

De novo design of synthetic prion domains

James A. Toombs^{a,1}, Michelina Petri^a, Kacy R. Paul^a, Grace Y. Kan^a, Asa Ben-Hur^b, and Eric D. Ross^{a,2}

Departments of ^aBiochemistry and Molecular Biology and ^bComputer Science, Colorado State University, Fort Collins, CO 80523

Edited by Reed B. Wickner, National Institutes of Health, Bethesda, MD, and approved March 2, 2012 (received for review November 28, 2011)

Prions are important disease agents and epigenetic regulatory elements. Prion formation involves the structural conversion of proteins from a soluble form into an insoluble amyloid form. In many cases, this structural conversion is driven by a glutamine/asparagine (Q/N)-rich prion-forming domain. However, our understanding of the sequence requirements for prion formation and propagation by Q/N-rich domains has been insufficient for accurate prion propensity prediction or prion domain design. By focusing exclusively on amino acid composition, we have developed a prion aggregation prediction algorithm (PAPA), specifically designed to predict prion propensity of Q/N-rich proteins. Here, we show not only that this algorithm is far more effective than traditional amyloid prediction algorithms at predicting prion propensity of Q/N-rich proteins, but remarkably, also that PAPA is capable of rationally designing protein domains that function as prions in vivo.

Sup35 | yeast | bioinformatics

Priions result from proteins that are capable of converting from a soluble, often intrinsically disordered native state into an infectious aggregated amyloid form. Amyloid fibrils are β -sheet-rich protein aggregates associated with numerous human diseases, including Alzheimer's disease and type II diabetes. Amyloid fibrils can also serve beneficial functions, acting as structural scaffolds or protein-only elements of inheritance (1). Therefore, amyloid-based prions could potentially be used for synthetic biology applications, allowing for the construction of posttranslational epigenetic regulatory elements. However, because developing methods to design stable protein folds has been a long-standing challenge, designing proteins capable of adopting two distinct stable states would seem to provide a near-impossible challenge.

Seven amyloid-based prions have been identified in yeast (2). In each case, an intrinsically disordered glutamine/asparagine (Q/N)-rich prion-forming domain (PFD) drives prion formation. Q/N-rich amyloid-like aggregates have been proposed to be involved in various diseases (3, 4), as well as to act as structural (5) and regulatory elements (1). The genetic tractability of yeast makes the yeast prions a powerful model system to explore the sequence requirements for amyloid formation by Q/N-rich domains and to test de novo designed PFDs.

Scrambling studies indicate that prion formation by Q/N-rich proteins is driven largely by amino acid composition, with only modest effects of primary sequence (6, 7). Therefore, we recently developed an in vivo method to score the prion propensity of each amino acid in the context of a Q/N-rich PFD. Combining these experimentally determined values with the disorder prediction algorithm FoldIndex (8), we developed an entirely composition-based prion prediction algorithm (9), which we now call the prion aggregation prediction algorithm (PAPA). We tested PAPA on a dataset generated by Alberti et al. (10), in which they identified the 100 proteins with greatest compositional similarity to known yeast prions and scored each domain for prion-like activity in four different assays. There was a strong correlation between their observed aggregation activity and prion propensity predicted by PAPA (9). This correlation was a unique demonstration of an algorithm that could effectively distinguish Q/N-rich proteins with and without prion activity.

Numerous other aggregation prediction algorithms have also been developed, including TANGO (11), Zyggregator (12), ZipperDB (13), and Waltz (14). However, none of these has yet shown the ability to distinguish between Q/N-rich proteins with and without prion activity (15). There are a few possible explanations for this failure. First, it is possible that some of these algorithms

might be effective for Q/N-rich domains, but just have not been thoroughly tested. For example, the ability of Waltz to identify predicted amyloid-prone hexapeptides within the PFD of the yeast prion protein Sup35 was used to argue for the utility of Waltz (14); however, this ability does not address whether the presence of these hexapeptides is predictive of a domain's amyloid propensity, as the control of examining whether such sequences are also found in non-amyloid-forming Q/N-rich sequences has not been reported. Second, some of these algorithms broadly predict aggregation propensity, but do not specifically predict amyloid formation propensity (14). Finally, our previous work (9) suggested that there may be two distinct classes of amyloid-forming proteins: those for which amyloid formation is driven by short, highly amyloidogenic, often hydrophobic, stretches (16) and those for which amyloid formation is driven by many weaker interactions across a large, intrinsically disordered domain (9). We hypothesized that whereas most prediction algorithms are designed for proteins in the first class, the Q/N-rich yeast PFDs fall into the second class (9). Consistent with the idea that there are two classes of amyloid proteins, PAPA accurately predicts prion propensity of Q/N-rich proteins, but does not accurately predict non-Q/N-rich prion proteins like Het-s and PrP (9). Here, we distinguish among these explanations by testing the ability of various algorithms to predict amyloid and prion formation by Q/N-rich proteins. We find that none of these other algorithms is as effective as PAPA at distinguishing between Q/N-rich domains with and without prion activity.

We then asked whether PAPA would be sufficient for de novo PFD design. Because the 100 proteins tested by Alberti et al. were all selected for their compositional similarity to known prions, they represent a relatively homogeneous pool. Therefore, the extent to which PAPA's predictive abilities would be limited to highly preselected proteins was unclear. Here, we have adapted PAPA for the more challenging task of de novo design of synthetic PFDs (sPFDs). Characterization of two sPFDs designed by this method reveals that these domains function in vivo similarly to naturally occurring PFDs. These results demonstrate the substantial progress that we have made in defining the sequence basis for prion formation.

Results

Predicting Aggregation Propensity of Q/N-Rich Proteins. Various strategies have been proposed for predicting aggregation propensity. We used the dataset of Alberti et al. (10) to test which of these strategies were effective for Q/N-rich proteins. Of the 100 Q/N-rich proteins tested by Alberti et al., 18 showed prion-like activity in all four assays, whereas 18 did not show prion-like activity in any assay. This result provides a useful dataset, with clear examples of non-prion-like and prion-like sequences. Importantly, the four assays include tests of both amyloid- and prion-forming ability; therefore, domains that fail all four tests not only fail to act as prions, but also show no detectable

Author contributions: J.A.T. and E.D.R. designed research; J.A.T., M.P., K.R.P., and G.Y.K. performed research; A.B.-H. contributed new reagents/analytic tools; J.A.T., M.P., K.R.P., G.Y.K., and E.D.R. analyzed data; and J.A.T. and E.D.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹Present address: Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114.

²To whom correspondence should be addressed. E-mail: Eric.Ross@colostate.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1119366109/-DCSupplemental.

Waltz identifies amyloid-prone hexapeptide segments on the basis of a scoring matrix. Using the “best overall performance” setting, Waltz identified amyloid-prone segments in 89% of the prion-like proteins and half of the nonprion proteins (Fig. 1C), clearly indicating that the presence of such segments is not sufficient for amyloid formation. Although on average there were more residues predicted to be in amyloid-forming segments among the prion proteins, no clear separation was seen between the two sets. Using the “high specificity” setting (which reduces false positives) did not significantly improve the results; 22% of the nonprions still scored positive, whereas only 55% of the prion-like proteins scored positive (Fig. S1D). Indeed, 4 of the 14 highest-scoring segments were in nonprion domains.

Other algorithms focus on the overall physical properties of peptide segments rather than specific primary sequence features. Chiti et al. demonstrated that a protein’s aggregation propensity could be predicted largely on the basis of its physico-chemical properties, such as hydrophobicity and charge (19). TANGO and Zyggregator use such properties to predict aggregation propensity (11, 12). Linding et al. previously reported that TANGO does not identify β -aggregation nuclei in the PFDs of the yeast prion proteins Sup35 and Ure2 (20). It likewise fails to detect β -aggregation nuclei in the PFDs of the Cyc8, Mot3, and Rnq1 prion proteins so is clearly not suited for prediction of Q/N-rich proteins. By contrast, for Zyggregator, there was a significant difference between average “optimal scores” for prion and nonprion sequences ($P = 2 \times 10^{-5}$ by ANOVA); however, there was substantial overlap between the prion and nonprion sets (Fig. 1D).

Rather than just focusing on the β -sheet propensity of individual stretches, PASTA (Prediction of Amyloid Structure Aggregation) incorporates pairwise interactions between neighboring β -strands (21). Trovato et al. previously reported that a PASTA energy prediction of < -4.0 was indicative of cross- β fibrillar aggregates (21). Such segments were found in 50% of prion-like proteins vs. only 22% of the nonprion proteins (Fig. 1E), and the PASTA score for the segment predicted to be most amyloid prone in each protein was much lower on average in the prion set than in the nonprion set (-4.64 vs. -2.92 ; $P = 0.004$ by ANOVA). However, a low PASTA score was not necessary for amyloid formation; for example, the most amyloid-prone segment identified by PASTA in Rnq1 had a PASTA score of -2.40 , well above the threshold predicted to form amyloid.

Finally, a recent study suggested that coiled coils play a key role in amyloid formation and that sequences capable of forming coiled coils are overrepresented among amyloid-forming sequences (22); however, this study did not include analysis of compositionally similar nonamyloid domains to assess whether this overrepresentation was simply due to compositional biases. In fact, there was no difference in the frequency of predicted coiled coils between the prion and nonprion set using Paircoil2 (ref. 23 and Fig. 1F). Using Coils (24), such sequences were actually slightly less common among prion-like domains (Fig. S1 E–G). Indeed, in both sets, most sequences did not have any regions predicted to form coiled coils by either prediction method. Thus, although it remains possible that formation of coiled coils is involved in prion formation for specific proteins, there is no evidence of a correlation between predicted coiled coils and prion propensity.

Correlations Between Algorithms. PAPA uses a large window size and is entirely composition based, whereas Waltz and ZipperDB look for small primary sequence motifs. Given the substantial differences in the basis for these algorithms, it was surprising that each showed statistically significant differences between the prion and nonprion sets. We therefore hypothesized that prion propensity of Q/N-rich proteins might be driven by a combination of global composition and local primary sequence; if so, combining multiple algorithms might improve predictive accuracy. However, pairing PAPA with each of the other algorithms had little effect on predictive accuracy (Fig. S2 and Table S2). For each pair of algorithms, we performed an unbiased analysis to determine the linear function that allowed for optimal separation of prion and nonprion sequences; when PAPA was paired with each of the other algorithms, optimal predictive accuracy was achieved when

little or no weight was given to the other algorithm (Table S2). Furthermore, no other pair of algorithms could match the predictive accuracy of PAPA alone (Table S2).

Instead, we found strong correlations between many of the pairs of algorithms (Fig. S2). Some of these correlations were not surprising. For example, both PAPA and Zyggregator give high scores to hydrophobic residues and low scores to charged residues. Consequently, there is a strong correlation ($P < 10^{-6}$ by Spearman’s rank analysis) between PAPA and Zyggregator scores (Fig. S2B). Modest correlations were seen for many of the other pairs of algorithms (Fig. S2B).

Surprisingly, although ZipperDB and Waltz look for specific small primary sequence motifs whereas PAPA assesses only composition, there was also a statistically significant correlation between the scores with PAPA and those of both ZipperDB ($P = 0.01$) and Waltz ($P = 0.004$). We hypothesized that these correlations are due to the position-independent aspects of each algorithm. For example, many of the residues that score high (I, V, W, F) or low (K, R, D, P) in PAPA are generally similarly favored/disfavored by Waltz (14); the difference is that Waltz adds position-specific effects. We hypothesized that Waltz gives higher scores on average to prion sequences than nonprion sequences simply because of compositional differences between the two sets and that the position-specific aspects of Waltz do nothing to add to its predictive accuracy for Q/N-rich proteins. To test this, we scrambled each of the prion and nonprion sequences in silico and reanalyzed the scrambled sequences with Waltz. No significant change was seen in the number of Waltz-positive amyloid stretches in either the prion or the nonprion set after scrambling (Fig. 2A). In other words, making Waltz blind to the original primary sequence of a domain does nothing to reduce prediction accuracy. This result argues that the higher Waltz scores seen in the prion set reflect compositional differences between the sets and not that the

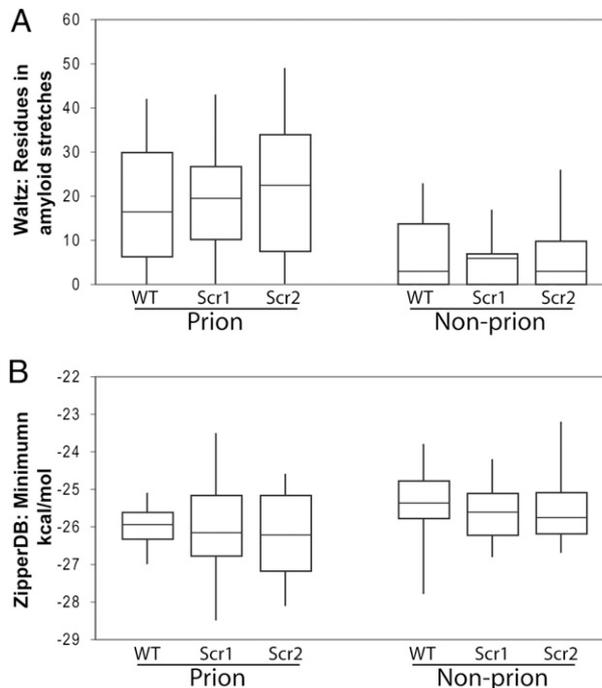


Fig. 2. Waltz and ZipperDB predictions of scrambled prion and nonprion domains. Each of the prion and nonprion domains analyzed in Fig. 1 was scrambled in silico such that the primary sequence was randomized while maintaining composition. This process was then repeated to generate a second set of scrambled constructs. Shown are box and whisker plots of the range of Waltz (A) and ZipperDB (B) predictions for the wild-type prion and nonprion domains before scrambling (WT) and for the first (Scr1) and second (Scr2) sets of scrambled sequences.

residues are preferentially arranged in specific sequence patterns in the prion set.

Interestingly, although scrambling did not significantly change the average ZipperDB scores in the prion and nonprion sets, it did change the distribution of scores (Fig. 2B). For each of the prion sequences before scrambling, ZipperDB predicted the most amyloid-prone segment to have a free energy between -25 and -27 kcal/mol. After scrambling, the range was much larger (-23.5 to -28.5), modestly reducing the separation between prion and nonprion sequences. This result suggests that there may be selection among the prion-forming sequences for primary sequences that confer a specific range of Rosetta energies.

Prion Domain Design. On the basis of the high prediction ability of PAPA, we tested whether it would be sufficient for de novo PFD design. We started with the yeast translation termination factor Sup35, which forms the $[PSI^+]$ prion (25). Sup35 contains three domains: an N-terminal glutamine/asparagine (Q/N)-rich prion-forming domain (amino acids 1–114) that is responsible for $[PSI^+]$ formation and propagation, a C-terminal domain that is responsible for Sup35's translation termination function, and a highly charged middle domain (26). PFDs from other prion proteins can substitute for the Sup35 PFD in supporting $[PSI^+]$ formation and propagation (27, 28). This ability to swap PFDs in a modular way provides a useful system for testing de novo designed PFDs.

We developed a simple computer algorithm to design sPFDs. Briefly, we simulated a small synthetic proteome (or "synthome") by having a computer algorithm randomly select 10^5 aa. Every amino acid except cysteine was represented in this synthome; to increase the likelihood of identifying PFDs, the composition was biased toward residues that balance prion propensity and disorder propensity. We used PAPA to scan this synthome to identify 112-aa segments that (i) were predicted to be entirely disordered, (ii) had identical Q/N content to the Sup35 PFD (52 Q/Ns in residues 3–114), and (iii) had high predicted prion propensity. We identified two segments that fit these criteria (Fig. 3; see Fig. S3 for prion propensity maps): sPFD-1 and sPFD-2. As negative controls, we selected three 112-aa segments (cPFD-1, -2, and -3) that fit the first two criteria, but that were predicted to have low prion propensity (Fig. 3 and Fig. S3).

sPFDs Form Stable, Curable Prions. To test each domain for its prion-forming ability, we took advantage of the fact that $[PSI^+]$ increases read-through of stop codons, allowing for $[PSI^+]$ detection by monitoring nonsense suppression of the *ade2-1* allele (29). In $[psi^-]$ (nonprion) cells, *ade2-1* mutants are unable to grow in the absence of adenine and turn red in the presence of limiting adenine; by contrast, $[PSI^+]$ cells can grow in the absence of adenine and are white on limiting adenine (29). Each sPFD and cPFD was inserted into Sup35 in place of the Sup35 PFD (residues 3–114). When expressed as the sole copy of Sup35, each protein efficiently

complemented a *sup35* deletion (Fig. 4A). When these cells were grown in the absence of adenine, formation of Ade^+ colonies was rare (Fig. 4B). Such rare Ade^+ colonies could result from either DNA mutation or $[PSI^+]$ formation. To distinguish between these possibilities, we tested the effect of PFD overexpression. The frequency of the Sup35 misfolding events that initiate $[PSI^+]$ formation is dependent on Sup35 concentration, whereas the frequency of chromosomal mutations is insensitive to Sup35 expression levels (25). Overexpression of the matching sPFD increased Ade^+ colony formation (Fig. 4B), suggesting that the Ade^+ phenotype was a result of a prion. Almost no Ade^+ colonies, and no increase upon overexpression, were seen for the cPFDs (Fig. 4B). To confirm that expression of the cPFDs had not caused loss of $[PIN^+]$, a prion required for efficient $[PSI^+]$ formation (30), a plasmid expressing wild-type Sup35 was reintroduced into each strain; after loss of the cPFD-expressing plasmid, cells were assayed for $[PSI^+]$ formation. In all cases, the strains maintained the ability to form $[PSI^+]$, demonstrating that they were still $[PIN^+]$ (Fig. S4).

The Ade^+ phenotype was further tested for curability, dominance, and transmissibility. Low concentrations of guanidine HCl cure yeast prions by inhibiting the chaperone Hsp104 (31, 32). For both sPFDs, we identified Ade^+ colonies that stably maintained the Ade^+ phenotype in the absence of guanidine and lost the Ade^+ phenotype after growth on medium containing guanidine (Fig. 4C and Table 1). When these curable Ade^+ cells were mated with $[psi^-]$ cells expressing the identical version of Sup35, the Ade^+ phenotype was generally dominant; however, when mated with $[psi^-]$ cells expressing wild-type Sup35, the Ade^+ phenotype was always recessive (Table 1). Furthermore, the Ade^+ phenotype was efficiently transferred by cytoduction into cells expressing the matching version of Sup35 (Table 1); cytoduction is a method that transfers cytoplasmic elements, but not nuclear

```

MSQQYNQNNLYQQGQQQNGEQSFWYQQNNLQQQGNLYQ
sPFD-1: QYNYTNGNNTSQTISQGGQNGGNQNNRQNNQNTA
PNSTSTSTNGYGASGHGRSTTSYGVQDHSGARIESA...
MSNQYNNVNWQIDQLQQNQNGYNNQNIQQQAQNTQN
sPFD-2: NSSNSMQHNTNFQYQRNQTQNSQANLNQPPQYAN
NQNSGSEHMTSHFHGELASRHSMTGSETGSTYTAT...
MSQAASTKQSTETQNGALNQTQDQNHQTPVGRNNQNGT
cPFD-1: QNPNYSEQPNQNNWTRNQSNNSAQNQPPQDQNNTRG
NQQQEPQQASGTSLAMNQHTKLNENNSQDFLQQMW...
MSTSQTMPMQEQLPEQDLQNNNGAHQQLNMSNNNASNMQ
cPFD-2: NKHQIQQKNQOMNKRKYSSNHTQQTSHNHWPPGAN
FLQQDNKQNNQGTQKLDQQNQGTQNPNSQNHNIQNN...
MSDQHKTNSSQQRQVNSPQHQQKQVQHEQNSDNRQNEQ
cPFD-3: FNNKQKTNKRSQQNIQNRQNYHQNPSAGQALQHQNHK
RYENNQAWHQQAQGHYSQKQQAGTKTNNSSQQTQNNQ...
    
```

Fig. 3. Sequences of sPFDs and cPFDs.

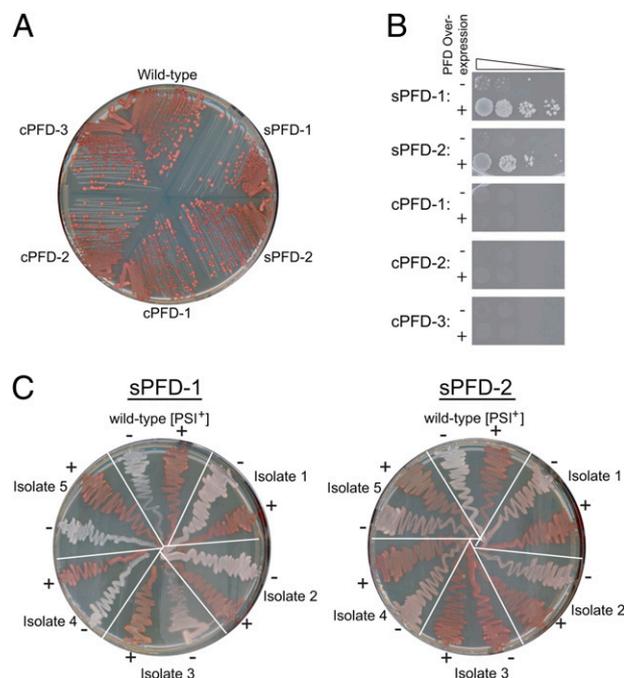


Fig. 4. Prion formation by sPFDs. (A) $[psi^-]$ strains expressing wild-type Sup35 or Sup35 in which the PFD was replaced with a sPFD or cPFD were streaked on YPD. (B) Induction assays for sPFDs and cPFDs. Yeast strains expressing a version of SUP35 in which the Sup35 PFD was replaced with the indicated sPFD or cPFD were transformed with either an empty vector (-) or with this vector modified to express the respective PFD from the GAL1 promoter (+). Cells were grown for 3 d in galactose/raffinose medium and serial dilutions were plated onto medium lacking adenine to select for $[PSI^+]$. (C) Curability of the sPFD's Ade^+ phenotype. 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.

genes (33). Collectively, these results demonstrate that the Ade⁺ phenotype is the result of a prion. By contrast, although extended incubation (10–12 d) of the cPFD inductions yielded rare Ade⁺ colonies, none were dominant or transmissible by cytoduction.

Discussion

Numerous algorithms have been developed to predict aggregation propensity. However, despite the importance of Q/N-rich domains both in disease and as regulatory elements, none of these algorithms has rigorously been validated on Q/N-rich proteins. We previously proposed that Q/N-rich amyloid proteins may represent a unique class of amyloid-forming proteins (9, 15). Specifically, although extensive evidence points to the importance of short, highly amyloidogenic stretches in driving amyloid formation by non-Q/N-rich proteins (13, 14, 16), we proposed that amyloid formation by Q/N-rich domains is driven by more diffuse sequence features, spread out over large, intrinsically disordered domains (9, 15). Our current data strongly support this hypothesis. First, PAPA, which scans proteins using a large window size and ignores primary sequence, is more effective at predicting amyloid/prion activity of Q/N-rich proteins than algorithms that focus on smaller sequence windows and that incorporate primary sequence requirements. Second, we show that PAPA is sufficient for PFD design.

Some evidence suggests that short stretches within Sup35 may act as important nucleation sites or interfaces within growing fibers (34, 35). However, this idea is compatible with the observation that prion formation is driven largely by amino acid composition. Sup35 has only a small number of both highly amyloid-prone and amyloid-inhibiting amino acids (9), and these residues are unevenly distributed; thus, their relative positioning naturally creates pockets of high amyloid propensity that could act as nucleating sites, explaining the apparent contradiction between the proposed importance of short stretches in nucleating [*PSI*⁺] formation and the relative insensitivity of Sup35 to scrambling. This insensitivity to scrambling, combined with the ability to design and predict prion propensity solely on the basis of composition, argues that any short nucleating sites are simply created by clustering of amyloid-prone residues and that any primary sequence requirements are so broad as to have almost no predictive value.

The sPFDs further suggest that the amyloid stretches predicted by Waltz and ZipperDB are not predictive of prion propensity. Among the sPFDs and cPFDs, all had predicted fiber-forming segments by ZipperDB (defined as a Rosetta energy below -23 kcal/mol), and all but cPFD-3 had Waltz-positive segments (Table S3). Indeed, Zyggregator was the only other algorithm to effectively distinguish between the cPFDs and sPFDs (Table S3). In short, algorithms that focus largely on composition (PAPA and Zyggregator) show far more predictive accuracy for Q/N-rich domains than those that focus on primary sequence, further supporting the idea that whereas primary sequence exerts subtle effects on prion formation, prion propensity is largely determined by composition.

Our ability to rationally design synthetic PFDs marks a critical milestone in our understanding of these unique protein moieties. Although other laboratories have assembled artificial PFDs by

modifying or combining primary sequence or compositional elements from other amyloid or prion proteins (6, 7, 36–38) or by modifying naturally occurring proteins (39), these experiments are unique in that they involve de novo designed PFDs. The ability to design sPFDs may facilitate the design of systems that posttranslationally control enzyme activity in a cell. Furthermore, because our PFDs were designed solely on the basis of three criteria—Q/N content, predicted prion propensity, and FoldIndex order propensity—it should be possible to design PFDs with a wide range of compositions, aiding in potential biotechnology applications of PFDs (40).

Excluding Q/N residues, the sPFDs are only 58% and 45% compositionally similar to the Sup35 PFD (Table S4). Glutamate, histidine, isoleucine, threonine, and tryptophan, although all absent from the Sup35 PFD, are all present in one or both of the sPFDs, demonstrating the flexible compositional requirements for prion formation and propagation. Furthermore, the cPFDs have a similar degree of compositional similarity to Sup35, with 45%, 40%, and 44% similarity, respectively, excluding Q/N residues, confirming that compositional similarity to known prions is a poor predictor of prion propensity (2, 9).

It is striking that the sPFDs not only form prions, but also can stably propagate these prions over many generations. We have previously identified Sup35 mutants that are able to form Ade⁺ colonies, but are unable to stably maintain the Ade⁺ phenotype without selection (7). The Sup35 PFD is composed of two subdomains: Amino acids 1–40 seem to drive amyloid nucleation and the ability to add to preexisting prion aggregates, whereas amino acids 41–114 are largely dispensable for these activities, but are required for prions to be propagated over multiple generations (36). These two regions have different compositional requirements, arguing that the ability to add to preexisting aggregates and the ability to propagate aggregates are driven by different compositional features (36, 41). The fact that we did not include separate criteria for each activity, yet the sPFDs had full prion activity, suggests that the compositional requirements for these two activities are quite broad.

Finally, it will be interesting in future experiments to dissect how each of our three design criteria separately affects prion propensity. For example, here we focused specifically on Q/N-rich proteins, because PAPA is uniquely well suited for this class of proteins. However, it remains possible that there may be a subset of non-Q/N-rich amyloid proteins that behave like Q/N-rich proteins (with amyloid formation driven by large disordered stretches), potentially allowing for design of non-Q/N-rich PFDs.

Materials and Methods

Yeast Media. All experiments were performed at 30 °C. Standard yeast media were as previously described (42), except that YPD contained 0.5% (wt/vol) yeast extract instead of the standard 1%. Galactose/raffinose dropout medium contained 2% (wt/vol) galactose and 1% (wt/vol) raffinose.

PAPA Algorithm. PAPA uses the prediction method described in Toombs et al. (2010). A detailed description of this method and instructions for using PAPA can be found in *SI Materials and Methods*. A Python script for running PAPA can be found at <http://combi.cs.colostate.edu/supplements/papa/>.

Table 1. Curability, dominance, and transmission of the Ade⁺ phenotype

	Curing*	Dominance [†]		Cytoduction [‡]	
		With matching PFD	With wild-type PFD	Isolate 1	Isolate 2
sPFD-1	16/20	14/16	0/16	11/12	12/12
sPFD-2	9/20	6/9	0/9	10/12	7/12

*For each sPFD, 20 Ade⁺ isolates were tested for curing. Numbers indicate the fraction of Ade⁺ isolates that maintained the Ade⁺ phenotype when grown in the absence of guanidine HCl, but lost the Ade⁺ phenotype when grown in the presence of guanidine HCl.

[†]All curable Ade⁺ colonies were mated with [*psi*⁻] cells expressing either the identical version of Sup35 or wild-type Sup35. Numbers indicate the fraction of isolates in which the Ade⁺ phenotype was dominant. Because dominance testing requires multiple handling steps, this assay both tests dominance and acts as a more stringent test of prion stability.

[‡]Two independent Ade⁺ isolates were used as cytoduction donors into cells expressing the same version of Sup35. Numbers indicate the fraction of recipient cells that were Ade⁺.

sPFD and cPFD Design and Construction. An Excel spreadsheet was used to simulate a synthetic proteome by randomly selecting 100,000 aa. To facilitate identification of disordered Q/N-rich domains with high predicted prion propensity, the proteome was biased in favor of Q/N residues and in favor of other residues that balance disorder and prion propensity. At each position in the proteome, each amino acid had the following probability of being selected: 23% Gln; 23% Asn; 8% Ser; 6% Gly; 6% Thr; 4% each of Ala, His, and Tyr; and 2% each of Asp, Glu, Phe, Leu, Ile, Lys, Met, Pro, Arg, Val, and Trp. The only amino acid excluded was Cys, to preclude the possibility of disulfide bonds. The random number feature in Excel was used to select each amino acid (for example, a random number between 0 and 0.04 indicated Ala, 0.04–0.06 indicated Asp, etc.).

This “proteome” was then scanned using PAPA (9). PAPA effectively uses an 81-aa window, but weights each amino acid in inverse proportion to its distance from the center of the window (see *SI Materials and Methods* for details). The predicted prion propensities of all 81-aa windows across the proteome were calculated. Potential 112-aa sPFDs and cPFDs were then identified as the highest- and lowest-scoring 81-aa windows, plus the 31 aa that followed the window in the proteome. Only domains with exactly 52 Q/N residues (the same number as in amino acids 3–114 of Sup35) were considered for further testing. Domains containing any windows predicted to be ordered by FoldIndex (8), using a 41-aa window size, were excluded from testing.

The two highest-scoring and three lowest-scoring domains were then built as synthetic genes with yeast optimized codons, using a modified version of the previously described method (7). Briefly, eight overlapping oligonucleotides were used to code for each domain (see *Table S5* for oligonucleotide sequences). These oligonucleotides were combined and amplified by PCR. PCR products were cotransformed with BamHI/HindIII-cut pJ526 (a *LEU2* cen plasmid; from Dan Masison, National Institutes of Health, Bethesda) into strain 780-1D (*MAT α kar1-1 SUQ5 ade2-1 his3 leu2 trp1 ura3 sup35*; from Dan

Masison) (43) carrying the *SUP35* maintainer plasmid pJ533 (*URA3*) (43). Transformations were selected on SC–Leu and then transferred to 5-fluoroorotic acid plates to select for loss of pJ533. Plasmids were then confirmed by DNA sequencing.

Building Induction Plasmids. The NM domain of each sPFD and cPFD was amplified by PCR, using the antisense oligonucleotide EDR969 paired with a primer unique to the specific sPFD or cPFD. EDR969 installs a stop codon and PstI restriction site at the end of the middle (M) domain. PCR products were digested with BamHI and PstI and inserted into BamHI/PstI-cut pKT24, a *TRP1* 2- μ m plasmid containing the *GAL1* promoter (7). Ligation products were transformed into *Escherichia coli* and analyzed by DNA sequencing.

Dominance Testing and Cytoduction. To generate recipient strains for dominance and cytoduction experiments, each sPFD and cPFD, along with the wild-type PFD, was PCR amplified using primers EDR261 and EDR301. PCR products were cotransformed with AatII/HindIII-cut pJ533 into YER282 (*MAT α kar1-1 SWQ5 ade2-1 his3 leu2 trp1 ura3 arg1::HIS3 sup35::KanMx pER186*) (7), selecting on SC–Ura. Transformants were then restreaked for single colonies on SC–Ura and replica plated to identify cells that had lost pER186.

To test for dominance, these cells were resuspended in water with Ade⁺ isolates from the induction experiments and spotted onto YPAD plates. After 48 h, the YPAD plates were replica plated to SD + Ade + Trp to select for diploids. Diploids were then replica plated onto SD + Trp and YPD to test for the Ade⁺ phenotype. Cytoduction experiments were performed as previously described (7).

ACKNOWLEDGMENTS. We thank P. Shing Ho, Olve Peersen, and members of the E.D.R. laboratory for helpful comments. This work was supported by National Science Foundation Grant MCB-1023771 (to E.D.R.).

- Fowler DM, Koulov AV, Balch WE, Kelly JW (2007) Functional amyloid—from bacteria to humans. *Trends Biochem Sci* 32:217–224.
- Maclea KS, Ross ED (2011) Strategies for identifying new prions in yeast. *Prion* 5:263–268.
- Cushman M, Johnson BS, King OD, Gitler AD, Shorter J (2010) Prion-like disorders: Blurring the divide between transmissibility and infectivity. *J Cell Sci* 123:1191–1201.
- Zoghbi HY, Orr HT (2000) Glutamine repeats and neurodegeneration. *Annu Rev Neurosci* 23:217–247.
- Decker CJ, Teixeira D, Parker R (2007) Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in *Saccharomyces cerevisiae*. *J Cell Biol* 179:437–449.
- Ross ED, Baxa U, Wickner RB (2004) Scrambled prion domains form prions and amyloid. *Mol Cell Biol* 24:7206–7213.
- Ross ED, Edskes HK, Terry MJ, Wickner RB (2005) Primary sequence independence for prion formation. *Proc Natl Acad Sci USA* 102:12825–12830.
- Prilusky J, et al. (2005) FoldIndex: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21:3435–3438.
- Toombs JA, McCarty BR, Ross ED (2010) Compositional determinants of prion formation in yeast. *Mol Cell Biol* 30:319–332.
- 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.
- 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.
- Tartaglia GG, et al. (2008) Prediction of aggregation-prone regions in structured proteins. *J Mol Biol* 380:425–436.
- Goldschmidt L, Teng PK, Riek R, Eisenberg D (2010) Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc Natl Acad Sci USA* 107:3487–3492.
- Maurer-Stroh S, et al. (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods* 7:237–242.
- Ross ED, Toombs JA (2010) The effects of amino acid composition on yeast prion formation and prion domain interactions. *Prion* 4:60–65.
- Esteras-Chopo A, Serrano L, López de la Paz M (2005) The amyloid stretch hypothesis: Recruiting proteins toward the dark side. *Proc Natl Acad Sci USA* 102:16672–16677.
- Nelson R, et al. (2005) Structure of the cross-beta spine of amyloid-like fibrils. *Nature* 435:773–778.
- Kuhlman B, Baker D (2000) Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci USA* 97:10383–10388.
- 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.
- Linding R, Schymkowitz J, Rousseau F, Diella F, Serrano L (2004) A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J Mol Biol* 342:345–353.
- Trovato A, Seno F, Tosatto SC (2007) The PASTA server for protein aggregation prediction. *Protein Eng Des Sel* 20:521–523.
- Fiumara F, Fioriti L, Kandel ER, Hendrickson WA (2010) Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* 143:1121–1135.
- McDonnell AV, Jiang T, Keating AE, Berger B (2006) Paircoil2: Improved prediction of coiled coils from sequence. *Bioinformatics* 22:356–358.
- Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science* 252:1162–1164.
- Wickner RB (1994) [URE3] as an altered URE2 protein: Evidence for a prion analog in *Saccharomyces cerevisiae*. *Science* 264:566–569.
- Ter-Avanesyan MD, Dagesamanskaya 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.
- Santoso A, Chien P, Osherochich LZ, Weissman JS (2000) Molecular basis of a yeast prion species barrier. *Cell* 100:277–288.
- Sondheimer N, Lindquist S (2000) Rnq1: An epigenetic modifier of protein function in yeast. *Mol Cell* 5:163–172.
- Cox BS (1965) PSI, a cytoplasmic suppressor of super-suppressor in yeast. *Heredity* 26:211–232.
- Derkatch IL, Bradley ME, Hong JY, Liebman SW (2001) Prions affect the appearance of other prions: The story of [PIN(+)]. *Cell* 106:171–182.
- 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.
- 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.
- Conde J, Fink GR (1976) A mutant of *Saccharomyces cerevisiae* defective for nuclear fusion. *Proc Natl Acad Sci USA* 73:3651–3655.
- Chen B, et al. (2010) Genetic and epigenetic control of the efficiency and fidelity of cross-species prion transmission. *Mol Microbiol* 76:1483–1499.
- Tessier PM, Lindquist S (2007) Prion recognition elements govern nucleation, strain specificity and species barriers. *Nature* 447:556–561.
- Osherochich LZ, Cox BS, Tuite MF, Weissman JS (2004) Dissection and design of yeast prions. *PLoS Biol* 2:E86.
- 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.
- Tank EM, Harris DA, Desai AA, True HL (2007) Prion protein repeat expansion results in increased aggregation and reveals phenotypic variability. *Mol Cell Biol* 27:5445–5455.
- Halfmann R, et al. (2011) Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins. *Mol Cell* 43:72–84.
- Scheibel T, et al. (2003) Conducting nanowires built by controlled self-assembly of amyloid fibers and selective metal deposition. *Proc Natl Acad Sci USA* 100:4527–4532.
- 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.
- Sherman F (1991) Getting started with yeast. *Methods Enzymol* 194:3–21.
- Song Y, et al. (2005) Role for Hsp70 chaperone in *Saccharomyces cerevisiae* prion seed replication. *Eukaryot Cell* 4:289–297.