Pathologic polyglutamine aggregation begins with a self-poisoning polymer crystal

Tej Kandola[1,2,*], Shriram Venkatesan[1,*], Jiahui Zhang[3], Brooklyn Lerbakken[1], Alex Von Schulze[1], Jillian F Blanck[1], Jianzheng Wu[1,4], Jay Unruh[1], Paula Berry[1], Jeffrey J. Lange[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 amyloid research has been to characterize the structural basis of the rate-determining nucleating event. However, the ephemeral nature of nucleation has made this goal unachievable with existing biochemistry, structural biology, and computational approaches. Here, we addressed that limitation for polyglutamine (polyQ), a polypeptide sequence that causes Huntington's and other amyloid-associated neurodegenerative diseases when its length exceeds a characteristic threshold. To identify essential features of the polyQ amyloid nucleus, we used a direct intracellular reporter of self-association to quantify nucleation frequencies as a function of concentration, conformational templates, and rational polyQ sequence permutations. We found that nucleation of pathologically expanded polyQ involves segments of three glutamine (Q) residues at every other position. We demonstrate using molecular simulations that this pattern encodes a four-stranded steric zipper with interdigitated Q side chains. Once formed, the zipper poisoned its own growth by engaging naive polypeptides on orthogonal faces, in a fashion characteristic of polymer crystals with intramolecular nuclei. We further show that preemptive oligomerization of polyQ inhibits amyloid nucleation. By uncovering the physical nature of the rate-limiting event for polyQ aggregation in cells, our findings elucidate the molecular etiology of polyQ diseases.

Introduction

Amyloids are highly ordered protein aggregates with self-templating activity. This activity drives the progression of multiple incurable diseases of aging, such as Alzheimer's (Chiti and Dobson, 2017; Huang et al., 2019). Understanding how amyloids start, or nucleate, is therefore fundamental to preventing these diseases. However, despite decades of intense research, we still lack a clear picture of even the gross anatomy of an amyloid nucleus.

Polyglutamine (polyQ) is an amyloid-forming sequence 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. Cells exhibiting polyQ pathology, whether in Huntington's disease patients, tissue culture, organotypic brain slices, or animal models, die independently and stochastically with a constant frequency (Clarke et al., 2000; Linsley et al., 2019). Two observations suggest that this emerges directly from the amyloid nucleation barrier. First, polyQ aggregation itself occurs stochastically in cells (Colby et al., 2006; Kakkar et al., 2016; Sinnige et al., 2021). Second, polyQ disease onset and progression are determined almost entirely by an *intramolecular* change, specifically, a genetically encoded expansion in the number of sequential glutamines beyond a protein-specific threshold of approximately 36 residues (Lieberman et al., 2019). Unlike

other amyloid diseases (Book et al., 2018; Kim et al., 2020; Selkoe and Hardy, 2016), polyQ disease severity generally does not worsen with gene dosage (Cubo et al., 2019; Lee et al., 2019; Wexler et al., 1987), implying that the rate-determining step in neuronal death occurs in a fixed minor fraction of the polyglutamine molecules. It is unclear how, and why, those molecules differ from the bulk. Hence, more so than for any other proteopathy, therapeutic progress against polyQ diseases awaits detailed knowledge of the very earliest steps of amyloid formation.

Amyloid nuclei cannot be observed directly by any existing experimental approach. This is because unlike mature amyloid fibrils, which are stable and amenable to structural biology, nuclei are unstable by definition. Their structures do not necessarily propagate into or correspond with the structures of 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). Moreover, 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 phase separation, which primarily concerns a loss of *intermolecular* entropy, amyloid formation additionally involves a major loss of *intramolecular* entropy. That is, nucleation 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, amyloid-forming proteins can accumulate to supersaturating concentrations while remaining soluble, thereby storing potential energy that will subsequently drive their aggregation following a stochastic nucleating event (Buell, 2017; Khan et al., 2018).

Due to the improbability that the requisite increases in both density and conformational ordering will occur spontaneously at the same time (i.e. *homogeneously*), amyloid nucleation tends to occur *heterogeneously*. In other words, amyloids tend to emerge via a progression of relatively less ordered but more probable, 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; Vekilov, 2012; Vitalis and Pappu, 2011; Yamaguchi et al., 2005; Zanjani et al., 2020). Heterogeneities divide the nucleation barrier into smaller, more probable steps, only one of which will be rate-limiting (**Fig. 1A**). Therefore, the occurrence of heterogeneities implies that the nature of the actual nucleus for a given amyloid can depend not only on the protein's sequence but also its *concentration* and cellular factors that influence its *conformation* (Bradley et al., 2002; Buell, 2017; Collinge and Clarke, 2007; Sanders et al., 2014; Törnquist et al., 2018).

Heterogeneities may be responsible for amyloid-associated proteotoxicity. Partially ordered species accumulate during the early stages of amyloid aggregation by all well-characterized pathological amyloids, but generally do not occur (or less so) during the formation of functional amyloids (Otzen and Riek, 2019). In the case of pathologically expanded polyQ, or the huntingtin protein containing pathologically expanded polyQ, amyloid-associated oligomers have been observed in vitro, in cultured cells, and in the brains of patients (Auer et al., 2008; Hsieh et al., 2017; Legleiter et al., 2010; Levin et al., 2014; Liang et al., 2018; Li et al., 2010; Olshina et al., 2010; Sathasivam et al., 2010; Sil et al., 2018; Takahashi et al., 2008; Vitalis and Pappu, 2011; Yamaguchi et al., 2005; Zanjani et al., 2020), and are likely culprits of proteotoxicity (Kim et al., 2016; Leitman et al., 2013; Lu and Palacino, 2013; Matlahov and van der Wel, 2019; Miller et al., 2011; Takahashi et al., 2008; Wetzel, 2020). In contrast, mature amyloid fibers are increasingly viewed as benign or even protective (Arrasate et al., 2004; Kim et al., 2016; Leitman et al., 2013; Lu and Palacino, 2013; Matlahov and van der Wel, 2019; Takahashi et al., 2008; Wetzel, 2020). The fact that polyQ disease kinetics is governed by amyloid nucleation, but the amyloids themselves appear not to be responsible, presents a paradox. A structural model of the polyQ amyloid nucleus can be expected to resolve this paradox by illuminating a path for the propagation of conformational order into distinct multimeric species that either cause or mitigate toxicity.

1

Structural features of amyloid nucleation may be deduced by studying the effects on amyloid kinetics of rational mutations made to the polypeptide (Thakur and Wetzel, 2002). The occurrence of density heterogeneities confounds this approach, however, as they blunt the dependence of amyloid kinetics on concentration that could otherwise reveal nucleus stoichiometry (Vitalis and Pappu, 2011). Resolving this problem would require an assay that can detect nucleation events independently of amyloid growth kinetics, and which can scale to accommodate the large numbers of mutations necessary to identify sequence-structure relationships. Classic assays of *in vitro* amyloid assembly kinetics are poorly suited to this task, both because of their limited throughput, and because even the smallest experimentally tractable reaction volumes are too large to observe discrete aggregation events (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. 1B**). 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), allowing amyloid to form in the same nucleation-limited fashion as it does in afflicted neurons (Colby et al., 2006; Wetzel, 2006). 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 point in time appropriate for the kinetics of amyloid formation (hours), DAmFRET probes the existence and magnitude of critical density fluctuations that may 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-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*⁻]) populate an amyloid conformation (Derkatch et al., 2001). [*PIN*⁺] serves as a partial template for amyloid formation by compositionally similar proteins including polyQ (Alexandrov et al., 2008; Duennwald et al., 2006a; Meriin et al., 2002; Serpionov et al., 2015). Known as cross-seeding, this phenomenon is analogous to, but much less efficient than, amyloid elongation by molecularly identical species (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 can uncouple the two components of the nucleation barrier and thereby relate specific sequence features to the nucleating conformation. Sequence changes that are specific to the sequence-encoded nucleus will more strongly influence nucleation in [*pin*⁻] cells than in [*PIN*⁺] cells.

PolyQ is an ideal polypeptide with which to deduce the physical nature of a pathologic amyloid nucleus. It has zero complexity, which profoundly simplifies the design and interpretation of sequence variants. Unlike other pathogenic protein amyloids, which occur as different structural polymorphs each plausibly with their own structural nucleus, polyQ amyloids have an invariant core structure under different assembly conditions and in the context of different flanking domains (Boatz et al., 2020; Galaz-Montoya et al., 2021; Lin et al., 2017; Schneider et al., 2011). This core contains antiparallel beta-sheets (Buchanan et al., 2014; Matlahov and van der Wel, 2019; Schneider et al., 2011), while most other amyloids, including all other Q-rich amyloids, have a parallel beta-sheet core (Eisenberg and Jucker, 2012; Margittai and Langen, 2008). This suggests that amyloid nucleation and pathogenesis of polyQ may follow from an ability to spontaneously acquire a *specific* nucleating

2

conformation. Any variant that decelerates nucleation can therefore be interpreted with respect to its effect on *that* conformation.

The experiments and results reported herein reveal that the polyQ amyloid nucleus is a steric zipper encoded by a pattern of approximately twelve Q residues in a single polypeptide molecule. We find that the clinical length threshold for polyQ disease is the minimum length that can encompass the pattern. Consistent with a monomeric nucleus, amyloid formation occurred less frequently at high concentrations or when polyQ was expressed in oligomeric form. We further found that the nucleus promotes its own kinetic arrest by templating competing dimensions of Q zipper ordering -- both along the amyloid axis and orthogonally to it. This leads to the accumulation of partially ordered aggregates.

Results

A hidden pattern of glutamines governs amyloid nucleation

To validate the use of DAmFRET for pathologic polyQ, we first surveyed the length-dependence for amyloid formation by polyQ tracts expressed as fusions to mEos3.1 in nondividing [*pin⁻*] yeast cells. We genetically compromised protein degradation in these cells to prevent differential turnover of potential polyQ heterogeneities, which could otherwise obscure the relationship of aggregation to concentration. Our data recapitulated the pathologic threshold -- Q lengths 35 and shorter lacked AmFRET, indicating a failure to aggregate or even appreciably oligomerize, while Q lengths 40 and longer did acquire AmFRET in a length and concentration-dependent manner **(Fig. 1C top, 1D, S1A)**. Specifically, the cells partitioned into two discontinuous populations: one with high AmFRET and the other with none (top and bottom populations in **Fig. 1C**). The two populations occurred with overlapping concentration ranges (red dashed box in **Fig. 1C top**), indicating a kinetic barrier to forming the high-FRET state. The two populations were well-resolved at high concentrations, where relatively few cells populated intermediate values of AmFRET. These features together indicate that aggregation of polyQ is rate-limited by a nucleation barrier associated with a rare conformational fluctuation.

Sixty residues proved to be the optimum length to observe both the pre- and post-nucleated states of polyQ in single experiments, and corresponds clinically to disease onset in early adulthood (Kuiper et al., 2017). We therefore focused on this length moving forward. The frequency of polyQ nucleation greatly exceeds that of other Q-rich amyloid-forming proteins of comparable length (Khan et al., 2018; Posey et al., 2021). To determine if the protein's exceptional amyloid propensity can be attributed to a property of the Q residues themselves, as opposed to its extremely low sequence complexity, we next tested a 60-residue homopolymer of asparagine (polyN), whose side chain is one methylene shorter but otherwise identical to that of glutamine. PolyN populated the high FRET state at much lower frequencies (typically undetectable) than polyQ even at the highest concentrations (**Fig. 1C bottom**, approximately 200 µM (Khan et al., 2018)).

The much larger nucleation barrier for polyN -- despite its physicochemical similarity to polyQ -- led us to consider whether substituting Qs for Ns at specific positions in polyQ might reveal patterns that uniquely encode the structure of the polyQ nucleus. We reasoned that a random screen of Q/N sequence space would be unlikely to yield informative patterns, however, given that there are more than one trillion combinations of Q and N for even a 40-residue peptide. We therefore devised a systematic approach to rationally sample Q/N sequence space. Specifically, we characterized by DAmFRET a series of related sequences with a single N residue inserted after every q Q residues over a total length of 60, for all values of q from 0 through 9

3

(designated Q_1N, Q_2N, etc. with the value of q subscripted; **Table S1**). The resulting dataset revealed a shockingly strong dependence of amyloid propensity on the exact sequence of Q residues, and specifically the following two determinants.

First, we observed amyloid formation for all values of $q \geq 6$ (**Fig. S1B**). Second, amyloid formation for values of $q < 6$ was limited to odd numbers 1, 3, and 5 (**Fig. 1E**). This simple pattern could be explained by a local requirement for Q at every other position along the sequence (i.e. i, $i+2$, $i+4$, etc.). To test this, we inserted a second N residue into the repeats for two such sequences. Indeed, these failed to aggregate (**Fig. S1C**).

To determine if this "odd" dependency results from a general preference of amyloid for homogeneity at every other position, irrespective of their amino acid identity, we created an analogous "polyN" series with a single Q placed after every first, second, third, or fourth N. This series showed no preference for even or odd values (**Fig. S1C**), confirming that Qs, specifically, are required at every other position. We next asked if the odd dependency concerns the Q side chains themselves, rather than their interaction with Ns, by replacing the Ns with residues of diverse physicochemistry -- either glycine, alanine, serine, or histidine. Again, nucleation was much more frequent for Q_3X and Q_5X than for Q_4X (**Fig. 1E, S1D**), confirming that the odd dependency resulted from the Q side chains interacting specifically with other Q side chains. The identity of X did, however, influence nucleation frequency particularly for Q_4X. Glycine proved just as detrimental as asparagine; histidine slightly less so; and alanine lesser still. Serine was relatively permissive. The relative indifference of Q_3X and Q_5X to the identity of X suggests that these sequences form a pure Q core that excludes every other side chain. In other words, the nucleus is encoded by segments of sequence with Qs at every other position.

To reduce the contribution of the conformational fluctuation to nucleation, we next performed DAmFRET experiments for all sequences in [PIN^+] cells. Amyloid formation broadly increased (**Fig. S1A-D**), with polyQ now nucleating in most [PIN^+] cells even at the lowest concentrations sampled by DAmFRET (~1 µM, (Khan et al., 2018)). Importantly, amyloid formation again occurred only for polyQ lengths exceeding the clinical threshold. Additionally, nucleation in [PIN^+] retained the structural constraints of de novo nucleation, as indicated by a) the relative impacts of different X substitutions, which again increased in the order serine < alanine < histidine < glycine/asparagine; and b) the persistence of an odd preference, with q = 1, 3, and 5 nucleating more frequently than q = 4. The latter failed entirely to form amyloid in the cases of asparagine and glycine. All N-predominant sequences and only two Q-predominant sequences (Q_2N and $(Q_3N)_2N$; introduced below) nucleated robustly in [PIN^+], even though they failed entirely to do so in [pin^-] cells. Their behavior resembles that of typical "prion-like" sequences (Khan et al., 2018; Posey et al., 2021), consistent with our deduction that these amyloids necessarily have a different nucleus structure than that of uninterrupted polyQ.

The pattern encodes a single, singular steric zipper

What structure does the pattern encode, and why is it so sensitive to Ns? We sought to uncover a physical basis for the different nucleation propensities between Q and N using state-of-the-art amyloid predictors (Charoenkwan et al., 2021; Keresztes et al., 2021; Prabakaran et al., 2021). Unfortunately, none of the predictors were able to distinguish Q_3N and Q_5N from Q_4N (Table S2), despite their very dissimilar experimentally-determined amyloid propensities. In fact, most predictors failed outright to detect amyloid propensity in this class of sequences. Apparently, polyQ is exceptional among the known amyloid-forming sequences on which these predictors were trained, again hinting at a specific nucleus structure.

Most pathogenic amyloids feature residues that are hydrophobic and/or have a high propensity for beta strands. Glutamine falls into neither of these categories (Fujiwara et al., 2012; Nacar, 2020; Simm et al., 2016).

4

We reasoned that nucleation of polyQ 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 (Hoop et al., 2016; Schneider et al., 2011; Sikorski and Atkins, 2005). The resulting structural element exhibits exceptional shape complementarity (even among amyloids) and is stabilized by the multitude of resulting van der Waals interactions. Additionally, the terminal amides of Q side chains form regular hydrogen bonds to the backbones of the opposing beta-sheet (Esposito et al., 2008; Hervas et al., 2020; Man et al., 2015; Schneider et al., 2011; Y. Zhang et al., 2016). To minimize confusion while referring to these distinguishing tertiary features of the all-Q amyloid core, relative to the similarly packed cores of virtually all other amyloids, which lack regular side chain-to-backbone H bonds (Eisenberg and Sawaya, 2017; Sawaya et al., 2007), we will henceforth refer to them as "Q zippers" (**Fig. 2A**).

We reasoned that the entropic cost of acquiring such an exquisitely ordered structure by the entirely disordered ensemble of naive polyQ (Chen et al., 2001; Moradi et al., 2012; Newcombe et al., 2018; Vitalis et al., 2007; Wang et al., 2006) may very well underlie the nucleation barrier. If so, then the odd dependency could relate to the fact that successive side chains alternate by 180° along a beta strand (**Fig. 2B-C**). This would force non-Q side chains into the zipper for all even values of q. Our experimental results imply that the intruding residue is most disruptive when it is an N, and much less so when it is an S. To determine if the Q zipper has such selectivity, we carried out fully atomistic molecular dynamics simulations on Q-substituted variants of a minimal Q zipper model (Y. Zhang et al., 2016). Specifically, we aligned four Q7 peptides into a pair of interdigitated two-stranded antiparallel β-sheets with interdigitated Q side chains. We then substituted Q residues in different positions with either N or S and allowed the structures to evolve for 200 ns. In agreement with the experimental data, the Q zipper was highly specific for Q side chains: it rapidly dissolved when any proximal pair of inward pointing Qs were substituted, as would occur for unilaterally discontinuous sequences such as Q_4X (**Fig. 2D, S2A**). In contrast, it remained intact when any number of outward-pointing Q residues were substituted, as would occur for Q_3X and Q_5X. Given the sensitivity of this minimal zipper to substitutions, we next constructed a zipper with twice the number of strands, where each strand contained a single inward pointing N or S substitution, and allowed the structures to evolve for 1200 ns. In this case, the S-containing zipper remained intact while the N-containing zipper dissolved (**Fig. S2B**), consistent with our experimental results.

A close examination of the simulation trajectories revealed how N disrupts Q zippers. When inside a Q zipper, the N side chain is too short to H-bond the opposing backbone. In our simulations, the terminal amide of N competed against 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. 2E-F**). The now detached Q side chain collided with adjacent Q side chains, causing them to also detach, ultimately unzipping the zipper (**Fig. S2C and Movies S1, S2**). We found that the side chain of S, which is slightly shorter and less bulky than that of N, does not intercept H-bonds from the side chains of opposing Q residues (**Fig. S2D**). As a result, order is maintained so long as multiple S side chains do not occur in close proximity inside the zipper.

We also considered an alternative mechanism whereby N side chains H-bonded to adjacent Q side chains on the same side of the strand (i.e. i + 2). This structure -- termed a "polar clasp" (Gallagher-Jones et al., 2018) -- would be incompatible with interdigitation by opposing Q side chains (**Fig. S2E**). We therefore 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. This analysis showed no evidence that Q side chains preferentially H-bond adjacent N side chains (**Fig. S2F-G**), ruling out the polar clasp mechanism of inhibition.

In sum, our simulations are consistent with polyQ amyloid nucleating from a single Q zipper comprising two β-sheets engaged across an interface of fully interdigitated Q side chains.

The Q zipper grows in two dimensions

Prior structural studies show that polyQ amyloids have a lamellar or "slab-like" architecture (Boatz et al., 2020; Galaz-Montoya et al., 2021; Nazarov et al., 2022; Sathasivam et al., 2010; Sharma et al., 2005). This indicates that a nucleus comprising a single Q zipper would have to propagate not only axially but also laterally as it matures toward amyloid, provided that both sides of each strand can form a Q zipper (**Fig. 3A**). We observed that when expressed to high concentrations, sequences with q values larger than 5 generally reached higher AmFRET values than those with q = 3 or 5 (**Fig. 3B**). This suggests that the former has a higher subunit density consistent with a multilamellar structure. We will therefore refer to such sequences as *bilaterally* contiguous or Q_B and sequences that are only capable of Q zipper formation on one side of a strand, such as Q_3X and Q_5X, as *unilaterally* contiguous or Q_U. AmFRET is a measure of total cellular FRET normalized by concentration. In a two phase regime, AmFRET scales with the fraction of protein in the assembled phase (Posey et al., 2021). At very high expression levels where approximately all the protein is assembled, AmFRET should theoretically approach a maximum value determined only by the proximity of fluorophores to one another in the assembled phase. This proximity in turn reflects the density of subunits in the phase as well as the orientation of the fluorophores on the subunits. To control for the latter, we varied the terminus and type of linker used to attach mEos3.1 to polyQ, Q_3N, and Q_7N. We found that polyQ and Q_7N achieved higher AmFRET than Q_3N regardless of linker terminus and identity (**Fig. S3A**). The high AmFRET level achieved by polyQ amyloids therefore results from the subunits occurring in closer proximity than in amyloids that lack the ability to form lamella.

To more directly investigate structural differences between amyloids of these sequences, we analyzed the size distributions of SDS-insoluble multimers using semi-denaturing detergent-agarose gel electrophoresis (SDD-AGE) (Halfmann and Lindquist, 2008; Kryndushkin et al., 2003). We found that Q_U amyloid particles were smaller than those of Q_B (**Fig. 3C, S3B**). This difference necessarily means that they either nucleate more frequently (resulting in more but smaller multimers at steady state), grow slower, and/or fragment more than polyQ amyloids (Knowles et al., 2009). The DAmFRET data do not support the first possibility (**Fig. S1D**). To compare growth rates, we examined AmFRET histograms through the bimodal region of DAmFRET plots for [*pin*-] cells. Cells with intermediate AmFRET values (i.e. cells in which amyloid had nucleated but not yet reached steady state) were less frequent for Q_U (**Fig. 3D, S3C**), which is inconsistent with the second possibility. Therefore, the SDD-AGE data are most consistent with the third possibility -- increased fragmentation. We attribute this increase to a difference in the structures rather than sequences of the two types of amyloids, because Q and N residues are not directly recognized by protein disaggregases (Alexandrov et al., 2012, 2008; Osherovich et al., 2004). Based on the fact that larger amyloid cores, such as those with multiple steric zippers, oppose fragmentation (Tanaka et al., 2006; Verges et al., 2011; Zanjani et al., 2020), these data are therefore consistent with a lamellar (**Fig. 3A**) architecture for polyQ.

Q zippers poison themselves

A close look at the DAmFRET data reveals a complex relationship of amyloid formation to concentration. While all sequences formed amyloid more frequently with concentration in the low regime, as concentrations increased further, the frequency of amyloid formation plateaued somewhat before increasing again at higher concentrations (**Fig. 4A**). A multiphasic dependence of amyloid formation on concentration implies that some

higher-order species inhibits nucleation and/or growth (Vitalis and Pappu, 2011). The non-amyloid containing cells in the lower population lacked AmFRET altogether, suggesting that the inhibitory species either have very low densities of the fluorophore (e.g. due to co-assembly with other proteins), and/or do not accumulate to detectable concentrations. Liquid-liquid phase separation by low-complexity sequences can give rise to condensates with low densities and heterogeneous compositions (Wei et al., 2017), and condensation has been shown to inhibit amyloid formation in some contexts (Gabryelczyk et al., 2022; Küffner et al., 2021; Lipiński et al., 2022). We therefore inspected the subcellular distribution of representative proteins in different regions of their respective DAmFRET plots. The amyloid-containing cells had large round or stellate puncta, as expected. To our surprise, however, cells lacking AmFRET contained exclusively diffuse protein (no detectable puncta), even at high expression (**Fig. 4B, S4A**). This means that naive polyQ does not itself phase separate under normal cellular conditions. The inhibitory species are therefore some form of soluble oligomer. That these oligomers are too sparse to detect by AmFRET, even at high total protein concentration, further suggests that they are dead-end (kinetically trapped) products of the Q zipper nuclei themselves. In other words, high concentrations of soluble polyQ stall the growth of nascent Q zippers.

We were struck by the similarity of this phenomenon to "self-poisoning" in the polymer crystal literature (Sadler, 1983; Ungar et al., 2005; Whitelam et al., 2016; M. Zhang et al., 2016; Zhang et al., 2021). Self-poisoning is a deceleration of crystal growth with increasing driving force. In short, when multiple molecules simultaneously engage the templating crystal surface, they tend to "trap" each other in partially ordered configurations that block the recruitment of subsequent molecules (**Fig. 4C, red**). This phenomenon requires conformational conversion on the templating surface to be slow relative to the arrival of new molecules. Indeed, disordered polyQ is an extremely viscous globule (Crick et al., 2006; Dougan et al., 2009; Kang et al., 2017; Walters and Murphy, 2009) whose slow conformational fluctuations limit the rate of amyloid elongation (Walters et al., 2012).

Self-poisoning is most extreme where phase boundaries converge, as when a polymer is equally compatible with either of two crystal polymorphs (Hu, 2018; Ungar et al., 2005). For example, the rate of polymer crystallization rises, falls, and then rises again as a function of polymer length, where the second minimum results from the polymer heterogeneously conforming to either a single-long-strand polymorph or a hairpin polymorph (**Fig. S4B**). This is a consequence of secondary nucleation *within* the molecule (Hu, 2018; Ungar et al., 2005; M. Zhang et al., 2016; Zhang et al., 2021). Remarkably, an analogous phenomenon manifested for Q zippers. Specifically, we noticed that while Q_5N formed amyloid robustly at low concentrations (i.e. prior to self-poisoning at higher concentrations), Q_7N did not (**Fig. 4A, purple arrow**). We were initially perplexed by this because all odd q sequences have a fully contiguous pattern of Qs at every other position that should allow for Q zipper nucleation. We then realized, however, that the anomalous solubility of Q_7N coincides with the expected enhancement of self-poisoning at the transition between single long Q zippers (favored by Q_U sequences) and short lamellar Q zippers (favored by Q_B sequences). In other words, the striking difference between Q_5N and Q_7N appears to result from the formation of *intramolecular* Q zippers with strands only six residues long, which form *contralaterally* to the intermolecular Q zipper interface (**Fig. 4C, blue**). As for other polymers, Q_B sequences increasingly escaped the poisoned state as the potential length of lamellar strands increased beyond the minimum, i.e. with increasing q values beyond 6 (**Fig. 4D**).

In our experimental system, ongoing protein translation will cause the polypeptide to become increasingly supersaturated with respect to the self-poisoned amyloid phase. However, as concentrations increase, the contributions of unstructured low affinity interactions (primarily H-bonds between Q side chains (Kang et al., 2018; Punihaole et al., 2018)) to intermolecular associations will increase relative to the contributions of Q

7

zipper elements. Disordered polyQ has a greater affinity for itself than does disordered polyN (Halfmann et al., 2011). The ability of Q_B sequences to escape poisoning with increasing concentrations may therefore be a simple consequence of their greater content of Qs. Indeed, polyQ amyloid (Walters et al., 2012), as for other polymer crystals (Zhang et al., 2021), has a relatively disordered growth front at high polymer concentrations.

Our finding that polyQ does not phase-separate prior to Q zipper nucleation implies that intermolecular Q zippers will be less stable than intramolecular Q zippers of the same length, because the effective concentration of Q zipper-forming segments will always be lower outside the intramolecular globule than inside it. Therefore, the critical strand length for growth of a single (non-lamellar) Q zipper must be longer than six, whereas strands of length six will only grow in the context of lamella. To test this prediction, we designed a series of sequences with variably restricted unilateral or bilateral contiguity (**Fig. S4C-D**). We found that, while all sequences formed amyloid with a detectable frequency in [*PIN*⁺] cells, only those with at least five unilaterally contiguous Qs, or at least six bilaterally contiguous Qs, did so in [*pin*⁻] cells. Given that the ability to nucleate in the absence of a pre-existing conformational template is characteristic of Q zippers, this result is consistent with a threshold strand length of nine or ten residues for single Q zippers to propagate, and six residues for lamellar Q zippers to propagate.

One consequence of self-poisoning is that crystallization rates accelerate with monomer depletion (Ungar et al., 2005). In our system, poisoning should be most severe in the early stages of amyloid formation when templates are too few for polypeptide deposition to outpace polypeptide synthesis, thereby maintaining concentrations at poisoning levels. For sufficiently fast nucleation rates and slow growth rates, this property predicts a bifurcation in AmFRET values as concentrations enter the poisoning regime, as we see for Q_B in [*PIN*⁺] cells. To determine if the mid-AmFRET cells in the bifurcated regime indeed have not yet reached steady state, we used translation inhibitors to stop the influx of new polypeptides -- and thereby relax self-poisoning -- for six hours prior to analysis. As predicted, treated AmFRET-positive cells in the bifurcated regime achieved higher AmFRET values than cells whose translation was not arrested (**Fig. 4E, S4E**).

The nucleus forms within a single molecule

Because amyloid formation is a transition in both conformational ordering *and* density, and the latter is not rate-limiting over our experimental time-scales (Khan et al., 2018), the critical conformational fluctuation for most amyloid-forming sequences is generally presumed to occur within disordered multimers (Auer et al., 2008; Buell, 2017; Serio et al., 2000; Vekilov, 2012; Vitalis and Pappu, 2011). In contrast, homopolymer crystals nucleate preferentially within monomers when the chain exceeds a threshold length (Hu, 2018; Xu et al., 2021). The possibility that polyQ amyloid may nucleate as a monomer has long been suspected to underlie the length threshold for polyQ pathology (Chen et al., 2002). A minimal intramolecular Q zipper would comprise a pair of two-stranded beta sheets connected by loops of approximately four residues (Chou and Fasman, 1977; Hennetin et al., 2006). If we take the clinical threshold of approximately 36 residues as the length required to form this structure, then the strands must be approximately six residues with three unilaterally contiguous Qs -- exactly the length we deduced for intramolecular strand formation in the previous section.

This interpretation leads to an important prediction concerning the inability of Q_4N to form amyloid. With segments of four unilaterally and bilaterally contiguous Qs, this sequence should be able to form the hypothetical intramolecular Q zipper, but fail to propagate it either axially or laterally as doing so requires at least five unilaterally or six bilaterally contiguous Qs, respectively. If the rate-limiting step for amyloid is the formation of a single short Q zipper, then appending Q_4N to another Q zipper amyloid-forming sequence will

facilitate its nucleation. If nucleation instead requires longer or lamellar zippers, then appending Q_4N will inhibit nucleation. To test this prediction, we employed Q_3N as a non-lamellar "sensor" of otherwise cryptic Q zipper strands donated by a 30 residue Q_4N tract appended to it. To control for the increased total length of the polypeptide, we separately appended a non Q zipper-compatible tract: Q_2N. As predicted, the Q_4N appendage increased the fraction of cells in the high-AmFRET population relative to those expressing Q_3N alone, and even more so relative to those expressing the Q_2N-appended protein (**Fig. 5A, S5A**). These data are consistent with a single short Q zipper as the amyloid nucleus.

To now determine if the nucleating zipper occurs preferentially in a single polypeptide chain, we genetically fused Q_U and Q_B to either of two homo-oligomeric modules: a coiled coil dimer (oDi, (Fletcher et al., 2012)), or a 24-mer (human FTH1, (Bracha et al., 2018)). Importantly, neither of these fusion partners contain β-structure that could conceivably template Q zippers. Remarkably, the oDi fusion reduced amyloid formation, and the FTH1 fusion all but eliminated it, for both Q_U and Q_B sequences (**Fig. 5B, S5B**). To exclude trivial explanations for this result, we additionally tested fusions to oDi on the opposite terminus, with a different linker, or with a mutation designed to block dimerization (designated Odi{X}). The inhibitory effect of oDi manifested regardless of the terminus tagged or the linker used, and the monomerizing mutation rescued amyloid formation (**Fig. S5C-F**), altogether confirming that preemptive oligomerization prevents Q zipper formation.

Finally, we designed a sequence representing the minimal "polyQ" amyloid nucleus. This sequence has the precise number and placement of Qs necessary for amyloid to form via intramolecular Q zipper nucleation: four tracts of six Qs linked by minimal loops of three glycines. The sequence itself is too short -- at 33 residues -- to form amyloid unless the glycine loops are correctly positioned (recall from **Fig. 1D** and **S1A** that pure polyQ shorter than 40 residues does not detectably aggregate). We found that the protein indeed formed amyloid robustly, and with a concentration-dependence and [*PIN*$^+$]-independence that is characteristic of Q_B (**Fig. 5C, S5G**). If the nucleus is indeed a monomer, then a defect in even one of the four strands should severely restrict amyloid formation. We therefore mutated a single Q residue to an N. Remarkably, this tiny change -- removing just one carbon atom from the polypeptide -- completely eliminated amyloid formation (**Fig. 5C, S5G**). We conclude that the polyQ amyloid nucleus is an intramolecular Q zipper.

Discussion

PolyQ amyloid begins *within* a molecule

The initiating molecular event leading to pathogenic aggregates is the most important yet intractable step in the progression of age-associated neurodegenerative diseases such as Huntington's. 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. Under the assumption of homogeneous nucleation, this implies that pathogenic nucleation occurs within a monomer (Bhattacharyya et al., 2005; Chen et al., 2002; Kar et al., 2011). However, that conclusion, along with the assumption it rests on, contrasts with observations that polyQ can form apparently non-amyloid multimers prior to amyloids in vitro (Crick et al., 2006; Vitalis and Pappu, 2011), and that phase separation can promote amyloid formation

9

(Babinchak and Surewicz, 2020; Camino et al., 2021; Shin and Brangwynne, 2017; Sprunger and Jackrel, 2021). By explicitly manipulating concentration, length, density and conformational heterogeneities, we found that, indeed, the rate-limiting step for pathologic polyQ amyloid formation occurs within a monomer. Importantly, it does so even in the dynamic, intermingled, living cellular environment.

Once formed, this minimal Q zipper germinates in all three dimensions to ultimately produce amyloid fibers with lamellar zippers longer than that of the nucleus proposed here. Both features have been confirmed by structural studies (Boatz et al., 2020; Galaz-Montoya et al., 2021; Hoop et al., 2016; Nazarov et al., 2022). The difference in critical lengths between intra- and intermolecular Q zippers follows from the fact that intramolecular segments are covalently constrained to much higher effective concentrations. In the case of polyQ, the short-stranded nucleus is merely a catalyst for longer, lamellar Q zippers that make up the amyloid core. That the nascent amyloid should become progressively more ordered is in keeping with decades of experimental and theoretical work by polymer physicists (Keller and O'Connor, 1957; Xu et al., 2021; M. Zhang et al., 2016). The minimally competent zipper identified here provides lower bounds on the widths of fibers -- they can be as thin as two sheets (1.6 nm) and as short as six residues. This hypothetical restriction is thinner and shorter than the cores of all known amyloid structures of full length polypeptides (Sawaya et al., 2021). Remarkably, observations from recent cryoEM tomography studies closely match this prediction: pathologic polyQ or huntingtin amyloids, whether in cells or in vitro, feature a slab-like architecture with restrictions as thin as 2 nm (Galaz-Montoya et al., 2021) and as short as five residues (Nazarov et al., 2022).

Multidimensional polymer crystal growth offers a simple explanation for the characteristically antiparallel arrangement of strands in polyQ amyloid. Each lamellum first nucleates then propagates along the templating lamellum by the back-and-forth folding of monomers (M. Zhang et al., 2016). In the context of polyQ, the end result is a stack of beta hairpins. If lamella are responsible for antiparallel strands, then Q_U sequences should form typical amyloid fibers with parallel beta strands. The only atomic resolution structure of a Q zipper amyloid happens to be of a protein with striking unilateral (but not bilateral) contiguity, and the strands are indeed arranged in parallel (Hervas et al., 2020).

Seven of the nine polyQ diseases have length thresholds at or above (Lieberman et al., 2019) the minimum length for an intramolecular Q zipper as deduced here. The two polyQ diseases with length thresholds below that of the intramolecular Q zipper, SCA2 (32 residues) and SCA6 (21 residues), are also atypical for this class of diseases in that disease onset is accelerated in homozygous individuals (Laffita-Mesa et al., 2012; Mariotti et al., 2001; Soga et al., 2017; Spadafora et al., 2007; Tojima et al., 2018). This is consistent with nucleation occurring in oligomers, as would necessarily be the case for such short polyQ tracts.

We expect that the monomeric nucleus of polyQ will prove to be unusual among pathologic amyloid-forming proteins. Among the hundreds of amyloid-forming proteins we have now characterized by DAmFRET (Khan et al., 2018; Posey et al., 2021), and unpublished), only Cyc8 PrD has a similar concentration-dependence. We now attribute this to self-poisoning due to its exceptionally long tract of 50 unilaterally contiguous Qs, primarily in the form of QA dipeptide repeats.

Proteotoxic multimers are likely to be self-poisoned polymer crystals

Soluble oligomers accumulate during the aggregation of pathologically lengthened 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). They 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 multimers, rather than the other way around. Notwithstanding one report to the contrary (Peskett et al., 2018) -- which we respectfully attribute to a known artifact of the FLAG tag when fused to polyQ (Duennwald et al., 2006a, 2006b; Jiang et al., 2017) -- our findings corroborate prior demonstrations that polyQ (absent flanking domains) 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).

Our findings that kinetically arrested aggregates emerge from the same nucleating event responsible for amyloid formation suggests a resolution to the paradox that polyQ diseases are rate-limited by amyloid nucleation despite the implication of non-amyloid species (Kim et al., 2016; Leitman et al., 2013; Lu and Palacino, 2013; Matlahov and van der Wel, 2019; Takahashi et al., 2008; Wetzel, 2020). Likewise, our discovery that polyQ amyloid formation is blocked by oligomerization or phase separation, together with demonstrations that full length Huntingtin protein with a pathologically lengthened polyQ tract can phase separate in cells (Aktar et al., 2019; de Mattos et al., 2022; Peskett et al., 2018; Wan et al., 2021), suggests a simple explanation for the otherwise paradoxical fact that Huntington's Disease onset and severity do not increase in homozygous individuals (Cubo et al., 2019; Lee et al., 2019; Wexler et al., 1987). Specifically, if mutant Huntingtin forms an endogenous condensate (if only in some microcompartment of the cell) when expressed from just one allele, then the buffering effect of phase separation (Klosin et al., 2020) will prevent a second allele from increasing the concentration of nucleation-competent monomers.

Our data indirectly illuminate the nature of these arrested multimers. For Q_U, the multimers failed to accumulate to a level detectable by AmFRET, and failed to coalesce to microscopic puncta. They therefore seem to involve only a small fraction of the total protein, presumably limited to soluble species that have acquired a nascent Q zipper. The disordered polyQ globule is extremely viscous and this causes very slow conformational conversion on the tips of growing fibers (Bhattacharyya et al., 2005; Walters et al., 2012). It should therefore be highly susceptible to self-poisoning as has been widely observed and studied for polymer crystals in vitro (Jiang et al., 2016; Ungar and Keller, 1987; Whitelam et al., 2016; Zhang et al., 2020, 2018). In short, the multimers are nascent amyloids that templated their own arrest in the earliest stages of growth following nucleation. We do not yet know if these aborted amyloids contribute to pathology. However, the role of contralateral zippers in their formation is consistent with the apparent protective effect of CAT trinucleotide insertions that preserve unilateral contiguity (QHQHQH) in the polyQ disease protein, SCA1 (Menon et al., 2013; Nethisinghe et al., 2018; Sen et al., 2003), and the absence of bilateral contiguity in the only known functional Q zipper, formed by the neuronal translational regulator, Orb2 (Hervas et al., 2020). We speculate that the toxicity of polyQ amyloid arises from its lamellar architecture.

Concluding remarks

The etiology of polyQ pathology has been elusive. Decades-long efforts by many labs have revealed precise measurements of polyQ aggregation kinetics, atomistic details of the amyloid structure, a catalog of proteotoxic candidates, and snapshots of the conformational preferences of disordered polyQ. What they have not yet led to are treatments. We synthesized those insights to recognize that pathogenesis likely begins with a very specific conformational fluctuation. We set out to characterize the nature of that event in the cellular milieu, deploying a technique we developed to do just that, and arrived at an intramolecular four-stranded polymer

11

crystal. Our findings rationalize key aspects of polyQ diseases, such as length thresholds, kinetics of progression, and involvement of pre-amyloid multimers. More importantly, they illuminate a new avenue for potential treatments. Current therapeutic efforts focused on lowering the levels of mutant Huntingtin have not been successful. As an admittedly radical alternative, we suggest that therapies designed to (further) oligomerize huntingtin *preemptively* will delay nucleation and thereby decelerate the disease.

Acknowledgments

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). See **Table S1** for all full-length protein sequences.

Yeast strain rhy3078a ([*PIN*+]) was constructed as follows. We first deleted *PDR5* and *ATG8* from strain rhy1713 (Khan et al., 2018) by sequentially mating and sporulating it with the respective strains from the *MATa* deletion collection (Open Biosystems). We then used PCR-based mutagenesis (Goldstein and McCusker, 1999) with template vector CX (Miller et al. submitted) to integrate BDFP1.6:1.6 prior to the stop codon of chromosomal *PGK1*. Yeast strain rhy3082 ([*pin*-]) is rhy3078a with the amyloid form of Rnq1 eliminated by passaging it four times on YPD plates containing 3 mM GdHCl, a prion-curing agent (Ferreira et al., 2001).

DAmFRET

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 1000 rpm at 30°C. Following overnight growth, cells were spun down 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. 75 µL of cells were then re-arrayed from a 96 well plate to a 384 well plate and photoconverted, while shaking at 800 rpm, 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.

High-throughput flow cytometry (10 uL/well, flow speed 1.8) was performed on the Bio-Rad ZE5 with a Propel automation setup and a GX robot (PAA Inc). 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 detectors. Acceptor signal was collected from a 561 nm laser set to 50 mW with voltage at 525, into a 589/15 detector. Autofluorescence was collected from a 405 nm laser at 100 mW and voltage 450, into a 460/22 detector.

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. Acceptor fluorescence intensities were divided by side scatter (SSC), a proxy for cell volume (Miller et al. submitted), to convert them to concentration (in arbitrary units).

Imaging flow cytometry was conducted as in previous work (Khan et al., 2018; Venkatesan et al., 2019).

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

13

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.

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. Line profiles of the protein smear were quantified using the following procedure. All images were processed in Fiji (https://imagej.net/software/fiji/). Gel images were first background subtracted using a rolling ball with a radius of 200 pixels. The images were then rotated so that the orientation was perfectly aligned. A user-defined line was drawn on the image down the first lane. A 10 pixel wide line profile was generated using an in house written plugin "polyline kymograph jru v1". This line was then programmatically shifted to every other lane and line profiles were generated for each. Each line profile was then normalized to the integral under the line profile using "normalize trajectories jru v1". From these line profiles, csv sheets were generated and these were imported to python to make line profile plots.

Amyloid Prediction

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.

Figures

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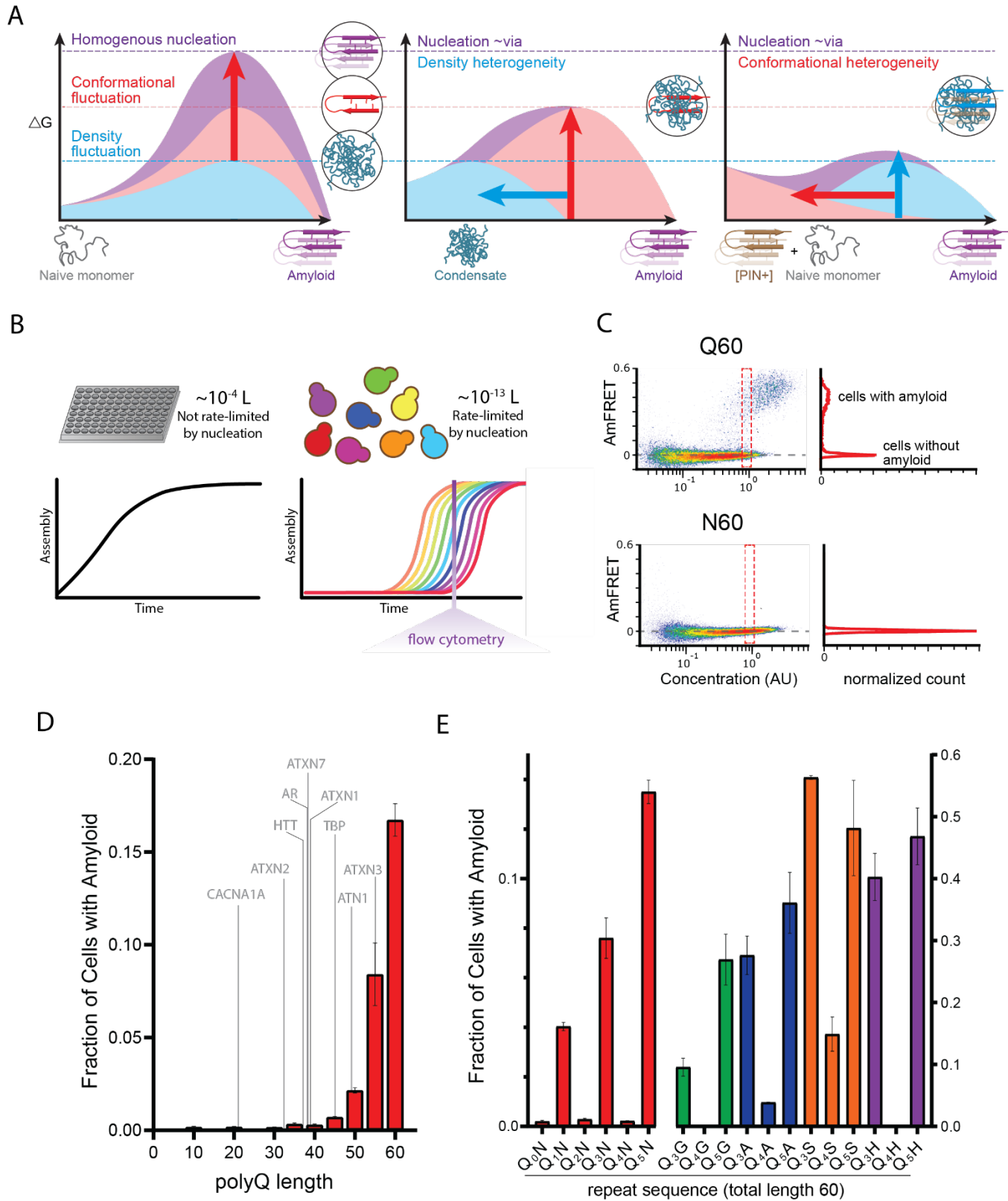
Figure 1. A hidden pattern of glutamines governs amyloid nucleation.

A. Reaction coordinate diagram schematizing the energy barriers for forming an amyloid nucleus homogeneously (left), facilitated by a density heterogeneity (middle), or facilitated by a conformational heterogeneity (right). Because amyloid involves a transition from disordered monomer to ordered multimer, the nucleation barrier (purple) results from a combination of high energy fluctuations in both density (blue) and conformation (red).

B. Cellular volumes quantize amyloid nucleation. Amyloid nuclei occur at such low concentrations that fewer than one exists in the femtoliter volumes of cells. This causes amyloid formation to be rate-limited by stochastic nucleation in individual yeast cells (right) but not in the microliter volumes of conventional in vitro kinetic assays (left). Taking a population-level snapshot of the extent of protein self-assembly as a function of concentration in each cell reveals heterogeneity attributable to the nucleation barrier.

C. DAmFRET plots, showing the extent of *de novo* self-assembly (AmFRET) as a function of protein concentration for polyglutamine (Q60) or polyasparagine (N60) in yeast cells, lacking endogenous amyloid ([*pin*-]). Cells expressing Q60 partition into distinct populations that either lack (no AmFRET) or contain (high AmFRET) amyloid. The bimodal distribution persists even among cells with the same concentration of protein, indicating that amyloid formation is rate-limited by nucleation. The nucleation barrier for N60 is so large that spontaneous amyloid formation occurs at undetectable frequencies. Insets show histograms of AmFRET values. AU, arbitrary units.

D. Bar plot of the fraction of [*pin*-] cells in the AmFRET-positive population for the indicated length variants of polyQ, along with the pathologic length thresholds for polyQ tracts in the indicated proteins. Shown are means +/- SEM of biological triplicates. CACNA1A, Cav2.1; ATXN2, Ataxin-2; HTT, Huntingtin; AR, Androgen Receptor; ATXN7, Ataxin-7; ATXN1, Ataxin-1; TBP, TATA-binding protein; ATN1, Atrophin-1; ATXN3, Ataxin-3.

E. Bar plot of the fraction of [*pin*-] cells in the AmFRET-positive population for the indicated sequences, showing that amyloid is inhibited when Q tracts are interrupted by a non-Q side chain at odd-numbered intervals. Shown are means +/- SEM of biological triplicates.

17
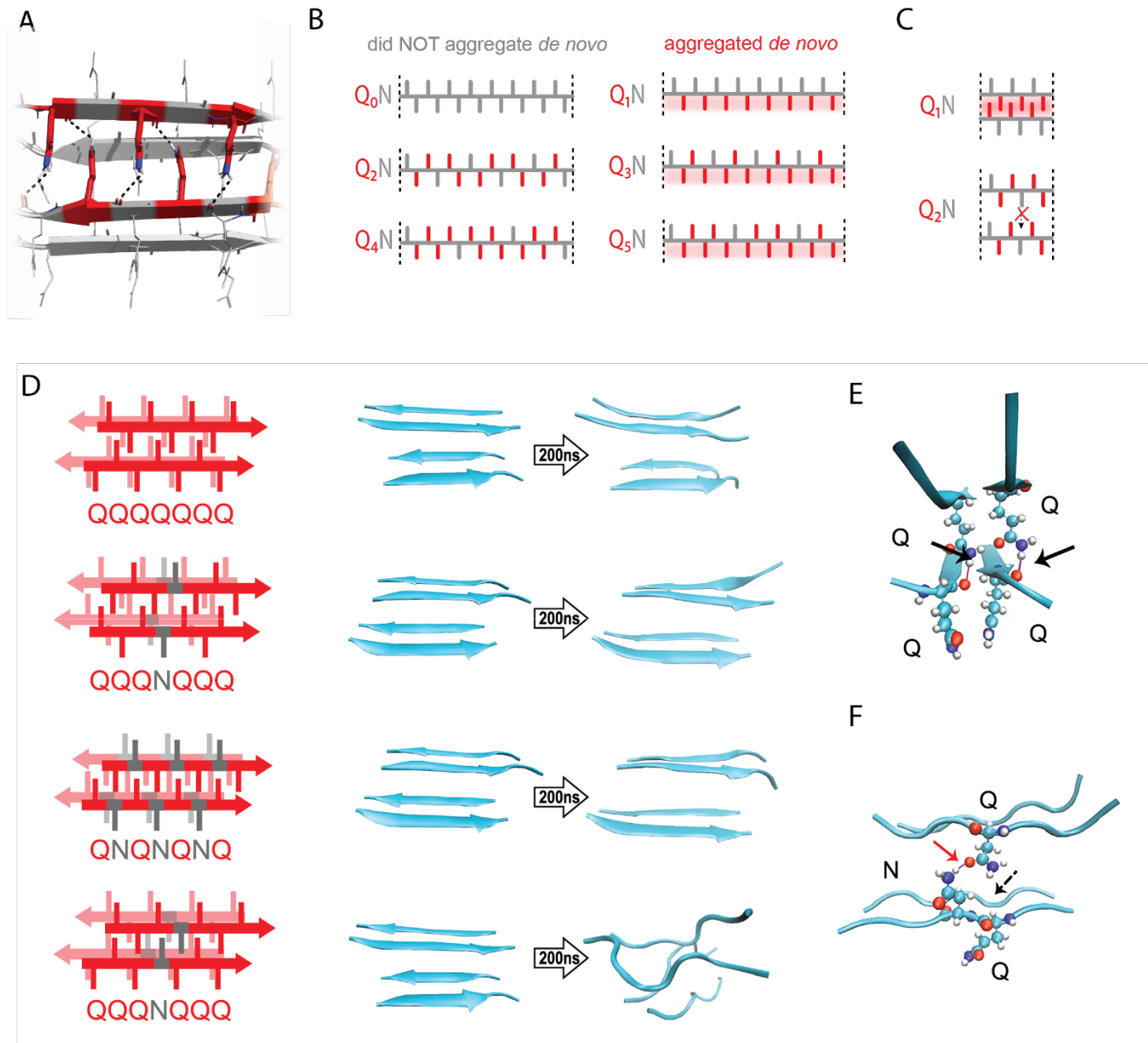
Figure 2

Figure 2. The pattern encodes a single, singular steric zipper.

A. A view down the axis of a local segment of all-glutamine steric zipper, or "Q zipper", between two antiparallel two-stranded sheets. Residues with internally facing side-chains on the top layer are colored red to emphasize interdigitation and H-bonding (dashed lines) between the terminal amides and the opposing backbone.

B. Schema of the "odd dependency" for Q zipper formation, showing side chain arrangements along a continuous beta strand for sequences composed of tandem repeats of Qs (red) interrupted by single Ns (gray). Shading highlights contiguous stretches of Qs that would occur in a continuous beta strand. Note that the illustrated strands will not necessarily be continuous in the context of the nucleus; i.e. the nucleus may contain shorter strands connected by loops.

C. Schema of the tertiary contacts between two beta strands, as in a steric zipper. The zipper can be formed only when the single interrupting non-Q residue follows an odd number of Qs (e.g. Q_1N), but not when it follows an even number of Qs (e.g. Q_2N).

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 even a single substitution faces 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
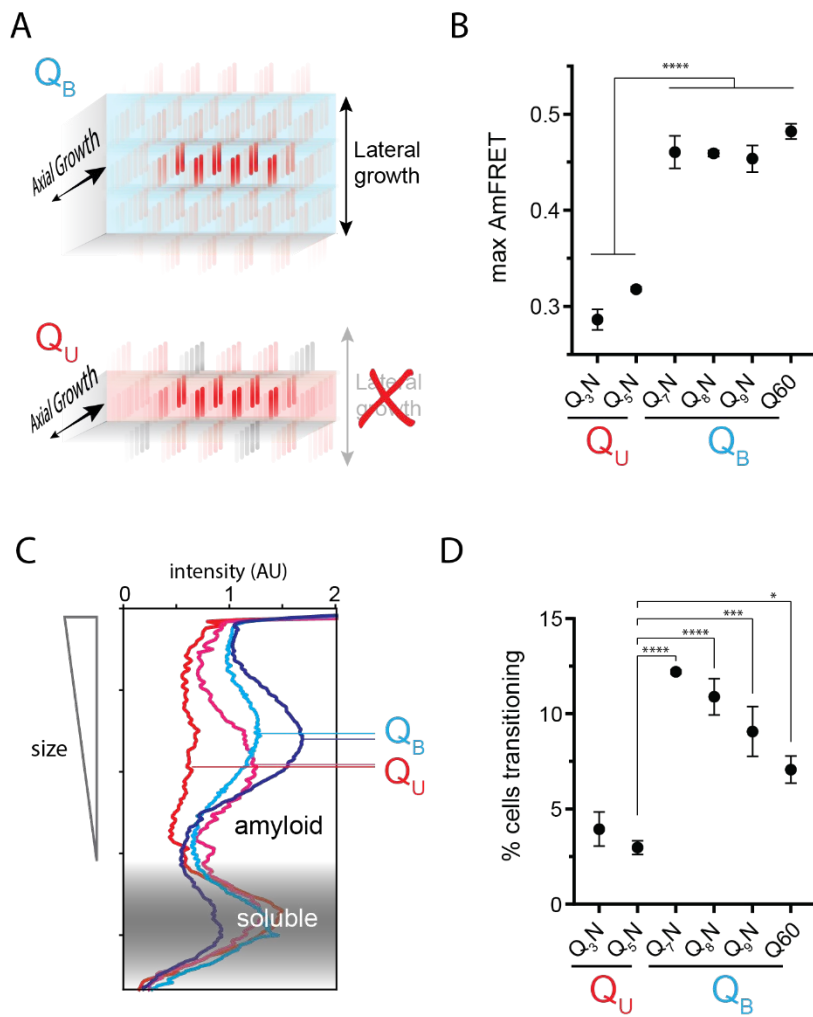
Figure 3. The Q zipper grows in two dimensions.

A. Schematic illustrating how sequences with bilaterally contiguous Qs (Q_B) can hypothetically allow for lateral growth (secondary nucleation) of Q zippers giving rise to lamellar amyloid fibers. In contrast, sequences with only unilateral contiguous Qs (Q_U) can form amyloids with only a single Q zipper.
B. Maximum AmFRET values for the indicated sequences in [*pin*⁻] cells, suggesting that Q_B amyloids have a greater subunit density. Shown are means +/- SEM of the median AmFRET values of triplicates. **** p < 0.0001; ANOVA and Dunnett's multiple comparison test.
C. Densitometric analysis of SDD-AGE characterizing amyloid length distributions for the indicated Q_U and Q_B amyloids, showing that Q_B amyloid particles are larger. Data are representative of multiple experiments.
D. Fraction of cells at intermediate AmFRET values for the indicated sequences in [*pin*⁻] cells, suggesting that Q_B amyloids grow slower. Shown are means +/- SEM of the percentage of cells between lower and upper populations, of triplicates. ****, ***, * p < 0.0001, < 0.001, < 0.05; ANOVA.

Figure 4
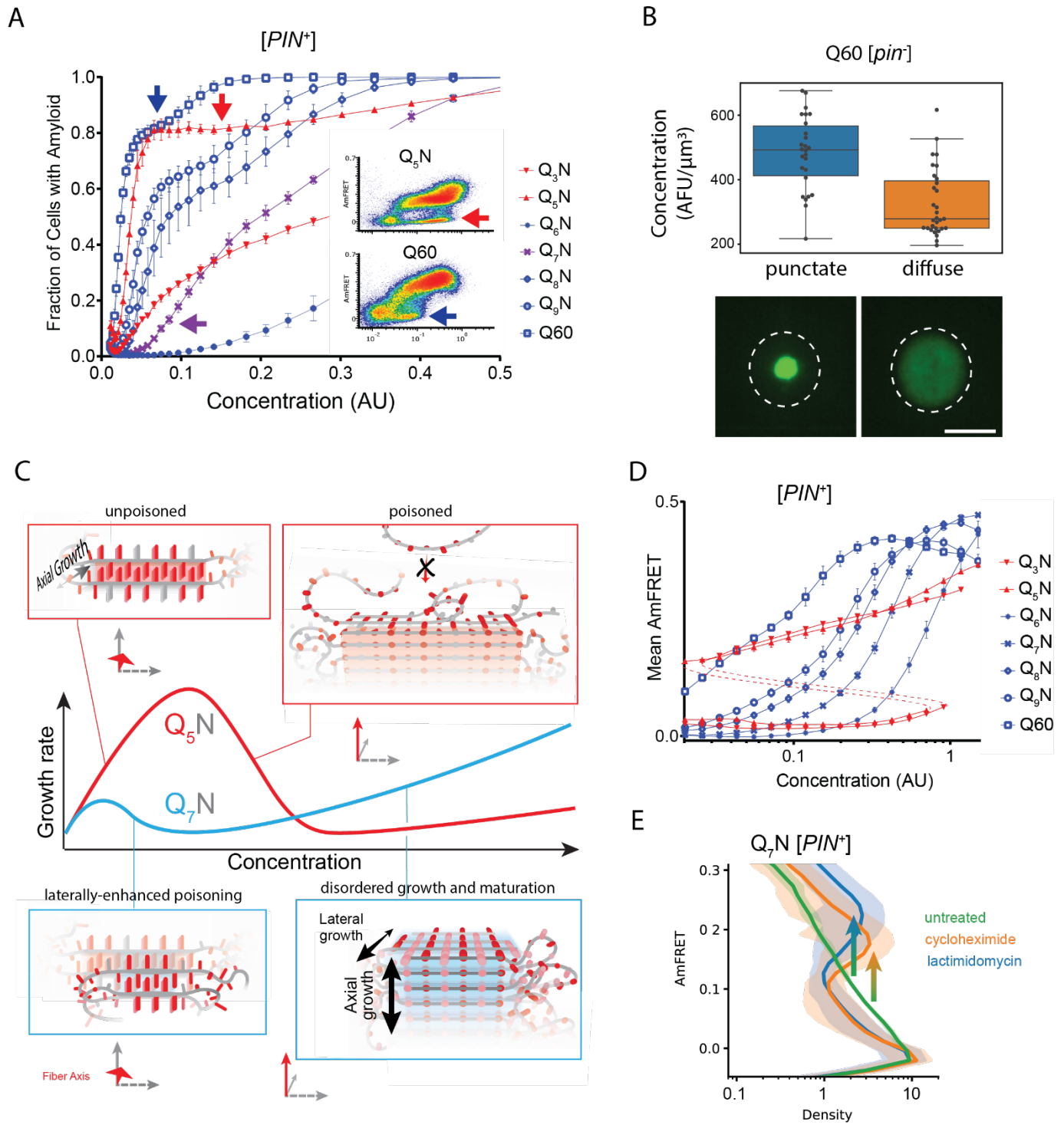
Figure 4. Q zippers poison themselves.

A. Fraction of [*PIN*[+]] cells in the AmFRET-positive population as a function of concentration for the indicated sequences. Arrows denote the population of cells with self-poisoned aggregation (inset), and the corresponding plateaus in the relationship of amyloid formation to concentration. The purple arrow highlights the sharp reduction in aggregation for Q_7N relative to Q_5N, which we attribute to enhanced poisoning as a result of intramolecular Q zipper formation. Shown are means +/- SEM of triplicates.

B. Distribution of cytosolic concentrations (AFU/μm^3) of Q60 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. Schematic illustrating self-poisoned growth as a function of concentration for Q zippers of Q_5N and Q_7N. Conformational conversion of Q_5N to amyloid decelerates (becomes poisoned) at high concentrations, as a consequence of polypeptides interfering with each other's conversion on the templating surface. This is illustrated here by the red trace and inset showing entangled, partially ordered polypeptides on the axial surface. The presence of contralaterally contiguous Qs in Q_7N exacerbates poisoning at low concentrations, as illustrated here by the blue trace and inset showing partially-ordered species immobilized with bilateral zippers. Growth resumes at high concentrations through relatively disordered deposition with subsequent slow ordering.

D. Graph of spline fits of AmFRET values for the indicated sequences in [*PIN*[+]] cells. The upper and lower populations of Q_3N and Q_5N were treated separately due to the extreme persistence of the low population for these sequences. The red dashed lines denote these are subpopulations of the same samples. The ability of amyloid to grow at low concentrations fell sharply with the onset of bilateral contiguity at Q_6N and then gradually increased with higher q values. Shown are means +/- SEM of triplicates.

E. Histogram of AmFRET values for Q_7N -expressing cells transitioning from the low to high populations (boxed region from DAmFRET plot in Fig. S4D) upon translation inhibition for six hours following 18 hours of expression. Shown are means +/- 95% CI of biological triplicates. Blocking new protein synthesis prior to analysis causes AmFRET to rise, whether by cycloheximide or lactimidomycin ($p < 0.01$, < 0.05, respectively, Dunnett's test).

Figure 5
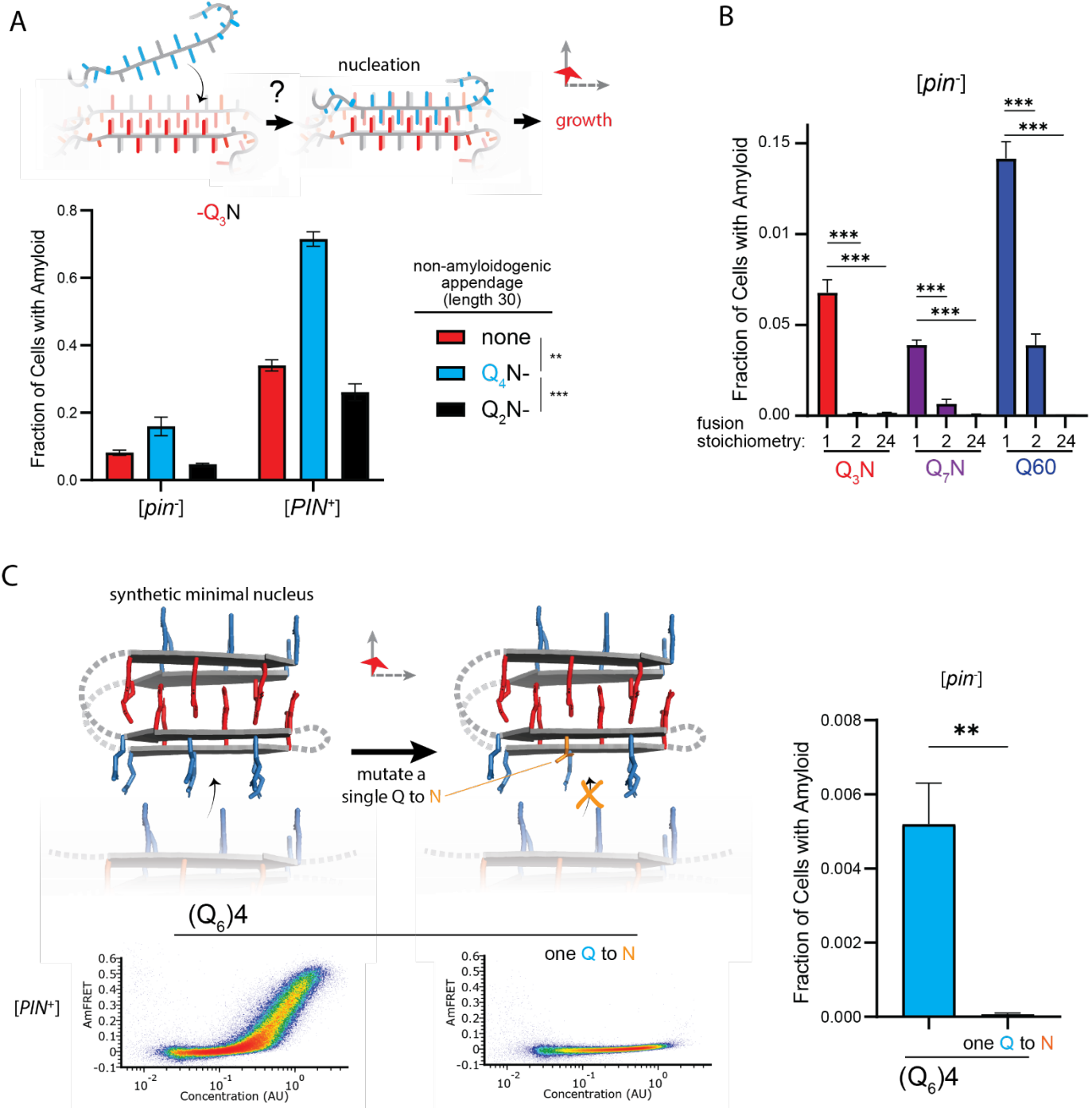
Figure 5. The nucleus forms within a single molecule.

A. Fraction of cells in the AmFRET-positive population when expressing Q_3N with the indicated non-amyloidogenic fusions. Shown are means +/- SEM of triplicates. ***, ** $p < 0.001$, < 0.01; ANOVA.

B. Fraction of cells in the AmFRET-positive population (with higher AmFRET than that of the oligomer itself) when expressing the indicated protein fused to proteins with the indicated stoichiometry. Shown are means +/- SEM of triplicates. *** $p < 0.001$; t-test.

C. Schema, DAmFRET plots, and quantitation of amyloid formation by a synthetic minimal polyQ amyloid-forming sequence. Q side chains in the nucleating zipper are colored red, while those necessary for growth of the zipper -- which requires lateral propagation due to its short length -- are colored blue. The three G3 loops are represented by dashed gray lines; the actual topology of the loops may differ. Mutating a single Q to N blocks amyloid formation. Shown are means +/- SEM of triplicates. ** $p < 0.01$; t-test.

Supplemental Files

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

Table S2. Amyloid predictor output.

Figure S1

A

B
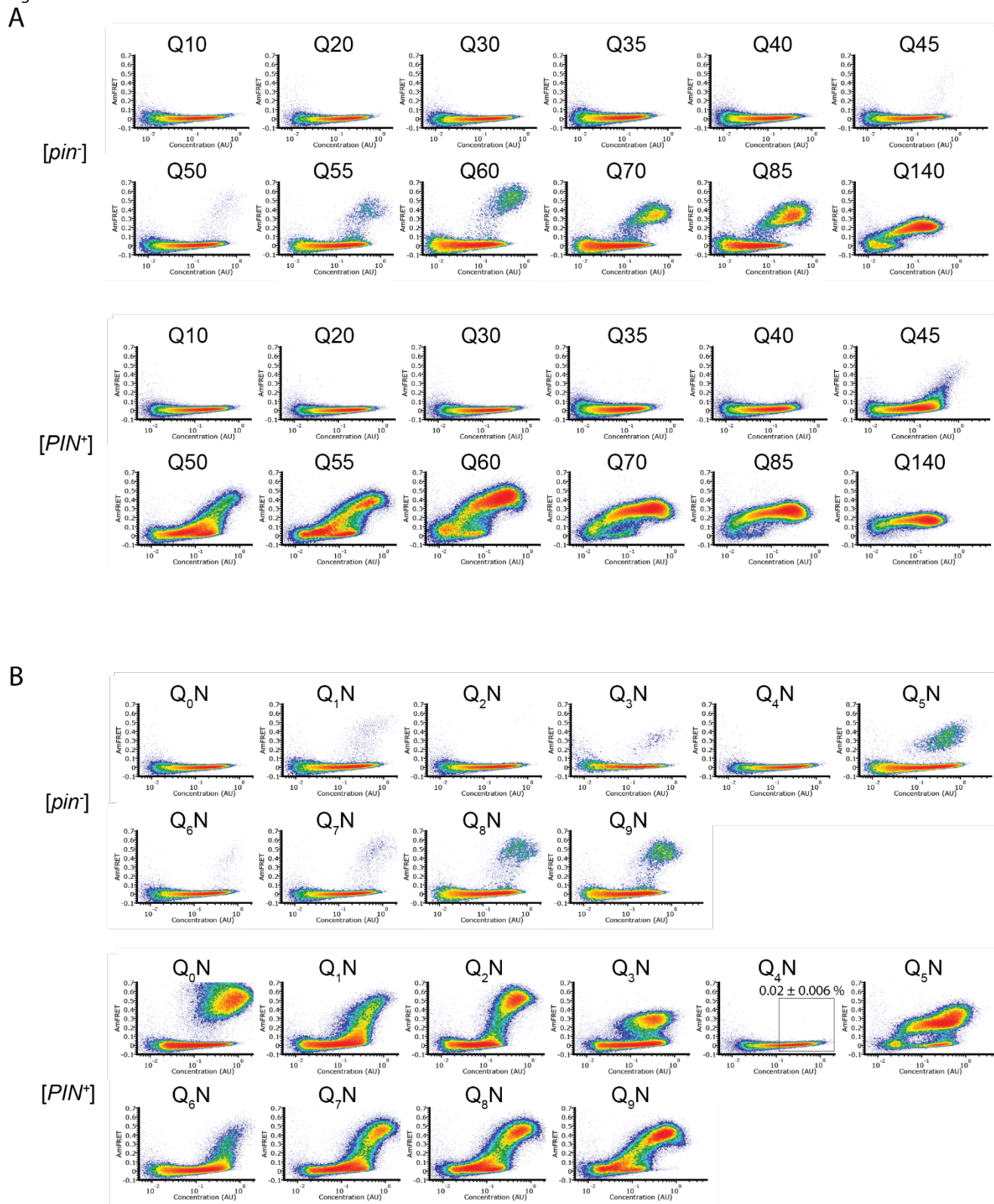
C
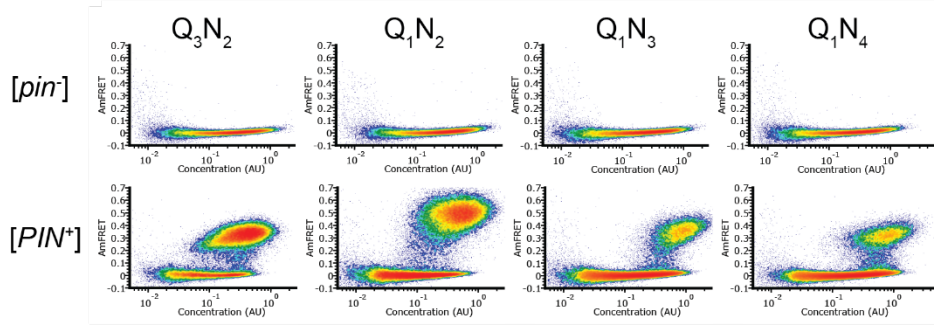
D
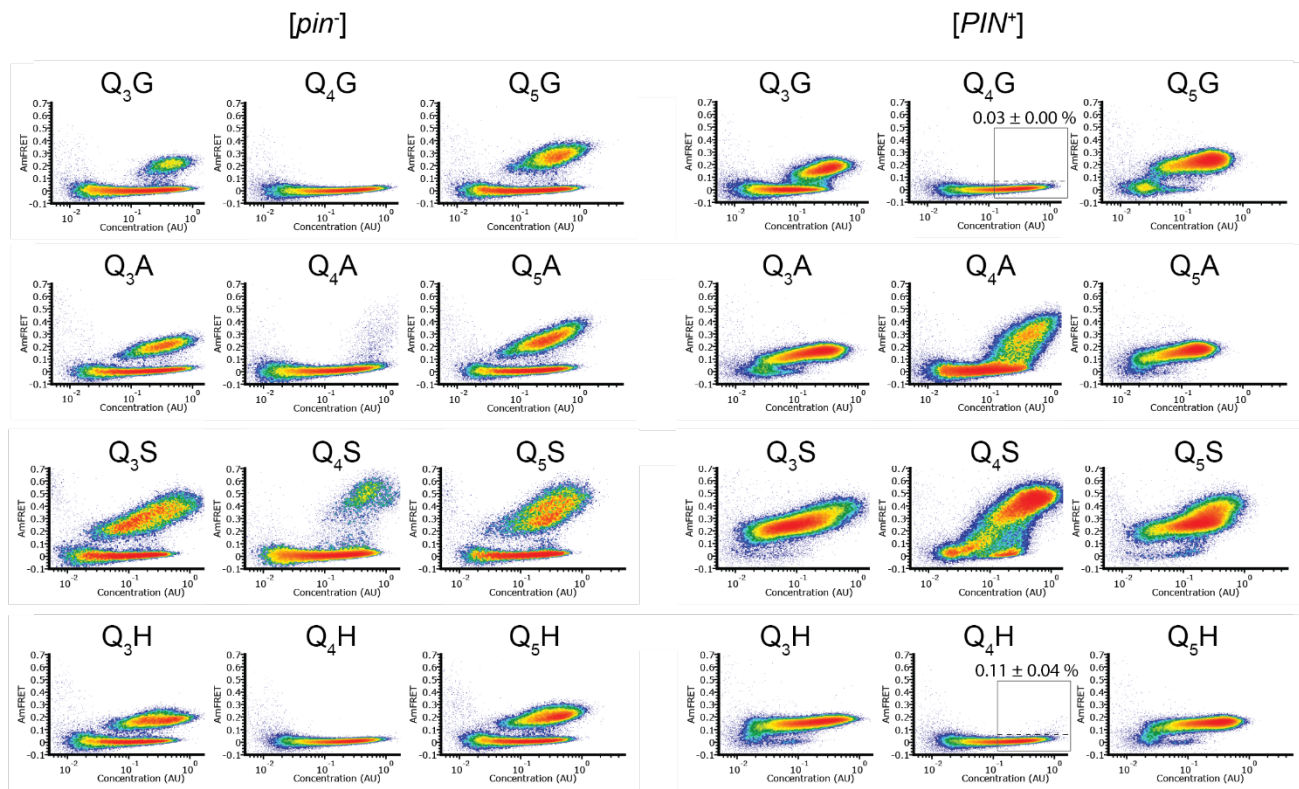
Figure S1.

A. DAmFRET plots of polyQ length variants. Shown are representative plots of biological triplicates.

B. DAmFRET plots of polypeptides composed of tandem repeats of q (subscripted) Qs separated by an N for a total length of 60 residues. Plots are representative of biological triplicates.

C. DAmFRET plots of polypeptides composed of tandem repeats of the indicated N-rich sequences, for a total length of 60 residues, showing negligible nucleation in the absence of a conformational template. Note that because the nominal pattern repeats, "Q_1N_2", "Q_1N_3", and "Q_1N_4" are synonymous to "N_2Q_1", "N_3Q_1", and "N_4Q_1", respectively. Plots are representative of biological triplicates.

D. DAmFRET plots of polypeptides composed of tandem repeats of the indicated sequences, for a total length of 60 residues, showing that Q_3X and Q_5X have a greater amyloid propensity than Q_4X regardless of the identity of X. Labels above the boxed regions of the [*PIN*$^+$] Q_4N, Q_4G, and Q_4H plots indicate the percentage of cells in the high-FRET region, revealing rare but significant nucleation for the latter (p = 0.046 versus Q_4N, one-tailed T-test). Plots are representative of biological triplicates.

Figure S2

A

QQQSQQQ

QQQSQQQ

B

top sheet
QQQQNQQ
QQNQQQQ
QQQQNQQ
QQNQQQQ

bottom sheet
QNQQQQQ
QQQQQNQ
QNQQQQQ
QQQQQNQ

top sheet
QQQQSQQ
QQSQQQQ
QQQQSQQ
QQSQQQQ

bottom sheet
QSQQQQQ
QQQQQSQ
QSQQQQQ
QQQQQSQ
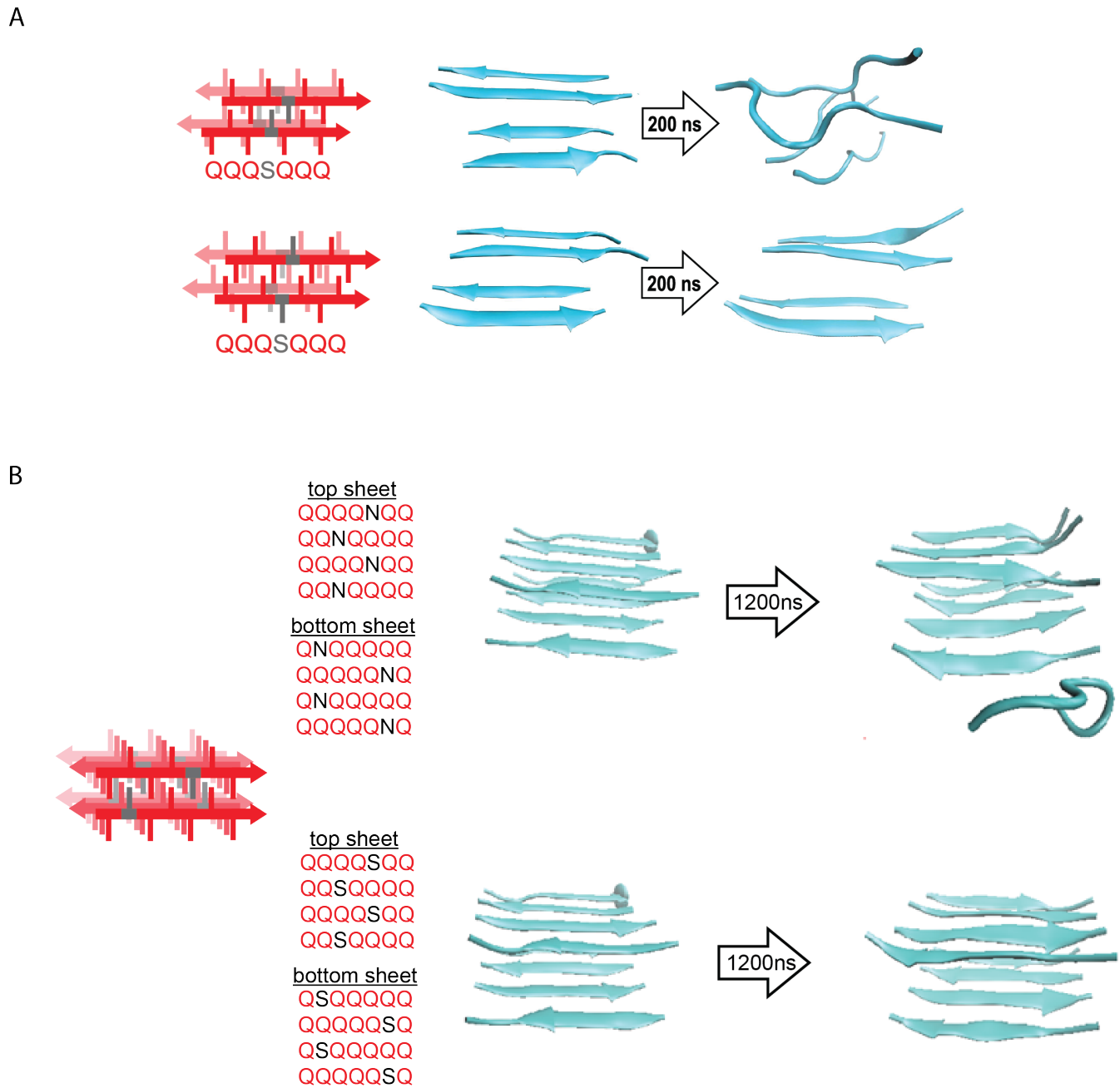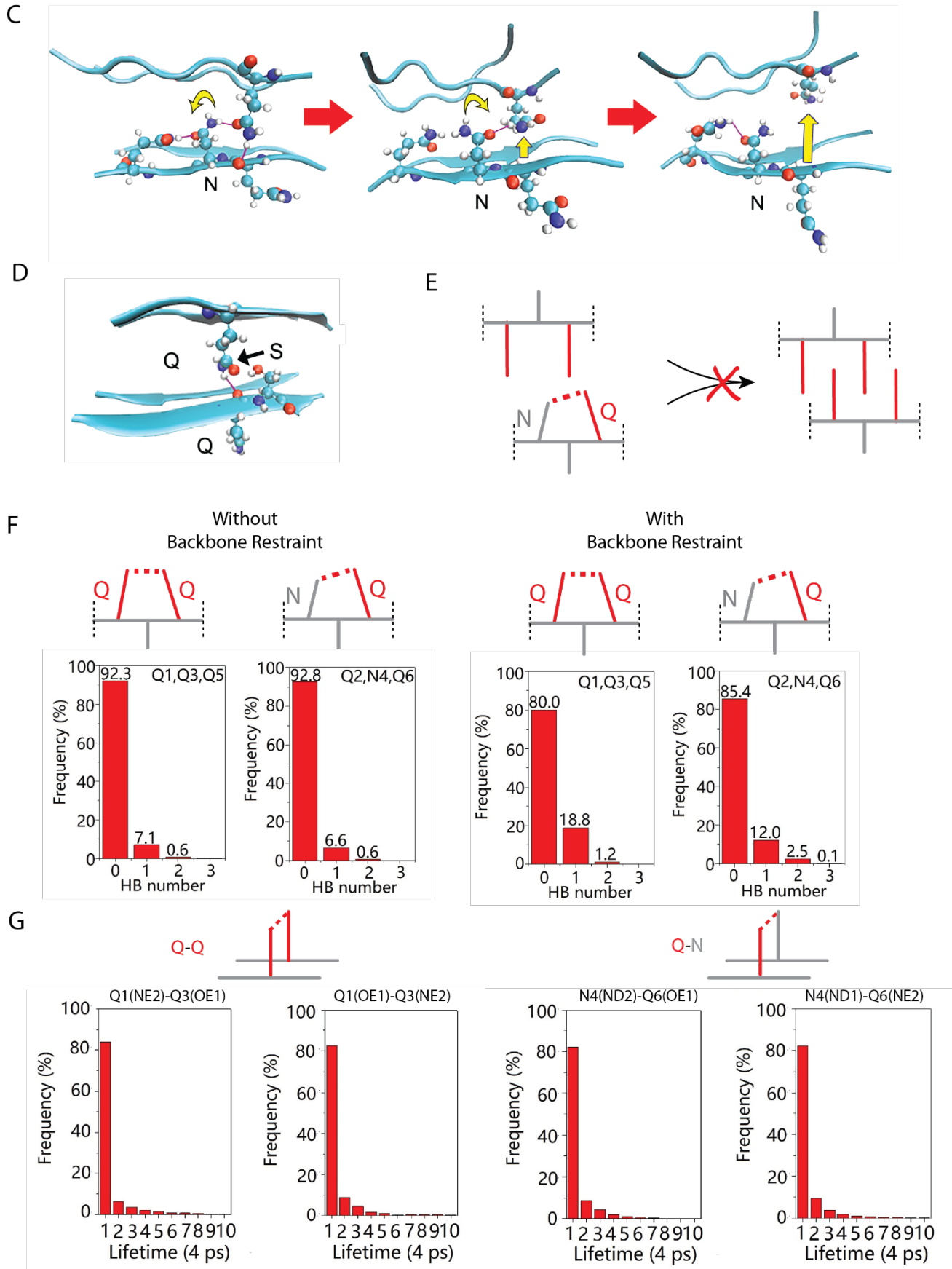
Figure S2.

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 antiparallel beta sheets, containing a single asparagine (top) or serine (bottom) residue per strand. The structure proved less stable in the case of asparagine.

C. 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.

D. As for N, the side chain of S is too short to H-bond with the opposing backbone. 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.

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

F. 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.

G. 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.

Figure S3
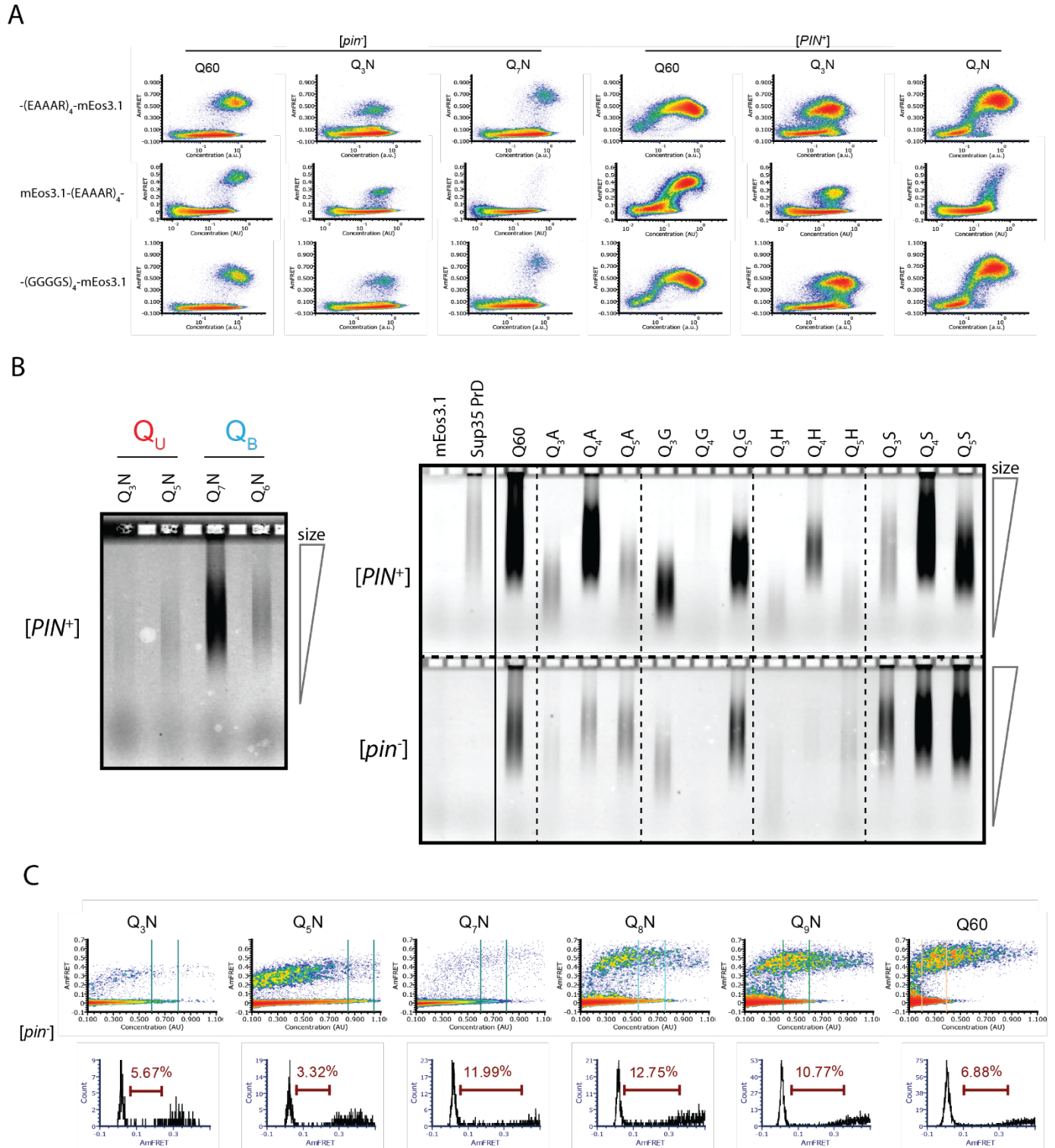
Figure S3.

A. DAmFRET plots of the indicated sequence variants with either a C-terminal (EAAAR)$_4$-mEos3.1 fusion, as used throughout this work, an N-terminal mEos3.1-(EAAAR)$_4$ fusion, or a C-terminal (GGGGS)$_4$-mEos3.1 fusion, showing that the sequence-specific differences in relative steady state AmFRET levels do not depend on the linker or terminus fused. Plots are representative of biological triplicates.

B. Fluorescence images of SDD-AGE gels showing the size distributions of SDS-resistant complexes of the indicated mEos3.1-tagged proteins. Left: raw data quantified in **Fig. 3C**. Right: Additional Q$_3$X, Q$_4$X, Q$_5$X proteins, showing that Q$_U$ amyloids are consistently smaller than other amyloids, such as those of Q60 and Sup35 PrD (which only nucleates in [*PIN*⁺] cells). The solid line shows where the image was spliced, although all lanes are from the same gel. Lysates were normalized by fluorescence to within 50% of each other prior to loading.

C. Histograms of AmFRET values for the indicated gates (at the respective approximate EC50s) for the indicated sequences. The brown gate on the histograms shows the percentage of transitioning cells.

Figure S4
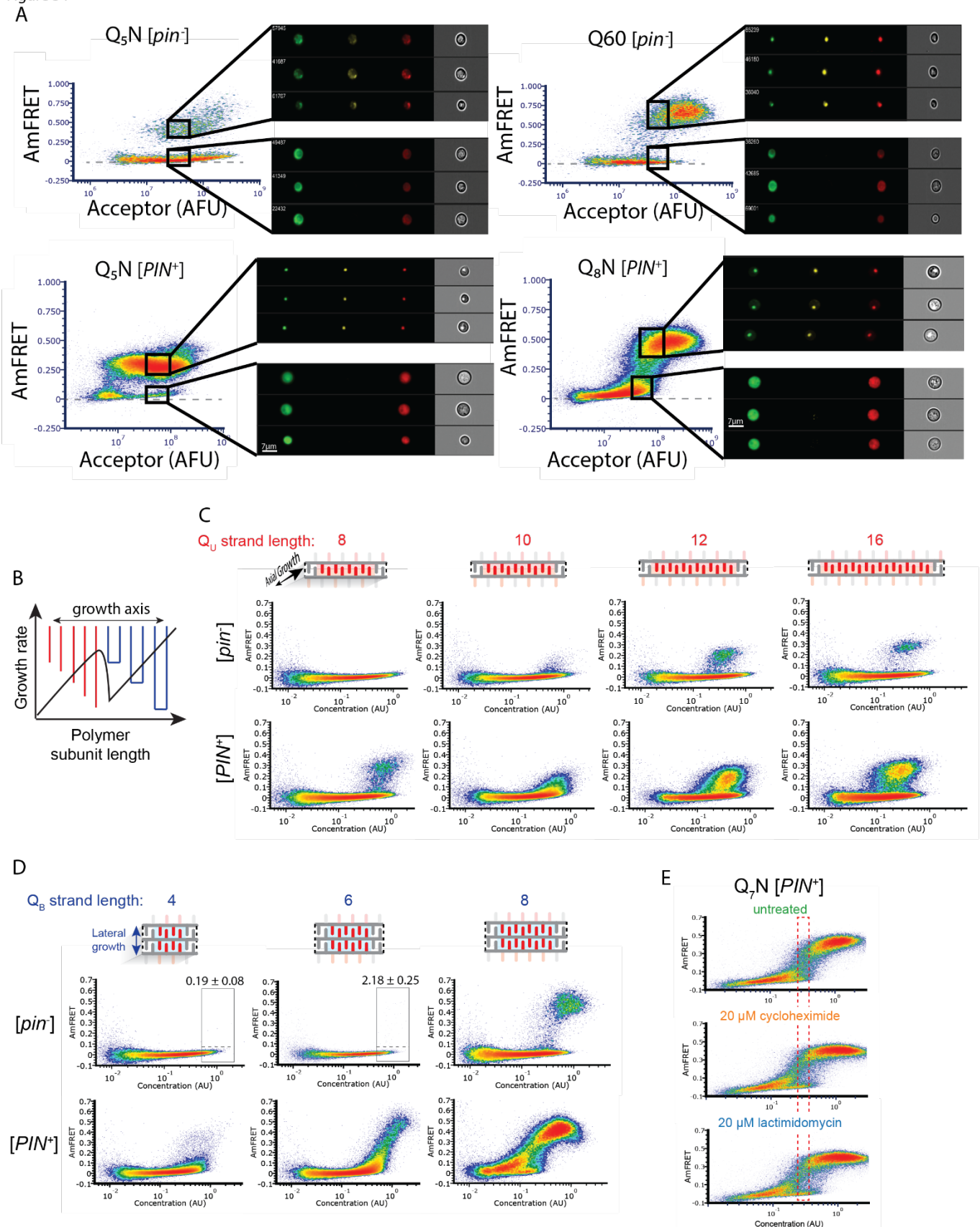
Figure S4.

A. DAmFRET plots of Q_5N, $Q60$, and 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. Schematic of the rate of polymer crystallization as a function of length, showing a sharp deceleration when the polymer length is equally compatible with either of two polymorphs. Adapted from (Ungar et al., 2005).
C. DAmFRET plots of [*pin*] cells expressing unilateral contiguity variants of the Q_3N base sequence, showing that at least five unilaterally contiguous glutamines (see schematic) are required for *de novo* nucleation of single long Q zipper amyloids. Plots are representative of biological triplicates.
D. DAmFRET plots of bilateral contiguity variants (Q_4N_2, Q_6N_2, Q_8N_2), showing that at least six bilaterally contiguous Qs are required for *de novo* amyloid formation. Numbers indicate the percentage of cells in the high-FRET boxed region, revealing significant nucleation by Q_6N_2 (p = 0.0004, T-test).
E. DAmFRET plots of Q_7N in [*pin*] cells treated as indicated for six hours prior to analysis. The boxed region was used to compute histograms of AmFRET in Fig. 4E.

Figure S5
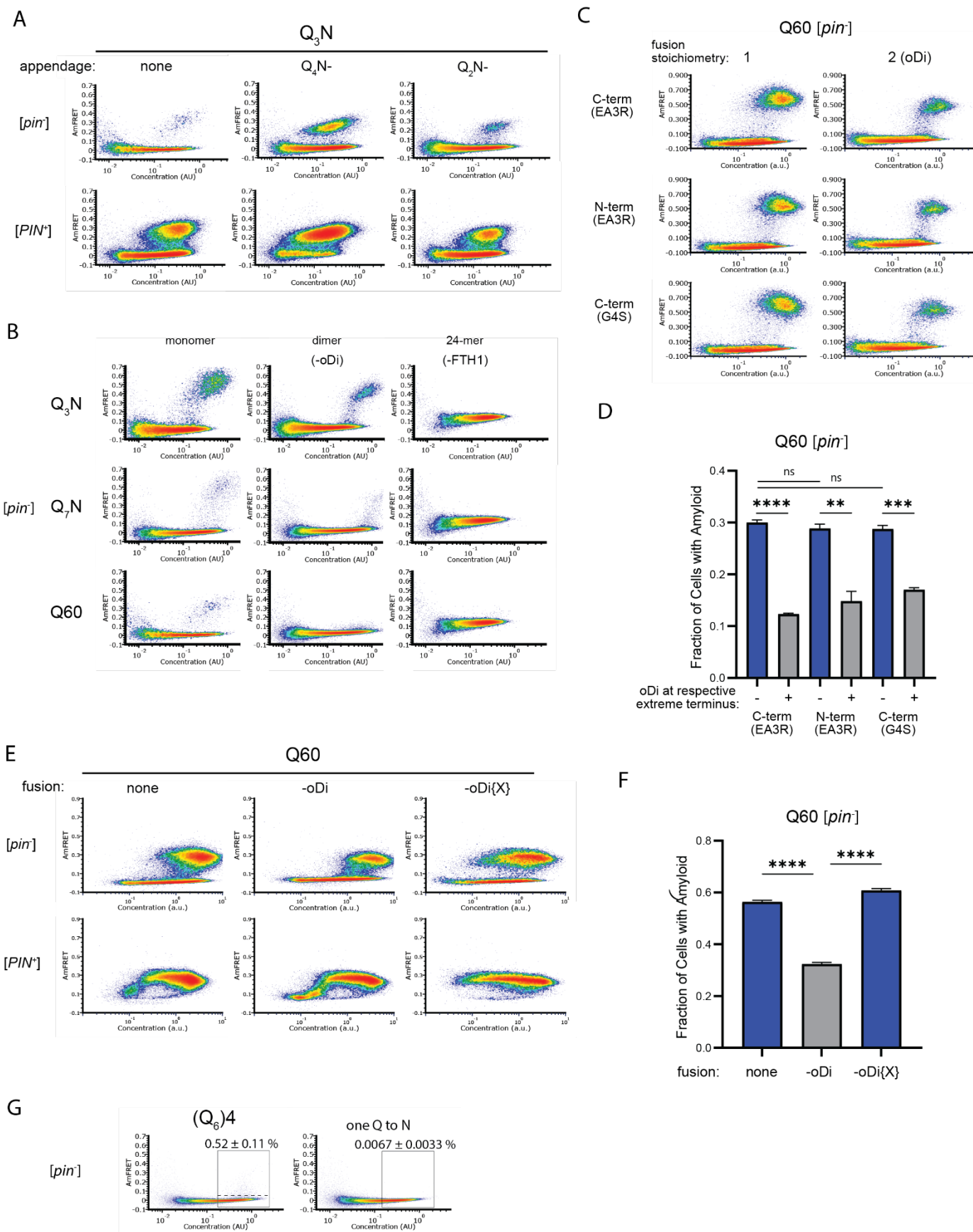
Figure S5.

A. DAmFRET plots of cells expressing Q_3N (length 60) with the indicated appendage (length 30). Plots are representative of biological triplicates.
B. DAmFRET plots of [*pin*⁻] and [*PIN*⁻] cells expressing the indicated sequences either unfused ("monomer") or fused to oDi ("dimer") for FTH1 (24-mer). Plots are representative of biological triplicates.
C. DAmFRET plots of cells expressing Q60 either with or without oDi and with the indicated linkers and termini of the fusion. Plots are representative of biological triplicates.
D. Quantification of the data in C), showing that oDi reduces Q60 nucleation irrespective of the linker and terminus it is fused to. Shown are means +/- SEM. **, ***, **** p < 0.01, < 0.001, < 0.0001; t-test.
E. DAmFRET plots of cells expressing Q60 either with or without oDi or a monomeric mutant of oDi (oDi{X}). Plots are representative of biological triplicates.
F. Quantification of the data in E), showing that the monomerizing mutation eliminates the amyloid-inhibiting effect of oDi. Shown are means +/- SEM. **** p < 0.0001; t-test.
G. DAmFRET plots of [*pin*⁻] cells expressing a synthetic minimal polyQ amyloid-forming sequence, or the same sequence with the tenth Q mutated to N. Quantitation is the same as in Fig. 5C. Plots are representative of biological triplicates.

Movie S1. Simulation of the pure Q zipper shown in **Fig. 2D** (top model), showing its persistence.

Movie S2. Simulation of the N-substituted Q zipper shown in **Fig. 2D** (bottom model), showing the motions of side chains summarized in **Fig. S2C**.

Bibliography

Aktar F, Burudpakdee C, Polanco M, Pei S, Swayne TC, Lipke PN, Emtage L. 2019. The huntingtin inclusion is a dynamic phase-separated compartment. *Life Sci Alliance* **2**. doi:10.26508/lsa.201900489

Alexandrov AI, Polyanskaya AB, Serpionov GV, Ter-Avanesyan MD, Kushnirov VV. 2012. The effects of amino acid composition of glutamine-rich domains on amyloid formation and fragmentation. *PLoS ONE* **7**:e46458. doi:10.1371/journal.pone.0046458

Alexandrov IM, Vishnevskaya AB, Ter-Avanesyan MD, Kushnirov VV. 2008. Appearance and propagation of polyglutamine-based amyloids in yeast: tyrosine residues enable polymer fragmentation. *J Biol Chem* **283**:15185–15192. doi:10.1074/jbc.M802071200

Arrasate M, Mitra S, Schweitzer ES, Segal MR, Finkbeiner S. 2004. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431**:805–810. doi:10.1038/nature02998

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

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

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

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

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

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 CA, 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 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

Cubo E, Martinez-Horta S-I, Santalo FS, Descalls AM, Calvo S, Gil-Polo C, Muñoz I, Llano K, Mariscal N, Diaz D, Gutierrez A, Aguado L, Ramos-Arroyo MA, European HD Network. 2019. Clinical manifestations of homozygote allele carriers in Huntington disease. *Neurology* **92**:e2101–e2108. doi:10.1212/WNL.0000000000007147

Derkatch IL, Bradley ME, Hong JY, Liebman SW. 2001. Prions affect the appearance of other prions: the story of [PIN(+)]. *Cell* **106**:171–182. doi:10.1016/s0092-8674(01)00427-5

de Mattos EP, Musskopf MK, Bergink S, Kampinga HH. 2022. *In vivo* suppression of polyglutamine aggregation via co-condensation of the molecular chaperone DNAJB6. *BioRxiv*. doi:10.1101/2022.08.23.504914

Dougan L, Li J, Badilla CL, Berne BJ, Fernandez JM. 2009. Single homopolypeptide chains collapse into mechanically rigid conformations. *Proc Natl Acad Sci USA* **106**:12605–12610. doi:10.1073/pnas.0900678106

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 D, Jucker M. 2012. The amyloid state of proteins in human diseases. *Cell* **148**:1188–1203. doi:10.1016/j.cell.2012.02.022

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

Esposito L, Paladino A, Pedone C, Vitagliano L. 2008. Insights into structure, stability, and toxicity of monomeric and aggregated polyglutamine models from molecular dynamics simulations. *Biophys J* **94**:4031–4040. doi:10.1529/biophysj.107.118935

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

Ferreira PC, Ness F, Edwards SR, Cox BS, Tuite MF. 2001. The elimination of the yeast [PSI+] prion by guanidine hydrochloride is the result of Hsp104 inactivation. *Mol Microbiol* **40**:1357–1369. doi:10.1046/j.1365-2958.2001.02478.x

Fletcher JM, Boyle AL, Bruning M, Bartlett GJ, Vincent TL, Zaccai NR, Armstrong CT, Bromley EHC, Booth PJ, Brady RL, Thomson AR, Woolfson DN. 2012. A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Synth Biol* **1**:240–250. doi:10.1021/sb300028q

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

Gabryelczyk B, Alag R, Philips M, Low K, Venkatraman A, Kannaian B, Shi X, Linder M, Pervushin K, Miserez A. 2022. In vivo liquid-liquid phase separation protects amyloidogenic and aggregation-prone peptides during overexpression in Escherichia coli. *Protein Sci* **31**:e4292. doi:10.1002/pro.4292

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

Goldstein AL, McCusker JH. 1999. Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. *Yeast*.

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

Hu W. 2018. The physics of polymer chain-folding. *Physics Reports* **747**:1–50. doi:10.1016/j.physrep.2018.04.004

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 X, Reiter G, Hu W. 2016. How Chain-Folding Crystal Growth Determines the Thermodynamic Stability of Polymer Crystals. *J Phys Chem B* **120**:566–571. doi:10.1021/acs.jpcb.5b09324

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

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, Luan B, Zhou R. 2018. Glassy dynamics in mutant huntingtin proteins. *J Chem Phys* **149**:072333. doi:10.1063/1.5029369

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, 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

Keller A, O'Connor A. 1957. Large Periods in Polyethylene: the Origin of Low-Angle X-ray Scattering. *Nature* **180**:1289–1290. doi:10.1038/1801289a0

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

Klosin A, Oltsch F, Harmon T, Honigmann A, Jülicher F, Hyman AA, Zechner C. 2020. Phase separation provides a mechanism to reduce noise in cells. *Science* **367**:464–468. doi:10.1126/science.aav6691

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

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

Küffner AM, Linsenmeier M, Grigolato F, Prodan M, Zuccarini R, Capasso Palmiero U, Faltova L, Arosio P. 2021. Sequestration within biomolecular condensates inhibits Aβ-42 amyloid formation. *Chem Sci* **12**:4373–4382. doi:10.1039/d0sc04395h

Kuiper EFE, de Mattos EP, Jardim LB, Kampinga HH, Bergink S. 2017. Chaperones in Polyglutamine Aggregation: Beyond the Q-Stretch. *Front Neurosci* **11**:145. doi:10.3389/fnins.2017.00145

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

Lipiński WP, Visser BS, Robu I, Fakhree MAA, Lindhoud S, Claessens MMAE, Spruijt E. 2022. Biomolecular condensates can both accelerate and suppress aggregation of α-synuclein. *Sci Adv* **8**:eabq6495. doi:10.1126/sciadv.abq6495

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

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

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

Margittai M, Langen R. 2008. Fibrils with parallel in-register structure constitute a major class of amyloid fibrils: molecular insights from electron paramagnetic resonance spectroscopy. *Q Rev Biophys* **41**:265–297. doi:10.1017/S0033583508004733

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

Menon RP, Nethisinghe S, Faggiano S, Vannocci T, Rezaei H, Pemble S, Sweeney MG, Wood NW, Davis MB, Pastore A, Giunti P. 2013. The role of interruptions in polyQ in the pathology of SCA1. *PLoS Genet* **9**:e1003648. doi:10.1371/journal.pgen.1003648

Meriin AB, Zhang X, He X, Newnam GP, Chernoff YO, Sherman MY. 2002. Huntington toxicity in yeast model depends on polyglutamine aggregation mediated by a prion-like protein Rnq1. *J Cell Biol* **157**:997–1004. doi:10.1083/jcb.200112104

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

Miller J, Arrasate M, Brooks E, Libeu CP, Legleiter J, Hatters D, Curtis J, Cheung K, Krishnan P, Mitra S, Widjaja K, Shaby BA, Lotz GP, Newhouse Y, Mitchell EJ, Osmand A, Gray M, Thulasiramin V, Saudou F, Segal M, Finkbeiner S. 2011. Identifying polyglutamine protein species in situ that best predict neurodegeneration. *Nat Chem Biol* **7**:925–934. doi:10.1038/nchembio.694

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

Nazarov S, Chiki A, Boudeffa D, Lashuel HA. 2022. Structural basis of huntingtin fibril polymorphism revealed by cryogenic electron microscopy of exon 1 HTT fibrils. *J Am Chem Soc*. doi:10.1021/jacs.2c00509

Nethisinghe S, Pigazzini ML, Pemble S, Sweeney MG, Labrum R, Manso K, Moore D, Warner J, Davis MB, Giunti P. 2018. Polyq tract toxicity in SCA1 is length dependent in the absence of CAG repeat interruption. *Front Cell Neurosci* **12**:200. doi:10.3389/fncel.2018.00200

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

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

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

Osherovich LZ, Cox BS, Tuite MF, Weissman JS. 2004. Dissection and design of yeast prions. *PLoS Biol* **2**:E86. doi:10.1371/journal.pbio.0020086

Otzen D, Riek R. 2019. Functional Amyloids. *Cold Spring Harb Perspect Biol* **11**. doi:10.1101/cshperspect.a033860

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

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, Jakubek RS, Workman RJ, Asher SA. 2018. Interaction enthalpy of side chain and backbone amides in polyglutamine solution monomers and fibrils. *J Phys Chem Lett* **9**:1944–1950. doi:10.1021/acs.jpclett.8b00348

Sadler DM. 1983. Roughness of growth faces of polymer crystals: Evidence from morphology and implications for growth mechanisms and types of folding. *Polymer* **24**:1401–1409. doi:10.1016/0032-3861(83)90220-3

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, Hughes MP, Rodriguez JA, Riek R, Eisenberg DS. 2021. The expanding amyloid family: Structure, stability, function, and pathogenesis. *Cell* **184**:4857–4873. doi:10.1016/j.cell.2021.08.013

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

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

Sen S, Dash D, Pasha S, Brahmachari SK. 2003. Role of histidine interruption in mitigating the pathological effects of long polyglutamine stretches in SCA1: A molecular approach. *Protein Sci* **12**:953–962. doi:10.1110/ps.0224403

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

Serpionov GV, Alexandrov AI, Antonenko YN, Ter-Avanesyan MD. 2015. A protein polymerization cascade mediates toxicity of non-pathological human huntingtin in yeast. *Sci Rep* **5**:18407. doi:10.1038/srep18407

Sharma D, Shinchuk LM, Inouye H, Wetzel R, Kirschner DA. 2005. Polyglutamine homopolymers having 8-45 residues form slablike beta-crystallite assemblies. *Proteins* **61**:398–411. doi:10.1002/prot.20602

Shin Y, Brangwynne CP. 2017. Liquid phase condensation in cell physiology and disease. *Science* **357**. doi:10.1126/science.aaf4382

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

Sprunger ML, Jackrel ME. 2021. Prion-Like Proteins in Phase Separation and Their Link to Disease. *Biomolecules* **11**. doi:10.3390/biom11071014

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

Ungar G, Keller A. 1987. Inversion of the temperature dependence of crystallization rates due to onset of chain folding. *Polymer* **28**:1899–1907. doi:10.1016/0032-3861(87)90298-9

Ungar G, Putra EGR, de Silva DSM, Shcherbina MA, Waddon AJ. 2005. The Effect of Self-Poisoning on Crystal Morphology and Growth Rates In: Allegra G, editor. Interphases and Mesophases in Polymer Crystallization I, Advances in Polymer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 45–87. doi:10.1007/b107232

Vekilov PG. 2012. Phase diagrams and kinetics of phase transitions in protein solutions. *J Phys Condens Matter* **24**:193101. doi:10.1088/0953-8984/24/19/193101

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

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

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 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

Wei M-T, Elbaum-Garfinkle S, Holehouse AS, Chen CC-H, Feric M, Arnold CB, Priestley RD, Pappu RV, Brangwynne CP. 2017. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat Chem* **9**:1118–1125. doi:10.1038/nchem.2803

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

Wetzel R. 2006. Nucleation of huntingtin aggregation in cells. *Nat Chem Biol* **2**:297–298. doi:10.1038/nchembio0606-297

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

Whitelam S, Dahal YR, Schmit JD. 2016. Minimal physical requirements for crystal growth self-poisoning. *J Chem Phys* **144**:064903. doi:10.1063/1.4941457

Xu J, Reiter G, Alamo R. 2021. Concepts of nucleation in polymer crystallization. *Crystals* **11**:304.

    doi:10.3390/cryst11030304

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

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 M, Guo B-H, Xu J. 2016. A review on polymer crystallization theories. *Crystals* **7**:4. doi:10.3390/cryst7010004

Zhang S, Wang Z, Guo B, Xu J. 2021. Secondary nucleation in polymer crystallization: A kinetic view. *Polymer Crystallization* **4**. doi:10.1002/pcr2.10173

Zhang X, Marxsen SF, Ortmann P, Mecking S, Alamo RG. 2020. Crystallization of Long-Spaced Precision Polyacetals II: Effect of Polymorphism on Isothermal Crystallization Kinetics. *Macromolecules* **53**:7899–7913. doi:10.1021/acs.macromol.0c01443

Zhang X, Zhang W, Wagener KB, Boz E, Alamo RG. 2018. Effect of Self-Poisoning on Crystallization Kinetics of Dimorphic Precision Polyethylenes with Bromine. *Macromolecules* **51**:1386–1397. doi:10.1021/acs.macromol.7b02745

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

Figure panel	plasmid name	Annotated Sequence	extra N-term tag	linker	N-term Tag	N-term linker	Inserted Peptide Sequence	linker	C-term Tag	linker	extra tag
3 C; S3 B	rhx0935	mEos3.1							mEos3.1		
3 C; S3 B	rhx0974	Sup35 PrD					MSDSNQGNNQ QNYQQYSQNG NQQQGNNRYQ GYQAYNAQAQ PAGGYYQNYQ GYSGYQQGGY QQYNPDAGYQ QQYNPQGGYQ QYNPQGGYQQ QFNPQGGRGN YKNFNYNNNL QGYQAGFQPQ SQGMSLNDFQ KQQKQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 D	rhx1177a	Q85					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 D	rhx1177d	Q20					QQQQQQQQQQ QQQQQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 D	rhx1177e	Q30					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 D	rhx1177f	Q70					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 D	rhx1177g	Q10					QQQQQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 D	rhx1177h	Q55					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 D	rhx1177j	Q140					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 D	rhx1177l	Q35					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 D	rhx1177m	Q40					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 D	rhx1177n	Q45					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 D	rhx1177o	Q50					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E; 3 B,D; 4 A,D	rhx2602	Q_3N					QQQNQQQNQQ NQQQNQQQNQ QNQQQNQQQN QQNQQQNQQQ NQQQNQQQNQ QNQQQNQQQN	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx2603	Q_4N					QQQQNQQQQN QQQQNQQQQN QQQQNQQQQN QQQQNQQQQN QQQQNQQQQN QQQQNQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E; 3 B,D; S4 A	rhx2604	Q_5N					QQQQQNQQQQ QNQQQQQNQQ QQQNQQQQQN QQQQQNQQQQ QNQQQQQNQQ QQQNQQQQQN	EAAAREAAAR EAAAREAAAR	mEos3.1		
3 B,D; 4 A,D,E	rhx2605	Q_7N					QQQQQQQNQQ QQQQQNQQQQ QQQNQQQQQQ QNQQQQQQQN QQQQQQQNQQ QQQQQNQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
4 A,D	rhx2606	Q_9N					QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 C,E	rhx2681	N60/ Q_0N					NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 C; 4 A,B,D	rhx2682	Q60					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx2683	Q_2N					QQNQQNQQNQ QNQQNQQNQQ NQQNQQNQQN QQNQQNQQNQ QNQQNQQNQQ NQQNQQNQQN	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx2684	Q_1N					QNQNQNQNQN QNQNQNQNQN QNQNQNQNQN QNQNQNQNQN QNQNQNQNQN QNQNQNQNQN	EAAAREAAAR EAAAREAAAR	mEos3.1		
S3 B; 4 A,D	rhx2685	Q_6N					QQQQQQNQQQ QQQNQQQQQQ NQQQQQQNQQ QQQQNQQQQQ QNQQQQQQNQ QQQQQNQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
S1 C	rhx2686	Q_2N_3					NNNQQNNNQQ NNNQQNNNQQ NNNQQNNNQQ NNNQQNNNQQ NNNQQNNNQQ NNNQQNNNQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
S1 C	rhx2687	Q_2N_4					NNNNQQNNNN QQNNNNQQNN NNQQNNNNQQ NNNNQQNNNN QQNNNNQQNN NNQQNNNNQQNN	EAAAREAAAR EAAAREAAAR	mEos3.1		
3 B,D; S3 B,C; 4 A,D	rhx3071	Q_8N					QQQQQQQQNQ QQQQQQQNQQ QQQQQQNQQQ QQQQQNQQQQ QQQQNQQQQQ QQQNQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx3117	$Q_{15}S$					QQQSQQQSQQ QSQQQSQQQS QQQSQQQSQQ QSQQQSQQQS QQQSQQQSQQ QSQQQSQQQS	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx3118	Q_4S					QQQQSQQQQS QQQQSQQQQS QQQQSQQQQS QQQQSQQQQS QQQQSQQQQS QQQQSQQQQS	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx3119	Q_5S					QQQQQSQQQQ QSQQQQQSQQ QQQSQQQQQS QQQQQSQQQQ QSQQQQQSQQ QQQSQQQQQS	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E; 3 C	rhx3265	Q_3A					QQQAQQQAQQ QAQQQAQQQA QQQAQQQAQQ QAQQQAQQQA QQQAQQQAQQ QAQQQAQQQA	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E; 3 C	rhx3266	Q_4A					QQQQAQQQQA QQQQAQQQQA QQQQAQQQQA QQQQAQQQQA QQQQAQQQQA QQQQAQQQQA	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E; 3 C	rhx3267	Q_5A					QQQQQAQQQQ AQQQQQAQQQ QAQQQQQAQQ QQAQQQQQAQ QQQAQQQQQA QQQQAQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx4276	Q_3H					QQQHQQQHQQ QHQQQHQQQH QQQHQQQHQQ QHQQQHQQQH QQQHQQQHQQ QHQQQHQQQH QQHQQQHQQQH QHQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx4277	Q_4H					QQQQHQQQQH QQQQHQQQQH QQQQHQQQQH QQQQHQQQQH QQQQHQQQQH QQQQHQQQQH QQQQHQQQQH QQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx4278	Q_5H					QQQQQHQQQQ QHQQQQQHQQ QQQHQQQQQH QQQQQHQQQQ QHQQQQQHQQ QQQHQQQQQH QQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx3453	Q_3G					QQQGQQQGQQ QGQQQGQQQG QQQGQQQGQQ QGQQQGQQQG QQQGQQQGQQ QGQQQGQQQG	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx3454	Q_4G					QQQQGQQQQG QQQQGQQQQG QQQQGQQQQG QQQQGQQQQG QQQQGQQQQG QQQQGQQQQG	EAAAREAAAR EAAAREAAAR	mEos3.1		
1 E	rhx3455	Q_5G					QQQQQGQQQQ GQQQQQGQQQ QQGQQQQQGQ QQQQGQQQQQ GQQQQQGQQQ QQGQQQQQGQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
S3 A	rhx3989	Q60			mEos3.1	EAAAREAAAR EAAAREAAAR	QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ				
S3 A	rhx4710	Q_3N			mEos3.1	EAAAREAAAR EAAAREAAAR	QQQNQQQNQQ NQQQNQQQNQ QNQQQNQQQN QQNQQQNQQQ NQQQNQQQNQ QNQQQNQQQN				
S3 A	rhx4068	Q_9N			mEos3.1	EAAAREAAAR EAAAREAAAR	QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN				
S3 A	rhx4584	Q60					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	GGGGSGGGGS GGGGSGGGGS	mEos3.1		
S3 A	rhx4585	Q_3N					QQQNQQQNQQ NQQQNQQQNQ QNQQQNQQQN QQNQQQNQQQ NQQQNQQQNQ QNQQQNQQQN	GGGGSGGGGS GGGGSGGGGS	mEos3.1		
S3 A	rhx4586	Q_9N					QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN	GGGGSGGGGS GGGGSGGGGS	mEos3.1		
S4 B	rhx3450	$Q_3NQ_3N_2$					QQQNQQQNNQ QQNQQQNNQQ QNQQQNNQQQ NQQQNNQQQN QQQNNQQQNQ QQNNQQQNQQ QNNQQQN				
S4 B	rhx3451	$Q_3NQ_3N Q_3N_2$					QQQNQQQNQQ QNNQQQNQQQ NQQQNNQQQN QQQNQQQNNQ QQNQQQNQQQ NNQQQNQQQN QQQNN				
S4 B	rhx3452	$Q_3NQ_3N Q_3N_2Q_3N_2$					QQQNQQQNQQ QNNQQQNNQQ QNQQQNQQQN NQQQNNQQQN QQQNQQQNNQ QQNNQQQNQQ QNQQQNN				
S4 C	rhx2608	Q_4N_2					QQQQNNQQQQ NNQQQQNNQQ QQNNQQQQNN QQQQNNQQQQ NNQQQQNNQQ QQNNQQQQNN	EAAAREAAAR EAAAREAAAR	mEos3.1		
S4 C	rhx3789	Q_5N_2					QQQQQNNQQQ QQNNQQQQQN NQQQQQNNQQ QQQNNQQQQQ NNQQQQQNNQ QQQQNNQQQQQN	EAAAREAAAR EAAAREAAAR	mEos3.1		
S4 C	rhx2609	Q_6N_2					QQQQQQNNQQ QQQQNNQQQQ QQNNQQQQQQ NNQQQQQQNN QQQQQQNNQQ QQQQNNQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
S1 C	rhx3947	Q_3N_3					NNNQQQNNNQ QQNNNQQQNN NQQQNNNQQQ NNNQQQNNNQ QQNNNQQQNN NQQQNNNQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
S1 C	rhx4709	Q_2N_2					QQNNQQNNQQ NNQQNNQQNN QQNNQQNNQQ NNQQNNQQNN QQNNQQNNQQ NNQQNNQQNN	EAAAREAAAR EAAAREAAAR	mEos3.1		
5 A	rhx3458	$(Q_9N)_6(Q_3N)_{15}$					QQQNQQQNQQ QNQQQNQQQN QQQNQQQNQQ QNQQQNQQQN QQQNQQQNQQ QNQQQNQQQN QNQQQNQQQN QNQQQNQQQN QNQQQNC	EAAAREAAAR EAAAREAAAR	mEos3.1		
5 A	rhx4612	$(Q_3N)_{18}(Q_3N)_{15}$					QQQNQQQNQQ QNQQQNQQQN QQQNQQQNQQ QNQQQNQQQN QQQNQQQNQQ QNQQQNQQQN QNQQQNQQQN QNQQQNQQQN QNQQQNC	EAAAREAAAR EAAAREAAAR	mEos3.1		
5 B	rhx4294	Q_3N-oDi					QQQNQQQNQQ NQQQNQQQNQ QNQQQNQQQN QQNQQQNQQQ NQQQNQQQNQ QNQQQNQQQN	EAAAREAAAR EAAAREAAAR	mEos3.1	GGGGSGGGGS GGGGSGGGGS	PEDEIAALKK EIAALKQENA ALKQEIAALK KEIAALKQG
5 B	rhx4633	Q_3N-FTH1					QQQNQQQNQQ NQQQNQQQNQ QNQQQNQQQN QQNQQQNQQQ NQQQNQQQNQ QNQQQNQQQN	EAAAREAAAR EAAAREAAAR	mEos3.1	GGGGSGGGGS GGGGSGGGGS	MTTASTSQVR QNYHQDSEAA INRQINLELY ASYVYVLSMSY YFDRDDVALK NFAKYFLHQS HEEREHAEKL MKLQNQRGGR IFLQDIKKPD CDDWESGLNA MECALHLEKN VNQSLLELHK LATDKNDPHL CDFIETHYLN EQVKAIKELG DHVTNLRKMG APESGLAEYL FDKHTLGDSD NES
5 B	rhx4293	Q_9N-oDi					QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN	EAAAREAAAR EAAAREAAAR	mEos3.1	GGGGSGGGGS GGGGSGGGGS	PEDEIAALKK EIAALKQENA ALKQEIAALK KEIAALKQG
5 B	rhx4631	Q_9N-FTH1					QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN QQQQQQQQQN	EAAAREAAAR EAAAREAAAR	mEos3.1	GGGGSGGGGS GGGGSGGGGS	MTTASTSQVR QNYHQDSEAA INRQINLELY ASYVYVLSMSY YFDRDDVALK NFAKYFLHQS HEEREHAEKL MKLQNQRGGR IFLQDIKKPD CDDWESGLNA MECALHLEKN VNQSLLELHK LATDKNDPHL CDFIETHYLN EQVKAIKELG DHVTNLRKMG APESGLAEYL FDKHTLGDSD NES
5 B	rhx4242	Q60-oDi					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1	GGGGSGGGGS GGGGSGGGGS	PEDEIAALKK EIAALKQENA ALKQEIAALK KEIAALKQG
5 B	rhx4632	Q60-FTH1					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	EAAAREAAAR EAAAREAAAR	mEos3.1	GGGGSGGGGS GGGGSGGGGS	MTTASTSQVR QNYHQDSEAA INRQINLELY ASYVYVLSMSY YFDRDDVALK NFAKYFLHQS HEEREHAEKL MKLQNQRGGR IFLQDIKKPD CDDWESGLNA MECALHLEKN VNQSLLELHK LATDKNDPHL CDFIETHYLN EQVKAIKELG DHVTNLRKMG APESGLAEYL FDKHTLGDSD NES
S5 C	rhx4584	Q60					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	GGGGSGGGGS GGGGSGGGGS	mEos3.1		
S5 C	rhx4345b	oDi-Q60	PEDEIAALKK EIAALKQENA ALKQEIAALK KEIAALKQG	GGGGSGGGGS GGGGSGGGGS	mEos3.1	EAAAREAAAR EAAAREAAAR	QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ				
S5 C,D	rhx4615	Q60-oDi					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	GGGGSGGGGS GGGGSGGGGS	mEos3.1	GGGGSGGGGS GGGGSGGGGS	PEDEIAALKK EIAALKQENA ALKQEIAALK KEIAALKQG
S5 D	rhx4619	Q60-oDi{X}					QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ QQQQQQQQQQ	GGGGSGGGGS GGGGSGGGGS	mEos3.1	GGGGSGGGGS GGGGSGGGGS	PEDEIAALKK EIAALKQENA ALKQEIAALK EIKAAKLQG
5 C	rhx4071	$(Q_6)4$					QQQQQQGGGG QQQQQQGGGG QQQQGGQQGG QQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		
5 C	rhx4134	$(Q_6)4$ "Q10N"					QQQQQQGGGG QQQQQQGGGG QQQQGGQQGG QQQ	EAAAREAAAR EAAAREAAAR	mEos3.1		

Table S2. Summary of results from publically available APR and aggregation propensity predictor webservers. All methods available as of April 2021.

Application Name	Link	Amino Acid Sequence					Reference
		$Q(60)$	$N(60)$	$Q_3N(60)$	$Q_4N(60)$	$Q_5N(60)$	
AGGRESCAN	http://bioinf.uab.es/aggrescan	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Chonchillo-Solï¿½ et al. (2007)
WALTZ	https://walktz.switchlab.org	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Maurer-Stroh et al. (2010)
FoldAmyloid	http://bioinfo.protres.ru/fold-amyloid/	amyloid	amyloid	amyloid	amyloid	amyloid	Garbuzynskiy et al. (2010)
3D Profile (ZipperDB)	http://services.mbi.ucla.edu/zipperdb/	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Thompson et al. (2006)
Amylogram	http://smorfland.uni.wroc.pl/shiny/AmyloGram	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Burdukiewicz et al. (2017)
ANuPP	https://web.iitm.ac.in/bioinfo2/ANuPP/	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Prabakaran et al. (2020)
TANGO	http://tango.crg.es/	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Fernandez-Escamilla et al. (2004)
SecStr	http://biophysics.biol.uoa.gr/SecStr/	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Hamodrakas et al. (2007)
BETASCAN	http://betascan.csail.mit.edu/	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Bryan et al. (2009)
Net-CSSP	http://cssp2.sookmyung.ac.kr/index.html	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Kim et al. (2009)
AmyloidMutants	http://amyloid.csail.mit.edu/	amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	O'Donnell et al. (2011)
PASTA 2	http://old.protein.bio.unipd.it/pasta2/	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Walsh et al. (2014)
GAP	http://www.iitm.ac.in/bioinfo/GAP/	amyloid	amyloid	amyloid	amyloid	amyloid	Thangakani et al. (2014)
MetAmyl	http://metamyl.genouest.org/	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Emily et al. (2013)
Budapest Amyloid Predictor	https://pitgroup.org/bap/	non-amyloid	non-amyloid	non-amyloid	non-amyloid	non-amyloid	Keresztes, et al. (2020)
iAMY-SCM	http://camt.pythonanywhere.com/iAMY-SCM	amyloid	amyloid	amyloid	amyloid	amyloid	Charoenkwan, et al. (2020)