. **(E)** Quantification of Ade^{+} colony formation. Serial dilutions of the galactose cultures from Panel D were plated onto full plates containing medium with and without adenine. The frequency of Ade^{+} colony formation was determined as the ratio of colonies formed with and without adenine. Data represent mean ± s.d.; n ≥ 3. **(F)** Curability of Ade^{+} colonies. For each mutant, eight individual Ade^{+} isolates were grown on YPD (-) or YPD plus 4 mM guanidine HCl (+). Cells were then restreaked onto YPD to test for loss of the Ade^{+} phenotype. **(G)** The Ade^{+} phenotype is associated with protein aggregation. For the indicated mutants, Ade^{-} and Ade^{+} cells were transformed with a plasmid expressing the matching PrLD-GFP mutant under control of the *GAL1* promoter. Cells were grown for 1 h in galactose dropout medium and visualized by confocal microscopy.

To test this hypothesis, we replaced the aspartic acid at the disease-associated position in hnRNPA2 with predicted prion-promoting amino acids (phenylalanine, isoleucine, and tyrosine), a prion-neutral amino acid (asparagine), and prion-inhibiting amino acids (arginine, glutamic acid, proline, or lysine). We tested these mutations in the hnRNPA2-Sup35 chimeric protein (Figure 2.2A). [*PSI*$^+$] formation can be assayed by monitoring nonsense suppression of *ade2-1* allele [45]. *ade2-1* mutants are unable to grow in the absence of adenine, and turn red on limited adenine due accumulation of a pigment derived from the substrate of the Ade2 protein. In [*PSI*$^+$] cells, Sup35 is sequestered into prion aggregates, resulting in occasional read through of the *ade2-1* premature stop codon; therefore, [*PSI*$^+$] are able to grow in the absence of adenine, and form white or pink colonies on limiting adenine. One hallmark of prion activity is that increasing protein concentration should increase the frequency of prion formation [34]. We therefore monitored the frequency of Ade$^+$ colony formation with and without overexpression of the matching prion domain fused to GFP.

The full-length fusions showed only modest differences in protein expression, although the D290Y mutant showed two bands, suggesting a possible post-translational modification (Figure 2.2B). Likewise, all of the PrLD-GFP fusions, except the one from the D290R mutant, showed similar levels of overexpression (Figure 2.2C). The single point mutations had profound effects on Ade$^+$ colony formation, with the mutants showing multiple orders-of-magnitude differences upon PrLD overexpression (Figure 2.2D,E). Strikingly, there was a strong correlation between the predicted effect of each mutation and the observed frequency of Ade$^+$ colony formation. Each mutation predicted to enhance prion activity (D290F, I, V, Y, and N) showed statistically significant increases in Ade$^+$ colony formation upon PrLD overexpression ($P<0.001$ by t test) relative to the wild-type fusion.

Ade$^+$ colony formation can result from either prion formation or from a nonsense suppressor mutation. For each of the prion-promoting mutations (D290F, I, V, Y, and N), the fact that the frequency of Ade$^+$ colony formation showed a multiple orders-of-magnitude increase upon PrLD overexpression strongly suggests that the Ade$^+$ phenotype is a result of prion formation, as the frequency of DNA mutation should be insensitive to expression levels [34]. Two assays were used to further confirm that these mutants were forming prions. First, we tested whether the Ade$^+$ phenotype could be cured by low concentrations of guanidine HCl. Guanidine HCl cures [*PSI*$^+$] [46] by inhibiting Hsp104 [47, 48]. For the D290F, I, V, Y, and N mutants, almost all tested Ade$^+$ colonies formed upon PrLD overexpression maintained a white phenotype in the absence of guanidine HCl, but turned red after treatment with guanidine HCl (Figure 2.2F), consistent with the Ade$^+$ phenotype resulting from prion formation. By contrast, none of the tested Ade$^+$ colonies formed by the D290R, E, P, and D mutants were not curable by guanidine HCl (data not shown), suggesting that the Ade$^+$ phenotype is likely a result of DNA mutation. The D290K mutant did have a small number of stable, curable Ade$^+$ colonies, although these occurred less frequently than for any of the aggregation-promoting mutations (data not shown). Second, we used a GFP assay [49] to confirm that the fusion proteins were aggregated in curable Ade$^+$ cells. When Sup35N-GFP is transiently overexpressed in [*psi*$^-$] cells, it initially shows diffuse cytoplasmic localization; by contrast, in [*PSI*$^+$] cells, Sup35N-GFP rapidly joins existing prion aggregates, and coalesces into foci [49]. Therefore, to test for the presence of prion aggregates, we transiently overexpressed PrLD-GFP fusions in Ade$^+$ and Ade$^-$ cells for each predicted prion-promoting mutant. In the Ade$^-$ cells, the GFP fusions remained diffuse, while in Ade$^+$ cells the fusions rapidly coalesced into foci (Figure 2.2G).

Additive and Compensatory Mutations

Each of the disease-associated mutations in hnRNPA2B1 and hnRNPA1 target a highly conserved aspartic acid within a motif that is conserved across much of the hnRNP A/B family [18], suggesting that this position may be a critical determinant of aggregation propensity; however, composition-based algorithms like PAPA predict that there is nothing unique about this specific aspartic acid, and that similar mutations at other positions should exert a similar effect. The hnRNPA2 PrLD contains very few predicted prion-inhibiting amino acids, but a second aspartic acid is found at amino acid 276 (Figure 2.1). An aspartic acid to valine substitution at this position also promoted Ade^+ colony formation, although to a lesser extent than the disease-associated mutations (Figure 2.3A,B). Additionally, combining mutations at both positions had an additive effect, generating a mutant that formed Ade^+ colonies efficiently even in the absence of PrLD overexpression (Figure 2.3A,B). For both the D276V mutant and the double mutant, the majority of Ade^+ colonies formed upon PrLD overexpression were curable by guanidine HCl, consistent with prion formation (data not shown). The strong additive effect of the mutations was not due to differences in protein expression; the double mutant actually had slightly lower levels of expression for the full-length hnRNPA2-Sup35 chimeric protein, and its PrLD-GFP fusion showed similar levels of expression to the wild-type and D290V mutant (Figure 2.3C).

PAPA also predicts that it should also be possible to design compensatory mutations that offset the effects of the disease-associated mutations. Tyrosines are predicted to be strongly prion-promoting [44]. As predicted, replacing Y283 with various prion-inhibiting amino acids (R, E, P, D, K) partially or completely offset the effects of the D290V mutation, while replacing this tyrosine with other prion-promoting amino acids had little effect (Figure 2.4A,B). Similar results were seen with a more limited panel of mutants at a second position (Y288; Figure 2.4C). For

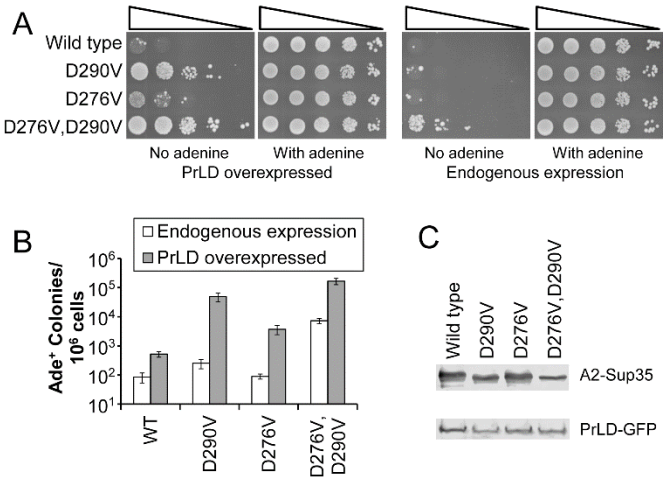


Figure 2.3: Additive mutations. (A) The indicated hnRNPA2-Sup35 mutants were tested for prion formation. D276V enhances Ade^+ colony formation, albeit less than the D290V mutation. The D276V/D290V double mutant shows substantially higher levels of Ade^+ colony formation than D290V alone, even forming Ade^+ colonies in the absence of PrLD overexpression. **(B)** Quantification of Ade^+ colony formation. Data represent mean ± s.d.; $n \geq 3$. **(C)** Western blot analysis of endogenous expression of full-length hnRNPA2-Sup35 chimeric proteins and overexpression of PrLD-GFP fusions.

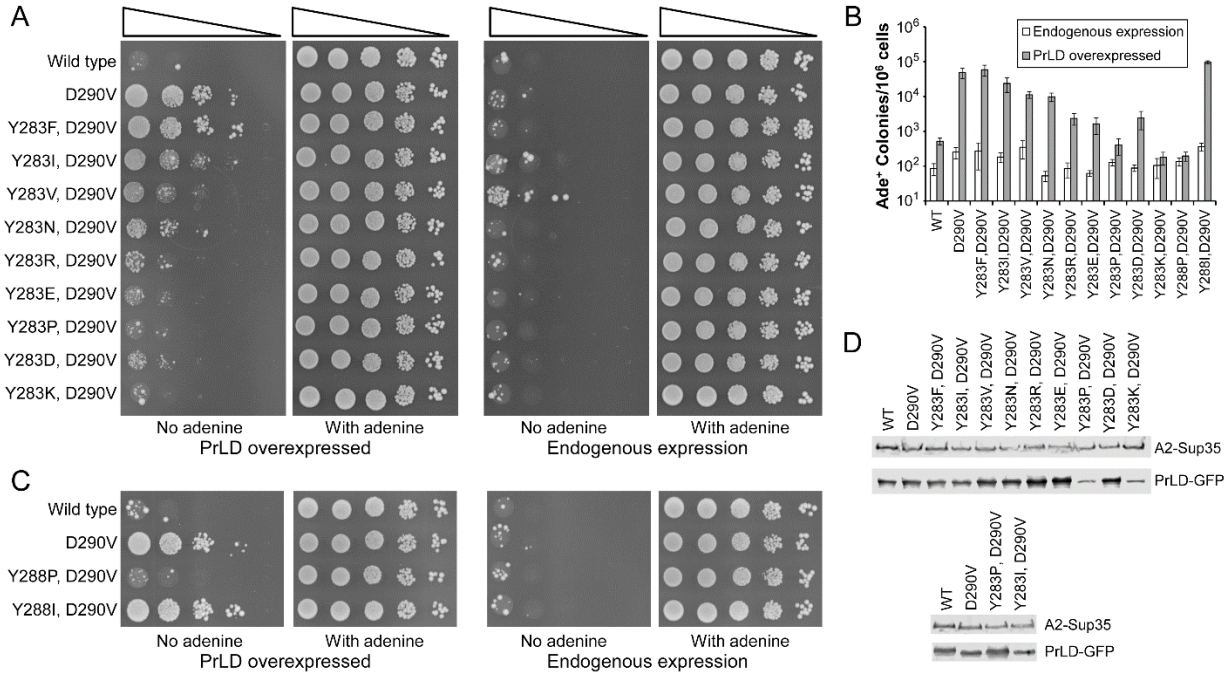


Figure 2.4: Compensatory mutations. (A) Prion-inhibiting mutations effectively offset the effects of the D290V mutation. Y283 in the hnRNPA2-Sup35 (D290V) fusion was replaced with either other prion promoting amino acids (F, I, V), a neutral amino acid (N), or prion-inhibiting amino acids (R, E, P, D, K). Each of the predicted prion-inhibiting amino acids partially or completely reversed the effects of the D290V mutation. **(B)** Quantification of Ade^+ colony formation. Data represent mean ± s.d.; $n \geq 3$. **(C)** Y288 in the hnRNPA2-Sup35 (D290V) fusion was replaced with either a prion-inhibiting proline or a prion-promoting isoleucine. **(D)** Western blot analysis of endogenous expression of full-length hnRNPA2-Sup35 chimeric proteins and overexpression of PrLD-GFP fusions.

each of the mutants in which Y was replaced with a prion-promoting amino acid, the majority of Ade^+ colonies formed upon PrLD overexpression were curable by guanidine HCl, consistent with prion formation (data not shown). Y283 and Y288 mutants showed only modest differences in expression of the full-length hnRNPA2-Sup35 chimeras, although two of the Y283 mutants (Y283P and Y283K) showed lower levels of PrLD-GFP overexpression (Figure 2.4D), potentially explaining why these two mutations showed the strongest aggregation-inhibiting effects.

Zipper Segments are Neither Necessary nor Sufficient for Prion Aggregation

Each of the disease-associated mutations in hnRNPA2B1 and A1 are predicted by ZipperDB to create strong steric zipper segments (Figure 2.5A,B; [18]). Each of the prion-promoting residues tested in Figure 2.2 are likewise predicted to create strong zipper segments, so it is unclear whether the mutations enhance prion formation solely because of compositional effects, or due to creation of a steric zipper. The presence of a strong zipper segment is clearly not sufficient for prion formation, as the compensatory mutations in Figure 2.4A prevent prion formation without disrupting the predicted zipper segment (Figure 2.5C and data not shown). We designed additional mutations to test whether zipper segments are necessary for prion formation by the hnRNPA2-Sup35 chimera. Because aspartic acid is predicted by PAPA to be strongly prion-inhibiting, deletion of aspartic acid is predicted to enhance prion activity. Indeed, deletion of one or both aspartic acids in the core A2 PrLDs strongly enhanced Ade^+ colony formation by the fusion proteins (Figure 2.5D, E), and the majority of Ade^+ colonies formed upon PrLD overexpression for these mutants were curable by guanidine HCl. This effect is not

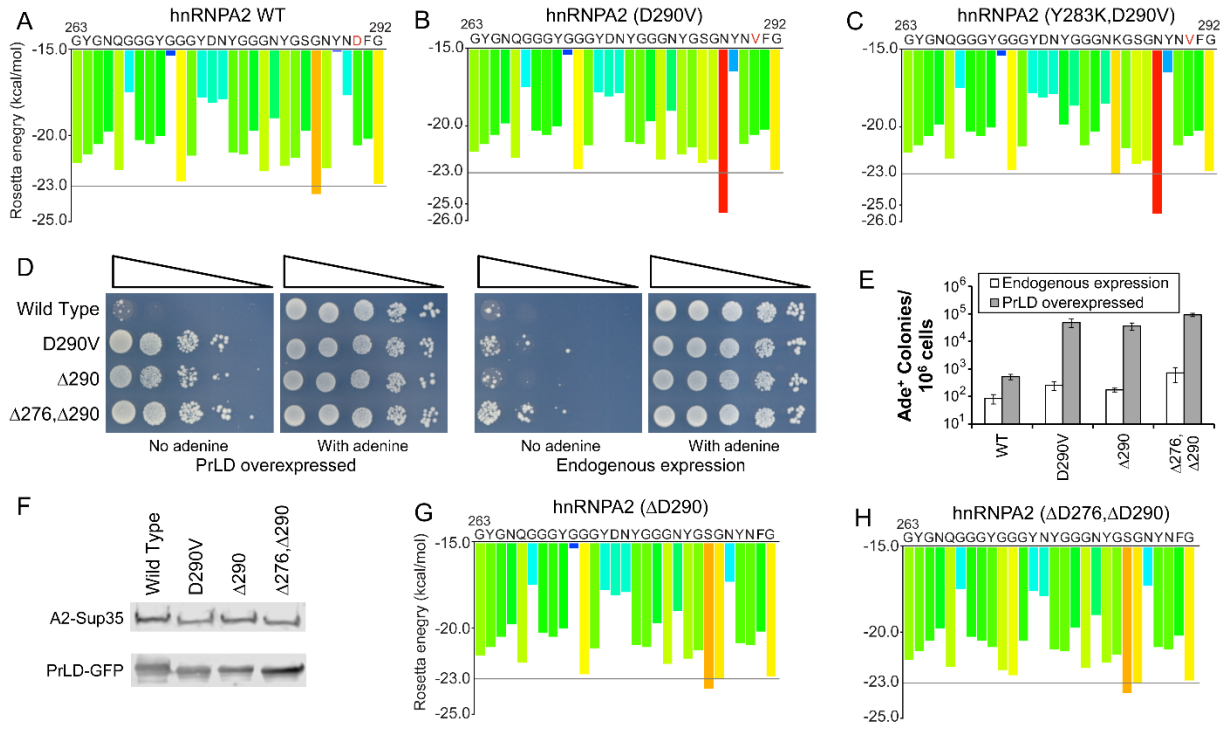


Figure 2.5: Predicted strong steric zipper segments are neither necessary nor sufficient for prion activity. (A, B) ZipperDB analysis [24] of the core PrLDs of hnRNPA2 wild-type and D290V. Segments with a Rosetta energy below -23.0 kcal/mol are predicted to form steric zippers. The D290V mutation creates a strong predicted steric zipper segment from amino acids 287-292. The remainder of the core PrLD is not scored by ZipperDB due to the presence of prolines at positions 262 and 298. Adapted from Kim et al. [18]. **(C)** The Y283K mutation blocks prion formation by the hnRNPA2-Sup35 (D290V) mutant (Figure 2.4), but does not affect the predicted strong steric zipper segment. **(D)** Both ΔD290 and a ΔD276/ΔD290 double mutant substantially increase Ade^+ colony formation by the hnRNPA2-Sup35 fusion. **(E)** Quantification of Ade^+ colony formation. Data represent mean ± s.d.; $n \geq 3$. **(F)** Western blot analysis of endogenous expression of full-length hnRNPA2-Sup35 chimeric proteins and overexpression of PrLD-GFP fusions. **(G, H)** Neither the ΔD290 nor ΔD276/ΔD290 mutations are predicted to create a strong steric zipper.

due to differences in expression level; the ΔD290 mutant showed similar levels of expression to wild-type for both the full-length hnRNPA2-Sup35 chimera and the PrLD-GFP fusion (Figure 2.5F), and although the double deletion showed modestly higher PrLD-GFP overexpression, this difference would be unlikely to explain the multiple orders of magnitude increase in Ade+ colony formation relative to the wild-type protein (Figure 2.5F). However, although they substantially increased prion formation, neither of these mutations is predicted to create a strong steric zipper segment (Figure 2.5G, H), indicating that strong zipper segments are neither necessary (Figure 2.5D-H) or sufficient (Figure 2.5C) for prion-like aggregation.

Effects of Mutations in Drosophila

For each of the mutations tested in yeast, prion activity closely correlated with PAPA predictions. Because hnRNPA2(D290V) primarily causes myopathy in humans [18], we were interested in whether our yeast results could accurately predict myopathy in a multi-cellular organism. Expression of aggregation-prone prion or prion-like proteins in muscle tissue of various model systems can cause muscle disorganization [18, 50, 51]. We previously showed that when expressed in *Drosophila*, wild-type hnRNPA2 localizes to the nucleus and is predominantly detergent soluble, whereas hnRNPA2(D290V) forms cytoplasmic inclusions, is largely detergent insoluble, and leads to muscle degeneration (Figure 2.6; [18]). Similar results were observed with two other antibodies: DP3B3 with untagged hnRNPA2, and anti-Flag antibody with Flag-tagged hnRNPA2 (data not shown). The cytoplasmic inclusions formed by hnRNPA2(D290V) are RNA granule assemblies, containing various RNA binding proteins [52].

To test whether other predicted prion-promoting mutations at the 290 position would also increase insolubility and promote formation of cytoplasmic foci, we expressed the

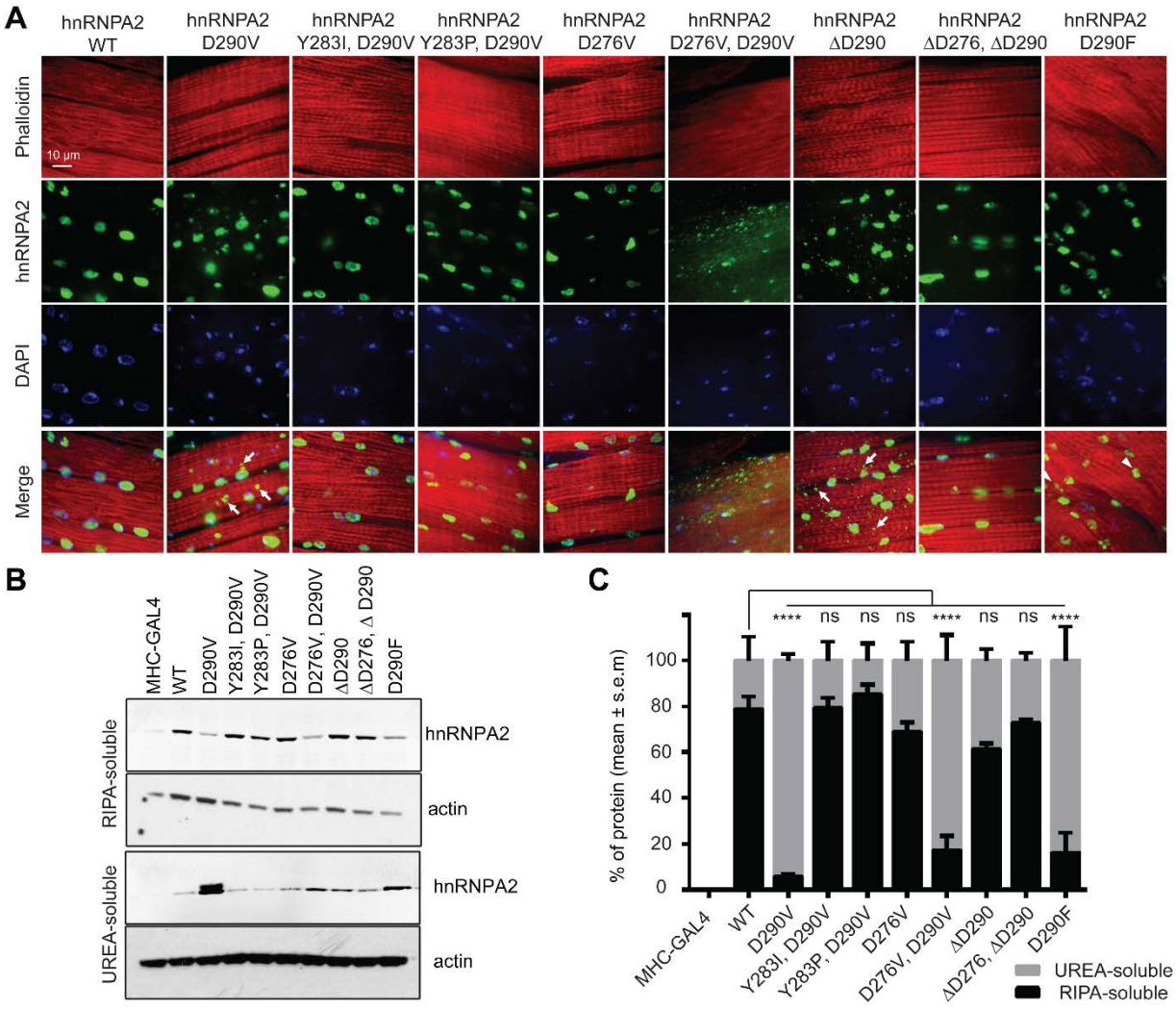


Figure 2.6: Effects of mutations in *Drosophila*. (A) Adult fly thoraces were stained with anti-hnRNPA2B1 (green), Texas Red-X phalloidin (red) and DAPI (blue). Wild type hnRNPA2 localizes exclusively to the nuclei, whereas the D290V mutant also forms cytoplasmic foci. The other mutants show a range of localization patterns, including much more substantial cytoplasmic foci for the D276V,D290V double mutant. Examples of cytoplasmic and nuclear foci are indicated with arrows and arrow heads, respectively. **(B)** Thoraces of adult flies were dissected and sequential extractions were performed to examine the solubility profile of hnRNPA2. **(C)** Quantification of the blot shown in (B). Data represent mean ± s.e.m.; n = 3; **** $P < 0.0001$; two-way ANOVA test with Bonferroni's post hoc test.

hnRNPA2(D290F) in *Drosophila*. While there was less RIPA-insoluble (urea-soluble) protein for hnRNPA2(D290F) than for hnRNPA2(D290V) (Figure 2.6B), hnRNPA2(D290F) was nevertheless largely RIPA-insoluble (Figure 2.6C); interestingly it predominantly formed nuclear foci, rather than the cytoplasmic foci seen for hnRNPA2(D290V) (Figure 2.6A).

To test whether mutations at other sites would mimic the D290V mutation, we expressed the hnRNPA2(D276V) and the hnRNPA2(D276V,D290V) mutants. As in yeast, the D276V mutation had a smaller effect than the D290V mutation. D276V slightly increased the fraction of detergent insoluble protein compared to wild-type hnRNPA2, although this increase was not statistically significant (Figure 2.6C). As in yeast, the double mutant had a strongly additive effect. The fraction of insoluble protein was actually slightly lower than for the hnRNPA2(D290V), likely because hnRNPA2(D290V) had higher protein levels, and was already almost entirely insoluble (Figure 2.6C); however, the double mutant had a much more dramatic immunohistological phenotype, with many small foci throughout the cytoplasm and nucleus (Figure 2.6A). Additionally, it showed clear disruption of muscle fibers, seen as a loss of the regular striations normally observed with phalloidin staining of healthy muscle (Figure 2.6A).

Two of the mutations showed different behavior in yeast and *Drosophila*. As expected, the predicted prion-inhibiting Y283P mutation largely offset the effect of the D290V mutation, restoring solubility and nuclear localization (Figure 2.6). However, the control Y283I mutation, which had little effect in yeast (Figure 2.4A), also offset the effect of the D290V mutation in *Drosophila* (Figure 2.6). As in yeast, the ΔD290 mutation decreased solubility of the protein *Drosophila*, albeit not statistically significantly (Figure 2.6C), and caused formation of

cytoplasmic inclusions (Figure 2.6A); however, the hnRNPA2(ΔD276,ΔD290) double mutant actually appeared to be more soluble.

One other striking difference was observed between hnRNPA2(D290V) and all other mutants tested: only hnRNPA2(D290V) showed two bands on the western blot, likely reflecting an uncharacterized post-translational modification. The significance of this second band is unclear; given that it was not observed in the hnRNPA2(D276V,D290V) double mutant, clearly it is not required for insolubility, mislocalization, or muscle pathology.

In Vitro Analysis of Mutants

Prediction algorithms like PAPA are generally designed to predict the intrinsic aggregation propensity of peptides or proteins. However, mutations can influence aggregation by affecting activities other than intrinsic aggregation propensity, including: altering interactions with other cellular factors or with other parts of the protein; changing expression levels or protein stability; or altering localization. For the mutants that showed divergent behavior in yeast and *Drosophila*, we hypothesized that this divergent behavior likely reflected effects of the mutation beyond intrinsic aggregation propensity. To test this hypothesis, we utilized an *in vitro* aggregation assay to examine the intrinsic aggregation propensity of these mutants in the absence of other cellular factors.

We generated 35-amino acid peptides from the core PrLDs. We tested each for amyloid aggregation using thioflavin T, a dye that fluoresces upon interaction with amyloid fibrils, but not soluble proteins or amorphous aggregates [53]. Most amyloid-forming proteins show sigmodal aggregation kinetics, with a lag time, followed by a growth phase in which there is a rapid increase in aggregation, and then a plateau, as soluble material is exhausted. For each

protein where the yeast and *Drosophila* results diverged, the *in vitro* aggregation kinetics mimicked the yeast results and PAPA predictions; higher frequencies of prion formation in yeast correlated with shorter lag times and a steeper growth phase. Specifically, the wild-type protein showed very slow aggregation kinetics, with a lag time of approximately 25 h (Figure 2.7A). The D290V mutation substantially accelerated aggregation, shortening the lag phase to about 4 h (Figure 2.7B). The ΔD290 likewise showed accelerated aggregation, which was further enhanced in the ΔD276/ΔD290 double mutant (Figure 2.7A). The Y283P mutant was largely able to offset the aggregation promoting effect of D290V, while the more conservative Y283I mutation had little effect (Figure 2.7B). In all cases, the increase in thioflavin T signal was associated with fiber formation (Figure 2.7C).

Collectively, these results indicate that both the PAPA prediction algorithm and yeast fusion system can accurately predict the effects of mutations on the intrinsic aggregation propensity of peptides, but that intrinsic aggregation propensity is an imperfect predictor of in vivo aggregation of the peptides in the context of their respective full-length proteins.

DISCUSSION

Although prion formation in humans is generally thought of as pathogenic, an emerging theory suggests that many prion-like domains may be evolved to form weak or transient interactions that mediate the formation of membrane-less organelles [9]. For example, P bodies and stress granules are two types of RNA-protein assemblies that regulate translation and mRNA turnover. The prion-like domain of TIA-1 helps mediate the formation of stress granules [12], and in yeast, the prion-like domain of Lsm4 is involved in P body formation [11]. This suggests

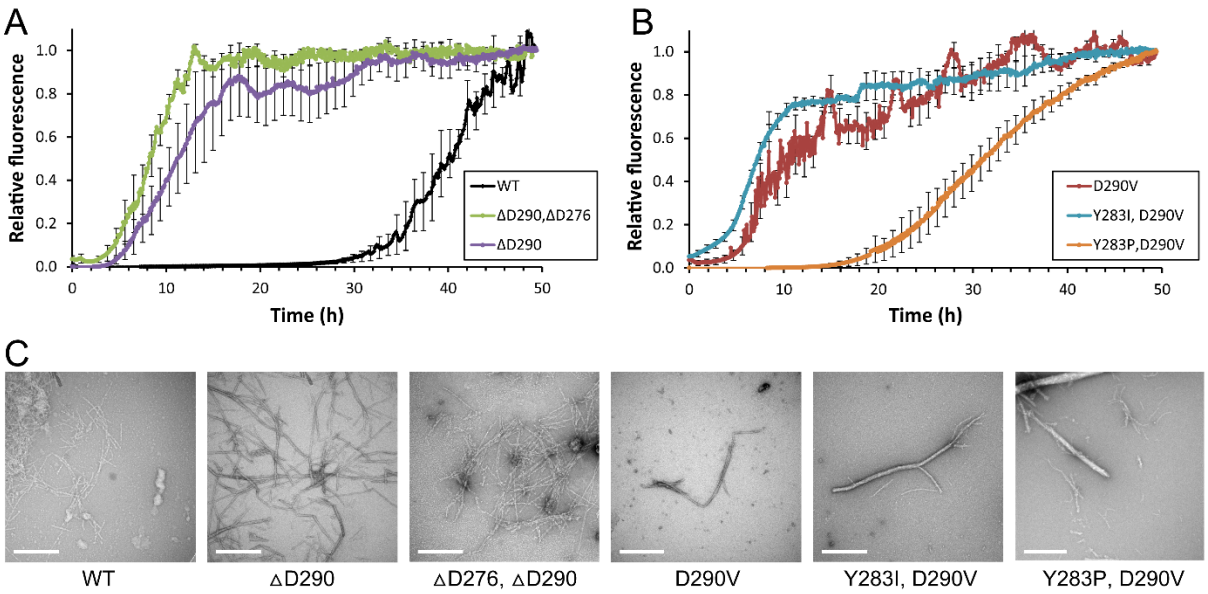


Figure 2.7: In vitro amyloid formation by hnRNPA2 mutants. (A, B) Synthetic 35-amino-acid peptides from the hnRNPA2 core PrLD were generated with the indicated mutations. Peptides were resuspended under denaturing conditions, and then diluted to initiate amyloid formation. Reactions were incubated at room temperature with intermittent shaking. Amyloid formation was monitored by thioflavin T fluorescence. Data represent the mean ± s.e.m., with error bars shown for every 10th data point; n = 3. **(C)** Electron micrographs of amyloid formation assays after 48 h. Scale bars, 500 nm.

the intriguing hypothesis that mutations may cause disease by disrupting the dynamics of these assemblies.

However, the exact relationship between PrLD aggregation propensity and disease is unclear. The cytoplasmic inclusions seen for disease-associated hnRNPA2B1 and A1 mutations are not just simple aggregates of these proteins, but instead are RNA granules [52]. Therefore, these complex structures are normally under regulatory control, so it is unclear whether simple increases in aggregation propensity are sufficient to cause disease. Examining this question is challenging, as our understanding of these diseases is currently based on a limited set of mutations. For example, only one disease-associated mutation has been characterized in hnRNPA2B1. Furthermore, targeted mutations to investigate the role of PrLD aggregation in protein function and pathology have often involved dramatic changes to protein sequence, such as deletion or replacement of the entire PrLD, so these mutations likely have effects beyond just changing aggregation propensity.

By contrast, we were able to cause profound changes in the aggregation propensity of hnRNPA2 with just single or double point mutations. This ability to rationally design more subtle mutations to alter aggregation propensity will provide a powerful tool to explore the role of PrLDs in functional and pathological aggregation. While not all mutations behaved as expected in *Drosophila* in the context of full-length hnRNPA2, most did, suggesting that it is relatively simple to design mutations to modulate aggregation propensity. This will facilitate experiments both to explore the normal role of functional aggregation and to test whether increasing aggregation propensity disrupts the dynamics of these aggregates and leads to disease.

Our success rate in predicting prion aggregation in yeast was surprising. We designed 13 mutations expected to increase prion formation relative to the wild-type hnRNPA2-Sup35 fusion

(D290F, I, Y, and N; D290V paired with Y283F, I, V, or N; D290V paired with Y288I; D276V; D276V,D290V; ΔD290; and ΔD276,ΔD290) and 10 mutations expected to decrease prion activity relative to the D290V mutant (D290R, E, P, and K; D290V paired with Y283R, E, P, D, K; and D290V paired with Y288P). Numerous factors can affect Ade^+ colony formation in our assay: our mutants showed subtle differences in protein expression, which will influence prion formation; for wild-type Sup35, many prions formed are toxic to cells [54], so the fraction of toxic prions for a given mutant will affect Ade^+ colony formation; the assay may not detect weak, poorly propagating prions; and interactions with other cellular proteins may influence prion activity. Despite all of these potentially confounding factors, we accurately predicted the direction of the effect relative to the wild-type or D290V reference for all 23 mutations tested. Although the strength of the effect of each mutation was not perfectly predictable, this remarkable prediction success despite the limitations of the assay demonstrates that even single point mutations can have a profound effect on aggregation propensity.

Nevertheless, our experiments also highlight remaining challenges for predicting the effects of mutations. In the past decade, mutations in numerous PrLD-containing RNA binding proteins, including FUS [55, 56], TDP-43 [57], TAF15 [58], EWSR1 [59], hnRNPDL [60], hnRNPA1 [18], and hnRNPA2 [18] have been linked to various degenerative diseases. While some algorithms have proven successful at identifying the PrLDs in these proteins [7], predicting the effects of mutations has proven more difficult [6, 7]. One issue is that while mutations in these proteins appear to cause disease by disrupting of RNA homeostasis, increasing the aggregation propensity of these RNA binding proteins is just one of many mechanisms by which RNA homeostasis could be disrupted. For example, for TDP-43, a subset of disease-associated mutations do not cause a detectable increase in aggregation propensity [61]. Likewise, for FUS

and hnRNPA1, some of the disease-associated mutations are found in a predicted nuclear localization signal [62, 63], so may lead to the formation of cytoplasmic inclusions by disrupting nuclear localization rather than directly increasing aggregation propensity.

Our results suggest an additional challenge in predicting the effects of mutations: while we clearly have made substantial strides in predicting the effects of mutations on the aggregation propensity of isolated PrLDs, this was an imperfect predictor of foci formation by the full-length protein in *Drosophila*. The reasons for this discrepancy are unclear. For natively folded proteins, native state stability can be a critical determinant of aggregation propensity [64], so the intrinsic aggregation propensity of a protein (i.e., the propensity of the protein to form aggregates from a denatured state) is an imperfect predictor of the aggregation propensity of the native protein. However, the hnRNPA2 PrLD is predicted to be intrinsically disordered, so native state stability should be less of an issue. For a disordered protein, other factors, including changes in localization, interactions with other proteins or nucleic acids, post-translational modifications, or expression levels could indirectly affect aggregation propensity. Because RNA granule formation is a highly regulated process, interactions of a mutation with the normal regulatory machinery may influence the effect of a mutation, potentially resulting in the observed disconnect between intrinsic aggregation propensity and observed foci formation and insolubility in cells. Examining these outliers may ultimately provide a more complete understanding of the factors that affect pathological protein aggregation. Furthermore, it remains to be determined whether intrinsic aggregation propensity (in yeast and *in vitro*) or foci formation in *Drosophila* is more predictive of disease.

These experiments also highlight the risk of using small isolated peptides to examine aggregation propensity. This was seen at two levels: as discussed above, the peptides tested *in*

vitro and in yeast were imperfect predictors of the behavior of full-length proteins; and ZipperDB, which was developed from analysis of 6-amino-acid peptides, was an imperfect predictor of longer ~35-amino-acid peptides. ZipperDB utilizes structure-based prediction, threading sequences into the crystal structure of a 6-amino-acid peptide in an amyloid-like conformation [24]. There are two main types of interactions that stabilize the structure: in-register parallel β-sheet interactions that run the length of the amyloid fibril, and steric zipper packing interactions between β-sheets. Because the structure is based on a 6-amino-acid peptide, the predicted steric zipper interactions are intermolecular, between two identical peptides. Therefore, when ZipperDB predicts the ability to form steric zippers, it is essentially predicting self-complementarity of a peptide. However, in the context of longer peptides, steric zipper interactions may be intramolecular, between different segments of a single peptide [65]. Such intramolecular zippers are seen in a recent high-resolution structure of amyloid-β fibers [66, 67]. These intramolecular zippers are currently not predicted by ZipperDB, potentially explaining why it inaccurately predicts some of the peptides here. However, it should be noted that while strong predicted zipper segments were clearly not sufficient for efficient aggregation in yeast (Figure 2.4, 5), in *Drosophila* (Figure 2.6), or *in vitro* (Figure 2.7), whether they are necessary in *Drosophila* is less clear. The high aggregation propensity of the ΔD276/ΔD290 double mutant *in vitro* (Figure 2.7) and in yeast (Figure 2.5) clearly shows that a strong predicted steric zipper is not necessary in these contexts; however, this double mutant showed low aggregation propensity in *Drosophila* (Figure 2.6), so we cannot rule out the possibility that a strong zipper segment is important in the context of the full-length protein.

REFERENCES

1. Kisilevsky R, Fraser PE. A beta amyloidogenesis: unique, or variation on a systemic theme? Crit Rev Biochem Mol Biol. 1997;32(5):361-404.
2. Sipe JD, Cohen AS. Review: history of the amyloid fibril. Journal of structural biology. 2000;130(2-3):88-98.
3. Liebman SW, Chernoff YO. Prions in yeast. Genetics. 2012;191(4):1041-72.
4. Wickner RB, Shewmaker FP, Bateman DA, Edskes HK, Gorkovskiy A, Dayani Y, et al. Yeast prions: structure, biology, and prion-handling systems. Microbiol Mol Biol Rev. 2015;79(1):1-17.
5. Du Z. The complexity and implications of yeast prion domains. Prion. 2011;5(4):311-6.
6. Cascarina SM, Ross ED. Yeast prions and human prion-like proteins: sequence features and prediction methods. Cell Mol Life Sci. 2014.
7. King OD, Gitler AD, Shorter J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain research. 2012:Epub ahead of print.
8. Oldfield CJ, Dunker AK. Intrinsically disordered proteins and intrinsically disordered protein regions. Annual review of biochemistry. 2014;83:553-84.
9. Ramaswami M, Taylor JP, Parker R. Altered Ribostasis: RNA-Protein Granules in Degenerative Disorders. Cell. 2013;154(4):727-36.
10. Wolozin B. Regulated protein aggregation: stress granules and neurodegeneration. Mol Neurodegener. 2012;7:56.
11. Decker CJ, Teixeira D, Parker R. Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in Saccharomyces cerevisiae. The Journal of cell biology. 2007;179(3):437-49.
12. Gilks N, Kedersha N, Ayodele M, Shen L, Stoecklin G, Dember LM, et al. Stress granule assembly is mediated by prion-like aggregation of TIA-1. Mol Biol Cell. 2004;15(12):5383-98.
13. Reijns MAM, Alexander RD, Spiller MP, Beggs JD. A role for Q/N-rich aggregation-prone regions in P-body localization. Journal of cell science. 2008;121(15):2463-72.
14. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell. 2012;149(4):753-67.
15. Lin Y, Protter DS, Rosen MK, Parker R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. Molecular cell. 2015;60(2):208-19.
16. Molliex A, Temirov J, Lee J, Coughlin M, Kanagaraj AP, Kim HJ, et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. Cell. 2015;163(1):123-33.
17. Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell. 2015;162(5):1066-77.
18. Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature. 2013;495(7442):467-73.

19. March ZM, King OD, Shorter J. Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. Brain research. 2016;1647:9-18.
20. Osherovich LZ, Cox BS, Tuite MF, Weissman JS. Dissection and design of yeast prions. PLoS Biol. 2004;2(4):E86.
21. Toombs JA, Petri M, Paul KR, Kan GY, Ben-Hur A, Ross ED. De novo design of synthetic prion domains. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(17):6519-24.
22. Toombs JA, McCarty BR, Ross ED. Compositional determinants of prion formation in yeast. Mol Cell Biol. 2010;30(1):319-32.
23. Ross ED, Maclea KS, Anderson C, Ben-Hur A. A bioinformatics method for identifying Q/N-rich prion-like domains in proteins. Methods in molecular biology. 2013;1017:219-28.
24. Goldschmidt L, Teng PK, Riek R, Eisenberg D. Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proc Natl Acad Sci USA 2010;107(8):3487-92.
25. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, et al. Structure of the cross-beta spine of amyloid-like fibrils. Nature. 2005;435(7043):773-8.
26. Sherman F. Getting started with yeast. Methods Enzymol. 1991;194:3-21.
27. MacLea KS, Paul KR, Ben-Musa Z, Waechter A, Shattuck JE, Gruca M, et al. Distinct amino acid compositional requirements for formation and maintenance of the [PSI(+)] prion in yeast. Molecular and cellular biology. 2015;35(5):899-911.
28. Gonzalez Nelson AC, Paul KR, Petri M, Flores N, Rogge RA, Cascarina SM, et al. Increasing prion propensity by hydrophobic insertion. PloS one. 2014;9(2):e89286.
29. Paul KR, Hendrich CG, Waechter A, Harman MR, Ross ED. Generating new prions by targeted mutation or segment duplication. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(28):8584-9.
30. Ross ED, Edskes HK, Terry MJ, Wickner RB. Primary sequence independence for prion formation. Proc Natl Acad Sci USA. 2005;102(36):12825-30.
31. Fredrickson EK, Gallagher PS, Clowes Candadai SV, Gardner RG. Substrate recognition in nuclear protein quality control degradation is governed by exposed hydrophobicity that correlates with aggregation and insolubility. The Journal of biological chemistry. 2013;288(9):6130-9.
32. Bagriantsev SN, Kushnirov VV, Liebman SW. Analysis of amyloid aggregates using agarose gel electrophoresis. Methods in enzymology. 2006;412:33-48.
33. Chernoff YO, Lindquist SL, Ono B, Inge-Vechtomov SG, Liebman SW. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. Science. 1995;268(5212):880-4.
34. Wickner RB. [URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae. Science. 1994;264(5158):566-9.
35. Liu JJ, Sondheimer N, Lindquist SL. Changes in the middle region of Sup35 profoundly alter the nature of epigenetic inheritance for the yeast prion [PSI+]. Proc Natl Acad Sci USA. 2002;99 Suppl 4:16446-53.
36. Ter-Avanesyan MD, Dagkesamanskaya AR, Kushnirov VV, Smirnov VN. The SUP35 omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant [psi+] in the yeast Saccharomyces cerevisiae. Genetics. 1994;137(3):671-6.

37. Ter-Avanesyan MD, Kushnirov VV, Dagkesamanskaya AR, Didichenko SA, Chernoff YO, Inge-Vechtomov SG, et al. Deletion analysis of the SUP35 gene of the yeast Saccharomyces cerevisiae reveals two non-overlapping functional regions in the encoded protein. Mol Microbiol. 1993;7(5):683-92.
38. Li L, Lindquist S. Creating a protein-based element of inheritance. Science. 2000;287(5453):661-4.
39. Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. Cell. 2009;137(1):146-58.
40. Osherovich LZ, Weissman JS. Multiple Gln/Asn-rich prion domains confer susceptibility to induction of the yeast [PSI(+)] prion. Cell. 2001;106(2):183-94.
41. Sondheimer N, Lindquist S. Rnq1: an epigenetic modifier of protein function in yeast. Mol Cell. 2000;5(1):163-72.
42. DePace AH, Santoso A, Hillner P, Weissman JS. A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. Cell. 1998;93(7):1241-52.
43. Toombs JA, Liss NM, Cobble KR, Ben-Musa Z, Ross ED. [PSI+] maintenance is dependent on the composition, not primary sequence, of the oligopeptide repeat domain. PloS one. 2011;6(7):e21953.
44. Ross ED, Toombs JA. The effects of amino acid composition on yeast prion formation and prion domain interactions. Prion. 2010;4(2):60-5.
45. Cox BS. PSI, a cytoplasmic suppressor of super-suppressor in yeast. Heredity. 1965;26:211-32.
46. Tuite MF, Mundy CR, Cox BS. Agents that cause a high frequency of genetic change from [psi+] to [psi-] in Saccharomyces cerevisiae. Genetics. 1981;98(4):691-711.
47. Ferreira PC, Ness F, Edwards SR, Cox BS, Tuite MF. The elimination of the yeast [PSI+] prion by guanidine hydrochloride is the result of Hsp104 inactivation. Mol Microbiol. 2001;40(6):1357-69.
48. Jung G, Masison DC. Guanidine hydrochloride inhibits Hsp104 activity in vivo: a possible explanation for its effect in curing yeast prions. Curr Microbiol. 2001;43(1):7-10.
49. Patino MM, Liu JJ, Glover JR, Lindquist S. Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. Science. 1996;273(5275):622-6.
50. Nussbaum-Krammer CI, Park KW, Li L, Melki R, Morimoto RI. Spreading of a prion domain from cell-to-cell by vesicular transport in Caenorhabditis elegans. PLoS genetics. 2013;9(3):e1003351.
51. Park KW, Li L. Cytoplasmic expression of mouse prion protein causes severe toxicity in Caenorhabditis elegans. Biochemical and biophysical research communications. 2008;372(4):697-702.
52. Li S, Zhang P, Freibaum BD, Kim NC, Kolaitis RM, Molliex A, et al. Genetic interaction of hnRNPA2B1 and DNAJB6 in a Drosophila model of multisystem proteinopathy. Human molecular genetics. 2016;25(5):936-50.
53. LeVine H. Quantification of beta-sheet amyloid fibril structures with thioflavin T. Method Enzymol. 1999;309:274-84.

54. McGlinchey RP, Kryndushkin D, Wickner RB. Suicidal [PSI+] is a lethal yeast prion. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(13):5337-41.
55. Kwiatkowski TJ, Jr., Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science. 2009;323(5918):1205-8.
56. Vance C, Rogelj B, Hortobagyi T, De Vos KJ, Nishimura AL, Sreedharan J, et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science. 2009;323(5918):1208-11.
57. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science. 2006;314(5796):130-3.
58. Couthouis J, Hart MP, Shorter J, Dejesus-Hernandez M, Erion R, Oristano R, et al. A yeast functional screen predicts new candidate ALS disease genes. Proceedings of the National Academy of Sciences of the United States of America. 2011;108:20881-90.
59. Couthouis J, Hart MP, Erion R, King OD, Diaz Z, Nakaya T, et al. Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. Human molecular genetics. 2012:Epub ahead of print.
60. Vieira NM, Naslavsky MS, Licinio L, Kok F, Schlesinger D, Vainzof M, et al. A defect in the RNA-processing protein HNRPDL causes limb-girdle muscular dystrophy 1G (LGMD1G). Human molecular genetics. 2014;23(15):4103-10.
61. Johnson BS, Snead D, Lee JJ, McCaffery JM, Shorter J, Gitler AD. TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. The Journal of biological chemistry. 2009;284(30):20329-39.
62. Da Cruz S, Cleveland DW. Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. Curr Opin Neurobiol. 2011;21(6):904-19.
63. Liu Q, Shu S, Wang RR, Liu F, Cui B, Guo XN, et al. Whole-exome sequencing identifies a missense mutation in hnRNPA1 in a family with flail arm ALS. Neurology. 2016;87(17):1763-9.
64. Kelly JW. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. Curr Opin Struct Biol. 1998;8(1):101-6.
65. Paul KR, Ross ED. Controlling the prion propensity of glutamine/asparagine-rich proteins. Prion. 2015;9(5):347-54.
66. Eisenberg DS, Sawaya MR. Implications for Alzheimer's disease of an atomic resolution structure of amyloid-beta(1-42) fibrils. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(34):9398-400.
67. Walti MA, Ravotti F, Arai H, Glabe CG, Wall JS, Bockmann A, et al. Atomic-resolution structure of a disease-relevant A beta(1-42) amyloid fibril. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(34):E4976-E84.

CHAPTER 3: SEQUENCE FEATURES GOVERNING AGGREGATION OR DEGRADATION OF PRION-LIKE PROTEINS[3]

INTRODUCTION

Protein misfolding disorders involve the conversion of native proteins into non-native, deleterious forms. Some misfolded proteins form highly ordered amyloid aggregates, stabilized by intermolecular cross-β sheets. Once formed, these aggregates can convert remaining soluble proteins to the aggregated form via a templated misfolding mechanism [1]. Harmful aggregates must be prevented, sequestered, disassembled, or degraded by cells to prevent disruption of essential cellular functions. Enhanced protein aggregation or impaired clearance of aggregates can lead to neurodegenerative disorders such as Alzheimer's Disease, Parkinson's Disease, Amyotrophic Lateral Sclerosis (ALS), and Huntington's Disease (for review, see [2-9]).

Prion diseases represent a unique sub-class of protein misfolding disorders in which protein aggregates are infectious. Prions can arise *de novo* through protein misfolding events that convert native proteins into the infectious form, or may be acquired through environmental encounter with the infectious form [10]. Although first described in humans, a number of prion proteins were later found to occur in budding yeast [11, 12].

Saccharomyces cerevisiae has since been used extensively as a model organism to study prions [11, 13]. Discovery and characterization of the first two yeast prion proteins, Ure2 and Sup35, revealed that both proteins contain remarkably glutamine/asparagine (Q/N)-rich prion domains [12, 14, 15]. The prion domains also contain relatively few charged and hydrophobic residues. Scrambling experiments demonstrated that the ability of Ure2 and Sup35 to form prions

[3] This chapter is adapted from a manuscript that has been submitted for publication. Kacy R Paul built and tested the phenotypes for the hydrophobic insertion mutations in the hnRNPA2 PrLD and the Sup35 ND (Figure 3.6).

is largely dependent on the amino acid composition of the prion domains, rather than the primary amino acid sequence [16, 17]. Methods for scanning the yeast proteome for additional proteins with similar compositional features resulted in successful identification of new yeast prions [18-20]. To date, nine yeast proteins have been demonstrated to form aggregation-mediated prions [12, 18, 21-27]. The majority of these proteins also contain prion domains with high Q/N content and low charged/hydrophobic content.

Examination of the human proteome with more sophisticated composition-based search algorithms revealed a number of human proteins with "prion-like domains" (PrLDs), defined as domains that compositionally resemble yeast prion domains [5, 28]. Remarkably, many of the top candidates (including TDP-43 [29, 30], FUS [31, 32], EWSR1 [33, 34], and TAF15 [34-36]) had already been associated with protein misfolding disorders, and others (including hnRNPA1 [37], hnRNPA2B1 [37], and TIA1 [38]) were later implicated in protein misfolding disorders. In addition to containing PrLDs, aggregates formed by these proteins are thought to spread throughout an individual in an infectious prion-like manner along a neuroanatomical path that parallels the progression of pathological symptoms [3, 39]. Furthermore, the PrLDs from hnRNPA1 and hnRNPA2B1 are able to support prion activity when substituted in place of the portion of the Sup35 prion domain that is responsible for nucleating prion activity [37, 40].

Although composition-based algorithms have been reasonably effective at identifying candidate yeast prion proteins and potential disease-associated human PrLDs, these algorithms are less effective at predicting the aggregation propensity of these domains or the effects of mutations [41]. One limitation of these methods is that while they assess the frequency with which amino acids occur in prion domains, this may not reflect the importance of each amino acid in prion formation. To address this knowledge gap, we previously used a quantitative

mutagenesis method to score the prion propensity of each amino acid in the context of a Q/N-rich prion domain [42]. Interestingly, although the yeast prions tend to be strikingly Q/N-rich, both glutamine and asparagine were found to have a neutral prion propensity scores [42, 43]. Instead, many of the non-aromatic hydrophobic amino acids (I, M, and V) and the aromatic amino acids (F, W, and Y) were observed to have a strong prion-promoting effect, implicating these amino acids as key nucleators of prion aggregation [42, 44, 45]. We then used these prion-propensity scores to create PAPA, a prion prediction algorithm optimized for yeast prion domains [46, 47]. PAPA is reasonably effective at predicting the prion propensity of Q/N-rich domains, as well as the effects of mutations on prion activity [45, 47, 48].

However, although the composition of the human PrLDs resembles the composition of yeast prion domains, the human PrLDs tend to be less Q/N-rich and contain a higher percentage of serine and glycine (for review, see [41]). Therefore, it is likely that prediction methods developed for yeast prion domains may not be optimized for human PrLDs. To understand how amino acid context (i.e. the starting composition) within PrLDs affects amino acid prion propensities, we sought to determine the prion propensity of each amino acid in the context of the glycine-rich human PrLDs from hnRNPA1 and A2. As in the context of Q/N-rich yeast prion domains, we found that aromatic amino acids were strongly prion-promoting in a G-rich context; however, the non-aromatic hydrophobic amino acids were not strongly prion-promoting. Instead, these non-aromatic hydrophobic amino acids served as a signal for targeted degradation of the G-rich PrLDs. This suggests that aromatic amino acids may have the unique capacity to increase the aggregation propensity of prion or prion-like domains while avoiding efficient detection by protein degradation systems. Furthermore, Q/N residues strongly inhibited degradation of the G-rich PrLDs, suggesting that they may help prevent degradation of prion and prion-like domains.

Indeed, many of the same sequences that led to degradation in the context of the G-rich PrLD had no effect on turnover of a Q/N-rich prion domain. These results broaden our understanding of the proteostatic regulation of aggregation-prone proteins, and shed light on the role of Q/N residues within prion domains.

MATERIALS AND METHODS

Strains and Media

Standard yeast media and methods were used as previously described [77], except that YPD plates contained 0.5% yeast extract rather than the standard 1%. YPAD for all experiments contained the standard 1% yeast extract, as well as 0.02% adenine hemisulfate. Prion curing assays were performed for individual *ADE*$^{+}$ isolates by streaking onto YPD with and without 4mM GuHCl, then re-streaking to YPD to test for loss of the *ADE*$^{+}$ phenotype. In all experiments, yeast were grown at 30^{o}C. The yeast strains used in this study were YER826/pER589 (α *kar1-1 SUQ5 ade2-1 his3 leu2 trp1 ura3 sup35*::KanMx) and YER1161 (α *kar1-1 SUQ5 ade2-1 his3 leu2 trp1 ura3 sup35*::KanMx *pdr5*::*HIS3*). pER589 expresses a truncated version of Sup35 lacking the prion domain (Sup35MC) as the sole copy of Sup35 in the cell. This plasmid was subsequently replaced by plasmid shuffling in order to assay activity of the full-length, randomly mutagenized hnRNP-Sup35 fusions.

Generating Mutant Libraries

The A1-Sup35 and A2-Sup35 fusion libraries were generated in a manner similar to MacLea *et al.* [44]. Briefly, the N-terminal end and C-terminal end of each gene were amplified from a plasmid containing either the A1-Sup35 fusion or the A2-Sup35 fusion (pER595 for hnRNPA1 and pER697 for hnRNPA2; [37]). Oligonucleotides (from Integrated DNA

Technologies) were used to re-amplify the respective products and incorporate a 24-nucleotide degenerate region in which each of the four nucleotides has a 25% probability of occurring at the first two positions of each codon, while C, G, and T each have a 33% probability of occurring at the final position of each codon. The N-terminal and mutagenized C-terminal products, which contain complementary segments, were mixed and re-amplified. The final PCR products were co-transformed with BamHI/HindIII-cut pJ526 into YER826 and plated on synthetic complete media lacking leucine (SC-Leu) to select for cells containing a recombined plasmid. Individual colonies were then picked and stamped onto media containing 5-Fluoroorotic acid (5-FOA) to select for loss of pER589.

Determination of Prion Propensity Scores and Degradation Propensity Scores

After 5-FOA treatment, cells were transferred to YPAD, YPD, and SC-ade. After three days at 30ºC, isolates for which more than 5 colonies appeared on SC-ade were identified and placed in a category (*ADE*$^+$ library) separate from those with fewer than 5 colonies (*ade*$^-$). Randomly selected representative isolates from both groups were sequenced to generate each library. The odds ratio for the *ADE*$^+$ phenotype (OR_A) for each amino acid was determined as follows:

$$OR_A = \left[\frac{f_D}{1-f_D}\right] / \left[\frac{f_N}{1-f_N}\right] \quad (3.1)$$

where f_D represents the per residue frequency of the amino acid among the isolates that were able to grow on SC-ade, and f_N represents the per residue frequency of the amino acid among the naïve isolates (i.e., those that were unable to grow on SC-ade). Final degradation propensity scores for each amino acid (DP_{aa}) were determined as follows:

$$DP_{aa} = \ln(OR_A) \quad (3.2)$$

In addition, prions isolates were identified as previously described [42, 44] and sequenced to generate the prion library. Briefly, the isolates that were initially unable to grow on SC-ade were pooled from the solid YPAD media and re-plated on SC-ade at approximately 10^6 and 10^5 cells per plate. After 3-5 days at 30ºC, individual colonies were streaked onto YPD and YPD plus 4mM GuHCl to assay for prion loss. Odds ratios for prion activity (OR_P) for each amino acid were determined as follows:

$$OR_P = \left[\frac{f_P}{1-f_P}\right] / \left[\frac{f_N}{1-f_N}\right] \quad (3.3)$$

where f_P represents the per residue frequency of the amino acid among the prion-forming isolates. Final prion propensity scores (PP_{aa}) were determined as follows:

$$PP_{aa} = \ln(OR_P) \quad (3.4)$$

Degradation Assays

Cells were diluted to an optical density of 0.75 in liquid YPD media and incubated with shaking at 30ºC for 1hr before treatment with CHX, or DMSO for untreated cells. Where applicable, MG-132 was added to a final concentration of 10µg/mL 1 hr prior to addition of CHX. After the treatment period, the optical densities of all cultures were measured. 10mL of the least-dense culture for each strain were harvested. Based on the optical densities, the approximate number of cells harvested for each of the remaining cultures was normalized to the least-dense culture within each unique strain. Cells were pelleted by centrifugation at 3,000rpm for 5 minutes at 4ºC. Cell pellets were lysed as previously described [55]. 30µL of prepared lysate were loaded onto a 12% polyacrylamide gel, transferred to a PVDF membrane, and probed with a monoclonal antibody that recognizes the C-domain of Sup35 (BE4 [78], kindly made available by Susan Liebman).

RESULTS

Non-Aromatic Hydrophobic Residues Promote Degradation in Human PrLDs but Not in a Yeast Prion Domain

The core PrLDs from the human RNA-binding proteins hnRNPA1 and hnRNPA2 were chosen as model substrates to examine the sequence requirements for aggregation within glycine-rich PrLDs. Both proteins contain a C-terminal G-rich PrLD. Mutations in these domains cause ALS and multisystem proteinopathy in humans, increase their aggregation propensity *in vitro*, and cause muscle degeneration when the proteins are expressed in *Drosophila* [37, 48, 49]. Additionally, the core PrLD from disease-associated mutants of hnRNPA1 and hnRNPA2 can support prion activity when substituted in place of a portion of the prion domain of the yeast prion protein Sup35 [37]. Sup35 contains three functionally distinct domains: an N-terminal prion domain that is necessary and sufficient for formation of prion aggregates; a C-terminal functional domain, which is involved in translation termination; and a highly charged middle domain [15, 50, 51]. The first 40 amino acids of the prion domain, referred to as the nucleation domain (ND), is very Q/N-rich and is responsible for nucleating prion aggregates [40]. When the Sup35 ND is replaced with the core PrLD hnRNPA1 and hnRNPA2, the fusion proteins show mutation-dependent prion activity.

These fusion proteins allowed us to use well-established Sup35 prion detection assays to probe the relationship between amino acid sequence and aggregation activity for the hnRNPA1 and hnRNPA2 PrLDs. Formation of [*PSI*$^{+}$], the prion form of Sup35, can be assayed by monitoring nonsense suppression of the *ade2-1* allele [52]. *ade2-1* mutants are unable to grow on medium lacking adenine (SC-ade), and grow red on medium containing limited adenine (YPD) due to accumulation of a pigment derived from the substrate of the Ade2 enzyme; [*PSI*$^{+}$]

formation results in a low level of read-through of the *ade2-1* premature stop codon, allowing for growth on SC-ade, and formation of white colonies on YPD.

We previously developed a method to quantitatively score the effects of mutations on the Sup35 prion domain with respect to prion activity [42, 44]. We replaced 8-amino acid segments of the prion domain with a random sequence, generating libraries of mutants. Each mutant was expressed as the sole copy of Sup35 in the cell. Randomly mutagenized libraries were plated onto medium lacking adenine to select for mutants that maintained the ability to form [*PSI*$^+$]. We have previously applied this method to various regions of wild-type and scrambled Sup35, including the Sup35 nucleation domain [42, 44]. Therefore, to examine how the sequence requirements for aggregation differ between Q/N-rich and G-rich PrLDs, we repeated this method, mutating the hnRNPA1-Sup35 and hnRNPA2-Sup35 fusions (herein referred to as A1-Sup35 and A2-Sup35 respectively; Figure 3.1A). As targets for mutagenesis, we selected segments with a mixture of predicted aggregation-promoting, aggregation-inhibiting, and neutral amino acids, near the site corresponding to a region previously mutagenized in Sup35 (Figure 3.1B).

Spontaneous [*PSI*$^+$] formation is typically a stochastic and very rare event, occurring at a rate of less than 10^{-6} per generation [53]. By contrast, mutations that reduce Sup35 activity without causing prion aggregation will result in a constitutive *ADE*$^+$ phenotype. Thus, to detect rare prion formation events from among a library of mutants, it is necessary to first eliminate mutants that have a constitutive *ADE*$^+$ phenotype (Figure 3.1A). In previous screens, such constitutive *ADE*$^+$ mutants were relatively rare, comprising ~5% of screened isolates ([42, 44]). Unexpectedly, for the mutagenized A2-Sup35 and A1-Sup35 fusions, approximately 30-40% of the isolates were able to grow in the absence of adenine. These *ADE*$^+$ isolates were not cured by

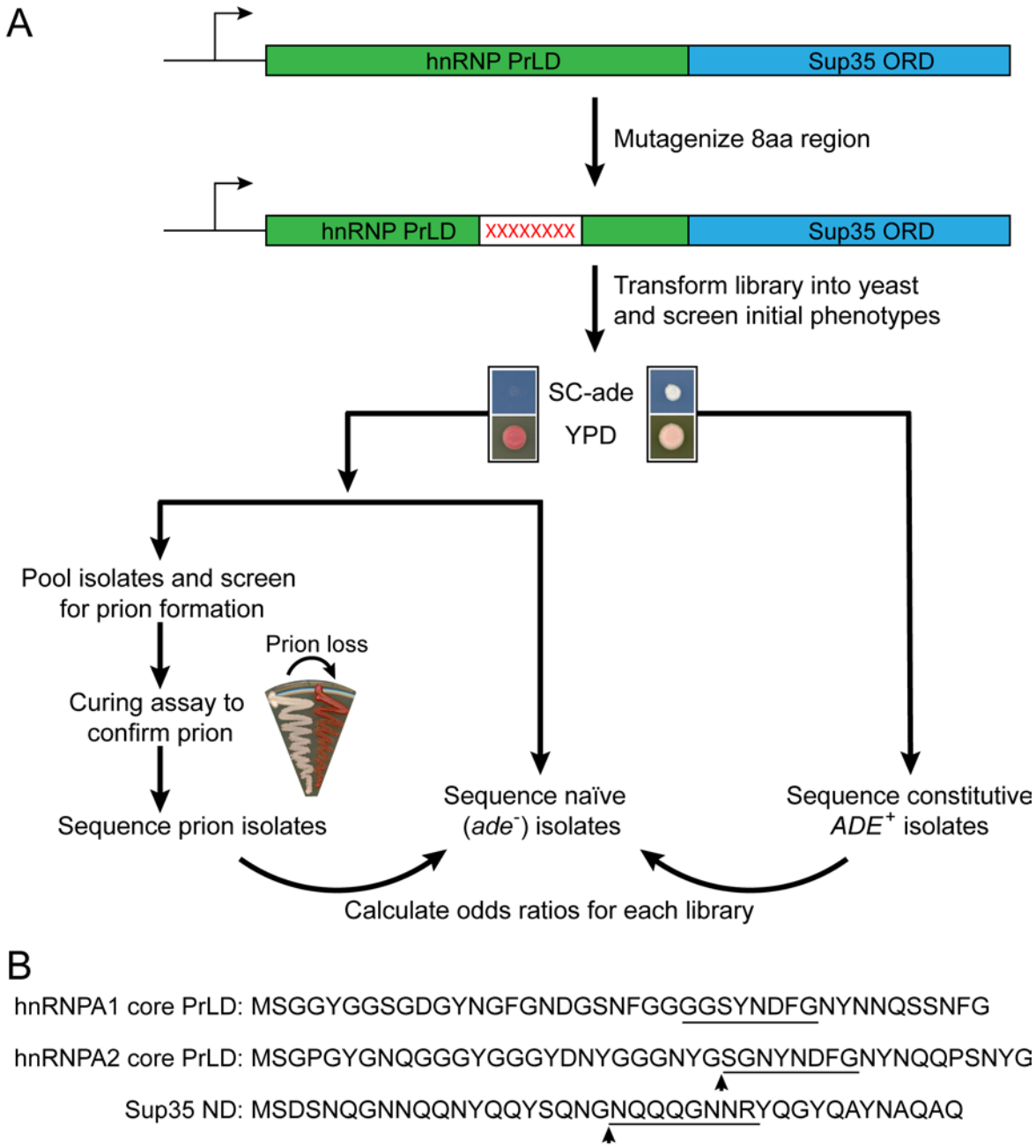


Figure 3.1: Mutagenesis method. (A) The Sup35 prion domain contains an N-terminal Q/N-rich prion nucleation domain, and an oligopeptide repeat domain. The nucleation domain in full-length Sup35 was replaced with the core PrLDs from hnRNPA1 and hnRNPA2. An 8-amino-acid segment in each of the fusion proteins was then randomly mutagenized. Mutants were expressed as the sole copy of Sup35 in the cell. Library members were screened for their initial adenine phenotype, and for the ability to form prions. **(B)** Sequences of the core PrLDs from hnRNPA1 and A2, and the nucleation domain of Sup35. The underlined segments of hnRNPA1 and A2 were mutagenized in this study, while the underlined segment of Sup35 was mutagenized previously [44]. Arrow heads indicate the sites of hydrophobic insertion in Figure 3.6A and 6B.

treatment with 4mM GuHCl, which typically cures [*PSI*$^+$] [54], suggesting that the growth on SC-ade resulted from non-prion-based inactivation of the hnRNP-Sup35 fusion proteins. Additionally, when plasmids expressing A1- or A2-Sup35 mutants were isolated from cells with a constitutive *ADE*$^+$ phenotype and shuffled back into the parent strain, the *ADE*$^+$ phenotype re-appeared, indicating that the phenotype resulted from loss of activity of the Sup35 fusion protein, not from mutations in other cellular proteins.

Therefore, we sought to determine the basis of Sup35 inactivation among these isolates. We sequenced the mutagenized region of the *A1/A2-SUP35* gene from randomly selected *ade*$^-$ and *ADE*$^+$ isolates to determine whether specific sequence features were correlated with the *ADE*$^+$ phenotype. For each amino acid, an odds ratio was calculated (equation 3.1), representing the degree of over-/under-representation of the amino acid among *ADE*$^+$ isolates (Table 3.1). For both libraries, each of the non-aromatic hydrophobic amino acids (I, L, M, and V) were over-represented among *ADE*$^+$ isolates, while glutamine, asparagine, and each of the charged amino acids (D, E, K, and R) were under-represented (Table 3.1; Figure 3.2). Individually, not all of these biases reached the threshold of statistical significance (Table 3.1). Grouping amino acids of similar physical properties can increase statistical significance by effectively increasing sample sizes. When considered as a group, the biases for hydrophobic amino acids, against charged amino acids, and against Q/N were each statistically significant in both libraries ($P<0.01$ in all cases; Table 3.1).

One possible explanation for the *ADE*$^+$ phenotype is that the Sup35 fusions could be poorly expressed or rapidly degraded, causing a decrease in steady state levels of the Sup35 fusion proteins. Exposed hydrophobic patches are known in some cases to trigger protein degradation [55, 56]. To test for this, four representative A2-Sup35 isolates that exemplified the

Table 3.1: Amino acid representation among *ADE*$^+$ and *ade*$^-$ isolates

Amino Acids	hnRNPA2-Sup35 Fusion Library Frequency				hnRNPA1-Sup35 Fusion Library Frequency			
	ADE$^+$ Library	*ade*$^-$ Library	ln(OR_A)[‡]	*P value*	*ADE*$^+$ Library	*ade*$^-$ Library	ln(OR_A)[‡]	*P value*
Valine	0.078	0.023	1.30	7.8×10^{-4}	0.083	0.042	0.74	0.049
Methionine	0.041	0.013	1.21	0.024	0.033	0.011	1.12	0.074
Leucine	0.125	0.045	1.11	1.7×10^{-4}	0.100	0.042	0.94	6.3×10^{-3}
Isoleucine	0.047	0.020	0.89	0.049	0.083	0.033	0.97	9.2×10^{-3}
Tyrosine	0.057	0.040	0.38	0.29	0.038	0.031	0.21	0.65
Alanine	0.051	0.040	0.25	0.58	0.038	0.036	0.039	1.00
Phenylalanine	0.047	0.040	0.18	0.71	0.067	0.042	0.50	0.19
Proline	0.057	0.053	0.10	0.87	0.063	0.064	-0.023	1.00
Threonine	0.057	0.055	0.046	1.00	0.054	0.061	-0.13	0.86
Serine	0.128	0.125	0.031	0.91	0.125	0.119	0.052	0.90
Tryptophan	0.024	0.025	-0.057	1.00	0.033	0.025	0.30	0.62
Aspartic Acid	0.041	0.053	-0.27	0.59	0.021	0.047	-0.85	0.12
Glycine	0.068	0.088	-0.28	0.39	0.033	0.075	-0.85	0.034
Histidine	0.034	0.045	-0.30	0.56	0.050	0.042	0.19	0.69
Lysine	0.020	0.028	-0.31	0.63	0.004	0.033	-2.11	0.019
Cysteine	0.030	0.050	-0.52	0.25	0.063	0.036	0.58	0.17
Arginine	0.054	0.110	-0.77	9.4×10^{-3}	0.058	0.125	-0.84	7.6×10^{-3}
Glutamic Acid	0.014	0.030	-0.81	0.20	0.017	0.019	-0.16	1.00
Glutamine	0.014	0.035	-0.97	0.093	0.017	0.039	-0.87	0.15
Asparagine	0.014	0.085	-1.91	1.5×10^{-5}	0.021	0.078	-1.38	2.9×10^{-3}
Groups of amino acids								
Aromatic (FWY)	0.128	0.105	0.23	0.34	0.138	0.097	0.39	0.15
Charged (DEKR)	0.128	0.220	-0.65	2.0×10^{-3}	0.100	0.225	-0.96	6.9×10^{-5}
Hydrophobic (ILMV)	0.291	0.100	1.30	1.7×10^{-10}	0.300	0.128	1.07	3.5×10^{-7}
Polar (GHST)	0.287	0.313	-0.12	0.50	0.263	0.297	-0.17	0.41
QN	0.027	0.120	-1.59	3.8×10^{-6}	0.038	0.117	-1.22	5.1×10^{-4}

[‡] Odds ratios (OR_A) were calculated as in equation 3.1, and represent the degree of overrepresentation or underrepresentation of each amino acid among constitutive *ADE*$^+$ isolates.

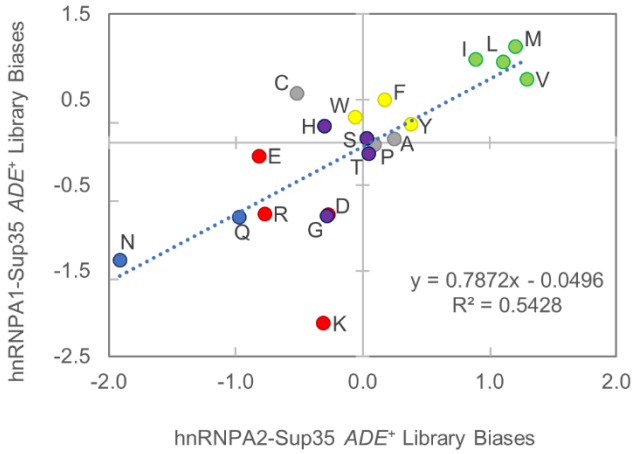


Figure 3.2: Similar amino acid biases govern the *ADE*$^{+}$ phenotype within the A1 and A2 PrLD's. Comparison of the log-odds ratio for each amino acid between the A1 and A2 libraries. The log-odds ratios reflects the degree of over-/under-representation of each amino acid among constitutive *ADE*$^{+}$ isolates, relative to *ade*$^{-}$ isolates. Colors correspond to the amino acid groups in Table 3.1.

amino acid biases among the *ADE*$^+$ library were selected for comparison with randomly selected isolates from the *ade*$^-$ library. The *ADE*$^+$ and *ade*$^-$ phenotypes originally observed for these isolates were confirmed by spotting onto SC-ade, YPD, and YPAD (Figure 3.3A).

Cycloheximide (CHX) globally inhibits translation by preventing translocation of the ribosome along mRNA, providing a convenient tool to assay protein turnover [57]. In the absence of CHX (0 hr), the A2-Sup35 fusion protein levels showed only slight differences among tested strains. However, after treatment with CHX the fusion proteins from *ADE*$^+$ isolates were rapidly degraded (Figure 3.3A). Three of the four *ADE*$^+$ isolates contained little or no detectable A2-Sup35 by 2.5 hours after addition of CHX, while the fourth showed a substantial decrease in A2-Sup35 levels throughout the 5 hour timecourse (Figure 3.3A). By contrast, A2-Sup35 levels remained relatively stable or decreased slightly over a period of 5 hours after addition of CHX for all of the *ade*$^-$ isolates, as well as for the wild-type A2-Sup35 fusion (Figure 3.3A). These results suggest that hydrophobic amino acids trigger degradation of the A2-Sup35 fusions.

Interestingly, random mutagenesis of the Sup35 prion domain yielded very few isolates with the degradation phenotype in the initial screen ([44], and data not shown), suggesting that the Sup35 prion domain can buffer the effects of degradation-promoting peptides. Indeed, when the degradation-promoting 8-amino acid sequences from the A2-Sup35 library were substituted into the corresponding region of the Sup35 prion domain, each of the proteins resulted in phenotypically *ade*$^-$ cells (Figure 3.3B). Furthermore, none of the peptides accelerated the degradation rate of Sup35 over 5 hours. Therefore, while the A2 PrLD is susceptible to the degradation-promoting effects of hydrophobic amino acids, the Sup35 prion domain can mask these effects and resist degradation.

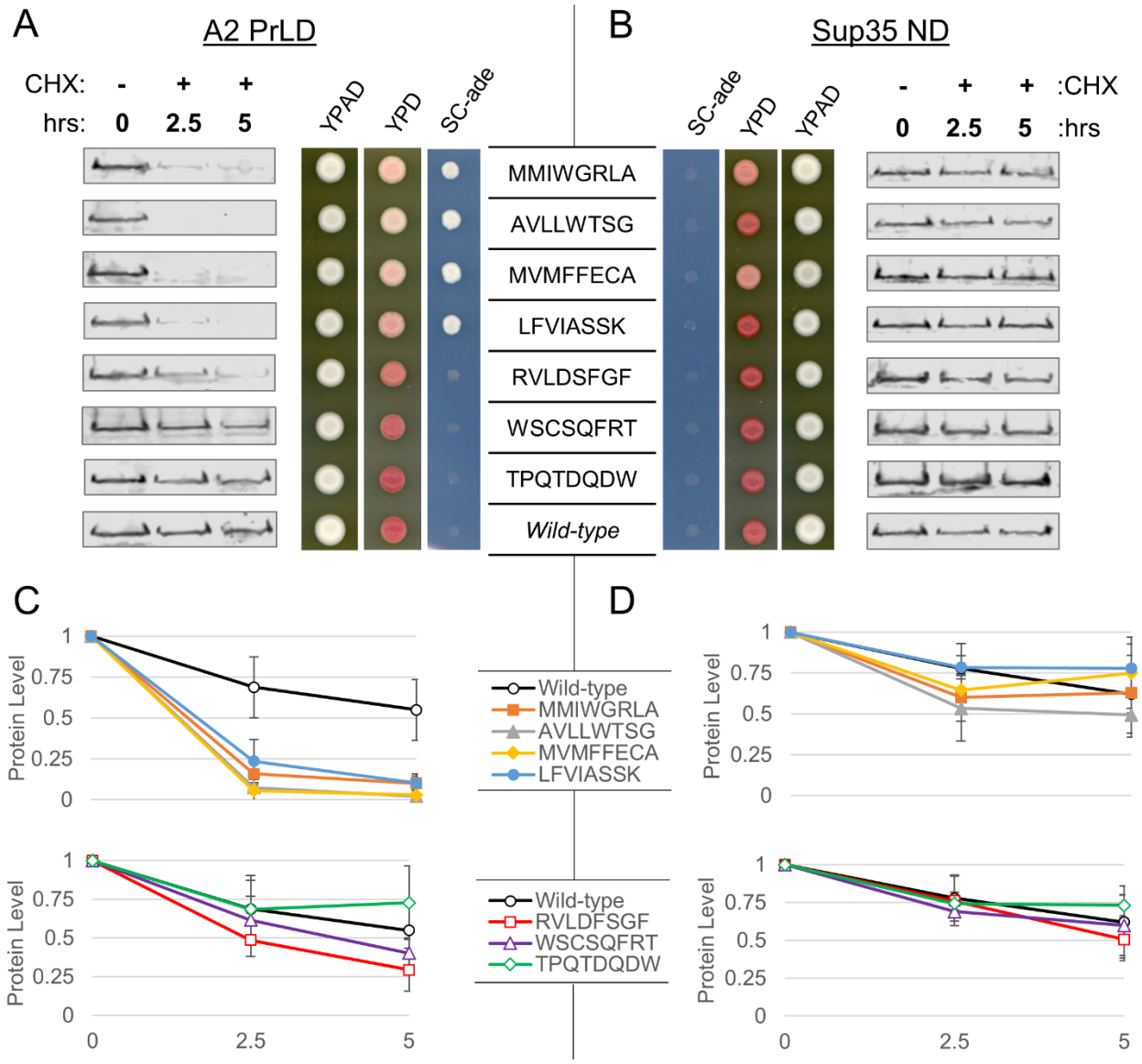


Figure 3.3: Hydrophobic peptides promote degradation of the A2 PrLD but not the Sup35 nucleation domain. (A) The *ADE*$^{+}$ phenotype for A2-Sup35 mutants is associated with increased protein turnover. *ADE*$^{+}$ and *ade*$^{-}$ isolates expressing the indicated A2-Sup35 fusion as the sole copy of Sup35 in the cell were plated on SC-ade and YPD to confirm phenotypes originally observed in the mutagenesis/screening method. Protein turnover was assessed by western blot after treatment with CHX. The *ADE*$^{+}$ phenotype is associated with accelerated degradation of the fusion protein. Wild-type sequences are the respective sequences from wild-type Sup35 and the A2 PrLD. **(B)** In the context of the Sup35 ND, all peptides conferred an *ade*$^{-}$ phenotype and did not accelerate degradation. Western blots were quantified for all A2 PrLD **(C)** and Sup35 ND **(D)** mutants. Data represent means ± SDs (n≥3).

Degradation of the A2 PrLD is Proteasome-Dependent

The ubiquitin-proteasome system is one of the main protein recycling pathways in eukaryotic cells. MG-132, a commonly used proteasome inhibitor, is effective in yeast lacking the pleiotropic drug resistance 5 gene (*Δpdr5*). Therefore, to assess whether degradation of the A2-Sup35 proteins occurs via the proteasome, *PDR5* was deleted from the genome, and the turnover of the A2-Sup35 proteins was assessed in the presence or absence of MG-132. Pre-treatment with MG-132 for 1 hour prior to addition of CHX had no effect on the turnover of the WT A2-Sup35 fusion protein over the 5 hour timecourse (Figure 3.4). By contrast, MG-132 pre-treatment resulted in nearly complete stabilization of the degradation-prone A2-Sup35 fusions over the 5 hour timecourse. These results suggest that the *ADE*$^{+}$ phenotype is due to enhanced turnover of the A2-Sup35 fusion proteins via the ubiquitin-proteasome system.

Degradation-Prone Sequences Can Be Predicted by Amino Acid Composition

Since degradation-promoting sequences failed to cause degradation of Sup35 (Figure 3.3B), we reasoned that some peptide sequences within a previously published Sup35 dataset [44], which did not cause degradation in the context of the Sup35 prion domain, would promote degradation of the A2-Sup35 fusion protein. To identify potential degradation-promoting sequences, each peptide from the library was scored by summing the log-odds ratios from Table 3.1 for the eight amino acids in the mutagenized region. Three sequences predicted to promote degradation (i.e., sequences enriched in non-aromatic hydrophobic residues, with few charged or Q/N residues) were selected from the dataset.

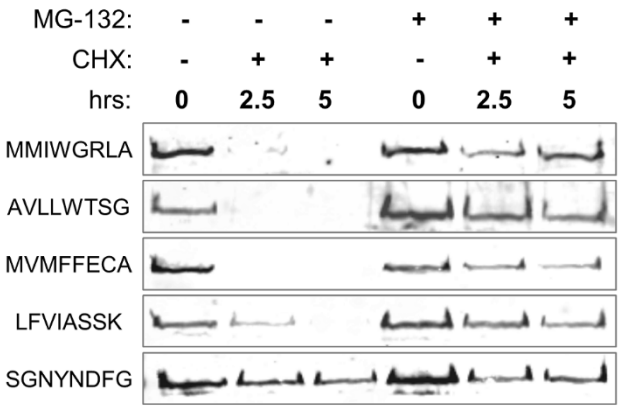


Figure 3.4: Degradation of A2 PrLDs occurs via the ubiquitin-proteasome system. Addition of MG-132 (+) 1hr prior to the addition of CHX prevents degradation of A2-Sup35 fusions (SGNYNDFG is the sequence in the corresponding region of the wild-type A2 PrLD).

When substituted into A2-Sup35, all three predicted degradation-promoting peptides led to enhanced turnover of A2-Sup35 and characteristic degradation phenotypes, albeit to varying degrees (Figure 3.5A). All three strains appeared white on YPD, whereas growth on SC-ade correlated qualitatively with the degree of degradation conferred by each peptide. Additionally, two sequences predicted to have no effect on A2-Sup35 turnover (i.e., sequences enriched in charged and polar residues) were chosen from the same dataset as controls. When substituted into A2-Sup35, neither peptide enhanced degradation, and both strains displayed the associated *ade*$^{-}$ phenotypes (Figure 3.5A). By contrast, all five peptides substituted into the Sup35 prion domain had little effect on turnover and resulted in the characteristic *ade*$^{-}$ phenotype (Figure 3.5B). These results demonstrate that the compositional biases originally observed in the *ADE*$^{+}$ libraries are sufficient to predictively categorize degradation-promoting and degradation-inhibiting sequences.

Hydrophobic Residues Induce Degradation or Prion Formation at a Similar Threshold

The sequences obtained through random mutagenesis are heterogeneous with respect to composition and sequence. To more rigorously define the minimum number of non-aromatic hydrophobic residues required to accelerate the rate of degradation or prion formation, hydrophobic content was progressively increased in WT A2-Sup35 and WT Sup35. Valine, leucine, and methionine (the hydrophobic residues most over-represented in the *A2-Sup35 ADE*$^{+}$ library) were inserted in an alternating fashion adjacent to the region targeted for random mutagenesis (Figure 3.1B and Figure 3.6).

As few as two hydrophobic residues were sufficient to slightly increase turnover of A2-Sup35, as indicated by western blot and the characteristic *ADE*$^{+}$ phenotype (Figure 3.6A). Three

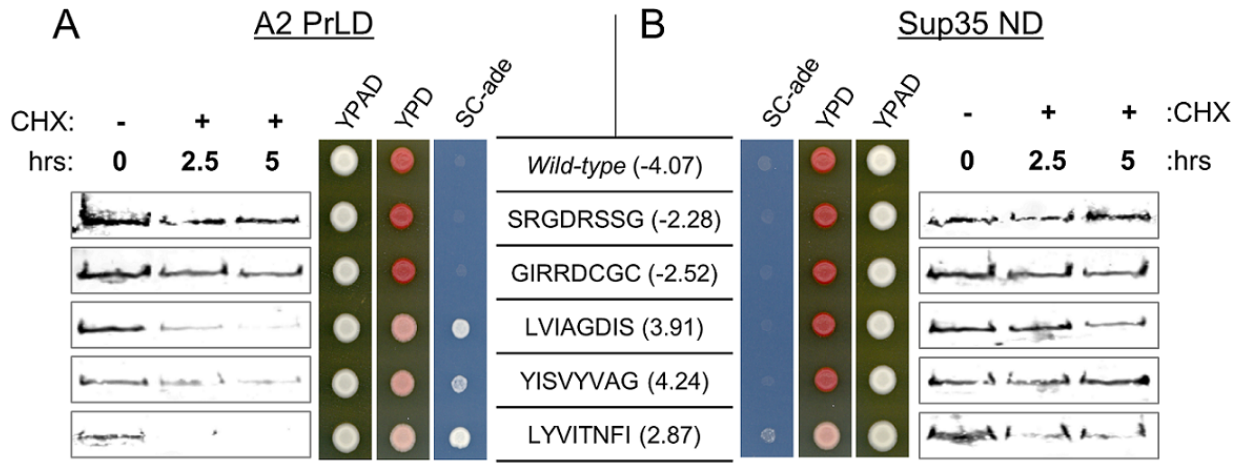


Figure 3.5: Amino acid degradation scores are sufficient to identify degradation-promoting and inhibiting peptides from an independent peptide library. (A) Predicted degradation-promoting (LVIAGDIS, YISVYVAG, and LYVITNFI) or inhibiting (SRGDRSSG and GIRRDCGC) peptides were substituted into the A2 PrLD (numbers in parentheses indicate the sum of the individual amino acid scores derived from the A2 PrLD degradation library). Predicted degradation-promoting peptides led to an *ADE*$^{+}$ phenotype and accelerated degradation of the A2 PrLD. Predicted degradation-inhibiting peptides led to the *ade*$^{-}$ and showed no increase in degradation rate of the A2-Sup35 fusion. **(B)** In the context of the Sup35 ND, all peptides were stable over 5 hrs and conferred a predominantly *ade*$^{-}$ phenotype.

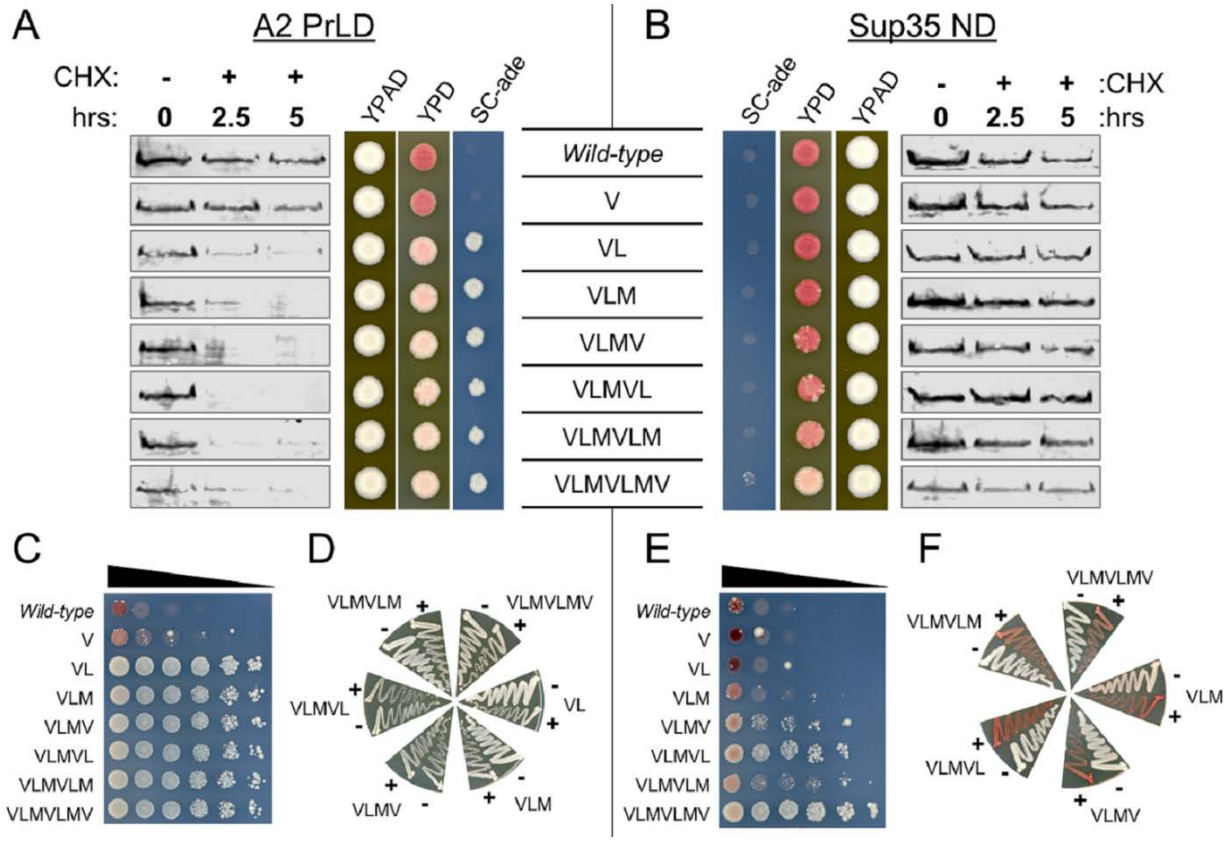


Figure 3.6: Degradation of the A2 PrLD and prion aggregation of Sup35 occur at similar hydrophobic content thresholds. (A) Two or more hydrophobic residues inserted into the A2 PrLD resulted in a robust *ADE*$^+$ phenotype and accelerated degradation of the A2 PrLD. **(B)** Insertion of three or more hydrophobic residues into Sup35 led to a progressive increase in the frequency of white sectors on YPD without affecting Sup35 turnover. **(C)** To quantify the frequency of *ADE*$^+$ colony formation, serial dilutions of cells expressing each A2-Sup35 fusion were plated onto SC-ade. Degradation of the A2 PrLD upon insertion of two or more hydrophobic residues was correlated with a binary-like switch from *ade*$^-$ to *ADE*$^+$. **(D)** *ADE*$^+$ isolates from the A2 PrLD mutants were not curable by GuHCl. To test for curability of the *ADE*$^+$ phenotype, individual *ADE*$^+$ colonies were streaked on YPD (-) or YPD plus 4mM GuHCl (+), and then re-streaked onto YPD to test for loss of the *ADE*$^+$ phenotype. **(E)** Insertion of three or more hydrophobic residues in the Sup35 ND leads to a progressive increase in *ADE*$^+$ growth. **(F)** *ADE*$^+$ isolates from the Sup35 mutants were curable by GuHCl, consistent with the *ADE*$^+$ phenotype resulting from prion formation.

hydrophobic residues further accelerate A2-Sup35 degradation, and four to seven hydrophobic residues caused almost complete loss of A2-Sup35 by 2.5 hours after the addition of CHX. Two or fewer hydrophobic residues inserted into Sup35 resulted in uniform *ade*$^{-}$ phenotypes, whereas three or more hydrophobic residues resulted in the appearance of white sectors, which are classical indications of prion formation (Figure 3.6B). Strikingly, the degree of sectoring increased in a dose-dependent fashion as hydrophobic content increased.

To more accurately quantify the frequency of *ADE*$^{+}$ colony formation by each mutant, all mutants were plated in serial dilution on SC-ade, starting from a higher density than originally assayed. Fewer than two hydrophobic residues in A2-Sup35 resulted in minor growth only at high cell density, whereas two or more hydrophobic residues resulted in robust growth even at very low cell density (Figure 3.6C). Treatment with GuHCl did not alter the color phenotype on YPD (Figure 3.6D), suggesting that the *ADE*$^{+}$ growth was not due to prion formation.

By contrast, three or more hydrophobic residues in Sup35 resulted in a progressive increase in the frequency of *ADE*$^{+}$ colonies, consistent with the progressive increase in sectoring observed on YPD for these mutants (Figure 3.6E). Treating the cells with GuHCl reverted the *ADE*$^{+}$ phenotype to an *ade*$^{-}$ phenotype (Figure 3.6F), confirming that growth on SC-ade was due to the formation of *bona fide* prions.

Collectively, these results suggest that Sup35 is highly resistant to degradation, and therefore sequence features that promote degradation in other contexts promote prion formation in Sup35.

Aromatic Amino Acids Increase Prion Propensity in the Human PrLDs Without Promoting Protein Turnover

Our A2-Sup35 results are consistent with previous studies indicating that the degradation machinery recognizes exposed hydrophobic segments, and that there is a strong correlation between the sequence features that promote aggregation and degradation [55, 56]. We were interested in whether this correlation is absolute, or whether there are sequence features that can promote aggregation of the G-rich PrLDs without promoting degradation. Our A1- and A2-Sup35 fusions provide a useful system for comparing the sequence requirements for degradation versus aggregation. To determine whether specific sequence features could promote prion aggregation without triggering degradation, isolates with an initial *ade*$^{-}$ phenotype were plated onto medium lacking adenine to screen for the ability to spontaneously form prions (Figure 3.1A). [*PRION*$^{+}$] isolates were confirmed by curing with GuHCl, and the mutagenized *A1/A2-SUP35* gene in each was sequenced. Sequences from each library were pooled, and the prion propensity scores for each amino acid were determined, as described previously ([42, 44]; equation 3.4).

Interestingly, aromatic amino acids were significantly more common among [*PRION*$^{+}$] isolates compared to the *ade*$^{-}$ isolates from both the A2-Sup35 and A1-Sup35 libraries, while the non-aromatic hydrophobic residues were approximately equally represented among [*PRION*$^{+}$] and *ade*$^{-}$ isolates (Table 3.2). By contrast, both the aromatic and the non-aromatic hydrophobic amino acid groups had a strong prion-promoting effect in the context of the Q/N-rich Sup35 nucleation domain [42, 44]. Furthermore, Q/N residues were significantly under-represented among A2-Sup35 [*PRION*$^{+}$] isolates, although their effects were mixed among A1-Sup35 [*PRION*$^{+}$] isolates. Together, these results suggest that a hitherto unappreciated property of

Table 3.2: Amino acid representation among [*PRION*$^+$] and *ade*$^-$ isolates

Amino Acids	hnRNPA2-Sup35 Fusion Library Frequency				hnRNPA1-Sup35 Fusion Library Frequency			
	[*PRION*$^+$] Library	*ade*$^-$ Library	ln(OR$_P$) ǂ	*P value*	[*PRION*$^+$] Library	*ade*$^-$ Library	ln(OR$_P$) ǂ	*P value*
Phenylalanine	0.093	0.040	0.90	0.011	0.069	0.042	0.53	0.13
Tyrosine	0.088	0.040	0.84	0.018	0.069	0.031	0.86	0.028
Tryptophan	0.037	0.025	0.41	0.45	0.020	0.025	-0.24	0.80
Threonine	0.079	0.055	0.38	0.30	0.049	0.061	-0.23	0.61
Valine	0.032	0.023	0.38	0.44	0.053	0.042	0.25	0.58
Isoleucine	0.028	0.020	0.34	0.58	0.026	0.033	-0.24	0.65
Alanine	0.051	0.040	0.25	0.54	0.036	0.036	0.0021	1.00
Proline	0.060	0.053	0.14	0.71	0.026	0.064	-0.93	0.026
Methionine	0.014	0.013	0.11	1.00	0.030	0.011	1.00	0.10
Histidine	0.046	0.045	0.030	1.00	0.056	0.042	0.31	0.47
Glycine	0.083	0.088	-0.053	1.00	0.049	0.075	-0.45	0.20
Glutamic Acid	0.028	0.030	-0.079	1.00	0.007	0.019	-1.10	0.19
Leucine	0.042	0.045	-0.080	1.00	0.053	0.042	0.25	0.58
Arginine	0.093	0.110	-0.19	0.58	0.092	0.125	-0.34	0.21
Serine	0.093	0.125	-0.34	0.29	0.145	0.119	0.22	0.36
Glutamine	0.023	0.035	-0.43	0.47	0.023	0.039	-0.54	0.27
Cysteine	0.032	0.050	-0.45	0.41	0.039	0.036	0.093	0.84
Aspartic Acid	0.032	0.053	-0.50	0.31	0.033	0.047	-0.38	0.43
Asparagine	0.037	0.085	-0.88	0.029	0.109	0.078	0.37	0.18
Lysine	0.009	0.028	-1.11	0.15	0.016	0.033	-0.72	0.22
Groups of amino acids								
Aromatic (FWY)	0.218	0.105	0.86	2.7×10^{-4}	0.158	0.097	0.55	0.025
Charged (DEKR)	0.162	0.220	-0.38	0.092	0.148	0.225	-0.51	0.013
Hydrophobic (ILMV)	0.116	0.100	0.16	0.58	0.161	0.128	0.27	0.22
Polar (GHST)	0.301	0.313	-0.054	0.78	0.299	0.297	0.010	1.00
Q/N	0.060	0.120	-0.76	0.017	0.132	0.117	0.14	0.64

ǂ Odds ratios (OR$_P$) were calculated as in equation 3.3, and represent the degree of overrepresentation or underrepresentation of each amino acid among prion isolates.

aromatic amino acids is the unique ability to promote protein aggregation of prion and prion-like domains, while avoiding detection by the degradation machinery.

Indeed, while there is a statistically significant (P=0.008 by Spearman rank analysis) correlation between the prion propensity (as scored by PAPA) of each amino acid and its propensity to promote degradation (Figure 3.7), there are five amino acids which have substantially lower degradation propensities than would be predicted by their prion propensities: the three aromatic amino acids, glutamine, and asparagine. Strikingly, these amino acids are all overrepresented among yeast prion proteins. While both aromatic and non-aromatic hydrophobic amino acids strongly promote prion formation [42], candidate prion domains with prion activity tend to contain more aromatic residues and fewer aliphatic residues than candidate prion domains with no detectable prion activity [44]. Likewise, although serine, glycine, threonine, glutamine, and asparagine each promote intrinsic disorder and have similar prion propensities [42], Q/N residues are far more common among yeast prion domains.

Collectively, these results suggest a possible explanation for the amino acid biases observed among yeast prion domains. Many components of protein quality control systems act specifically to antagonize protein aggregation. Therefore, proteins that form observable protein aggregates must possess mechanisms to avoid or outcompete antagonistic proteostasis machinery. Yeast prion domains tend to favor amino acids that promote aggregation while being poorly recognized by the degradation machinery.

Q/N Residues Stabilize Sup35

These results may also provide an explanation for Sup35's resistance to degradation. Q/N residues were among the lowest scoring amino acids in the degradation libraries. The human

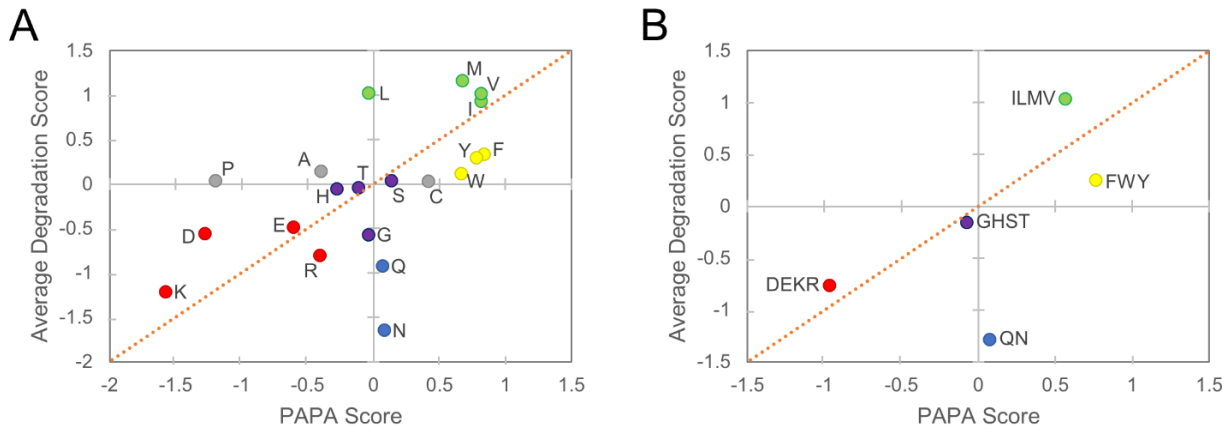


Figure 3.7: Yeast prion domains are enriched in amino acids that are prion-prone but not degradation-promoting. Average degradation scores from the A1 PrLD and A2 PrLD libraries are plotted against yeast prion propensity scores for individual amino acids **(A)** or amino acid groups **(B)**. Within native yeast prion domains, commonly occurring amino acids (Q, N, and aromatic amino acids) exhibit a combination of high prion propensity and low degradation propensity. Colors correspond to the amino acid groups in Table 3.1 and Figure 3.2.

PrLDs and the Sup35 ND differ most notably in their Q/N content; the Sup35 ND contains a much higher percentage of Q/N-residues, while the A1 and A2 core PrLDs are more G-rich. This points to the simple hypothesis that the high Q/N-content of the Sup35 ND may protect highly aggregation-prone features from recognition by components of the proteostasis machinery. To test this hypothesis, two of the degradation-prone members of the A2 library and their Sup35 counterparts were chosen as initial substrates for mutagenesis. To examine the relationship between Q/N content and degradation, we mutated some or all of the Q/N's in the Sup35 nucleation domain to G's (Figure 3.8A). Similarly, we mutated some or all of the G's in the A2 PrLDs to Q/N.

The rate of degradation of Sup35 correlated with Q/N-content in a dose-dependent manner. Partial substitution of Q/N-residues for G's increased the turnover rate of each Sup35 derivative and resulted in the emergence of the degradation phenotype (Figure 3.8B). Substitution of the remaining Q/N's for G's further enhanced the rate of Sup35 degradation. However, other sequence features of the A2 PrLD besides Q/N content must contribute to its sensitivity to degradation, as replacing some or all of the G's in the A2-Sup35 derivatives with Q/N residues resulted in only partial stabilization of the proteins (Figure 3.8B). Therefore, in addition to their role in prion formation, Q/N residues help prion domains resist degradation by intracellular anti-aggregation systems.

DISCUSSION

Protein misfolding is a selective challenge faced by all cellular life. Misfolded proteins can result in proteotoxicity, either through loss-of-function of the native protein or through a

A

Domain	Sequence
Original Sup35 ND	MSDSNQGNNQQNYQQYSQNGxxxxxxxxYQGYQAYNAQAQ
Sup35 ND Partial Substitution	MSDSGQGGGGGNYQGYSGGGxxxxxxxxYGGYQAYGAGAG
Sup35 ND Full Substitution	MSDSGGGGGGGGYGGYSGGGxxxxxxxxYGGYGAYGAGAG
Original A2 PrLD	MSGPGYGNQGGGYGGGYDNYGGGNYGxxxxxxxxNYNQQPSNYG
A2 PrLD Partial Substitution	MSGPQYQNQQQNYQGQYDNYQQGNYQxxxxxxxxNYNQQPSNYQ
A2 PrLD Full Substitution	MSQPQYQNQQQNYQNQYDNYQQQNYQxxxxxxxxNYNQQPSNYQ

B

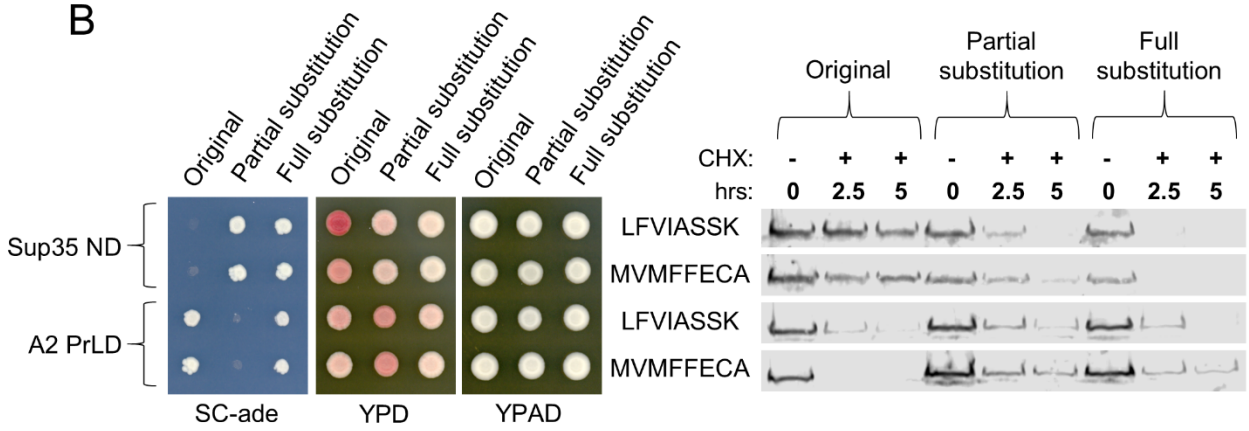


Figure 3.8: High Q/N content of Sup35 antagonizes degradation. (A) Sequences of the partial or full Q/N/G substitution constructs. Glycines are indicated in red, while glutamine and asparagine are in green. **(B)** Partial or full substitution of Q/N residues for G within the Sup35 nucleation domain resulted in the *ADE*$^{+}$ degradation phenotype (*left*), and a step-wise increase in degradation rate (*right*). Partial substitution of G residues for Q/N residues within the A2 PrLD resulted in an *ade*$^{-}$ phenotype (*left*), and corresponding decrease in degradation rate (*right*). Full substitution of remaining G residues for QN residues had mixed effects.

toxic gain-of-function of the misfolded species. To address these selective challenges, eukaryotic cells possess extensive proteostasis machinery, which constitutively act to procure and maintain pools of natively folded proteins. The proteostasis machinery broadly consists of three main systems: 1) the protein chaperone network, which aids in nascent protein folding as well as the re-folding of partially or fully denatured proteins, 2) the ubiquitin-proteasome system, and 3) the autophagy system, which together aid in the destruction of aged, terminally misfolded, or aggregated proteins (for review, see [58, 59]).

Despite the constant surveillance of protein quality control systems, numerous diseases result from misfolding and aggregation of proteins. Additionally, a variety of proteins form functional aggregates that are involved in the regulation of various cellular processes [60-62]. Therefore, understanding how the proteostasis machinery detects misfolded proteins, and how some aggregation-prone proteins evade this detection, may provide insight into both functional and pathogenic aggregation.

Previous work has shown that one way that proteostasis network components achieve specificity for misfolded proteins is by recognizing patches of solvent-exposed hydrophobicity [63-71]. Hydrophobic patches are generally buried in the interior of folded proteins [72], so exposed hydrophobicity can act as a signal of protein misfolding. Additionally, there is a strong correlation between hydrophobicity and aggregation propensity [73], so recognizing exposed hydrophobicity would seem to be an effective mechanism to recognize aggregation-prone misfolded proteins.

One well-characterized example that uses this mechanism is the yeast E3 ubiquitin ligase San1, a nuclear protein involved in the ubiquitin-proteasome degradation system [74]. San1 is a largely disordered protein that is particularly adept at targeting toxic misfolded proteins for

degradation [75], primarily by recognizing exposed hydrophobic residues in substrates [56]. Interestingly, San1 recognition of these substrates tends to correlate with their insolubility, demonstrating the effectiveness of targeting hydrophobicity to prevent protein aggregation [55]. It should be noted that degradation of the A1- and A2-Sup35 fusions was independent of San1 (data not shown), so additional work will be required to identify the cellular factors responsible for recognition and degradation of these PrLDs.

Although our data is generally consistent with the idea that exposed hydrophobic amino acids promote recognition by the proteostasis machinery, our results provide some additional unexpected insights. First, in contrast to what has been proposed for San1, we show that aggregation propensity and recognition by the proteostasis machinery can be uncoupled in a composition-dependent manner. The aromatic amino acids within PrLDs increase aggregation propensity without substantially enhancing recognition by the proteostasis machinery. Second, the ability of the proteostasis machinery to recognize hydrophobic patches was highly context dependent.

Our results raise the intriguing possibility that Q/N residues may potentiate the aggregation of prion domains in part by protecting aggregation-prone features from the proteostasis machinery. Although the exact mechanism by which Sup35 resists degradation is unclear, high Q/N-content appears to play an important role. The Sup35 prion domain and the A1/A2 PrLDs are each predicted to be intrinsically disordered. However, the Sup35 prion domain is thought to form a collapsed but disordered structure [76], which may hide hydrophobic patches from the proteostasis machinery. High Q/N content may help mask hydrophobic patches by promoting a collapsed but disordered structure, or by shielding hydrophobic amino acids within these structures. Alternatively, rather than preventing the initial

recognition of hydrophobic patches by the proteostasis machinery, Q/N residues may inhibit a downstream step in the subsequent events leading to degradation. Furthermore, features of the Sup35 prion domain besides Q/N content seem well-suited to avoid detection by the degradation machinery. We previously showed that six amino acids are highly prion-promoting: F, Y, W, I, V, and M [42]. Sup35 contains 23 of these highly prion-promoting amino acids, yet all except the initiating methionine are aromatic. Additionally, the prion-promoting amino acids are well-dispersed. There is only one position where two occur adjacent to each other, and almost all have adjacent Q/N residues. Thus, Sup35 avoids the hydrophobic patches that could trigger degradation.

Numerous labs have made extensive progress in defining how the amino acid sequence of a protein affects its intrinsic aggregation propensity. However, our results highlight that intrinsic aggregation propensity is only a small piece of the puzzle. A more complete understanding of functional and pathogenic protein aggregation requires a clearer view of how amino acid sequence affects interactions with other cellular proteins. Our results provide one unexpected piece to this puzzle, demonstrating that specific sequence features can promote protein aggregation, while simultaneously hiding from the proteostasis machinery.

REFERENCES

1. Walker LC, LeVine H, 3rd. Corruption and spread of pathogenic proteins in neurodegenerative diseases. The Journal of biological chemistry. 2012;287(40):33109-15.
2. Brundin P, Melki R, Kopito R. Prion-like transmission of protein aggregates in neurodegenerative diseases. Nature reviews Molecular cell biology. 2010;11(4):301-7.
3. Costanzo M, Zurzolo C. The cell biology of prion-like spread of protein aggregates: mechanisms and implication in neurodegeneration. Biochem J. 2013;452(1):1-17.
4. Holmes BB, Diamond MI. Cellular mechanisms of protein aggregate propagation. Curr Opin Neurol. 2012;25(6):721-6.
5. King OD, Gitler AD, Shorter J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain Res. 2012:Epub ahead of print.
6. Li YR, King OD, Shorter J, Gitler AD. Stress granules as crucibles of ALS pathogenesis. The Journal of cell biology. 2013;201(3):361-72.
7. Ling SC, Polymenidou M, Cleveland DW. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. Neuron. 2013;79(3):416-38.
8. Ramaswami M, Taylor JP, Parker R. Altered Ribostasis: RNA-Protein Granules in Degenerative Disorders. Cell. 2013;154(4):727-36.
9. Taylor JP, Brown RH, Jr., Cleveland DW. Decoding ALS: from genes to mechanism. Nature. 2016;539(7628):197-206.
10. Prusiner SB. Prions. Proceedings of the National Academy of Sciences of the United States of America. 1998;95(23):13363-83.
11. Liebman SW, Chernoff YO. Prions in yeast. Genetics. 2012;191(4):1041-72.
12. Wickner RB. [URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae. Science. 1994;264(5158):566-9.
13. Wickner RB. Yeast and Fungal Prions. Cold Spring Harbor perspectives in biology. 2016;8(9).
14. Masison DC, Wickner RB. Prion-inducing domain of yeast Ure2p and protease resistance of Ure2p in prion-containing cells. Science. 1995;270(5233):93-5.
15. Ter-Avanesyan MD, Dagkesamanskaya AR, Kushnirov VV, Smirnov VN. The SUP35 omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant [psi+] in the yeast Saccharomyces cerevisiae. Genetics. 1994;137(3):671-6.
16. Ross ED, Baxa U, Wickner RB. Scrambled Prion Domains Form Prions and Amyloid. Mol Cell Biol. 2004;24(16):7206-13.
17. Ross ED, Edskes HK, Terry MJ, Wickner RB. Primary sequence independence for prion formation. Proc Natl Acad Sci USA. 2005;102(36):12825-30.
18. Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. Cell. 2009;137(1):146-58.
19. Harrison PM, Gerstein M. A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes. Genome Biol. 2003;4(6):R40.
20. Michelitsch MD, Weissman JS. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. Proc Natl Acad Sci USA. 2000;97(22):11910-5.

21. Derkatch IL, Bradley ME, Hong JY, Liebman SW. Prions affect the appearance of other prions: the story of [PIN(+)]. Cell. 2001;106(2):171-82.
22. Du Z, Park KW, Yu H, Fan Q, Li L. Newly identified prion linked to the chromatin-remodeling factor Swi1 in Saccharomyces cerevisiae. Nat Genet. 2008;40(4):460-5.
23. Halfmann R, Wright J, Alberti S, Lindquist S, Rexach M. Prion formation by a yeast GLFG nucleoporin. Prion. 2012;6(4).
24. Patel BK, Gavin-Smyth J, Liebman SW. The yeast global transcriptional co-repressor protein Cyc8 can propagate as a prion. Nature cell biology. 2009;11(3):344-9.
25. Rogoza T, Goginashvili A, Rodionova S, Ivanov M, Viktorovskaya O, Rubel A, et al. Non-Mendelian determinant [ISP+] in yeast is a nuclear-residing prion form of the global transcriptional regulator Sfp1. Proc Natl Acad Sci USA. 2010;107(23):10573-7.
26. Sondheimer N, Lindquist S. Rnq1: an epigenetic modifier of protein function in yeast. Mol Cell. 2000;5(1):163-72.
27. Suzuki G, Shimazu N, Tanaka M. A yeast prion, Mod5, promotes acquired drug resistance and cell survival under environmental stress. Science. 2012;336(6079):355-9.
28. Lancaster AK, Nutter-Upham A, Lindquist S, King OD. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. Bioinformatics. 2014;30(17):2501-2.
29. Arai T, Hasegawa M, Akiyama H, Ikeda K, Nonaka T, Mori H, et al. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochemical and biophysical research communications. 2006;351(3):602-11.
30. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science. 2006;314(5796):130-3.
31. Kwiatkowski TJ, Jr., Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science. 2009;323(5918):1205-8.
32. Vance C, Rogelj B, Hortobagyi T, De Vos KJ, Nishimura AL, Sreedharan J, et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science. 2009;323(5918):1208-11.
33. Couthouis J, Hart MP, Erion R, King OD, Diaz Z, Nakaya T, et al. Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. Human molecular genetics. 2012:Epub ahead of print.
34. Neumann M, Bentmann E, Dormann D, Jawaid A, DeJesus-Hernandez M, Ansorge O, et al. FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. Brain. 2011;134(Pt 9):2595-609.
35. Couthouis J, Hart MP, Shorter J, Dejesus-Hernandez M, Erion R, Oristano R, et al. A yeast functional screen predicts new candidate ALS disease genes. Proceedings of the National Academy of Sciences of the United States of America. 2011;108:20881-90.
36. Ticozzi N, Vance C, Leclerc AL, Keagle P, Glass JD, McKenna-Yasek D, et al. Mutational analysis reveals the FUS homolog TAF15 as a candidate gene for familial amyotrophic lateral sclerosis. Am J Med Genet B Neuropsychiatr Genet. 2011;156B(3):285-90.

37. Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature. 2013;495(7442):467-73.
38. Klar J, Sobol M, Melberg A, Mabert K, Ameur A, Johansson AC, et al. Welander distal myopathy caused by an ancient founder mutation in TIA1 associated with perturbed splicing. Hum Mutat. 2013;34(4):572-7.
39. Ravits J. Focality, stochasticity and neuroanatomic propagation in ALS pathogenesis. Exp Neurol. 2014;262 Pt B:121-6.
40. Osherovich LZ, Cox BS, Tuite MF, Weissman JS. Dissection and design of yeast prions. PLoS Biol. 2004;2(4):E86.
41. Cascarina SM, Ross ED. Yeast prions and human prion-like proteins: sequence features and prediction methods. Cell Mol Life Sci. 2014.
42. Toombs JA, McCarty BR, Ross ED. Compositional determinants of prion formation in yeast. Mol Cell Biol. 2010;30(1):319-32.
43. Ross ED, Toombs JA. The effects of amino acid composition on yeast prion formation and prion domain interactions. Prion. 2010;4(2):60-5.
44. MacLea KS, Paul KR, Ben-Musa Z, Waechter A, Shattuck JE, Gruca M, et al. Distinct amino acid compositional requirements for formation and maintenance of the [PSI(+)] prion in yeast. Molecular and cellular biology. 2015;35(5):899-911.
45. Gonzalez Nelson AC, Paul KR, Petri M, Flores N, Rogge RA, Cascarina SM, et al. Increasing prion propensity by hydrophobic insertion. PloS one. 2014;9(2):e89286.
46. Ross ED, Maclea KS, Anderson C, Ben-Hur A. A bioinformatics method for identifying Q/N-rich prion-like domains in proteins. Methods in molecular biology. 2013;1017:219-28.
47. Toombs JA, Petri M, Paul KR, Kan GY, Ben-Hur A, Ross ED. De novo design of synthetic prion domains. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(17):6519-24.
48. Paul KR, Molliex A, Cascarina S, Boncella AE, Taylor JP, Ross ED. Effects of mutations on the aggregation propensity of the human prion-like protein hnRNPA2B1. Molecular and cellular biology. 2017;37(8) e00652-16.
49. Li S, Zhang P, Freibaum BD, Kim NC, Kolaitis RM, Molliex A, et al. Genetic interaction of hnRNPA2B1 and DNAJB6 in a Drosophila model of multisystem proteinopathy. Human molecular genetics. 2016;25(5):936-50.
50. Liu JJ, Sondheimer N, Lindquist SL. Changes in the middle region of Sup35 profoundly alter the nature of epigenetic inheritance for the yeast prion [PSI+]. Proc Natl Acad Sci USA. 2002;99 Suppl 4:16446-53.
51. Ter-Avanesyan MD, Kushnirov VV, Dagkesamanskaya AR, Didichenko SA, Chernoff YO, Inge-Vechtomov SG, et al. Deletion analysis of the SUP35 gene of the yeast Saccharomyces cerevisiae reveals two non-overlapping functional regions in the encoded protein. Mol Microbiol. 1993;7(5):683-92.
52. Cox BS. PSI, a cytoplasmic suppressor of super-suppressor in yeast. Heredity. 1965;26:211-32.
53. Lancaster AK, Bardill JP, True HL, Masel J. The spontaneous appearance rate of the yeast prion [PSI+] and its implications for the evolution of the evolvability properties of the [PSI+] system. Genetics. 2010;184(2):393-400.

54. Tuite MF, Mundy CR, Cox BS. Agents that cause a high frequency of genetic change from [psi+] to [psi-] in Saccharomyces cerevisiae. Genetics. 1981;98(4):691-711.
55. Fredrickson EK, Gallagher PS, Clowes Candadai SV, Gardner RG. Substrate recognition in nuclear protein quality control degradation is governed by exposed hydrophobicity that correlates with aggregation and insolubility. The Journal of biological chemistry. 2013;288(9):6130-9.
56. Fredrickson EK, Rosenbaum JC, Locke MN, Milac TI, Gardner RG. Exposed hydrophobicity is a key determinant of nuclear quality control degradation. Mol Biol Cell. 2011;22(13):2384-95.
57. Schneider-Poetsch T, Ju J, Eyler DE, Dang Y, Bhat S, Merrick WC, et al. Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nat Chem Biol. 2010;6(3):209-17.
58. Hipp MS, Park SH, Hartl FU. Proteostasis impairment in protein-misfolding and -aggregation diseases. Trends Cell Biol. 2014;24(9):506-14.
59. Labbadia J, Morimoto RI. The biology of proteostasis in aging and disease. Annual review of biochemistry. 2015;84:435-64.
60. Falsone A, Falsone SF. Legal but lethal: functional protein aggregation at the verge of toxicity. Frontiers in cellular neuroscience. 2015;9:45.
61. Fowler DM, Koulov AV, Balch WE, Kelly JW. Functional amyloid--from bacteria to humans. Trends in biochemical sciences. 2007;32(5):217-24.
62. Silva JL, Cordeiro Y. The "Jekyll and Hyde" Actions of Nucleic Acids on the Prion-like Aggregation of Proteins. The Journal of biological chemistry. 2016;291(30):15482-90.
63. Flynn GC, Pohl J, Flocco MT, Rothman JE. Peptide-binding specificity of the molecular chaperone BiP. Nature. 1991;353(6346):726-30.
64. Karagoz GE, Duarte AMS, Akoury E, Ippel H, Biernat J, Luengo TM, et al. Hsp90-Tau Complex Reveals Molecular Basis for Specificity in Chaperone Action. Cell. 2014;156(5):963-74.
65. Karagoz GE, Rudiger SGD. Hsp90 interaction with clients. Trends in biochemical sciences. 2015;40(2):117-25.
66. Li YL, Gao XF, Chen LL. GroEL Recognizes an Amphipathic Helix and Binds to the Hydrophobic Side. Journal of Biological Chemistry. 2009;284(7):4324-31.
67. Rudiger S, Germeroth L, Schneider-Mergener J, Bukau B. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. The EMBO journal. 1997;16(7):1501-7.
68. Rudiger S, Schneider-Mergener J, Bukau B. Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. Embo Journal. 2001;20(5):1042-50.
69. Saio T, Guan X, Rossi P, Economou A, Kalodimos CG. Structural Basis for Protein Antiaggregation Activity of the Trigger Factor Chaperone. Science. 2014;344(6184): 1250494.
70. Willmund F, del Alamo M, Pechmann S, Chen T, Albanese V, Dammer EB, et al. The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis. Cell. 2013;152(1-2):196-209.
71. Rudiger S, Buchberger A, Bukau B. Interaction of Hsp70 chaperones with substrates. Nature structural biology. 1997;4(5):342-9.

72. Moelbert S, Emberly E, Tang C. Correlation between sequence hydrophobicity and surface-exposure pattern of database proteins. Protein Sci. 2004;13(3):752-62.
73. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature. 2003;424(6950):805-8.
74. Gardner RG, Nelson ZW, Gottschling DE. Degradation-mediated protein quality control in the nucleus. Cell. 2005;120(6):803-15.
75. Rosenbaum JC, Fredrickson EK, Oeser ML, Garrett-Engele CM, Locke MN, Richardson LA, et al. Disorder targets misorder in nuclear quality control degradation: a disordered ubiquitin ligase directly recognizes its misfolded substrates. Molecular cell. 2011;41(1):93-106.
76. Mukhopadhyay S, Krishnan R, Lemke EA, Lindquist S, Deniz AA. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(8):2649-54. Epub 007 Feb 13.
77. Sherman F. Getting started with yeast. Methods Enzymol. 1991;194:3-21.
78. Bagriantsev SN, Kushnirov VV, Liebman SW. Analysis of amyloid aggregates using agarose gel electrophoresis. Methods in enzymology. 2006;412:33-48.

CHAPTER 4: INVESTIGATING CONTEXTUAL PRION-LIKE DOMAIN FEATURES AND CELLULAR FACTORS INVOLVED IN THE DEGRADATION OF PRION-LIKE DOMAINS[4]

INTRODUCTION

In Chapter 3, I presented empirical estimates of the degradation propensities and prion propensities of the 20 canonical amino acids in two human PrLDs. Although these results suggest that the composition of the mutagenized regions is the primary determinant of degradation or prion formation, these effects are clearly influenced by neighboring amino acids within the PrLD or PFD. Although the G-rich PrLDs and Q/N-rich PFDs appear to modulate the effects of degradation-promoting amino acids, it is unclear which sequence features within these domains are the strongest contributors. Therefore, in this chapter, I further define contextual features that influence the degradation rates of the A2 PrLD and the Sup35 ND.

In addition, many components of the proteostasis network, including protein chaperones and ubiquitin-proteasome factors, efficiently recognize exposed hydrophobic residues for re-folding or degradation of the substrate protein (for extensive discussion, refer to Chapter 1 and Chapter 3). Therefore, in this chapter, I also examine the possible involvement of native yeast proteostasis factors in the recognition and degradation of the A2 PrLD.

[4] Kacy Paul built and tested phenotypes for the hydrophobic insertions at alternative locations in the ND, ORD, and M-domain, as well as the PrLD/ND hybrid (Figures 4.1-4.3). Western blots for these mutants were performed by Satoshi Machihara and myself. Scrambled PrLDs and NDs were designed, built, and tested by Satoshi Machihara and myself (Table 4.1 and Figure 4.4). Kacy Paul built and tested phenotypes for the PrLD-ND and ND-PrLD chimeras (Figure 4.5). Western blots for these mutants were performed by Satoshi Machihara (Figure 4.5). Genetic knockouts were generated and tested by Satoshi Machihara and myself (Table 4.2).

MATERIALS AND METHODS

Strains, Media, and Degradation Assays

Strains, media, and growth conditions were as described in Chapter 3. Degradation assays were also performed as described in Chapter 3, using the same antibodies and immunoblotting procedure.

Scrambling Mutagenesis

Randomizations of the A2 PrLD and Sup35 ND primary sequences were performed *in silico* using Microsoft Excel's random number function to re-assign amino acid positions. Scrambled sequences were built synthetically by sequential PCR reactions containing mutagenic oligonucleotides encoding the scrambled sequences, with 5' and 3' extensions facilitating cloning by homologous recombination into our standard cloning vector and strain (Chapter 3).

Analysis of G/Q/N Distribution within the A2 PrLD

Computation was carried out using an in-house Python script. The A2 PrLD was scanned using a 5-amino acid sliding window (windows smaller than 5 amino acids were not scored). For each window, the percentages of G and Q/N residues were calculated and assigned to the first residue in the window. G and Q/N percentages were plotted in Microscoft Excel against amino acid position.

RESULTS

Hydrophobic Residues Enhance Degradation Specifically in the Context of the A2 PrLD

Our initial experiments focused on understanding the amino acids leading to degradation of the A2 PrLD but not the Sup35 ND. While these mutations provided crucial insight into the features leading to degradation or prion formation, they are limited to a single site within the A2PrLD. However, our results suggest that these processes are predominantly governed by amino acid composition, rather than primary sequence. Therefore, we reasoned that the location of the hydrophobic residues would not dramatically influence their effects within the A2 PrLD or Sup35 ND. To test this directly, hydrophobic residues were inserted either 10 or 30 residues from the N-terminus (herein referred to as N10 and N30 insertions) in Sup35. N10 insertion of two or four hydrophobic residues in Sup35 resulted in a predominantly *ade*$^-$ phenotype and did not accelerate degradation of the expressed Sup35 protein, whereas N10 insertion of six hydrophobic residues resulted in slightly enhanced degradation and a pink phenotype on YPD (Figure 4.1A, *top*). N30 insertion of two hydrophobic amino acids in Sup35 did not enhance the turnover rate of the protein and resulted in a predominantly *ade*$^-$ but "sectored" phenotype, indicative of spontaneous prion formation (Figure 4.1A, *bottom*). N30 insertion of four or six hydrophobic residues also did not noticeably affect the stability of Sup35 over five hours, and resulted in a pink phenotype on YPD but little or no growth on SC-ade. This suggests that the Sup35 PFD is capable of masking the degradation-promoting effects of hydrophobic residues regardless of their location.

Analogous insertions were also generated in the A2 PrLD either 11 or 33 residues from the N-terminus (herein referred to as N11 and N33 insertions). N11 insertion of two or more hydrophobic residues in the A2 PrLD resulted in rapid degradation of A2-Sup35 and a uniform

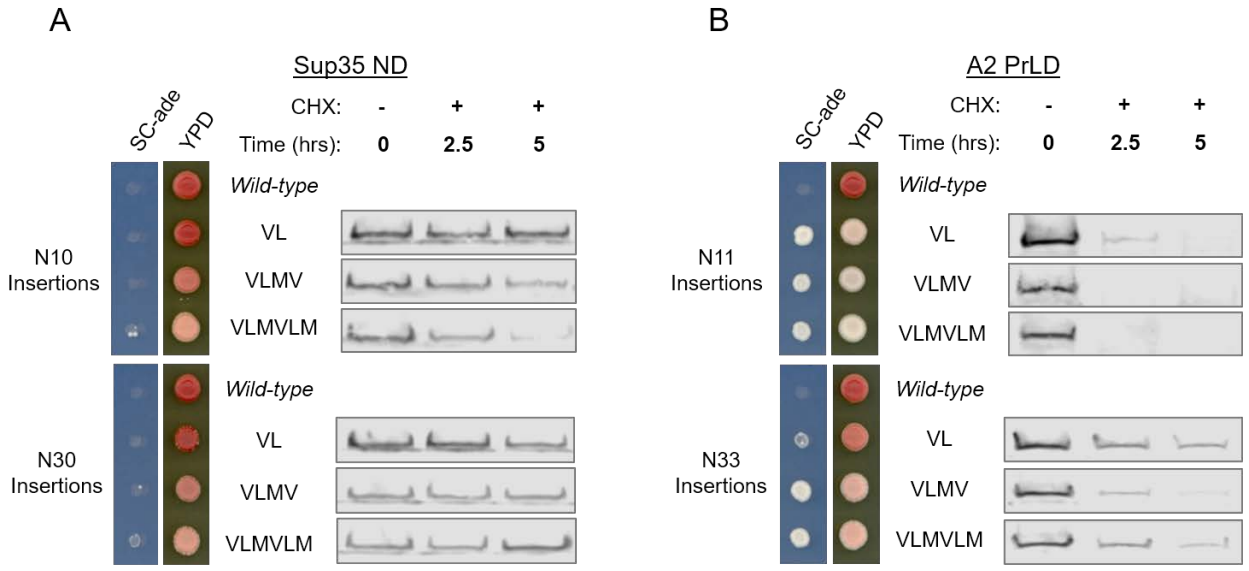


Figure 4.1: Degradation of the A2 PrLD occurs irrespective of the position of hydrophobic residues within the PrLD. (A) Hydrophobic insertions in the Sup35 ND at the N10 (*top*) and N30 (*bottom*) positions generally result in an *ade*$^{-}$ phenotype and do not strongly affect Sup35 stability. Six hydrophobic residues at the N10 position result in a slight increase in the degradation rate of Sup35. **(B)** Insertion of as few as two hydrophobic residues at the N11 position in the A2 PrLD results in a strong *ADE*$^{+}$ phenotype and rapid degradation of the A2 PrLD (*top*). Insertion of four or six hydrophobic residues at the N33 position results in a strongly *ADE*$^{+}$ phenotype and modest destabilization of the A2 PrLD (*bottom*).

ADE$^+$ phenotype (Figure 4.1B, *top*). N33 insertion of two hydrophobic residues slightly destabilized A2-Sup35 (although to a lesser degree than the analogous N11 insertion) and resulted in a pink phenotype on YPD with minor growth on SC-ade (Figure 4.1B, *bottom*). N33 insertion of four or six hydrophobic residues enhanced the *ADE*$^+$ phenotype and A2 PrLD turnover, although to differing degrees. These results suggest that hydrophobic residues trigger degradation of the A2 PrLD in a largely position-independent manner, although N33 insertions near the Sup35 portion of the A2-Sup35 fusion may facilitate some degree of protection against the degradation machinery.

Based on the evidence presented thus far, it is clear that hydrophobic residues act synergistically with the A2 PrLD to promote degradation. However, it is unclear whether these two features contribute independently to PrLD turnover, or if the hydrophobic residues must be embedded within the A2 PrLD for a cooperative effect on degradation. In order to examine this possibility, hydrophobic residues were inserted into two regions common to the A2-Sup35 fusion and the native Sup35 protein but outside of the PrLD/ND region. The oligopeptide repeat domain (ORD), spanning amino acids 41-97 of Sup35, consists of five and a half imperfect repeats enriched in Q/N/G/Y residues. The middle (M)-domain, spanning amino acids 124-253 of Sup35, is a highly-charged domain enriched predominantly in E/K residues. Notably, both domains are predicted to be intrinsically disordered, although they likely adopt distinct disordered conformations [1, 2], which could affect hydrophobic residue accessibility.

Progressive insertion of hydrophobic residues in the ORD (up to eight hydrophobic residues) and M-domain (up to six hydrophobic residues) of WT Sup35 resulted in an *ade*$^-$ phenotype and did not accelerate degradation of the Sup35 protein (Figure 4.2A). Insertion of four or fewer hydrophobic residues in the ORD or M-domain of the A2-Sup35 fusion also

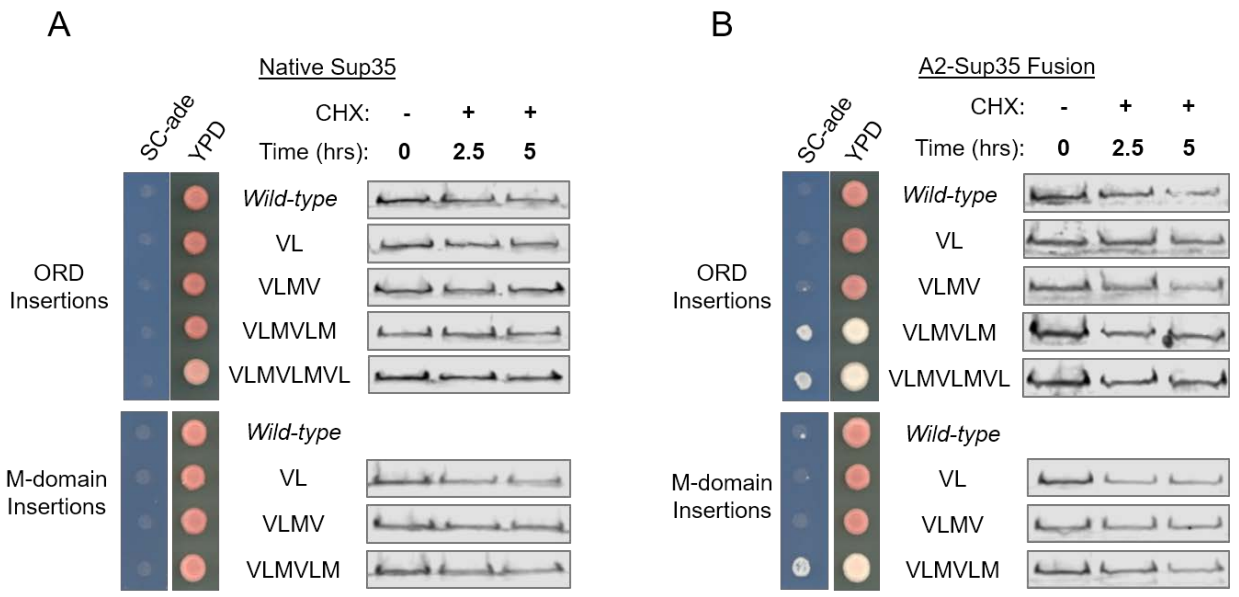


Figure 4.2: Hydrophobic residues enhance degradation only within the A2 PrLD. (A) Insertion of hydrophobic residues into the ORD (*top*) or M-domain (*bottom*) of Sup35 leads to an *ade*$^{-}$ phenotype and no apparent increase in the rate of Sup35 degradation. **(B)** Insertion of four or fewer hydrophobic residues into the ORD (*top*) or M-domain (*bottom*) of the A2-Sup35 fusion leads to an *ade*$^{-}$ phenotype and no apparent increase in the A2-Sup35 degradation rate. Insertion of 6 or 8 hydrophobic residues in the ORD (*top*), or 6 hydrophobic residues in the M-domain (*bottom*) of the A2-Sup35 fusion results in an *ADE*$^{+}$ phenotype, although it is not due to an increase in degradation rate.

resulted in an *ade*$^{-}$ phenotype and no increase in the degradation rate of the A2-Sup35 protein (Figure 4.2B). This is in stark contrast to analogous insertions in the A2 nucleation domain itself, where as few as two hydrophobic residues were sufficient to enhance the degradation rate of the A2-Sup35 fusion (Chapter 3, Figure 3.6). Interestingly, insertion of six or eight hydrophobic residues in the ORD of the A2-Sup35 fusion resulted in an *ADE*$^{+}$ phenotype, yet no apparent difference in the degradation rate over the five hour timecourse (Figure 4.2B, *top*). A similar effect was observed for insertion of six hydrophobic residues in the M-domain of the A2-Sup35 fusion (Figure 4.2B, *bottom*). Although the function of the A2-Sup35 fusion is compromised by inserting hydrophobic residues outside of the A2 nucleation domain, the observed loss-of-function is not the result of an increase in turnover rate. This suggests that degradation is promoted by the cooperative effects of hydrophobic residues specifically in the context of the A2 PrLD.

Degradation-Enhancement by Hydrophobic Residues Requires an Extended A2 PrLD Context

In order to further dissect the regions within the A2 PrLD that facilitate degradation, 11- or 22-amino acid segments from the A2 PrLD were substituted into the corresponding region of the Sup35 nucleation domain (ND) to generate chimeric hybrids of the original A2 PrLD and Sup35 ND (Figure 4.3). Hydrophobic residues (two or four residues) were then inserted into each of the PrLD/ND hybrids. Substitution of the first 11 A2 PrLD residues for the analogous region of the Sup35 ND resulted in an *ADE*$^{+}$ phenotype even in the absence of hydrophobic residue insertions, but did not alter the apparent protein stability (data not shown). For each of the remaining hybrids, expression of the hybrid protein without added hydrophobic residues result in an *ade*$^{-}$ phenotype and relatively high stability over the five hour timecourse (Figure 4.3).

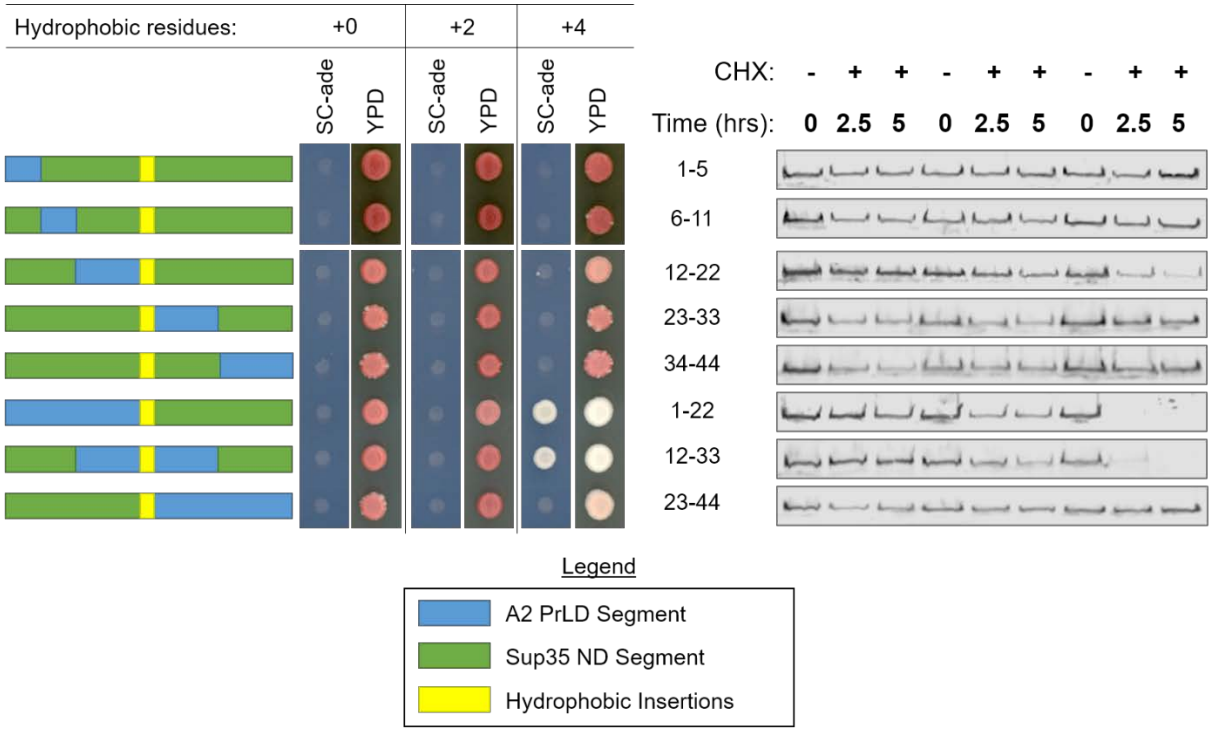


Figure 4.3: An extended A2 PrLD region is required to facilitate degradation mediated by hydrophobic residues. Various regions of the A2 PrLD were substituted for analogous regions within the Sup35 ND to generate PrLD/ND chimeras. In each case 0, 2, or 4 hydrophobic residues were also inserted at the original hydrophobic insertion site in the Sup35 ND (Chapter 3). Degradation of the A2 PrLD requires at least 20 amino acids from the A2 PrLD surrounding the hydrophobic residues to facilitate degradation. Numbers indicate the residues from the A2 PrLD present in each chimera (*middle*).

Although two hydrophobic residues were sufficient to accelerate degradation of the full A2 PrLD (Chapter 3, Figure 3.6), addition of two hydrophobic residues in each of the hybrid PrLD/ND's results in an *ade*$^-$ phenotype and no apparent increase in the degradation rate of any of the hybrid proteins. Insertion of four hydrophobic residues resulted in an *ADE*$^+$ phenotype and rapid degradation of the hybrids containing residues 1-22 or 12-33 derived from the A2 PrLD. Interestingly, 11-amino acid substitutions in each of the regions corresponding to those covered by the 22-amino acid substitutions (1-11, 12-22, and 23-33) did not facilitate degradation with the same hydrophobic insertions (Figure 4.3, and data not shown). This suggests that efficient degradation requires a relatively long (~20 amino acid) segment of the A2 PrLD surrounding the hydrophobic residues. It is also interesting to note that, for all of the strains expressing hybrids for which no change in protein stability occurs upon insertion of hydrophobic residues (23-33, 34-44, and 23-44), the phenotypes are predominantly *ade*$^-$ yet sectored, indicating that when sufficiently stable, the A2 PrLD may contribute to prion aggregation.

Stabilization by the Sup35 ND is Independent of Primary Amino Acid Sequence

A number of observations suggest that amino acid composition (rather than primary amino acid sequence) is a key contextual feature leading to stabilization of the Sup35 ND or degradation of the A2 PrLD. First, while the A2 PrLD shares many compositional features with yeast prion domains, the most prominent difference between the A2 PrLD and the Sup35 ND is the percent composition of G, Q, and N residues. Our degradation propensity scores suggest that Q/N residues disfavor on increase in degradation rate of the PrLD's, while G residues do not strongly influence the degradation rate. Therefore, the Q/N content of the Sup35 ND may serve to stabilize otherwise degradation-prone sequence features. Indeed, upon insertion of

hydrophobic residues within the A2 PrLD and Sup35 ND, substitution of G's for Q/N's within the A2 PrLD tends to decrease its degradation rate, while substitution of Q/N's for G's within the Sup35 ND tends to accelerate its degradation (Chapter 3, Figure 3.8). Second, the location of hydrophobic residues within the A2 PrLD or Sup35 ND exerts little effect on the stability of the resulting proteins (Figure 4.1), indicating that degradation is facilitated or prevented by broad regions of the A2 PrLD and Sup35 ND, respectively. Finally, substituting small portions of the A2 PrLD (11 amino acid regions) for analogous regions of the Sup35 ND failed to facilitate degradation of the resulting chimeric proteins. Rather, an extended region of the A2 PrLD (at least 22 amino acids) surrounding the hydrophobic residues was required for degradation, indicating that no individual single feature present in the smaller 11 amino acid regions enhanced degradation. Collectively, this suggests that amino acid composition (particularly, with respect to G/Q/N content) rather than primary sequence is one of the key features influencing the effects of non-aromatic hydrophobic amino acids.

To directly examine the effects of primary sequence, two degradation-promoting peptides in the context of the A2 PrLD and the Sup35 ND were chosen as starting substrates for primary sequence randomization. In each case, the A2 PrLD or Sup35 ND was independently scrambled twice, resulting in eight new combinations of degradation-promoting peptide and surrounding PrLD/ND primary sequence (Table 4.1).

As demonstrated in Chapter 3, peptides enriched in hydrophobic residues do not promote degradation in the context of the native Sup35 ND sequence, and result in an *ade*$^{-}$ phenotype (Figure 4.4). Additionally, all four scrambled Sup35 ND sequences resulted in an *ade*$^{-}$ phenotype, with no apparent decrease in the stability of the Sup35 ND. By contrast, while the degradation-promoting peptides destabilize the native A2 PrLD sequence, all four scrambled A2

Table 4.1: Degradation of the A2 PrLD occurs irrespective of the position of hydrophobic residues within the PrLD. Amino acid sequences of the native A2 PrLD and Sup35 ND are indicated. Selected *ADE*$^{+}$ isolate sequences from the A2 PrLD degradation library (Chapter 3), as well as their Sup35 ND counterparts, were each scrambled twice. For reference, the original library mutagenesis region is underlined, Q/N's are highlighted in green, and G's are highlighted in red.

		Sequence
Native A2 PrLD		MSGPGYGNQGGGYGGGYDNYGGGNYGSGNYNDFGNYNQQPSNYG
A2 Peptide 1 Scrambles	**#1**	MSGGGDGQYSGNYYGPNNGGGQYGYQMVMFFECAGYNYGGNGPN
	#2	MSNGYNGYYQGGGNQGYSGGGPNNYYMVMFFECAPGGGGGDQYN
A2 Peptide 2 Scrambles	**#1**	MSGYQNNGGQGGPNGGYGYQPSYYGGLFVIASSKGNNYNGYGGD
	#2	MSQNYGGYGNDGGNGGGYGNNSGGQPLFVIASSKYNGYYYGPQG
Native Sup35 ND		MSDSNQGNNQQNYQQYSQNGNQQQGNNRYQGYQAYNAQAQ
Sup35 Peptide 1 Scrambles	**#1**	MSQQNYQANADNYQYSNYQNMVMFFECAGQASNQQQGGYQ
	#2	MSNQNYDGYGAAQNQYQQAYMVMFFECAQQQSNGSNQYQN
Sup35 Peptide 2 Scrambles	**#1**	MSQNYYSAQNGDQNSYAQYQLFVIASSKNQGQANQYNQQG
	#2	MSDQQQNYSGQAYQQQQQYNLFVIASSKAGQNNYNGSANY

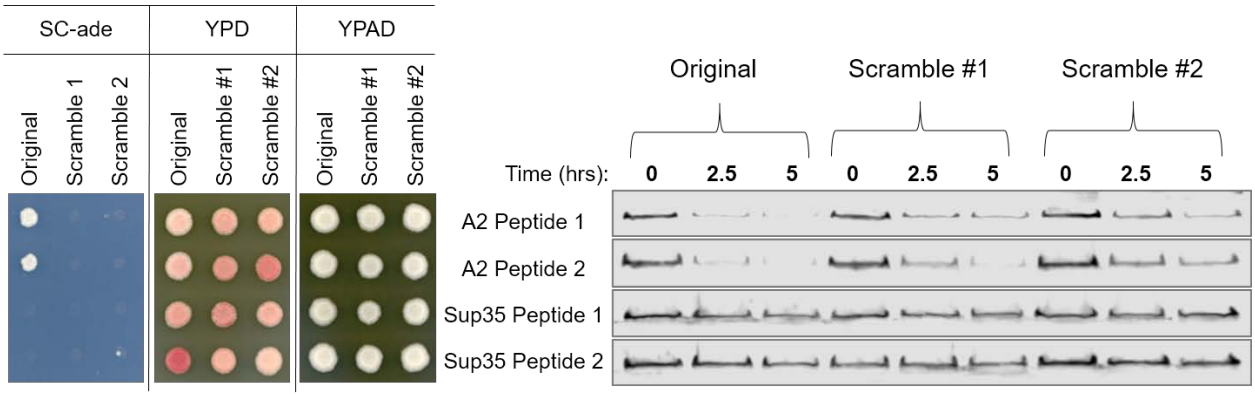


Figure 4.4: The anti-degradation activity of the Sup35 ND is primary sequence independent. Scrambling the A2 PrLD results in a phenotypic change from *ADE*$^{+}$ to *ade*$^{-}$ (*left*) and stabilization of the A2 PrLD to varying degrees (*right*). Scrambling the Sup35 ND does not lead to a change in phenotype or Sup35 ND stability.

PrLD sequences result in an *ade*$^{-}$ phenotype and improve the stability of the A2 PrLD, albeit to differing degrees. Therefore, while composition of the Sup35 ND is predominantly responsible for counteracting the degradation-promoting effects of hydrophobic peptides, degradation of the A2 PrLD is at least partially dependent on surrounding primary sequence features.

In light of the aforementioned evidence indicating composition as a key contextual feature facilitating degradation of the A2 PrLD, these results prompted further examination of the scrambled sequences to elucidate potential primary sequence features that were disrupted by scrambling the A2 PrLD. While the A2 PrLD is predominantly G-rich, it also contains a smaller percentage of Q/N residues. In the native A2 PrLD, these residues are non-uniformly distributed – the G residues tend to cluster on the N-terminal side of the 8-amino acid mutagenesis region, while Q/N residues tend to cluster on the C-terminal side (Figure 4.5). In all cases, scrambling the A2 PrLD resulted in a redistribution of G/Q/N residues, with more Q/N residues found on the N-terminal side and more G residues found on the C-terminal side of the mutagenesis region relative to the wild-type sequence. Interestingly, of the scrambled A2 PrLD's, the protein exhibiting the fastest degradation rate (A2 Peptide 2 Scramble #1; Figure 4.4) has a G/Q/N distribution that most closely resembles the original A2 PrLD sequence, with G residues dominating the sequence immediately preceding the random mutagenesis window, and Q/N residues preferentially clustering on the C-terminal side of the mutagenesis window. Therefore, this analysis suggests that the distribution of G/Q/N residues near hydrophobic sequences can influence degradation rates, and may reflect a combination of primary amino acid sequence and amino acid composition effects. However, further experiments are required to fully resolve the importance of G/Q/N distribution.

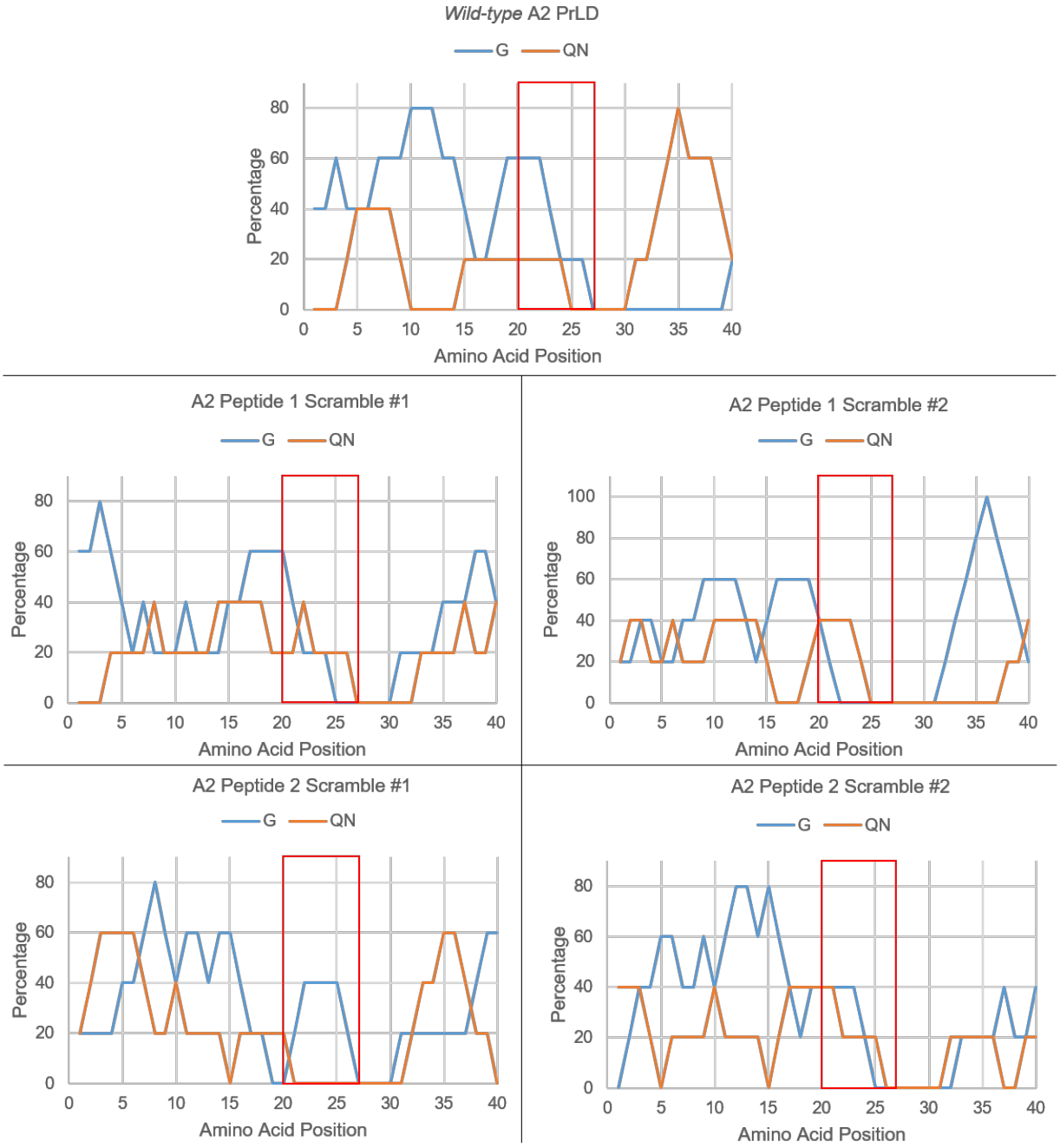


Figure 4.5: Scrambling disrupts the non-uniform distribution of G/Q/N residues within the native A2 PrLD. Percentages of G (blue) and Q/N (orange) residues within 5-amino acid sliding windows were assigned to the first residue in each window and plotted as a function of amino acid position within the A2 PrLD. For reference, the 7 amino acids immediately preceding the peptide from the A2 PrLD degradation library are indicated (red box). Sliding window percentages were computed using an in-house Python script.

The Sup35 ND Exerts a Dominant Stabilizing Effect and Facilitates Prion Formation by the A2 PrLD

Hydrophobic residues inserted within the Sup35 ND fail to elicit an increase in degradation rate. It is possible that the Sup35ND is able to mask embedded degradation-promoting features, or that the Sup35ND is simply inherently degradation-resistant. To examine these possibilities, the Sup35 ND was inserted on the N-terminal or C-terminal end of the A2 PrLD in the A2-Sup35 fusion. Two or four hydrophobic residues were then inserted in either the A2 PrLD portion or the Sup35 ND portion, resulting in 10 possible combinations of PrLD-ND orientation and hydrophobic insertion location.

In the absence of hydrophobic insertions, expression of the PrLD-ND and ND-PrLD resulted in an *ade*$^{-}$ phenotype, and each protein was relatively stable over the five hour timecourse (Figure 4.6). Interestingly, simply having the two domains present in the same fusion (regardless of their orientation) resulted in sectoring on YPD, which indicates an increase in the rate of spontaneous prion formation relative to the A2-Sup35 fusion or the WT Sup35 protein. Insertion of two or four hydrophobic residues within the Sup35 ND portion of either fusion enhanced the degree of sectoring in a dose-dependent manner, similar to what was observed for the same insertions in the Sup35 ND in the absence of the A2 PrLD (Chapter 3, Figure 3.6). Strikingly, insertion of two hydrophobic residues in the A2 PrLD in either fusion also resulted in a predominantly red phenotype with sectoring on YPD, minor growth on SC-ade, and no apparent increase in the rate of degradation. Inserting four hydrophobic residues resulted in a complete *ADE*$^{+}$ phenotype on YPD and SC-ade, yet no reduction in protein stability. Collectively, this suggests that the Sup35 ND exerts a dominant effect on the stability of the

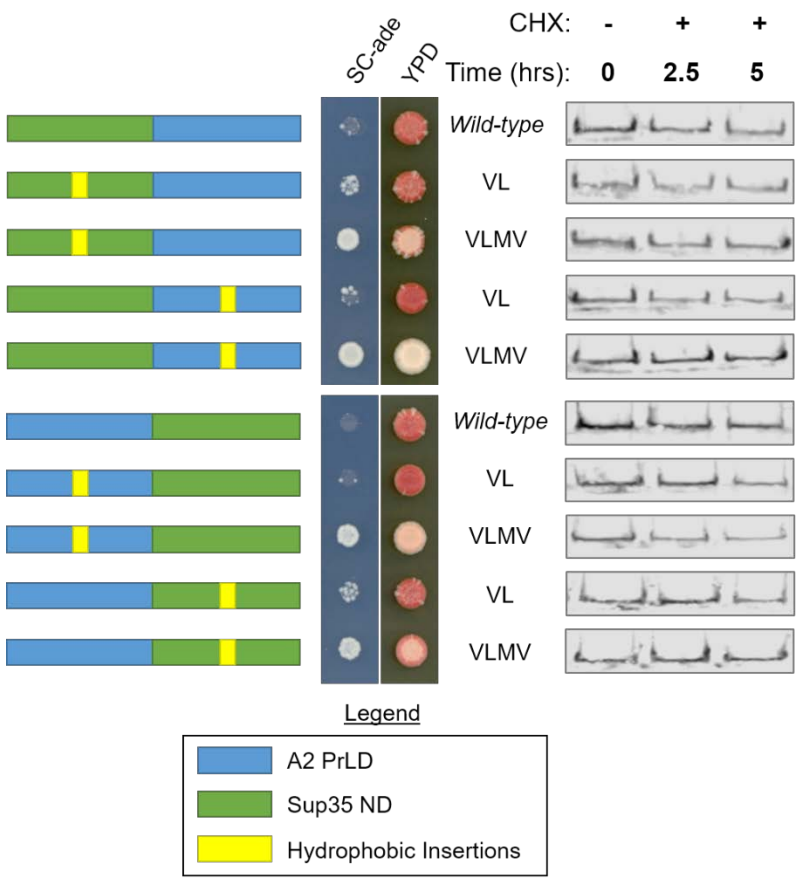


Figure 4.6: The Sup35 ND prevents degradation of the A2 PrLD irrespective of orientation relative to the PrLD or and location of hydrophobic residues. The A2 PrLD was inserted adjacent to the Sup35 ND on the N-terminal end or the C-terminal end. In each orientation 0, 2, or 4 hydrophobic residues were inserted either in the A2 PrLD or the Sup35 PrLD. In all cases, insertion of up to 4 hydrophobic residues does not increase the degradation rate of the fusion protein. Furthermore, in the absence of hydrophobic residues, both PrLD/ND orientations result in a sectored phenotype on YPD, indicating spontaneous prion formation. The degree of sectoring/*ADE*$^{+}$ growth generally increases upon insertion of hydrophobic residues.

fusion proteins, which facilitates spontaneous prion formation by an otherwise degradation-prone PrLD.

Investigation of Potential Proteostasis Factors Involved in Degradation of the A2 PrLD

As described in Chapter 1 and Chapter 3, abnormal protein features are often targeted by components of the proteostasis network for re-folding or degradation. In order to elucidate the proteostasis factors involved in the targeted degradation of the A2 PrLD, we deleted a number of proteostasis factor genes and examined potential genetic interactions with degradation-prone A2 PrLD mutants. These genes were selected on the basis of 1) well-characterized involvement in protein folding, disaggregation, or protein degradation pathways, or 2) the reported ability to specifically recognize short, solvent-exposed hydrophobic sequences (Table 4.2).

The majority of gene deletions did not alter the phenotype typically associated with degradation-prone PrLD's. Deletion of two genes (HUL5, which encodes a proteasome-bound ubiquitin ligase, and STI1, which encodes an Hsp70 and Hsp90 co-chaperone) caused a subset of phenotypically ADE^{+} strains to revert to an ade^{-} phenotype. However, subsequent western blots showed no detectable increase in the stability of the A2 PrLD, suggesting that these genetic interactions were not related to A2 PrLD recognition and degradation.

The simplest interpretation of these results is that none of the factors tested are involved in degradation of the A2 PrLD. However, these data should be interpreted with caution, as some proteostasis factors have overlapping or compensatory functions [3]. Further investigation (including inactivation of essential genes that cannot be deleted) may be required to uncover the factor or factors primarily responsible.

Table 4.2: Deletion of candidate proteostasis factors does not influence the stability of the A2 PrLD

Gene Deletions	Molecular Function	Phenotype Reversion?	A2 PrLD Stabilization?
SAN1	Ubiquitin Ligase	-	-
UBR1	Ubiquitin Ligase	-	n.d.
HUL5	Ubiquitin Ligase	+	-
DOA10	Ubiquitin Ligase	-	n.d.
LTN1	Ubiquitin Ligase	-	n.d.
SSA1	Protein Chaperone (Hsp70 family)	-	n.d.
SSA2	Protein Chaperone (Hsp70 family)	-	n.d.
SSA3	Protein Chaperone (Hsp70 family)	-	n.d.
SSA4	Protein Chaperone (Hsp70 family)	-	n.d.
SSA1/SSA2	Protein Chaperones (Hsp70 family)	-	n.d.
SSA1/SSA3	Protein Chaperones (Hsp70 family)	-	n.d.
SSA1/SSA4	Protein Chaperones (Hsp70 family)	-	n.d.
SSB1	Protein Chaperone (Hsp70 family)	-	n.d.
SSB2	Protein Chaperone (Hsp70 family)	-	n.d.
SSB1/SSB2	Protein Chaperones (Hsp70 family)	-	n.d.
HSP82	Protein Chaperone (Hsp70 family)	-	n.d.
HSC82	Protein Chaperone (Hsp70 family)	-	n.d.
HSP104	Protein Disaggregase (Hsp100 family)	-	n.d.
YDJ1	Co-chaperone (Hsp40 family)	-	n.d.
STI1	Co-chaperone	+	-
CUE5	Autophagy Adaptor	-	n.d.
RNQ1	Prion-inducing Factor	-	n.d.

CONCLUSION

Our genetic screen in yeast allowed us to ascertain the effects of each amino acid on the molecular fates of two aggregation-prone domains (Chapter 3). Our data indicated that non-aromatic hydrophobic amino acids can lead to triage of PFDs/PrLDs into distinct pathways, resulting in either proteasome-mediated degradation or prion aggregation. However, these trajectories were clearly dependent upon other features within the PFD/PrLD. Therefore, we sought to expand our understanding of the contextual features that influence the prion propensities and degradation propensities of non-aromatic hydrophobic residues.

In Chapter 3, I demonstrated that degradation of the A2 PrLD (but not the Sup35ND) is accelerated upon insertion of hydrophobic residues. In order to resolve whether the A2 PrLD and the non-aromatic hydrophobic residues enhance degradation independently or synergistically, we tested alternative locations for hydrophobic residues both within and outside of the A2 PrLD/Sup35ND. We find that hydrophobic residues enhance degradation when inserted at multiple locations within the A2 PrLD but not outside it. This suggests that degradation is governed by a combination of degradation-promoting amino acids as well as a PrLD context which supports or enables recognition of these features by the proteostasis machinery. Furthermore, the degradation-promoting effects of hydrophobic residues were suppressed at multiple locations both within and outside of the Sup35ND in the native Sup35 protein, perhaps indicating that the Sup35ND provides broad resistance to recognition of aggregation-prone features by the proteostasis machinery.

In Chapter 3, I also present evidence that the degree of susceptibility of the A2 PrLD and Sup35 ND could be altered by changing the G/Q/N content of the region surrounding the hydrophobic residues. Scrambling the Sup35ND did not increase its susceptibility to

degradation, suggesting that primary sequence features do not play a critical role in preventing hydrophobic residues from triggering degradation. By contrast, the A2 PrLD exhibits a decrease in degradation rate upon randomization of the primary sequence. Although we have not yet determined the primary sequence features of the A2 PrLD important for susceptibility to degradation, computational analysis of the original and scrambled A2 PrLD sequences indicate that G/Q/N distribution within the A2 PrLD (which was disrupted upon scrambling) may play a role. Scrambling smaller regions of the A2 PrLD would allow us to directly test this hypothesis.

In order to more fully map the regions of the A2 PrLD and Sup35 ND required for instability or stability, respectively, we constructed a series of chimeric PrLD-NDs in which portions of the Sup35 ND were replaced by the corresponding regions of the A2 PrLD. Individual substitution of 11-amino acid segments along the entire Sup35 ND sequence did not affect Sup35 ND stability. However, substitution of 20-amino acid segments surrounding or adjacent to the hydrophobic residues resulted in a reduction in Sup35 stability. Collectively, this suggested that an extended portion of the A2 PrLD surrounding degradation-promoting hydrophobic residues confers susceptibility to enhanced protein turnover.

Shockingly, fusion proteins containing both the A2 PrLD and the Sup35 ND were degradation-resistant regardless of the orientation of the domains or the location of the hydrophobic residues. Therefore, the Sup35 ND exerts a dominant stabilizing effect even when hydrophobic residues are inserted into the A2 PrLD in a position that potently induces degradation in the original A2 PrLD. Once sufficiently stabilized, these A2 PrLD's supported spontaneous prion formation in the absence of hydrophobic residues. Furthermore, the frequency of spontaneous prion formation was augmented by insertion of increasing numbers of hydrophobic residues. This suggests that protein stability is an underappreciated determinant of

prion propensity – in some cases, sufficient degradation-resistance may even be a prerequisite for aggregate formation.

These results have important implications for understanding prion formation in the presence of intracellular pro-degradation and anti-aggregation systems. We find that similarly aggregation-prone domains can differ in their susceptibility to intracellular degradation machinery, which is dependent upon short sequence motifs, the surrounding prion domain context, and even relatively distant yet dominantly stabilizing domains. Further exploration of the features that are simultaneously aggregation-prone and degradation-resistant may aid in the identification of new prion candidates.

REFERENCES

1. Das, R.K. and R.V. Pappu, *Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues.* Proc Natl Acad Sci U S A, 2013. **110**(33): p. 13392-7.
2. Mukhopadhyay, S., et al., *A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures.* Proc Natl Acad Sci U S A, 2007. **104**(8): p. 2649-54. Epub 2007 Feb 13.
3. Albanese, V., et al., *Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells.* Cell, 2006. **124**(1): p. 75-88.

CHAPTER 5: CONCLUSION

ADVANCES IN PRION PREDICTION AND VALIDATION *IN VITRO* AND *IN VIVO*

The first three yeast prion proteins, Sup35, Ure2, and Rnq1, were discovered in rapid succession in the mid- to late-1990's [1-4]. In each case, the domain responsible for prion formation was found to be unusually Q/N-rich, at the time suggesting that high Q/N content is an important feature of prion domains. Further exploration of the protein features contributing to prion formation in yeast suggested that amino acid composition of prion domains is the predominant determinant [5, 6]. Consequently, first-generation prion prediction algorithms focused predominantly on compositional similarity to known yeast prion domains [7-9]. This important first step led to the identification of candidate prion proteins [7-9], many of which were later experimentally confirmed as *bona fide* prion proteins [7, 10-13]. However, extensive empirical testing of prion candidates also revealed that compositional similarity alone does not effectively predict prion activity among high-scoring candidates [7, 14]. Collectively, while these results represented a monumental advance in understanding and predicting prion activity, they also suggested that some features controlling prion formation remained incompletely understood.

These algorithms assume that all deviations in amino acid composition from known yeast prion domains will disfavor prion formation. Therefore, equivalently-scored candidate prion domains may differ dramatically in their actual prion propensities depending upon the identities of the amino acids that constitute the compositional differences. Our group adopted an alternative approach in an attempt to empirically estimate the prion propensities of the 20 canonical amino acids within the context of a Q/N-rich prion domain [14]. Amino acids with similar physicochemical properties clustered remarkably well when ranked on the basis of the

resulting prion propensity scores, suggesting that basic physicochemical properties of the amino acid groups were largely responsible for their effects on prion formation. These prion propensity scores allowed us to develop a prion prediction algorithm (which we named the Prion Aggregation Prediction Algorithm, or PAPA) that was not predicated on the compositions of known prion domains, and which provides novel predictions for mutations within prion domains [14, 15]. Applying PAPA to the candidates from the Alberti and Halfmann *et al.* screen showed that we could effectively distinguish candidates with prion activity in each of the four assays from candidates with no prion activity in any of the assays [15], suggesting that we had successfully identified the second layer of features controlling prion formation that is missed by other prion prediction methods.

As our collective understanding of prion processes in yeast expanded, prion prediction algorithms were subsequently applied to the human proteome [16]. A number of the top-scoring human proteins with prion-like domains (PrLDs) have recently been linked to a variety of degenerative disorders [16], suggesting that prion-like phenomena may play a role in disease development or progression. Two interesting examples are the related RNA-binding proteins hnRNPA1 and hnRNPA2, which both contain glycine-rich PrLDs. Single amino acid substitutions in either PrLD can lead to aberrant protein aggregation and multisystem proteinopathy in humans [17]. In both proteins, the substitution is an aspartic acid-to-valine substitution, which is predicted by PAPA to increase aggregation propensity. While this represents a critical piece of evidence suggesting that PAPA may provide accurate predictions for human PrLDs, it is limited to a single substitution in related PrLDs.

TEST-DRIVING PRION PREDICTION ALGORITHMS IN MULTICELLULAR EUKARYOTES: WHAT WE CAN LEARN FROM SUCCESS AND FAILURE

In order to more comprehensively investigate our prediction accuracy, we recently tested additional types of mutations and alternative mutation locations [18]. According to our empirical prion propensity estimates, aspartic acid is not expected to be uniquely prion inhibiting. Likewise, valine is not expected to be uniquely prion-promoting. Additionally, since prion formation is predominantly composition-dependent, the precise location of a given mutation should not dramatically alter its effects, provided the mutation still lies within the PrLD. We chose a range of single mutations (as well as combinations of mutations) categorized as prion-promoting, neutral, or prion-inhibiting mutations based on PAPA's predictions.

As demonstrated previously, the hnRNPA2 PrLD with the native D290 displays little to no prion activity, whereas the disease-associated D290V mutation dramatically enhances its prion formation in yeast [17]. We tested eight additional substitutions at the 290 position, and found in all cases that PAPA accurately predicted their effects of prion formation. Furthermore, PAPA successfully predicted the combinatorial effects of various mutations at alternative positions. This demonstrates that PAPA is a good predictor of intrinsic prion propensity of human PrLDs when expressed in yeast, even when the mutations differ from those observed to cause disease in humans.

Additionally, PAPA was able to predict the aggregation-promoting effects (and associated cellular pathology) of most hnRNPA2 mutations in the context of the full hnRNPA2 protein expressed in a complex multicellular organism, *Drosophila melanogaster*. However, in some cases, mutations resulted in unexpected deviation from PAPA predictions and from our experiments in yeast. For example, PAPA predicts that tyrosine and isoleucine will have similar

effects on prion formation. Indeed, the "neutral" Y283I substitution (in combination with the D290V mutation) maintained high rates of amyloid formation *in vitro* and prion formation in yeast. However, the Y283I mutation completely restored the solubility of hnRNPA2 D290V in *Drosophila*. Deletion of one (Δ290) or both (Δ276/Δ290) prion inhibiting residues in the hnRNPA2 PrLD resulted in step-wise increases in aggregation rates *in vitro* and prion formation in yeast. However, the Δ276/Δ290 actually appeared to increase the solubility of hnRNPA2 in *Drosophila*. This suggests that PAPA accurately predicts inherent aggregation propensity *in vitro* and prion propensity in yeast – features which are generally sufficient to make accurate predictions of aggregation in *Drosophila* as well. Therefore, while most mutants predictably affected aggregation *in vitro*, in yeast, and in *Drosophila*, a small subset of PrLD mutants do not behave as predicted in *Drosophila*. Identification of the principles governing these near-miss predictions may improve prion prediction for complex multicellular eukaryotes.

We also note a subset of PrLD mutants for which aggregation propensity itself followed PAPA predictions, but aggregate handling, dynamics, or modification differed. For example, the D290F mutation enhanced the frequency of prion formation (to a level comparable to the D290V mutation) and facilitated coalescence into cytoplasmic PrLD aggregates in yeast. In *Drosophila*, the D290F mutation also enhanced aggregation, which is consistent with PAPA predictions and its effect on prion formation in yeast. However, these aggregates appeared to localize to the nucleus, suggesting that the mutation had affected nucleocytoplasmic transport dynamics in addition to intrinsic aggregation propensity. Another example was the D276V/D290V double mutant. As predicted, these prion-promoting mutations had an additive effect on the rate of prion formation in yeast (relative to the single mutations), and resulted in the formation of cytoplasmic aggregates in *Drosophila*. However, the aggregates in *Drosophila* were greater in number and

smaller in size compared with those formed by the D290V single mutant, suggesting that the dynamics of aggregate formation had been fundamentally altered. This could result from changes in intrinsic aggregation properties, such as the rate of aggregate nucleation or aggregate growth (which may be indirectly accounted for in the "lump-sum" predictions provided by current prion prediction algorithms), or from extrinsic factors such as aggregate remodeling by proteostasis machinery (which is not currently accounted for in any prion prediction algorithm). Finally, certain PrLD mutants appear to be preferentially post-translationally modified, and modifications depended on a combination of the specific amino acid mutation and whether they are expressed in yeast or *Drosophila*. For example, the insoluble fraction of hnRNPA2 D290V appears as two distinct bands by western blot, indicating a post-translational modification. However, none of the other hnRNPA2 mutants indicate a modification (including the D276V/D290V double mutant), and this modification is not detectable when the D290V PrLD is expressed in yeast. Likewise, the D290Y mutant appears to be modified in yeast (not tested in *Drosophila*), while none of the other mutants expressed in yeast indicate a modification. While, neither modification prevented aggregation or prion formation, post-translational modifications in other locations or other PrLDs would conceivably affect aggregation-propensity *in vivo*.

PrLD COMPOSITION INFLUENCES AMINO ACID PRION PROPENSITIES AND DEGRADATION TRIAGE DECISIONS

While single amino acid changes provide a crude approximation of the prion propensities of select amino acids within the human PrLDs (which were generally accurately predicted by PAPA), we complemented this approach with our random mutagenesis and prion screening pipeline to empirically resolve prion propensities for the 20 canonical amino acids in the context

of the hnRNPA1 and hnRNPA2 PrLDs. Most amino acids had similar prion-promoting or -inhibiting effects in the Sup35 prion domain [14] and the human PrLDs (Chapter 3, Table 3.2), suggesting that their prion propensities are not strongly influenced by the amino acid composition of the surrounding prion domain. For example, aromatic residues strongly promote prion formation in both the yeast and human prion/prion-like domains, while charged residues strongly inhibit prion formation in the yeast and human prion/prion-like domains.

Unexpectedly, in the same screen we found that short stretches of non-aromatic hydrophobic residues can accelerate degradation of the hnRNPA2 PrLD in yeast. This was particularly surprising, given that we had already shown that the aspartic acid-to-valine substitutions with the hnRNP PrLDs increase prion propensity [17]. However, progressively increasing the number of hydrophobic residues within the hnRNPA2 PrLD (via insertions) allowed us to resolve the minimum number of hydrophobic residues required for enhanced degradation. A detectable increase in the degradation rate of the hnRNPA2 PrLD was first observed upon insertion of as few as two adjacent hydrophobic residues, and was further enhanced with three or more hydrophobic residues. This indicates that, although the hnRNPA2 PrLD may be susceptible to targeted degradation via hydrophobic residues, the D290V mutant and the D276V/D290V double mutant (which contains two distant hydrophobic residues) may effectively evade degradation and form aggregates. Interestingly, aromatic amino acids increased the prion propensity of the hnRNPA2 PrLD but did not lead to an apparent increase in the degradation rate, suggesting that aromatic amino acids may be aggregation-prone yet poorly detected by the degradation machinery.

These results were also surprising to us because hydrophobic residues potently promote prion formation in the Sup35 prion domain [14, 15, 19, 20]. This suggested that Sup35 might

resist the degradation-promoting effects of short hydrophobic sequences. Indeed, progressive insertion of hydrophobic residues into the Sup35 prion domain did not lead to a detectable increase in its rate of degradation. Rather, as we progressively increased the hydrophobic content of Sup35, we observed a dose-dependent increase in the rate of spontaneous prion formation. This suggests that the Sup35 prion domain can effectively mask aggregation-prone features from recognition by the degradation machinery, and that its ability to stabilize aggregation-prone features is an important prerequisite for prion formation.

Hydrophobic residues are highly aggregation-prone [21] but preferentially partition to the interior of natively folded proteins [22]. Therefore, these features are common targets of proteostasis factors, including chaperones [23-31] and components of the ubiquitin-proteasome degradation pathway [32, 33], which work to maintain proper protein folds, prevent protein aggregation, and degrade damaged or misfolded proteins. Our results demonstrate that the ability of hydrophobic residues to stimulate targeted degradation or prion aggregation is highly context-dependent, even among similarly aggregation-prone domains. Collectively, this indicates that PrLDs may differ in their susceptibility to proteostatic regulation, which could ultimately affect their ability to aggregate *in vivo*.

FUTURE CHALLENGES IN PRION PREDICTION: PRION FORMATION IN A CELLULAR CONTEXT

We offer a perspective oriented toward future refinement of prion prediction methods. Current prediction methods focus predominantly on intrinsic protein sequence features. Despite their simplicity, these algorithms have been quite successful at predicting prion activity in yeast [7, 14, 15] and humans [16, 17], demonstrating that these methods clearly possess an important

piece of the prion prediction puzzle. However, our recent studies highlight both the strengths and the limitations of our current prediction method.

PAPA predictions closely correlate with intrinsic aggregation propensity in vitro and prion propensity in yeast, which generally extends to accurate predictions of aggregation propensity in Drosophila. However, our observations also highlight a number of in vivo considerations that could conceivably affect prion formation yet are not currently incorporated into any prion prediction algorithm. It is worth noting that some prediction algorithms were founded upon sequences that demonstrate bona fide prion activity in yeast. Therefore, to some degree, they indirectly integrate the cumulative influence of an intracellular environment into their predictions. But this may also limit their adaptability to other organisms or tissues. While it is not feasible to take into account the totality of potential prion-modifying factors on a protein-by-protein basis, we emphasize potentially generalizable cell biological principles that may shed light on in vivo prion propensities and lead to better-informed prion predictions (Figure 5.1).

Protein Expression and Abundance

Aggregation is a concentration-dependent process, so intracellular protein concentrations could affect the probability of self-association and prion formation. By extension, the free intracellular protein concentrations of a given protein might be affected by the concentrations and affinities of molecular interacting partners. Protein levels and molecular interacting partners vary from organism to organism, from tissue to tissue, and even between subcellular compartments, which could affect the final molecular outcome (i.e. soluble or insoluble), or the manifestation of pathology (i.e. the tissue affected). Indeed, hnRNPA1 and hnRNPA2 are most highly expressed in brain and muscle tissue [17], which are both affected by the disease.

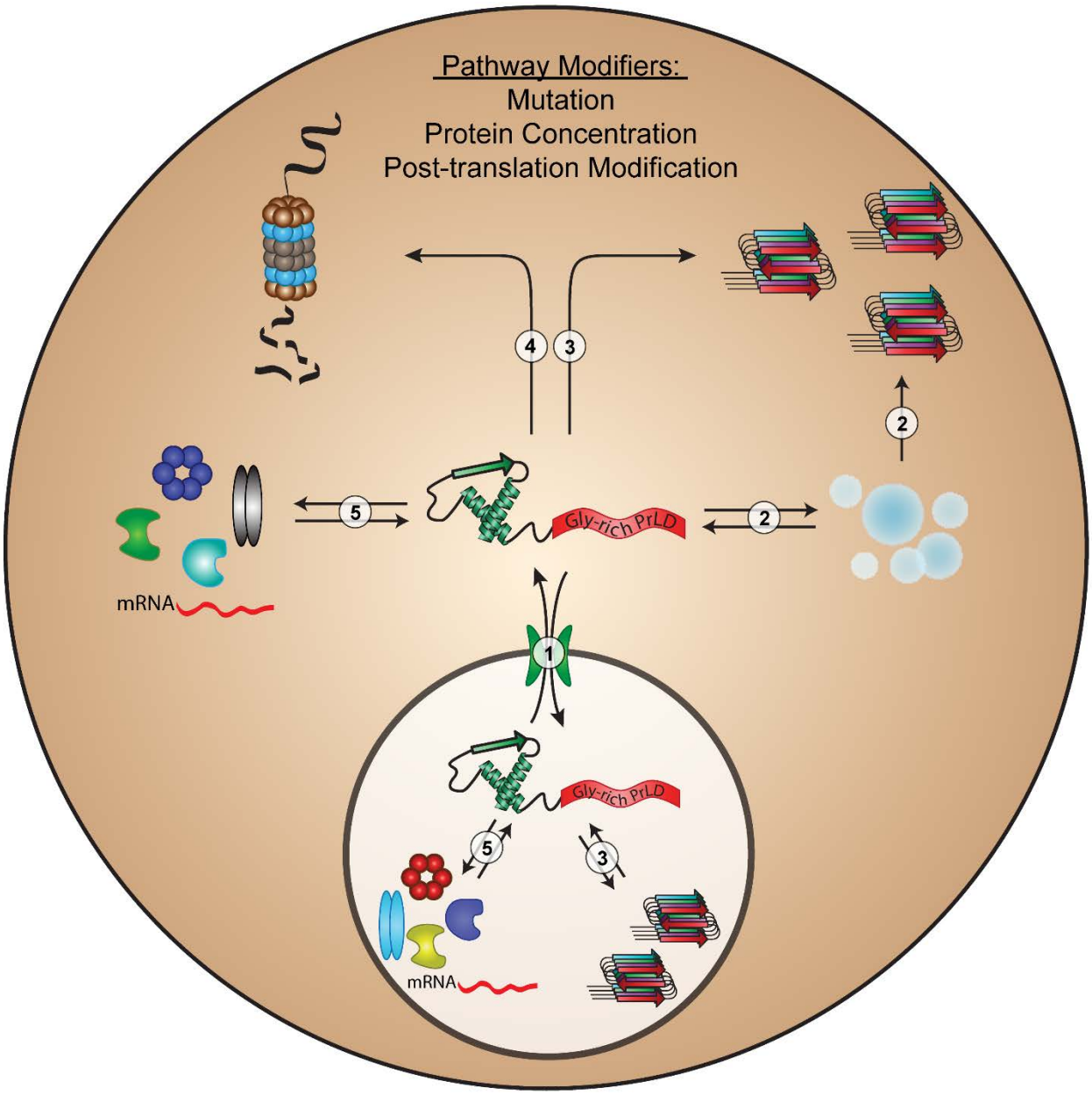


Figure 5.1: Intracellular processes commonly affecting PrLD aggregation. Mutations, changes in protein concentration, or post-translational modifications in PrLDs can result in disruptions in nucleocytoplasmic transport **(1)**, liquid-liquid phase separation/stress granule dynamics **(2)**, intrinsic prion propensity **(3)**, proteasome-mediated degradation **(4)** or alterations in organism-, tissue-, or compartment-specific intermolecular interactions **(5)**.

Liquid-Liquid Phase Separation

A number of recent studies demonstrate that many proteins with PrLDs undergo concentration-dependent liquid-liquid phase separation under a variety of conditions, including changes in salt concentration [34-37], RNA concentration [34-39], poly(ADP)-ribose [40], pH [38], and temperature [34, 36, 38]. In many cases, the PrLDs play a direct role in liquid-liquid de-mixing [34-37, 39-41], or affect de-mixing indirectly [38]. This biophysical process has been proposed as the basis for the formation of various non-membrane-bound organelles, and can occur in both the cytoplasm (e.g., stress granules and P-bodies) and the nucleus (e.g. nucleolus, nucleolar subcompartments, cajal bodies, and nuclear paraspeckles, etc.) – for review, see [42-44]. Protein de-mixing greatly enhances the local concentrations of incorporated proteins, which may increase the likelihood of forming stable amyloid-like aggregates. Mutations within PrLDs may also perturb the dynamics of these assemblies [34, 36, 38-41], which can result in the formation of off-pathway pathological aggregates. It is also unclear whether differences in subcellular environments (particularly, between the nucleus and the cytoplasm) can affect PrLDs differently based on their sequence and composition. Further investigation is necessary to develop a more complete mechanistic understanding of the biophysical conditions and protein features that governing liquid-liquid phase separation, which may ultimately inform prion prediction methods.

Aggregation, Proteostasis, and Aging

For the vast majority of proteins, misfolding and aggregation represents a cytotoxic threat to cells. Accordingly, cells possess an arsenal of protein quality control factors to maintain proper proteostasis. We've begun to dissect the protein features that make prion and prion-like

domains susceptible or resistant to proteasome-mediated degradation. The ability to evade the proteostasis machinery may be an important protein trait when considering aggregation propensity *in vivo*. Furthermore, cellular proteostasis typically declines with age [45]. Therefore, aggregation-promoting mutations in PrLDs coupled with an age-related decline in cellular proteostasis may contribute to the late-onset nature of many neurodegenerative diseases. A deeper understanding of the rules governing cellular recognition and handling systems of aggregation-prone proteins may facilitate improvements in prion prediction methods.

Nucleocytoplasmic Shuttling of Prion-Like Proteins

Many of the disease-linked PrLD-containing proteins are predominantly nuclear, but form cytoplasmic inclusions in diseased patients. This has led some to propose that disruption of nucleocytoplasmic transport is a key mediator of disease pathology (for review, see [46-48]). Aberrant transport dynamics of PrLD-containing proteins could lead to accumulation of the protein in the cytosol, which may increase aggregation risk. A particularly well-characterized example is the PrLD-containing protein, fused in sarcoma (FUS); many ALS-associated mutations in FUS directly impair its nuclear import and lead to accumulation in cytoplasmic aggregates [49-51]. Moreover, expansion of a hexanucleotide repeat in the *C9orf72* gene (the most common genetic cause of familial ALS) leads to cytotoxic disruptions in nucleocytoplasmic transport [52-55], perturbation of stress granule dynamics [56-61], and proteasome impairment [62-64], which in many cases were shown to affect the localization or aggregation of PrLD-containing proteins. Although it is now clear that nucleocytoplasmic transport is affected in disease, it is currently unclear whether subcellular localization and trafficking dynamics directly influence or simply coincide with protein aggregation. Understanding how mutations alter

intracellular trafficking dynamics may shed light on why many of these proteins form cytoplasmic inclusions.

Post-Translational Regulation of Prion-Like Proteins

Post-translational modifications within or near PrLDs could affect intrinsic aggregation propensity, liquid-liquid phase separation, intracellular trafficking, targeted degradation, or a variety of other processes. The human PrLDs tend to contain a high proportion of glutamine, asparagine, serine, and tyrosine residues [65], which can be modified enzymatically or oxidatively. Additionally, hnRNPA1 and hnRNPA2 PrLDs (as well as other PrLD-containing proteins, including FUS, EWSR1, and TAF15 – for review, see [66]) possess a small number of arginine residues within or near the PrLD's that are asymmetrically di-methylated [67, 68]. Methylation of these residues in the hnRNP's or similar PrLD-containing proteins has been shown to regulate nucleocytoplasmic shuttling [69-73], liquid-liquid phase separation [74], stress granule assembly [75], and pathological protein aggregation [49, 70]. Therefore, post-translational modifications can simultaneously affect a multitude of intracellular processes that relate to protein aggregation either directly or indirectly.

CONCLUSION

Our intention is to highlight the interplay between intrinsic prion domain features and prion formation in a cellular context. A generalized understanding of the features governing these prion-modifying pathways may offer a way to explicitly weight prion predictions with additional intracellular considerations. This will be especially important when extending prion

predictions to multicellular eukaryotic organisms, in which organism-specific and tissue-specific intracellular conditions may differ substantially.

REFERENCES

1. Masison, D.C. and R.B. Wickner, *Prion-inducing domain of yeast Ure2p and protease resistance of Ure2p in prion-containing cells.* Science, 1995. **270**(5233): p. 93-5.
2. Sondheimer, N. and S. Lindquist, *Rnq1: an epigenetic modifier of protein function in yeast.* Mol. Cell, 2000. **5**(1): p. 163-72.
3. Ter-Avanesyan, M.D., et al., *The SUP35 omnipotent suppressor gene is involved in the maintenance of the non-Mendelian determinant [psi+] in the yeast Saccharomyces cerevisiae.* Genetics, 1994. **137**(3): p. 671-6.
4. Wickner, R.B., *[URE3] as an altered URE2 protein: evidence for a prion analog in Saccharomyces cerevisiae.* Science, 1994. **264**(5158): p. 566-9.
5. Ross, E.D., U. Baxa, and R.B. Wickner, *Scrambled Prion Domains Form Prions and Amyloid.* Mol. Cell. Biol., 2004. **24**(16): p. 7206-7213.
6. Ross, E.D., et al., *Primary sequence independence for prion formation.* Proc. Natl. Acad. Sci. USA, 2005. **102**(36): p. 12825-12830.
7. Alberti, S., et al., *A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins.* Cell, 2009. **137**(1): p. 146-158.
8. Harrison, P.M. and M. Gerstein, *A method to assess compositional bias in biological sequences and its application to prion-like glutamine/asparagine-rich domains in eukaryotic proteomes.* Genome Biol, 2003. **4**(6): p. R40.
9. Michelitsch, M.D. and J.S. Weissman, *A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions.* Proc. Natl. Acad. Sci. USA, 2000. **97**(22): p. 11910-5.
10. Du, Z., et al., *Newly identified prion linked to the chromatin-remodeling factor Swi1 in Saccharomyces cerevisiae.* Nat Genet, 2008. **40**(4): p. 460-5.
11. Halfmann, R., et al., *Prion formation by a yeast GLFG nucleoporin.* Prion, 2012. **6**(4).
12. Patel, B.K., J. Gavin-Smyth, and S.W. Liebman, *The yeast global transcriptional co-repressor protein Cyc8 can propagate as a prion.* Nat Cell Biol, 2009. **11**(3): p. 344-9.
13. Rogoza, T., et al., *Non-Mendelian determinant [ISP+] in yeast is a nuclear-residing prion form of the global transcriptional regulator Sfp1.* Proc. Natl. Acad. Sci. USA, 2010. **107**(23): p. 10573-7.
14. Toombs, J.A., B.R. McCarty, and E.D. Ross, *Compositional determinants of prion formation in yeast.* Mol. Cell. Biol., 2010. **30**(1): p. 319-332.
15. Toombs, J.A., et al., *De novo design of synthetic prion domains.* Proc Natl Acad Sci U S A, 2012. **109**(17): p. 6519-6524.
16. King, O.D., A.D. Gitler, and J. Shorter, *The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease.* Brain Res, 2012: p. Epub ahead of print.
17. Kim, H.J., et al., *Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS.* Nature, 2013. **495**(7442): p. 467-73.
18. Paul, K.R., et al., *The effects of mutations on the aggregation propensity of the human prion-like protein hnRNPA2B1.* Mol Cell Biol, 2017.
19. Gonzalez Nelson, A.C., et al., *Increasing prion propensity by hydrophobic insertion.* PLoS One, 2014. **9**(2): p. e89286.

20. MacLea, K.S., et al., *Distinct amino acid compositional requirements for formation and maintenance of the [PSI(+)] prion in yeast.* Mol Cell Biol, 2015. **35**(5): p. 899-911.
21. Chiti, F., et al., *Rationalization of the effects of mutations on peptide and protein aggregation rates.* Nature, 2003. **424**(6950): p. 805-8.
22. Moelbert, S., E. Emberly, and C. Tang, *Correlation between sequence hydrophobicity and surface-exposure pattern of database proteins.* Protein Science, 2004. **13**(3): p. 752-762.
23. Flynn, G.C., et al., *Peptide-binding specificity of the molecular chaperone BiP.* Nature, 1991. **353**(6346): p. 726-30.
24. Karagoz, G.E., et al., *Hsp90-Tau Complex Reveals Molecular Basis for Specificity in Chaperone Action.* Cell, 2014. **156**(5): p. 963-974.
25. Karagoz, G.E. and S.G.D. Rudiger, *Hsp90 interaction with clients.* Trends in Biochemical Sciences, 2015. **40**(2): p. 117-125.
26. Li, Y.L., X.F. Gao, and L.L. Chen, *GroEL Recognizes an Amphipathic Helix and Binds to the Hydrophobic Side.* Journal of Biological Chemistry, 2009. **284**(7): p. 4324-4331.
27. Rudiger, S., A. Buchberger, and B. Bukau, *Interaction of Hsp70 chaperones with substrates.* Nat Struct Biol, 1997. **4**(5): p. 342-9.
28. Rudiger, S., et al., *Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries.* EMBO J, 1997. **16**(7): p. 1501-7.
29. Rudiger, S., J. Schneider-Mergener, and B. Bukau, *Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone.* Embo Journal, 2001. **20**(5): p. 1042-1050.
30. Saio, T., et al., *Structural Basis for Protein Antiaggregation Activity of the Trigger Factor Chaperone.* Science, 2014. **344**(6184): p. 597-+.
31. Willmund, F., et al., *The cotranslational function of ribosome-associated Hsp70 in eukaryotic protein homeostasis.* Cell, 2013. **152**(1-2): p. 196-209.
32. Fredrickson, E.K., et al., *Substrate recognition in nuclear protein quality control degradation is governed by exposed hydrophobicity that correlates with aggregation and insolubility.* J Biol Chem, 2013. **288**(9): p. 6130-9.
33. Fredrickson, E.K., et al., *Exposed hydrophobicity is a key determinant of nuclear quality control degradation.* Mol Biol Cell, 2011. **22**(13): p. 2384-95.
34. Conicella, A.E., et al., *ALS Mutations Disrupt Phase Separation Mediated by alpha-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain.* Structure, 2016. **24**(9): p. 1537-49.
35. Lin, Y., et al., *Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins.* Mol Cell, 2015. **60**(2): p. 208-19.
36. Molliex, A., et al., *Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization.* Cell, 2015. **163**(1): p. 123-33.
37. Zhang, H., et al., *RNA Controls PolyQ Protein Phase Transitions.* Mol Cell, 2015. **60**(2): p. 220-30.
38. Riback, J.A., et al., *Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response.* Cell, 2017. **168**(6): p. 1028-1040 e19.
39. Xiang, S., et al., *The LC Domain of hnRNPA2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets, and Nuclei.* Cell, 2015. **163**(4): p. 829-39.
40. Patel, A., et al., *A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation.* Cell, 2015. **162**(5): p. 1066-77.

41. Murakami, T., et al., *ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function.* Neuron, 2015. **88**(4): p. 678-690.
42. Bergeron-Sandoval, L.P., N. Safaee, and S.W. Michnick, *Mechanisms and Consequences of Macromolecular Phase Separation.* Cell, 2016. **165**(5): p. 1067-1079.
43. Brangwynne, C.P., P. Tompa, and R.V. Pappu, *Polymer physics of intracellular phase transitions.* Nature Physics, 2015. **11**(11): p. 899-904.
44. Hyman, A.A., C.A. Weber, and F. Julicher, *Liquid-liquid phase separation in biology.* Annu Rev Cell Dev Biol, 2014. **30**: p. 39-58.
45. Labbadia, J. and R.I. Morimoto, *The biology of proteostasis in aging and disease.* Annu Rev Biochem, 2015. **84**: p. 435-64.
46. Boeynaems, S., et al., *Inside out: the role of nucleocytoplasmic transport in ALS and FTLD.* Acta Neuropathol, 2016. **132**(2): p. 159-73.
47. Dormann, D. and C. Haass, *TDP-43 and FUS: a nuclear affair.* Trends Neurosci, 2011. **34**(7): p. 339-48.
48. Jovicic, A., J.W. Paul, 3rd, and A.D. Gitler, *Nuclear transport dysfunction: a common theme in amyotrophic lateral sclerosis and frontotemporal dementia.* J Neurochem, 2016. **138 Suppl 1**: p. 134-44.
49. Dormann, D., et al., *ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import.* EMBO J, 2010. **29**(16): p. 2841-57.
50. Ito, D., et al., *Nuclear transport impairment of amyotrophic lateral sclerosis-linked mutations in FUS/TLS.* Ann Neurol, 2011. **69**(1): p. 152-62.
51. Vance, C., et al., *ALS mutant FUS disrupts nuclear localization and sequesters wild-type FUS within cytoplasmic stress granules.* Hum Mol Genet, 2013. **22**(13): p. 2676-88.
52. Boeynaems, S., et al., *Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD.* Sci Rep, 2016. **6**: p. 20877.
53. Freibaum, B.D., et al., *GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport.* Nature, 2015. **525**(7567): p. 129-33.
54. Zhang, K., et al., *The C9orf72 repeat expansion disrupts nucleocytoplasmic transport.* Nature, 2015. **525**(7567): p. 56-61.
55. Jovicic, A., et al., *Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS.* Nature Neuroscience, 2015. **18**(9): p. 1226-+.
56. Boeynaems, S., et al., *Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics.* Mol Cell, 2017. **65**(6): p. 1044-1055 e5.
57. Dafinca, R., et al., *C9orf72 Hexanucleotide Expansions Are Associated with Altered Endoplasmic Reticulum Calcium Homeostasis and Stress Granule Formation in Induced Pluripotent Stem Cell-Derived Neurons from Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia.* Stem Cells, 2016. **34**(8): p. 2063-78.
58. Lee, K.H., et al., *C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles.* Cell, 2016. **167**(3): p. 774-788 e17.
59. Maharjan, N., et al., *C9ORF72 Regulates Stress Granule Formation and Its Deficiency Impairs Stress Granule Assembly, Hypersensitizing Cells to Stress.* Mol Neurobiol, 2016.
60. Tao, Z., et al., *Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity.* Hum Mol Genet, 2015. **24**(9): p. 2426-41.

61. Wen, X., et al., *Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death.* Neuron, 2014. **84**(6): p. 1213-25.
62. Gupta, R., et al., *The Proline/Arginine Dipeptide from Hexanucleotide Repeat Expanded C9ORF72 Inhibits the Proteasome.* eNeuro, 2017. **4**(1).
63. Yamakawa, M., et al., *Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS.* Hum Mol Genet, 2015. **24**(6): p. 1630-45.
64. Zhang, Y.J., et al., *Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress.* Acta Neuropathol, 2014. **128**(4): p. 505-24.
65. Cascarina, S.M. and E.D. Ross, *Yeast prions and human prion-like proteins: sequence features and prediction methods.* Cell Mol Life Sci, 2014.
66. Thandapani, P., et al., *Defining the RGG/RG motif.* Mol Cell, 2013. **50**(5): p. 613-23.
67. Beyer, A.L., et al., *Identification and characterization of the packaging proteins of core 40S hnRNP particles.* Cell, 1977. **11**(1): p. 127-38.
68. Wilk, H.E., et al., *The core proteins of 35S hnRNP complexes. Characterization of nine different species.* Eur J Biochem, 1985. **146**(1): p. 71-81.
69. Araya, N., et al., *Transcriptional down-regulation through nuclear exclusion of EWS methylated by PRMT1.* Biochem Biophys Res Commun, 2005. **329**(2): p. 653-60.
70. Fujii, S., et al., *Treatment with a Global Methyltransferase Inhibitor Induces the Intranuclear Aggregation of ALS-Linked FUS Mutant In Vitro.* Neurochem Res, 2016. **41**(4): p. 826-35.
71. Nichols, R.C., et al., *The RGG domain in hnRNP A2 affects subcellular localization.* Exp Cell Res, 2000. **256**(2): p. 522-32.
72. Tradewell, M.L., et al., *Arginine methylation by PRMT1 regulates nuclear-cytoplasmic localization and toxicity of FUS/TLS harbouring ALS-linked mutations.* Hum Mol Genet, 2012. **21**(1): p. 136-49.
73. Suarez-Calvet, M., et al., *Monomethylated and unmethylated FUS exhibit increased binding to Transportin and distinguish FTLD-FUS from ALS-FUS.* Acta Neuropathol, 2016. **131**(4): p. 587-604.
74. Nott, T.J., et al., *Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles.* Mol Cell, 2015. **57**(5): p. 936-47.
75. Wall, M.L. and S.M. Lewis, *Methylarginines within the RGG-Motif Region of hnRNP A1 Affect Its IRES Trans-Acting Factor Activity and Are Required for hnRNP A1 Stress Granule Localization and Formation.* J Mol Biol, 2017. **429**(2): p. 295-307.

APPENDIX I: INCREASING PRION PROPENSITY BY HYDROPHOBIC INSERTION[5]

INTRODUCTION

Prions are protein-based infectious agents, caused by proteins capable of adopting an alternate, self-propagating amyloid-like structure. In mammals, misfolding of the prion protein PrP is responsible for the transmissible spongiform encephalopathies (TSEs), all of which are incurable and fatal [1]. Additionally, many other non-infectious diseases also involve the aggregation of proteins into amyloid deposits. In fungi, a number of proteins can adopt a prion state. The filamentous fungus *P. anserina* carries a prion protein, Het-S [2], that acts as part of a heterokaryon incompatibility mechanism. The yeast *Saccharomyces cerevisiae* carries at least nine proteins that convert to a prion state [3].

Yeast prions provide a useful model system for examining how amino acid sequence affects amyloid and prion propensity. For all but one of the amyloid-based yeast prion proteins, a glutamine/asparagine (Q/N) rich prion-forming domain (PFD) drives prion formation. Intriguingly, in the past few years, a number of proteins with prion-like domains (domains compositionally resembling the yeast PFDs) have been linked to various age-related degenerative disorders [4]: cytoplasmic inclusions containing FUS and TDP-43 are seen in both ALS and some forms of FTLD, and mutations in these proteins have been linked to some familial cases of ALS [5,6,7]; TAF15 and EWSR1 have separately been connected to ALS and FTLD [8,9,10]; mutations in hnRNPA1 and hnRNPA2/B1 cause IBMPFD/ALS (inclusion body myopathy with frontotemporal dementia, Paget's disease of bone, and ALS; [11]); and mutations

[5] This chapter has been reformatted from the following publication: Gonzalez-Nelson AC, Paul KR, Petri M, Flores N, Rogge RA, Cascarina SM, Ross ED. PLoS ONE. (2014) 9(2):e89286. My contribution consisted of development and implementation of all western blot assays examining protein expression levels.

in TIA1 cause Welander distal myopathy [12]. A better understanding of how sequence and composition affect the amyloid propensity of prion-like domains would permit a better understanding of the mechanism of aggregation in these diseases. It would also allow for bioinformatics searches to identify new prion-like domains.

The yeast prion protein Sup35, which forms the [*PSI*$^{+}$] prion, is an essential subunit of the translation termination complex. Sup35 has three functionally distinct domains [13,14,15]. The N-terminal PFD (residues 1-114) is an intrinsically disordered domain that is necessary and sufficient for prion aggregation [13,14,15]. Like other yeast prions, it has high Q/N content and few hydrophobic residues [16]. The M domain (residues 114-253) is a highly charged, intrinsically disordered region that is not required for either prion formation or translation termination activity, but that stabilizes [*PSI*$^{+}$] [17]. The C domain (residues 253-685) is a structured region that is necessary and sufficient for translation termination.

Scrambling the PFD of Sup35 does not prevent prion formation, demonstrating that composition is a dominant variable affecting prion propensity [18]. A number of search algorithms to identify new prion proteins have been developed that take advantage of this fact by testing for compositional similarity to known PFDs [16,19,20]. Several prions were discovered using these methods [19,21,22]. However, this approach has limitations. Alberti *et al*. identified 100 yeast domains that had the greatest compositional similarity to existing PFDs and tested them for amyloid and prion-like activity using four different assays [19]. Remarkably, eighteen behaved as prions in all four assays. However, there was almost no correlation between the prion-forming ability of the 100 tested domains and their compositional similarity to existing PFDs [23,24]. Therefore, while this algorithm is very effective at identifying prion candidates, it was ineffective at distinguishing among these candidates.

To better understand how composition affects prion propensity, we developed a method to quantify the prion-forming propensity of each amino acid in the context of a Q/N-rich PFD [24]. We replaced an eight amino acid segment from a scrambled Sup35 with a random library of sequences. We then selected for the subset of sequences that could form prions; the prion propensity of each amino acid was determined by comparing the frequency of occurrence of the amino acid among the prion-forming sequences to the frequency of the amino acid in the starting library. These prion propensity values were then used to build the prediction algorithm PAPA (Prion Aggregation Prediction Algorithm; [25,26]).

PAPA is quite effective at discriminating between Q/N-rich domains with and without prion activity [24]. However, some of the individual prion propensity values for specific amino acids were quite surprising. As expected, charged residues and prolines were under-represented among prion-forming clones, consistent with their relative rarity in yeast PFDs [24]. Unexpectedly, Q/N residues were relatively neutral despite their prevalence in yeast PFDs, while hydrophobic residues, which are rare in yeast PFDs [16], were strongly over-represented among prion-forming clones, suggesting that they strongly promote prion activity.

This high predicted prion propensity for hydrophobic residues is particularly intriguing. The strong under-representation of hydrophobic residues in yeast PFDs would seem to suggest that these residues inhibit prion activity in the context of Q/N-rich domains. Indeed, any algorithm that uses compositional similarity to known PFDs to identify new prion proteins is predicated on the assumption that compositional changes that reduce the biases seen in known PFDs will reduce prion propensity; thus, such algorithms assume that increasing hydrophobic content will reduce prion propensity. At the same time, hydrophobic residues have long been thought to promote amyloid formation in the context of non-Q/N-rich proteins [27], although the

applicability of these results to Q/N-rich proteins is unclear. Specifically, a number of algorithms, including Waltz [28], Zyggregator [29], ZipperDB [30], and TANGO [31], have been developed that can accurately predict the aggregation propensity of non-Q/N-rich domains; each of these algorithms favors hydrophobic residues, yet none of these algorithms are able to distinguish between Q/N-rich proteins with and without prion activity, making it unclear the extent to which results from non-Q/N-rich amyloid proteins can be applied to Q/N-rich proteins.

A recent study raised further doubts about the ability of hydrophobic residues to promote prion activity. Although expanded poly-glutamine tracts show high aggregation propensity, they do not propagate efficiently as prions in yeast because they are poorly fragmented by the chaperone machinery [32]; such fragmentation is required to maintain prions over multiple generations of cell division. Alexandrov *et al*. recently showed that insertion of aromatic residues into poly-Q tracts promotes fiber fragmentation, but that non-aromatic hydrophobic residues do not exert the same positive effect [33].

There are a number of possible hypotheses that could explain why non-aromatic hydrophobic residues are so rare in yeast PFDs and fail to promote prion activity when inserted into poly-Q tracts, yet showed high prion propensities in the screen used to develop PAPA. The simplest explanation is that the predicted prion-propensity values are either an artifact of the region tested or simply inaccurate. For example, because the prion propensity values for each amino acid were derived by random sampling, these values have large confidence intervals, so the non-aromatic hydrophobic residues may simply be less prion-prone than we predicted [24]. However, we hypothesized a more nuanced explanation. Aggregation and prion maintenance are distinct activities that appear to have distinct compositional requirements [34]. The Alexandrov experiments focused on prion maintenance. By contrast, the PAPA scores do not separate these

two activities, so likely reflect some combination of the two. While non-aromatic hydrophobic residues appear unable to promote fiber fragmentation, they may still promote prion formation; in this case, aromatic hydrophobic residues may simply be favored in yeast PFDs because they can serve a dual role, promoting both prion formation and prion maintenance. To test this hypothesis, we specifically examined the effects of non-aromatic hydrophobic residues on prion formation by Sup35. We found that non-aromatic hydrophobic residues can promote prion formation to a remarkable degree. These results, combined with bioinformatics analysis of prion and non-prion Q/N-rich domains, provide insight into a number of unanswered questions about the sequence basis for prion activity.

EXPERIMENTAL PROCEDURES

Strains and Media

Standard yeast media and methods were used, as described previously [35], except that yeast extract-peptone-dextrose (YPD) contained 0.5% yeast extract instead of the standard 1%. In all experiments, yeast were grown at 30°C. Experiments were performed with *Saccharomyces cerevisiae* strain YER632/pJ533 (α *kar1-1 SUQ5 ade2-1 his3 leu2 trp1 ura3 sup35*::KanMx [*psi*$^-$] [*PIN*$^+$]; pJ533 expresses *SUP35* from a *URA3* plasmid as the sole copy of *SUP35* in the cell), a [*psi*$^-$] version of 780-1D/pJ533 [36].

Design of the Mutants

For the hydrophobic insertions, the Excel random number function was used to select positions for insertion between amino acids 8-24 of Sup35. In each case, an equal number of

isoleucines and valines were inserted. For the tyrosine deletions, the Excel random number function was likewise used to select which tyrosines should be deleted.

Cloning

CEN plasmids expressing full-length Sup35 mutants from the *SUP35* promoter were generated using homologous recombination. The mutations were inserted into the N domain of *SUP35* in two steps. For each mutant, two PCR reactions were set up. The N-terminal portion of *SUP35* was amplified with EDR302 and a mutant-specific primer, while the C-terminal portion of *SUP35* was amplified with EDR262 and a second mutant-specific primer (see Supplemental Table S1 for a complete list of primer sequences). Products of these two reactions were combined and reamplified with EDR301 and EDR262. The final PCR products were co-transformed with HindIII/BamHI-cut pJ526 [37] into yeast strain YER632/pJ533. Transformations were selected on SC-Leu, and then transferred to FOA plates to select for loss of pJ533.

To generate induction plasmids, the NM domain of each mutant was amplified by PCR using primers EDR1008 and EDR1084. EDR1084 installs a stop codon and XhoI restriction site at the end of the middle (M) domain, while EDR1008 installs a BamHI restriction site before the Sup35 start codon. PCR products were digested with BamHI and XhoI, and then inserted into BamHI/XhoI cut pKT24, a *TRP1* 2μm plasmid containing the *GAL1* promoter [37]. Ligation products were transformed into Escherichia coli and analyzed by DNA sequencing.

To generate vectors expressing GFP fusions, first the cassette containing the *GAL1* promoter and *ADH1* terminator was amplified from pKT24 using primers EDR1747 and EDR1748, which install SphI and EcoRI sites, respectively. This product was digested with SphI

and EcoRI and then inserted into SphI/EcoRI cut YEplac112 [38] to generate plasmid pER687. Yeast-optimized GFP was then amplified from pYGFP [39] using primers EDR1898 and EDR1899, which add BamHI and SalI restriction sites, respectively, to the 5' and 3' ends of GFP. PCR products were digested with BamHI and SalI, and then inserted into BamHI/XhoI cut pER687, generating plasmid pER760. The NM domain of the Sup35 mutants were then amplified with EDR1008 and EDR1924, which add BamHI and XhoI restriction sites, respectively, to the 5' and 3' ends of the Sup35 NM. PCR products were digested with BamHI and XhoI, and then inserted into BamHI/XhoI cut pER760.

Western Blot

Western blots were performed as previously described ([40]), using a monoclonal antibody against Sup35's C-terminal domain (BE4 [41], from Cocalico Biologicals, kindly made available by Susan Liebman).

[*PSI*$^+$] *Formation*

For all prion formation assays except those for the hydrophobic rearrangement constructs, strains were transformed with either pKT24 or with a derivative of pKT24 in which the respective PFD was inserted under control of the *GAL1* promoter. Strains were grown for 3 days in galactose/raffinose dropout medium lacking tryptophan to select for pKT24 or the pKT24 derivative. It is not necessary to maintain selection for the plasmid expressing the full-length Sup35 mutant, because this plasmid expresses the only copy of *SUP35* in the cells, and *SUP35* is an essential gene. Serial 10-fold dilutions were spotted onto SC-ade medium to select for [*PSI*$^+$] cells and grown for 5 days. Although new colonies will continue to appear after 5 days, we find

that these colonies tend to be unable to propagate the Ade^+ phenotype when removed from selection.

The hydrophobic rearrangement constructs were only tested under uninduced conditions. These strains were grown in YPAD for 2 days, and then serial 10-fold dilutions were spotted onto SC-ade medium to select for [PSI^+] cells and grown for 5 days.

Protein Expression and Purification

The NM domain of Sup35 was recombinantly expressed on a pET-17b expression vector in BL21 CodonPlus competent cells (Agilent Technologies, CAT#230245). A one liter 2xYT culture was grown to an OD600 of 1.0. Cells were induced with 0.5 mM IPTG for four hours. Cultures were centrifuged and pellets stored at -70˚C. Protein was purified under denaturing conditions in two steps. First, cells were resuspended in lysis buffer (6M GuHCl, 0.1 M KH2PO4, 10 mM Tris Base, 0.05% Tween 20, pH 8) and lysed by sonication. The lysate was loaded onto a Ni-NTA sepharose column (GE Healthcare, 17-5286-01). Sup35NM was eluted with imidizole buffer (6M Urea, 0.1 M KH2PO4, 10 mM Tris Base, 0.05% Tween 20, 0.5 M Imidazole, pH 8). Second, fractions containing protein were pooled and diluted 1:4 into loading buffer (6 M Urea, 50 mM MES pH 6.0). Sup35NM was loaded onto an SP Sepharose Ion exchange column and eluted in high salt (6 M Urea, 50 mM MES pH 6.0, 1 M NaCl). Fractions containing Sup35NM protein were pooled and concentrated using an Amicon Ultra centrifugal filter (Fisher, UFC901008). Protein was stored in urea at -70^o C.

In Vitro Amyloid Aggregation Assay

The *in vitro* assays were performed using a protocol adapted from Collins *et al*. [42]. Briefly, reactions were set up as follows: A 96-well plate (Fisher, 07-200-567) was treated with 5% casein solution for five minutes at room temperature, then rinsed with DI water and allowed to dry. Protein and thioflavin-T stock solution were diluted to a final concentration of 5 and 25 µM, respectively, in 50 mM glycine buffer, with a final reaction volume of 200 µl. Fluorescence was monitored in a Victor3 Perkin Elmer fluorescence plate reader, with excitation and emissions wavelengths of 460 and 490 nm, respectively. Reactions were monitored for 48 h. Between readings, reactions were incubated without agitation for 3 minutes, and then shaken for 10 sec. The fraction aggregated was calculated by normalizing relative to the final fluorescence of the well.

Bioinformatics Analysis of the Yeast Proteome

The complete set of systematically-named *Saccharomyces cerevisiae* open reading frames was downloaded from the Saccharomyces Genome Database (http://downloads.yeastgenome.org/sequence/S288C_reference/orf_protein/). To generate a histogram of the compositional distribution of the yeast proteome, the proteome was scanned using a 100 amino acid window size, scoring the amino acid composition of each window.

RESULTS

Insertion of Hydrophobic Residues Increases Prion Formation

Various studies suggest that a key difference between aromatic and hydrophobic residues in the context of Q/N-rich domains is that aromatic residues facilitate the chaperone-dependent

fragmentation that is required for prion maintenance [33,43]. However, the relative effects of aromatic and hydrophobic residues on prion formation are less clear. To specifically focus on the effects of hydrophobic residues on prion formation, we took advantage of the fact that the prion formation and maintenance activities of the Sup35 PFD largely reside in separate regions of the PFD [44]. The first 40 amino acids are highly enriched in Q/N residues and are required for prion nucleation and fiber growth, while amino acids 40-114 are thought to be primarily involved in prion maintenance [34,44,45,46,47]. To test the effect of non-aromatic hydrophobic residues on prion formation, we generated four constructs in which we inserted isoleucine or valine at random positions between residues 8-24 of Sup35, a region of the nucleation domain that is particularly important for prion activity [45]. Isoleucine and valine were chosen because they score as the most prion-promoting non-aromatic amino acids according to PAPA; leucine actually scores as slightly prion-inhibiting, likely due to its low β-sheet propensity [24]. Four *SUP35* mutants were generated (Figure 6.1A, B): two in which two hydrophobic residues were inserted into random locations in the nucleating domain of Sup35p (called +2HydA and +2HydB), and two in which six hydrophobic residues were inserted (+6HydA and +6HydB).

Each mutant was cloned into a CEN plasmid under the control of the *SUP35* promoter. These plasmids were shuffled into a yeast strain that lacks an endogenous copy of *SUP35*, but carries a maintainer copy expressed from a *URA3* plasmid. After selection for loss of the maintainer plasmid, strains were tested for their propensity to convert to [*PSI*$^{+}$]. [*PSI*$^{+}$] was detected by monitoring nonsense suppression of the *ade2-1* allele [48]. [*psi*$^{-}$] *ade2-1* mutants cannot grow in the absence of adenine and form red colonies in the presence of limiting adenine due to accumulation of a pigment derived from the substrate of Ade2. However, [*PSI*$^{+}$] allows

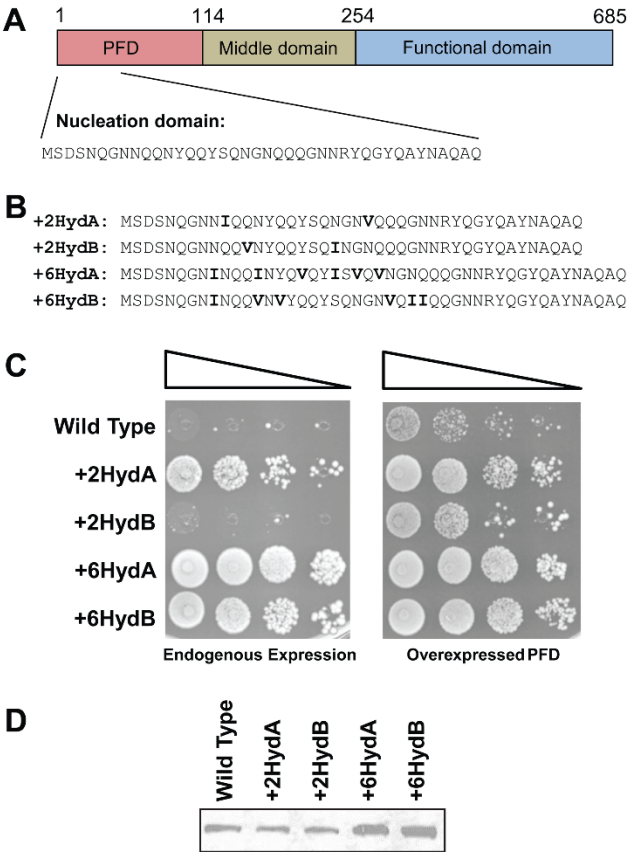


Figure 6.1: Insertion of hydrophobic residues increases prion formation. (A) Schematic of Sup35. The sequence of the nucleation domain (amino acids 1-40) is shown. **(B)** Sequences of the nucleation domains of each of the hydrophobic-addition constructs. Inserted hydrophobic residues are indicated in bold. For each, the remainder of the protein is the same as wild-type Sup35. **(C)** Prion formation by each construct. Strains expressing the indicated Sup35 mutants as the sole copy of Sup35 were transformed either with an empty vector (*left*) or with a plasmid expressing the matching Sup35 mutant under control of the *GAL1* promoter (*right*). All strains were cultured for three days in galactose/raffinose dropout medium, and then 10-fold serial dilutions were plated onto medium lacking adenine to select for [*PSI*$^{+}$]. **(D)** Western blot of wild-type and mutant Sup35.

for occasional read through of the *ade2-1* nonsense mutation. Thus, [*PSI*$^+$] cells can grow in the absence of adenine, and grow white in the presence of limiting adenine.

Insertion of hydrophobic residues substantially increased the frequency of Ade$^+$ colony formation (Figure 6.1C). Spontaneous wild-type prion formation is an extremely rare event, occurring in approximately one cell per million when Sup35 is expressed at endogenous levels [49]. Efficient [*PSI*$^+$] formation requires PFD overexpression, which increases the pool of soluble protein, thereby increasing the probability of the nucleation events that initiate prion formation (Figure 6.1C, right versus left panel) [50]. By contrast, the addition of six hydrophobic residues generated strains that appeared to be constantly [*PSI*$^+$], even in the absence of PFD overexpression (Figure 6.1C, left panel). This prion-promoting effect was so strong that the cells were even able to form [*PSI*$^+$] in the absence of [*PIN*$^+$], a prion required for wild-type Sup35 to form prions (Figure S1) [51]. Even just two additional hydrophobic residues caused a significant increase in frequency of Ade$^+$ colony formation, with roughly one in ten cells expressing the +2HydA construct forming Ade$^+$ colonies in the absence of PFD overexpression (Figure 6.1C, right versus left panel). This increase was not due to changes in protein levels; although the +6Hyd constructs both showed modestly higher protein levels by western blot than wild-type Sup35, protein levels for the +2Hyd constructs were similar to wild-type Sup35 (Figure 6.1D).

Despite having identical amino acid compositions, the +2HydA and +2HydB constructs showed substantial differences in frequency of prion formation. This supports the idea that while amino acid composition is the dominant factor affecting prion propensity, primary sequence also exerts an effect [18,19,37].

To confirm that the Ade$^+$ colonies were due to prion formation, Ade$^+$ isolates from each mutant were tested for stability and curability. Guanidine hydrochloride cures yeast prions by

disrupting the activity of Hsp104 [52,53], a chaperone protein involved in prion propagation [54,55,56]. Individual Ade^+ isolates were streaked on YPD, with and without the addition of 4 mM guanidine. Cells were then tested for loss of [PSI^+] by re-streaking onto medium containing limiting adenine (Figure S2). The majority of the Ade^+ isolates from each of the hydrophobic addition constructs were stably Ade^+ in the absence of guanidine, but lost the Ade^+ phenotype after growth on guanidine (Figure S2), demonstrating that the phenotype was the result of a prion. Therefore, addition of hydrophobic residues dramatically increases the frequency of prion formation, without interfering with prion propagation.

Interestingly, some of the constructs rapidly reverted to the [PSI^+] after curing. The most extreme was the +6A construct. It formed predominantly weak prions, as indicated by a pink phenotype. Although these cells were fully red on guanidine medium (data not shown), upon restreaking onto non-selective medium, they rapidly converted to a mixture of red, white, pink, and sectored colonies.

To ensure that the observed differences in prion formation were due to changes in prion propensity, rather than an artifact such as mislocalization, differences in toxicity, or alteration of a prion-modifying protein-protein interaction, the mutants were purified and assayed for amyloid formation *in vitro* (Figure 6.2). Amyloid aggregation was monitored using Thioflavin T, a dye that forms fluorescent complexes with amyloid fibrils, but not with soluble proteins or amorphous aggregates [57]. In each case, the rate of aggregation *in vitro* correlated well with prion formation in vivo (Figure 6.1C). As expected, wild-type Sup35 had a lag phase lasting approximately 9 hours before aggregation and increased in a roughly sigmoid fashion. Remarkably, +6HydA and +6HydB each showed no detectable lag phase and plateaued within three hours.

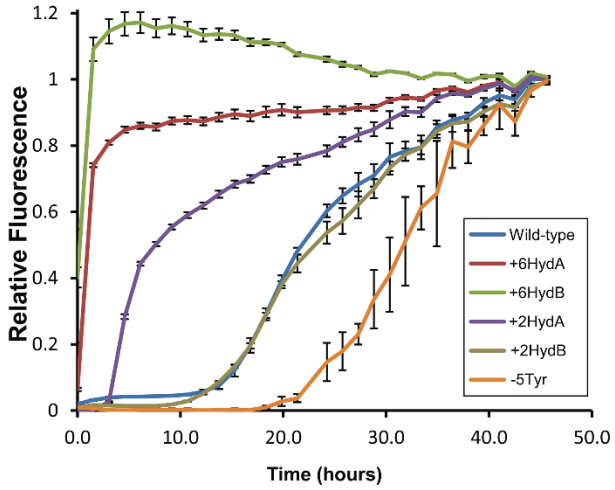


Figure 6.2: In vitro amyloid aggregation of the mutant prion forming domains. Aggregation of purified PFDs was monitored using thioflavin T. Reactions were incubated with intermittent shaking for 48 h. Fluorescent readings were taken approximately every 90 min. Error bars represent the standard deviations of three samples

The Effect of Primary Sequence on Prion Formation

Both +2HydA and +2HydB carry an extra isoleucine and valine. The observed prion formation differences between these compositionally identical constructs demonstrate that small changes in primary sequence can exert substantial effects on prion formation. Therefore, these constructs provide a useful system to explore the basis for such primary sequence effects.

However, systematically repositioning the isoleucine and valine did not reveal any clear trend (Figure 6.3A,B). The constructs showed substantial differences in both the number of Ade^+ colonies observed and the fraction of these colonies that propagated as stable, guanidine-curable prions (Figure 6.3B). Western blot showed only small expression differences among the mutants, and neither the frequency of Ade^+ colony formation nor the stability of the Ade^+ phenotype consistently correlated with expression levels (Figure 6.3C). Unexpectedly, the +2HydD mutant showed two bands: a predominant band at the expected size, and a minor band running at a higher molecular weight. This raises the possibility that a subset of the +2HydD protein pool could be undergoing modification, suggesting that prion formation levels for this mutant should be interpreted with caution.

These large differences in prion activity are not predicted by any of the commonly-used aggregation prediction algorithms. Not surprisingly, composition-based algorithms such as PAPA or Zyggregator were not effective at distinguishing among these constructs. However, while the prion propensity of Q/N-rich domains is predominantly determined by amino acid composition, a variety of evidence suggests that short sequence motifs may play a critical role in nucleating prion formation by Sup35 [58,59,60]. Thus, the mutations may affect prion activity by creating or disrupting amyloid-promoting primary sequence motifs. The Serrano group has used both computational and experimental techniques to determine a consensus hexameric sequence

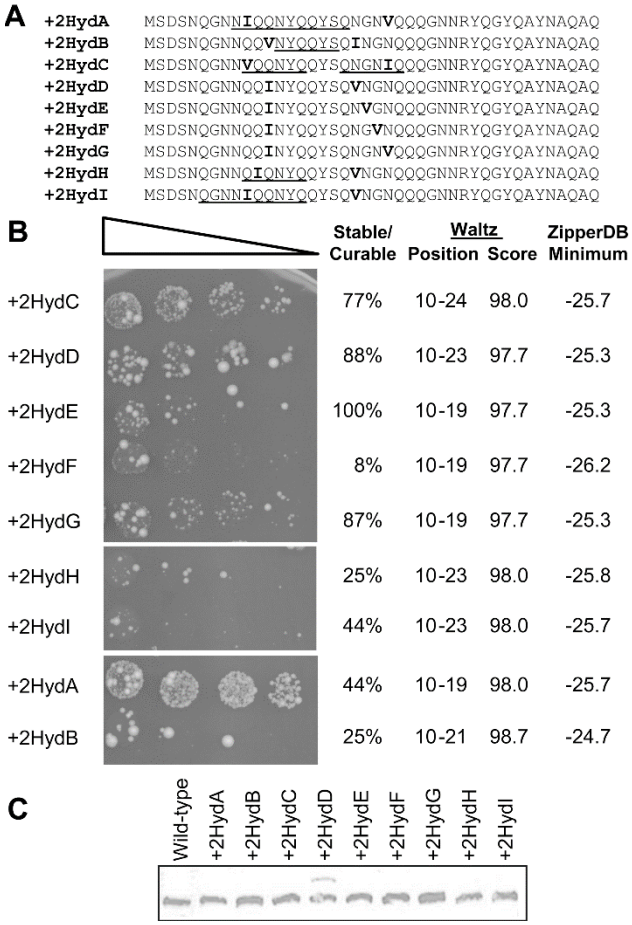


Figure 6.3: Effects of primary sequence on prion formation. (A) Amino acid sequences of constructs in which two additional hydrophobic residues were added at various positions within the Sup35 nucleation domain. For each, the remainder of the protein is the same as wild-type Sup35. Amyloid stretches, as predicted by Lopez de la Paz and Serrano [62], are underlined. The inserted hydrophobic residues are indicated in bold. **(B)** Prion formation by each of the constructs. Strains expressing the indicated Sup35 mutants as the sole copy of Sup35 were grown in YPAD medium for two days, and then 10-fold serial dilutions were plated onto medium lacking adenine to select for [*PSI*$^+$]. For each construct, the position and scores of amyloid stretches predicted by Waltz [28], as well as the minimum ZipperDB score [63], are indicated. Individual Ade$^+$ colonies we picked from each plate and tested for stability and curability, as in Figure S2. Colonies were considered stable and curable if they maintained a white/pink phenotype on YPD, but were red after treatment with guanidine HCl. **(C)** Western blot of expression levels of wild-type and mutant Sup35s.

that promotes amyloid formation [61,62]. Interestingly, the only such stretch in the nucleation domain of Sup35 overlaps with the region mutated in these constructs (Figure 6.3A). However, there did not appear to be any correlation between the presence of such stretches and prion activity (Figure 6.3A,B).

Other prediction algorithms were no more effective. A recent, more comprehensive study has expanded the definition of the hexameric amyloid stretch [28]. The prediction algorithm Waltz utilizes this broader definition and provides quantitative scores for different stretches. However, no correlation was seen between Waltz scores and Ade^+ colony formation. All of the proteins had Waltz-positive segments; although there were differences in the length of these segments, there was no clear correlation between the length or score of the predicted amyloid stretch and observed prion activity. The same was true using the "High Specificity" setting, which is intended to reduce false positives; again, all constructs had Waltz-positive segments overlapping with the mutated region, and there was no correlation between the length of the predicted amyloid stretch and observed prion activity (data not shown).

Similar results were seen for ZipperDB, another algorithm that looks for 6-amino-acid aggregation-prone segments. ZipperDB is a structure-based prediction method. Sequences are threaded into a known NNQQNY amyloid-forming hexapeptide crystal structure and the energetic fit is determined [30,63]. Segments with a free energy below -23 kcal/mol are considered to have high fibrillation propensity; insertion of a single such sequence into a loop region of RNase A was sufficient to cause amyloid formation [64]. All of the +2Hyd constructs had segments well below -23 kcal/mol that overlapped with the mutated region; however, there was no correlation between the predicted free energy and the observed frequency of prion

activity (Figure 6.3B). In short, while it is clear that primary sequence effects do exist, none of the commonly used amyloid prediction algorithms successfully predict these effects.

Deletion of Tyrosine Residues Reduces Prion Formation and Aggregation

The Sup35 nucleation domain shows a striking under-representation of highly hydrophobic residues, completely lacking F, I, L, M, W or V; however, it does contain five tyrosines. Three mutants were generated in which either two (-2TyrA and -2TyrB) or five of these tyrosines (-5Tyr) were eliminated from the nucleation domain (Figure 6.4A). Each was expressed at levels comparable to wild-type Sup35 (Figure 6.4B). None of these mutants showed detectable Ade^+ colony formation when expressed at endogenous levels (Figure 6.4C).

However, because even wild-type Sup35 only rarely forms prions without Sup35 overexpression, it remained possible that the tyrosine deletion mutants were simply forming prions at a frequency below the threshold of detection. Indeed, transient over-expression of the corresponding PFD increased Ade^+ colony formation by each of the strains (Figure 6.4C), suggesting that each of the mutants is capable of prion formation. However, the -5Tyr construct showed substantially reduced frequencies of Ade^+ colony formation (Figure 6.4C). When tested for stability and curability, all Ade^+ colonies isolated from the tyrosine deletion mutants were red after growth both with and without guanidine, indicating that the Ade^+ phenotype is unstable (Figure S3).

Additionally, PFD-GFP fusions showed substantially reduced foci formation. For wild-type Sup35, over-expression of PFD-GFP fusions results in the formation of fluorescent foci (Figure 6.5A). Likewise, large foci were consistently observed for each of the hydrophobic addition constructs (Figure 6.5B). However, no foci were observed in cells expressing -2TyrA or

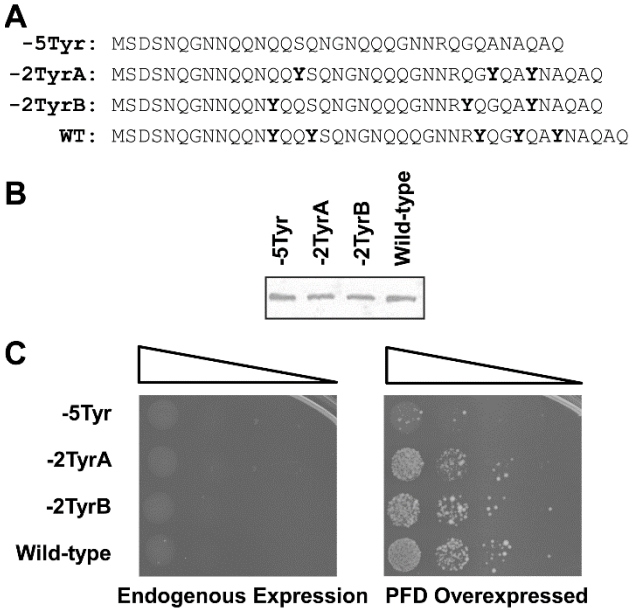


Figure 6.4: Deletion of tyrosine residues reduces prion formation. (A) Amino acid sequences of constructs in which tyrosines were deleted from various positions within the Sup35 nucleation domain. Tyrosines are indicated in bold. **(B)** Western blot of expression levels of wild-type and mutant Sup35s. **(C)** Prion formation by each construct. Strains expressing the indicated Sup35 mutants as the sole copy of Sup35 were transformed either with an empty vector (*left*) or with a plasmid expressing the matching Sup35 mutant under control of the *GAL1* promoter (*right*). All strains were cultured for three days in galactose/raffinose dropout medium, and then 10-fold serial dilutions were plated onto medium lacking adenine to select for [*PSI*$^{+}$].

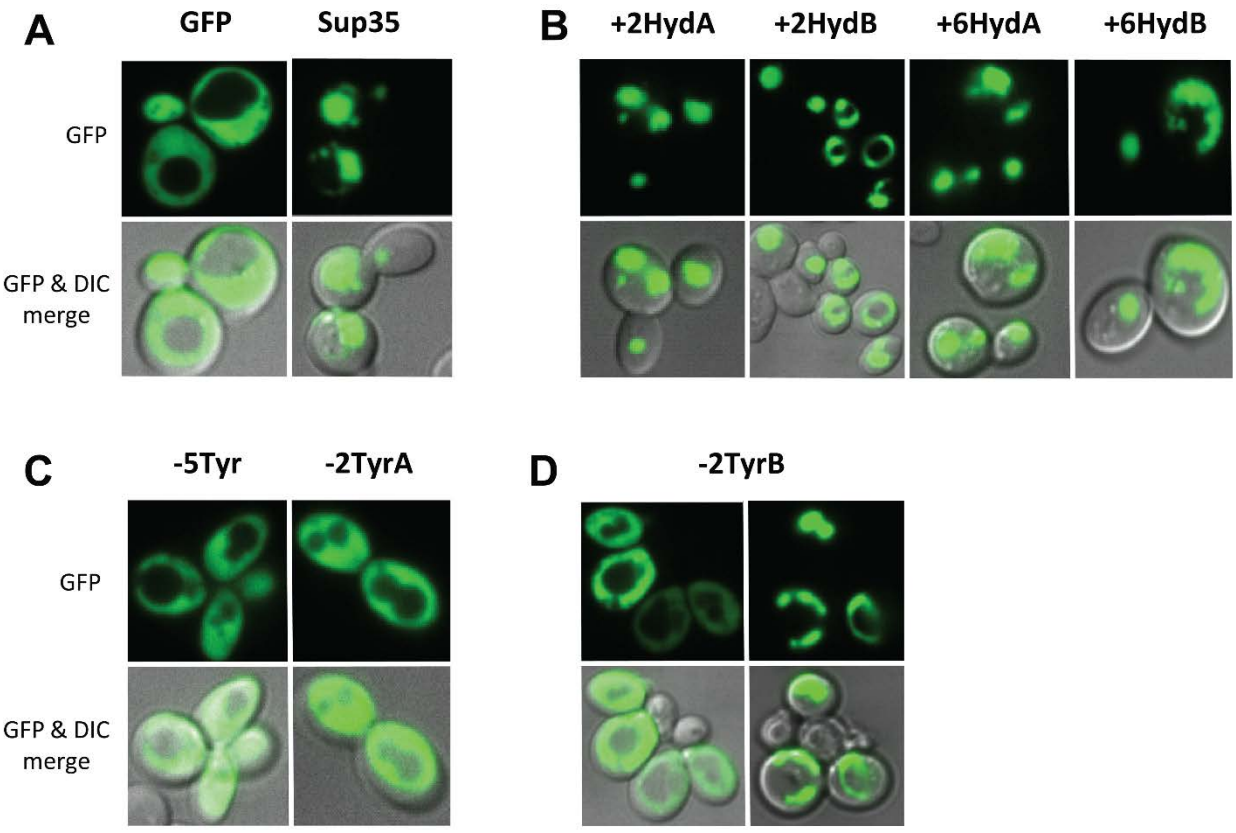


Figure 6.5: Tyrosine and hydrophobic residues promote foci formation. (A) The Sup35 PFD promotes formation of fluorescent foci. GFP or the NM domain from wild-type Sup35 fused to GFP were expressed under control of the *GAL1* promoter. Cells were grown in galactose/raffinose dropout medium for 24 h, and then visualized by confocal microscopy. **(B)** The hydrophobic insertion constructs each support formation of fluorescent foci. Conditions were as described in (A). **(C)** The -5Try and -2TyrA constructs fail to form fluorescent foci. **(D)** The -2TyrB construct forms foci in a fraction of cells.

5Tyr (Figure 6.5C), while foci were observed in only a subset of the cells expressing -2TyrB (Figure 6.5D). Therefore, the tyrosine deletion mutants show substantially reduced in vivo aggregation, and appear completely unable to form stable [*PSI*$^{+}$] prions.

Some prion variants formed by wild-type Sup35 are either deleterious or lethal to yeast cells [65]. Therefore, the reduced prion formation and aggregation by the -5Tyr mutant could theoretically result from an artifact such as an increase in prion toxicity. However, the -5Tyr construct also showed substantially reduced aggregation kinetics *in vitro* (Figure 6.2), suggesting that tyrosine deletion directly affects aggregation propensity.

The Effect of Aromatic Residues on Prion Formation

Although there are very few highly hydrophobic residues in yeast PFDs, tyrosines are over-represented among a subset of yeast PFDs [16]. We directly compared the ability of hydrophobic and aromatic residues to drive prion formation by replacing the five native tyrosines in the Sup35 nucleation domain with leucines, isoleucines, or valines. All three constructs were able to form Ade^{+} colonies, albeit at different frequencies, and all three showed more Ade^{+} colony formation with PFD over-expression than without, consistent with the Ade^{+} colonies resulting from prion formation (Figure 6.6A). However, there were substantial differences both in the frequency of Ade^{+} colony formation (Figure 6.6A) and the fraction of these colonies that propagated as stable, curable prions (Figure 6.6B-E). The valine substitution construct showed the highest frequency of Ade^{+} colony formation (Figure 6.6A); however, the Ade^{+} colonies were consistently unstable, rapidly losing the Ade^{+} phenotype upon growth on non-selective medium (Figure 6.6D). While the construct with isoleucine substitutions showed less Ade^{+} colony formation, the majority of the Ade^{+} colonies were stable and curable (Figure 6.6E). The

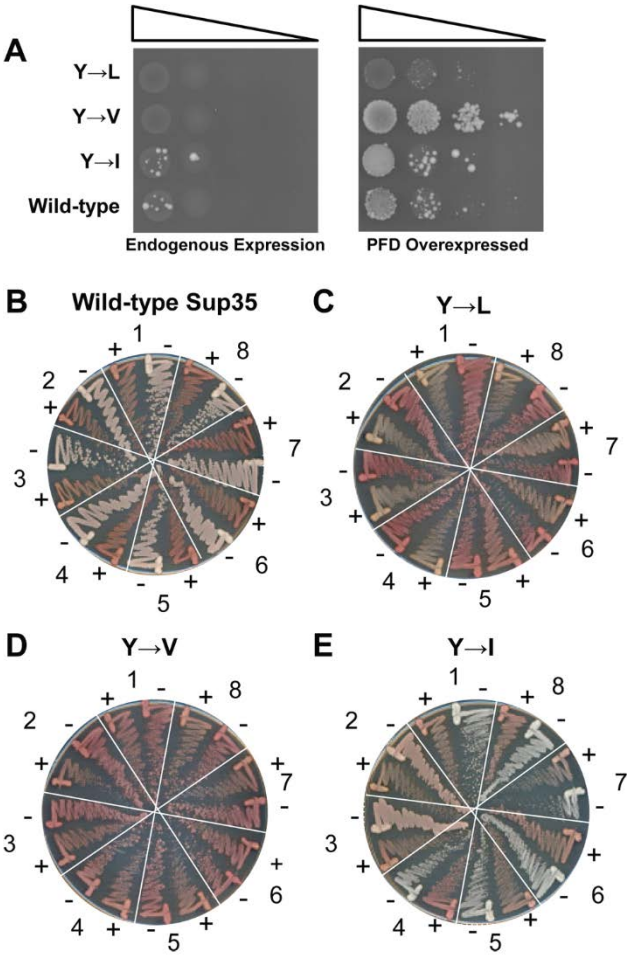


Figure 6.6: Aromatic residues are not required in the Sup35 nucleation domain. (A) The five tyrosines in the Sup35 nucleation domain (amino acids 1-40) were replaced with either leucines, isoleucines or valines. Strains were transformed either with an empty vector (*left*) or with a plasmid expressing the matching Sup35 mutant under control of the *GAL1* promoter (*right*). All strains were cultured for three days in galactose/raffinose dropout medium, and then 10-fold serial dilutions were plated onto medium lacking adenine to select for [*PSI*$^+$]. **(B-E)** Stability and curability of the Ade$^+$ phenotype in cells expressing wild-type Sup35 **(B)**, or Sup35 in which the five tyrosines in the nucleation domain were replaced with leucine **(C)**, valine **(D)** or isoleucine **(E)**. For each mutant, eight individual Ade$^+$ isolates were grown on YPD (-) and YPD plus 4 mM guanidine HCl (+). Cells were then restreaked onto YPD to test for loss of the Ade$^+$ phenotype.

construct with leucine substitutions only formed small Ade+ colonies, and none of the Ade+ isolates were able to maintain the Ade+ phenotype upon non-selective growth (Figure 6.6C). This result is consistent with the relatively low prion propensity of leucine compared to the other hydrophobic residues [24], although it is possible that this failure to form stable prions is a result of a less direct effect, such as increased toxicity of prions formed by this mutant. Therefore, while the identity of the hydrophobic residues within the nucleation domain affects both prion formation and prion stability, there is no strict requirement for aromatic residues within the prion-nucleating domain of Sup35.

Compositional Biases in Glutamine/Asparagine Rich Domains.

The substantial effects of insertion or deletion of hydrophobic and aromatic residues in Sup35 highlight the narrow prion-propensity window required for a protein to act as a prion. According to PAPA, there are six strongly prion-promoting amino acids: F, I, V, Y, M and W. These amino acids have similar prion propensity scores, and are all predicted to be substantially more prion-prone than any other amino acid [24]. The 114-amino-acid Sup35 PFD contains 23 of these prion-promoting amino acids, representing 20.2% of the PFD; increasing this number to 24.2% in the +6Hyd constructs almost completely eliminated the soluble, functional state. It is likely that the exact number of prion-promoting residues required for prion activity is somewhat context-dependent; however, based on this dramatic effect of hydrophobic insertions, we hypothesized that it would be unlikely that any Q/N-rich regions in yeast would contain substantially more prion-promoting residues than Sup35.

Indeed, this appears to be true. Harrison and Gerstein developed an algorithm to identify regions of high compositional bias [16]. They identified 170 regions in the yeast proteome with

strong Q/N bias. There is substantial diversity in these regions; they vary from 25 to 886 amino acids long, and range from 16.8 to 96% Q/N content. Nevertheless, only one of these 170 Q/N-rich domains has more than the 24.2% F, I, V, Y, M and W that is found in the +6Hyd constructs (Figure 6.7A); the lone exception is a fragment from New1, which has 25.5%. By contrast, when the yeast proteome is scanned with a 100-amino-acid window size, over half of all protein fragments have more than the 24.2% F, I, V, Y, M and W. Moreover, although F, I, V, Y, M and W constitute 23.1% of the yeast proteome, the Q/N-rich regions contain on average only 11.9% of these residues.

There are also subtle differences in the frequencies of strongly prion-promoting residues between the prion-forming and non-prion Q/N-rich domains. The Harrison and Gerstein data set includes fragments from 22 proteins with clear prion activity. Eight of these have been proven to act as prions, while an additional fourteen were shown by Alberti *et al*. to have prion-like activity in four independent assays [19]. The non-prion Q/N-rich sequences show both lower average frequencies of strongly prion-promoting residues (Figure 6.7C) and a broader range (Figure 6.7B). Additionally, there appear to be differences in which strongly prion-promoting residues are found in the prion versus non-prion sequences. While the two sets have similar numbers of non-aromatic prion-promoting residues (I, V and M), the prion sequences have substantially more aromatic residues (Figure 6.7C).

The substantial bias against strongly prion-promoting residues in non-prion Q/N-rich domains could simply be a result of high Q/N content. Because these domains average about 45% Q/N residues, the high Q/N content may simply crowd out other residues. However, when the frequency of strongly prion promoting residues is calculated as a percentage of the total number of non-Q/N residues, the prion-promoting amino acids are still slightly under-

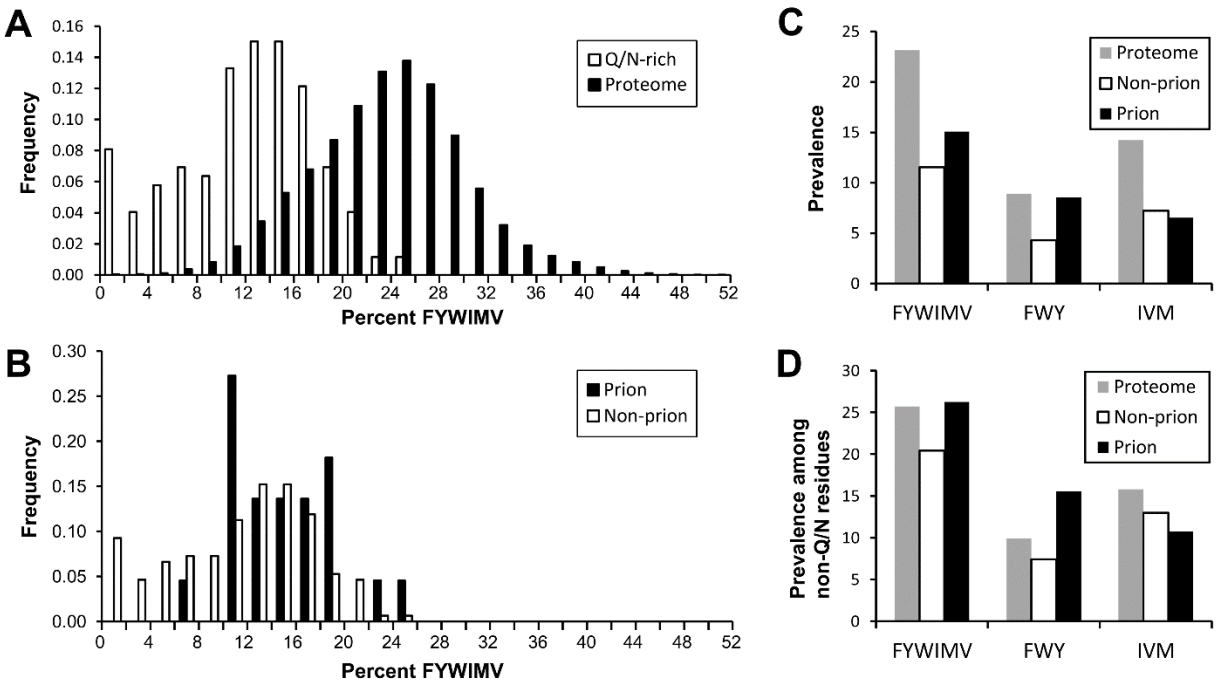


Figure 6.7: Amino acid composition of prion and non-prion Q/N-rich domains. (A) Histogram of the prevalence of strongly prion-promoting residues (FYWIMV) among Q/N-rich proteins (open bars) and among peptide fragments from the yeast proteome (black bars). For the Q/N-rich proteins, each of the regions of the yeast proteome identified by Harrison and Gerstein [16] as having high Q/N-bias were scored for the fraction of strongly prion-promoting amino acids. For the proteomic data, the yeast proteome was scanned using a 100 amino acid window size; each 100-amino-acid window was scored for the fraction of strongly prion-promoting amino acids. **(B)** Histogram of the prevalence of strongly prion-promoting amino acids among yeast prion and non-prion Q/N-rich domains. The black bars include Q/N-rich regions (as identified by Harrison and Gerstein) from yeast proteins shown to act as prions, as well as from proteins containing domains shown by Alberti *et al*. to have prion-like activity in four independent assays [19]. Open bars represent all other yeast Q/N-rich regions identified by Harrison and Gerstein. **(C)** Amino acid prevalence in Q/N-rich domains. Grey bars represent the prevalence of different groups of amino acids in the yeast proteome. Black bars represent the average frequency of these amino acids among Q/N-rich regions from both proteins shown to act as prions and proteins containing domains shown by Alberti *et al*. to have prion-like activity in four independent assays. Open bars represent the average frequency of these amino acids among all other yeast Q/N-rich domains identified by Harrison and Gerstein. **(D)** The prevalence of different groups of amino acids, plotted as a fraction of non-Q/N residues.

represented in non-prion Q/N-rich domains. In the yeast genome, F, I, V, Y, M and W constitute 25.7% of the non-Q/N residues (Figure 6.7D). This is similar to their average frequency in Q/N-rich PFDs; by contrast, the average among non-prion Q/N-rich domains is 20.4%. Thus, even when positions occupied by Q/N residues are excluded from the analysis, strongly prion-promoting amino acids are still slightly under-represented in non-prion Q/N-rich domains.

There are other datasets of Q/N-rich proteins that could have been used for the analysis in Figure 6.7, each with unique strengths and weaknesses. The Harrison and Gerstein data set is useful, because it identifies regions with strong Q/N-bias without imposing any additional compositional requirements. Nevertheless, it has the disadvantage that, because it simply identifies regions with strong statistical bias for Q/N residues, it includes some very large regions that are only modestly enriched for Q/N. However, very similar results were observed with a second data set; Michelitsch and Weissman developed a search algorithm called DIANA. DIANA identified every yeast protein that contains an 80 residue window with at least 30 Q/Ns; then, within these proteins it identified the most Q/N-rich 80 amino-acid segment. While this method is effective for identifying prion candidates, it was not as ideal for our purposes; because the window size is fixed at 80 amino acids, in some cases only a portion of the 80 amino acid segment is Q/N-rich, while in other cases the algorithm may capture only a portion of a long Q/N-rich segment. Nevertheless, the same basic trends were seen in this data set as in the Harrison and Gerstein set (Figure S4).

DISCUSSION

We previously scored the prion propensity of each amino acid in the context of a Q/N-rich PFD [24], and were surprised to find that there was little correlation between the amino

acids that most strongly support prion formation and those that are actually found in yeast PFDs. Here, we provide an explanation for this apparent contradiction.

We first confirmed that non-aromatic hydrophobic residues do strongly promote prion formation. We found that both aromatic residues and non-aromatic hydrophobic residues (with the exception of leucine) all promote prion nucleation, albeit to varying degrees, demonstrating that our previous prion-propensity estimates were not an artifact of the region tested or a product of sampling error. This effect was even stronger than we anticipated, and suggests that prion formation can easily be controlled by modifying the number and position of hydrophobic residues. However, the question remained, if these residues promote prion formation, why are they so rare in actual PFDs?

Combined with our experimental data, our bioinformatics analysis suggests an answer to this question. Strongly prion-promoting residues (F, W, Y, I, V and M) are under-represented among both prion and non-prion Q/N-rich domains (Figure 6.7), most likely because too many of these residues would make proteins excessively aggregation-prone. A variety of evidence indicates that aromatic residues facilitate prion maintenance [33,43]. This requirement for aromatic residues, coupled with a limit on the number of strongly prion-promoting residues tolerated in Q/N-rich domains, likely leads to the exclusion of non-aromatic residues from yeast PFDs. It should be noted that one study suggests that another difference between aromatic and non-aromatic hydrophobics is that aromatic residues, but not non-aromatic hydrophobic residues, can make contacts that facilitate the early oligomerization steps in prion formation [66]. However, in that study, leucine was used as the non-aromatic hydrophobic residue; leucine is uniquely non-prion-prone among the hydrophobic residues [24], presumably due to its low β-sheet propensity [67], so additional studies will be needed to determine whether this result

applies to all hydrophobic residues. Regardless, our results strongly argue that hydrophobic residues are rare not because they inhibit prion formation, but because aromatic residues are equally able to support prion formation, and can also contribute to other steps in prion activity.

Indeed, while it has been well-documented that non-aromatic hydrophobic residues are under-represented in yeast PFDs [16,68], the fact that these residues are almost equally rare among non-prion Q/N-rich domains is often ignored. This highlights a key point: the sequence features that most clearly distinguish Q/N-rich PFDs from the entire proteome may not be the same features that most effectively distinguish between prion and non-prion Q/N-rich domains. This distinction likely explains why algorithms designed to identify new prions based on compositional similarity to existing prions are very effective at identifying prion candidates, yet are far less effective at ranking the top candidates [24]. Unfortunately, this distinction continues to be missed. For example, a recent paper by Espinosa Angarica *et al*. argued that because C, W and E are rare among yeast PFDs, these residues must make an "unfavorable contribution" to prion activity [68]; however, this analysis ignores the fact that while W is rare among yeast PFDs, it is even more rare among non-prion Q/N-rich domains.

One key caveat with these experiments is that different regions of PFDs may have different sequence requirements, based on their respective roles in prion activity. For Sup35, the nucleation domain and remainder of the prion domain (termed the oligopeptide repeat domain, due to the presence of a series of imperfect peptide repeats) have both distinct roles in prion activity [44] and distinct compositional requirements [34]. By focusing on the nucleation domain, we were able to specifically isolate the effects of hydrophobic residues on prion formation. In the nucleation domain, non-aromatic hydrophobic residues and aromatic residues seem at least partially interchangeable. However, we have separately begun to systematically

examine the distinct sequence requirements for prion formation versus maintenance, and have found that aromatic residues appear to play a more essential role within the ORD (unpublished data), consistent with their proposed functions in prion maintenance [33,43].

Our results explain a number of other conundrums in the prion field. There has been substantial debate about the role of short sequence motifs in yeast prion formation. For example, a variety of evidence suggests that a short segment of the Sup35 PFD spanning amino acids 8-24 acts as a key nucleating site for prion formation, and point mutations in this region can prevent addition to prion aggregates and can substantially affect efficiency of cross-species transmission [58,60]. However, much larger fragments are required for prion activity [44]; the shortest region from any prion protein that has be shown to support prion activity is 37 amino acids from Swi1 [69]. Furthermore, the fact that PFDs can be scrambled without blocking prion formation seems to argue against the importance of short sequence motifs. Our data provide a simple explanation for this apparent contradiction. If yeast PFDs contain relatively few strongly prion-promoting amino acids, then wherever these amino acids are located will naturally act as potential nucleating sites. Indeed, residues 8-24 of Sup35 contain two strongly prion promoting amino acids, and no strongly inhibiting amino acids (charged residues or prolines). In fact, the longest segment in the Sup35 PFD without a prion-inhibiting amino acid spans residues 4-27. Thus, the key role of this segment in prion nucleation may be explainable solely based on composition.

Interestingly, while composition is the dominant factor in determining prion activity, our data clearly demonstrate that primary sequence can exert a substantial effect on both the frequency of prion formation and the stability of the prion phenotype. The basis for this effect is unclear. Composition-based algorithms such as PAPA clearly do not predict such a strong effect of primary sequence. However, even algorithms designed to detect primary-sequence motifs

appear to be no more effective (Figure 6.3B). While this set of mutants is likely too small to extract the exact relationship between primary sequence and prion propensity, this data set could provide a useful tool for testing future primary-sequence-based algorithms. However, it is important to note that the observed differences may not be due solely to differences in prion propensity. Because some prion variants can be deleterious or lethal [65], a mutation that shifts the distribution of variants formed by the protein to more toxic variants (or increases the toxicity of common variants) could give the appearance of reducing prion propensity. Therefore, more detailed studies will be needed to untangle the basis for this primary sequence effect.

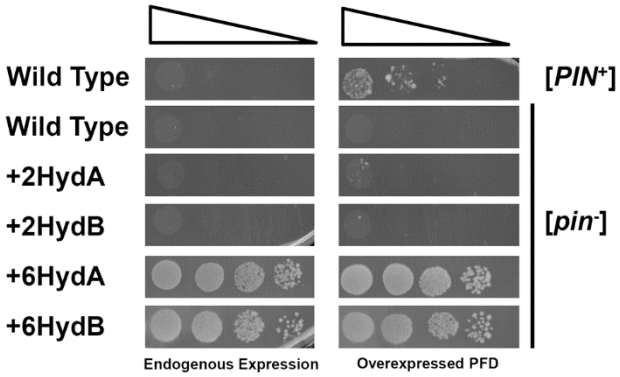


Figure S6.1: Strain YER632/pJ533 was streaked for three consecutive passages on YPAD +4 mM guanidine HCl. This strain was then transformed with plasmids expressing the indicated Sup35 mutants. After FOA selection for loss of pJ533, the strains were transformed with empty vector (*left*) or with a plasmid expressing the matching Sup35 mutant under control of the *GAL1* promoter (*right*). All strains were cultured for three days in galactose/raffinose dropout medium, and then 10-fold serial dilutions were plated onto medium lacking adenine to select for [*PSI*$^+$]. As a control, prion formation by YER632/pJ533 before (wild-type), [*PIN*$^+$]) and after (wild-type, [*pin*$^-$]) guanidine treatment is shown.

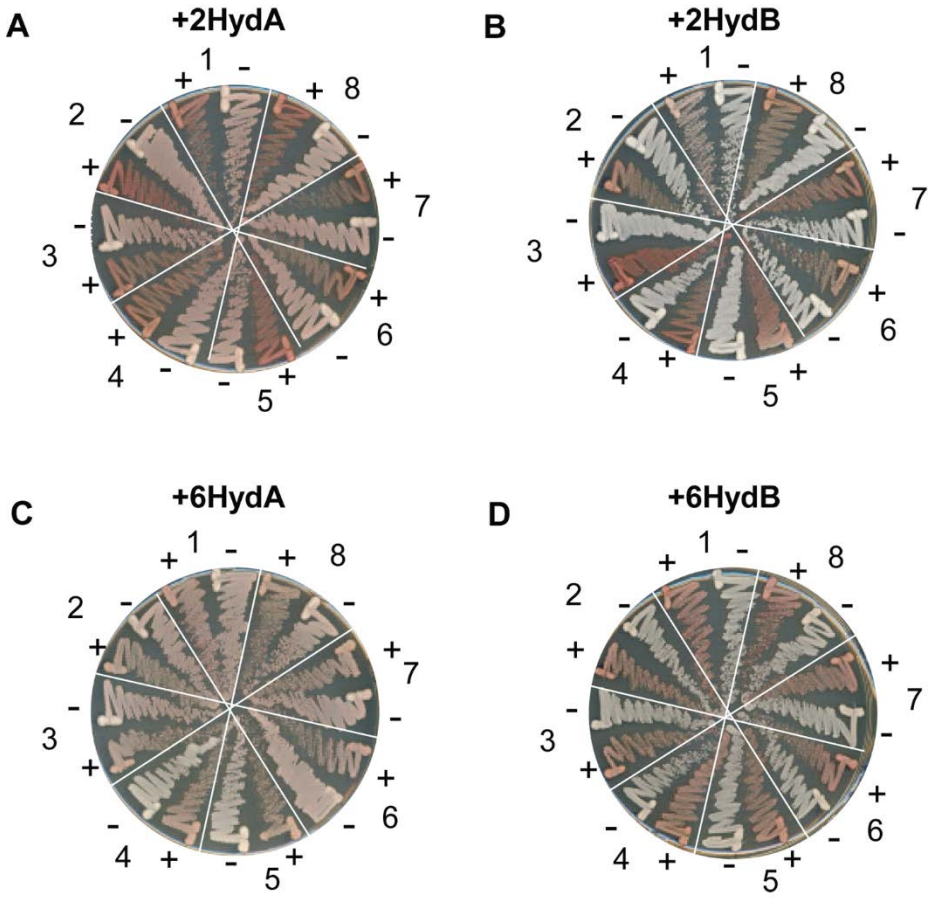


Figure S6.2: Hydrophobic addition constructs form curable prions. For +2HydA **(A)**, +2HydB **(B)**, +6HydA **(C)**, and +6HydB **(D)**, eight individual Ade^+ isolates were grown on YPD (-) and YPD plus 4 mM guanidine HCl (+). Cells were restreaked onto YPD to test for loss of the Ade^+ phenotype.

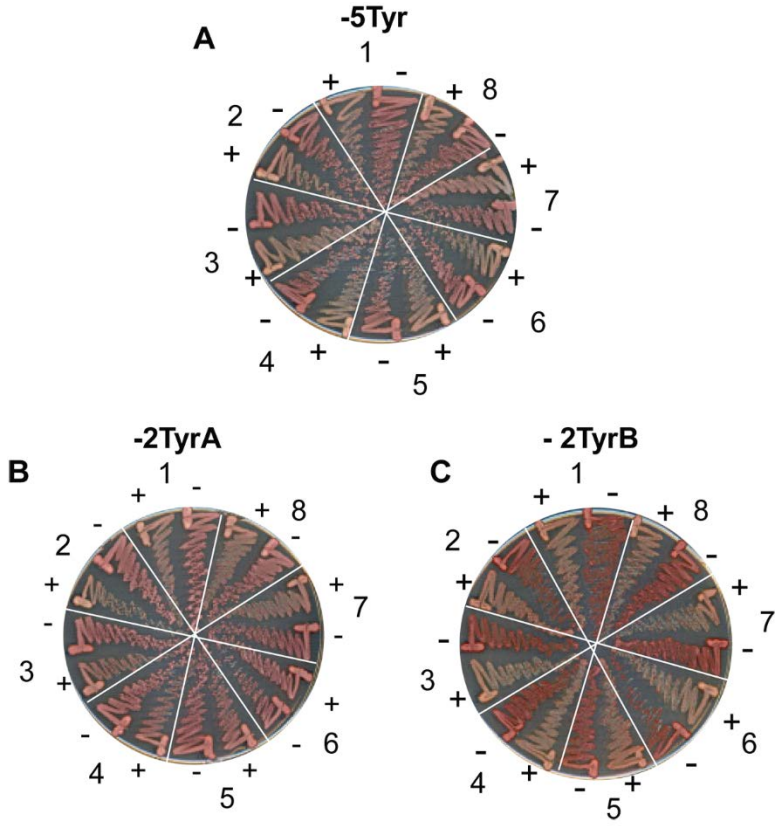


Figure S6.3: Stability and curability of Ade$^+$ colonies formed by tyrosine deletion constructs. For −5Tyr **(A)**, −2TyrA **(B)**, and −2TyrB **(C)**, eight individual Ade$^+$ isolates were grown on YPD (−) and YPD plus 4 mM guanidine HCl (+). Cells were then restreaked onto YPD to test for loss of the Ade$^+$ phenotype.

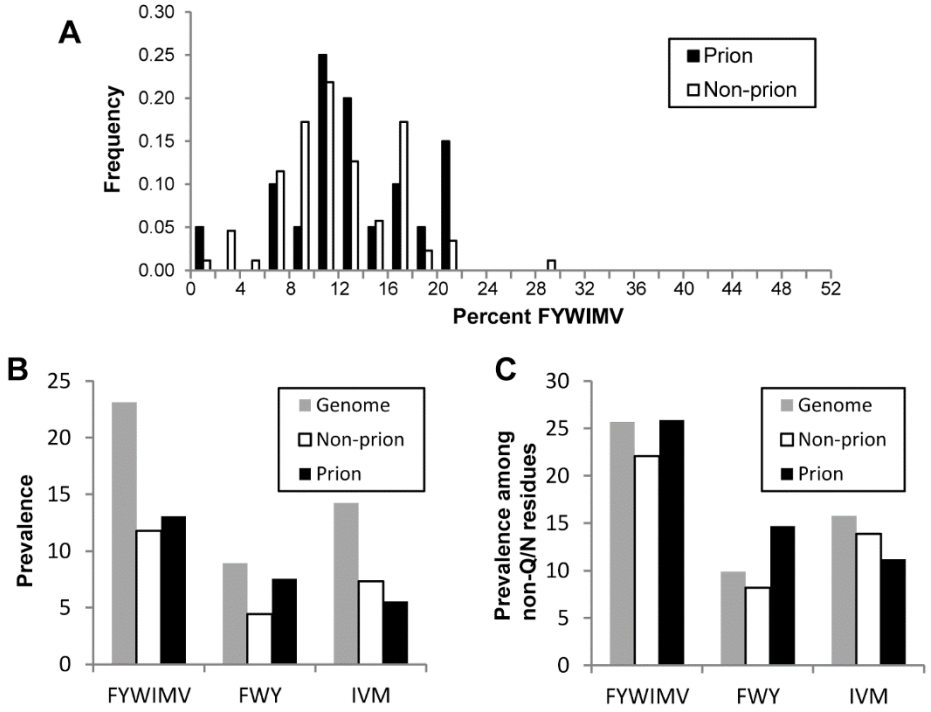


Figure S6.4: Amino acid composition of prion and non-prion Q/N-rich domains, using the data set of Michelitsch and Weissman. (A) Histogram of the prevalence of strongly prion-promoting amino acids (FYWIMV) among regions of the yeast proteome identified by Michelitsch and Weissman as being highly enriched in Q/N-residues. The black bars include Q/N-rich regions from proteins shown to act as prions, as well as from proteins containing domains shown by Alberti et al. to have prion-like activity in four independent assays. Open bars represent all other Q/N-rich regions identified by Michelitsch and Weissman. **(B)** The prevalence of different groups of amino acids in the yeast genome (grey bars) compared to the average frequency of these amino acids among Q/N-rich prion (black bars) and non-prion (open bars). **(C)** The prevalence of different groups of amino acids, plotted as a fraction of non-Q/N residues.

Table S1: Oligonucleotides used in this study.

Name	Description	Sequence
		GTCGGATTCAAACCAAGGCAACAATATCCAGCAAAACTACCAGC AATACAGCCAGAACGGTAACGTTCAACAACAAGGTAACAACAG
EDR1263	Sense primer to build +2HydA	ATACC
EDR1264	Antisense primer to build +2HydA, C-I	GTTGCCTTGGTTTGAATCCGAC
		GGATTCAAACCAAGGCAACAATCAGCAAGTCAACTACCAGCAAT ACAGCCAGATTAACGGTAACCAACAACAAGGTAACAACAGATA
EDR1265	Sense primer to build +2HydB	CC
EDR1266	Antisense primer to build +2HydB	GACTTGCTGATTGTTGCCTTGGTTTGAATCC
		GTCGGATTCAAACCAAGGCAACATCAATCAGCAAATTAACTACC AGGTTCAATACATTAGCGTCCAGGTTAACGGTAACCAACAACAA
EDR1267	Sense primer to build +6HydA	GGTAACAACAG
EDR1268	Antisense primer to build +6HydA	GTTGCCTTGGTTTGAATCCG
		GGCAACATCAATCAGCAAGTCAACGTTTACCAGCAATACAGCCA GAACGGTAACGTTCAAATTATCCAACAAGGTAACAACAGATACC
EDR1269	Sense primer to build +6HydB	AAGG
EDR1270	Antisense primer to build +6HydB	GACTTGCTGATTGATGTTGCCTTGGTTTGAATCCGAC
		GCAACAATCAGCAAAACCAGCAAAGCCAGAACGGTAACCAACA ACAAGGTAACAACAGACAAGGTCAAGCTAATGCTCAAGCCCAAC
EDR1257	Sense primer to build -5Tyr	CTGCAG
EDR1258	Antisense primer to build -5Tyr	GCTTTGCTGGTTTTGCTGATTGTTGCCTTGGTTTGAATCC
		CAAACCAAGGCAACAATCAGCAAAACCAGCAATACAGCCAGAA CGGTAACCAACAACAAGGTAACAACAGACAAGGTTATCAAGCTT
EDR1259	Sense primer to build -2TyrA	ACAATGCTCAAGC
EDR1260	Antisense primer to build -2TyrA	GTTTTGCTGATTGTTGCCTTGGTTTGAATCC
		GCAACAATCAGCAAAACTACCAGCAAAGCCAGAACGGTAACCA ACAACAAGGTAACAACAGATACCAAGGTCAAGCTTACAATGCTC
EDR1261	Sense primer to build -2TyrB	AAGCCC
EDR1262	Antisense primer to build -2TyrB	GGCTTTGCTGGTAGTTTTGCTGATTGTTGC

EDR1409	Sense primer to build +2HydC	GTCGGATTCAAACCAAGGCAACAATGTTCAGCAAAACTACCAGC AATACAGCCAGAACGGTAACATCCAACAACAAGGTAACAACAG ATACC
EDR1403	Sense primer to build +2HydD	GTCGGATTCAAACCAAGGCAACAATCAGCAAATCAACTACCAGC AATACAGCCAGGTTAACGGTAACCAACAACAAGGTAACAACAG ATACC
EDR1404	Sense primer to build +2HydE	GTCGGATTCAAACCAAGGCAACAATCAGCAAATCAACTACCAGC AATACAGCCAGAACGTTGGTAACCAACAACAAGGTAACAACAG ATACC
EDR1405	Sense primer to build +2HydF	GTCGGATTCAAACCAAGGCAACAATCAGCAAATCAACTACCAGC AATACAGCCAGAACGGTGTTAACCAACAACAAGGTAACAACAG ATACC
EDR1406	Sense primer to build +2HydG	GTCGGATTCAAACCAAGGCAACAATCAGCAAATCAACTACCAGC AATACAGCCAGAACGGTAACGTTCAACAACAAGGTAACAACAG ATACC
EDR1407	Sense primer to build +2HydH	GTCGGATTCAAACCAAGGCAACAATCAGATCCAAAACTACCAGC AATACAGCCAGGTTAACGGTAACCAACAACAAGGTAACAACAG ATACC
EDR1408	Sense primer to build +2HydI	GTCGGATTCAAACCAAGGCAACAATATCCAGCAAAACTACCAGC AATACAGCCAGGTTAACGGTAACCAACAACAAGGTAACAACAG ATACC
EDR1308	Sense primer to build Sup35(Y→L)	CAGATTGCAAGGTTTACAAGCTCTGAATGCTCAAGCCCAACCTG CAG
EDR1309	Antisense primer to build Sup35(Y→L)	CAGAGCTTGTAAACCTTGCAATCTGTTGTTACCTTGTTGTTGGTT ACCGTTCTGGCTCAATTGCTGTAAGTTTTGCTGATTGTTGCCTTG GTTTGAATCC
EDR1310	Sense primer to build Sup35(Y→V)	CAGAGTTCAAGGTGTACAAGCTGTGAATGCTCAAGCCCAACCTG CAG
EDR1311	Antisense primer to build Sup35(Y→V)	CACAGCTTGTACACCTTGAACTCTGTTGTTACCTTGTTGTTGGTTA CCGTTCTGGCTAACTTGCTGAACGTTTTGCTGATTGTTGCCTTGGT TTGAATCC
EDR1312	Sense primer to build Sup35(Y→I)	CAGAATCCAGGGTATTCAAGCTATCAATGCTCAAGCCCAACCTG CAG

EDR1313	Antisense primer to build Sup35(Y→I)	GATAGCTTGAATACCCTGGATTCTGTTGTTACCTTGTTGTTGGTTACCGTTCTGGCTTATTTGCTGAATGTTTTGCTGATTGTTGCCTTGGTTTGAATCC
EDR1747	Sense primer to make pER687	GCTTCCAATGATGCATGCATGGACGCAAAGAAGTTTAATAATCATATTACATGG
EDR1748	Antisense primer to make pER687	GGACCTCAAGATGAATTCCCATTCAGGCTGCGCAACTG
EDR1898	Sense primer to make pER760	GAGCTACTGGATCCACAATGTCTGGAGCTCCCTCGAGTGGAGGTAGCTACTCTAAAGGTGAAGAATTATTCACTGGTG
EDR1899	Antisense primer to make pER760	GGATGTCAGTTGTCGACTTTATTTGTACAATTCATCCATACCATGGG
EDR1924	Antisense primer to make NM-GFP fusions	GTCGATGCTACTCGAGTCGTTAACAACTTCGTCATCCACTTC
EDR1084	Common primer to build induction plasmids and NM-GFP fusions	CGATGCTACTCGAGTTTACATATCGTTAACAACTTCGTCATCCAC
EDR1008	Common primer to build induction plasmids	GAGCTACTGGATCCACAATGTCGGATTCAAACCAAGGCAAC
EDR262	Antisense primer to Sup35	GCATCAGCACTGGTAACATTGG
EDR301	Sense primer binding upstream of Sup35	CGTCACAGTGTTCGAGTCTG

REFERENCES

1. Wadsworth JD, Collinge J (2007) Update on human prion disease. Biochim Biophys Acta 1772: 598-609.
2. Coustou V, Deleu C, Saupe S, Begueret J (1997) The protein product of the het-s heterokaryon incompatibility gene of the fungus Podospora anserina behaves as a prion analog. Proc Natl Acad Sci USA 94: 9773-9778.
3. Maclea KS, Ross ED (2011) Strategies for identifying new prions in yeast. Prion 5: 263-268.
4. Li YR, King OD, Shorter J, Gitler AD (2013) Stress granules as crucibles of ALS pathogenesis. J Cell Biol 201: 361-372.
5. Da Cruz S, Cleveland DW (2011) Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. Curr Opin Neurobiol 21: 904-919.
6. Geser F, Martinez-Lage M, Kwong LK, Lee VM, Trojanowski JQ (2009) Amyotrophic lateral sclerosis, frontotemporal dementia and beyond: the TDP-43 diseases. J Neurol 256: 1205-1214.
7. Weihl CC, Temiz P, Miller SE, Watts G, Smith C, *et al*. (2008) TDP-43 accumulation in inclusion body myopathy muscle suggests a common pathogenic mechanism with frontotemporal dementia. J Neurol Neurosurg Psychiatry 79: 1186-1189.
8. Couthouis J, Hart MP, Erion R, King OD, Diaz Z, *et al*. (2012) Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. Hum Mol Genet: Epub ahead of print.
9. Couthouis J, Hart MP, Shorter J, Dejesus-Hernandez M, Erion R, *et al*. (2011) A yeast functional screen predicts new candidate ALS disease genes. Proc Natl Acad Sci U S A 108: 20881-20890.
10. Neumann M, Bentmann E, Dormann D, Jawaid A, DeJesus-Hernandez M, *et al*. (2011) FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. Brain 134: 2595-2609.
11. Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, *et al*. (2013) Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature 495: 467-473.
12. Klar J, Sobol M, Melberg A, Mabert K, Ameur A, *et al*. (2013) Welander distal myopathy caused by an ancient founder mutation in TIA1 associated with perturbed splicing. Hum Mutat 34: 572-577.
13. Bradley ME, Liebman SW (2004) The Sup35 domains required for maintenance of weak, strong or undifferentiated yeast [*PSI*$^{+}$] prions. Mol Microbiol 51: 1649-1659.
14. Ter-Avanesyan MD, Dagkesamanskaya AR, Kushnirov VV, 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: 671-676.
15. Ter-Avanesyan MD, Kushnirov VV, Dagkesamanskaya AR, Didichenko SA, Chernoff YO, *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: 683-692.

16. 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: R40.
17. Liu JJ, Sondheimer N, Lindquist SL (2002) Changes in the middle region of Sup35 profoundly alter the nature of epigenetic inheritance for the yeast prion [*PSI*$^{+}$]. Proc Natl Acad Sci USA 99 Suppl 4: 16446-16453.
18. Ross ED, Baxa U, Wickner RB (2004) Scrambled Prion Domains Form Prions and Amyloid. Mol Cell Biol 24: 7206-7213.
19. Alberti S, Halfmann R, King O, Kapila A, Lindquist S (2009) A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. Cell 137: 146-158.
20. 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: 11910-11915.
21. Sondheimer N, Lindquist S (2000) Rnq1: an epigenetic modifier of protein function in yeast. Mol Cell 5: 163-172.
22. Halfmann R, Wright J, Alberti S, Lindquist S, Rexach M (2012) Prion formation by a yeast GLFG nucleoporin. Prion 6.
23. Ross ED, Toombs JA (2010) The effects of amino acid composition on yeast prion formation and prion domain interactions. Prion 4: 60-65.
24. Toombs JA, McCarty BR, Ross ED (2010) Compositional determinants of prion formation in yeast. Mol Cell Biol 30: 319-332.
25. Ross ED, Maclea KS, Anderson C, Ben-Hur A (2013) A bioinformatics method for identifying Q/N-rich prion-like domains in proteins. Methods Mol Biol 1017: 219-228.
26. Toombs JA, Petri M, Paul KR, Kan GY, Ben-Hur A, *et al*. (2012) De novo design of synthetic prion domains. Proc Natl Acad Sci U S A 109: 6519-6524.
27. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 424: 805-808.
28. Maurer-Stroh S, Debulpaep M, Kuemmerer N, Lopez de la Paz M, Martins IC, *et al*. (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. Nature Methods 7: 237-242.
29. Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F, *et al*. (2008) Prediction of aggregation-prone regions in structured proteins. J Mol Biol 380: 425-436.
30. Goldschmidt L, Teng PK, Riek R, Eisenberg D (2010) Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proc Natl Acad Sci USA 107: 3487-3492.
31. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 22: 1302-1306.
32. Salnikova AB, Kryndushkin DS, Smirnov VN, Kushnirov VV, Ter-Avanesyan MD (2005) Nonsense suppression in yeast cells overproducing Sup35 (eRF3) is caused by its non-heritable amyloids. J Biol Chem 280: 8808-8812.
33. Alexandrov AI, Polyanskaya AB, Serpionov GV, Ter-Avanesyan MD, Kushnirov VV (2012) The effects of amino acid composition of glutamine-rich domains on amyloid formation and fragmentation. PLoS ONE 7: e46458.

34. 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.
35. Sherman F (1991) Getting started with yeast. Methods Enzymol 194: 3-21.
36. Song Y, Wu YX, Jung G, Tutar Y, Eisenberg E, *et al*. (2005) Role for Hsp70 chaperone in *Saccharomyces cerevisiae* prion seed replication. Eukaryot Cell 4: 289-297.
37. Ross ED, Edskes HK, Terry MJ, Wickner RB (2005) Primary sequence independence for prion formation. Proc Natl Acad Sci USA 102: 12825-12830.
38. Gietz RD, Sugino A (1988) New yeast-Escherichia coli shuttle vectors constructed with *in vitro* mutagenized yeast genes lacking six-base pair restriction sites. Gene 74: 527-534.
39. Cormack BP, Bertram G, Egerton M, Gow NA, Falkow S, *et al*. (1997) Yeast-enhanced green fluorescent protein (yEGFP)a reporter of gene expression in Candida albicans. Microbiology 143: 303-311.
40. 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: 929-940.
41. Bagriantsev SN, Kushnirov VV, Liebman SW (2006) Analysis of amyloid aggregates using agarose gel electrophoresis. Methods Enzymol 412: 33-48.
42. Collins SR, Douglass A, Vale RD, Weissman JS (2004) Mechanism of Prion Propagation: Amyloid Growth Occurs by Monomer Addition. PLoS Biology 2: e321.
43. 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: 15185-15192.
44. Osherovich LZ, Cox BS, Tuite MF, Weissman JS (2004) Dissection and design of yeast prions. PLoS Biol 2: E86.
45. 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: 1241-1252.
46. Parham SN, Resende CG, Tuite MF (2001) Oligopeptide repeats in the yeast protein Sup35p stabilize intermolecular prion interactions. EMBO J 20: 2111-2119.
47. Shkundina IS, Kushnirov VV, 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: 827-835.
48. Cox BS (1965) *PSI*, a cytoplasmic suppressor of super-suppressor in yeast. Heredity 26: 211-232.
49. Lancaster AK, Bardill JP, True HL, Masel J (2010) The spontaneous appearance rate of the yeast prion [*PSI*+] and its implications for the evolution of the evolvability properties of the [*PSI*+] system. Genetics 184: 393-400.
50. Wickner RB (1994) [URE3] as an altered URE2 protein: evidence for a prion analog in *Saccharomyces cerevisiae*. Science 264: 566-569.
51. Derkatch IL, Bradley ME, Zhou P, Chernoff YO, Liebman SW (1997) Genetic and Environmental Factors Affecting the de novo Appearance of the [*PSI*(+)] Prion in *Saccharomyces cerevisiae*. Genetics 147: 507-519.
52. Ferreira PC, Ness F, Edwards SR, Cox BS, Tuite MF (2001) The elimination of the yeast [*PSI*+] prion by guanidine hydrochloride is the result of Hsp104 inactivation. Mol Microbiol 40: 1357-1369.

53. Jung G, Masison DC (2001) Guanidine hydrochloride inhibits Hsp104 activity in vivo: a possible explanation for its effect in curing yeast prions. Curr Microbiol 43: 7-10.
54. Ness F, Ferreira P, Cox BS, Tuite MF (2002) Guanidine hydrochloride inhibits the generation of prion "seeds" but not prion protein aggregation in yeast. Mol Cell Biol 22: 5593-5605.
55. Paushkin SV, Kushnirov VV, 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. EMBO J 15: 3127-3134.
56. Wegrzyn RD, Bapat K, Newnam GP, Zink AD, Chernoff YO (2001) Mechanism of prion loss after Hsp104 inactivation in yeast. Mol Cell Biol 21: 4656-4669.
57. LeVine H, 3rd (1999) Quantification of beta-sheet amyloid fibril structures with thioflavin T. Methods Enzymol 309: 274-284.
58. Chen B, Bruce KL, Newnam GP, Gyoneva S, Romanyuk AV, *et al*. (2010) Genetic and epigenetic control of the efficiency and fidelity of cross-species prion transmission. Mol Microbiol 76: 1483-1499.
59. Santoso A, Chien P, Osherovich LZ, Weissman JS (2000) Molecular basis of a yeast prion species barrier. Cell 100: 277-288.
60. Tessier PM, Lindquist S (2007) Prion recognition elements govern nucleation, strain specificity and species barriers. Nature 447: 556-561. Epub 2007 May 2009.
61. Pastor MT, Esteras-Chopo A, Serrano L (2007) Hacking the code of amyloid formation: the amyloid stretch hypothesis. Prion 1: 9-14. Epub 2007 Jan 2005.
62. Lopez de la Paz M, Serrano L (2004) Sequence determinants of amyloid fibril formation. Proc Natl Acad Sci U S A 101: 87-92. Epub 2003 Dec 2022.
63. Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D, *et al*. (2006) The 3D profile method for identifying fibril-forming segments of proteins. Proc Natl Acad Sci U S A 103: 4074-4078.
64. Teng PK, Eisenberg D (2009) Short protein segments can drive a non-fibrillizing protein into the amyloid state. Protein Eng Des Sel 22: 531-536.
65. McGlinchey RP, Kryndushkin D, Wickner RB (2011) Suicidal [*PSI*+] is a lethal yeast prion. Proc Natl Acad Sci U S A 108: 5337-5341.
66. 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 6: 225-230.
67. Street AG, Mayo SL (1999) Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. Proc Natl Acad Sci U S A 96: 9074-9076.
68. Espinosa Angarica V, Ventura S, Sancho J (2013) Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. BMC Genomics 14: 316.
69. Crow ET, Du Z, Li L (2011) A small, glutamine-free domain propagates the [SWI(+)] prion in budding yeast. Mol Cell Biol 31: 3436-3444.

APPENDIX II: INVESTIGATING THE POTENTIAL ROLE OF PRIMARY SEQUENCE FEATURES IN THE DEGRADATION OF PRION-LIKE DOMAINS[6]

INTRODUCTION

Our observations suggest that amino acid composition of the mutagenized regions in the hnRNP PrLDs is the predominant determinant of degradation or prion formation (see Chapter 3). However, I also sought to investigate whether primary sequence features could be influencing the degradation susceptibility of the PrLDs, which would not be predicted on the basis of composition alone.

MATERIALS AND METHODS

Amino Acid Position Preference Analysis

The number of observed occurrences for each amino acid at each position were calculated from each library using in-house Python scripts. To estimate the expected number of occurrences at each position within the 8-amino acid randomly mutagenized window, the total number of occurrences for a given amino acid within a sequence library was divided by the number of possible positions within each individual sequence (8 positions in all cases). Additionally, 20 scrambled versions of each library (i.e. the A1, A2, and combined degradation and *ade*$^{-}$ libraries) were generated *in silico* (resulting in a total of 120 scrambled sequence libraries), where each 8-amino acid sequence was scrambled individually.

[6] Original library mutagenesis, assay development, data collection, and computational analyses were performed by myself. Additional data collection to expand the sequence libraries was performed by Lindsey Brookbank and Mikaela Elder (Figure 7.1 and Table 7.1).

Dipeptide Occurrence Analysis

Library sequences were evaluated using in-house Python scripts. 400 unique dipeptide arrangements can be generated using the 20 canonical amino acids. For all possible dipeptide arrangements, the number of observed occurrences for each dipeptide was calculated from the sequence library. The number of expected dipeptide occurrences (D_{exp}), assuming a sequence library of the same size as the one being evaluated, was calculated as follows:

$$\mathrm{D_{exp}} = \frac{\mathrm{n_{aa1}}}{\mathrm{n_{total}}} \times \frac{\mathrm{n_{aa2}}}{\mathrm{n_{total}}} \times \mathrm{D_{total}} \quad (4.1)$$

where n_{aa1} represents the number of times the first amino acid in the dipeptide occurred in the sequence library, n_{aa2} represents the number of times the second amino acid in the dipeptide occurred in the sequence library, n_{total} represents the total number of amino acids in the sequence library, and D_{total} represents the total number of dipeptides in the sequence library. For each sequence library, the number of expected occurrences was plotted against the number of observed occurrences for the corresponding dipeptide arrangement using Microsoft Excel. Additionally, for each library, each 8-amino acid sequence was scrambled individually *in silico*, and the scrambled libraries were re-analyzed as detailed above.

RESULTS

Investigation of Amino Acid Position Preferences as a Potential Contributing Factor in the Degradation or Stability of the hnRNP PrLDs

The sequence libraries generated by our random mutagenesis methodology are highly heterogeneous with respect to primary sequence, suggesting that the amino acid biases observed among degradation-promoting or degradation-inhibiting sequences are the result of compositional effects, rather than specific primary sequence arrangements. Nevertheless, we

examined the possibility that susceptibility or resistance could, in some cases, result from primary sequence features within the *ADE*$^{+}$ or *ade*$^{-}$ libraries.

In order to examine whether any amino acids exhibited position preferences within the 8-amino acid randomly mutagenized region, we first expanded our original degradation and *ade*$^{-}$ libraries for both A1 and A2. Expansion of the libraries is necessary to increase the sample sizes for each amino acid at each of the 8 positions within the randomly mutagenized window. Importantly, there was a strong correlation between the degradation scores derived from the original degradation libraries and the expanded degradation libraries (Figure 7.1 and Table 7.1), indicating that our phenotype screening methodology yields consistent and reproducible results. The number of observed occurrences for each amino acid at each position were calculated for the A1, A2, and combined degradation libraries and *ade*$^{-}$ libraries.

In principle, a very large sequence library consisting of the 20 canonical amino acids with no position preferences would result in an equal distribution of each amino acid among the 8 positions. In order to estimate the number of expected occurrences for each amino acid at each position (assuming theoretically equal distributions), the total number of occurrences for each amino acid within the given library was divided by the number of positions in the randomly mutagenized region. The ratio of the number of observed occurrences to the number of expected occurrences (O/E ratio) serves as a useful comparison to examine enrichment or depletion of a given amino acid at each position: as the distributions of each amino acid among the positions approaches a random distribution, this ratio would approach a value of 1.

In general, no strong position preferences were observed in the degradation and *ade*$^{-}$ libraries for both A2 and A1 (Figure 7.2A-D). The large majority of O/E values fall between 0.5 and 2.0, suggesting that the amino acids are roughly equally distributed among the 8 positions in

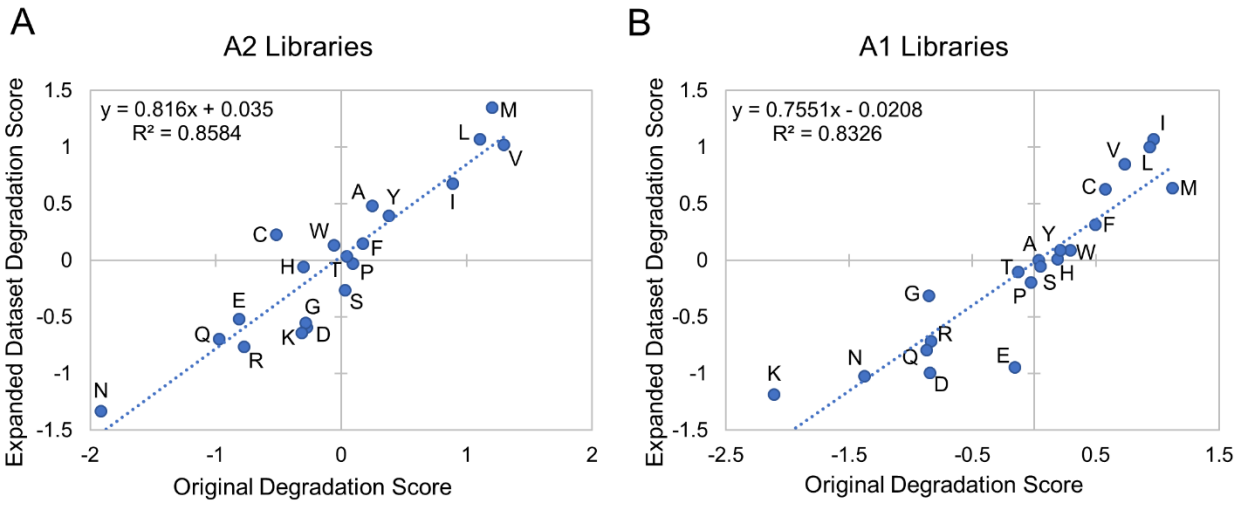
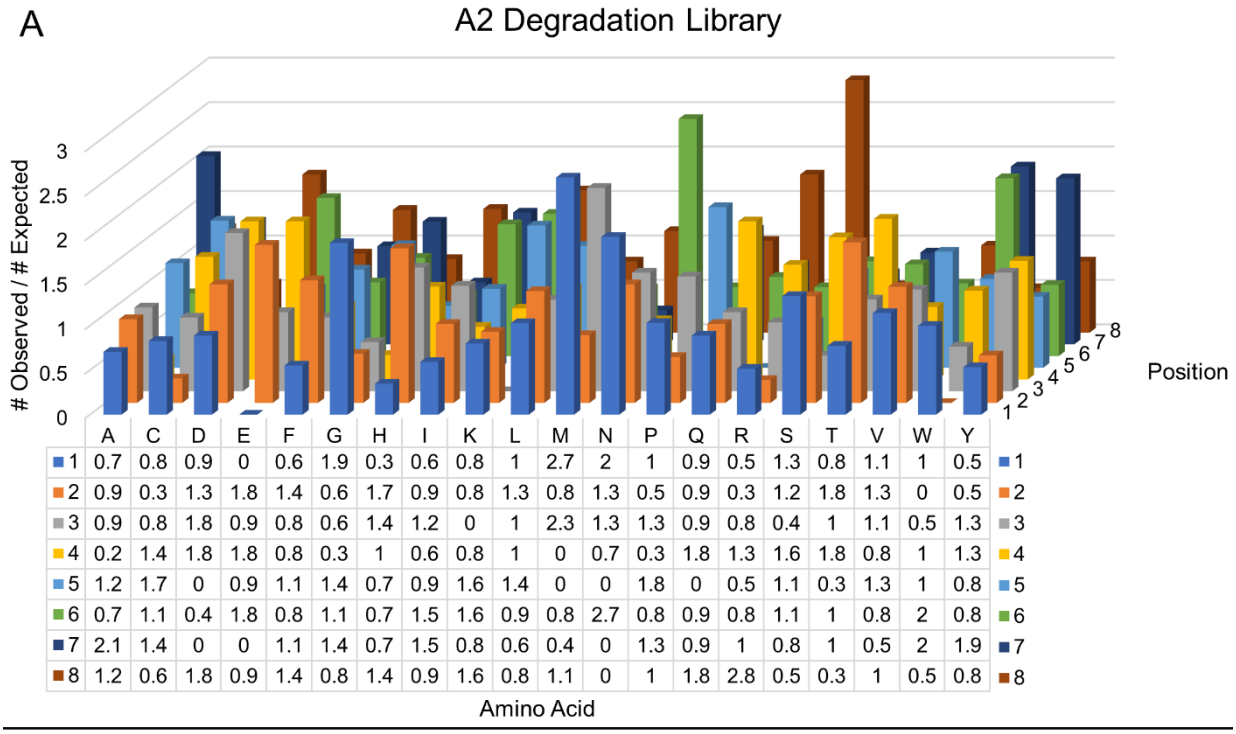


Figure 7.1: Expansion of the degradation and naïve datasets does not dramatically affect amino acid degradation scores. Degradation scores derived from the expanded dataset were plotted against the original degradation scores for A2 **(A)** and A1 **(B)**. In both cases, a strong correlation is observed.

Table 7.1: Comparison of the degradation scores derived from the original and expanded datasets

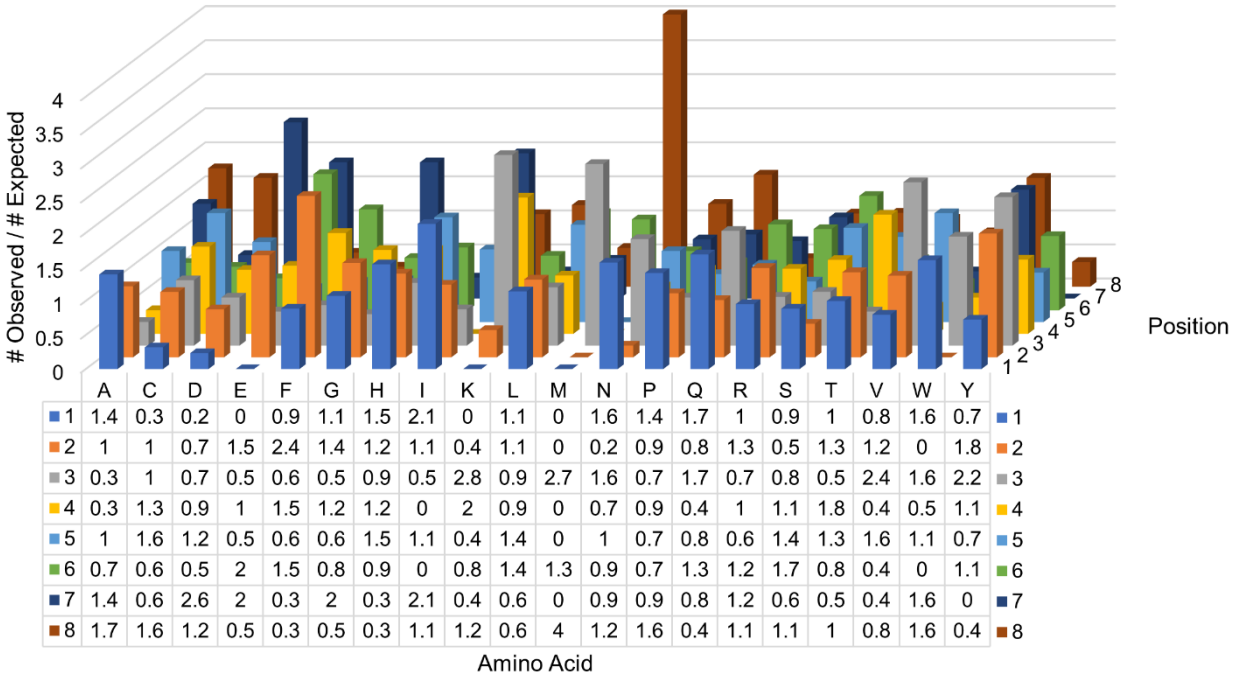
	Degradation Score			
	A2		A1	
Amino Acid	Original	Expanded	Original	Expanded
Valine	1.30	1.02	0.74	0.85
Methionine	1.21	1.35	1.12	0.64
Leucine	1.11	1.07	0.94	1.00
Isoleucine	0.89	0.68	0.97	1.07
Tyrosine	0.38	0.40	0.21	0.089
Alanine	0.25	0.48	0.039	0.0022
Phenylalanine	0.18	0.15	0.50	0.31
Proline	0.10	-0.026	-0.023	-0.20
Threonine	0.046	0.038	-0.13	-0.10
Serine	0.031	-0.26	0.052	-0.052
Tryptophan	-0.057	0.14	0.30	0.092
Aspartic Acid	-0.27	-0.59	-0.85	-0.99
Glycine	-0.28	-0.55	-0.85	-0.31
Histidine	-0.30	-0.057	0.19	0.0099
Lysine	-0.31	-0.64	-2.11	-1.18
Cysteine	-0.52	0.23	0.58	0.63
Arginine	-0.77	-0.76	-0.84	-0.71
Glutamic Acid	-0.81	-0.52	-0.16	-0.94
Glutamine	-0.97	-0.69	-0.87	-0.79
Asparagine	-1.91	-1.33	-1.38	-1.02

A



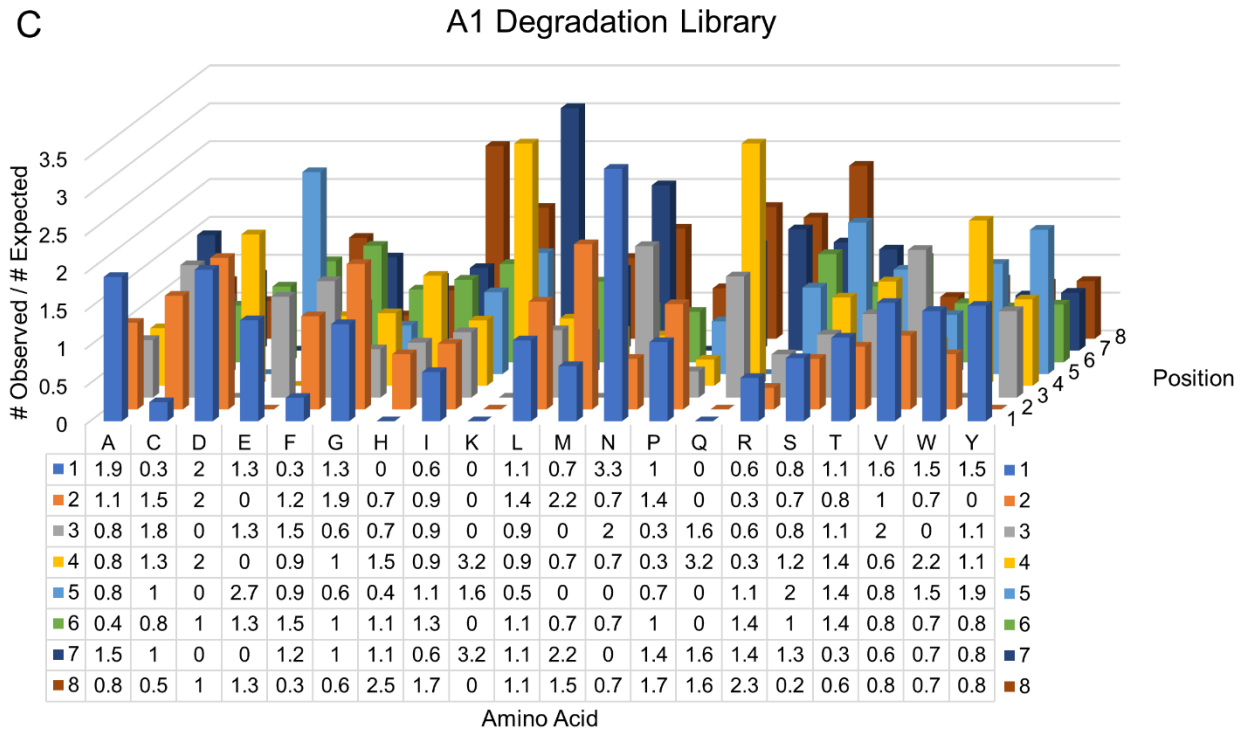
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
1	0.7	0.8	0.9	0	0.6	1.9	0.3	0.6	0.8	1	2.7	2	1	0.9	0.5	1.3	0.8	1.1	1	0.5	1
2	0.9	0.3	1.3	1.8	1.4	0.6	1.7	0.9	0.8	1.3	0.8	1.3	0.5	0.9	0.3	1.2	1.8	1.3	0	0.5	2
3	0.9	0.8	1.8	0.9	0.8	0.6	1.4	1.2	0	1	2.3	1.3	1.3	0.9	0.8	0.4	1	1.1	0.5	1.3	3
4	0.2	1.4	1.8	1.8	0.8	0.3	1	0.6	0.8	1	0	0.7	0.3	1.8	1.3	1.6	1.8	0.8	1	1.3	4
5	1.2	1.7	0	0.9	1.1	1.4	0.7	0.9	1.6	1.4	0	0	1.8	0	0.5	1.1	0.3	1.3	1	0.8	5
6	0.7	1.1	0.4	1.8	0.8	1.1	0.7	1.5	1.6	0.9	0.8	2.7	0.8	0.9	0.8	1.1	1	0.8	2	0.8	6
7	2.1	1.4	0	0	1.1	1.4	0.7	1.5	0.8	0.6	0.4	0	1.3	0.9	1	0.8	1	0.5	2	1.9	7
8	1.2	0.6	1.8	0.9	1.4	0.8	1.4	0.9	1.6	0.8	1.1	0	1	1.8	2.8	0.5	0.3	1	0.5	0.8	8

B



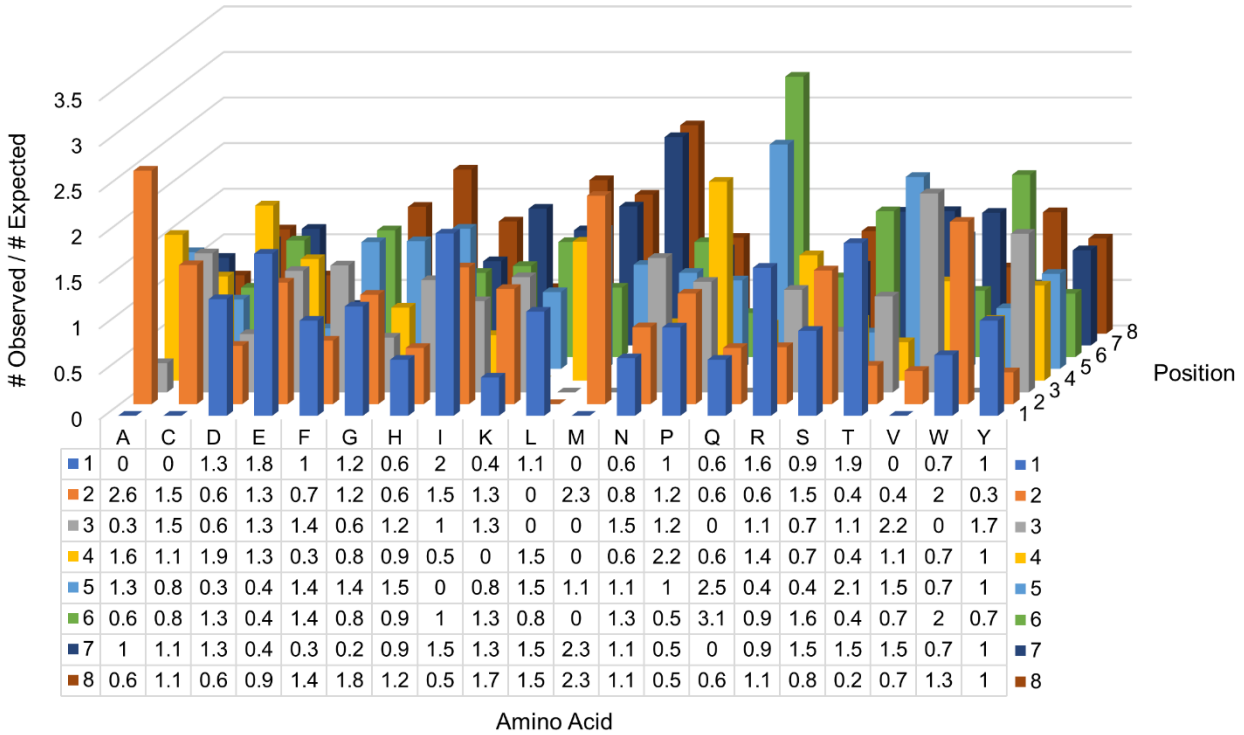
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
1	1.4	0.3	0.2	0	0.9	1.1	1.5	2.1	0	1.1	0	1.6	1.4	1.7	1	0.9	1	0.8	1.6	0.7	1
2	1	1	0.7	1.5	2.4	1.4	1.2	1.1	0.4	1.1	0	0.2	0.9	0.8	1.3	0.5	1.3	1.2	0	1.8	2
3	0.3	1	0.7	0.5	0.6	0.5	0.9	0.5	2.8	0.9	2.7	1.6	0.7	1.7	0.7	0.8	0.5	2.4	1.6	2.2	3
4	0.3	1.3	0.9	1	1.5	1.2	1.2	0	2	0.9	0	0.7	0.9	0.4	1	1.1	1.8	0.4	0.5	1.1	4
5	1	1.6	1.2	0.5	0.6	0.6	1.5	1.1	0.4	1.4	0	1	0.7	0.8	0.6	1.4	1.3	1.6	1.1	0.7	5
6	0.7	0.6	0.5	2	1.5	0.8	0.9	0	0.8	1.4	1.3	0.9	0.7	1.3	1.2	1.7	0.8	0.4	0	1.1	6
7	1.4	0.6	2.6	2	0.3	2	0.3	2.1	0.4	0.6	0	0.9	0.9	0.8	1.2	0.6	0.5	0.4	1.6	0	7
8	1.7	1.6	1.2	0.5	0.3	0.5	0.3	1.1	1.2	0.6	4	1.2	1.6	0.4	1.1	1.1	1	0.8	1.6	0.4	8

C

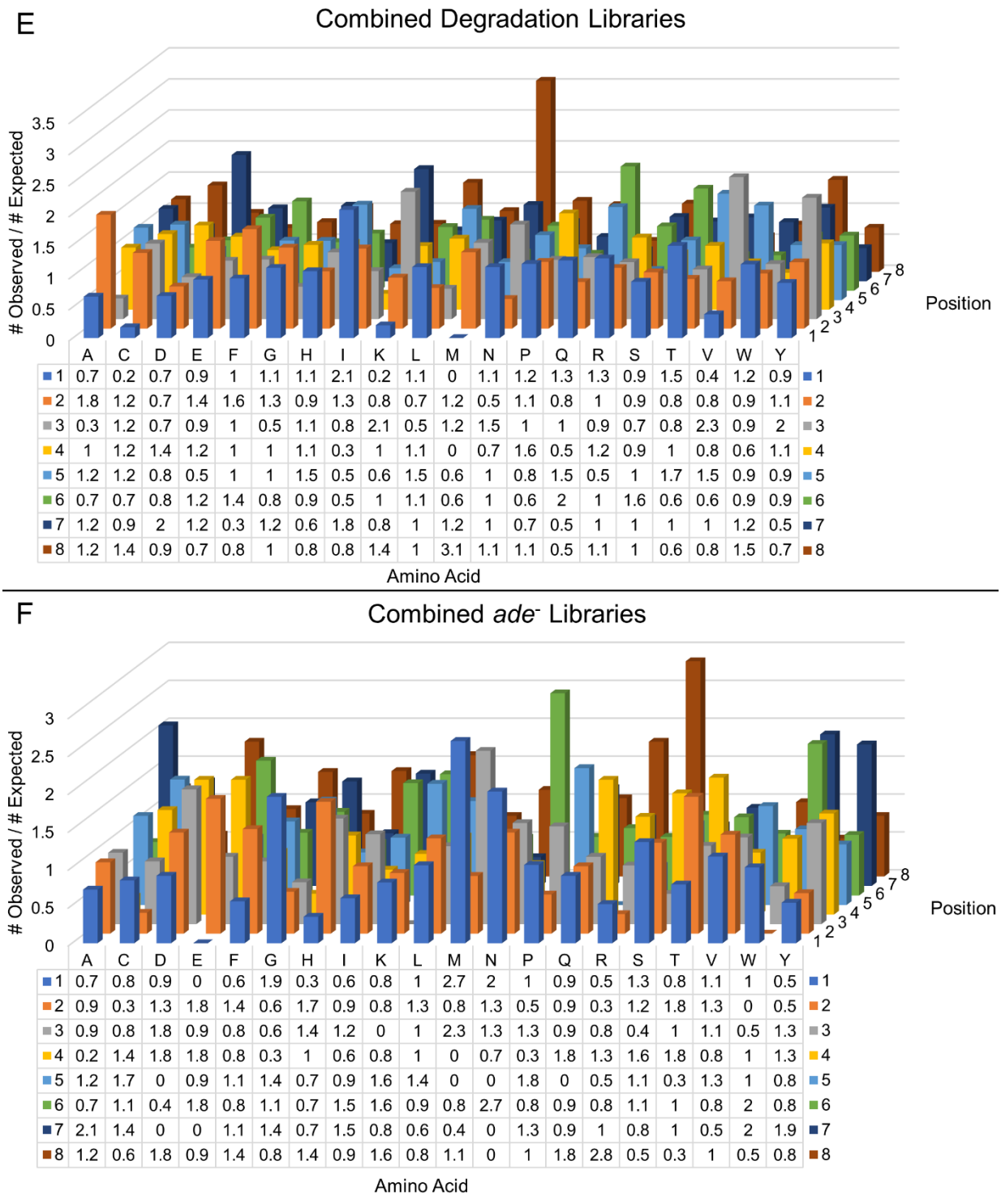


	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
1	1.9	0.3	2	1.3	0.3	1.3	0	0.6	0	1.1	0.7	3.3	1	0	0.6	0.8	1.1	1.6	1.5	1.5	1
2	1.1	1.5	2	0	1.2	1.9	0.7	0.9	0	1.4	2.2	0.7	1.4	0	0.3	0.7	0.8	1	0.7	0	2
3	0.8	1.8	0	1.3	1.5	0.6	0.7	0.9	0	0.9	0	2	0.3	1.6	0.6	0.8	1.1	2	0	1.1	3
4	0.8	1.3	2	0	0.9	1	1.5	0.9	3.2	0.9	0.7	0.7	0.3	3.2	0.3	1.2	1.4	0.6	2.2	1.1	4
5	0.8	1	0	2.7	0.9	0.6	0.4	1.1	1.6	0.5	0	0	0.7	0	1.1	2	1.4	0.8	1.5	1.9	5
6	0.4	0.8	1	1.3	1.5	1	1.1	1.3	0	1.1	0.7	0.7	1	0	1.4	1	1.4	0.8	0.7	0.8	6
7	1.5	1	0	0	1.2	1	1.1	0.6	3.2	1.1	2.2	0	1.4	1.6	1.4	1.3	0.3	0.6	0.7	0.8	7
8	0.8	0.5	1	1.3	0.3	0.6	2.5	1.7	0	1.1	1.5	0.7	1.7	1.6	2.3	0.2	0.6	0.8	0.7	0.8	8

D



	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
1	0	0	1.3	1.8	1	1.2	0.6	2	0.4	1.1	0	0.6	1	0.6	1.6	0.9	1.9	0	0.7	1	1
2	2.6	1.5	0.6	1.3	0.7	1.2	0.6	1.5	1.3	0	2.3	0.8	1.2	0.6	0.6	1.5	0.4	0.4	2	0.3	2
3	0.3	1.5	0.6	1.3	1.4	0.6	1.2	1	1.3	0	0	1.5	1.2	0	1.1	0.7	1.1	2.2	0	1.7	3
4	1.6	1.1	1.9	1.3	0.3	0.8	0.9	0.5	0	1.5	0	0.6	2.2	0.6	1.4	0.7	0.4	1.1	0.7	1	4
5	1.3	0.8	0.3	0.4	1.4	1.4	1.5	0	0.8	1.5	1.1	1.1	1	2.5	0.4	0.4	2.1	1.5	0.7	1	5
6	0.6	0.8	1.3	0.4	1.4	0.8	0.9	1	1.3	0.8	0	1.3	0.5	3.1	0.9	1.6	0.4	0.7	2	0.7	6
7	1	1.1	1.3	0.4	0.3	0.2	0.9	1.5	1.3	1.5	2.3	1.1	0.5	0	0.9	1.5	1.5	1.5	0.7	1	7
8	0.6	1.1	0.6	0.9	1.4	1.8	1.2	0.5	1.7	1.5	2.3	1.1	0.5	0.6	1.1	0.8	0.2	0.7	1.3	1	8



E	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
1	0.7	0.2	0.7	0.9	1	1.1	1.1	2.1	0.2	1.1	0	1.1	1.2	1.3	1.3	0.9	1.5	0.4	1.2	0.9	1
2	1.8	1.2	0.7	1.4	1.6	1.3	0.9	1.3	0.8	0.7	1.2	0.5	1.1	0.8	1	0.9	0.8	0.8	0.9	1.1	2
3	0.3	1.2	0.7	0.9	1	0.5	1.1	0.8	2.1	0.5	1.2	1.5	1	1	0.9	0.7	0.8	2.3	0.9	2	3
4	1	1.2	1.4	1.2	1	1	1.1	0.3	1	1.1	0	0.7	1.6	0.5	1.2	0.9	1	0.8	0.6	1.1	4
5	1.2	1.2	0.8	0.5	1	1	1.5	0.5	0.6	1.5	0.6	1	0.8	1.5	0.5	1	1.7	1.5	0.9	0.9	5
6	0.7	0.7	0.8	1.2	1.4	0.8	0.9	0.5	1	1.1	0.6	1	0.6	2	1	1.6	0.6	0.6	0.9	0.9	6
7	1.2	0.9	2	1.2	0.3	1.2	0.6	1.8	0.8	1	1.2	1	0.7	0.5	1	1	1	1	1.2	0.5	7
8	1.2	1.4	0.9	0.7	0.8	1	0.8	0.8	1.4	1	3.1	1.1	1.1	0.5	1.1	1	0.6	0.8	1.5	0.7	8

F	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
1	0.7	0.8	0.9	0	0.6	1.9	0.3	0.6	0.8	1	2.7	2	1	0.9	0.5	1.3	0.8	1.1	1	0.5	1
2	0.9	0.3	1.3	1.8	1.4	0.6	1.7	0.9	0.8	1.3	0.8	1.3	0.5	0.9	0.3	1.2	1.8	1.3	0	0.5	2
3	0.9	0.8	1.8	0.9	0.8	0.6	1.4	1.2	0	1	2.3	1.3	1.3	0.9	0.8	0.4	1	1.1	0.5	1.3	3
4	0.2	1.4	1.8	1.8	0.8	0.3	1	0.6	0.8	1	0	0.7	0.3	1.8	1.3	1.6	1.8	0.8	1	1.3	4
5	1.2	1.7	0	0.9	1.1	1.4	0.7	0.9	1.6	1.4	0	0	1.8	0	0.5	1.1	0.3	1.3	1	0.8	5
6	0.7	1.1	0.4	1.8	0.8	1.1	0.7	1.5	1.6	0.9	0.8	2.7	0.8	0.9	0.8	1.1	1	0.8	2	0.8	6
7	2.1	1.4	0	0	1.1	1.4	0.7	1.5	0.8	0.6	0.4	0	1.3	0.9	1	0.8	1	0.5	2	1.9	7
8	1.2	0.6	1.8	0.9	1.4	0.8	1.4	0.9	1.6	0.8	1.1	0	1	1.8	2.8	0.5	0.3	1	0.5	0.8	8

Figure 7.2: Susceptibility or resistance to degradation is not position-dependent. The ratio of observed to expected occurrences is plotted for each amino acid at each position within the random mutagenesis window for A2 **(A, B)**, A1 **(C, D)**, and the combined libraries **(E, F)**. The majority of values range from 0.5 to 2.0, suggesting that the degradation-promoting or degradation-inhibiting effects of the amino acids is generally not position-dependent.

the degradation libraries and *ade*$^-$ libraries. This is particularly true for amino acids that are well-represented in the original sequence libraries (Chapter 3, Table 3.1). Many of the amino acids for which apparent position preferences (i.e. relatively large or small O/E values) are observed tended to have fewer total observations in the original sequence libraries, due in part to the limited number of codons encoding these amino acids (e.g. E, K, M, N, Q, and W; Chapter 3, Table 3.1). Therefore, the sample sizes for these amino acids may be insufficient to accurately evaluate position preferences.

Similar compositional biases were observed the degradation libraries for both hnRNP PrLDs, suggesting that the features controlling degradation of the PrLDs largely overlap (Chapter 3, Figure 3.2). Therefore, combining the degradation libraries and combining the *ade*$^-$ libraries from the A1 and A2 random mutagenesis experiments may be a way to increase the effective amino acid sample sizes. As with the individual libraries, the large majority of O/E values for the combined degradation library and combined *ade*$^-$ library fall between 0.5 and 2.0 (Figure 7.2E, F).

In order to more rigorously explore whether the range of occurrences observed in our sequence libraries falls within the range of occurrences that can be expected by random chance in libraries of equivalent size, each sequence library was scrambled *in silico* (see Materials and Methods). From each scrambled library, the minimum and maximum number of occurrences observed for each amino acid across the 8 positions was calculated. For each library, the scrambling was repeated 19 additional times, along with calculation of the minimum and maximum number of occurrences observed for each amino acid. After completing the iterative scrambling, the smallest minimum occurrence value and largest maximum occurrence value

were calculated for each set of 20 scrambled sequence libraries, which together approximate likely lower and upper occurrence bounds assuming no amino acid position preference.

The vast majority of minimum and maximum occurrence values observed in the original sequence libraries fall within the range of occurrence values observed in the scrambled library sets (Table 7.2). Of the 640 total occurrence observations in the A1 and A2 degradation and *ade*$^{-}$ libraries (20 amino acids x 8 positions x 4 individual libraries), only five occurrence values fall outside the lower and upper occurrence bounds observed in the respective scrambled sequence library sets. For the combined degradation library and combined *ade*$^{-}$ library, only 6 of the 320 total occurrence observations fall outside of the respective scramble set occurrence bounds. Furthermore, 5 of the 6 outliers differ by no more than 2 occurrences from the scrambled set occurrence bounds. Interestingly, the single outlier (out of the 960 total occurrence observations considering all libraries collectively) that differs by more than 2 occurrences is a preference for arginine at position 8 in both the A1 and A2 degradation libraries. While sequences enriched in hydrophobic residues can lead to San1-mediated degradation in yeast (although it does not appear to be involved in the degradation of the hnRNP PrLDs; see Chapter 4, Table 4.2), a single arginine substitution adjacent to hydrophobic San1 degrons results in partial San1-independent degradation [1]. It is possible that arginine residues at specific locations or near strong degron determinants promote degradation by this currently uncharacterized degradation pathway. However, collectively these results suggest that susceptibility or resistance to degradation is driven by amino acid composition in a largely position-independent manner.

Table 7.2: Comparison of the minimum and maximum occurrences observed in the original and scrambled sets for each sequence library. For the original sequence libraries, the minimum and maximum values are derived from the position with the smallest and largest number of occurrences respectively. For the scrambled library sets, the minimum and maximum values represent smallest minimum value and largest maximum value observed within a scrambled library set. Highlighted boxes indicate occurrence values observed in the original libraries that fall outside of the lower or upper occurrence bounds derived from the scrambled sequence sets.

	A2 Degradation Library				A2 *ade*$^-$ Library				A1 Degradation Library				A1 *ade*$^-$ Library				Combined Degradation Libraries				Combined *ade*$^-$ Libraries			
	Original		Scrambled		Original		Scrambled		Original		Scrambled		Original		Scrambled		Original		Scrambled		Original		Scrambled	
Amino Acid	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
A	1	9	0	10	1	5	0	7	1	5	0	7	0	8	0	7	3	13	1	13	2	11	1	12
C	1	6	0	10	1	5	0	9	1	7	1	9	0	4	0	8	4	10	1	15	1	8	0	12
D	0	4	0	7	1	11	0	10	0	2	0	4	1	6	0	8	0	6	0	8	5	15	2	13
E	0	2	0	4	0	4	0	6	0	2	0	4	1	4	0	6	0	3	0	5	2	6	0	9
F	2	5	0	9	1	8	0	9	1	5	0	8	1	4	0	7	3	9	1	13	2	10	1	13
G	1	7	0	9	3	13	2	12	2	6	0	7	1	9	0	11	4	11	2	13	6	15	3	22
H	1	5	0	7	1	5	0	9	0	7	0	7	2	5	1	7	1	11	0	12	4	10	2	13
I	2	5	0	8	0	4	0	7	3	8	0	9	0	4	0	5	5	11	1	15	1	8	0	10
K	0	2	0	4	0	7	0	7	0	2	0	2	0	4	0	6	0	3	0	5	1	10	0	12
L	5	12	1	18	2	5	0	10	3	8	1	12	0	4	0	6	11	19	6	23	3	9	1	14
M	0	7	0	8	0	3	0	3	0	3	0	5	0	2	0	4	0	8	1	9	0	5	0	5

N	0	4	0	6	1	9	0	12	0	5	0	5	3	7	0	11	0	8	0	10	5	16	3	19
P	1	7	0	8	3	7	1	11	1	5	0	8	2	9	0	11	2	9	2	15	5	13	2	15
Q	0	2	0	5	1	4	0	7	0	2	0	3	0	5	0	5	0	4	0	6	2	8	0	8
R	1	11	0	8	5	11	1	16	1	8	0	8	3	13	2	18	2	19	2	13	8	21	9	27
S	3	12	1	14	5	17	2	19	1	12	0	11	3	12	2	13	5	20	6	22	13	29	9	27
T	1	7	0	8	2	7	0	9	1	5	0	8	1	10	0	11	3	12	2	15	5	15	2	17
V	3	8	1	13	1	6	0	7	3	10	0	12	0	6	0	6	6	17	4	20	2	12	0	12
W	0	4	0	6	0	3	0	6	0	3	0	5	0	3	0	5	1	5	0	10	2	5	0	8
Y	2	7	0	8	0	6	0	8	0	5	0	7	1	5	0	6	2	9	1	14	3	11	1	13

Evaluation of the Potential Role of Dipeptides as Simple Primary Sequence Elements Affecting PrLD Degradation

One of the simplest possible primary sequence elements is the dipeptide. With the 20 canonical amino acids, 400 unique dipeptide arrangements are theoretically possible. In order to examine whether any dipeptides were enriched in the degradation libraries or naïve libraries, the number of observed occurrences for each possible dipeptide was calculated from the degradation library or the *ade*$^{-}$ library for each hnRNP PrLD individually, or with the PrLD libraries combined. For each dipeptide, the expected number of occurrences in a library of equivalent size was estimated as the product of the individual amino acid frequencies within a given library times the number of dipeptides in the library (equation 7.1).

In principle, if dipeptide elements do not affect PrLD degradation, a large library of sequences would be expected to approach a perfect correlation between the observed and expected number of occurrences for each dipeptide. We find a strong correlation between the observed and expected dipeptide occurrences in the degradation and *ade*$^{-}$ libraries for each of the individual PrLDs (Figure 7.3A-D, and Table 7.3). As mentioned previously, the features controlling degradation of the PrLDs overlap substantially (Chapter 3, Figure 3.2). Therefore, if primary sequence features contribute to the PrLDs' susceptibility or resistance to degradation, combining the libraries could result in stronger dipeptide biases. However, combining the libraries also resulted in a strong correlation between observed and expected dipeptide occurrences (Figure 7.3E, F).

Additionally, if primary sequence elements do exert an effect on the degradation of the PrLDs, scrambling the original sequence datasets should disrupt these features and improve the correlation between observed and expected dipeptide occurrences. For the A1, A2, and combined

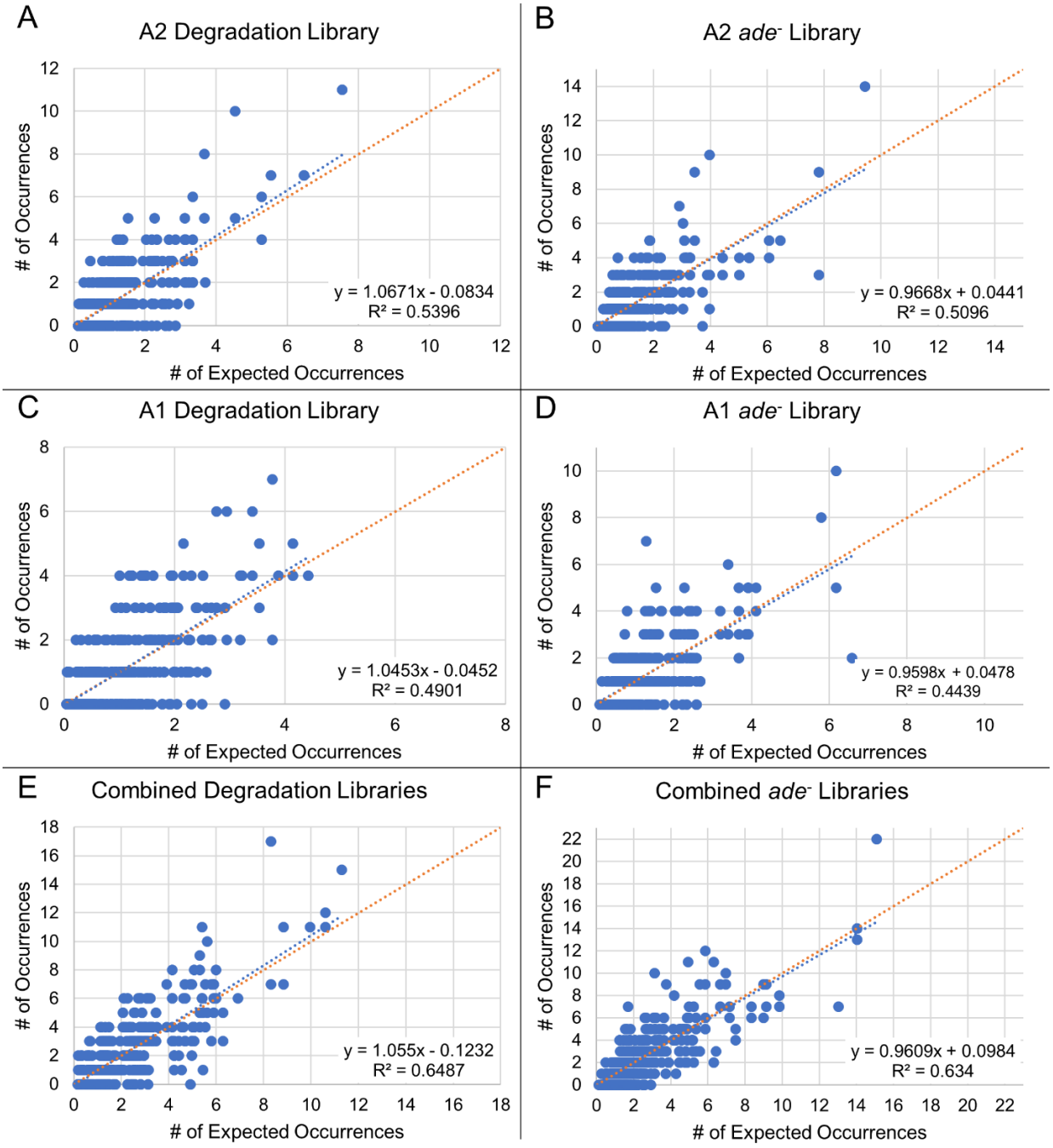


Figure 7.3: Evaluation of dipeptides as potential determinants of degradation susceptibility. For each dipeptide, the observed number of occurrences was plotted against the expected number of occurrences for the A2 **(A, B)** and A1 **(C, D)** individual libraries as well as combined **(E, F)** libraries. In each case, the degradation libraries (*left*) and *ade*- libraries (*right*) were evaluated. In all graphs, the blue line represents the best linear fit, whereas the orange line indicates a theoretical perfect correlation between expected and observed occurrences.

Table 7.3: Ratios of the number of observed occurrences (Obs) to the number of expected (Exp) occurrences for all 400 possible dipeptide arrangements in the individual PrLD and combined libraries.

	A2 Degradation Library		A2 *ade*$^-$ Library		A1 Degradation Library		A1 *ade*$^-$ Library		Combined Degradation Libraries		Combined *ade*$^-$ Libraries	
Dipeptide	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
AA	0	1.78	1	0.76	0	0.85	2	1.01	0	2.58	3	1.75
AC	1	1.52	1	0.83	3	1.29	2	0.84	4	2.87	3	1.68
AD	1	0.94	1	1.13	0	0.32	2	1.01	1	1.22	3	2.15
AE	1	0.47	0	0.53	0	0.24	1	0.72	1	0.70	1	1.24
AF	5	1.52	1	0.89	0	1.05	0	0.92	5	2.58	1	1.82
AG	0	1.52	2	1.72	1	1.01	2	1.61	1	2.54	4	3.35
AH	4	1.20	1	0.86	2	0.89	1	1.05	6	2.11	2	1.90
AI	2	1.41	0	0.50	2	1.49	0	0.64	4	3.01	0	1.13
AK	0	0.52	0	0.66	0	0.20	1	0.76	0	0.70	1	1.42
AL	5	3.67	0	0.93	1	1.81	1	0.84	6	5.40	1	1.79
AM	2	1.10	0	0.20	0	0.44	1	0.28	2	1.50	1	0.47
AN	0	0.63	1	1.52	1	0.48	2	1.53	1	1.13	3	3.06
AP	2	1.62	2	1.13	0	0.93	1	1.33	2	2.54	3	2.44
AQ	0	0.47	1	0.63	0	0.20	1	0.52	0	0.66	2	1.17
AR	1	1.62	2	2.22	3	1.13	0	2.57	4	2.77	2	4.78
AS	4	3.14	2	2.68	4	1.93	2	2.41	8	5.08	4	5.14
AT	0	1.62	2	1.06	1	1.17	2	1.53	1	2.82	4	2.55
AV	1	2.57	0	0.66	1	1.65	0	0.88	2	4.23	0	1.53
AW	0	0.84	1	0.50	0	0.44	0	0.48	0	1.27	1	0.98
AY	0	1.57	0	0.73	0	0.85	2	0.92	0	2.40	2	1.64
CA	3	1.52	0	0.83	2	1.29	1	0.84	5	2.87	1	1.68
CC	1	1.30	0	0.90	3	1.96	1	0.71	4	3.18	1	1.61
CD	1	0.80	1	1.22	1	0.49	0	0.84	2	1.36	1	2.06
CE	0	0.40	0	0.58	0	0.37	2	0.61	0	0.78	2	1.19
CF	1	1.30	1	0.97	0	1.60	1	0.78	1	2.87	2	1.75
CG	0	1.30	3	1.87	0	1.54	0	1.35	0	2.81	3	3.21
CH	2	1.03	0	0.94	0	1.35	1	0.88	2	2.35	1	1.82
CI	1	1.21	0	0.54	2	2.27	0	0.54	3	3.34	0	1.08
CK	0	0.45	0	0.72	2	0.31	0	0.64	2	0.78	0	1.36
CL	2	3.13	1	1.01	6	2.76	0	0.71	8	5.99	1	1.71
CM	0	0.94	0	0.22	0	0.68	0	0.24	0	1.67	0	0.45
CN	2	0.54	2	1.66	0	0.74	1	1.28	2	1.25	3	2.93
CP	0	1.38	0	1.22	0	1.41	1	1.11	0	2.81	1	2.34
CQ	0	0.40	0	0.68	0	0.31	2	0.44	0	0.73	2	1.12

CR	1	1.38	2	2.41	2	1.72	3	2.16	3	3.08	5	4.58
CS	4	2.68	7	2.91	6	2.95	4	2.03	10	5.63	11	4.93
CT	4	1.38	3	1.15	2	1.78	1	1.28	6	3.13	4	2.45
CV	3	2.19	0	0.72	4	2.52	0	0.74	7	4.69	0	1.47
CW	1	0.71	0	0.54	0	0.68	0	0.41	1	1.41	0	0.94
CY	1	1.34	0	0.79	0	1.29	0	0.78	1	2.66	0	1.57
DA	0	0.94	2	1.13	0	0.32	0	1.01	0	1.22	2	2.15
DC	1	0.80	1	1.22	1	0.49	2	0.84	2	1.36	3	2.06
DD	1	0.50	2	1.66	0	0.12	0	1.01	1	0.58	2	2.64
DE	0	0.25	0	0.78	0	0.09	0	0.72	0	0.33	0	1.52
DF	0	0.80	1	1.32	1	0.40	1	0.92	1	1.22	2	2.24
DG	3	0.80	2	2.54	0	0.38	3	1.61	3	1.20	5	4.12
DH	0	0.64	1	1.27	0	0.34	0	1.05	0	1.00	1	2.33
DI	0	0.75	1	0.73	1	0.57	1	0.64	1	1.42	2	1.39
DK	2	0.28	2	0.98	0	0.08	2	0.76	2	0.33	4	1.75
DL	1	1.94	1	1.37	0	0.69	2	0.84	1	2.55	3	2.20
DM	0	0.58	0	0.29	0	0.17	1	0.28	0	0.71	1	0.58
DN	0	0.33	1	2.25	0	0.18	2	1.53	0	0.53	3	3.76
DP	1	0.86	1	1.66	0	0.35	1	1.33	1	1.20	2	3.00
DQ	1	0.25	1	0.93	0	0.08	1	0.52	1	0.31	2	1.43
DR	0	0.86	4	3.28	0	0.43	2	2.57	0	1.31	6	5.87
DS	1	1.66	1	3.96	1	0.74	1	2.41	2	2.40	2	6.32
DT	2	0.86	4	1.57	2	0.45	2	1.53	4	1.33	6	3.14
DV	0	1.36	0	0.98	1	0.63	1	0.88	1	2.00	1	1.88
DW	0	0.44	2	0.73	0	0.17	0	0.48	0	0.60	2	1.21
DY	1	0.83	2	1.08	0	0.32	1	0.92	1	1.13	3	2.02
EA	1	0.47	1	0.53	0	0.24	3	0.72	1	0.70	4	1.24
EC	1	0.40	1	0.58	0	0.37	1	0.61	1	0.78	2	1.19
ED	0	0.25	1	0.78	0	0.09	1	0.72	0	0.33	2	1.52
EE	0	0.12	0	0.37	0	0.07	0	0.52	0	0.19	0	0.88
EF	0	0.40	0	0.62	0	0.30	0	0.67	0	0.70	0	1.29
EG	1	0.40	1	1.20	0	0.29	2	1.16	1	0.69	3	2.38
EH	0	0.32	2	0.60	0	0.25	1	0.75	0	0.58	3	1.34
EI	0	0.37	0	0.35	0	0.43	1	0.46	0	0.82	1	0.80
EK	0	0.14	0	0.46	0	0.06	1	0.55	0	0.19	1	1.01
EL	0	0.97	0	0.64	1	0.52	0	0.61	1	1.47	0	1.27
EM	1	0.29	0	0.14	0	0.13	0	0.20	1	0.41	0	0.34
EN	0	0.17	1	1.06	0	0.14	0	1.10	0	0.31	1	2.17
EP	2	0.43	1	0.78	1	0.26	0	0.96	3	0.69	1	1.73
EQ	0	0.12	0	0.44	1	0.06	0	0.38	1	0.18	0	0.83
ER	0	0.43	2	1.54	1	0.32	2	1.85	1	0.76	4	3.38
ES	1	0.83	5	1.87	1	0.55	0	1.74	2	1.38	5	3.64
ET	0	0.43	0	0.74	0	0.33	0	1.10	0	0.77	0	1.81

EV	1	0.68	0	0.46	0	0.47	2	0.64	1	1.15	2	1.08
EW	0	0.22	0	0.35	0	0.13	0	0.35	0	0.35	0	0.70
EY	0	0.42	0	0.51	0	0.24	2	0.67	0	0.65	2	1.16
FA	1	1.52	3	0.89	2	1.05	2	0.92	3	2.58	5	1.82
FC	1	1.30	0	0.97	2	1.60	1	0.78	3	2.87	1	1.75
FD	0	0.80	0	1.32	0	0.40	0	0.92	0	1.22	0	2.24
FE	1	0.40	2	0.62	0	0.30	0	0.67	1	0.70	2	1.29
FF	2	1.30	1	1.05	3	1.30	1	0.85	5	2.58	2	1.90
FG	2	1.30	2	2.02	0	1.25	0	1.48	2	2.54	2	3.49
FH	2	1.03	0	1.01	2	1.10	0	0.96	4	2.11	0	1.97
FI	0	1.21	1	0.58	3	1.85	0	0.59	3	3.01	1	1.18
FK	0	0.45	2	0.78	1	0.25	1	0.70	1	0.70	3	1.48
FL	3	3.13	1	1.09	2	2.25	1	0.78	5	5.40	2	1.86
FM	1	0.94	1	0.23	2	0.55	1	0.26	3	1.50	2	0.49
FN	0	0.54	1	1.79	0	0.60	1	1.41	0	1.13	2	3.19
FP	1	1.38	2	1.32	1	1.15	1	1.22	2	2.54	3	2.54
FQ	0	0.40	1	0.74	0	0.25	1	0.48	0	0.66	2	1.22
FR	3	1.38	3	2.60	0	1.40	4	2.37	3	2.77	7	4.98
FS	0	2.68	4	3.15	3	2.39	1	2.22	3	5.08	5	5.35
FT	2	1.38	0	1.24	1	1.45	2	1.41	3	2.82	2	2.66
FV	2	2.19	1	0.78	0	2.05	1	0.81	2	4.23	2	1.60
FW	1	0.71	0	0.58	2	0.55	1	0.44	3	1.27	1	1.03
FY	2	1.34	1	0.85	1	1.05	0	0.85	3	2.40	1	1.71
GA	2	1.52	3	1.72	1	1.01	1	1.61	3	2.54	4	3.35
GC	1	1.30	5	1.87	2	1.54	1	1.35	3	2.81	6	3.21
GD	2	0.80	2	2.54	0	0.38	2	1.61	2	1.20	4	4.12
GE	0	0.40	0	1.20	1	0.29	1	1.16	1	0.69	1	2.38
GF	2	1.30	2	2.02	0	1.25	1	1.48	2	2.54	3	3.49
GG	2	1.30	3	3.89	1	1.20	0	2.57	3	2.49	3	6.43
GH	0	1.03	2	1.95	1	1.06	4	1.67	1	2.08	6	3.63
GI	1	1.21	1	1.12	0	1.77	0	1.03	1	2.95	1	2.17
GK	0	0.45	3	1.50	1	0.24	2	1.22	1	0.69	5	2.73
GL	4	3.13	3	2.10	5	2.16	0	1.35	9	5.31	3	3.42
GM	0	0.94	2	0.45	1	0.53	0	0.45	1	1.48	2	0.91
GN	0	0.54	5	3.44	1	0.58	4	2.44	1	1.11	9	5.87
GP	0	1.38	3	2.54	1	1.10	2	2.12	1	2.49	5	4.68
GQ	0	0.40	2	1.42	0	0.24	1	0.84	0	0.65	3	2.24
GR	3	1.38	3	5.01	0	1.34	4	4.12	3	2.72	7	9.15
GS	2	2.68	4	6.06	0	2.30	3	3.86	2	4.98	7	9.85
GT	3	1.38	0	2.39	0	1.39	2	2.44	3	2.77	2	4.89
GV	4	2.19	3	1.50	4	1.97	0	1.42	8	4.15	3	2.93
GW	0	0.71	2	1.12	0	0.53	0	0.77	0	1.25	2	1.89
GY	0	1.34	1	1.65	4	1.01	3	1.48	4	2.35	4	3.14

HA	1	1.20	0	0.86	1	0.89	1	1.05	2	2.11	1	1.90
HC	2	1.03	3	0.94	2	1.35	1	0.88	4	2.35	4	1.82
HD	0	0.64	0	1.27	0	0.34	2	1.05	0	1.00	2	2.33
HE	0	0.32	1	0.60	0	0.25	0	0.75	0	0.58	1	1.34
HF	0	1.03	0	1.01	1	1.10	1	0.96	1	2.11	1	1.97
HG	2	1.03	3	1.95	0	1.06	1	1.67	2	2.08	4	3.63
HH	0	0.81	0	0.97	1	0.93	1	1.09	1	1.73	1	2.05
HI	0	0.96	0	0.56	2	1.56	0	0.67	2	2.46	0	1.22
HK	0	0.35	0	0.75	0	0.21	1	0.79	0	0.58	1	1.54
HL	2	2.48	1	1.05	0	1.90	0	0.88	2	4.42	1	1.94
HM	1	0.74	0	0.22	0	0.46	1	0.29	1	1.23	1	0.51
HN	1	0.43	3	1.72	0	0.51	1	1.59	1	0.92	4	3.32
HP	1	1.10	1	1.27	0	0.97	0	1.38	1	2.08	1	2.65
HQ	0	0.32	0	0.71	2	0.21	1	0.54	2	0.54	1	1.26
HR	2	1.10	3	2.51	4	1.18	1	2.68	6	2.27	4	5.17
HS	2	2.13	6	3.03	1	2.03	3	2.51	3	4.15	9	5.57
HT	2	1.10	3	1.20	0	1.22	1	1.59	2	2.31	4	2.76
HV	2	1.74	1	0.75	1	1.73	2	0.92	3	3.46	3	1.66
HW	0	0.57	0	0.56	0	0.46	2	0.50	0	1.04	2	1.07
HY	1	1.06	0	0.82	0	0.89	2	0.96	1	1.96	2	1.78
IA	3	1.41	0	0.50	0	1.49	1	0.64	3	3.01	1	1.13
IC	2	1.21	1	0.54	2	2.27	0	0.54	4	3.34	1	1.08
ID	0	0.75	1	0.73	0	0.57	1	0.64	0	1.42	2	1.39
IE	1	0.37	1	0.35	0	0.43	1	0.46	1	0.82	2	0.80
IF	1	1.21	0	0.58	3	1.85	0	0.59	4	3.01	0	1.18
IG	2	1.21	1	1.12	2	1.77	2	1.03	4	2.95	3	2.17
IH	1	0.96	0	0.56	3	1.56	0	0.67	4	2.46	0	1.22
II	2	1.12	0	0.32	2	2.63	1	0.41	4	3.50	1	0.73
IK	1	0.42	0	0.43	0	0.35	2	0.49	1	0.82	2	0.92
IL	1	2.91	0	0.60	2	3.19	0	0.54	3	6.29	0	1.15
IM	0	0.87	0	0.13	0	0.78	0	0.18	0	1.75	0	0.31
IN	0	0.50	0	0.99	1	0.85	1	0.98	1	1.31	1	1.98
IP	1	1.29	0	0.73	1	1.63	1	0.85	2	2.95	1	1.58
IQ	1	0.37	0	0.41	0	0.35	0	0.33	1	0.77	0	0.75
IR	3	1.29	2	1.45	1	1.99	4	1.65	4	3.23	6	3.08
IS	1	2.50	4	1.75	6	3.41	0	1.54	7	5.91	4	3.32
IT	1	1.29	0	0.69	3	2.06	0	0.98	4	3.28	0	1.65
IV	0	2.04	1	0.43	0	2.91	1	0.57	0	4.92	2	0.99
IW	2	0.67	0	0.32	2	0.78	0	0.31	4	1.48	0	0.64
IY	1	1.25	2	0.47	1	1.49	0	0.59	2	2.79	2	1.06
KA	0	0.52	1	0.66	0	0.20	0	0.76	0	0.70	1	1.42
KC	0	0.45	2	0.72	1	0.31	1	0.64	1	0.78	3	1.36
KD	0	0.28	0	0.98	0	0.08	1	0.76	0	0.33	1	1.75

KE	1	0.14	1	0.46	1	0.06	1	0.55	2	0.19	2	1.01
KF	0	0.45	0	0.78	1	0.25	0	0.70	1	0.70	0	1.48
KG	0	0.45	0	1.50	0	0.24	2	1.22	0	0.69	2	2.73
KH	0	0.35	0	0.75	0	0.21	0	0.79	0	0.58	0	1.54
KI	0	0.42	0	0.43	0	0.35	1	0.49	0	0.82	1	0.92
KK	0	0.15	1	0.58	0	0.05	0	0.58	0	0.19	1	1.16
KL	0	1.08	0	0.81	0	0.43	1	0.64	0	1.47	1	1.45
KM	0	0.32	0	0.17	1	0.11	0	0.21	1	0.41	0	0.39
KN	1	0.18	1	1.32	0	0.12	1	1.16	1	0.31	2	2.49
KP	0	0.48	1	0.98	0	0.22	2	1.01	0	0.69	3	1.98
KQ	0	0.14	1	0.55	0	0.05	0	0.40	0	0.18	1	0.95
KR	1	0.48	0	1.93	0	0.27	2	1.96	1	0.76	2	3.88
KS	1	0.92	3	2.33	0	0.46	1	1.83	1	1.38	4	4.18
KT	0	0.48	3	0.92	0	0.28	1	1.16	0	0.77	4	2.07
KV	1	0.75	3	0.58	1	0.39	1	0.67	2	1.15	4	1.24
KW	0	0.25	0	0.43	0	0.11	0	0.37	0	0.35	0	0.80
KY	3	0.46	0	0.63	0	0.20	0	0.70	3	0.65	0	1.33
LA	8	3.67	1	0.93	3	1.81	2	0.84	11	5.40	3	1.79
LC	3	3.13	1	1.01	3	2.76	2	0.71	6	5.99	3	1.71
LD	3	1.94	2	1.37	0	0.69	1	0.84	3	2.55	3	2.20
LE	0	0.97	1	0.64	0	0.52	0	0.61	0	1.47	1	1.27
LF	5	3.13	0	1.09	1	2.25	0	0.78	6	5.40	0	1.86
LG	4	3.13	2	2.10	4	2.16	1	1.35	8	5.31	3	3.42
LH	1	2.48	1	1.05	2	1.90	2	0.88	3	4.42	3	1.94
LI	1	2.91	0	0.60	4	3.19	0	0.54	5	6.29	0	1.15
LK	0	1.08	2	0.81	0	0.43	2	0.64	0	1.47	4	1.45
LL	11	7.55	1	1.13	4	3.89	0	0.71	15	11.30	1	1.82
LM	1	2.26	0	0.24	0	0.95	0	0.24	1	3.14	0	0.48
LN	2	1.29	3	1.85	0	1.04	2	1.28	2	2.36	5	3.13
LP	2	3.34	2	1.37	3	1.99	0	1.11	5	5.31	2	2.49
LQ	1	0.97	2	0.77	0	0.43	0	0.44	1	1.38	2	1.19
LR	4	3.34	3	2.70	3	2.42	1	2.16	7	5.80	4	4.88
LS	7	6.47	4	3.26	4	4.14	3	2.03	11	10.61	7	5.25
LT	3	3.34	0	1.29	2	2.50	0	1.28	5	5.90	0	2.61
LV	4	5.28	1	0.81	3	3.54	1	0.74	7	8.84	2	1.56
LW	1	1.73	0	0.60	1	0.95	0	0.41	2	2.65	0	1.00
LY	2	3.24	0	0.89	2	1.81	0	0.78	4	5.01	0	1.67
MA	1	1.10	0	0.20	0	0.44	0	0.28	1	1.50	0	0.47
MC	2	0.94	0	0.22	0	0.68	0	0.24	2	1.67	0	0.45
MD	0	0.58	0	0.29	0	0.17	1	0.28	0	0.71	1	0.58
ME	1	0.29	0	0.14	0	0.13	0	0.20	1	0.41	0	0.34
MF	2	0.94	0	0.23	1	0.55	0	0.26	3	1.50	0	0.49
MG	0	0.94	0	0.45	0	0.53	0	0.45	0	1.48	0	0.91

MH	0	0.74	0	0.22	1	0.46	1	0.29	1	1.23	1	0.51
MI	1	0.87	0	0.13	1	0.78	0	0.18	2	1.75	0	0.31
MK	1	0.32	0	0.17	0	0.11	0	0.21	1	0.41	0	0.39
ML	3	2.26	0	0.24	0	0.95	0	0.24	3	3.14	0	0.48
MM	1	0.68	0	0.05	0	0.23	0	0.08	1	0.88	0	0.13
MN	0	0.39	1	0.40	0	0.25	1	0.43	0	0.66	2	0.83
MP	0	1.00	0	0.29	1	0.49	0	0.37	1	1.48	0	0.66
MQ	1	0.29	0	0.16	0	0.11	0	0.15	1	0.38	0	0.32
MR	1	1.00	1	0.58	2	0.59	1	0.72	3	1.61	2	1.29
MS	0	1.94	1	0.70	1	1.01	1	0.68	1	2.95	2	1.39
MT	0	1.00	0	0.28	1	0.61	0	0.43	1	1.64	0	0.69
MV	1	1.59	0	0.17	1	0.87	0	0.25	2	2.46	0	0.41
MW	1	0.52	0	0.13	0	0.23	0	0.14	1	0.74	0	0.27
MY	2	0.97	0	0.19	0	0.44	0	0.26	2	1.39	0	0.44
NA	1	0.63	1	1.52	0	0.48	1	1.53	1	1.13	2	3.06
NC	0	0.54	0	1.66	2	0.74	0	1.28	2	1.25	0	2.93
ND	0	0.33	4	2.25	0	0.18	5	1.53	0	0.53	9	3.76
NE	1	0.17	0	1.06	0	0.14	1	1.10	1	0.31	1	2.17
NF	2	0.54	3	1.79	1	0.60	0	1.41	3	1.13	3	3.19
NG	0	0.54	9	3.44	1	0.58	3	2.44	1	1.11	12	5.87
NH	0	0.43	3	1.72	0	0.51	1	1.59	0	0.92	4	3.32
NI	1	0.50	2	0.99	1	0.85	1	0.98	2	1.31	3	1.98
NK	0	0.18	0	1.32	0	0.12	0	1.16	0	0.31	0	2.49
NL	1	1.29	3	1.85	3	1.04	7	1.28	4	2.36	10	3.13
NM	0	0.39	1	0.40	0	0.25	0	0.43	0	0.66	1	0.83
NN	0	0.22	2	3.05	0	0.28	1	2.32	0	0.49	3	5.36
NP	1	0.57	0	2.25	0	0.53	1	2.02	1	1.11	1	4.27
NQ	0	0.17	1	1.26	0	0.12	1	0.79	0	0.29	2	2.04
NR	0	0.57	3	4.44	1	0.64	3	3.91	1	1.21	6	8.36
NS	3	1.11	4	5.36	1	1.11	2	3.67	4	2.21	6	9.00
NT	1	0.57	1	2.12	0	0.67	1	2.32	1	1.23	2	4.47
NV	0	0.91	0	1.32	1	0.94	4	1.34	1	1.85	4	2.68
NW	1	0.30	1	0.99	0	0.25	0	0.73	1	0.55	1	1.72
NY	0	0.55	1	1.46	0	0.48	1	1.41	0	1.05	2	2.87
PA	2	1.62	3	1.13	3	0.93	0	1.33	5	2.54	3	2.44
PC	1	1.38	0	1.22	3	1.41	2	1.11	4	2.81	2	2.34
PD	1	0.86	1	1.66	0	0.35	1	1.33	1	1.20	2	3.00
PE	0	0.43	1	0.78	0	0.26	1	0.96	0	0.69	2	1.73
PF	0	1.38	1	1.32	1	1.15	2	1.22	1	2.54	3	2.54
PG	0	1.38	1	2.54	0	1.10	4	2.12	0	2.49	5	4.68
PH	3	1.10	1	1.27	2	0.97	3	1.38	5	2.08	4	2.65
PI	1	1.29	0	0.73	1	1.63	1	0.85	2	2.95	1	1.58
PK	0	0.48	1	0.98	0	0.22	1	1.01	0	0.69	2	1.98

PL	3	3.34	1	1.37	1	1.99	0	1.11	4	5.31	1	2.49
PM	3	1.00	0	0.29	0	0.49	0	0.37	3	1.48	0	0.66
PN	1	0.57	1	2.25	1	0.53	3	2.02	2	1.11	4	4.27
PP	2	1.48	3	1.66	0	1.02	1	1.75	2	2.49	4	3.41
PQ	1	0.43	2	0.93	1	0.22	1	0.69	2	0.65	3	1.63
PR	2	1.48	4	3.28	1	1.24	3	3.40	3	2.72	7	6.67
PS	2	2.87	3	3.96	1	2.12	3	3.18	3	4.98	6	7.18
PT	2	1.48	2	1.57	1	1.28	1	2.02	3	2.77	3	3.56
PV	3	2.34	0	0.98	1	1.81	0	1.17	4	4.15	0	2.14
PW	0	0.76	0	0.73	0	0.49	2	0.64	0	1.25	2	1.37
PY	0	1.43	2	1.08	1	0.93	2	1.22	1	2.35	4	2.29
QA	0	0.47	0	0.63	0	0.20	0	0.52	0	0.66	0	1.17
QC	1	0.40	0	0.68	0	0.31	0	0.44	1	0.73	0	1.12
QD	1	0.25	2	0.93	1	0.08	0	0.52	2	0.31	2	1.43
QE	0	0.12	0	0.44	0	0.06	0	0.38	0	0.18	0	0.83
QF	1	0.40	2	0.74	0	0.25	0	0.48	1	0.66	2	1.22
QG	0	0.40	2	1.42	0	0.24	0	0.84	0	0.65	2	2.24
QH	0	0.32	1	0.71	0	0.21	2	0.54	0	0.54	3	1.26
QI	1	0.37	1	0.41	0	0.35	0	0.33	1	0.77	1	0.75
QK	0	0.14	0	0.55	0	0.05	1	0.40	0	0.18	1	0.95
QL	1	0.97	0	0.77	0	0.43	0	0.44	1	1.38	0	1.19
QM	1	0.29	0	0.16	0	0.11	0	0.15	1	0.38	0	0.32
QN	0	0.17	0	1.26	0	0.12	0	0.79	0	0.29	0	2.04
QP	1	0.43	1	0.93	1	0.22	0	0.69	2	0.65	1	1.63
QQ	0	0.12	0	0.52	0	0.05	0	0.27	0	0.17	0	0.78
QR	0	0.43	4	1.83	1	0.27	1	1.34	1	0.71	5	3.18
QS	0	0.83	0	2.21	1	0.46	4	1.25	1	1.29	4	3.43
QT	0	0.43	3	0.88	0	0.28	4	0.79	0	0.72	7	1.70
QV	0	0.68	1	0.55	0	0.39	0	0.46	0	1.08	1	1.02
QW	0	0.22	0	0.41	0	0.11	0	0.25	0	0.32	0	0.66
QY	0	0.42	1	0.60	0	0.20	0	0.48	0	0.61	1	1.09
RA	3	1.62	1	2.22	1	1.13	4	2.57	4	2.77	5	4.78
RC	1	1.38	3	2.41	2	1.72	1	2.16	3	3.08	4	4.58
RD	0	0.86	4	3.28	0	0.43	1	2.57	0	1.31	5	5.87
RE	0	0.43	4	1.54	1	0.32	1	1.85	1	0.76	5	3.38
RF	2	1.38	3	2.60	1	1.40	3	2.37	3	2.77	6	4.98
RG	0	1.38	4	5.01	1	1.34	5	4.12	1	2.72	9	9.15
RH	0	1.10	2	2.51	0	1.18	1	2.68	0	2.27	3	5.17
RI	4	1.29	2	1.45	2	1.99	1	1.65	6	3.23	3	3.08
RK	1	0.48	1	1.93	0	0.27	1	1.96	1	0.76	2	3.88
RL	2	3.34	3	2.70	1	2.42	2	2.16	3	5.80	5	4.88
RM	0	1.00	1	0.58	1	0.59	1	0.72	1	1.61	2	1.29
RN	1	0.57	4	4.44	0	0.64	3	3.91	1	1.21	7	8.36

RP	2	1.48	3	3.28	4	1.24	6	3.40	6	2.72	9	6.67
RQ	0	0.43	2	1.83	0	0.27	0	1.34	0	0.71	2	3.18
RR	1	1.48	5	6.46	0	1.50	2	6.59	1	2.97	7	13.03
RS	2	2.87	3	7.81	3	2.58	10	6.18	5	5.44	13	14.03
RT	0	1.48	5	3.09	1	1.56	5	3.91	1	3.02	10	6.97
RV	0	2.34	3	1.93	1	2.20	5	2.26	1	4.54	8	4.18
RW	0	0.76	1	1.45	1	0.59	2	1.24	1	1.36	3	2.69
RY	1	1.43	4	2.12	0	1.13	1	2.37	1	2.57	5	4.48
SA	2	3.14	1	2.68	2	1.93	3	2.41	4	5.08	4	5.14
SC	3	2.68	3	2.91	2	2.95	3	2.03	5	5.63	6	4.93
SD	2	1.66	10	3.96	1	0.74	1	2.41	3	2.40	11	6.32
SE	1	0.83	2	1.87	1	0.55	1	1.74	2	1.38	3	3.64
SF	1	2.68	4	3.15	3	2.39	2	2.22	4	5.08	6	5.35
SG	3	2.68	5	6.06	4	2.30	3	3.86	7	4.98	8	9.85
SH	3	2.13	1	3.03	3	2.03	2	2.51	6	4.15	3	5.57
SI	3	2.50	1	1.75	4	3.41	3	1.54	7	5.91	4	3.32
SK	2	0.92	3	2.33	0	0.46	1	1.83	2	1.38	4	4.18
SL	7	6.47	2	3.26	5	4.14	0	2.03	12	10.61	2	5.25
SM	1	1.94	1	0.70	2	1.01	1	0.68	3	2.95	2	1.39
SN	1	1.11	4	5.36	1	1.11	5	3.67	2	2.21	9	9.00
SP	2	2.87	3	3.96	1	2.12	4	3.18	3	4.98	7	7.18
SQ	1	0.83	2	2.21	1	0.46	2	1.25	2	1.29	4	3.43
SR	0	2.87	9	7.81	1	2.58	5	6.18	1	5.44	14	14.03
SS	7	5.55	14	9.44	4	4.42	8	5.79	11	9.97	22	15.10
ST	2	2.87	0	3.73	2	2.67	4	3.67	4	5.54	4	7.50
SV	10	4.53	0	2.33	7	3.78	3	2.12	17	8.31	3	4.50
SW	2	1.48	3	1.75	0	1.01	2	1.16	2	2.49	5	2.89
SY	3	2.77	2	2.56	3	1.93	1	2.22	6	4.71	3	4.82
TA	1	1.62	0	1.06	0	1.17	1	1.53	1	2.82	1	2.55
TC	0	1.38	1	1.15	0	1.78	1	1.28	0	3.13	2	2.45
TD	1	0.86	1	1.57	1	0.45	0	1.53	2	1.33	1	3.14
TE	1	0.43	3	0.74	0	0.33	1	1.10	1	0.77	4	1.81
TF	2	1.38	1	1.24	4	1.45	4	1.41	6	2.82	5	2.66
TG	0	1.38	3	2.39	4	1.39	3	2.44	4	2.77	6	4.89
TH	2	1.10	0	1.20	0	1.22	3	1.59	2	2.31	3	2.76
TI	1	1.29	1	0.69	2	2.06	1	0.98	3	3.28	2	1.65
TK	0	0.48	2	0.92	0	0.28	2	1.16	0	0.77	4	2.07
TL	6	3.34	3	1.29	0	2.50	3	1.28	6	5.90	6	2.61
TM	0	1.00	0	0.28	1	0.61	0	0.43	1	1.64	0	0.69
TN	0	0.57	1	2.12	0	0.67	2	2.32	0	1.23	3	4.47
TP	2	1.48	2	1.57	1	1.28	3	2.02	3	2.77	5	3.56
TQ	1	0.43	0	0.88	0	0.28	0	0.79	1	0.72	0	1.70
TR	1	1.48	4	3.09	3	1.56	5	3.91	4	3.02	9	6.97

TS	4	2.87	2	3.73	3	2.67	3	3.67	7	5.54	5	7.50
TT	2	1.48	1	1.47	4	1.61	0	2.32	6	3.08	1	3.72
TV	4	2.34	2	0.92	2	2.28	0	1.34	6	4.61	2	2.23
TW	1	0.76	1	0.69	1	0.61	2	0.73	2	1.38	3	1.44
TY	1	1.43	0	1.01	1	1.17	3	1.41	2	2.61	3	2.39
VA	0	2.57	0	0.66	1	1.65	1	0.88	1	4.23	1	1.53
VC	1	2.19	1	0.72	2	2.52	0	0.74	3	4.69	1	1.47
VD	0	1.36	1	0.98	1	0.63	0	0.88	1	2.00	1	1.88
VE	1	0.68	0	0.46	0	0.47	2	0.64	1	1.15	2	1.08
VF	0	2.19	2	0.78	3	2.05	2	0.81	3	4.23	4	1.60
VG	1	2.19	0	1.50	2	1.97	1	1.42	3	4.15	1	2.93
VH	2	1.74	4	0.75	3	1.73	1	0.92	5	3.46	5	1.66
VI	4	2.04	0	0.43	3	2.91	0	0.57	7	4.92	0	0.99
VK	1	0.75	2	0.58	1	0.39	0	0.67	2	1.15	2	1.24
VL	6	5.28	1	0.81	5	3.54	1	0.74	11	8.84	2	1.56
VM	2	1.59	0	0.17	1	0.87	0	0.25	3	2.46	0	0.41
VN	0	0.91	4	1.32	1	0.94	2	1.34	1	1.85	6	2.68
VP	3	2.34	1	0.98	3	1.81	2	1.17	6	4.15	3	2.14
VQ	1	0.68	0	0.55	0	0.39	0	0.46	1	1.08	0	1.02
VR	3	2.34	1	1.93	1	2.20	3	2.26	4	4.54	4	4.18
VS	5	4.53	0	2.33	2	3.78	2	2.12	7	8.31	2	4.50
VT	3	2.34	0	0.92	1	2.28	2	1.34	4	4.61	2	2.23
VV	2	3.70	0	0.58	4	3.23	1	0.78	6	6.92	1	1.34
VW	3	1.21	0	0.43	1	0.87	0	0.42	4	2.08	0	0.86
VY	5	2.26	1	0.63	2	1.65	0	0.81	7	3.92	1	1.44
WA	0	0.84	1	0.50	0	0.44	1	0.48	0	1.27	2	0.98
WC	1	0.71	0	0.54	1	0.68	1	0.41	2	1.41	1	0.94
WD	0	0.44	0	0.73	0	0.17	1	0.48	0	0.60	1	1.21
WE	0	0.22	0	0.35	0	0.13	0	0.35	0	0.35	0	0.70
WF	0	0.71	1	0.58	0	0.55	1	0.44	0	1.27	2	1.03
WG	1	0.71	0	1.12	0	0.53	0	0.77	1	1.25	0	1.89
WH	1	0.57	0	0.56	0	0.46	0	0.50	1	1.04	0	1.07
WI	0	0.67	0	0.32	0	0.78	0	0.31	0	1.48	0	0.64
WK	0	0.25	0	0.43	0	0.11	0	0.37	0	0.35	0	0.80
WL	2	1.73	0	0.60	0	0.95	0	0.41	2	2.65	0	1.00
WM	0	0.52	0	0.13	1	0.23	1	0.14	1	0.74	1	0.27
WN	0	0.30	2	0.99	0	0.25	0	0.73	0	0.55	2	1.72
WP	2	0.76	2	0.73	1	0.49	0	0.64	3	1.25	2	1.37
WQ	0	0.22	0	0.41	0	0.11	0	0.25	0	0.32	0	0.66
WR	2	0.76	2	1.45	1	0.59	2	1.24	3	1.36	4	2.69
WS	1	1.48	3	1.75	1	1.01	2	1.16	2	2.49	5	2.89
WT	1	0.76	0	0.69	2	0.61	1	0.73	3	1.38	1	1.44
WV	1	1.21	1	0.43	0	0.87	0	0.42	1	2.08	1	0.86

WW	1	0.39	0	0.32	1	0.23	0	0.23	2	0.62	0	0.55
WY	2	0.74	0	0.47	2	0.44	0	0.44	4	1.18	0	0.92
YA	2	1.57	0	0.73	0	0.85	1	0.92	2	2.40	1	1.64
YC	3	1.34	1	0.79	0	1.29	1	0.78	3	2.66	2	1.57
YD	3	0.83	0	1.08	1	0.32	1	0.92	4	1.13	1	2.02
YE	0	0.42	0	0.51	1	0.24	1	0.67	1	0.65	1	1.16
YF	1	1.34	1	0.85	1	1.05	1	0.85	2	2.40	2	1.71
YG	1	1.34	2	1.65	1	1.01	2	1.48	2	2.35	4	3.14
YH	1	1.06	2	0.82	2	0.89	0	0.96	3	1.96	2	1.78
YI	2	1.25	1	0.47	4	1.49	1	0.59	6	2.79	2	1.06
YK	1	0.46	1	0.63	0	0.20	0	0.70	1	0.65	1	1.33
YL	1	3.24	3	0.89	3	1.81	0	0.78	4	5.01	3	1.67
YM	0	0.97	0	0.19	0	0.44	0	0.26	0	1.39	0	0.44
YN	0	0.55	0	1.46	1	0.48	3	1.41	1	1.05	3	2.87
YP	2	1.43	0	1.08	1	0.93	3	1.22	3	2.35	3	2.29
YQ	0	0.42	0	0.60	0	0.20	1	0.48	0	0.61	1	1.09
YR	1	1.43	2	2.12	1	1.13	3	2.37	2	2.57	5	4.48
YS	3	2.77	2	2.56	0	1.93	0	2.22	3	4.71	2	4.82
YT	0	1.43	1	1.01	2	1.17	0	1.41	2	2.61	1	2.39
YV	3	2.26	1	0.63	1	1.65	0	0.81	4	3.92	1	1.44
YW	0	0.74	1	0.47	0	0.44	0	0.44	0	1.18	1	0.92
YY	3	1.39	3	0.70	0	0.85	2	0.85	3	2.22	5	1.54

degradation libraries and *ade*$^-$ libraries, each 8-amino acid sequence was scrambled *in silico*. Scrambling each sequence individually has the advantage of maintaining the overall compositional profile of each individual sequence that promotes either degradation or stability while still disrupting potential primary sequence elements. The scrambled sequence libraries were re-evaluated for potential dipeptide biases. As expected, the correlation between the observed and expected dipeptide occurrences in the scrambled degradation and *ade*$^-$ libraries approaches a perfect 1:1 correlation (Figure 7.4, and Table 7.4). However, the values for the slopes of the best fit lines are comparable to those observed for the number of observed versus expected dipeptides in each of the original (non-scrambled) datasets (Figure 7.3). Furthermore, the R^2 values for the scrambled datasets are comparable to those observed for the original datasets, indicating a similar degree of spread in the data.

DISCUSSION

Primary amino acid sequence and amino acid composition are inevitably intertwined, and their effects on protein function or fate can be difficult to disentangle. Furthermore, the difficulty in evaluating primary sequence features in our sequence libraries is compounded by limitations in sample sizes for each primary sequence element. For example, evaluating individual amino acid position preferences requires calculating observed and expected occurrences for each individual amino acid at each of 8 possible positions within the mutagenized region. Analogously, with 400 possible 2-amino acid combinations, evaluating dipeptide enrichment/depletion requires a very large dataset to reach an adequate sample size for each dipeptide.

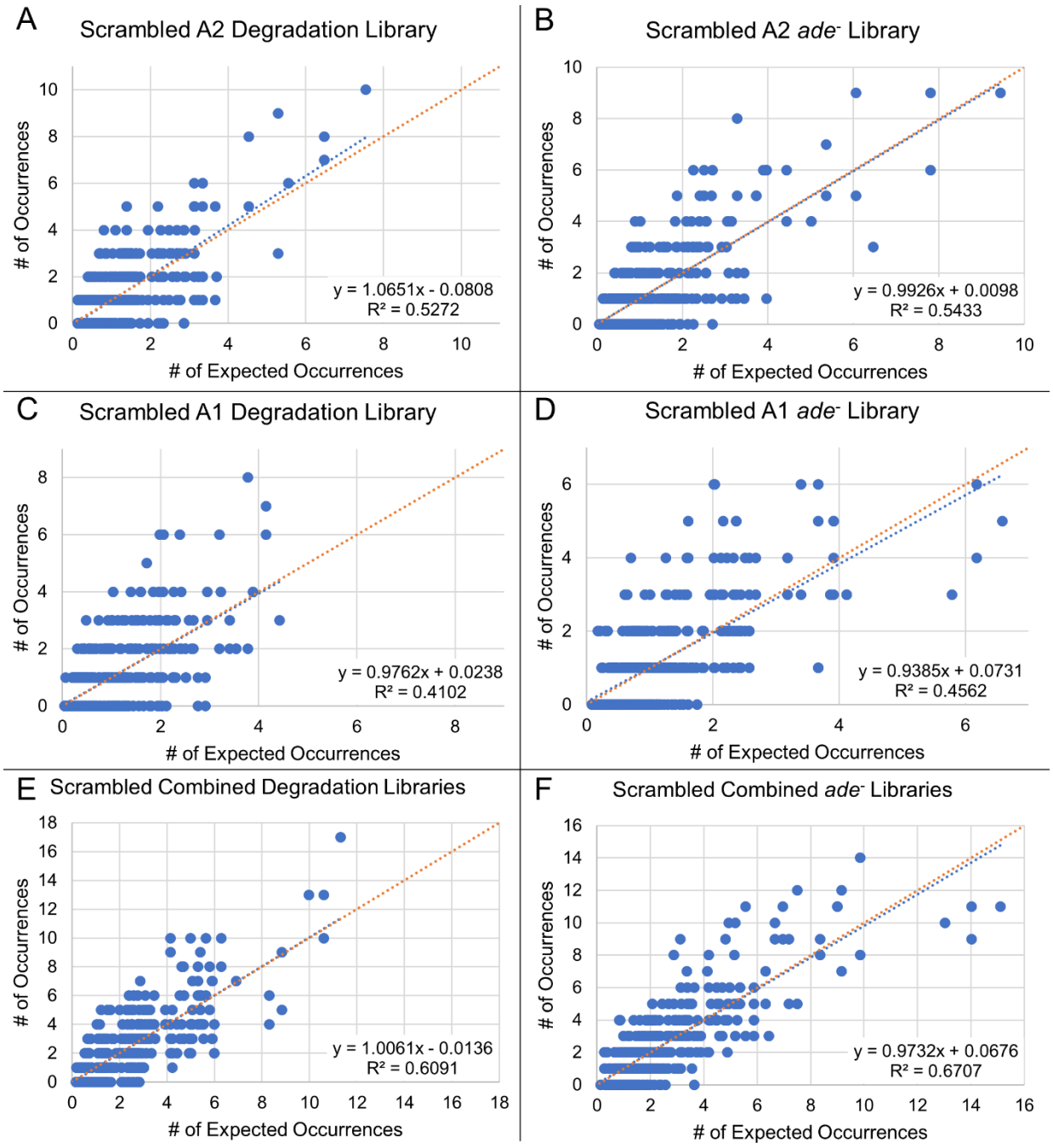


Figure 7.4: Evaluation of dipeptides in scrambled sequence libraries. For each dipeptide, the observed number of occurrences was plotted against the expected number of occurrences for the scrambled A2 **(A, B)** and A1 **(C, D)** individual libraries as well as combined **(E, F)** libraries. In each case, the degradation libraries (*left*) and *ade*- libraries (*right*) were evaluated. In all graphs, the blue line represents the best linear fit, whereas the orange line indicates a theoretical perfect correlation between expected and observed occurrences.

Table 7.4: Ratios of the number of observed occurrences (Obs) to the number of expected (Exp) occurrences for all 400 possible dipeptide arrangements in the scrambled sequence libraries.

	Scrambled Combined Degradation Libraries		Scrambled Combined *ade*$^-$ Libraries		Scrambled A2 Degradation Library		Scrambled A2 *ade*$^-$ Library		Scrambled A1 Degradation Library		Scrambled A1 *ade*$^-$ Library	
Dipeptide	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
AA	5	2.58	2	1.75	1	1.78	0	0.76	1	0.85	3	1.01
AC	4	2.87	2	1.68	2	1.52	1	0.83	1	1.29	2	0.84
AD	0	1.22	3	2.15	0	0.94	1	1.13	0	0.32	0	1.01
AE	1	0.70	1	1.24	0	0.47	1	0.53	1	0.24	0	0.72
AF	3	2.58	2	1.82	0	1.52	4	0.89	0	1.05	0	0.92
AG	3	2.54	4	3.35	3	1.52	1	1.72	1	1.01	4	1.61
AH	0	2.11	3	1.90	2	1.20	1	0.86	2	0.89	0	1.05
AI	1	3.01	0	1.13	0	1.41	0	0.50	2	1.49	0	0.64
AK	0	0.70	2	1.42	0	0.52	1	0.66	0	0.20	0	0.76
AL	9	5.40	1	1.79	5	3.67	2	0.93	0	1.81	1	0.84
AM	0	1.50	1	0.47	0	1.10	1	0.20	0	0.44	0	0.28
AN	2	1.13	3	3.06	0	0.63	1	1.52	3	0.48	1	1.53
AP	5	2.54	1	2.44	3	1.62	1	1.13	0	0.93	1	1.33
AQ	1	0.66	1	1.17	1	0.47	1	0.63	0	0.20	0	0.52
AR	4	2.77	4	4.78	2	1.62	1	2.22	0	1.13	4	2.57
AS	4	5.08	8	5.14	4	3.14	1	2.68	0	1.93	1	2.41
AT	1	2.82	0	2.55	3	1.62	1	1.06	1	1.17	1	1.53
AV	1	4.23	2	1.53	2	2.57	0	0.66	3	1.65	0	0.88
AW	0	1.27	0	0.98	1	0.84	1	0.50	0	0.44	1	0.48
AY	2	2.40	1	1.64	2	1.57	2	0.73	1	0.85	3	0.92
CA	1	2.87	2	1.68	3	1.52	1	0.83	0	1.29	2	0.84
CC	3	3.18	4	1.61	0	1.30	1	0.90	0	1.96	1	0.71
CD	0	1.36	4	2.06	0	0.80	0	1.22	0	0.49	1	0.84
CE	0	0.78	0	1.19	0	0.40	1	0.58	1	0.37	1	0.61
CF	7	2.87	0	1.75	2	1.30	1	0.97	4	1.60	1	0.78
CG	1	2.81	4	3.21	2	1.30	3	1.87	1	1.54	1	1.35
CH	1	2.35	2	1.82	0	1.03	1	0.94	1	1.35	0	0.88
CI	4	3.34	1	1.08	0	1.21	0	0.54	1	2.27	0	0.54
CK	3	0.78	1	1.36	0	0.45	1	0.72	2	0.31	0	0.64
CL	4	5.99	1	1.71	6	3.13	0	1.01	0	2.76	0	0.71

CM	1	1.67	0	0.45	1	0.94	0	0.22	0	0.68	0	0.24
CN	2	1.25	4	2.93	1	0.54	1	1.66	0	0.74	0	1.28
CP	2	2.81	2	2.34	0	1.38	0	1.22	3	1.41	0	1.11
CQ	0	0.73	0	1.12	0	0.40	0	0.68	0	0.31	1	0.44
CR	4	3.08	5	4.58	1	1.38	3	2.41	2	1.72	2	2.16
CS	10	5.63	5	4.93	4	2.68	3	2.91	4	2.95	6	2.03
CT	2	3.13	2	2.45	2	1.38	2	1.15	3	1.78	2	1.28
CV	3	4.69	1	1.47	3	2.19	0	0.72	1	2.52	0	0.74
CW	2	1.41	1	0.94	0	0.71	0	0.54	0	0.68	0	0.41
CY	2	2.66	0	1.57	2	1.34	0	0.79	1	1.29	0	0.78
DA	1	1.22	3	2.15	1	0.94	3	1.13	0	0.32	1	1.01
DC	2	1.36	3	2.06	0	0.80	2	1.22	0	0.49	1	0.84
DD	1	0.58	3	2.64	1	0.50	0	1.66	0	0.12	0	1.01
DE	0	0.33	1	1.52	1	0.25	0	0.78	0	0.09	2	0.72
DF	1	1.22	1	2.24	0	0.80	1	1.32	1	0.40	1	0.92
DG	3	1.20	2	4.12	1	0.80	2	2.54	0	0.38	0	1.61
DH	0	1.00	2	2.33	0	0.64	2	1.27	1	0.34	0	1.05
DI	3	1.42	1	1.39	2	0.75	2	0.73	0	0.57	3	0.64
DK	1	0.33	1	1.75	0	0.28	2	0.98	0	0.08	1	0.76
DL	1	2.55	4	2.20	4	1.94	2	1.37	2	0.69	1	0.84
DM	0	0.71	1	0.58	0	0.58	0	0.29	0	0.17	0	0.28
DN	1	0.53	3	3.76	0	0.33	3	2.25	0	0.18	1	1.53
DP	0	1.20	1	3.00	0	0.86	0	1.66	1	0.35	1	1.33
DQ	0	0.31	1	1.43	1	0.25	1	0.93	0	0.08	0	0.52
DR	1	1.31	3	5.87	1	0.86	2	3.28	0	0.43	2	2.57
DS	3	2.40	7	6.32	1	1.66	1	3.96	1	0.74	1	2.41
DT	1	1.33	6	3.14	0	0.86	3	1.57	2	0.45	1	1.53
DV	2	2.00	2	1.88	0	1.36	2	0.98	0	0.63	2	0.88
DW	0	0.60	1	1.21	0	0.44	1	0.73	0	0.17	0	0.48
DY	3	1.13	1	2.02	3	0.83	1	1.08	0	0.32	2	0.92
EA	1	0.70	2	1.24	1	0.47	1	0.53	0	0.24	1	0.72
EC	0	0.78	0	1.19	1	0.40	1	0.58	0	0.37	1	0.61
ED	0	0.33	1	1.52	0	0.25	0	0.78	0	0.09	0	0.72
EE	0	0.19	1	0.88	1	0.12	1	0.37	1	0.07	0	0.52
EF	1	0.70	1	1.29	0	0.40	0	0.62	0	0.30	1	0.67
EG	1	0.69	3	2.38	1	0.40	1	1.20	1	0.29	2	1.16
EH	0	0.58	0	1.34	0	0.32	0	0.60	1	0.25	1	0.75
EI	1	0.82	1	0.80	1	0.37	0	0.35	1	0.43	0	0.46

EK	0	0.19	1	1.01	0	0.14	0	0.46	0	0.06	0	0.55
EL	0	1.47	2	1.27	0	0.97	1	0.64	0	0.52	3	0.61
EM	1	0.41	0	0.34	0	0.29	1	0.14	0	0.13	0	0.20
EN	1	0.31	2	2.17	1	0.17	2	1.06	0	0.14	0	1.10
EP	1	0.69	3	1.73	0	0.43	1	0.78	0	0.26	2	0.96
EQ	0	0.18	1	0.83	0	0.12	1	0.44	0	0.06	0	0.38
ER	1	0.76	6	3.38	1	0.43	3	1.54	1	0.32	1	1.85
ES	2	1.38	0	3.64	1	0.83	1	1.87	0	0.55	1	1.74
ET	0	0.77	1	1.81	0	0.43	1	0.74	0	0.33	1	1.10
EV	0	1.15	2	1.08	0	0.68	0	0.46	0	0.47	1	0.64
EW	0	0.35	1	0.70	0	0.22	0	0.35	0	0.13	0	0.35
EY	2	0.65	1	1.16	0	0.42	0	0.51	1	0.24	0	0.67
FA	6	2.58	4	1.82	2	1.52	1	0.89	0	1.05	1	0.92
FC	3	2.87	2	1.75	3	1.30	0	0.97	4	1.60	0	0.78
FD	2	1.22	3	2.24	0	0.80	0	1.32	0	0.40	1	0.92
FE	0	0.70	2	1.29	1	0.40	0	0.62	0	0.30	1	0.67
FF	1	2.58	4	1.90	0	1.30	0	1.05	0	1.30	0	0.85
FG	1	2.54	5	3.49	1	1.30	3	2.02	1	1.25	0	1.48
FH	2	2.11	1	1.97	1	1.03	4	1.01	0	1.10	1	0.96
FI	3	3.01	0	1.18	0	1.21	0	0.58	2	1.85	0	0.59
FK	1	0.70	1	1.48	2	0.45	0	0.78	1	0.25	0	0.70
FL	4	5.40	0	1.86	5	3.13	2	1.09	1	2.25	1	0.78
FM	5	1.50	0	0.49	1	0.94	0	0.23	2	0.55	0	0.26
FN	1	1.13	3	3.19	0	0.54	0	1.79	1	0.60	3	1.41
FP	0	2.54	2	2.54	2	1.38	2	1.32	0	1.15	1	1.22
FQ	0	0.66	0	1.22	0	0.40	0	0.74	1	0.25	1	0.48
FR	1	2.77	6	4.98	1	1.38	3	2.60	2	1.40	3	2.37
FS	7	5.08	6	5.35	2	2.68	4	3.15	2	2.39	2	2.22
FT	5	2.82	1	2.66	0	1.38	3	1.24	0	1.45	3	1.41
FV	5	4.23	2	1.60	2	2.19	2	0.78	4	2.05	2	0.81
FW	0	1.27	1	1.03	1	0.71	1	0.58	0	0.55	0	0.44
FY	2	2.40	2	1.71	3	1.34	0	0.85	0	1.05	0	0.85
GA	3	2.54	7	3.35	1	1.52	2	1.72	2	1.01	2	1.61
GC	3	2.81	3	3.21	0	1.30	5	1.87	3	1.54	1	1.35
GD	1	1.20	7	4.12	4	0.80	4	2.54	0	0.38	5	1.61
GE	2	0.69	3	2.38	0	0.40	0	1.20	0	0.29	1	1.16
GF	4	2.54	2	3.49	3	1.30	1	2.02	0	1.25	1	1.48
GG	1	2.49	3	6.43	0	1.30	6	3.89	2	1.20	3	2.57

GH	3	2.08	3	3.63	0	1.03	0	1.95	0	1.06	1	1.67
GI	4	2.95	2	2.17	0	1.21	2	1.12	1	1.77	2	1.03
GK	1	0.69	3	2.73	1	0.45	1	1.50	0	0.24	0	1.22
GL	6	5.31	3	3.42	5	3.13	1	2.10	3	2.16	0	1.35
GM	0	1.48	0	0.91	0	0.94	1	0.45	1	0.53	0	0.45
GN	0	1.11	6	5.87	0	0.54	2	3.44	0	0.58	2	2.44
GP	1	2.49	6	4.68	1	1.38	2	2.54	1	1.10	3	2.12
GQ	0	0.65	1	2.24	0	0.40	2	1.42	0	0.24	1	0.84
GR	0	2.72	7	9.15	1	1.38	4	5.01	0	1.34	3	4.12
GS	3	4.98	14	9.85	2	2.68	5	6.06	3	2.30	3	3.86
GT	4	2.77	5	4.89	5	1.38	3	2.39	2	1.39	1	2.44
GV	4	4.15	3	2.93	0	2.19	0	1.50	4	1.97	2	1.42
GW	2	1.25	2	1.89	0	0.71	1	1.12	0	0.53	0	0.77
GY	2	2.35	2	3.14	0	1.34	3	1.65	1	1.01	1	1.48
HA	4	2.11	2	1.90	3	1.20	0	0.86	1	0.89	2	1.05
HC	4	2.35	2	1.82	3	1.03	1	0.94	2	1.35	0	0.88
HD	0	1.00	0	2.33	0	0.64	2	1.27	0	0.34	1	1.05
HE	0	0.58	1	1.34	0	0.32	1	0.60	1	0.25	1	0.75
HF	3	2.11	2	1.97	0	1.03	1	1.01	0	1.10	2	0.96
HG	0	2.08	6	3.63	0	1.03	0	1.95	0	1.06	1	1.67
HH	1	1.73	1	2.05	0	0.81	0	0.97	1	0.93	1	1.09
HI	3	2.46	2	1.22	0	0.96	0	0.56	1	1.56	0	0.67
HK	0	0.58	1	1.54	0	0.35	0	0.75	0	0.21	0	0.79
HL	4	4.42	1	1.94	4	2.48	1	1.05	2	1.90	0	0.88
HM	0	1.23	0	0.51	0	0.74	0	0.22	1	0.46	2	0.29
HN	0	0.92	4	3.32	1	0.43	2	1.72	0	0.51	3	1.59
HP	3	2.08	2	2.65	1	1.10	2	1.27	0	0.97	1	1.38
HQ	2	0.54	2	1.26	1	0.32	1	0.71	1	0.21	1	0.54
HR	2	2.27	10	5.17	1	1.10	6	2.51	1	1.18	3	2.68
HS	3	4.15	3	5.57	1	2.13	3	3.03	1	2.03	4	2.51
HT	5	2.31	2	2.76	2	1.10	0	1.20	3	1.22	1	1.59
HV	4	3.46	1	1.66	0	1.74	2	0.75	2	1.73	0	0.92
HW	4	1.04	0	1.07	1	0.57	2	0.56	1	0.46	0	0.50
HY	2	1.96	2	1.78	2	1.06	1	0.82	2	0.89	1	0.96
IA	5	3.01	1	1.13	2	1.41	0	0.50	3	1.49	0	0.64
IC	2	3.34	0	1.08	2	1.21	0	0.54	4	2.27	0	0.54
ID	1	1.42	2	1.39	0	0.75	0	0.73	0	0.57	1	0.64
IE	0	0.82	0	0.80	0	0.37	1	0.35	0	0.43	1	0.46

IF	3	3.01	0	1.18	0	1.21	0	0.58	4	1.85	0	0.59
IG	3	2.95	4	2.17	0	1.21	0	1.12	0	1.77	0	1.03
IH	2	2.46	1	1.22	1	0.96	0	0.56	2	1.56	1	0.67
II	3	3.50	1	0.73	1	1.12	1	0.32	2	2.63	1	0.41
IK	1	0.82	2	0.92	0	0.42	2	0.43	0	0.35	0	0.49
IL	10	6.29	1	1.15	2	2.91	0	0.60	2	3.19	1	0.54
IM	3	1.75	0	0.31	3	0.87	0	0.13	2	0.78	2	0.18
IN	1	1.31	0	1.98	2	0.50	1	0.99	1	0.85	1	0.98
IP	1	2.95	3	1.58	0	1.29	1	0.73	3	1.63	1	0.85
IQ	0	0.77	1	0.75	1	0.37	2	0.41	0	0.35	0	0.33
IR	4	3.23	3	3.08	2	1.29	2	1.45	1	1.99	1	1.65
IS	3	5.91	3	3.32	3	2.50	0	1.75	2	3.41	3	1.54
IT	5	3.28	2	1.65	0	1.29	0	0.69	2	2.06	0	0.98
IV	3	4.92	1	0.99	3	2.04	2	0.43	1	2.91	2	0.57
IW	3	1.48	0	0.64	2	0.67	1	0.32	0	0.78	0	0.31
IY	4	2.79	1	1.06	1	1.25	0	0.47	2	1.49	0	0.59
KA	0	0.70	0	1.42	0	0.52	2	0.66	0	0.20	2	0.76
KC	0	0.78	1	1.36	0	0.45	2	0.72	0	0.31	0	0.64
KD	1	0.33	2	1.75	1	0.28	0	0.98	0	0.08	1	0.76
KE	1	0.19	1	1.01	1	0.14	0	0.46	0	0.06	1	0.55
KF	2	0.70	0	1.48	1	0.45	1	0.78	0	0.25	1	0.70
KG	1	0.69	2	2.73	0	0.45	1	1.50	0	0.24	2	1.22
KH	0	0.58	0	1.54	0	0.35	1	0.75	0	0.21	0	0.79
KI	1	0.82	0	0.92	0	0.42	1	0.43	0	0.35	1	0.49
KK	0	0.19	3	1.16	0	0.15	2	0.58	0	0.05	0	0.58
KL	2	1.47	3	1.45	0	1.08	3	0.81	0	0.43	1	0.64
KM	1	0.41	2	0.39	0	0.32	0	0.17	0	0.11	0	0.21
KN	0	0.31	2	2.49	0	0.18	0	1.32	0	0.12	0	1.16
KP	0	0.69	2	1.98	0	0.48	1	0.98	0	0.22	0	1.01
KQ	0	0.18	1	0.95	0	0.14	1	0.55	0	0.05	0	0.40
KR	0	0.76	2	3.88	1	0.48	2	1.93	0	0.27	3	1.96
KS	1	1.38	8	4.18	1	0.92	1	2.33	1	0.46	2	1.83
KT	0	0.77	1	2.07	0	0.48	0	0.92	0	0.28	1	1.16
KV	4	1.15	1	1.24	1	0.75	0	0.58	2	0.39	0	0.67
KW	0	0.35	2	0.80	0	0.25	0	0.43	0	0.11	0	0.37
KY	1	0.65	1	1.33	0	0.46	1	0.63	0	0.20	0	0.70
LA	6	5.40	3	1.79	1	3.67	1	0.93	2	1.81	1	0.84
LC	2	5.99	1	1.71	2	3.13	2	1.01	1	2.76	1	0.71

LD	3	2.55	3	2.20	0	1.94	0	1.37	0	0.69	1	0.84
LE	1	1.47	1	1.27	2	0.97	2	0.64	0	0.52	1	0.61
LF	5	5.40	1	1.86	1	3.13	2	1.09	1	2.25	0	0.78
LG	7	5.31	5	3.42	3	3.13	4	2.10	2	2.16	2	1.35
LH	2	4.42	4	1.94	1	2.48	1	1.05	3	1.90	1	0.88
LI	8	6.29	1	1.15	3	2.91	1	0.60	6	3.19	0	0.54
LK	0	1.47	2	1.45	0	1.08	0	0.81	0	0.43	0	0.64
LL	17	11.30	2	1.82	10	7.55	0	1.13	4	3.89	1	0.71
LM	3	3.14	1	0.48	3	2.26	0	0.24	0	0.95	1	0.24
LN	0	2.36	2	3.13	1	1.29	1	1.85	0	1.04	2	1.28
LP	8	5.31	1	2.49	6	3.34	2	1.37	3	1.99	0	1.11
LQ	1	1.38	1	1.19	0	0.97	0	0.77	2	0.43	0	0.44
LR	8	5.80	2	4.88	6	3.34	0	2.70	4	2.42	3	2.16
LS	13	10.61	4	5.25	8	6.47	1	3.26	6	4.14	1	2.03
LT	3	5.90	3	2.61	2	3.34	1	1.29	2	2.50	2	1.28
LV	5	8.84	3	1.56	9	5.28	2	0.81	2	3.54	1	0.74
LW	2	2.65	0	1.00	3	1.73	0	0.60	2	0.95	0	0.41
LY	2	5.01	2	1.67	1	3.24	1	0.89	1	1.81	2	0.78
MA	2	1.50	0	0.47	1	1.10	0	0.20	1	0.44	0	0.28
MC	3	1.67	0	0.45	1	0.94	0	0.22	0	0.68	0	0.24
MD	2	0.71	0	0.58	1	0.58	0	0.29	0	0.17	0	0.28
ME	1	0.41	0	0.34	0	0.29	0	0.14	0	0.13	0	0.20
MF	2	1.50	0	0.49	1	0.94	0	0.23	2	0.55	0	0.26
MG	0	1.48	2	0.91	0	0.94	0	0.45	2	0.53	1	0.45
MH	2	1.23	0	0.51	1	0.74	0	0.22	1	0.46	0	0.29
MI	3	1.75	0	0.31	0	0.87	0	0.13	1	0.78	0	0.18
MK	1	0.41	0	0.39	0	0.32	0	0.17	0	0.11	2	0.21
ML	3	3.14	0	0.48	0	2.26	0	0.24	1	0.95	0	0.24
MM	0	0.88	0	0.13	3	0.68	0	0.05	0	0.23	0	0.08
MN	0	0.66	0	0.83	0	0.39	0	0.40	0	0.25	1	0.43
MP	0	1.48	1	0.66	1	1.00	1	0.29	0	0.49	1	0.37
MQ	1	0.38	1	0.32	0	0.29	0	0.16	0	0.11	0	0.15
MR	1	1.61	2	1.29	2	1.00	0	0.58	1	0.59	0	0.72
MS	2	2.95	2	1.39	0	1.94	2	0.70	0	1.01	2	0.68
MT	1	1.64	1	0.69	1	1.00	1	0.28	1	0.61	0	0.43
MV	3	2.46	0	0.41	1	1.59	1	0.17	1	0.87	0	0.25
MW	0	0.74	1	0.27	1	0.52	0	0.13	0	0.23	0	0.14
MY	1	1.39	1	0.44	2	0.97	1	0.19	0	0.44	0	0.26

NA	1	1.13	1	3.06	1	0.63	1	1.52	0	0.48	0	1.53
NC	1	1.25	3	2.93	0	0.54	1	1.66	1	0.74	1	1.28
ND	0	0.53	4	3.76	1	0.33	3	2.25	0	0.18	1	1.53
NE	0	0.31	1	2.17	1	0.17	0	1.06	0	0.14	0	1.10
NF	1	1.13	3	3.19	0	0.54	1	1.79	0	0.60	1	1.41
NG	2	1.11	4	5.87	1	0.54	1	3.44	0	0.58	3	2.44
NH	0	0.92	4	3.32	0	0.43	2	1.72	0	0.51	4	1.59
NI	2	1.31	1	1.98	1	0.50	1	0.99	1	0.85	0	0.98
NK	0	0.31	1	2.49	0	0.18	2	1.32	0	0.12	1	1.16
NL	1	2.36	9	3.13	2	1.29	3	1.85	4	1.04	3	1.28
NM	3	0.66	1	0.83	2	0.39	0	0.40	0	0.25	0	0.43
NN	0	0.49	5	5.36	0	0.22	4	3.05	0	0.28	3	2.32
NP	2	1.11	4	4.27	0	0.57	0	2.25	0	0.53	3	2.02
NQ	1	0.29	1	2.04	0	0.17	0	1.26	0	0.12	0	0.79
NR	1	1.21	9	8.36	2	0.57	6	4.44	2	0.64	5	3.91
NS	0	2.21	11	9.00	0	1.11	5	5.36	1	1.11	6	3.67
NT	0	1.23	4	4.47	0	0.57	3	2.12	1	0.67	4	2.32
NV	2	1.85	2	2.68	1	0.91	0	1.32	0	0.94	1	1.34
NW	1	0.55	1	1.72	0	0.30	1	0.99	0	0.25	0	0.73
NY	1	1.05	8	2.87	0	0.55	3	1.46	0	0.48	0	1.41
PA	1	2.54	1	2.44	1	1.62	1	1.13	2	0.93	1	1.33
PC	2	2.81	4	2.34	2	1.38	0	1.22	2	1.41	0	1.11
PD	2	1.20	3	3.00	1	0.86	2	1.66	0	0.35	2	1.33
PE	2	0.69	3	1.73	1	0.43	1	0.78	1	0.26	1	0.96
PF	2	2.54	2	2.54	3	1.38	1	1.32	0	1.15	1	1.22
PG	3	2.49	3	4.68	1	1.38	1	2.54	0	1.10	3	2.12
PH	2	2.08	3	2.65	4	1.10	0	1.27	2	0.97	2	1.38
PI	3	2.95	1	1.58	2	1.29	1	0.73	1	1.63	0	0.85
PK	1	0.69	3	1.98	1	0.48	0	0.98	0	0.22	1	1.01
PL	4	5.31	2	2.49	1	3.34	0	1.37	6	1.99	0	1.11
PM	2	1.48	1	0.66	0	1.00	0	0.29	1	0.49	0	0.37
PN	0	1.11	5	4.27	1	0.57	6	2.25	1	0.53	1	2.02
PP	5	2.49	2	3.41	0	1.48	2	1.66	1	1.02	0	1.75
PQ	1	0.65	2	1.63	0	0.43	1	0.93	0	0.22	0	0.69
PR	3	2.72	10	6.67	1	1.48	2	3.28	1	1.24	6	3.40
PS	3	4.98	5	7.18	4	2.87	6	3.96	0	2.12	3	3.18
PT	2	2.77	1	3.56	2	1.48	3	1.57	0	1.28	6	2.02
PV	9	4.15	1	2.14	1	2.34	0	0.98	1	1.81	1	1.17

PW	0	1.25	4	1.37	1	0.76	0	0.73	0	0.49	1	0.64
PY	2	2.35	3	2.29	1	1.43	1	1.08	3	0.93	2	1.22
QA	1	0.66	0	1.17	0	0.47	0	0.63	1	0.20	0	0.52
QC	2	0.73	1	1.12	0	0.40	1	0.68	1	0.31	0	0.44
QD	1	0.31	2	1.43	0	0.25	2	0.93	0	0.08	0	0.52
QE	1	0.18	4	0.83	0	0.12	1	0.44	0	0.06	1	0.38
QF	0	0.66	3	1.22	0	0.40	0	0.74	1	0.25	1	0.48
QG	0	0.65	2	2.24	1	0.40	1	1.42	1	0.24	1	0.84
QH	0	0.54	0	1.26	0	0.32	1	0.71	0	0.21	1	0.54
QI	1	0.77	0	0.75	1	0.37	0	0.41	0	0.35	2	0.33
QK	0	0.18	0	0.95	0	0.14	0	0.55	0	0.05	1	0.40
QL	1	1.38	0	1.19	0	0.97	1	0.77	0	0.43	0	0.44
QM	0	0.38	0	0.32	0	0.29	0	0.16	0	0.11	0	0.15
QN	0	0.29	2	2.04	0	0.17	1	1.26	0	0.12	0	0.79
QP	1	0.65	1	1.63	2	0.43	2	0.93	0	0.22	1	0.69
QQ	0	0.17	1	0.78	0	0.12	0	0.52	0	0.05	0	0.27
QR	0	0.71	2	3.18	1	0.43	4	1.83	1	0.27	1	1.34
QS	1	1.29	5	3.43	2	0.83	2	2.21	0	0.46	3	1.25
QT	1	0.72	3	1.70	0	0.43	0	0.88	0	0.28	0	0.79
QV	3	1.08	0	1.02	1	0.68	0	0.55	0	0.39	0	0.46
QW	0	0.32	2	0.66	0	0.22	1	0.41	0	0.11	0	0.25
QY	0	0.61	0	1.09	0	0.42	2	0.60	0	0.20	0	0.48
RA	4	2.77	4	4.78	2	1.62	3	2.22	3	1.13	2	2.57
RC	3	3.08	3	4.58	2	1.38	5	2.41	5	1.72	5	2.16
RD	0	1.31	5	5.87	0	0.86	8	3.28	1	0.43	1	2.57
RE	1	0.76	5	3.38	0	0.43	0	1.54	0	0.32	2	1.85
RF	4	2.77	6	4.98	4	1.38	3	2.60	1	1.40	5	2.37
RG	0	2.72	12	9.15	0	1.38	4	5.01	0	1.34	3	4.12
RH	3	2.27	3	5.17	2	1.10	5	2.51	2	1.18	4	2.68
RI	5	3.23	4	3.08	1	1.29	0	1.45	2	1.99	1	1.65
RK	1	0.76	3	3.88	1	0.48	0	1.93	1	0.27	3	1.96
RL	5	5.80	4	4.88	2	3.34	6	2.70	2	2.42	1	2.16
RM	5	1.61	2	1.29	1	1.00	1	0.58	1	0.59	1	0.72
RN	5	1.21	8	8.36	1	0.57	4	4.44	0	0.64	5	3.91
RP	2	2.72	9	6.67	3	1.48	5	3.28	1	1.24	3	3.40
RQ	0	0.71	3	3.18	1	0.43	1	1.83	0	0.27	2	1.34
RR	1	2.97	10	13.03	1	1.48	3	6.46	1	1.50	5	6.59
RS	6	5.44	9	14.03	3	2.87	9	7.81	3	2.58	4	6.18

RT	3	3.02	9	6.97	0	1.48	4	3.09	0	1.56	4	3.91
RV	3	4.54	6	4.18	2	2.34	1	1.93	2	2.20	2	2.26
RW	1	1.36	3	2.69	0	0.76	0	1.45	2	0.59	1	1.24
RY	2	2.57	4	4.48	2	1.43	0	2.12	0	1.13	5	2.37
SA	3	5.08	5	5.14	5	3.14	5	2.68	1	1.93	2	2.41
SC	6	5.63	10	4.93	3	2.68	1	2.91	3	2.95	3	2.03
SD	0	2.40	5	6.32	1	1.66	6	3.96	3	0.74	1	2.41
SE	2	1.38	3	3.64	0	0.83	1	1.87	0	0.55	1	1.74
SF	5	5.08	6	5.35	3	2.68	1	3.15	6	2.39	1	2.22
SG	10	4.98	8	9.85	1	2.68	9	6.06	1	2.30	3	3.86
SH	10	4.15	11	5.57	1	2.13	3	3.03	0	2.03	2	2.51
SI	7	5.91	3	3.32	3	2.50	0	1.75	3	3.41	1	1.54
SK	1	1.38	2	4.18	1	0.92	2	2.33	1	0.46	1	1.83
SL	10	10.61	5	5.25	7	6.47	1	3.26	7	4.14	3	2.03
SM	1	2.95	1	1.39	1	1.94	1	0.70	0	1.01	0	0.68
SN	3	2.21	11	9.00	2	1.11	7	5.36	1	1.11	5	3.67
SP	4	4.98	9	7.18	0	2.87	6	3.96	3	2.12	4	3.18
SQ	3	1.29	4	3.43	1	0.83	4	2.21	0	0.46	4	1.25
SR	4	5.44	11	14.03	0	2.87	6	7.81	3	2.58	6	6.18
SS	13	9.97	11	15.10	6	5.55	9	9.44	3	4.42	3	5.79
ST	4	5.54	12	7.50	4	2.87	5	3.73	2	2.67	1	3.67
SV	6	8.31	4	4.50	8	4.53	3	2.33	2	3.78	4	2.12
SW	3	2.49	1	2.89	1	1.48	1	1.75	2	1.01	2	1.16
SY	3	4.71	4	4.82	1	2.77	4	2.56	1	1.93	4	2.22
TA	3	2.82	2	2.55	2	1.62	0	1.06	0	1.17	0	1.53
TC	5	3.13	5	2.45	0	1.38	1	1.15	0	1.78	2	1.28
TD	0	1.33	1	3.14	2	0.86	1	1.57	1	0.45	1	1.53
TE	1	0.77	2	1.81	0	0.43	1	0.74	0	0.33	1	1.10
TF	0	2.82	4	2.66	2	1.38	1	1.24	0	1.45	2	1.41
TG	6	2.77	4	4.89	2	1.38	4	2.39	4	1.39	2	2.44
TH	2	2.31	3	2.76	1	1.10	0	1.20	0	1.22	2	1.59
TI	3	3.28	2	1.65	1	1.29	2	0.69	6	2.06	1	0.98
TK	0	0.77	5	2.07	0	0.48	3	0.92	0	0.28	2	1.16
TL	7	5.90	1	2.61	5	3.34	1	1.29	2	2.50	1	1.28
TM	0	1.64	0	0.69	0	1.00	0	0.28	0	0.61	0	0.43
TN	3	1.23	4	4.47	1	0.57	2	2.12	0	0.67	2	2.32
TP	0	2.77	5	3.56	3	1.48	1	1.57	3	1.28	4	2.02
TQ	0	0.72	3	1.70	1	0.43	1	0.88	0	0.28	2	0.79

TR	1	3.02	11	6.97	1	1.48	2	3.09	1	1.56	3	3.91
TS	3	5.54	5	7.50	2	2.87	5	3.73	3	2.67	1	3.67
TT	6	3.08	2	3.72	0	1.48	0	1.47	1	1.61	1	2.32
TV	8	4.61	2	2.23	0	2.34	0	0.92	3	2.28	1	1.34
TW	0	1.38	1	1.44	1	0.76	1	0.69	0	0.61	2	0.73
TY	3	2.61	1	2.39	2	1.43	0	1.01	1	1.17	1	1.41
VA	2	4.23	1	1.53	3	2.57	0	0.66	0	1.65	1	0.88
VC	8	4.69	2	1.47	3	2.19	1	0.72	1	2.52	0	0.74
VD	5	2.00	0	1.88	3	1.36	0	0.98	1	0.63	1	0.88
VE	2	1.15	1	1.08	0	0.68	1	0.46	1	0.47	0	0.64
VF	3	4.23	2	1.60	3	2.19	1	0.78	1	2.05	0	0.81
VG	2	4.15	4	2.93	5	2.19	1	1.50	3	1.97	1	1.42
VH	6	3.46	2	1.66	2	1.74	0	0.75	2	1.73	1	0.92
VI	2	4.92	3	0.99	2	2.04	2	0.43	0	2.91	1	0.57
VK	2	1.15	1	1.24	1	0.75	1	0.58	0	0.39	0	0.67
VL	9	8.84	1	1.56	3	5.28	0	0.81	2	3.54	2	0.74
VM	2	2.46	0	0.41	2	1.59	0	0.17	1	0.87	0	0.25
VN	1	1.85	3	2.68	1	0.91	2	1.32	1	0.94	2	1.34
VP	3	4.15	0	2.14	2	2.34	2	0.98	2	1.81	0	1.17
VQ	0	1.08	2	1.02	1	0.68	0	0.55	1	0.39	1	0.46
VR	6	4.54	4	4.18	2	2.34	2	1.93	2	2.20	2	2.26
VS	4	8.31	5	4.50	5	4.53	2	2.33	8	3.78	2	2.12
VT	4	4.61	3	2.23	3	2.34	0	0.92	2	2.28	2	1.34
VV	7	6.92	0	1.34	2	3.70	0	0.58	4	3.23	0	0.78
VW	3	2.08	0	0.86	1	1.21	0	0.43	0	0.87	1	0.42
VY	5	3.92	2	1.44	2	2.26	1	0.63	3	1.65	0	0.81
WA	1	1.27	0	0.98	0	0.84	0	0.50	0	0.44	0	0.48
WC	0	1.41	0	0.94	1	0.71	0	0.54	0	0.68	0	0.41
WD	1	0.60	2	1.21	0	0.44	1	0.73	0	0.17	1	0.48
WE	0	0.35	0	0.70	0	0.22	0	0.35	0	0.13	1	0.35
WF	1	1.27	2	1.03	0	0.71	2	0.58	2	0.55	0	0.44
WG	0	1.25	0	1.89	3	0.71	0	1.12	0	0.53	1	0.77
WH	1	1.04	1	1.07	2	0.57	0	0.56	0	0.46	0	0.50
WI	2	1.48	0	0.64	1	0.67	0	0.32	0	0.78	0	0.31
WK	0	0.35	0	0.80	0	0.25	0	0.43	0	0.11	1	0.37
WL	1	2.65	1	1.00	0	1.73	0	0.60	2	0.95	0	0.41
WM	0	0.74	2	0.27	0	0.52	0	0.13	0	0.23	0	0.14
WN	2	0.55	2	1.72	0	0.30	3	0.99	1	0.25	0	0.73

WP	2	1.25	1	1.37	0	0.76	0	0.73	0	0.49	2	0.64
WQ	0	0.32	2	0.66	1	0.22	2	0.41	0	0.11	0	0.25
WR	2	1.36	5	2.69	1	0.76	2	1.45	1	0.59	1	1.24
WS	4	2.49	4	2.89	2	1.48	2	1.75	3	1.01	1	1.16
WT	3	1.38	0	1.44	0	0.76	0	0.69	0	0.61	2	0.73
WV	1	2.08	4	0.86	3	1.21	0	0.43	1	0.87	0	0.42
WW	0	0.62	0	0.55	0	0.39	0	0.32	0	0.23	0	0.23
WY	2	1.18	1	0.92	1	0.74	0	0.47	0	0.44	1	0.44
YA	1	2.40	2	1.64	1	1.57	0	0.73	2	0.85	1	0.92
YC	3	2.66	0	1.57	1	1.34	0	0.79	1	1.29	0	0.78
YD	2	1.13	3	2.02	0	0.83	1	1.08	1	0.32	1	0.92
YE	0	0.65	2	1.16	1	0.42	1	0.51	0	0.24	0	0.67
YF	6	2.40	0	1.71	2	1.34	3	0.85	1	1.05	0	0.85
YG	1	2.35	5	3.14	0	1.34	1	1.65	3	1.01	3	1.48
YH	2	1.96	1	1.78	1	1.06	0	0.82	2	0.89	0	0.96
YI	2	2.79	2	1.06	0	1.25	0	0.47	1	1.49	1	0.59
YK	0	0.65	1	1.33	1	0.46	1	0.63	0	0.20	4	0.70
YL	4	5.01	2	1.67	2	3.24	1	0.89	1	1.81	0	0.78
YM	2	1.39	0	0.44	1	0.97	1	0.19	1	0.44	1	0.26
YN	0	1.05	5	2.87	0	0.55	0	1.46	0	0.48	1	1.41
YP	5	2.35	1	2.29	1	1.43	1	1.08	0	0.93	1	1.22
YQ	0	0.61	0	1.09	0	0.42	0	0.60	0	0.20	0	0.48
YR	0	2.57	6	4.48	1	1.43	3	2.12	0	1.13	2	2.37
YS	6	4.71	9	4.82	3	2.77	3	2.56	1	1.93	1	2.22
YT	3	2.61	0	2.39	3	1.43	0	1.01	2	1.17	2	1.41
YV	4	3.92	0	1.44	4	2.26	1	0.63	1	1.65	1	0.81
YW	2	1.18	0	0.92	1	0.74	1	0.47	2	0.44	1	0.44
YY	3	2.22	1	1.54	1	1.39	0	0.70	0	0.85	0	0.85

In an attempt to address the potential role of primary sequence elements in the degradation of the PrLDs, we expanded our original sequence library dataset. Using simple computational approaches, we observe strong correlations between the primary sequence features observed in our experimentally-obtained sequence libraries and those that would be expected by random chance in sequence libraries of equivalent size. While this does not conclusively rule out the possibility that a small number of specific primary sequence features may actually contribute to PrLD degradation or stability, these results suggest that amino acid composition is at least the predominant determinant of degradation susceptibility, rather than primary sequence features.

REFERENCES

1. Fredrickson EK, Gallagher PS, Clowes Candadai SV, Gardner RG. Substrate recognition in nuclear protein quality control degradation is governed by exposed hydrophobicity that correlates with aggregation and insolubility. The Journal of biological chemistry. 2013;288(9):6130-9.