

DOI: 10.1002/cbic.200900373

Amyloids Go Genomic: Insights Regarding the Sequence Determinants of Prion Formation from Genome-Wide Studies

Hilal A. Lashuel*^[a] and Rohit V. Pappu*^[b]

Protein misfolding is the primary cause of several systemic and neurodegenerative diseases^[1,2] and a major challenge in the development of protein-based therapeutics,^[3] Figure 1 top. In misfolding diseases such as cystic fibrosis and α 1-antitrypsin deficiency, degradation and/or mistrafficking of specific proteins causes loss of protein function. A second class

of misfolding disease includes systemic amyloidoses, type II diabetes, Alzheimer's disease, and Parkinson's disease. These are caused by gains of toxic function resulting from the aggregation of specific misfolded proteins into highly ordered structures, termed amyloid fibrils, Figure 1. Amyloid-like protein aggregates that act as infectious agents by replicat-

aggregates isolated from tissues or prepared in vitro from purified recombinant proteins share a common core structure (cross- β -sheet), have fibrillar morphologies, and show dye-binding properties that are akin to classical amyloids. The transmission of prion diseases through misfolded and/or aggregated form(s) of the prion protein as an infectious agent distinguishes them from other amyloid diseases.

Given their link to disease, research interest in amyloids and prions has been centered on understanding the molecular mechanisms that govern their formation and toxicity.^[2] The goal is to identify molecules or agents that can inhibit and/or reverse these processes in vitro and disease progression in animal models. During the last decade, the debate on amyloids has shifted from whether they are the cause or consequence of disease^[6] to whether they might even have a protective physiological role. The idea of amyloids as being a protective rather than a toxic species is anchored in the recognition that amyloid formation appears to be a property that is generic to many proteins. Furthermore, amyloid fibrils and prions appear to be essential constituents of many living organisms and are associated with a number of important biological functions.^[1,7] In bacterial cells, the process of amyloid formation has been linked to biofilm formation on their surfaces.^[8] In mammalian cells, amyloid formation mediates melanin formation,^[9] synaptic changes involved in memory storage,^[10] and trait inheritance.^[11]

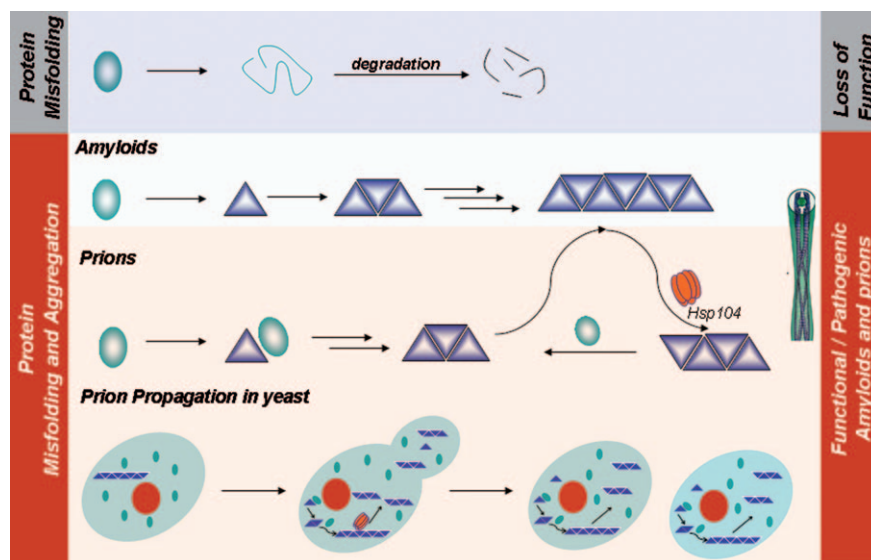


Figure 1. Schematic illustration of the proposed mechanisms of protein misfolding, prion formation, and prion propagation in yeast. The molecular chaperone Hsp104 plays a critical role in prion propagation by fragmenting prion particles, thereby creating additional seeds that nucleate the formation of additional prion particles, which are transmitted through the daughter cells.

[a] Prof. H. A. Lashuel
Brain Mind Institute (BMI), Ecole Polytechnique
Fédérale de Lausanne (EPFL)
1015 Lausanne (Switzerland)
Fax: (+41)21-693-1780
E-mail: hilal.lashuel@epfl.ch

[b] Prof. R. V. Pappu
Department of Biomedical Engineering and
Center for Computational Biology
Washington University in St. Louis
One Brookings Drive, Campus Box 1097
St. Louis, MO 63130 (USA)
E-mail: pappu@wustl.edu

ing and transmitting their misfolded state are called prions,^[4] Figure 1. Prion diseases include Creutzfeldt–Jakob disease (CJD), Gertsmann–Straussler syndrome (GSS), and fatal familial insomnia (FFI) in humans, bovine spongiform encephalopathy (BSE) in cattle, scrapie in sheep, and chronic wasting disease in elk. These diseases are characterized by the presence of an abnormal form of the prion protein (PrP) in the brain.^[5] Prion

Generating Inventories of Functional Amyloids

The full extent to which amyloid formation is used in biological functions remains unknown. Clearly, this knowledge will be useful for understanding the mechanisms by which cells process or utilize pathological amyloid formation. It also has relevance for understanding the role of amyloid formation in the regulation of different biological processes. To fill these gaps in our knowledge, it is useful to start with an inventory of putative amyloid-forming sequences. Annotations of these sequences will provide hints regarding the functional/dysfunctional roles of amyloid formation. The availability of fully sequenced genomes provides a useful starting point for identifying putative amyloid-forming sequences through genome-wide scans. With an inventory in hand, one can assess the amyloid-forming potential and the functional consequences of amyloid formation for each sequence.

Recent Progress

Although the approach outlined above is easy to articulate in theory, there are numerous hurdles to overcome, not the least of which is the lack of clarity regarding the signatures of amyloid-forming sequences. The recent work of Lindquist and co-workers^[12] provides the first proof-of-concept study that starts with genome-wide scanning to generate an inventory of putative amyloid formers (yeast prions in this case) and then goes beyond this inventory to identify the bona fide prions. Starting with prior knowledge regarding the signatures of prion-forming domains in Sup35, Ure2p, and Rnq1, Alberti et al. trained a hidden Markov model and carried out a proteome-wide scan of the *S. cerevisiae* genome.^[12] The scan yielded about 200 putative proteins with prion-forming domains. The training set biased the search to identify proteins with lengthy domains that are rich in either glutamine and/or asparagine residues. Based on these sequence characteristics, they identified 200 candidate proteins containing prion-coding sequences, which were tested by using an array of func-

tional *in vivo* assays and *in vitro* biophysical studies. Alberti et al. selected the top 100 candidate proteins from their scan to determine if they possess prion-like properties. Specifically, they tested the ability of these proteins to 1) form cytosolic foci, 2) form ordered aggregates that are resistant to detergent, 3) form amyloids *in vitro* as monitored by using thioflavin-T (ThT) fluorescence, 4) generate so-called $[PSI^+]$ states by using chimeras of candidate prion-forming domains with the N- and C-terminal domains of Sup35.^[12]

Not all Primary Amides are the Same

Prior to this study, it was believed that, although the glutamine and asparagine content is important in determining the prion-forming propensity of a given sequence, the actual ratio of glutamine to asparagine plays a minor role. The assessments of Alberti et al. yielded a rather surprising result, because they noticed that the aggregation-prone prion-forming domains tended to be enriched in asparagine residues. Conversely, the false-positive putative prions were enriched in glutamine residues as well as prolines and charged residues (Figure 2). One might argue that this result is the consequence of enrichment in charged residues and amyloid breakers such as proline. Lindquist and co-workers have initiated studies to understand the preference for asparagine-rich—as opposed to glutamine-rich—regions in strongly amyloidogenic prion-forming sequences.

Understanding the Origins of the Observed Specificities

Polyglutamine-rich regions have occupied the attention of researchers in the amyloid field because of their association with nine neurodegenerative diseases.^[13,14] In these diseases, there is a striking inverse correlation between the lengths of polyglutamine expansions within specific proteins and the ages-of-onset for diseases associated with the aggregation of these proteins or their fragments.^[15] Biophysical studies have

shed light on several important aspects of polyglutamine tracts.^[16–21] In aqueous milieu, monomeric polyglutamine forms collapsed structures to minimize the interface with the milieu.^[17,22] In short polyglutamine tracts, this preference for collapsed structures can be reversed by the addition of multiple lysine residues at the N and C termini.^[23] Initial studies suggested that polyglutamine aggregation proceeds through a homogeneous nucleation mechanism involving the formation of a monomeric, β -sheet conformation.^[24] Recent investigations indicate that the mechanism of polyglutamine aggregation is considerably more complex and in all likelihood involves the spontaneous formation of either spherical aggregates^[21] or large linear aggregates^[20] depending on the presence or absence of flanking lysine residues. The formation of ordered aggregates—which might be questionable for polyglutamine alone—appears to be a slow step that involves conformational rearrangements of small numbers of molecules within droplets that can be referred to as molten oligomers.^[25,26] Polyglutamine molecules collapse on themselves partly because the side-chain primary amides solvate backbone secondary amides; similarly, soluble molten oligomers form primarily through favorable interactions between side-chain primary amides.

If the preference for asparagine-rich regions in putative prion-forming domains originates in intrinsic differences between asparagine and glutamine residues, then it should be possible to tease out these differences by comparing the conformational ensembles for monomeric and oligomeric polyglutamine and polyasparagine, respectively. Alternatively, the differences observed by Alberti et al. might reflect the effects of different sequence contexts, which modulate the common intrinsic preferences of asparagine and glutamine residues (Figure 2). Ongoing studies from different laboratories that utilize combinations of techniques will allow us to dissect the origin for the observed specificity for asparagine over glutamine in putative prion-forming domains. Of particular interest is the role of sequence context and coarse-grain characteristics such as the charge/hydrophobicity of flanking sequences in de-

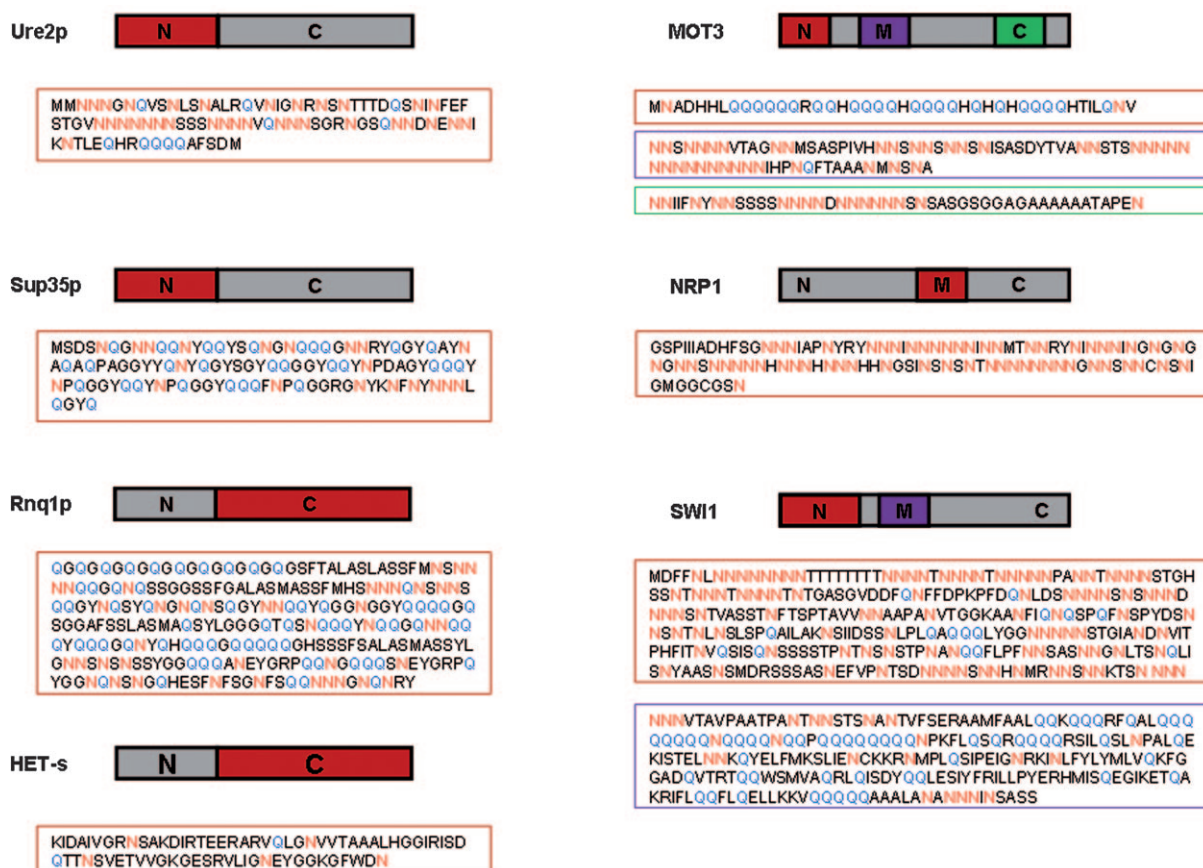


Figure 2. Schematic illustration of the primary sequence of the prionogenic glutamine- and/or asparagine-rich domains of the prion proteins Ure2p (Swiss-Prot P23202), Sup35NM (Swiss-Prot P05453), Rnq1p (Swiss-Prot P25367), HET-s (Swiss-Prot B8M215) and three of the prion candidates identified by Alberti et al. MOT3 (Swiss-Prot P54785), NRP1 (Swiss-Prot P32770), and SW1 (Swiss-Prot P09547). Glutamine residues are shown in red and asparagine residues are in blue.

termining the differential prion-/amyloid-forming abilities of glutamine- and asparagine-rich regions.

Looking Ahead

The ability of proteins to adopt functional and pathogenic states sharing similar structural properties suggests that formation of the functional state is tightly regulated and occurs under defined conditions. It has been hypothesized that proteins do not have to choose between an amyloidogenic/prionogenic and a functional state; rather, they can switch back and forth between the two states in response to cellular cues and/or environmental stresses. This hypothesis is supported by increasing evidence demonstrating the dynamic nature and reversibility of amyloid formation.^[27,28] If this is true, then what determines whether an amyloid or prionogenic protein is pathogenic or beneficial? Solving this

puzzle requires detailed analysis of the prion-coding domains; we also need a mapping of proteins' intracellular localization and their roles in different biological pathways. The ability of proteins to form distinct fibrillar morphologies from the same precursor protein, that is, prion strains, also remains a mystery.^[29] This concept has important implications for the diversity in function and pathogenicity of amyloid- and prion-forming proteins and might constitute the molecular basis underlying the ability of these proteins to switch between the different states.

The work of Alberti et al. is encouraging and constitutes an important first step toward more large-scale studies designed to answer precise questions regarding the amyloid-forming abilities of different sequences, not just those that contain glutamine- and asparagine-rich regions. In particular, systematic studies are needed to assess the validity of vari-

ous predictors of amyloidogenicity that have been developed by different laboratories.^[30–32] Such studies will provide us with an improved understanding of comparative driving forces and mechanisms of protein aggregation and are likely to be useful in developing quantitative models that enable us to understand how cells process deleterious or functional aggregates. From a biological standpoint, the work of Alberti et al. is encouraging because it is suggestive of routes that can be explored to discover novel amyloid- and prion-forming candidates in the genomes of other organisms, especially those exposed to stressful environments where the ability to switch between different functional conformations might be advantageous. A better understanding of the sequence determinants of amyloid and prion formation and propagation and the molecular determinants that govern the reversibility between functional and patho-

genic amyloids and prions will improve our understanding of how nature maintains the delicate balance between the two states to generate novel biological functions and protect against disease.

Acknowledgements

The authors would like to thank Bruno Fauvet for preparing Figure 2. This work was supported by grants from the Ecole Polytechnique Federale de Lausanne and Swiss National Science Foundation (H.A.L.) and the National Institutes of Health 5R01NS056114 (R.V.P.).

Keywords: aggregation · amyloid fibrils · asparagine · glutamine · misfolding · prions

- [1] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2006**, *75*, 333.
- [2] P. T. Lansbury, H. A. Lashuel, *Nature* **2006**, *443*, 774.
- [3] N. Wang, W. F. Smith, B. R. Miller, D. Aivazian, A. A. Lugovskoy, M. E. Reff, S. M. Glaser, L. J. Croner, S. J. Demarest, *Proteins* **2009**, *76*, 99.
- [4] S. B. Prusiner, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 13363.
- [5] A. Aguzzi, C. Sigurdson, M. Heikenwaelder, *Annu. Rev. Pathol. Mech. Dis.* **2008**, *3*, 11.
- [6] B. Caughey, P. T. Lansbury, *Annu. Rev. Neurosci.* **2003**, *26*, 267.
- [7] D. M. Fowler, A. V. Koulov, W. E. Balch, J. W. Kelly, *Trends Biochem. Sci.* **2007**, *32*, 217.
- [8] M. M. Barnhart, M. R. Chapman, *Annu. Rev. Microbiol.* **2006**, *60*, 131.
- [9] D. M. Fowler, A. V. Koulov, C. Alory-Jost, M. S. Marks, W. E. Balch, J. W. Kelly, *PLoS Biol.* **2006**, *4*, e6.
- [10] K. Si, S. Lindquist, E. R. Kandel, *Cell* **2003**, *115*, 879.
- [11] J. Shorter, S. Lindquist, *Nat. Rev. Genet.* **2005**, *6*, 435.
- [12] S. Alberti, R. Halfmann, O. King, A. Kapila, S. Lindquist, *Cell* **2009**, *137*, 146.
- [13] M. E. MacDonald, C. M. Ambrose, M. P. Duyao, R. H. Myers, C. Lin, L. Srinidhi, G. Barnes, S. A. Taylor, M. James, N. Groot, H. MacFarlane, B. Jenkins, M. A. Anderson, N. S. Wexler, J. F. Gusella, G. P. Bates, S. Baxendale, H. Hummerich, S. Kirby, *Cell* **1993**, *72*, 971.
- [14] C. J. Cummings, H. Y. Zoghbi, *Hum. Mol. Genet.* **2000**, *9*, 909.
- [15] C. A. Ross, M. A. Poirier, *Nat. Rev. Neurosci.* **2004**, *5*, 10.
- [16] S. Chen, V. Berthelie, W. Yang, R. Wetzel, *J. Mol. Biol.* **2001**, *311*, 173.
- [17] S. L. Crick, M. Jayaraman, C. Frieden, R. Wetzel, R. V. Pappu, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16764.
- [18] A. M. Bhattacharyya, A. K. Thakur, R. Wetzel, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15400.
- [19] A. Bhattacharyya, A. K. Thakur, V. M. Chellgren, G. Thiagarajan, A. D. Williams, B. W. Chellgren, T. P. Creamer, R. Wetzel, *J. Mol. Biol.* **2006**, *355*, 524.
- [20] C. C. Lee, R. H. Walters, R. M. Murphy, *Biochemistry* **2007**, *46*, 12810.
- [21] A. Vitalis, X. Wang, R. V. Pappu, *J. Mol. Biol.* **2008**, *384*, 279.
- [22] A. Vitalis, X. Wang, R. V. Pappu, *Biophys. J.* **2007**, *93*, 1923.
- [23] V. R. Singh, L. J. Lapidus, *J. Phys. Chem. B* **2008**, *112*, 13172.
- [24] S. M. Chen, V. Berthelie, J. B. Hamilton, B. O'Nuallain, R. Wetzel, *Biochemistry* **2002**, *41*, 7391.
- [25] R. Krishnan, S. L. Lindquist, *Nature* **2005**, *435*, 765.
- [26] R. V. Pappu, X. Wang, A. Vitalis, S. L. Crick, *Arch. Biochem. Biophys.* **2008**, *469*, 132.
- [27] R. Mimna, M. S. Camus, A. Schmid, G. Tuchscherer, H. A. Lashuel, M. Mutter, *Angew. Chem. Int. Ed.* **2007**, *46*, 2681.
- [28] D. Schenk, R. Barbour, W. Dunn, G. Gordon, H. Grajeda, T. Guido, K. Hu, J. Huang, K. Johnson-Wood, K. Khan, D. Kholodenko, M. Lee, Z. Liao, I. Lieberburg, R. Motter, L. Mutter, F. Soriano, G. Shopp, N. Vasquez, C. Vandevent, S. Walker, M. Wogulis, T. Yednock, D. Games, P. Seubert, *Nature* **1999**, *400*, 173.
- [29] A. Aguzzi, *Nat. Cell Biol.* **2004**, *6*, 290.
- [30] G. G. Tartaglia, A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti, M. Vendruscolo, *J. Mol. Biol.* **2008**, *380*, 425.
- [31] A. P. Pawar, K. F. Dubay, J. Zurdo, F. Chiti, M. Vendruscolo, C. M. Dobson, *J. Mol. Biol.* **2005**, *350*, 379.
- [32] A. M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, *Nat. Biotechnol.* **2004**, *22*, 1302.

Received: June 15, 2009
Published online on July 13, 2009