The polyglutamine amyloid nucleus in living cells is a monomer with competing dimensions of order

Tej Kandola[1,2,*], Shriram Venkatesan[1,*], Jiahui Zhang[3], Brooklyn Lerbakken[1], Jillian F Blanck[1], Jianzheng Wu[1,4], Jay Unruh[1], Paula Berry[1], Jeffrey J. Lange[1], Alex Von Schulze[1], Andrew Box[1], Malcolm Cook[1], Celeste Sagui[3], Randal Halfmann[1,4]

[1]Stowers Institute for Medical Research, Kansas City, MO 64110, USA, [2] The Open University, Milton Keyes, MK7 6AA, UK, [3]Department of Physics, North Carolina State University, Raleigh, NC 27695, USA, [4]Department of Molecular and Integrative Physiology, University of Kansas Medical Center, Kansas City, KS 66160, USA
*These authors contributed equally.

Abstract

A long-standing goal of the study of amyloids has been to characterize the structural basis of the rate-determining nucleating event. However, the ephemeral nature of that event has made it inaccessible to classical biochemistry, structural biology, and computational approaches. Here, we addressed that limitation by measuring the dependence of amyloid formation on concentration and conformational templates in living cells, whose volumes are sufficiently small to resolve the outcomes of independent nucleation events. We characterized numerous rationally designed sequence variants of polyglutamine (polyQ), a polypeptide that precipitates Huntington's and other amyloid-associated neurodegenerative diseases when its length exceeds a characteristic threshold. This effort uncovered a pattern of approximately twelve Qs, only for polypeptides exceeding the clinical length threshold, that allow for amyloid nucleation to occur spontaneously within single polypeptides. Nucleation was inhibited by intermolecular phase separation. Using atomistic molecular dynamics simulations, we found that the pattern encodes a minimal steric zipper of interdigitated side chains. Lateral growth of the steric zipper competed with axial growth to produce "pre-amyloid" oligomers. By illuminating the structural mechanism of polyQ amyloid formation in cells, our findings reveal a potential molecular etiology for polyQ diseases, and may provide a roadmap for the design of new therapies.

Introduction

Amyloids are highly ordered protein aggregates. They are widely associated with aging and neurodegenerative diseases (Chiti and Dobson, 2017; Huang et al., 2019), but also have functional roles in programmed cell death and subcellular compartmentalization (Boke et al., 2016; Hervas et al., 2020; Maji et al., 2009; Majumdar et al., 2012; O'Carroll et al., 2020; Rodríguez Gama et al., 2021; Vogler et al., 2018). These activities derive from an inherent tendency of amyloids to catalyze their own formation, a phenomenon driven by supersaturation resulting from sequence-encoded kinetic barriers to nucleation. The structural basis of nucleation remains mysterious despite decades of intense research.

Polyglutamine (polyQ) is an amyloid-forming module common to eukaryotic proteomes (Mier et al., 2020). In humans, it is responsible for nine invariably fatal neurodegenerative diseases, the most prevalent of which is Huntington's Disease. A remarkable feature of polyQ pathology -- found equally in Huntington's disease patients, cell culture, organotypic brain slices, and animal models -- is that neuronal loss occurs with a constant frequency over time, implying that death happens independently and stochastically in each cell (Clarke et al., 2000; Linsley et al., 2019). Two consistent observations suggest that this feature emerges directly from the nucleation barrier. First, the kinetics of polyQ aggregation in vitro, in cells, and in organismal models is consistent with primary nucleation as the rate-limiting step (Chen et al., 2002; Colby et al., 2006; Kakkar et al., 2016; Sinnige et al., 2021). Second, polyQ diseases result almost exclusively from an *intramolecular* change, specifically, an expansion in the number of sequential glutamines beyond a protein-specific threshold centered around 38 residues (Lieberman et al., 2019). In contrast to other amyloid diseases, which are frequently brought on by gene duplication, misplicing, and post-translational processing (Book et al., 2018; Kim et al., 2020; Selkoe and Hardy, 2016), polyQ diseases have relatively few *intermolecular* modifiers (Lee et al., 2019; Wexler et al., 1987). Hence, more so than for any other proteopathy, progress against polyQ diseases awaits detailed knowledge of the amyloid nucleus itself.

Despite its extraordinary relevance to both basic biology and human health, no empirically determined structure of an amyloid nucleus yet exists. This is because nuclei cannot be observed directly by any existing experimental approach. Unlike mature amyloid fibrils, which are stable and amenable to structural biology, nuclei are by definition unstable. Their structures do not necessarily persist into the mature amyloids that arise from them (Auer et al., 2008; Buell, 2017; Hsieh et al., 2017; Levin et al., 2014; Liang et al., 2018; Li et al., 2010; Sil et al., 2018; Yamaguchi et al., 2005; Zanjani et al., 2020). Nucleation occurs far too infrequently, and involves far too many degrees of freedom, to simulate computationally from a naive state (Barrera et al., 2021; Kar et al., 2011; Strodel, 2021). Unlike for phase separation, whose nucleation concerns primarily a loss of *intermolecular* entropy, amyloid nucleation also involves a major loss of *intramolecular* entropy (**Fig. S1A**), i.e. selects for a specific combination of backbone and side chain torsion angles (Khan et al., 2018; Vitalis and Pappu, 2011; Zhang and Schmit, 2016). As a consequence of the latter requirement, amyloid-forming proteins can accumulate to supersaturating concentrations while still remaining soluble, thereby storing potential energy that will subsequently drive their aggregation following a stochastic nucleation event (Buell, 2017; Khan et al., 2018). Once a nucleus does form, additional molecules join with relative ease, depleting the soluble level of the protein and changing its activity throughout the cell or organism. It is

for this reason that amyloid phenomena tend to progress toward predetermined fates, such as programmed cell death (Majumdar et al., 2012; Rodríguez Gama et al., 2021), neurodegeneration and aging (Chiti and Dobson, 2017; Huang et al., 2019).

Due to the improbability of critical fluctuations occurring simultaneously in both density and conformation (i.e. *homogeneous* nucleation), amyloid nucleation tends to involve *heterogeneities*, or metastable intermediates of varying stoichiometry and conformation (Auer et al., 2008; Buell, 2017; Hsieh et al., 2017; Levin et al., 2014; Liang et al., 2018; Li et al., 2010; Serio et al., 2000; Sil et al., 2018; Vitalis and Pappu, 2011; Yamaguchi et al., 2005; Zanjani et al., 2020). These heterogeneities erode the barrier to nucleation. The occurrence of heterogeneities in the nucleation pathway also implies that the nature of the rate-limiting structure for a given amyloid will depend both on the protein's *concentration* (**Fig. S1B**) and the existence of other cellular factors that influence the protein's *conformation* (**Fig. S1C**) (Bradley et al., 2002; Buell, 2017; Collinge and Clarke, 2007; Sanders et al., 2014; Törnquist et al., 2018).

In principle, however, the sequence-encoded structural features of an amyloid nucleus may be deduced by studying the effects on nucleation of iterative rational mutations made to the nucleating polypeptide. Mutations that raise or lower the barrier independently of concentration provide information by which to discriminate the possible structural models. With enough constraints, the conformation representing the saddle point of the lowest energy path over the nucleation barrier, will emerge. In practice, this approach would require an assay capable of distinguishing kinetic from equilibrium effects on amyloid formation for a large number of mutations. Classic assays of *in vitro* amyloid assembly kinetics cannot achieve this goal, both because of their limited throughput, and because experimentally tractable reaction volumes are far too large for nucleation to be rate-limiting for biologically relevant amyloids (Michaels et al., 2017).

We recently developed an assay to circumvent these limitations (Khan et al., 2018; Posey et al., 2021; Venkatesan et al., 2019). Distributed Amphifluoric FRET (DAmFRET) uses a photoconvertible fusion tag and high throughput flow cytometry to treat living cells as femtoliter-volume test tubes, thereby providing the large numbers of independent reaction vessels of exceptionally small volume that are required to discriminate independent nucleation events under physiological conditions (**Fig. 1A**). Compared to conventional reaction volumes for protein self-assembly assays, the budding yeast cells employed in DAmFRET increase the dependence of amyloid formation on nucleation by nine orders of magnitude (the difference in volumes). DAmFRET employs a high-variance expression system to induce a hundred-fold range of concentrations of the protein of interest. By taking a single snapshot of the protein's extent of self-association in each cell, at each concentration, at a single time point appropriate for the timescale of amyloid formation (hours), we probe the magnitudes of critical fluctuations in both density and conformation that govern nucleation.

The yeast system additionally allows for orthogonal experimental control over the critical conformational fluctuation. This is because yeast cells normally contain exactly one cytosolic amyloid species -- a prion state formed by the low abundance Q- and N-rich endogenous protein, Rnq1 (Kryndushkin et al., 2013; Nizhnikov et al., 2014). The prion state can be gained or eliminated experimentally to produce cells whose sole difference is whether the Rnq1 protein does ([*PIN*$^+$]) or does not ([*pin*$^-$]) exist in an amyloid state (Derkatch et al., 2001). [*PIN*$^+$] serves as an imperfect template that stabilizes amyloid

conformations in compositionally similar proteins. Known as cross-seeding, this phenomenon is believed to be analogous to, but much less efficient than, homotypic amyloid elongation (Keefer et al., 2017; Khan et al., 2018; Serio, 2018). By evaluating nucleation frequencies as a function of concentration in both [*pin*$^-$] and [*PIN*$^+$] cells, we uncouple the two components of the nucleation barrier and thereby relate specific sequence features to the nucleating conformation.

Thus equipped, we sought to deduce the structure of the polyQ amyloid nucleus. Fortunately, polyQ may be an ideal model to elucidate nucleus structure. It has zero sequence complexity, which profoundly simplifies the design and interpretation of sequence variants. Additionally, unlike other pathogenic protein amyloids, which are notoriously polymorphic at a structural level, polyQ amyloids have an invariant core structure even under different assembly conditions and with different flanking domains (Boatz et al., 2020; Galaz-Montoya et al., 2021; Lin et al., 2017). This means that amyloid nucleation, and pathogenesis, of polyQ emerges directly from its ability to spontaneously acquire a *specific* nucleating conformation, and any variant that decelerates nucleation can be interpreted with respect to its effect on *that* conformation.

Results

A cryptic pattern of glutamines governs amyloid nucleation

We used our DAmFRET approach to quantify polyQ nucleation events among approximately one hundred thousand cells, per experiment, as a function of two variables that separately govern the density and conformational components of the nucleation barrier -- concentration and Rnq1 amyloid status.

In [*pin*$^-$] cells, we found that cells expressing polyQ partitioned into two discontinuous populations: one containing only monomers, as indicated by an absence of AmFRET **(Fig. 1B)**, and the other containing self-assemblies of the protein, as indicated by high AmFRET **(Fig. 1B)**. The two populations occurred in overlapping windows of expression (~10^7-10^8 AFU, Fig. 2B left), with very few cells at intermediate AmFRET levels. This pronounced discontinuity in the relationship between self-assembly and concentration is characteristic for prion-like amyloids, and confirms the existence of a rate-limiting conformational fluctuation. Once nucleation occurs, however, the protein rapidly polymerizes to a steady state level (Buell, 2017; Khan et al., 2018).

The tendency of Q/N-rich proteins to form amyloid increases with their length (Alberti et al., 2009). That polyQ nucleated robustly even at the relatively short length of 60 distinguishes it from dozens of physicochemically similar (Q-rich) amyloids now characterized by DAmFRET (Khan et al., 2018; Posey et al., 2021) and unpublished. This suggests that nucleation is promoted either by the exceptionally low sequence complexity of pure polyQ and/or by an emergent property of Q side chains, specifically. To discriminate between these possibilities, we tested an equivalent length homopolymer of asparagine, whose side chain is one methylene shorter but otherwise identical to that of glutamine. Remarkably, the protein failed to acquire AmFRET even at the highest level of expression (**Fig. S2A**, 60:N, [*pin*$^-$]) -- approximately 200 µM (Khan et al., 2018). To determine if this can be attributed to a larger conformational nucleation barrier than that of polyQ, we expressed both proteins in cells harboring the

heterogeneous conformational template, [*PIN*$^+$]. Indeed, polyN now nucleated with a comparable frequency and concentration-dependence to well-characterized Q/N-rich prion-forming proteins (compare **Fig. S2A**, 60:N, [*PIN*$^+$] with plots in Posey et al. 2021). Interestingly, for polyQ, the overlapping expression range between monomer- and amyloid-containing cell populations was eliminated by [*PIN*$^+$], indicating that polyQ is exceptionally susceptible to this entity (**Fig. 1B**, 60:Q, [*PIN*$^+$]). We conclude that the polyQ amyloid nucleus contains a conformational element that is stabilized by cooperative interactions unique to Q side chains.

The surprising inability of N to form this nucleating element, despite its being the most physicochemically similar residue to Q, led us to consider whether a genetic screen of N-substituted polyQ variants might reveal patterns that uniquely encode the structure of the nucleus. We reasoned that a random screen of such sequences would be unlikely to yield informative patterns, however, given that the universe of Q/N sequences, even if limited to just 40 residues (slightly above the pathogenic length threshold for Huntington's Disease), exceeds one trillion.

We therefore employed a systematic approach to rationally sample Q/N sequence space. Each sequence in the resulting series comprises a tandem repeat of *a* Qs separated by *b* Ns for a total length of *L* residues (L:Q_aN_b, where *a* and/or *b* = 1; **Table S1**). We characterized the self-assembly of each sequence via DAmFRET in both [*pin*$^-$] and [*PIN*$^+$] cells. The resulting dataset revealed three sequence determinants of amyloid nucleation.

First, repeats that were at least ⅓ Ns (i.e. for *a* ≤ 2) resembled polyN rather than polyQ: they nucleated in a highly concentration-dependent fashion, and only in [*PIN*$^+$] cells (**Fig. S2A**). This type of amyloid is clearly distinct from that of polyQ, perhaps reflecting the preference of polyN for parallel rather than antiparallel arrangements (Zhang et al., 2016). We did not characterize it further here. Second, de novo nucleation only occurred for repeats with at least 50% Qs (i.e. for *b* = 1). Third, and quite unexpectedly, [*pin*$^-$] nucleation just above this compositional threshold only occured for the subset of repeats where *a* is an odd value (60:$Q_{\underline{1}}N_1$, 60:$Q_{\underline{3}}N_1$, 60:$Q_{\underline{5}}N_1$), and not for those where *a* is an even value (60:$Q_{\underline{2}}N_1$, 60:$Q_{\underline{4}}N_1$) (**Fig. 1C** and see boxed region in **S2B**). In other words, while nucleation increases with Q density, there exists a rule related to the spacing of Qs in the sequence.

To determine if this rule concerns the glutamine side chains themselves, rather than their interaction with asparagines, we uniformly substituted Ns with alternative residues (i.e. L:Q_aX_1, with the subscript "1" omitted hereafter) of diverse physicochemistry -- either glycine, alanine, serine, or histidine. In all cases, the even/odd dependence persisted. That is, de novo nucleation was much more frequent for repeats of Q_3X and Q_5X (hereafter, $Q_{(3,5)}X$) than for Q_4X (**Fig. 1D, S2C**). The difference was, however, less pronounced in the case of S, all three repeat variants of which nucleated more robustly than for other identities of X. Unlike Q_4N and Q_4G, the Q_4S and to a lesser extent Q_4A and Q_4H variants did detectably nucleate de novo, albeit in a concentration-dependent manner resembling that of longer even-numbered variants ($Q_{(6+)}N$). The trends persisted, with $Q_{(3,5)}X$ nucleating better than Q_4X, even in the presence of a heterogeneous conformational template, [*PIN*$^+$] **(S2D).**

Unilaterally contiguous glutamines encode an exclusive zipper structure

As a starting point to uncover possible physical explanations for the dependence on even/odd spacing, we analyzed representative sequences -- 60:Q_3N, 60:Q_4N, 60:Q_5N, 60:Q, 60:N -- with state-of-the-art amyloid predictors (Charoenkwan et al., 2021; Keresztes et al., 2021; Prabakaran et al., 2021). Using their respective default parameters, we found that most predictors failed entirely to detect amyloid propensity among these sequences. Remarkably, none of the predictors distinguished 60:Q_4N relative to 60:$Q_{(3,5)}$N (Table S2), despite their completely dissimilar amyloid propensities. Apparently, the amyloid propensity of polyQ is exceptional among the known amyloid-forming sequences on which these predictors were trained and validated, hinting at a very specific nucleation mechanism.

Most pathogenic amyloids feature residues that are hydrophobic and/or have a high propensity for beta strands. Glutamine is neither (Fujiwara et al., 2012; Nacar, 2020; Simm et al., 2016). We reasoned that nucleation therefore corresponds to the formation of a specific tertiary structure unique to Q residues. Structural investigations of predominantly Q-containing amyloid cores reveal an exquisitely ordered tertiary motif wherein columns of Q side chains from each sheet fully extend and interdigitate with those of the opposing sheet, to form a so-called steric zipper (Hervas et al., 2020; Hoop et al., 2016; Schneider et al., 2011; Sikorski and Atkins, 2005). The resulting structural element exhibits extraordinary shape complementarity and an outsized role for very weak (van der Waals) interactions that cumulatively stabilize the structure. Additionally, the terminal amide of each Q side chain in the all-Q steric zipper forms a hydrogen bond to the opposing main chain amide. To minimize confusion while referring to the distinguishing features of all-Q steric zippers, rather than similarly packed but less ordered steric zippers of non-Q residues that lack regular side chain-to-main chain H bonds (Eisenberg and Sawaya, 2017; Sawaya et al., 2007), we will henceforth refer to them as “Q zippers” (**Fig. 2A**).

The high entropic cost of forming an exquisitely ordered Q zipper from the disordered ensemble of polyQ (Chen et al., 2001; Moradi et al., 2012; Newcombe et al., 2018; Vitalis et al., 2007; Wang et al., 2006) stands to be the major contributor to the nucleation barrier. If so, the spacing dependence we identified could be explained by the fact that a single Q zipper would involve only one of the two faces of each beta sheet composed of unbroken beta strands. Because the orientation of successive side chains alternates by 180° along a beta strand, a single Q zipper can be encoded entirely from every other residue in sequence, with the intervening residues having no direct role. Therefore, for short-repeat polypeptides, a Q zipper will be possible only when non-Q side chains occur exclusively after odd-lengthed Q tracts. In contrast, substitutions after even-lengthed tracts would place non-Q side chains on both sides of all potential beta strands and thereby preclude zipper formation (**Fig. 2B**). For both Q_2X and Q_4X repeats, steric zippers cannot be formed without including non-glutamine residues. In contrast, even the most substituted odd repeat (Q_1X) can form a continuous zipper of Qs (**Fig. 2C**).

To determine if the dependence on even/odd spacing and the resulting insertion/exclusion of substituted residues into a Q zipper indeed results from an inability of the zipper to accommodate non-Q side chains, we carried out fully atomistic molecular dynamics simulations with explicit solvent and state-of-the-art force fields to probe the structural stability and conformational dynamics of model polyQ steric zippers. PolyQ amyloids feature antiparallel rather than parallel beta sheets (Matlahov and van der Wel, 2019), and this preference consistently manifests in molecular simulations (Man et al., 2015; Punihaole et al., 2016; Zhang et al., 2016). The simulated amyloid core of polyQ (but not polyN)

recapitulates the fully extended, side chain-to-main chain H-bonded steric zipper (Man et al., 2015; Zhang et al., 2016) deduced by structural measurements. We therefore started most simulations as a zipped pair of two-stranded antiparallel β-sheets of 7:Q peptides (model 2O10 (Zhang et al., 2016)). We substituted Q residues in different positions with either asparagine or serine residues, and allowed the structures to evolve for up to 400 ns.

The Q zipper proved to be highly intolerant of non-Q side chains. Substituting any proximal pair of inward pointing Qs led to its dissolution, whereas the structure remained stable when any number of outward-pointing Q residues were substituted **(Fig. 2D, S3A)**. We also ran simulations from parallel beta sheet models that we previously found to be marginally stable for larger sheets (Man et al., 2015; Zhang et al., 2016), but were unable to find any configuration of parallel strands that tolerated substitutions either inside or outside the Q zipper **(Fig. S3B)**, indicating that the nucleating Q zipper cannot be composed of parallel beta-sheets.

Notably, all of the simulated S-containing structures remained intact for longer than the corresponding N-containing structures (not shown), in agreement with the DAmFRET results showing less disruption of nucleation by the former. To gain additional insight into the nucleating structure, we therefore explored potential explanations for the instability imparted by N side chains. As for Q, but not for S, N side chains can simultaneously donate and accept H-bonds. A recent sub-ångström-resolution amyloid microcrystal structure revealed that N side chains can form a pair of H-bonds (termed a “polar clasp”) with adjacent Q side chains on the same side of the strand (i.e. *i* + 2) (Gallagher-Jones et al., 2018). Because this structure is incompatible with interdigitation by opposing Q side chains (**Fig. S3C**), we reasoned that its formation could prevent zipper formation. In order to test this, we quantified the frequencies and durations of side chain-side chain H-bonds in simulated singly N-substituted polyQ strands in explicit solvent, both in the presence and absence of restraints locking the backbone into a beta strand. In short, we found no evidence that Q side chains preferentially H-bond adjacent N side chains (**Fig. S3D,E**), ruling out the possibility of disruptive polar clasps.

The simulations nevertheless did reveal the nature of the destabilizing effect of N side chains. In a perfect Q zipper, the Q side chains form hydrogen bonds both with adjacent side chains and with the backbone of the opposing sheet (**Fig. 2A,E**). When an N side chain falls inside the Q zipper, because it is one methyl group shorter, it cannot make this side chain-to-main chain H-bond. Instead, the side chain of N disrupts the packing of the opposing Q side chains in two steps. First, its terminal amide competes with the adjacent backbone amides for H-bonding with the terminal amide of an opposing Q side-chain, blocking that side chain from fully extending as required for Q zipper stability (**Fig. 2F**). Second, the change in conformation for the detached Q side chain results in steric collisions with adjacent Q side chains. This destabilization propagates through the network of side chains like falling dominos, ultimately unzipping the zipper (**Fig. S3F**).

The side chain of S is shorter than that of N and its terminal group less bulky. Hence it does not intercept H-bonds from the side chains of opposing Q residues (**Fig. S3G**). As a result, the ordering of Q side chains is maintained, and the structure can form, so long as multiple S side chains do not occur in close proximity inside the zipper. Additionally, the shorter side chain of S lessens the entropic barrier to ordered assembly. We suspect this may account for the comparable nucleation frequency of Q_aS constructs relative to pure polyQ **(Fig. 1B, S2C-D)**. In sum, our simulations confirm that $Q_{(2,4)}N$ repeats

fail to nucleate de novo because the highly ordered structure of Q zippers cannot internally accommodate non-Q side chains in general, and N side chains in particular. Therefore, the conformational nucleus for “polyQ” amyloids is a Q zipper that can only be formed by segments of sequence with unilaterally contiguous Qs.

Bilaterally contiguous glutamines encode shorter, laminated zippers

We observed that the dependence on spacing for amyloid formation declined sharply beyond 5 consecutive Qs. Above this apparent threshold, nucleation occurred de novo irrespective of the spacing of non-Q residues **(Fig. 3A, S2B)**, suggesting the onset of an alternative nucleating structure composed of strands extending approximately six residues. The fraction of cells transitioning to the high-FRET state increased both with the expression level and consecutive Q length above this threshold (**Fig. 3A, Fig. S2B**).

Because steric zipper ordering on one side of a beta strand configures the backbone for steric zipper ordering on the other side of the strand, provided both sides have contiguous Qs, the acquisition of an alternative amyloid structure with increased density by $Q_{(4,6+)}$ repeats could be explained by the formation of short Q zippers on *both* sides of the beta strands. This would result in a laminated architecture, with multiple layers of beta sheets bonded by Q zippers (**Fig. 3B**).

Consistent with this interpretation, we observed that cells containing $Q_{(6+)}$ amyloids reached higher AmFRET than those containing $Q_{(3,5)}$ amyloids (**Fig. 3C**, **Fig. S2B-D**), suggesting that the former amyloids have a higher density (number of subunits in cross-section) than those of the latter.

As a more direct probe of structural differences between the two classes of Q zipper amyloids, we exploited the resistance of amyloids to denaturation in strong detergents (Nielsen et al., 2007; Villali et al., 2020), to compare their distributions of lengths using semi-denaturing detergent-agarose gel electrophoresis (SDD-AGE) (Halfmann and Lindquist, 2008; Kryndushkin et al., 2003). We observed a dramatic and consistent dependence of amyloid lengths on Q contiguity (**Fig. 3D, S4A**). Specifically, Q_3 amyloids ran the lowest, followed by Q_5. $Q_{(4,6+)}$ repeats ran the highest. These differences cannot be explained by different elongation rates of the corresponding amyloids, because amyloid formation by the former achieves steady state more rapidly than the latter. Nor are they consistent with different primary nucleation rates (Knowles et al., 2009), given the extended region of bimodal DAmFRET for the former, which implies a greater barrier to primary nucleation. Therefore, the length differences must be attributed to a secondary process(es) such as fragmentation. Based on established relationships of fragmentation to structure (Tanaka et al., 2006; Verges et al., 2011), the much shorter lengths of $Q_{(3,5)}$ amyloids suggest a thinner, more fragile architecture than those of $Q_{(4,6+)}$, consistent with lamination of the latter.

To investigate the lengths of continuously ordered beta strands in each type of zipper, we designed variants of 60:Q_3N and 60:Q_4N with differing lengths of unilateral or contralateral contiguity, respectively. We found that in the absence of contralateral Qs, de novo amyloid formation retained a low concentration-dependence only for six or more unilaterally contiguous Qs **(Fig. 3E, S4B)**. In contrast, only three unilaterally contiguous Qs were required in the presence of contralateral Qs **(Fig. 3F, S2B,S4C)**. These findings imply that the beta strands in the single Q zipper nucleus extend for 11-

12 residues, while those in the laminated Q zipper nucleus extend for only 6 residues. This latter number is consistent with the ideal lengths of antiparallel strands (Stanger et al., 2001), and is remarkably close to the critical strand length of 7 that had been deduced for apparently monomeric nucleation in vitro (Thakur and Wetzel, 2002). The slightly longer threshold in vitro can be attributed to charge repulsion from the lysine pair that had been appended to both ends of the polyQ tract (Walters and Murphy, 2009). Our constructs lack this element, as do endogenous pathologic polyQ tracts.

The different strand lengths required to form single or laminated Q zipper nuclei explain why 60:Q_4N cannot form either type of nucleus. This is illustrated by comparing the three polypeptides with identical length, composition, and unilateral contiguity (60; 48 Qs, 12 Ns; and 4, respectively). In the case where contralateral zipper-forming frames were maximally offset (60:Q_4N), amyloid did not form. Shifting the registry by one position (60:QNQ_6NQ) allowed for very sporadic amyloid formation only at very high concentrations, and shifting the registry by one more position (60:Q_8N_2) -- to maximize overlap between contralateral zippers -- further increased amyloid formation **(Fig. 3F, S2B,S4C)**. We conclude that the extraordinary resilience of Q_4N against amyloid formation can be attributed to its being the highest possible density of Qs that still lacks sufficient contiguity both unilaterally and bilaterally.

Q zipper nucleation occurs within monomers

Another intriguing difference between unilaterally and bilaterally contiguous sequences is that the former nucleated with a greater efficiency at low concentrations (**Fig. 4A**). For example, the fraction of [*pin*$^-$] cells in which 60:Q_5N formed amyloid rose sharply with expression up to 10^7 AFU, where it plateaued at approximately 25%. In contrast, the fraction of cells in which 60:$Q_{(6+)}N$ formed amyloid rose steadily with expression even beyond 10^8 AFU. We were able to envision two, not mutually exclusive mechanisms whereby nucleation could be compromised at high concentrations. First, there may exist within the cell a limiting amount of some nucleation-promoting factor, such as a molecular chaperone. As the concentration of the protein of interest increases, this factor becomes saturated. Second, the protein may partition into “off-pathway” oligomers or condensates at high concentration, within which the nucleating conformation is disfavored.

PolyQ has been observed to form non-amyloid oligomers and condensates, which has led to the prevailing paradigm that phase separation precedes and facilitates the nucleation of conformational order (Borcherds et al., 2021; Camino et al., 2021; Crick et al., 2013, 2006; Dignon et al., 2020; Fisher et al., 2021; Halfmann et al., 2011; Peskett et al., 2018; Posey et al., 2018; Vitalis and Pappu, 2011; Yang and Yang, 2020). The second explanation above would seem to contradict this paradigm. To investigate whether phase separation may be occurring in our system, we inspected the subcellular localization of representative proteins in different regions of their respective DAmFRET plots in both [*pin*$^-$] and [*PIN*$^+$] cells. Using imaging flow cytometry, we found that all high AmFRET cells contained large round or stellate puncta, as expected for amyloid deposition, whereas virtually all cells with low or no AmFRET contained entirely diffuse protein, even at high expression in a bimodal regime (**Fig. S5A**). Next, we used confocal fluorescence microscopy to compare total cytosolic concentrations of 60:Q expressed in [*pin*$^-$] cells that either lacked or contained puncta, and found that the distributions of concentrations overlapped (**Fig. 4B**), suggesting the protein remained dispersed prior to amyloid formation even when deeply supersaturated. We conclude that polyQ, even when expressed to high

micromolar concentrations, does not visibly phase separate (as detectable by fluorescence microscopy) prior to amyloid formation in living cells.

We next probed the role of oligomerization in amyloid formation, by relaxing the dependence of oligomerization on protein expression level. To do so, we lowered solubility by expressing select proteins as fusions to a well-characterized modular moiety, µNS, that forms coiled coil-mediated multimers at low concentration and does not itself form amyloid (Brandariz-Nuñez et al., 2010; Kroschwald et al., 2015; Schmitz et al., 2009). As a control for an oligomeric nucleus, we used the human inflammasome scaffold protein, ASC. The pyrin domain of ASC self-associates via electrostatic interfaces into a monomorphic three-stranded polymer (Lu et al., 2014) that is physiologically rate limited by nucleation (Cai et al., 2014; Khan et al., 2018). We found that genetically fusing µNS to ASC pyrin drove it to nucleate in every [*pin*$^-$] cell, as expected (**Fig. 4C, S5B**). Note that amyloid formation by µNS-fused proteins results in a higher AmFRET value than that of µNS condensates themselves, and the latter does not depend on fusion partner (**Fig. S5B-C**, and other data not shown). Note also that amyloids of fusion proteins tend to have lower AmFRET than that of the corresponding non-fused proteins; we attribute this to the fact that larger proteins occupy more volume per fluorophore unit. We next asked how µNS-driven condensation influenced nucleation for Q zipper amyloids. In striking contrast to ASC pyrin, representative Q_aX proteins fused to µNS nucleated de novo at *reduced* frequencies to their unfused counterparts (**Fig. 4C, S5B**). This effect was limited to the nucleation step, as the µNS-fusions went on to populate a uniform high-FRET state in every cell that contained [*PIN*$^+$]. That is, fusion-mediated oligomerization appears to have driven to amyloid the kinetically arrested species that had previously populated the low arm of the sigmoid. To confirm that the reduction in nucleation resulted from the protein's local condensation rather than a specific effect of µNS, we also performed experiments with a separate fusion partner, human ferritin heavy chain, which forms discrete 24-mer homo-oligomers (Bracha et al., 2018). Again, nucleation was reduced in [*pin*$^-$], specifically, relative to non-fused proteins (not shown). These results reveal that oligomerization of disordered polyQ impairs the rate-limiting conformational fluctuation required for it to acquire a minimal Q zipper. Once that zipper has nucleated, however, oligomerization drives it toward amyloid. In other words, the sequence-encoded homogeneous nucleus contains an exclusively intramolecular Q zipper.

Lamination competes with Q zipper lengthening

$Q_{(6+)}$ repeats also recapitulated the susceptibility of pure polyQ to a conformational template. Whereas [*PIN*$^+$] cells expressing $Q_{(3,5)}$ polypeptides populated both high- and low-AmFRET values over a large window of expression, those expressing $Q_{(6+)}$ populated a much narrower range of AmFRET values for a given expression level **(Fig. S2B-D)**. These data suggest that lamination specifically reduces a critical density fluctuation, which is more limiting in the presence of [*PIN*$^+$]. In other words, lamination stabilizes oligomers with relatively unstable Q zippers, and this compensates for imperfections in the Q zipper that must occur at the interface of polyQ and the [*PIN*$^+$] amyloid template. It also explains why lamination lowers the strand length requirement. In support of this interpretation, we observed that $Q_{(6+)}$ transitioned from low- to high-AmFRET in a continuous, sigmoidal fashion, indicating that elongation or maturation of the nascent amyloids, rather than nucleation per se, had become limiting. Continuous DAmFRET transitions can be interpreted as the formation of a distribution of multimers wherein the extent of assembly, in terms of both amount and order, increases with expression level (Posey et al.,

2021). To identify trends between AmFRET distributions that lack bimodal regimes, we plotted the mean of AmFRET values in serial bins of protein expression level (**Fig. 5A**). The lower arms of the resulting curves represent cells with low but non-negligible levels of AmFRET. We observed that, as the consecutive Q length increased, the midpoint of the distribution shifted to lower expression levels (**Fig. 5A**).

Despite the reduced nucleation barrier for $Q_{(6+)}$, the level of AmFRET achieved at low expression (e.g. < 10^8 AFU for 60:Q_6N and < 10^7 AFU for Q:60), was reduced relative to $Q_{(3,5)}$ amyloids (**Fig. 5A, S2B-D**), suggesting that the laminated oligomers are both off-pathway to, and preclude the formation of, long Q zipper species. That their formation is nucleation-limited means that this inhibitory activity cannot be attributed simply to a depletion of naive monomers. Rather, they must share a critical intermediate with long Q zipper amyloid. This intermediate is presumably the structural element common to both of them, namely, a single short Q zipper. This short zipper, once formed, can *either* extend to form amyloids, or laminate to stabilize "pre-amyloid" oligomers.

To explore this notion, we destabilized oligomers by expressing proteins as fusions to a well-established solubilizing moiety -- maltose binding protein (MBP). MBP does not interfere with the folding of its fusion partner but, due to its bulk, sterically limits its access to other protein molecules (Raran-Kurussi and Waugh, 2012). Fusion to MBP should therefore amplify the density component of nucleation barriers, rendering nucleation more concentration-dependent. It will also restrict the dimensions of self-assemblies, resulting in lower AmFRET. Finally and most importantly, it should bias amyloid conformers toward longer Q zippers.

To validate the approach, we compared DAmFRET statistics for nucleation by ASC, without and with fusion to MBP. The C_{min} statistic theoretically corresponds to the phase boundary; the EC_{50} statistic represents the concentration at which nucleation has occurred in half the cells, and the □ (delta) statistic describes the independence of nucleation on concentration, where conformationally-limited nucleation has higher □ values (Khan et al., 2018). As expected, we found that MBP increased both C_{min} and EC_{50}, but did not change □ (**Fig. S6A**). It also reduced AmFRET levels by approximately 30%. We then tested the effect of MBP on 60:Q as well as Q_3S polypeptides below or above the length threshold (32 and 60, as determined below) for intramolecular Q zipper nucleation. We found that MBP completely eliminated detectable assemblies for 60:Q and 32:Q_3S. In contrast, MBP-60:Q_3S retained a bimodal distribution, albeit with increased concentration-dependence and reduced AmFRET of the nucleated assemblies (**Fig. 5B, S6B**). To minimize any confounding effect of MBP on oligomerization post nucleation that may preclude a detectable increase in AmFRET, we also performed DAmFRET on an unmodified 60:Q_3N expressed in trans, to report on nucleation occurring in non-fluorescent versions of the above MBP fusions. As expected, MBP-60:Q_3S, but not MBP-32:Q_3S nor MBP-60:Q, promoted 60:Q_3N amyloid formation **(Fig. S6C)**. Together, these results confirm that lamination competes with Q zipper lengthening en route to mature polyQ amyloid, and that oligomerization relaxes that competition.

Finally, we probed the effect of MBP on the switch between single and laminated Q zipper preferences that occurs at Q_7N. This sequence represents the minimal consecutive Q length with continuous unilateral contiguity that nevertheless exhibited the continuous sigmoidal DAmFRET of laminated Q zipper-forming sequences. That 60:Q_7N fails to nucleate single zipper amyloids suggests that the alternative, laminated Q zipper has a lower nucleation barrier and therefore involves a less extreme

conformational fluctuation. If so, the ability of Q_7N to nucleate a single zipper amyloid should be exposed either by selectively rewarding conformational fluctuations -- by expressing it in the presence of orthogonal long zipper amyloids. We therefore expressed 60:Q_7N in [*PIN*$^+$] cells that also expressed nonfluorescent MBP fusions of 32:Q_3S or 60:Q_3S in trans. MBP-32:Q_3S had no effect on 60:Q_7N. In contrast, MBP-60:Q_3S allowed the cells to populate *both* the intermediate- and high-AmFRET states (**Fig. 5C, S6C**), indicative of single and laminated Q zipper amyloids, respectively. Importantly, low expression of 60:Q_7N favored the former, while high expression favored the latter, consistent with their relative dependencies on oligomerization. We conclude that the MBP-60:Q_3S seeds unmasked a latent ability of 60:Q_7N to self-assemble (once nucleated) into amyloid with a single Q zipper structure. This result confirms that 60:Q_7N can form either type of amyloid, but is kinetically prevented from nucleating long Q zippers de novo due to the relative ease with which short Q zippers are stabilized through lamination.

Altogether, our experiments to manipulate critical fluctuations in density and conformation reveal the existence of a bifurcated pathway from polyQ monomer to amyloid. Specifically, the critical conformational fluctuation occurs within a disordered monomer to produce a minimal intramolecular Q zipper, which can then either laminate or lengthen in a manner governed by the presence or absence, respectively, of contralateral Qs (**Fig. 6A**). Low concentrations and solubilization favor assemblies with longer strands and fewer sheets, while high concentrations and condensation favor assemblies with shorter strands and more sheets (**Fig. 6B**). The former manifest with reduced AmFRET levels and polymer lengths, relative to the latter. [*PIN*$^+$] accelerates the rate-limiting monomer nucleation step and thereby amplifies these differences at the population level. The proposed pathway explains multiple features of our DAmFRET and SDD-AGE data not yet discussed: while $Q_{(3,5)}X$ polypeptides nucleated the predominant lower-AmFRET form at low concentrations, a higher AmFRET form consistent with laminated amyloid emerged at high concentrations (for example, compare expression levels greater than, versus less than, 10^8 AFU for 60:Q_5N and 60:Q_3S in **Fig. S2B** and **S2C**, respectively). [*PIN*$^+$] inhibited this transition (**Fig. S2B-D**), and reduced the lengths of SDS-resistant species (**Fig. S4A**).

The Q zipper nucleus exhibits the pathologic length threshold

If the rate-limiting Q zipper indeed occurs within a single polypeptide, nucleation will become sharply more concentration-dependent for sequences that are too short to form that element (**Fig. 7A**). To evaluate potential length thresholds for single Q zipper nuclei, we generated a series of length variants for Q_3S (S was chosen for its relatively modest nucleation defect that afforded a wider window of informative length variants) and, for each, we evaluated the dependence of nucleation on concentration by comparing the EC_{50} and □ statistics derived from DAmFRET (**Table S3**). We observed a positive linear relationship between □ and EC_{50}, as expected, as the length of the polypeptide fell from 40 to 36 (**Fig. 7B, S7A**). The relationship became negative below 36, implying an increasing dependence of nucleation on intermolecular interactions. We interpret this value as an approximation of the minimum length of polypeptide, in the absence of other interacting domains, required for a Q zipper to form within a single polypeptide.

We next explored the length-dependence for amyloid formation by uninterrupted polyQ. We found that amyloid failed to form below 35:Q, formed to an almost imperceptible extent at 35:Q, and formed robustly at 40:Q and above **(Fig. 7C, S7B-D)**. The sigmoidal transition from low to high AmFRET in

[*PIN*$^+$] cells shifted to lower concentrations with increasing polypeptide length. Given the inclination of polyQ toward metastable intermediates, this relationship suggests an accelerated progression of laminated zippers to amyloid by longer polyQ. Note, however, that the continuous nature of the transition from low to high AmFRET in this case means that these results cannot be interpreted with respect to a threshold length for the nucleus itself. In other words, we cannot exclude from this experiment that nucleation may occur below 35:Q and simply fail to elongate sufficiently over the accessible concentration range to produce detectable AmFRET.

Therefore, to directly interrogate the length threshold for the polyQ amyloid nucleus itself, we needed to eliminate the concentration-dependence of laminated Q zipper maturation. To do so, we pre-condensed the polyQ length variants to high local concentration by expressing them as fusions to µNS in [*PIN*$^+$] cells (**Fig. 7D**). By partitioning the proteins into a condensate, this fusion minimizes any contribution of polyQ length to density heterogeneities. We found that polyQ nucleated in almost all cells at lengths 38 and above, but in a decreasing fraction of cells at lengths 36 and below (**Fig. 7E, S7E**). These data suggest that Q zipper formation is compromised when the continuous chain length falls below 36, which provides a lower bound on the length threshold of an intramolecular Q zipper.

Structural model of the polyQ amyloid nucleus

Collectively, our data constitute an empirical survey of relevant points on the energy landscape of polyQ amyloid formation, which we illustrate in **Fig. 8**. Excursions from naive disordered monomers, represented by the basin at bottom left, are described in terms of intermolecular density and intramolecular order. Mature amyloids featuring lengthened and laminated Q zippers lie in a deep well in the opposite corner from the disordered monomers. A rare conformational fluctuation within a monomer (dashed red arrow) creates an embryonic Q zipper.

Do the constraints identified here offer a glimpse of the embryonic Q zipper structure? Our data indicate that the critical Q zipper can be accessed by polypeptides as short as approximately 36 residues and has strands of length 6. Allowing for connecting turns and arcs of two to four residues each (Chou and Fasman, 1977; Hennetin et al., 2006), it can contain no more than four β-strands. If we further grant from pioneering in vitro experiments (Bhattacharyya et al., 2005; Buchanan et al., 2014; Chen et al., 2002; Kar et al., 2013, 2011; Sikorski and Atkins, 2005) that the formation of a hairpin -- and not an arch -- is rate-limiting, then the four strands are connected by a single turn and two arcs. This allows for only two possible non-equivalent Q zipper configurations. One has the first and fourth strands unpaired and flanking either side of the central hairpin. We simulated this configuration and found it to be extremely unstable (**Fig. S8A**) and therefore discarded it from further consideration. The other configuration has the first and fourth strands paired and flanking one side of the central hairpin (**Fig. S8B**). This hypothetical fold of polyQ, along with the alternative configuration having two hairpins and one arch, had previously been simulated by multiple groups and found to be more stable than any single sheet or parallel stranded zipper configuration (Man et al., 2015; Miettinen et al., 2012). We therefore conclude that the nucleating monomer most likely comprises a pair of two-stranded antiparallel sheets in a super-secondary topology of β-(arch, hairpin, arch), and will be hereafter referred to as “AHA”.

Our observations suggest that the AHA monomer does not appreciably polymerize, and so is not yet "amyloid". We suspect the axial surface is too short or unstable to serve as an intermolecular template for converting the conformations of new subunits, a process that for polyQ takes considerably longer than the rate of naive molecules encountering the surface (Bhattacharyya et al., 2005). Our results imply that the emergence of an amyloid axis of polymerization requires that the Q zipper first grow in either of two orthogonal directions to produce a larger and more stable templating surface. This occurs via oligomerization coupled to either lamination (**Fig. 8**, dashed blue arrow), or lengthening (**Fig. 8**, dashed fuschia arrow). The latter is associated with a larger nucleation barrier and is therefore a more extreme conformational fluctuation. Consequently, early Q zipper growth proceeds through lamination, resulting in an accumulation of amyloid-like oligomers. Subsequent iterations of subunit addition present new opportunities for the Q zipper to lengthen and reinforce the axis of polymerization. Consequently, our model predicts a directional progression during polymerization (dashed purple arrows) toward multiply laminated fibers with fully extended hairpins, consistent with the latest models for mature polyQ amyloids (Boatz et al., 2020).

To understand the nature of the second energy barrier associated with Q zipper lengthening, we used the constraints derived herein to create a structural model for the minimal hypothetical subunit of long zipper amyloid. Our data imply that this species can be accessed by polypeptides as short as approximately 36 residues and has unilaterally contiguous strands of length 11. It therefore contains three beta strands connected by a single turn and arc. We performed molecular simulations to investigate the stability of this structure as an isolated monomer and found that it fell apart within 200 ns (**Fig. S8C**). Together with the above simulation, this result reveals that a Q zipper cannot be maintained by just three strands even when those strands are the length required for long Q zipper amyloid. We then asked if it persisted for longer when joined to a second such conformer across the hypothetical axial interface. Indeed, the resulting homodimer not only persisted, but became slightly *more* ordered over the duration of the simulation (**Fig S8D**), consistent with the onset of energetically downhill polymerization.

Discussion

PolyQ amyloid begins *within* a molecule

The initiating molecular event leading to pathogenic aggregates is the most important yet unknown and difficult to study step in the progression of age-associated neurodegenerative diseases including polyglutaminopathies such as Huntington's Disease. Here we used our recently developed assay to identify sequence features that govern the dependence of nucleation frequency on concentration and conformational templates, to deduce that amyloid formation by pathologically expanded polyQ begins with the formation of a minimal Q zipper within a single polypeptide molecule.

Kinetic studies performed both with synthetic peptides in vitro (Bhattacharyya et al., 2005; Chen et al., 2002; Kar et al., 2011) and recombinant protein in animals (Sinnige et al., 2021) and neuronal cell culture (Colby et al., 2006) have found that polyQ aggregates with a reaction order of approximately one. This has been interpreted as evidence that pathogenic nucleation occurs within a monomer (Bhattacharyya et al., 2005; Chen et al., 2002; Kar et al., 2011). However, this interpretation assumes

homogeneous nucleation, which is invalidated by observations of non-amyloid multimers prior to amyloids in vitro (Crick et al., 2006; Vitalis and Pappu, 2011). We found that even when accounting for heterogeneous nucleation, the original conclusion of Chen et al. holds. Specifically, the rate-limiting step for amyloid formation from soluble monomers *does* occur within a monomer, and the relevant multimeric heterogeneities *follow* rather than precede that event. Because they interact via the defining secondary and tertiary elements of “amyloid” conformation, these “pre-amyloid” oligomers are not condensates in the sense of first order phase separation. Rather, they are equivalent to extremely early amyloid polymorphisms that have not yet achieved a dominant axis of polymerization.

Given that monomers can nucleate, and that there are thousands of nucleation-competent monomers in each neuron of a patient’s brain, why does aggregation and disease onset still take years? The reason lies in the peculiar properties of polyglutamine. Glutamine residues lack a strong preference for any particular secondary structure and strongly solvate the polypeptide backbone. Hence, polyglutamine populates a collapsed ensemble of disordered states (Crick et al., 2006; Kang et al., 2017; Walters and Murphy, 2009). The Q zipper is plausibly the only stable structure that can be formed by polyQ, as it satisfies the side chain H-bonding potential without disrupting backbone H-bonding. Critically, the strong tendency toward disorder for polyQ persists even when the entropic penalties for beta structure have been prepaid (Vitalis and Pappu, 2011), suggesting that it is the formation of the defining tertiary element of polyQ amyloid, i.e. the Q zipper, rather than the beta sheets themselves, that imposes the entropic bottleneck narrow enough to delay nucleation over the timescales of human lifespan.

Pioneering studies revealed that the nucleation event involves the formation of a beta hairpin (Bhattacharyya et al., 2005; Chen et al., 2002; Kar et al., 2013, 2011; Sikorski and Atkins, 2005). In our model, and consistent with early proposals by Wetzel and colleagues (Kar et al., 2011), the termini of the hairpin fold back toward the beta turn. Unlike other models for the polyQ nucleus (Chen and Wolynes, 2017; Chen et al., 2016; Miettinen et al., 2014; T. T. M. Phan and Schmit, 2020), the AHA monomer contains an embryo of exquisitely ordered Q zipper structure necessary for the entropic bottleneck. We note that the AHA conformer is one of a small number of configurations for a minimal tertiary motif known as a beta-arcade, which was previously proposed to serve as an amyloid nucleus based on its favorable energetics and characteristic occurrence in amyloid fibrils (Kajava et al., 2010).

While our model may be impossible to confirm unequivocally -- as nuclei cannot be observed directly -- it nevertheless makes testable predictions. For example, it suggests that individual protofilaments can be as narrow as two or three zippered sheets -- approximately 1.6 or 2.4 nm, respectively. It also predicts spatial and temporal heterogeneities within polyQ aggregates, as a direct consequence of structural evolution, from short Q zippers with abundant connecting loops, to relatively long zippers with few loops. Specifically, at sufficiently high concentrations, spherulite-type aggregates can be expected with relatively disordered, soft interiors and ordered, hard exteriors. At sufficiently low concentrations, incipient protofilaments can be expected to grow in a unidirectional fashion. Some of these predictions have been borne out while this paper was in preparation. A cryoEM tomography study revealed a slab-like architecture of 51:Q amyloids in cells, with occasional restrictions to an apparent protofilament width of approximately 2 nm (Galaz-Montoya et al., 2021). Fluorescence microscopy studies of aging *C. elegans* neurons expressing Q:128, fused to the oligomerizing N-terminus of huntingtin and C-

terminal CFP, revealed spherical aggregates with an apparently soft core and hard shell (Fisher et al., 2021).

That polyQ amyloid has a rate-limiting monomeric nucleus that is inhibited by oligomerization has important implications for disease etiology. First, it provides a rationale for why the age of onset for polyQ diseases appears not to depend on expression level nor copy number of the expanded polyQ allele. Second, it explains why the diseases instead depend on the length of the polyQ tract. Length thresholds for polyglutaminopathies tend to be in the 35–45 range (Matlahov and van der Wel, 2019). In agreement, we find that intramolecular nucleation occurs only for polyQ exceeding approximately 36 residues, corresponding to the length required to form a minimal intramolecular Q zipper.

Interplay of Q zipper structure and surface tension may drive the accumulation of pre-amyloid oligomers

Soluble oligomers accumulate during the aggregation of pathologic length polyQ and/or Htt in vitro (Auer et al., 2008; Hsieh et al., 2017; Levin et al., 2014; Liang et al., 2018; Li et al., 2010; Sil et al., 2018; Vitalis and Pappu, 2011; Yamaguchi et al., 2005; Zanjani et al., 2020), in cultured cells (Olshina et al., 2010; Takahashi et al., 2008), and in the brains of patients (Legleiter et al., 2010; Sathasivam et al., 2010), and are likely culprits of proteotoxicity (Kim et al., 2016; Leitman et al., 2013; Lu and Palacino, 2013; Matlahov and van der Wel, 2019; Takahashi et al., 2008; Wetzel, 2020).

We found that Q zipper nucleation precedes the accumulation of soluble oligomers, rather than the other way around. These appeared to increase in size and order with the expression level of the protein, implying an evolution of the Q zipper structure toward mature amyloid. This progression rationalizes why polyQ diseases are rate-limited by primary nucleation despite amyloids -- the ultimate product of nucleation -- having benign or even protective roles (Kim et al., 2016; Leitman et al., 2013; Lu and Palacino, 2013; Matlahov and van der Wel, 2019; Takahashi et al., 2008; Wetzel, 2020).

As lamination maintains a lower aspect ratio than lengthening, the preference for polyQ, in a manner dependent on contralateral Qs, to oligomerize rather than elongate at low concentrations is plausibly mediated by the surface tension inherent to polyQ's self-solvating property (Crick et al., 2006; Kang et al., 2017; Walters and Murphy, 2009). We speculate that this is a consequence of the fact that the opposition of surface tension to anisotropic growth falls above a critical curvature of the surface (Williamson et al., 2010). Hence, by inhibiting growth in one dimension, the removal of contralateral Qs increases the aspect ratio of Q zipper propagation and accelerates its escape from the surface tension-dominated regime. Flanking domains are likely to influence this balance (Crick et al., 2013; Williamson et al., 2010), and this presents an intriguing avenue for further research.

A specific role for phase separation

Much recent work has revealed a tendency of amyloid-forming proteins, including those with Q-rich low complexity sequences, to undergo liquid-liquid phase separation under permissive conditions (Babinchak and Surewicz, 2020; Camino et al., 2021). PolyQ itself lacks both hydrophobic and charged residues that would confer the requisite affinity for phase separation in cellular milieu, which is rife with

solubilizing competing interactors (Boncella et al., 2020; Gao et al., 2018; Wang et al., 2018; Xing et al., 2018). The physiological concentration of huntingtin in the human brain, estimated to be low nanomolar (Macdonald et al., 2014), is likely too low for liquid-liquid phase separation in the presence of even one binding partner at its physiological concentration (Posey et al., 2018). Notwithstanding limited reports to the contrary (Fisher et al., 2021; Peskett et al., 2018; Wan et al., 2021) -- which we respectfully attribute to known effects of flanking domain interactions and partitioning into pre-existing protein compartments (Duennwald et al., 2006a, 2006b; Jiang et al., 2017) -- our findings corroborate prior demonstrations that polyQ does not phase separate prior to amyloid formation, whether expressed in human neuronal cells (Colby et al., 2006; Kakkar et al., 2016), *C. elegans* body wall muscle cells (Sinnige et al., 2021), or [*pin*$^{-}$] yeast cells (Duennwald et al., 2006a, 2006b; Jiang et al., 2017).

Counter to the prevailing paradigm, we found that oligomerization of naive polypeptides, whether from high concentration or fusion to a self-associating moiety, actually decelerated the rate-limiting step in amyloid formation -- nucleation of an incipient Q zipper. This likely relates to the fact that polyQ solvates itself better than water and consequently is already condensed intramolecularly (Crick et al., 2006; Walters and Murphy, 2009). That coalescence of those monomers into larger globules inhibited nucleation suggests that surface tension amplifies conformational fluctuations or even biases the conformational ensemble toward Q zippers. Prior molecular simulation studies reveal evidence of such effects. For example, polyQ polypeptides relax to a more expanded state within globules, with fewer reversals of chain direction, as a consequence of their becoming solvated by identical amides from other molecules rather than their own (Vitalis et al., 2009). Additionally, the side chains of Qs at the surface of polyQ globules are more dynamic and tend to orient toward the interior (Kang et al., 2017). Whether these changes correspond to backbone and side chain configurations that are more permissive to Q zipper formation remain to be determined.

We did find, however, that once an incipient Q zipper formed, phase separation strongly promoted its maturation to amyloid. This is consistent with lamination producing a multi-subunit templating surface; with the fact that increasing concentrations stabilize short beta strand interactions (T. M. Phan and Schmit, 2020); and with observations that polyQ amyloid elongation can occur without full consolidation to the fibril structure (Walters et al., 2012). To the extent that amyloid-associated soluble polyQ species drive disease, condensation would be more likely to ameliorate than exacerbate proteotoxicity.

Concluding remarks

The etiology of polyQ pathology has been elusive. Tedious decades-long efforts by many labs have revealed precise measurements of polyQ aggregation kinetics, atomistic details of the amyloid structure, a catalogue of proteotoxic candidates, and snapshots of the conformational preferences of disordered polyQ. What those efforts notably have not led to, are treatments. Here, we synthesized those insights to recognize that pathogenesis likely begins with a very specific, and exceptionally rare, conformational fluctuation, and that our relative ignorance of that event may be blocking progress against polyQ diseases. We therefore set out to characterize the polyQ amyloid nucleus. Deploying a technique we recently developed to do just that, we arrived at a model for the single most important molecular species governing polyQ pathogenicity in vivo. We found that the model can explain key

aspects of disease, such as length thresholds, kinetics of progression, and involvement of soluble amyloid-like species.

Acknowledgments

We thank Viet Man and Jonathon Russell for assistance with molecular simulations; Jeremy Schmit, Ammon Posey, Rohit Pappu, Wei-feng Xue, and Patrick van der Wel for helpful discussions; Alejandro Rodriguez Gama for assistance with figure preparation; and Megan Halfmann for generous ancillary support. We thank Alexander Alexandrov for plasmids encoding Q_4X repeats, early results from which stimulated this project's inception. This work was funded by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM130927 (to RH) and the Stowers Institute for Medical Research. A portion of this work was done to fulfill, in part, requirements for a PhD thesis research for T.S.K. as a student registered with the Open University, UK at the Stowers Institute for Medical Research Graduate School, USA. Original data underlying this manuscript can be accessed from the Stowers Original Data Repository at http://www.stowers.org/research/publications/libpb-1494

Materials and Methods

Plasmid and Strain Construction

ORFs were codon optimized for expression in *S. cerevisiae*, synthesized, and cloned into vector V08 (Khan et al., 2018) by Genscript (Piscataway, NJ, USA).

Plasmids expressing fusions to MBP were created by inserting yeast-optimized MBP-SUMO to the N-terminus of the protein of interest. Plasmids expressing fusions to µNS were created by insertion of yeast-optimized µNS to the C-terminus of mEos3.1. See Table S1 for all full-length protein sequences.

Isogenic yeast strains rhy1713 ([*PIN*$^+$]) and rhy1852 ([*pin*$^-$]) were previously described (Khan et al., 2018). Yeast strain rhy2068 was created by homology-directed integration at the *ho* locus of rhy1713, a construct expressing *K. lactis URA3* from a tTA-tetO7 promoter (Bellí et al., 1998).

Yeast strains rhy2956 (MBP-32:Q_3S), rhy2957 (MBP-60:Q_3S), and rhy2958 (MBP-60:Q) were created by homologous recombination-mediated replacement of the *K. lactis URA3* ORF in rhy2068 with a PCR product of the coding sequence (minus mEos3.1) from plasmids rhx3735, rhx3736, and rhx3788, respectively.

Yeast Preparation for DAmFRET, SDD-AGE and FACS

The yeast strains were transformed using a standard lithium acetate protocol with plasmids encoding the sequence to be tested as a fusion to the indicated tags (Table S1) under the control of the *GAL1* promoter.

Individual colonies were picked and incubated in 200µL of a standard synthetic media containing 2% dextrose (SD -ura) overnight while shaking on a Heidolph Titramax-1000 at 1000rpm at 30°C. Following overnight growth, cells were washed twice with sterile water and resuspended in a synthetic induction media containing 2% galactose (SGal -ura). Cells were induced for 16 hours while shaking before being resuspended in fresh 2% SGal -ura for 4 hours to reduce autofluorescence. For DAmFRET, 75 µL of cells were then transferred to a 384 well plate for analysis on the cytometer.

DAmFRET Cytometric Assay

Following 16 hours of total induction, cells were re-arrayed from a 96 well plate to a 384 well plate and photoconverted while shaking at 800rpm for 25 minutes using an OmniCure S1000 fitted with a 320-500 nm (violet) filter and a beam collimator (Exfo), positioned 45 cm above the plate. Cells were then transferred to a.Bio-Rad ZE5 cell analyzer.

For screening flow cytometry on the ZE5, plates were run in high-throughput mode (10 uL/well, flow speed 1.8) on the Bio-Rad ZE5 with Propel automation. Donor and FRET signals were collected from a 488 nm laser set to 100 mW, with voltages 351 and 370 respectively, into 525/35 and 593/52 filters. Acceptor signal was collected from a 561 nm laser set to 50 mW with voltage at 525, into a 589/15

filter. Autofluorescence was collected from a 405 nm laser at 100 mW and voltage 450, into a 460/22 filter.

Compensation was performed manually, collecting files for non-photoconverted mEos3.1 for pure donor fluorescence and dsRed2 to represent acceptor signal as it has a similar spectrum to the red form of mEos3.1. FRET is compensated only in the direction of donor and acceptor fluorescence out of FRET channels, as there is not a pure FRET control.

Imaging flow cytometry was conducted as in previous work (Khan et. al 2018).

DAmFRET Automated Analysis

FCS 3.1 files resulting from assay were gated using an automated R-script running in flowCore. Prior to gating, the forward scatter (FS00.A, FS00.W, FS00.H), side scatter (SS02.A), donor fluorescence (FL03.A) and autofluorescence (FL17.A) channels were transformed using a logicle transform in R. Gating was then done using flowCore by sequentially gating for cells using FS00.A vs SS02.A then selecting for single cells using FS00.H vs FS00.W and finally selecting for expressing cells using FL03.A vs FL17.A.

Gating for cells was done using a rectangular gate with values of Xmin = 2.7, Xmax = 4.8, Ymin = 2.7, Ymax = 4.7). Gating for single cells was done using a rectangular gate with values of Xmin = 4.45, Xmax = 4.58, Ymin = 2.5, Ymax = 4.4. Gating of expressing cells was done using a polygon gate with x/y vertices of (1,0.1), (1.8, 2), (5, 2), (5,0.1). Cells falling within all of these gates were then exported as FCS3.0 files for further analysis.

The FCS files resulting from the autonomous gating in Step 1 were then utilized for the JAVA-based quantification of a curve similar to the analysis found in the original assay (Khan et al 2018). Specifically:

The quantification procedure first divides a defined negative control DAmFRET histogram into 64 logarithmically spaced bins across a pre-determined range large enough to accommodate all potential data sets. Then upper gate values were determined for each bin as the 99th percentile of the DamFRET distribution in that bin. For bins at very low and very high acceptor intensities, there are not enough cells to accurately calculate this gate value. As a result, for bins above the 99th acceptor percentile and bins below 2 million acceptor intensity units, the upper gate value was set to the value of the nearest valid bin. The upper gate profile was then smoothed by boxcar smoothing with a width of 5 bins and shifted upwards by 0.028 DAmFRET units to ensure that the negative control signal lies completely within the negative FRET gate. The lower gate values for all bins were set equal to -0.2 DamFRET units. For all samples, then, cells falling above this negative FRET gate can be said to contain assembled (FRET-positive) protein. A metric reporting the gross percentage of the expressing cells containing assembled proteins is therefore reported as fgate which is a unitless statistic between 0 and 1.

This gate is then applied to all DAmFRET plots to define cells containing proteins that are either positive (self-assembled) or negative (monomeric). In each of the 64 gates, the fraction of cells in the assembled population were plotted as a ratio to total cells in the gate.

These values were then fit to a Weibull function from which the statistics EC_{50}, δ and their respective errors (reported as the square of the residuals) were extracted (Khan et al. 2018). Mean δ and EC50 values are presented as the mean of four biological replicates with errors being represented by the standard deviation between replicates. We verified for all samples where conclusions are drawn from δ and/or EC50 that the Weibull fit closely approximated the raw DAmFRET plot.

In addition to these quantitations, we also utilized the mean raw AmFRET in each logarithmic window to detect even small changes in FRET between similar peptides agnostic to whether the assemblies were continuous or discontinuous.

Microscopy

Cells were imaged using a CSU-W1 spinning disc Ti2 microscope (Nikon) and visualized through a 100x Plan Apochromat objective (NA 1.45). mEos3.1 was excited at 488 nm and emission was collected for 50 ms per frame through a ET525/36M bandpass filter on to a Flash 4 camera (Hamamatsu). Full z stacks of all cells were acquired over ~15 µm total distance with z spacing of 0.2 µm. Transmitted light was collected at the middle of the z stack for reference. To quantify the total intensity of each cell, the z stacks were processed using Fiji (https://imagej.net/software/fiji/). Images were first converted to 32-bit and sum projected in Z. Regions of interest (ROIs) were hand drawn around each cell using the ellipse tool in Fiji on the transmitted light image. These ROIs were then used to measure the area, mean, standard deviation, and integrated density of each cell on the fluorescence channel. Each cell was then classified as being diffuse or punctate by calculating the coefficient of variance (CV, standard deviation divided by the square root of the mean intensity) of the fluorescence. Cells were only considered punctate if their CV was greater than 30. Cells from each category that had equivalent integrated densities were directly compared and the images were plotted on the same intensity scale. The volume of each cell was calculated by fitting the transmitted light hand drawn ROI to an ellipse. The cell was then assumed to be a symmetric ellipsoid with the parameters of the fit ellipse from Fiji. The volume was calculated by 4/3*pi*major*minor*minor, where major and minor are the major and minor axes of the Fiji ellipse fit to the hand drawn ellipse ROI. The volumes reported are the volumes of the 3D ellipsoids. Concentrations were calculated by dividing the integrated densities by the calculated volume in μm^3, yielding units of fluorescence per μm^3.

SDD-AGE

Semi-denaturing detergent agarose gel electrophoresis (SDD-AGE) was done as in (Khan et al., 2018). The gel was imaged directly using a GE Typhoon Imaging System using a 488 laser and 525(40) BP filter. Images were then loaded into ImageJ for contrast adjustment. Images were gaussian blurred with a radius of 1 and then background subtracted with a 200 pixel rolling ball. Representative samples were then cropped from the original image for emphasis. Dotted or solid lines denote where different regions of the same gel were aligned for comparison.

Bioinformatic Analysis

Comparative analyses were performed using default settings at the listed, public web servers as were available on April 1st, 2021. See Table S2 for further details.

Molecular Simulations

The simulations were carried out using the AMBER 20 package (Case et al., 2020) with ff14SB force field (Maier et al., 2015) and explicit TIP3P water model (Price and Brooks, 2004). The aggregates were placed in a cubic solvation box. In each aggregate, the minimum distance from the aggregate to the box boundary was set to be 12 nm to avoid self-interactions. An 8 Å cutoff was applied for the nonbonded interactions, and the particle mesh Ewald (PME) method (Essmann et al., 1995) was used to calculate the electrostatic interaction with cubic-spline interpolation and a grid spacing of approximately 1 Å. Once the box was set up, we performed a structural optimization with aggregates fixed, and on a second step allowed them to move, followed by a graduate heating procedure (NVT) with individual steps of 200 ps from 0 K to 300 K. Finally, the production runs were carried out using NPT Langevin dynamics with constant pressure of 1 atm at 300 K.

Figure legends

Figure 1. A cryptic pattern of glutamines governs amyloid nucleation.

A. Volumes of reaction vessels influence the kinetics of amyloid formation. The concentrations of amyloid nuclei are so low that amyloid formation is rate-limited by nucleation in the femtoliter volumes of individual yeast cells (right) but not in the microliter volumes of conventional in vitro kinetic assays (left).
B. DAmFRET provides a snapshot of the progress of amyloid formation as a function of expression level in each of several thousand cells at a single time point. DAmFRET profiles, showing the extent of self-assembly (AmFRET) as a function of protein expression level (AFU, arbitrary fluorescence units) of polyglutamine (60:Q) in the absence ([*pin*$^-$]) and presence ([*PIN*$^+$]) of endogenous amyloids of a different Q/N-rich protein. PolyQ has a large nucleation barrier in [*pin*$^-$] cells, as evidenced by the window of concentration with a bimodal distribution of AmFRET values, wherein cells either lack (no AmFRET) or contain (high AmFRET) amyloid. [*PIN*$^+$] eliminates the nucleation barrier, as evidenced by all cells shifting to the high AmFRET population. Insets show histograms of AmFRET values.
C. Scatter plot of the fraction of [*pin*$^-$] cells with amyloid versus the fraction of glutamines in the protein's sequence. Red dots denote repeats with an even number of glutamines followed by an asparagine. Shown are means +/- SD of biological quadruplicates (independent yeast transformants).
D. Bar plot of the fraction of [*pin*$^-$] cells with amyloid for Q_3X, Q_4X, and Q_5X sequences where X = N, G, A, S, or H. Shown are means +/- SD of biological quadruplicates. *Length of the QxH sequences is 75; all others are 60.

Figure 2. Unilaterally contiguous glutamines encode an exclusive steric zipper structure.

A. A view down the axis of a local segment of all-glutamine steric zipper, or "Q zipper", between two antiparallel two-stranded sheets. Only the top four internally facing side-chains are shown, to emphasize interdigitation and H-bonding between the terminal amides and the opposing backbone.
B. Schema showing side chain arrangements along a continuous beta strand for sequences composed of tandem repeats of Qs (blue) interrupted by single non-Q residues (X, red). Shading indicates contiguous stretches of Qs on one (blue) or both (yellow) sides of the strand. Note that bilateral contiguity is bounded by X on either side of the strand. Note also that the illustrated strands will not necessarily be continuous in the context of the nucleus; i.e. parts of the sequence in the 60 residue-long polypeptide must form loops that interconnect strand elements.
C. Schema of the tertiary contacts between two beta strands, as in a steric zipper. The zipper can be formed only from Q residues for repeats that have an odd number of Qs (e.g. Q_1X), but not for repeats with an even number of Qs (e.g. Q_2X).

D. Molecular simulations of model Q zippers formed by a pair of two-stranded antiparallel beta sheets, wherein non-Q residues (in red) face either inward or outward. The schema are oriented so the viewer is looking down the axis between two sheets. The zipper is stable for pure polyQ (QQQQQQQ, top simulation), or when substitutions face outward (QQQNQQQ, second simulation; and QNQNQNQ, fourth simulation), but not when substitutions face inward (QQQNQQQ, third simulation).
E. Snapshot from the uninterrupted Q zipper simulation, showing H-bonds (black arrows) between internal extended Q side chains and the opposing backbones.
F. Snapshot from the internally interrupted Q zipper simulation, illustrating that the side chain of N is too short to H-bond the opposing backbone. However, the N side chain is long enough to H-bond the opposing Q side chain (red arrow), thereby intercepting the side chain-backbone H-bond that would otherwise occur (dashed arrow) between that Q side chain and the backbone amide adjacent to the N. This leads to dissolution of the zipper.

Figure 3. Bilaterally contiguous glutamines encode shorter, laminated zippers.

A. Scatter plot of the fraction of [*pin*$^-$] cells with amyloid versus the fraction of glutamines in 60:Q_aN proteins, showing that for even values of *a* (red dots), nucleation only occurs for *a* ≥ 6. Shown are means +/- SD of quadruplicates.
B. Schematic of single Q zipper and laminated Q zipper architectures.
C. AmFRET values of [*pin*$^-$] cells containing amyloids of the indicated sequence variants. Shown are means +/- SD of the median AmFRET values of quadruplicates. *** $p < 0.001$, t-test.
D. SDD-AGE characterizing amyloid length distributions for mEos3.1-tagged 60:Q_3A, Q_4A and Q_5A polypeptides. Note that the lower boundary of the amyloid smear is lower for Q_3A and Q_5A relative to Q_4A and Sup35 PrD, a typical Q/N-rich amyloid. Data are representative of multiple experiments.
E. Fraction of [*PIN*$^+$] cells with amyloid for 60:Q_3N-derived sequences. Sequences are frame-shifted to give unilaterally contiguous lengths of 4, 6 or 8 Qs. Inset shows side-chain arrangements as in 3A. Shown are means +/- SD of quadruplicates.
F. Fraction of [*PIN*$^+$] or [*pin*$^-$] cells with amyloid for 60:Q_4N-derived sequences. Sequences are frame-shifted to give bilaterally contiguous lengths of 2, 3, or 4 Qs (i.e. on both sides of the beta strand). Inset shows side-chain arrangements as in 3A. Shown are means +/- SD of quadruplicates.

Figure 4. Q zipper nucleation occurs within monomers

A. Fraction of cells AmFRET-positive as a function of expression level, for [*pin*$^-$] cells expressing 60:$Q_{(1-5)}$ (left) or 60:$Q_{(6+)}$ (right). Note the plateau in 60:$Q_{(3,5)}$ profiles suggestive of inhibitory oligomerization.

B. Distribution of cytosolic concentrations (AFU/μm^3) of 60:Q-mEos3.1 expressed in [*pin*$^-$] cells either lacking or containing puncta, showing that the protein remains diffuse even when supersaturated relative to amyloid. Representative diffuse or punctate cells (N = 31 and 26, respectively) of equivalent total concentration are shown. Scale bar: 5 µm.
C. Fraction of cells with amyloid for the indicated proteins expressed with or without genetic fusion to µNS. Artificial condensation sufficed to drive nucleation for all cells expressing ASC PyD, which has an oligomeric nucleus. It likewise did so for 60:Q, 60:Q_3N, and 60:Q_7N specifically in [*PIN*$^+$] cells, whereas it inhibited their nucleation in [*pin*$^-$] cells. Shown are means +/- SD of triplicates. *** $p < 0.001$, **** $p < 0.0001$; t-test.

Figure 5. Q zipper nucleation occurs within monomers

A. Graph of spline fits of AmFRET values for 60:$Q_{(6+)}$N and the upper and lower populations of 60:Q_3N, in [*PIN*$^+$] cells, showing that $Q_{(6+)}$N but not Q_3N have appreciable AmFRET at low expression, that the $Q_{(6+)}$N AmFRET level is lower than that of the upper Q_3N population at low expression, and higher than that of the upper Q_3N population at high expression. The red dashed line denotes that the upper and lower traces of 60:Q_3N represent subpopulations of the same samples. Shown are means +/- SD of quadruplicates.
B. Fraction of [*pin*$^-$] cells with amyloid for the indicated variants either lacking or containing a genetic fusion to MBP.
C. Histograms of AmFRET values of 60:Q_7N at low expression (<$2x10^7$ AFU) showing the appearance of a mid-FRET population, indicative of single Q zipper amyloids, only when seeded by MBP-60:Q_3S.

Figure 6. Reaction coordinate diagrams for a bifurcated amyloid nucleation pathway

A. Reaction coordinate diagrams for heterogeneous amyloid nucleation, whereby the disordered monomer acquires a minimal Q zipper (red rectangle), whose subsequent oligomerization extends the Q zipper either by lamination or lengthening, to produce short zipper amyloid-like oligomers or long zipper amyloids. The former transition is promoted by density fluctuations and contralateral Qs. Providing a conformational template, in the form of [*PIN*$^+$] (bottom row), accelerates Q zipper nucleation and consequently amplifies these preferences.
B. Decreasing solubility by genetic fusion to µNS, does not accelerate Q zipper nucleation (top left), but does accelerate the maturation of a pre-existing Q zipper to amyloid, analogous to nucleation in [*PIN*$^+$] cells (bottom left). Increasing solubility by genetic fusion to MBP decelerates Q zipper lamination (top right), and this effect is most pronounced in [*PIN*$^+$] cells (bottom right).

Figure 7. The Q zipper nucleus exhibits the pathologic length threshold

A. Reaction coordinate diagram schematizing the relationship of polypeptide length to heterogeneous amyloid nucleation barriers. As length decreases below that required to nucleate intramolecularly, the nucleation barrier will increase because the limiting conformational fluctuation now becomes dependent on a density fluctuation.
B. Scatter plot of the concentration-dependence of nucleation (delta) versus the concentration at which the protein has nucleated in half of the cells (EC_{50}), obtained from Weibull fits of DAmFRET profiles of different length variants of Q_3S expressed in [*PIN*$^+$] cells. Shown are means +/- SD of quadruplicates.
C. Spline fits of DAmFRET data for polyQ length variants 35 and longer expressed in [*PIN*$^+$] cells.
D. Reaction coordinate diagram schematizing the relationship of polyQ length to the nucleating conformational fluctuation. The proteins are fused to μNS to eliminate density fluctuations from the nucleation barrier, and expressed in [*PIN*$^+$] cells to achieve detectable nucleation frequencies. The conformational energy barrier increases, and acquires a darker shade of red, with shorter lengths of polyQ.
E. Fraction of [*PIN*$^+$] cells with amyloid for the indicated polyQ length variants fused to μNS.

Figure 8. Structural model of the polyQ amyloid nucleus.

The hypothetical surface of free energy in the plane of intramolecular (Q zipper length, red axis) and intermolecular order (density, blue axis) in a supersaturated solution of polyQ molecules. Qualitative topological features of the landscape, but not absolute heights and positions, are as deduced herein. Naive monomers exist in a local energy minimum at maximum disorder, while mature amyloid exists in a global energy minimum with long laminated Q zippers (image from Boatz et al. 2020, with permission). The route from monomer to amyloid proceeds through either of two horizontal basins. The lower basin represents short Q zippers and features the nucleating AHA monomer at the shallow end. The basin descends gradually toward amyloid via lamination-driven oligomerization. The upper basin represents long Q zippers, and descends more steeply toward amyloid due to the increased intermolecular connectivity of long zippers. Transitions from the lower to upper basin involve high-energy lengthening of short Q zippers.

Supplemental Files

Table S1. **List of plasmids and sequences.**

Table S2. Amyloid predictor output.

Table S3. Weibull fit statistics for L:Q_3S variants.

Figure S1. Anatomy of an amyloid nucleation barrier.

A. Reaction coordinate diagram schematizing the energy barrier for forming a hypothetical homogeneous amyloid nucleus (blue circle with red square). Because amyloid involves a transition from disordered monomer (left) to ordered multimer (right), the nucleation barrier (purple) is a combination of high energy fluctuations in both density (blue) and conformation (red).
B. Increased local concentrations of the protein make the critical fluctuation in density, specifically, more favorable (gray arrows indicate reduction in ΔG), increasing the frequency of nucleation while making it more dependent on conformational fluctuations.
C. Conformational templates, such as amyloids of a different protein ([*PIN*$^+$]), make the critical fluctuations in conformation, specifically, more favorable, increasing the frequency of nucleation while making it more dependent on density fluctuations.

Figure S2.

A. DAmFRET plots of [*pin*$^-$] and [*PIN*$^+$] cells expressing sequence variants with greater than 50% N, showing negligible nucleation in the absence of a conformational template. Shown are representative plots from the same experiment.
B. DAmFRET plots of [*pin*$^-$] and [*PIN*$^+$] cells expressing polypeptides composed of tandem repeats of *a* Qs separated by an N for a total length of 60 residues. Shown are representative plots from the same experiment.The plots surrounded by a purple dashed box illustrate the striking sequence parity of de novo nucleation potential.
C. DAmFRET plots of [*pin*$^-$] cells expressing Q_3X, Q_4X, and Q_5X sequences where X = N, G, A, S, or H. H-containing repeat sequences are length 75; all others are length 60. Shown are representative plots from the same experiment.
D. DAmFRET plots of [*PIN*$^+$] cells expressing Q_3X, Q_4X, and Q_5X sequences where X = N, G, A, S, or H. H-containing repeat sequences are length 75; all others are length 60. Shown are representative plots from the same experiment

Figure S3.

A. Molecular simulations of model Q zippers formed by a pair of two-stranded antiparallel beta sheets, containing a single serine residue (QQQSQQQ) per strand. The structure is unstable when the S side chains face inward (top), but not when the S side chains face outward (bottom).
B. Simulations of model steric zippers formed by a pair of four-stranded parallel beta sheets, containing various numbers and orientations of non-Q residues. From top to bottom: one asparagine inside, one serine inside, multiple asparagines outside, multiple serines outside. The

structures proved unstable in all cases, indicating that amyloid fibers formed by $Q_{(<8)}X$ repeat polypeptides are unlikely to have a parallel beta sheet arrangement.

C. Schematic demonstrating how a polar clasp (red dashed line) would preclude Q zipper formation.

D. Schema and frequencies of polar clasps occurring between two unilaterally adjacent Q side chains (left) or a unilaterally adjacent N and Q side chain (right) within a QQQNQQQ peptide, simulated either with (top) or without (bottom) the backbone restrained in a beta conformation. The bar graphs show that polar clasps between Q and N occur less frequently than between Q and Q, indicating that the mechanism of Q zipper destabilization by N side chains cannot be attributed to polar clasps.

E. Schema and lifetimes of H-bonds between exterior stacked (axially adjacent) Q side chains (top) or N and Q side chains (bottom) in the Q zipper simulated in Fig. 2D, showing no difference in stabilities between axial H-bonds between Q and Q/N side chains.

F. As a consequence of the N side chain's interception of the opposing Q side chain's H-bond, the Q is no longer anchored in the outstretched configuration and sterically interferes with the ordering of adjacent Qs. This effect propagates through the zipper, resulting in its dissolution.

G. As for N, the side chain of S is too short to H-bond with the opposing main chain. Unlike for N, however, the S side chain is also too short to intercept the opposing Q side chain's H-bond, allowing the Q to H-bond (black arrow) the backbone amide adjacent to the S. Therefore, whereas Q zippers cannot accommodate internal N residues, they can accommodate sparse internal S residues.

Figure S4.

A. SDD-AGE of the indicated mEos3.1-tagged proteins expressed in both [*PIN*$^+$] and [*pin*$^-$] cells. Lysates were normalized by fluorescence to within 50% of each other prior to loading. Sample buffer was added to a final concentration of 2% either SDS (top) or sarkosyl (bottom). Sarkosyl is a milder detergent that allows for detection of amyloids that are denatured by SDS.

B. DAmFRET plots of [*pin*$^-$] cells expressing unilateral contiguity variants of the 60:Q_3N base sequence, showing that at least six unilaterally contiguous glutamines (see schematic) are required for *de novo* nucleation of single long Q zipper amyloid with low concentration-dependence.

C. DAmFRET plots of contralateral contiguity variants of the 60:Q_4N base sequence, showing that at least six bilaterally contiguous Qs are required for robust amyloid formation.

Figure S5.

A. DAmFRET plots of 60:Q_5N and 60:Q_8N acquired using imaging flow cytometry, showing gates at high expression for both high- or low-AmFRET populations. Insets show from left to right the distribution of donor, FRET, and acceptor fluorescence, respectively, in representative cells from each gate.

B. DAmFRET plots of [*pin*$^-$] cells expressing the indicated proteins with and without genetic fusion to µNS.
C. DAmFRET plots of [*PIN*$^-$] cells expressing the indicated proteins with and without genetic fusion to µNS.

Figure S6.

A. DAmFRET plots of [*PIN*$^+$] cells expressing ASC, with or without fusion to MBP. Right plot shows the fraction of cells FRET-positive and the corresponding Weibull fits for triplicates, with inset reporting statistics of the Weibull fits +/- error. MBP specifically increased the C_{min} and EC_{50} values.
B. DAmFRET plots of [*pin*$^-$] and [*PIN*$^+$] cells expressing the indicated sequences, with or without fusion to MBP.
C. DAmFRET plots of 60:Q_3N (top) or 60:Q_7N (bottom) in [*PIN*$^+$] cells either lacking or expressing non-fluorescent MBP-32:Q_3S or MBP-60:Q_3S in trans. 60:Q_3N nucleates at lower concentration specifically when seeded by MBP-60:Q_3S. Black arrow shows the appearance of a mid-FRET population for 60:Q_7N, indicative of single Q zipper amyloids, only when seeded by MBP-60Q_3S. Dashed boxes correspond to the histograms in Fig. 6D.

Figure S7.

A. DAmFRET plots (left) and corresponding output from automated Weibull fitting (middle and right) of the indicated Q_3S length variants expressed in [*PIN*$^+$] cells. Note that the Weibull fits in this case assume monomeric nucleation, i.e. the C_{min} parameter was constrained to 0, in order to tighten the confidence intervals for the other two parameters. The region outlined in yellow in the middle plots designates the non-FRET population used to distinguish whether nucleation has (outside) or has not (inside) occurred in each cell. The black trace in the right plots shows the fraction of cells nucleated in different concentration windows, and the blue trace shows the corresponding Weibull fit.
B. DAmFRET plots of polyQ length variants in [*PIN*$^+$] cells.
C. DAmFRET plots of polyQ length variants in [*pin*$^-$] cells.
D. Spline fits of DAmFRET data for polyQ lengths 8-35 expressed in [*PIN*$^+$] and [*pin*$^-$] cells. Only at 35:Q does AmFRET significantly increase relative to 8:Q ($p < 0.000001$; t-test), and only in [*PIN*$^+$] cells.
E. DAmFRET plots of µNS-fused polyQ length variants.

Figure S8. Minimal Q zipper conformers of polyQ compatible with our experimental data. Left columns contain a side view of the structures after equilibration, and a scheme of the topology of the beta strands as seen from the left end (cross, strand into the plane; dot, strand coming out of the plane).

Vertical dashed lines represent rows of hydrogen bonds within sheets (H-bonds involving side chains are not shown). Right columns show the time evolution of the structures, initial configurations are shown after energy minimization and equilibration. In all cases, the fibril axes are in the direction of the hydrogen bonds (vertical dashed lines).

A. One of the hypothetical folds consistent with all constraints identified for the polyQ amyloid nucleus. This fold differs from B in having the first and fourth strands unpaired and flanking either side of the central hairpin. The initial conformer in this simulation had 11 residues per strand. Note that the visualization package does not plot unpaired strands as ribbons.
B. Another hypothetical fold consistent with all constraints identified for the polyQ amyloid nucleus. This fold, dubbed AHA, differs from A in having the first and fourth strands paired and flanking only one side of the central hairpin. The initial conformer in this simulation had 8 residues per strand; the stable conformer (run for 1 □s) had 6 residues per strand.
C. Minimal monomeric subunit of single Q zipper amyloid. It comprises a beta hairpin and a flanking third extended strand. The initial conformer in this simulation had 11 residues per strand.
D. Axial homodimer of C, representing the smallest hypothetically possible Q zipper amyloid dimer.

Bibliography

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. doi:10.1016/j.cell.2009.02.044

Auer S, Meersman F, Dobson CM, Vendruscolo M. 2008. A generic mechanism of emergence of amyloid protofilaments from disordered oligomeric aggregates. *PLoS Comput Biol* **4**:e1000222. doi:10.1371/journal.pcbi.1000222

Babinchak WM, Surewicz WK. 2020. Liquid-Liquid Phase Separation and Its Mechanistic Role in Pathological Protein Aggregation. *J Mol Biol* **432**:1910–1925. doi:10.1016/j.jmb.2020.03.004

Barrera EE, Zonta F, Pantano S. 2021. Dissecting the role of glutamine in seeding peptide aggregation. *Comput Struct Biotechnol J* **19**:1595–1602. doi:10.1016/j.csbj.2021.02.014

Bellí G, Garí E, Piedrafita L, Aldea M, Herrero E. 1998. An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic Acids Res* **26**:942–947. doi:10.1093/nar/26.4.942

Bhattacharyya AM, Thakur AK, Wetzel R. 2005. polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction. *Proc Natl Acad Sci USA* **102**:15400–15405. doi:10.1073/pnas.0501651102

Boatz JC, Piretra T, Lasorsa A, Matlahov I, Conway JF, van der Wel PCA. 2020. Protofilament structure and supramolecular polymorphism of aggregated mutant huntingtin exon 1. *J Mol Biol* **432**:4722–4744. doi:10.1016/j.jmb.2020.06.021

Boke E, Ruer M, Wühr M, Coughlin M, Lemaitre R, Gygi SP, Alberti S, Drechsel D, Hyman AA, Mitchison TJ. 2016. Amyloid-like Self-Assembly of a Cellular Compartment. *Cell* **166**:637–650. doi:10.1016/j.cell.2016.06.051

Boncella AE, Shattuck JE, Cascarina SM, Paul KR, Baer MH, Fomicheva A, Lamb AK, Ross ED. 2020. Composition-based prediction and rational manipulation of prion-like domain recruitment to stress granules. *Proc Natl Acad Sci USA* **117**:5826–5835. doi:10.1073/pnas.1912723117

Book A, Guella I, Candido T, Brice A, Hattori N, Jeon B, Farrer MJ, SNCA Multiplication Investigators of the GEoPD Consortium. 2018. A Meta-Analysis of α-Synuclein Multiplication in Familial Parkinsonism. *Front Neurol* **9**:1021. doi:10.3389/fneur.2018.01021

Borcherds W, Bremer A, Borgia MB, Mittag T. 2021. How do intrinsically disordered protein regions encode a driving force for liquid-liquid phase separation? *Curr Opin Struct Biol* **67**:41–50. doi:10.1016/j.sbi.2020.09.004

Bracha D, Walls MT, Wei MT, Zhu L, Kurian M, Avalos JL, Toettcher JE, Brangwynne CP. 2018. Mapping local and global liquid phase behavior in living cells using photo-oligomerizable seeds. *Cell* **175**:1467-1480.e13. doi:10.1016/j.cell.2018.10.048

Bradley ME, Edskes HK, Hong JY, Wickner RB, Liebman SW. 2002. Interactions among prions and prion “strains” in yeast. *Proc Natl Acad Sci USA* **99 Suppl 4**:16392–16399. doi:10.1073/pnas.152330699

Brandariz-Nuñez A, Menaya-Vargas R, Benavente J, Martinez-Costas J. 2010. Avian reovirus microNS protein forms homo-oligomeric inclusions in a microtubule-independent fashion, which involves specific regions of its C-terminal domain. *J Virol* **84**:4289–4301. doi:10.1128/JVI.02534-09

Buchanan LE, Carr JK, Fluitt AM, Hoganson AJ, Moran SD, de Pablo JJ, Skinner JL, Zanni MT. 2014.

Structural motif of polyglutamine amyloid fibrils discerned with mixed-isotope infrared spectroscopy. *Proc Natl Acad Sci USA* **111**:5796–5801. doi:10.1073/pnas.1401587111

Buell AK. 2017. The Nucleation of Protein Aggregates - From Crystals to Amyloid Fibrils. *Int Rev Cell Mol Biol* **329**:187–226. doi:10.1016/bs.ircmb.2016.08.014

Cai X, Chen J, Xu H, Liu S, Jiang Q-X, Halfmann R, Chen ZJ. 2014. Prion-like polymerization underlies signal transduction in antiviral immune defense and inflammasome activation. *Cell* **156**:1207–1222. doi:10.1016/j.cell.2014.01.063

Camino JD, Gracia P, Cremades N. 2021. The role of water in the primary nucleation of protein amyloid aggregation. *Biophys Chem* **269**:106520. doi:10.1016/j.bpc.2020.106520

Case DA, Belfon K, Ben-Shalom I, Brozell SR, Cerutti D, Cheatham T, Cruzeiro VWD, Darden T, Duke RE, Giambasu G, Gilson M, Gohlke H, Götz A, Harris R, Izadi S, Измайлов СА, Kasavajhala K, Kovalenko A, Krasny R, Kurtzman T, Kollman PA. 2020. Amber 2020.

Charoenkwan P, Kanthawong S, Nantasenamat C, Hasan MM, Shoombuatong W. 2021. iAMY-SCM: Improved prediction and analysis of amyloid proteins using a scoring card method with propensity scores of dipeptides. *Genomics* **113**:689–698. doi:10.1016/j.ygeno.2020.09.065

Chen M, Tsai M, Zheng W, Wolynes PG. 2016. The aggregation free energy landscapes of polyglutamine repeats. *J Am Chem Soc* **138**:15197–15203. doi:10.1021/jacs.6b08665

Chen M, Wolynes PG. 2017. Aggregation landscapes of Huntingtin exon 1 protein fragments and the critical repeat length for the onset of Huntington's disease. *Proc Natl Acad Sci USA* **114**:4406–4411. doi:10.1073/pnas.1702237114

Chen S, Berthelier V, Yang W, Wetzel R. 2001. Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J Mol Biol* **311**:173–182. doi:10.1006/jmbi.2001.4850

Chen S, Ferrone FA, Wetzel R. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc Natl Acad Sci USA* **99**:11884–11889. doi:10.1073/pnas.182276099

Chiti F, Dobson CM. 2017. Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annu Rev Biochem* **86**:27–68. doi:10.1146/annurev-biochem-061516-045115

Chou PY, Fasman GD. 1977. β-turns in proteins. *J Mol Biol* **115**:135–175. doi:10.1016/0022-2836(77)90094-8

Clarke G, Collins RA, Leavitt BR, Andrews DF, Hayden MR, Lumsden CJ, McInnes RR. 2000. A one-hit model of cell death in inherited neuronal degenerations. *Nature* **406**:195–199. doi:10.1038/35018098

Colby DW, Cassady JP, Lin GC, Ingram VM, Wittrup KD. 2006. Stochastic kinetics of intracellular huntingtin aggregate formation. *Nat Chem Biol* **2**:319–323. doi:10.1038/nchembio792

Collinge J, Clarke AR. 2007. A general model of prion strains and their pathogenicity. *Science* **318**:930–936. doi:10.1126/science.1138718

Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV. 2006. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc Natl Acad Sci USA* **103**:16764–16769. doi:10.1073/pnas.0608175103

Crick SL, Ruff KM, Garai K, Frieden C, Pappu RV. 2013. Unmasking the roles of N- and C-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. *Proc Natl Acad Sci USA* **110**:20075–20080. doi:10.1073/pnas.1320626110

Derkatch IL, Bradley ME, Hong JY, Liebman SW. 2001. Prions affect the appearance of other prions: the story of [PIN(+)]. *Cell* **106**:171–182.

Dignon GL, Best RB, Mittal J. 2020. Biomolecular phase separation: from molecular driving forces to macroscopic properties. *Annu Rev Phys Chem* **71**:53–75. doi:10.1146/annurev-physchem-071819-113553

Duennwald ML, Jagadish S, Giorgini F, Muchowski PJ, Lindquist S. 2006a. A network of protein interactions determines polyglutamine toxicity. *Proc Natl Acad Sci USA* **103**:11051–11056. doi:10.1073/pnas.0604548103

Duennwald ML, Jagadish S, Muchowski PJ, Lindquist S. 2006b. Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc Natl Acad Sci USA* **103**:11045–11050. doi:10.1073/pnas.0604547103

Eisenberg DS, Sawaya MR. 2017. Structural studies of amyloid proteins at the molecular level. *Annu Rev Biochem* **86**:69–95. doi:10.1146/annurev-biochem-061516-045104

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995. A smooth particle mesh Ewald method. *J Chem Phys* **103**:8577. doi:10.1063/1.470117

Fisher RS, Jimenez RM, Soto E, Kalev D, Elbaum-Garfinkle S. 2021. An apparent core/shell architecture of polyQ aggregates in the aging Caenorhabditis elegans neuron. *Protein Sci* **30**:1482–1486. doi:10.1002/pro.4105

Fujiwara K, Toda H, Ikeguchi M. 2012. Dependence of α-helical and β-sheet amino acid propensities on the overall protein fold type. *BMC Struct Biol* **12**:18. doi:10.1186/1472-6807-12-18

Galaz-Montoya JG, Shahmoradian SH, Shen K, Frydman J, Chiu W. 2021. Cryo-electron tomography provides topological insights into mutant huntingtin exon 1 and polyQ aggregates. *Commun Biol* **4**:849. doi:10.1038/s42003-021-02360-2

Gallagher-Jones M, Glynn C, Boyer DR, Martynowycz MW, Hernandez E, Miao J, Zee C-T, Novikova IV, Goldschmidt L, McFarlane HT, Helguera GF, Evans JE, Sawaya MR, Cascio D, Eisenberg DS, Gonen T, Rodriguez JA. 2018. Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp. *Nat Struct Mol Biol* **25**:131–134. doi:10.1038/s41594-017-0018-0

Gao A, Shrinivas K, Lepeudry P, Suzuki HI, Sharp PA, Chakraborty AK. 2018. Evolution of weak cooperative interactions for biological specificity. *Proc Natl Acad Sci USA* **115**:E11053–E11060. doi:10.1073/pnas.1815912115

Halfmann R, Alberti S, Krishnan R, Lyle N, O'Donnell CW, King OD, Berger B, Pappu RV, Lindquist S. 2011. Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins. *Mol Cell* **43**:72–84. doi:10.1016/j.molcel.2011.05.013

Halfmann R, Lindquist S. 2008. Screening for amyloid aggregation by Semi-Denaturing Detergent-Agarose Gel Electrophoresis. *J Vis Exp* **17**:e838. doi:10.3791/838

Hennetin J, Jullian B, Steven AC, Kajava AV. 2006. Standard conformations of beta-arches in beta-solenoid proteins. *J Mol Biol* **358**:1094–1105. doi:10.1016/j.jmb.2006.02.039

Hervas R, Rau MJ, Park Y, Zhang W, Murzin AG, Fitzpatrick JAJ, Scheres SHW, Si K. 2020. Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in Drosophila. *Science* **367**:1230–1234. doi:10.1126/science.aba3526

Hoop CL, Lin H-K, Kar K, Magyarfalvi G, Lamley JM, Boatz JC, Mandal A, Lewandowski JR, Wetzel R, van der Wel PCA. 2016. Huntingtin exon 1 fibrils feature an interdigitated β-hairpin-based polyglutamine core. *Proc Natl Acad Sci USA* **113**:1546–1551. doi:10.1073/pnas.1521933113

Hsieh M-C, Liang C, Mehta AK, Lynn DG, Grover MA. 2017. Multistep conformation selection in amyloid assembly. *J Am Chem Soc* **139**:17007–17010. doi:10.1021/jacs.7b09362

Huang C, Wagner-Valladolid S, Stephens AD, Jung R, Poudel C, Sinnige T, Lechler MC, Schlörit N, Lu

M, Laine RF, Michel CH, Vendruscolo M, Kaminski CF, Kaminski Schierle GS, David DC. 2019. Intrinsically aggregation-prone proteins form amyloid-like aggregates and contribute to tissue aging in Caenorhabditis elegans. *elife* **8**. doi:10.7554/eLife.43059

Jiang Y, Di Gregorio SE, Duennwald ML, Lajoie P. 2017. Polyglutamine toxicity in yeast uncovers phenotypic variations between different fluorescent protein fusions. *Traffic* **18**:58–70. doi:10.1111/tra.12453

Kajava AV, Baxa U, Steven AC. 2010. Beta arcades: recurring motifs in naturally occurring and disease-related amyloid fibrils. *FASEB J* **24**:1311–1319. doi:10.1096/fj.09-145979

Kakkar V, Månsson C, de Mattos EP, Bergink S, van der Zwaag M, van Waarde MAWH, Kloosterhuis NJ, Melki R, van Cruchten RTP, Al-Karadaghi S, Arosio P, Dobson CM, Knowles TPJ, Bates GP, van Deursen JM, Linse S, van de Sluis B, Emanuelsson C, Kampinga HH. 2016. The S/T-Rich Motif in the DNAJB6 Chaperone Delays Polyglutamine Aggregation and the Onset of Disease in a Mouse Model. *Mol Cell* **62**:272–283. doi:10.1016/j.molcel.2016.03.017

Kang H, Vázquez FX, Zhang L, Das P, Toledo-Sherman L, Luan B, Levitt M, Zhou R. 2017. Emerging β-Sheet Rich Conformations in Supercompact Huntingtin Exon-1 Mutant Structures. *J Am Chem Soc* **139**:8820–8827. doi:10.1021/jacs.7b00838

Kar K, Hoop CL, Drombosky KW, Baker MA, Kodali R, Arduini I, van der Wel PCA, Horne WS, Wetzel R. 2013. β-hairpin-mediated nucleation of polyglutamine amyloid formation. *J Mol Biol* **425**:1183–1197. doi:10.1016/j.jmb.2013.01.016

Kar K, Jayaraman M, Sahoo B, Kodali R, Wetzel R. 2011. Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat Struct Mol Biol* **18**:328–336. doi:10.1038/nsmb.1992

Keefer KM, Stein KC, True HL. 2017. Heterologous prion-forming proteins interact to cross-seed aggregation in Saccharomyces cerevisiae. *Sci Rep* **7**:5853. doi:10.1038/s41598-017-05829-5

Keresztes L, Szögi E, Varga B, Farkas V, Perczel A, Grolmusz V. 2021. The budapest amyloid predictor and its applications. *Biomolecules* **11**. doi:10.3390/biom11040500

Khan T, Kandola TS, Wu J, Venkatesan S, Ketter E, Lange JJ, Rodríguez Gama A, Box A, Unruh JR, Cook M, Halfmann R. 2018. Quantifying nucleation in vivo reveals the physical basis of prion-like phase behavior. *Mol Cell* **71**:155-168.e7. doi:10.1016/j.molcel.2018.06.016

Kim G, Gautier O, Tassoni-Tsuchida E, Ma XR, Gitler AD. 2020. ALS genetics: gains, losses, and implications for future therapies. *Neuron* **108**:822–842. doi:10.1016/j.neuron.2020.08.022

Kim YE, Hosp F, Frottin F, Ge H, Mann M, Hayer-Hartl M, Hartl FU. 2016. Soluble Oligomers of PolyQ-Expanded Huntingtin Target a Multiplicity of Key Cellular Factors. *Mol Cell* **63**:951–964. doi:10.1016/j.molcel.2016.07.022

Knowles TPJ, Waudby CA, Devlin GL, Cohen SIA, Aguzzi A, Vendruscolo M, Terentjev EM, Welland ME, Dobson CM. 2009. An analytical solution to the kinetics of breakable filament assembly. *Science* **326**:1533–1537. doi:10.1126/science.1178250

Kroschwald S, Maharana S, Mateju D, Malinovska L, Nüske E, Poser I, Richter D, Alberti S. 2015. Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. *elife* **4**:e06807. doi:10.7554/eLife.06807

Kryndushkin D, Pripuzova N, Burnett BG, Shewmaker F. 2013. Non-targeted identification of prions and amyloid-forming proteins from yeast and mammalian cells. *J Biol Chem* **288**:27100–27111. doi:10.1074/jbc.M113.485359

Kryndushkin DS, Alexandrov IM, Ter-Avanesyan MD, Kushnirov VV. 2003. Yeast [PSI+] prion

aggregates are formed by small Sup35 polymers fragmented by Hsp104. *J Biol Chem* **278**:49636–49643. doi:10.1074/jbc.M307996200

Lee J-M, Correia K, Loupe J, Kim K-H, Barker D, Hong EP, Chao MJ, Long JD, Lucente D, Vonsattel JPG, Pinto RM, Abu Elneel K, Ramos EM, Mysore JS, Gillis T, Wheeler VC, MacDonald ME, Gusella JF, McAllister B, Massey T, Myers RH. 2019. CAG repeat not polyglutamine length determines timing of huntington's disease onset. *Cell* **178**:887-900.e14. doi:10.1016/j.cell.2019.06.036

Legleiter J, Mitchell E, Lotz GP, Sapp E, Ng C, DiFiglia M, Thompson LM, Muchowski PJ. 2010. Mutant huntingtin fragments form oligomers in a polyglutamine length-dependent manner in vitro and in vivo. *J Biol Chem* **285**:14777–14790. doi:10.1074/jbc.M109.093708

Leitman J, Ulrich Hartl F, Lederkremer GZ. 2013. Soluble forms of polyQ-expanded huntingtin rather than large aggregates cause endoplasmic reticulum stress. *Nat Commun* **4**:2753. doi:10.1038/ncomms3753

Levin A, Mason TO, Adler-Abramovich L, Buell AK, Meisl G, Galvagnion C, Bram Y, Stratford SA, Dobson CM, Knowles TPJ, Gazit E. 2014. Ostwald's rule of stages governs structural transitions and morphology of dipeptide supramolecular polymers. *Nat Commun* **5**:5219. doi:10.1038/ncomms6219

Liang C, Hsieh M-C, Li NX, Lynn DG. 2018. Conformational evolution of polymorphic amyloid assemblies. *Curr Opin Struct Biol* **51**:135–140. doi:10.1016/j.sbi.2018.04.004

Lieberman AP, Shakkottai VG, Albin RL. 2019. Polyglutamine repeats in neurodegenerative diseases. *Annu Rev Pathol* **14**:1–27. doi:10.1146/annurev-pathmechdis-012418-012857

Linsley JW, Tripathi A, Epstein I, Schmunk G, Mount E, Campioni M, Oza V, Barch M, Javaherian A, Nowakowski TJ, Samsi S, Finkbeiner S. 2019. Automated four-dimensional long term imaging enables single cell tracking within organotypic brain slices to study neurodevelopment and degeneration. *Commun Biol* **2**:155. doi:10.1038/s42003-019-0411-9

Lin H-K, Boatz JC, Krabbendam IE, Kodali R, Hou Z, Wetzel R, Dolga AM, Poirier MA, van der Wel PCA. 2017. Fibril polymorphism affects immobilized non-amyloid flanking domains of huntingtin exon1 rather than its polyglutamine core. *Nat Commun* **8**:15462. doi:10.1038/ncomms15462

Li J, Browning S, Mahal SP, Oelschlegel AM, Weissmann C. 2010. Darwinian evolution of prions in cell culture. *Science* **327**:869–872. doi:10.1126/science.1183218

Lu A, Magupalli VG, Ruan J, Yin Q, Atianand MK, Vos MR, Schröder GF, Fitzgerald KA, Wu H, Egelman EH. 2014. Unified polymerization mechanism for the assembly of ASC-dependent inflammasomes. *Cell* **156**:1193–1206. doi:10.1016/j.cell.2014.02.008

Lu B, Palacino J. 2013. A novel human embryonic stem cell-derived Huntington's disease neuronal model exhibits mutant huntingtin (mHTT) aggregates and soluble mHTT-dependent neurodegeneration. *FASEB J* **27**:1820–1829. doi:10.1096/fj.12-219220

Macdonald D, Tessari MA, Boogaard I, Smith M, Pulli K, Szynol A, Albertus F, Lamers MBAC, Dijkstra S, Kordt D, Reindl W, Herrmann F, McAllister G, Fischer DF, Munoz-Sanjuan I. 2014. Quantification assays for total and polyglutamine-expanded huntingtin proteins. *PLoS ONE* **9**:e96854. doi:10.1371/journal.pone.0096854

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**:3696–3713. doi:10.1021/acs.jctc.5b00255

Maji SK, Perrin MH, Sawaya MR, Jessberger S, Vadodaria K, Rissman RA, Singru PS, Nilsson KPR,

Simon R, Schubert D, Eisenberg D, Rivier J, Sawchenko P, Vale W, Riek R. 2009. Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **325**:328–332. doi:10.1126/science.1173155

Majumdar A, Cesario WC, White-Grindley E, Jiang H, Ren F, Khan MR, Li L, Choi EM-L, Kannan K, Guo F, Unruh J, Slaughter B, Si K. 2012. Critical role of amyloid-like oligomers of Drosophila Orb2 in the persistence of memory. *Cell* **148**:515–529. doi:10.1016/j.cell.2012.01.004

Man VH, Roland C, Sagui C. 2015. Structural determinants of polyglutamine protofibrils and crystallites. *ACS Chem Neurosci* **6**:632–645. doi:10.1021/cn500358g

Matlahov I, van der Wel PC. 2019. Conformational studies of pathogenic expanded polyglutamine protein deposits from Huntington's disease. *Exp Biol Med (Maywood)* **244**:1584–1595. doi:10.1177/1535370219856620

Michaels TCT, Liu LX, Meisl G, Knowles TPJ. 2017. Physical principles of filamentous protein self-assembly kinetics. *J Phys Condens Matter* **29**:153002. doi:10.1088/1361-648X/aa5f10

Mier P, Elena-Real C, Urbanek A, Bernadó P, Andrade-Navarro MA. 2020. The importance of definitions in the study of polyQ regions: A tale of thresholds, impurities and sequence context. *Comput Struct Biotechnol J* **18**:306–313. doi:10.1016/j.csbj.2020.01.012

Miettinen MS, Knecht V, Monticelli L, Ignatova Z. 2012. Assessing polyglutamine conformation in the nucleating event by molecular dynamics simulations. *J Phys Chem B* **116**:10259–10265. doi:10.1021/jp305065c

Miettinen MS, Monticelli L, Nedumpully-Govindan P, Knecht V, Ignatova Z. 2014. Stable polyglutamine dimers can contain β-hairpins with interdigitated side chains-but not α-helices, β-nanotubes, β-pseudohelices, or steric zippers. *Biophys J* **106**:1721–1728. doi:10.1016/j.bpj.2014.02.027

Moradi M, Babin V, Roland C, Sagui C. 2012. Are long-range structural correlations behind the aggregration phenomena of polyglutamine diseases? *PLoS Comput Biol* **8**:e1002501. doi:10.1371/journal.pcbi.1002501

Nacar C. 2020. Propensities of amino acid pairings in secondary structure of globular proteins. *Protein J* **39**:21–32. doi:10.1007/s10930-020-09880-6

Newcombe EA, Ruff KM, Sethi A, Ormsby AR, Ramdzan YM, Fox A, Purcell AW, Gooley PR, Pappu RV, Hatters DM. 2018. Tadpole-like Conformations of Huntingtin Exon 1 Are Characterized by Conformational Heterogeneity that Persists regardless of Polyglutamine Length. *J Mol Biol* **430**:1442–1458. doi:10.1016/j.jmb.2018.03.031

Nielsen MM, Andersen KK, Westh P, Otzen DE. 2007. Unfolding of beta-sheet proteins in SDS. *Biophys J* **92**:3674–3685. doi:10.1529/biophysj.106.101238

Nizhnikov AA, Alexandrov AI, Ryzhova TA, Mitkevich OV, Dergalev AA, Ter-Avanesyan MD, Galkin AP. 2014. Proteomic screening for amyloid proteins. *PLoS ONE* **9**:e116003. doi:10.1371/journal.pone.0116003

O'Carroll A, Coyle J, Gambin Y. 2020. Prions and Prion-like assemblies in neurodegeneration and immunity: The emergence of universal mechanisms across health and disease. *Semin Cell Dev Biol* **99**:115–130. doi:10.1016/j.semcdb.2019.11.012

Olshina MA, Angley LM, Ramdzan YM, Tang J, Bailey MF, Hill AF, Hatters DM. 2010. Tracking mutant huntingtin aggregation kinetics in cells reveals three major populations that include an invariant oligomer pool. *J Biol Chem* **285**:21807–21816. doi:10.1074/jbc.M109.084434

Peskett TR, Rau F, O'Driscoll J, Patani R, Lowe AR, Saibil HR. 2018. A liquid to solid phase transition underlying pathological huntingtin exon1 aggregation. *Mol Cell* **70**:588-601.e6.

doi:10.1016/j.molcel.2018.04.007
Phan TM, Schmit JD. 2020. Conformational entropy limits the transition from nucleation to elongation in amyloid aggregation. *BioRxiv*. doi:10.1101/2020.06.22.165423
Phan TTM, Schmit JD. 2020. Thermodynamics of huntingtin aggregation. *Biophys J* **118**:2989–2996. doi:10.1016/j.bpj.2020.05.013
Posey AE, Ruff KM, Harmon TS, Crick SL, Li A, Diamond MI, Pappu RV. 2018. Profilin reduces aggregation and phase separation of huntingtin N-terminal fragments by preferentially binding to soluble monomers and oligomers. *J Biol Chem* **293**:3734–3746. doi:10.1074/jbc.RA117.000357
Posey AE, Ruff KM, Lalmansingh JM, Kandola TS, Lange JJ, Halfmann R, Pappu RV. 2021. Mechanistic inferences from analysis of measurements of protein phase transitions in live cells. *J Mol Biol* **433**:166848. doi:10.1016/j.jmb.2021.166848
Prabakaran R, Rawat P, Thangakani AM, Kumar S, Gromiha MM. 2021. Protein aggregation: in silico algorithms and applications. *Biophys Rev* **13**:71–89. doi:10.1007/s12551-021-00778-w
Price DJ, Brooks CL. 2004. A modified TIP3P water potential for simulation with Ewald summation. *J Chem Phys* **121**:10096–10103. doi:10.1063/1.1808117
Punihaole D, Workman RJ, Hong Z, Madura JD, Asher SA. 2016. Polyglutamine Fibrils: New Insights into Antiparallel β-Sheet Conformational Preference and Side Chain Structure. *J Phys Chem B* **120**:3012–3026. doi:10.1021/acs.jpcb.5b11380
Raran-Kurussi S, Waugh DS. 2012. The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated. *PLoS ONE* **7**:e49589. doi:10.1371/journal.pone.0049589
Rodríguez Gama A, Miller T, Halfmann R. 2021. Mechanics of a molecular mousetrap-nucleation-limited innate immune signaling. *Biophys J* **120**:1150–1160. doi:10.1016/j.bpj.2021.01.007
Sanders DW, Kaufman SK, DeVos SL, Sharma AM, Mirbaha H, Li A, Barker SJ, Foley AC, Thorpe JR, Serpell LC, Miller TM, Grinberg LT, Seeley WW, Diamond MI. 2014. Distinct tau prion strains propagate in cells and mice and define different tauopathies. *Neuron* **82**:1271–1288. doi:10.1016/j.neuron.2014.04.047
Sathasivam K, Lane A, Legleiter J, Warley A, Woodman B, Finkbeiner S, Paganetti P, Muchowski PJ, Wilson S, Bates GP. 2010. Identical oligomeric and fibrillar structures captured from the brains of R6/2 and knock-in mouse models of Huntington's disease. *Hum Mol Genet* **19**:65–78. doi:10.1093/hmg/ddp467
Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apostol MI, Thompson MJ, Balbirnie M, Wiltzius JJW, McFarlane HT, Madsen AØ, Riekel C, Eisenberg D. 2007. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**:453–457. doi:10.1038/nature05695
Schmitz AM, Morrison MF, Agunwamba AO, Nibert ML, Lesser CF. 2009. Protein interaction platforms: visualization of interacting proteins in yeast. *Nat Methods* **6**:500–502. doi:10.1038/nmeth.1337
Schneider R, Schumacher MC, Mueller H, Nand D, Klaukien V, Heise H, Riedel D, Wolf G, Behrmann E, Raunser S, Seidel R, Engelhard M, Baldus M. 2011. Structural characterization of polyglutamine fibrils by solid-state NMR spectroscopy. *J Mol Biol* **412**:121–136. doi:10.1016/j.jmb.2011.06.045
Selkoe DJ, Hardy J. 2016. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol Med* **8**:595–608. doi:10.15252/emmm.201606210
Serio TR, Cashikar AG, Kowal AS, Sawicki GJ, Moslehi JJ, Serpell L, Arnsdorf MF, Lindquist SL. 2000. Nucleated conformational conversion and the replication of conformational information by a prion

determinant. *Science* **289**:1317–1321. doi:10.1126/science.289.5483.1317
Serio TR. 2018. [PIN+]ing down the mechanism of prion appearance. *FEMS Yeast Res* **18**. doi:10.1093/femsyr/foy026
Sikorski P, Atkins E. 2005. New model for crystalline polyglutamine assemblies and their connection with amyloid fibrils. *Biomacromolecules* **6**:425–432. doi:10.1021/bm0494388
Sil TB, Sahoo B, Bera SC, Garai K. 2018. Quantitative characterization of metastability and heterogeneity of amyloid aggregates. *Biophys J* **114**:800–811. doi:10.1016/j.bpj.2017.12.023
Simm S, Einloft J, Mirus O, Schleiff E. 2016. 50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification. *Biol Res* **49**:31. doi:10.1186/s40659-016-0092-5
Sinnige T, Meisl G, Michaels TCT, Vendruscolo M, Knowles TPJ, Morimoto RI. 2021. Kinetic analysis reveals that independent nucleation events determine the progression of polyglutamine aggregation in C. elegans. *Proc Natl Acad Sci USA* **118**. doi:10.1073/pnas.2021888118
Stanger HE, Syud FA, Espinosa JF, Giriat I, Muir T, Gellman SH. 2001. Length-dependent stability and strand length limits in antiparallel beta -sheet secondary structure. *Proc Natl Acad Sci USA* **98**:12015–12020. doi:10.1073/pnas.211536998
Strodel B. 2021. Amyloid aggregation simulations: challenges, advances and perspectives. *Curr Opin Struct Biol* **67**:145–152. doi:10.1016/j.sbi.2020.10.019
Takahashi T, Kikuchi S, Katada S, Nagai Y, Nishizawa M, Onodera O. 2008. Soluble polyglutamine oligomers formed prior to inclusion body formation are cytotoxic. *Hum Mol Genet* **17**:345–356. doi:10.1093/hmg/ddm311
Tanaka M, Collins SR, Toyama BH, Weissman JS. 2006. The physical basis of how prion conformations determine strain phenotypes. *Nature* **442**:585–589. doi:10.1038/nature04922
Thakur AK, Wetzel R. 2002. Mutational analysis of the structural organization of polyglutamine aggregates. *Proc Natl Acad Sci USA* **99**:17014–17019. doi:10.1073/pnas.252523899
Törnquist M, Michaels TCT, Sanagavarapu K, Yang X, Meisl G, Cohen SIA, Knowles TPJ, Linse S. 2018. Secondary nucleation in amyloid formation. *Chem Commun* **54**:8667–8684. doi:10.1039/c8cc02204f
Venkatesan S, Kandola TS, Rodríguez-Gama A, Box A, Halfmann R. 2019. Detecting and characterizing protein self-assembly in vivo by flow cytometry. *J Vis Exp*. doi:10.3791/59577
Verges KJ, Smith MH, Toyama BH, Weissman JS. 2011. Strain conformation, primary structure and the propagation of the yeast prion [PSI+]. *Nat Struct Mol Biol* **18**:493–499. doi:10.1038/nsmb.2030
Villali J, Dark J, Brechtel TM, Pei F, Sindi SS, Serio TR. 2020. Nucleation seed size determines amyloid clearance and establishes a barrier to prion appearance in yeast. *Nat Struct Mol Biol* **27**:540–549. doi:10.1038/s41594-020-0416-6
Vitalis A, Lyle N, Pappu RV. 2009. Thermodynamics of beta-sheet formation in polyglutamine. *Biophys J* **97**:303–311. doi:10.1016/j.bpj.2009.05.003
Vitalis A, Pappu RV. 2011. Assessing the contribution of heterogeneous distributions of oligomers to aggregation mechanisms of polyglutamine peptides. *Biophys Chem* **159**:14–23. doi:10.1016/j.bpc.2011.04.006
Vitalis A, Wang X, Pappu RV. 2007. Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories. *Biophys J* **93**:1923–1937. doi:10.1529/biophysj.107.110080
Vogler TO, Wheeler JR, Nguyen ED, Hughes MP, Britson KA, Lester E, Rao B, Betta ND, Whitney ON, Ewachiw TE, Gomes E, Shorter J, Lloyd TE, Eisenberg DS, Taylor JP, Johnson AM, Olwin BB,

Parker R. 2018. TDP-43 and RNA form amyloid-like myo-granules in regenerating muscle. *Nature* **563**:508–513. doi:10.1038/s41586-018-0665-2

Walters RH, Jacobson KH, Pedersen JA, Murphy RM. 2012. Elongation kinetics of polyglutamine peptide fibrils: a quartz crystal microbalance with dissipation study. *J Mol Biol* **421**:329–347. doi:10.1016/j.jmb.2012.03.017

Walters RH, Murphy RM. 2009. Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. *J Mol Biol* **393**:978–992. doi:10.1016/j.jmb.2009.08.034

Wang J, Choi J-M, Holehouse AS, Lee HO, Zhang X, Jahnel M, Maharana S, Lemaitre R, Pozniakovsky A, Drechsel D, Poser I, Pappu RV, Alberti S, Hyman AA. 2018. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**:688-699.e16. doi:10.1016/j.cell.2018.06.006

Wang X, Vitalis A, Wyczalkowski MA, Pappu RV. 2006. Characterizing the conformational ensemble of monomeric polyglutamine. *Proteins* **63**:297–311. doi:10.1002/prot.20761

Wan Q, Mouton SN, Veenhoff LM, Boersma AJ. 2021. A precise and general FRET-based method for monitoring structural transitions in protein self-organization. *BioRxiv*. doi:10.1101/2021.02.25.432866

Wetzel R. 2020. Exploding the Repeat Length Paradigm while Exploring Amyloid Toxicity in Huntington's Disease. *Acc Chem Res* **53**:2347–2357. doi:10.1021/acs.accounts.0c00450

Wexler NS, Young AB, Tanzi RE, Travers H, Starosta-Rubinstein S, Penney JB, Snodgrass SR, Shoulson I, Gomez F, Ramos Arroyo MA. 1987. Homozygotes for Huntington's disease. *Nature* **326**:194–197. doi:10.1038/326194a0

Williamson TE, Vitalis A, Crick SL, Pappu RV. 2010. Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin. *J Mol Biol* **396**:1295–1309. doi:10.1016/j.jmb.2009.12.017

Xing W, Muhlrad D, Parker R, Rosen MK. 2018. A quantitative inventory of yeast P body proteins reveals principles of compositional specificity. *BioRxiv*. doi:10.1101/489658

Yamaguchi K-I, Takahashi S, Kawai T, Naiki H, Goto Y. 2005. Seeding-dependent propagation and maturation of amyloid fibril conformation. *J Mol Biol* **352**:952–960. doi:10.1016/j.jmb.2005.07.061

Yang J, Yang X. 2020. Phase transition of huntingtin: factors and pathological relevance. *Front Genet* **11**:754. doi:10.3389/fgene.2020.00754

Zanjani AAH, Reynolds NP, Zhang A, Schilling T, Mezzenga R, Berryman JT. 2020. Amyloid evolution: antiparallel replaced by parallel. *Biophys J* **118**:2526–2536. doi:10.1016/j.bpj.2020.03.023

Zhang L, Schmit JD. 2016. Pseudo-one-dimensional nucleation in dilute polymer solutions. *Phys Rev E* **93**:060401. doi:10.1103/PhysRevE.93.060401

Zhang Y, Man VH, Roland C, Sagui C. 2016. Amyloid Properties of Asparagine and Glutamine in Prion-like Proteins. *ACS Chem Neurosci* **7**:576–587. doi:10.1021/acschemneuro.5b00337

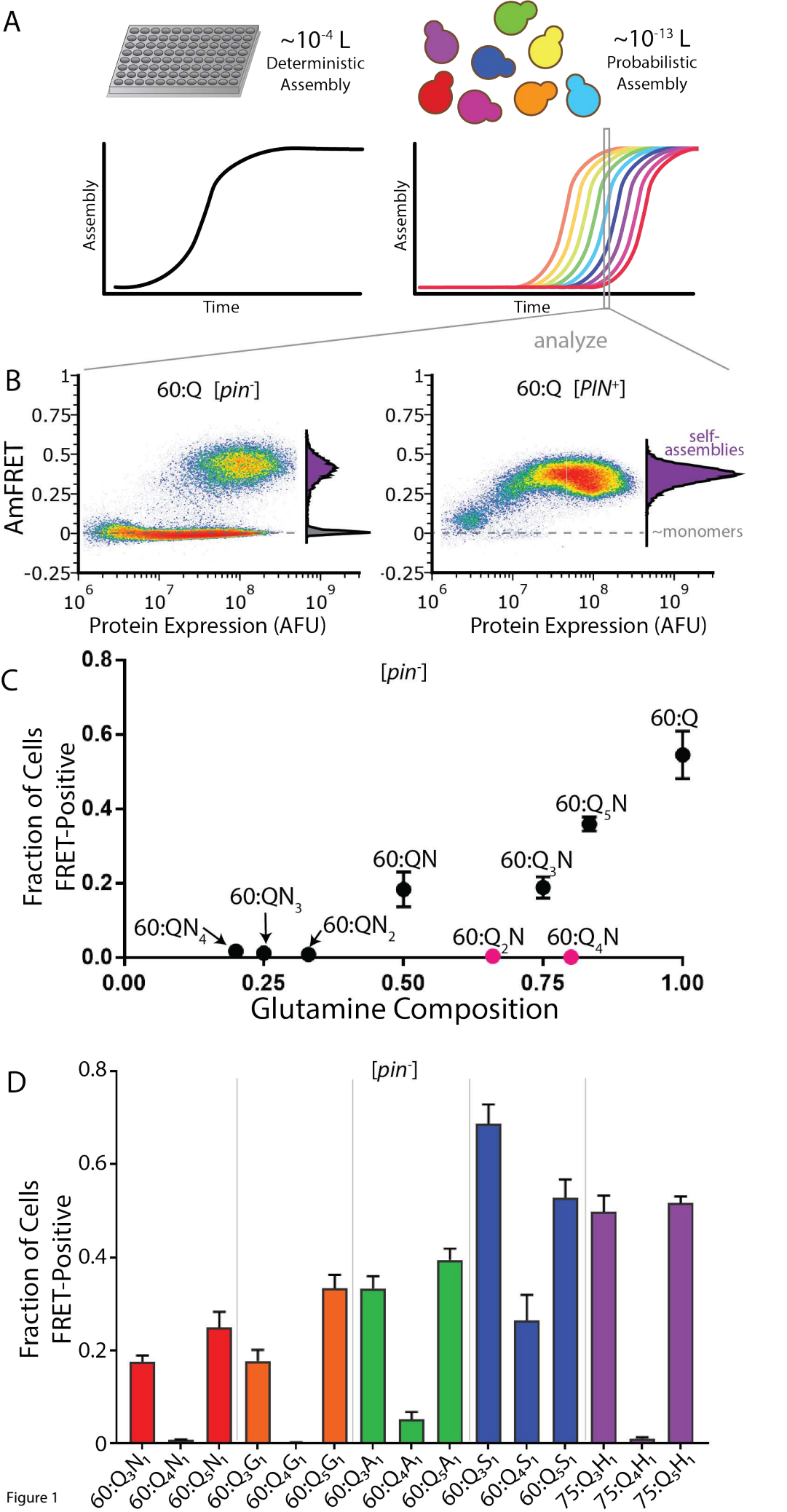
A
~10^-4 L
Deterministic
Assembly
~10^-13 L
Probabilistic
Assembly
Assembly
Time
Assembly
Time
analyze
B
60:Q [pin-]
60:Q [PIN+]
AmFRET
self-
assemblies
~monomers
Protein Expression (AFU)
Protein Expression (AFU)
C
[pin-]
Fraction of Cells
FRET-Positive
60:Q
60:Q5N
60:QN
60:Q3N
60:QN3
60:QN4
60:QN2
60:Q2N
60:Q4N
Glutamine Composition
D
[pin-]
Fraction of Cells
FRET-Positive
60:Q3N1
60:Q4N1
60:Q5N1
60:Q3G1
60:Q4G1
60:Q5G1
60:Q3A1
60:Q4A1
60:Q5A1
60:Q3S1
60:Q4S1
60:Q5S1
75:Q3H1
75:Q4H1
75:Q5H1

Figure 1

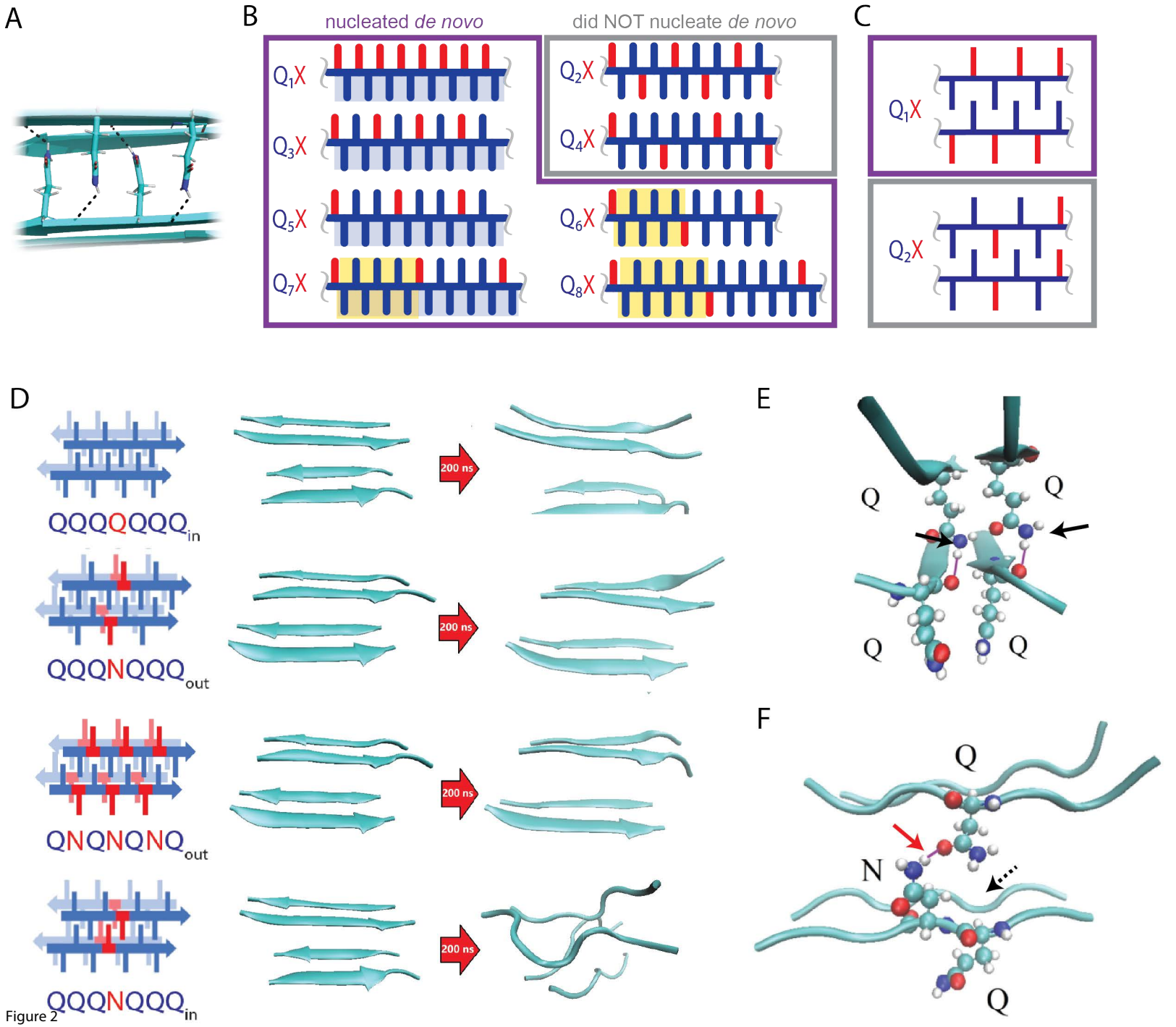
A
B
nucleated de novo
did NOT nucleate de novo
Q1X
Q2X
Q3X
Q4X
Q5X
Q6X
Q7X
Q8X
C
Q1X
Q2X
D
QQQQQQQin
QQQNQQQout
QNQNQNQout
QQQNQQQin
200 ns
200 ns
200 ns
200 ns
E
Q
Q
Q
Q
F
Q
N
Q

Figure 2

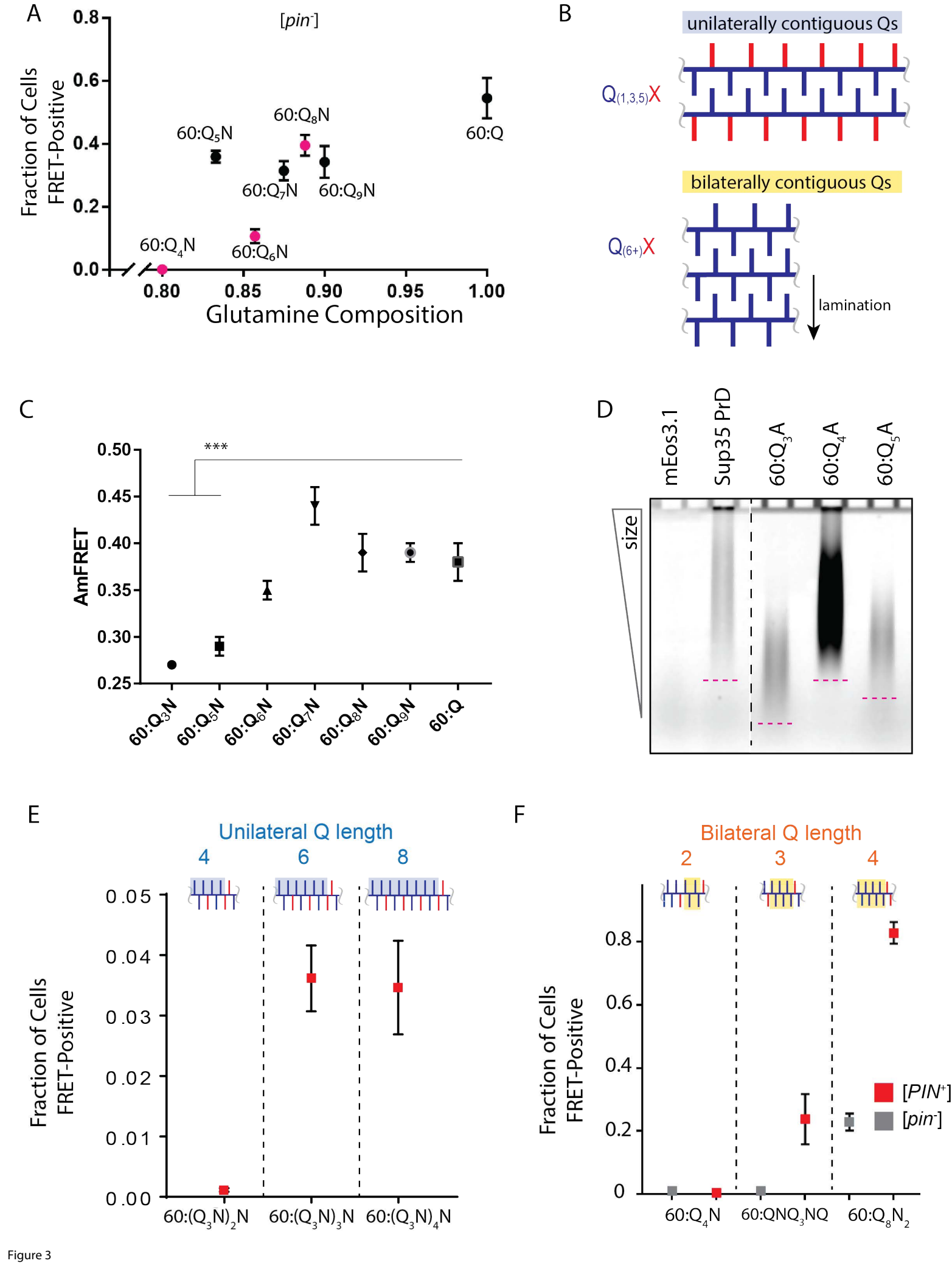
A
[pin-]
Fraction of Cells FRET-Positive
0.8
0.6
0.4
0.2
0.0
60:Q5N
60:Q8N
60:Q
60:Q7N
60:Q9N
60:Q4N
60:Q6N
0.80
0.85
0.90
0.95
1.00
Glutamine Composition
B
unilaterally contiguous Qs
Q(1,3,5)X
bilaterally contiguous Qs
Q(6+)X
lamination
C
***
AmFRET
0.50
0.45
0.40
0.35
0.30
0.25
60:Q3N
60:Q5N
60:Q6N
60:Q7N
60:Q8N
60:Q9N
60:Q
D
mEos3.1
Sup35 PrD
60:Q3A
60:Q4A
60:Q5A
size
E
Unilateral Q length
4
6
8
Fraction of Cells FRET-Positive
0.05
0.04
0.03
0.02
0.01
0.00
60:(Q3N)2N
60:(Q3N)3N
60:(Q3N)4N
F
Bilateral Q length
2
3
4
Fraction of Cells FRET-Positive
0.8
0.6
0.4
0.2
0
[PIN+]
[pin-]
60:Q4N
60:QNQ3NQ
60:Q8N2

Figure 3

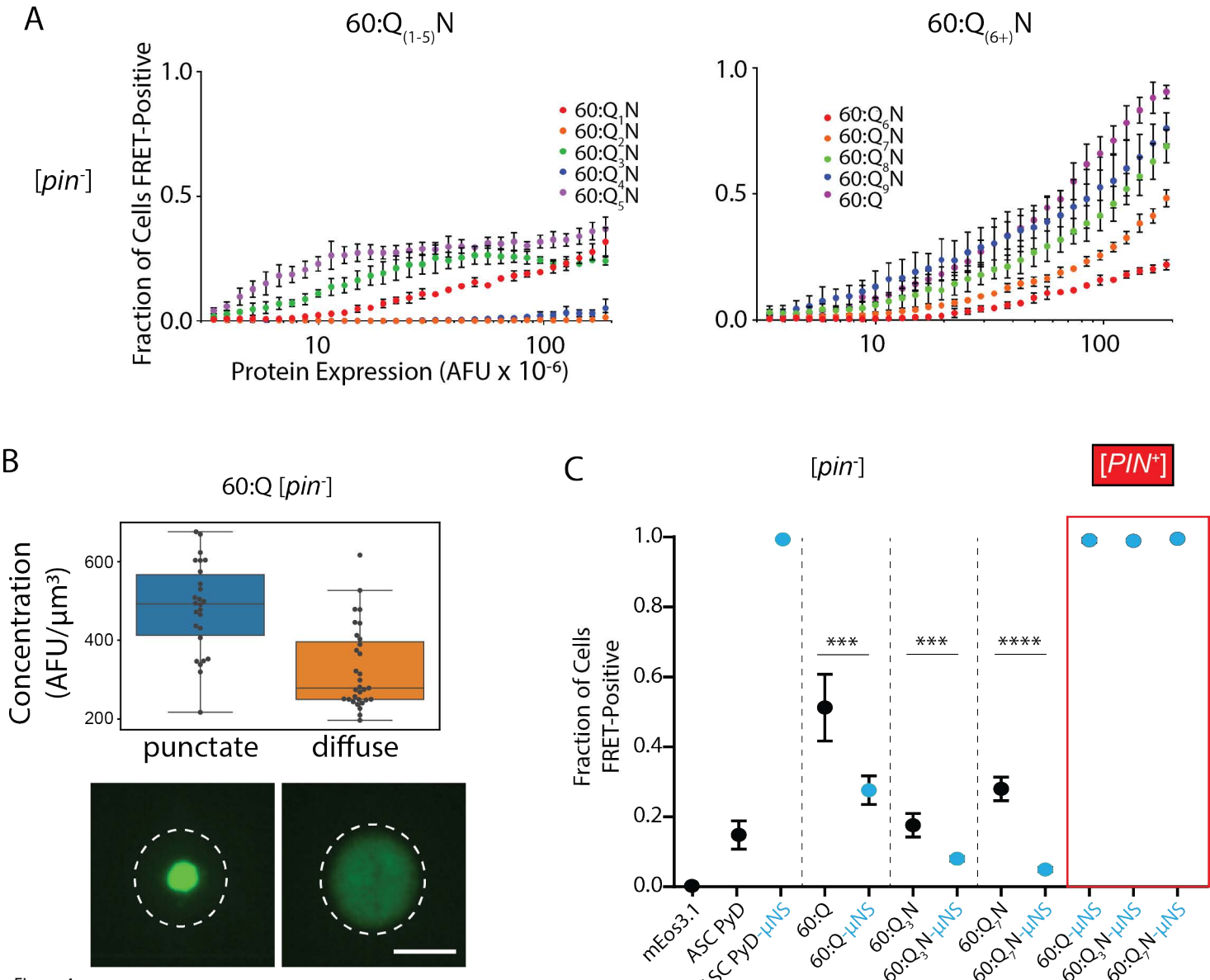
A
60:Q(1-5)N
60:Q(6+)N
[pin-]
Fraction of Cells FRET-Positive
1.0
0.5
0.0
10
100
Protein Expression (AFU x 10-6)
60:Q1N
60:Q2N
60:Q3N
60:Q4N
60:Q5N
60:Q6N
60:Q7N
60:Q8N
60:Q9N
60:Q
B
60:Q [pin-]
Concentration (AFU/µm3)
600
400
200
punctate
diffuse
C
[pin-]
[PIN+]
Fraction of Cells FRET-Positive
1.0
0.8
0.6
0.4
0.2
0.0
***
***
****
mEos3.1
ASC PyD
ASC PyD-µNS
60:Q
60:Q-µNS
60:Q3N
60:Q3N-µNS
60:Q7N
60:Q7N-µNS
60:Q-µNS
60:Q3N-µNS
60:Q7N-µNS

Figure 4

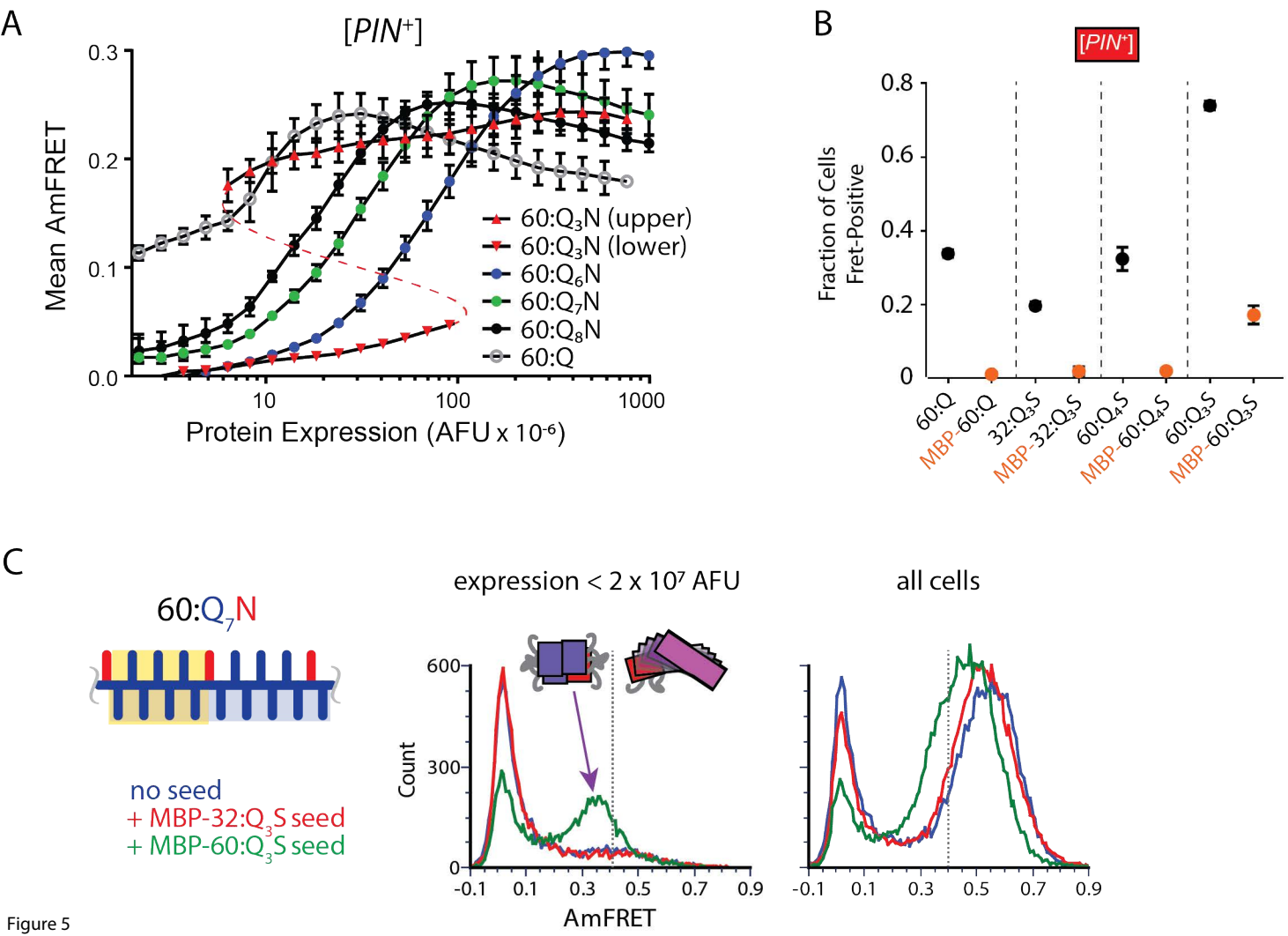
A
[PIN+]
Mean AmFRET
0.3
0.2
0.1
0.0
10
100
1000
Protein Expression (AFU x 10-6)
60:Q3N (upper)
60:Q3N (lower)
60:Q6N
60:Q7N
60:Q8N
60:Q
B
[PIN+]
Fraction of Cells Fret-Positive
0.8
0.6
0.4
0.2
0
60:Q
MBP-60:Q
32:Q3S
MBP-32:Q3S
60:Q4S
MBP-60:Q4S
60:Q3S
MBP-60:Q3S
C
60:Q7N
no seed
+ MBP-32:Q3S seed
+ MBP-60:Q3S seed
expression < 2 x 107 AFU
all cells
Count
600
300
0
-0.1
0.1
0.3
0.5
0.7
0.9
AmFRET

Figure 5

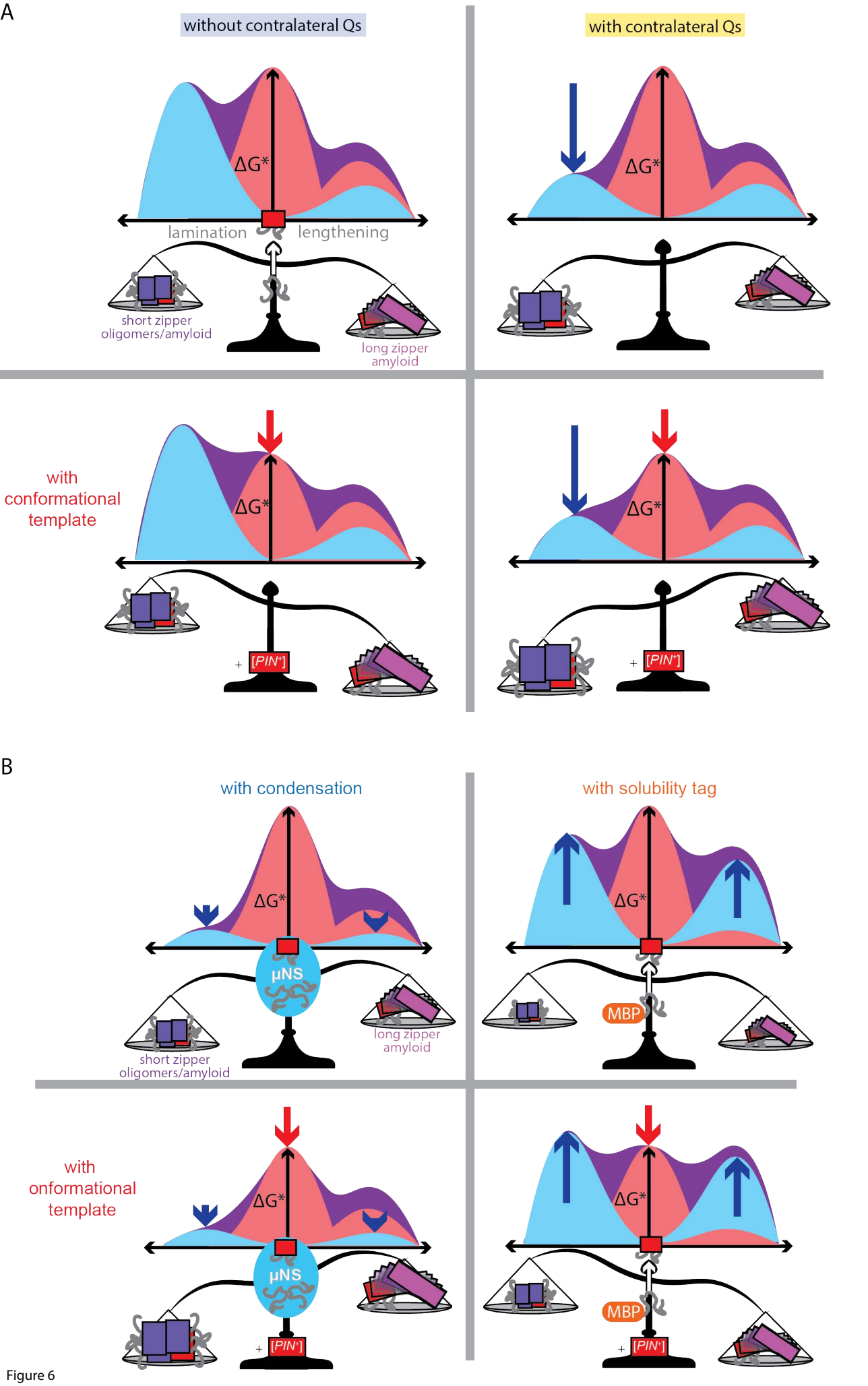
A
without contralateral Qs
with contralateral Qs
ΔG*
lamination
lengthening
short zipper oligomers/amyloid
long zipper amyloid
with conformational template
[PIN+]
B
with condensation
with solubility tag
μNS
MBP
with onformational template

Figure 6

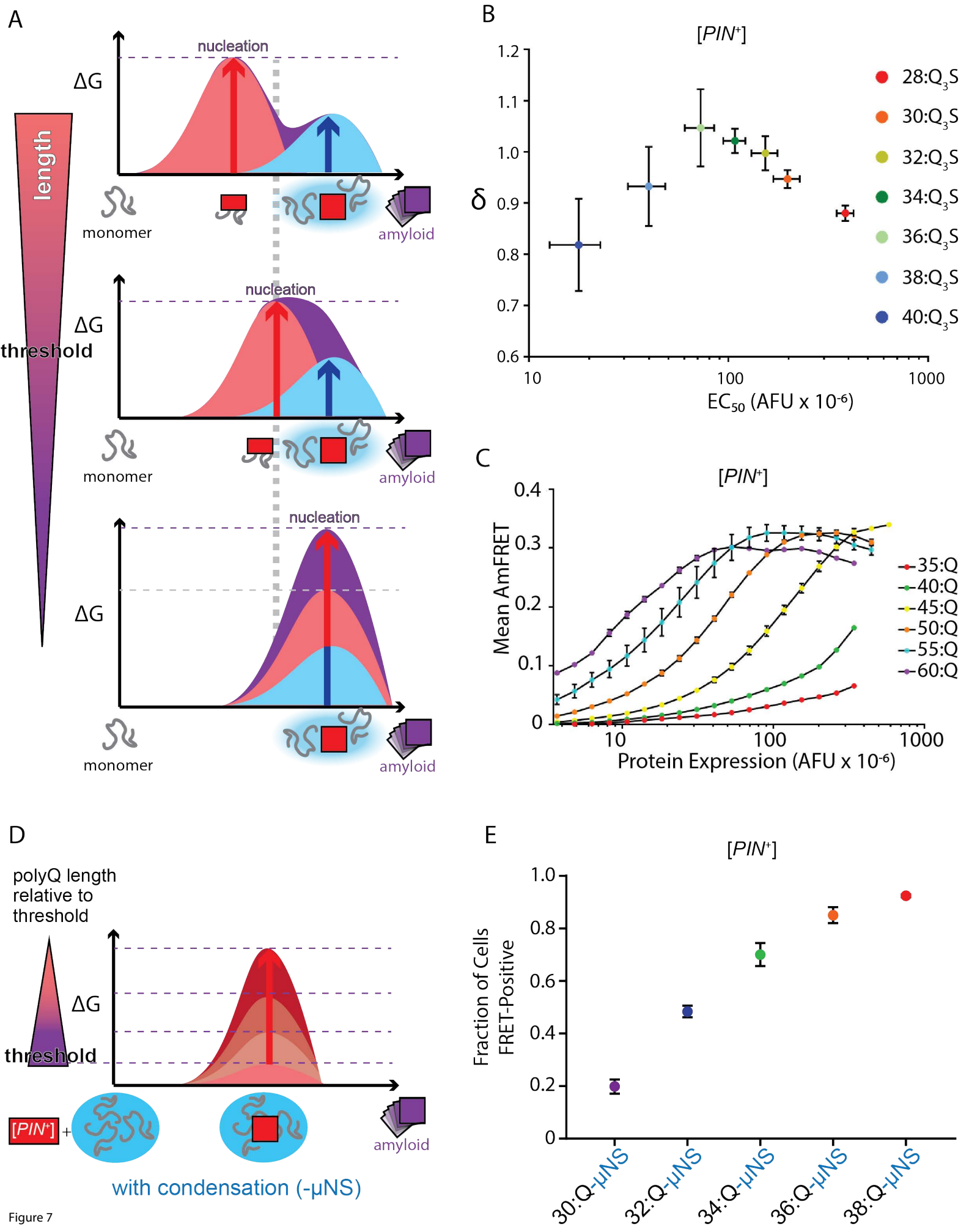
A
nucleation
ΔG
length
monomer
amyloid
threshold
B
[PIN+]
δ
28:Q3S
30:Q3S
32:Q3S
34:Q3S
36:Q3S
38:Q3S
40:Q3S
EC50 (AFU x 10-6)
C
[PIN+]
Mean AmFRET
35:Q
40:Q
45:Q
50:Q
55:Q
60:Q
Protein Expression (AFU x 10-6)
D
polyQ length relative to threshold
[PIN+]
with condensation (-μNS)
E
[PIN+]
Fraction of Cells FRET-Positive
30:Q-μNS
32:Q-μNS
34:Q-μNS
36:Q-μNS
38:Q-μNS

Figure 7

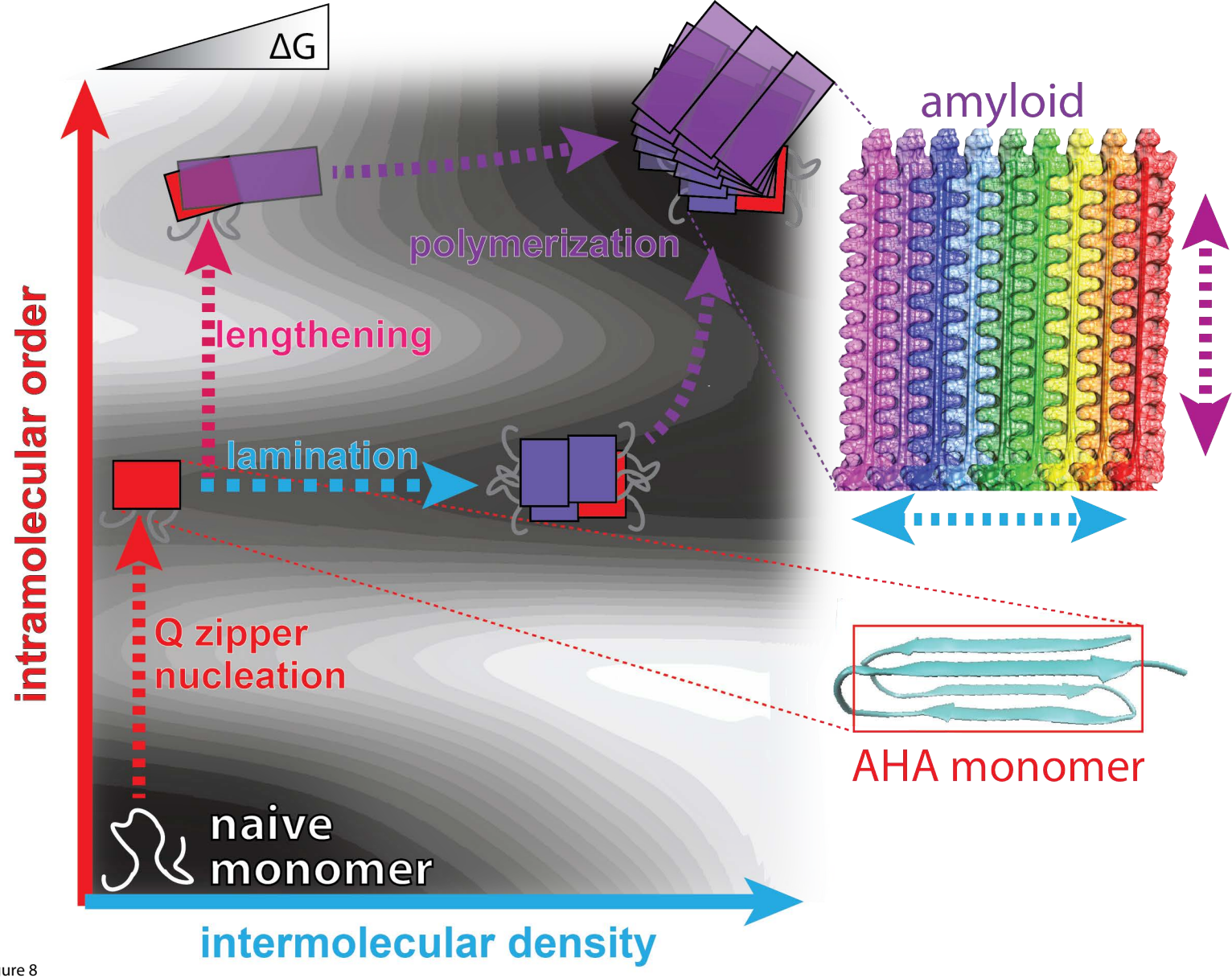
ΔG
amyloid
polymerization
lengthening
lamination
Q zipper
nucleation
naive
monomer
AHA monomer
intramolecular order
intermolecular density

Figure 8