doi:10.1016/j.jmb.2007.01.070 J. Mol. Biol. (2007) 368, 273–282

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Evolution of Budding Yeast Prion-determinant Sequences Across Diverse Fungi

Luke B. Harrison[1], Zhan Yu[1], Jason E. Stajich[2] Fred S. Dietrich[2] and Paul M. Harrison[1]*

[1]*Department of Biology McGill University Stewart Biology Building 1205 Docteur Penfield Ave Montreal, QC Canada H3A 1B1*

[2]*Department of Molecular Genetics and Microbiology Duke University, Durham NC 27710, USA*

**Corresponding author*

Prions are transmissible self-replicating alternative states of proteins. Four prions ([PSI+], [URE3], [RNQ+] and [NU+]) can be inherited cytoplasmically in *Saccharomyces cerevisiae* laboratory strains. In the case of [PSI+], there is increasing evidence that prion formation may engender mechanisms to uncover hidden genetic variation. Here, we have analysed the evolution of the prion-determinant (PD) domains across 21 fungi, focusing on compositional biases, repeats and substitution rates. We find evidence for constraint on all four PD domains, but each domain has its own evolutionary dynamics. For [PSI+], the Q/N bias is maintained in fungal clades that diverged one billion years ago, with purifying selection observed within the *Saccharomyces* species. The degree of Q/N bias is correlated with the degree of local homology to prion-associated repeats, which occur rarely in other proteins (<1% of sequences for the proteomes studied). The evolutionary conservation of Q/N bias in Sup35p is unusual, with only eight other *S. cerevisiae* proteins showing similar, phylogenetically deep patterns of bias conservation. The [URE3] PD domain is unique to *Hemiascomycota*; part of the PD domain shows purifying selection, whereas another part engenders bias changes between clades. Also, like for Sup35p, the [RNQ+] and [NU+] PD domains show purifying selection in *Saccharomyces* species. Additionally, in each proteome, we observe on average several hundred yeast-prion-like domains, with fewest in fission yeast. Our findings on yeast prion evolution provide further support for the functional significance of these molecules.

Keywords: prion; amyloid; bias; yeast; evolution

Introduction

Prions are alternative, propagating, transmissible states of normal cellular proteins. Prions were originally defined as the causative agent in mammalian neurodegenerative diseases, including scrapie in sheep and Creutzfeldt-Jakob disease in humans.[1] In *Saccharomyces cerevisiae*, prions were first identified as cytoplasmic elements inherited in a non-Mendelian fashion.[2–4] There are four known: [PSI+], [URE3], [NU+] and [RNQ+].[2,3,5–7] [PSI+] arises from the propagation of a misfolded form of Sup35p, part of the translation termination complex. Thus, formation of [PSI+] prions reduces the efficiency of translation termination and increases levels of nonsense-codon readthrough.[3,8,9] Such readthrough has been demonstrated to be a potential mechanism to uncover cryptic genetic variation.[10,11] [URE3], the prion form of the nitrogen catabolism protein Ure2p, functions to upregulate poor nitrogen source usage, even when rich sources are available.[2,4,5] Two other prions, [RNQ+] and [NU+] can function to allow [PSI+] induction by Sup35p overexpression.[12,13]

A defining characteristic of the known yeast prions is a region with a pronounced bias for glutamine (Q) and/or asparagine (N) residues.[7,14] Mutation of these residues diminishes or abolishes prion formation.[15,16] Previously, we have shown that, for three of the prions ([PSI+], [URE3] and [RNQ+]), Q/N-biased regions defined using a simple binomial probability algorithm are congruent with prion-determinant (PD) domains found in experiments.[14] For the fourth prion [NU+], the algorithmically

Abbreviations used: PD, prion determinant; LPS, lowest probability sub-sequence; YPL, yeast-prion-like.

E-mail address of the corresponding author: paul.harrison@mcgill.ca

defined Q/N-biased region coincides with a region necessary for [NU+] propagation.[7] Also, repetitive sequence motifs in the PD domain are structurally important in prion formation.[17–19] Specifically, decreasing/increasing the copy number of [PSI+] prion peptide repeats retards/induces prion formation.[17–19] PD domains for [PSI+] and [URE3] form β–sheet amyloid fibrils.[20,21] Crystallographic data for a [PSI+] peptide repeat[21] indicates stabilization by polar-zipper side-chain hydrogen bonds amongst glutamine and asparagine residues, and aromatic π interactions between tyrosine residues.

There is a growing body of evidence for conservation of prion-forming ability in yeasts. Homologues of *S. cerevisiae* PD domains from other yeasts can also form prions, either in *S. cerevisiae*, or in cells of their own species.[22–25] Also, "scrambled" forms of the Ure2p and Sup35p PD domains that maintain the amino acid composition, can form prions in *S. cerevisiae*, indicating that prion formation is primarily dependent on the composition of PD domains.[26,27]

Here, we survey the evolution of four known prion-determinant domains from budding yeast, across 21 diverse fungi. We demonstrate that prion-associated biases are maintained in fungi that are estimated to have diverged from each other about one billion years ago; furthermore, there is evidence for purifying selection, to varying degrees, on different prion domains and sub-domains.

Results and Discussion

What are the evolutionary dynamics of PD domains in yeasts and other fungi? Is there evidence of evolutionary constraint? We hypothesize that there might be such constraint, and so we examined PD evolution in two ways: (i) the conservation of the degree of Q/N compositional bias in the PD domains, using a method previously described, which calculates the lowest probability sub-sequences (LPS) within a given sequence;[14] and (ii) ratios of non-synonymous and synonymous codon substitution rates (i.e. K_a/K_s values), which is an indicator of positive selection (if K_a/K_s is significantly >1) or purifying selection ($K_a/K_s<1$). We studied the four known prions (Table 1), and yeast-prion-like biases, in 21 recently sequenced fungal genomes (Figure 1).

[PSI+]/Sup35p

Firstly, for Sup35p, we examined the degree of Q/N compositional bias in the PD domain across a diverse set of fungi, from the *Basidiomycota* and *Euascomycota*, as well as the *Hemiascomycota*. A maximum-likelihood phylogenetic tree was calculated for Sup35p orthologs (Figure 2). This tree generally follows a reference phylogeny that was derived for fungi (Supplementary Data, Figure 1). The degree of compositional bias for glutamine and asparagine was calculated as described, using the LPS method.[14] The biases were marked on the tree with binomial *P*-values (see the legend to Figure 2 and Methods for details). For each species, the number of regions in the proteome that are as biased as the PD domain homolog is also an indicator of bias maintenance (Supplementary Data, Table 2). Strikingly, the Q/N bias is conserved even in the evolutionarily distant fungal groups *Basidiomycota* and *Euascomycota*, with only about one region in every 100 sequences or so, at least as biased as the Sup35p PD domain (Figure 2 and Supplementary Data, Table 2). The last common ancestor with *Basidiomycota* and *Euascomycota* has been estimated using molecular clock analysis to be ~1 to 1.2 billion years ago.[28] The biases are strongest not in the *S. cerevisiae* sequence, but in Sup35p from the other hemi-ascomycetes *Candida glabrata*, *Saccharomyces castellii* and *Candida albicans*. The latter has been shown experimentally to form prions in its own cells.[23]

How unusual is Sup35p's maintenance of Q/N bias in the *Basidiomycota* and *Euascomycota*? To check this, we took the list of *S. cerevisiae* proteins which have as much Q/N bias as the Sup35p prion determinant (totalling 52 sequences). We then searched for orthologs that maintain a yeast-prion-like (YPL) region in *Basidiomycota* and *Euascomycota*

Table 1. Domains in the prion sequences, with analysis of K_a/K_s

Name	Location of PD	Aligned range[a]	Mean pairwise K_a/K_s (+/−s.d.)[b]	$2\Delta L$[c]
Sup35p/[PSI+] PD domain	1–123	1–123	0.146 (+/−0.123)	110 (4)
Sup35p/[PSI+] C-terminal globular domain	–	237–685	0.027 (+/−0.015)	434 (4)
Ure2p/[URE3] PD domain	1–65	7–45	0.024 (+/−0.029)	212 (12)
Ure2p/[URE3] C-terminal globular domain	–	109–351	0.019 (+/−0.012)	1876 (12)
Rnq1p/[RNQ+] PD domain	153–405	153–405	0.084 (+/−0.029)	358 (4)
New1p/[NU+] PD domain	1–153	2–153	0.015 (+/−0.012)	142 (6)

[a] Start and end points of the domains analysed in the relevant *S. cerevisiae* protein sequence; these serve to indicate the subsequences used to calculate K_a/K_s, guided by MUSCLE multiple alignments (Supplementary Data, Table 4).
[b] Pairwise maximum-likelihood–derived K_a/K_s values were calculated using PAML[32] (see Methods for details).
[c] For all of the sequences listed for each prion above, branch-specific K_a/K_s values were calculated using maximum likelihood, and the package PAML.[32] The log likelihood ratio test for selection (K_a/K_s significantly < or >1.0) was performed as described.[33] The test statistic, $2\Delta L$ for the log likelihood ratio test is tabulated, with the number of degrees of freedom in parentheses (all results were significant at $P<0.001$).