
Figures and figure supplements

Pathologic polyglutamine aggregation begins with a self-poisoning polymer crystal

Tej Kandola and Shriram Venkatesan et al.

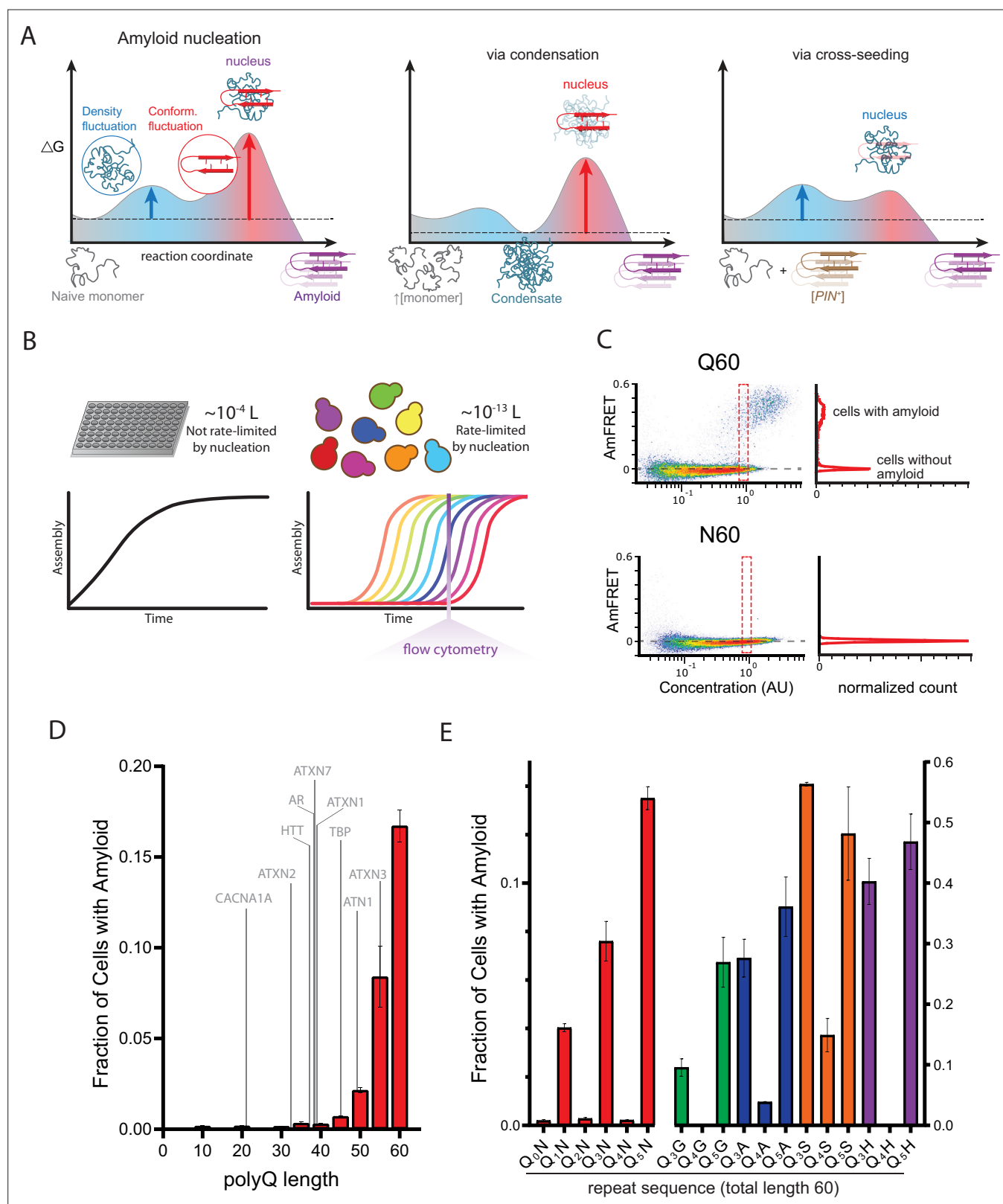


Figure 1. A hidden pattern of glutamines governs amyloid nucleation. **(A)** Reaction coordinate diagrams schematizing the energy barriers governing amyloid nucleation. Because amyloid involves a transition from disordered monomer to ordered multimer, the nucleation barrier results from a combination of high energy fluctuations in both density (blue) and conformation (red). The present study manipulates either component of the nucleation barrier by changing the protein's concentration (middle panel, illustrating a higher concentration surpassing a phase boundary for

Figure 1 continued on next page

Figure 1 continued

condensation) or the presence/absence of a conformational heterogeneity (right panel, with the amyloid template, $[PIN^+]$). Nucleation then depends on the remaining fluctuation, as illustrated by vertical arrows. Cartoon images illustrate the relevant reactant species (naïve monomers, condensate, $[PIN^+]$), products (amyloid), and transition states (nuclei). Note that while the relative heights of the blue and red barriers are arbitrary, we illustrate the latter as higher in keeping with the findings of **Khan et al., 2018** for prion-like amyloid nucleation. **(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 \pm 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 \pm SEM of biological triplicates.

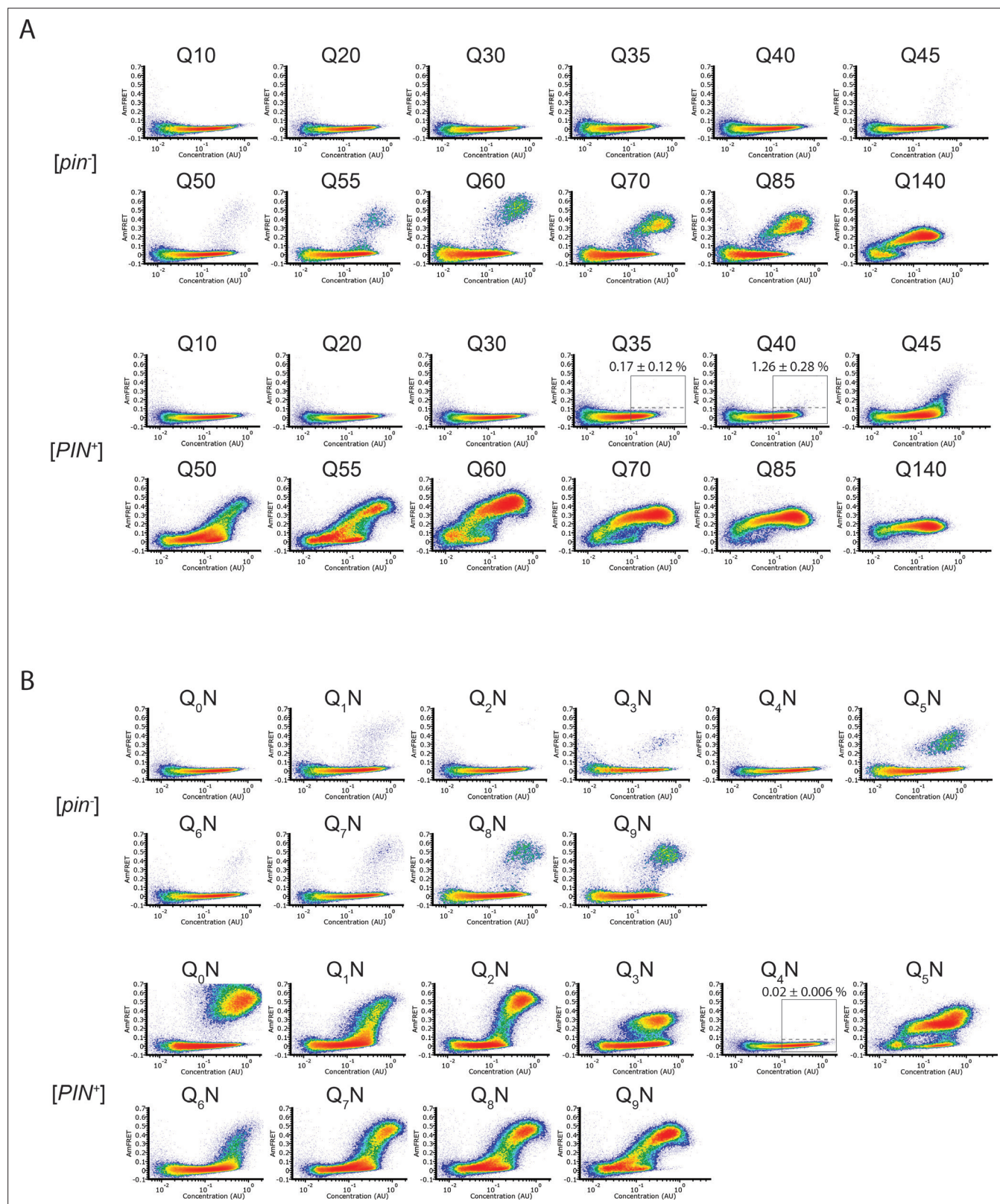


Figure 1—figure supplement 1. DAmFRET plots of polyQ sequence variants. **(A)** DAmFRET plots of polyQ length variants. Labels above the boxed regions of Q35 and Q40 in $[PIN^+]$ indicate the percentage of cells in the high-FRET region, revealing infrequent but significant nucleation for the latter ($P=0.004$, one-tailed T-test). 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.

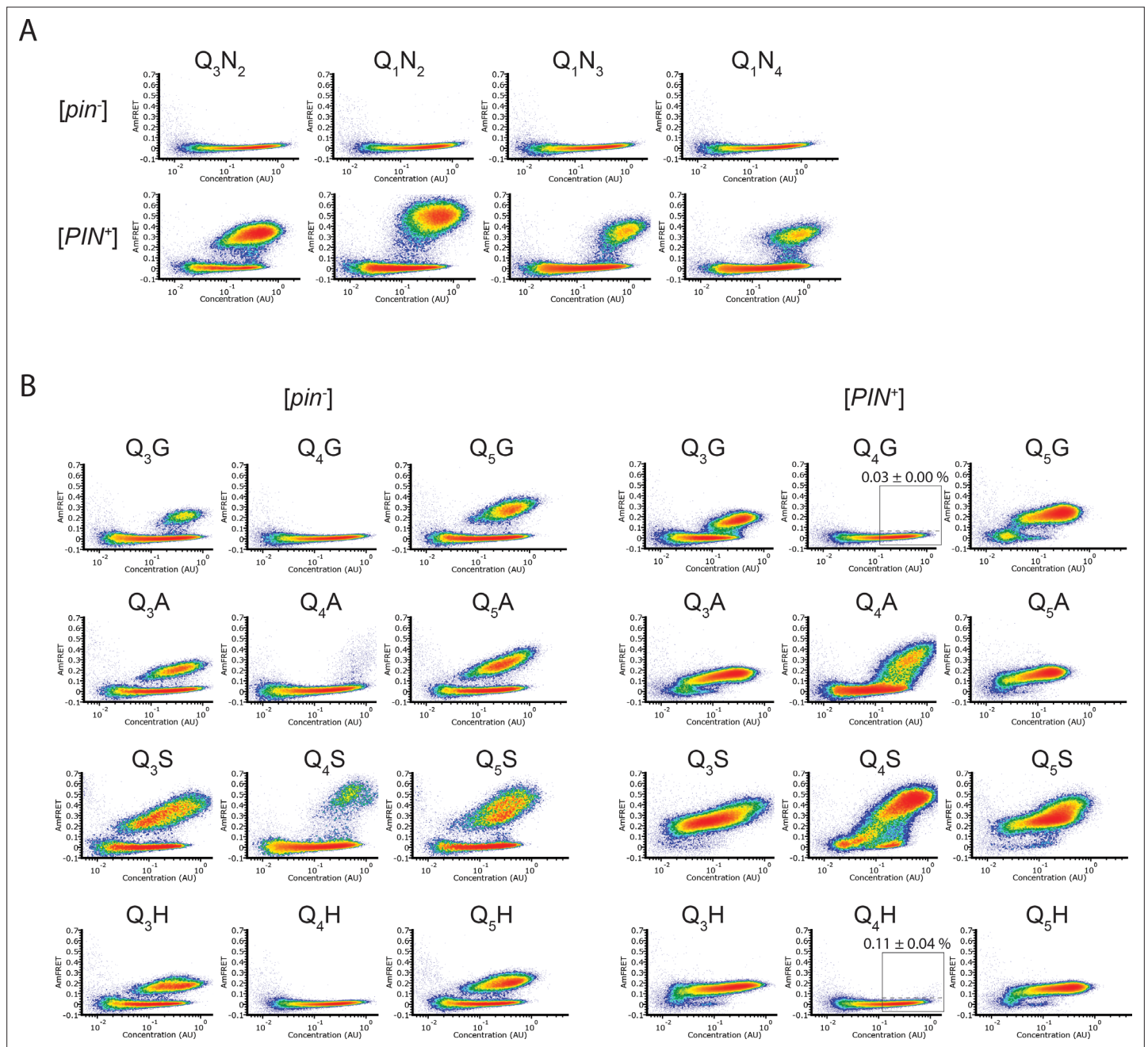


Figure 1—figure supplement 2. DAmFRET plots of N-rich and QnX sequence variants. **(A)** DAmFRET plots of polypeptides composed of tandem repeats of the indicated N-rich sequences, for a total length of 60 residues, showing negligible nucleation in the absence of a conformational template. Note that the nominal pattern repeats, “Q₁N₂”, “Q₁N₃”, and “Q₁N₄” are synonymous to “N₂Q₁”, “N₃Q₁”, and “N₄Q₁”, respectively. Plots are representative of biological triplicates. **(B)** DAmFRET plots of polypeptides composed of tandem repeats of the indicated sequences, for a total length of 60 residues, showing that Q₃X and Q₅X have a greater amyloid propensity than Q₄X regardless of the identity of X. Labels above the boxed regions of the [PIN⁺] Q₄N, Q₄G, and Q₄H plots indicate the percentage of cells in the high-FRET region, revealing rare but significant nucleation for the latter ($p=0.046$ versus Q₄N, one-tailed T-test). Plots are representative of biological triplicates.

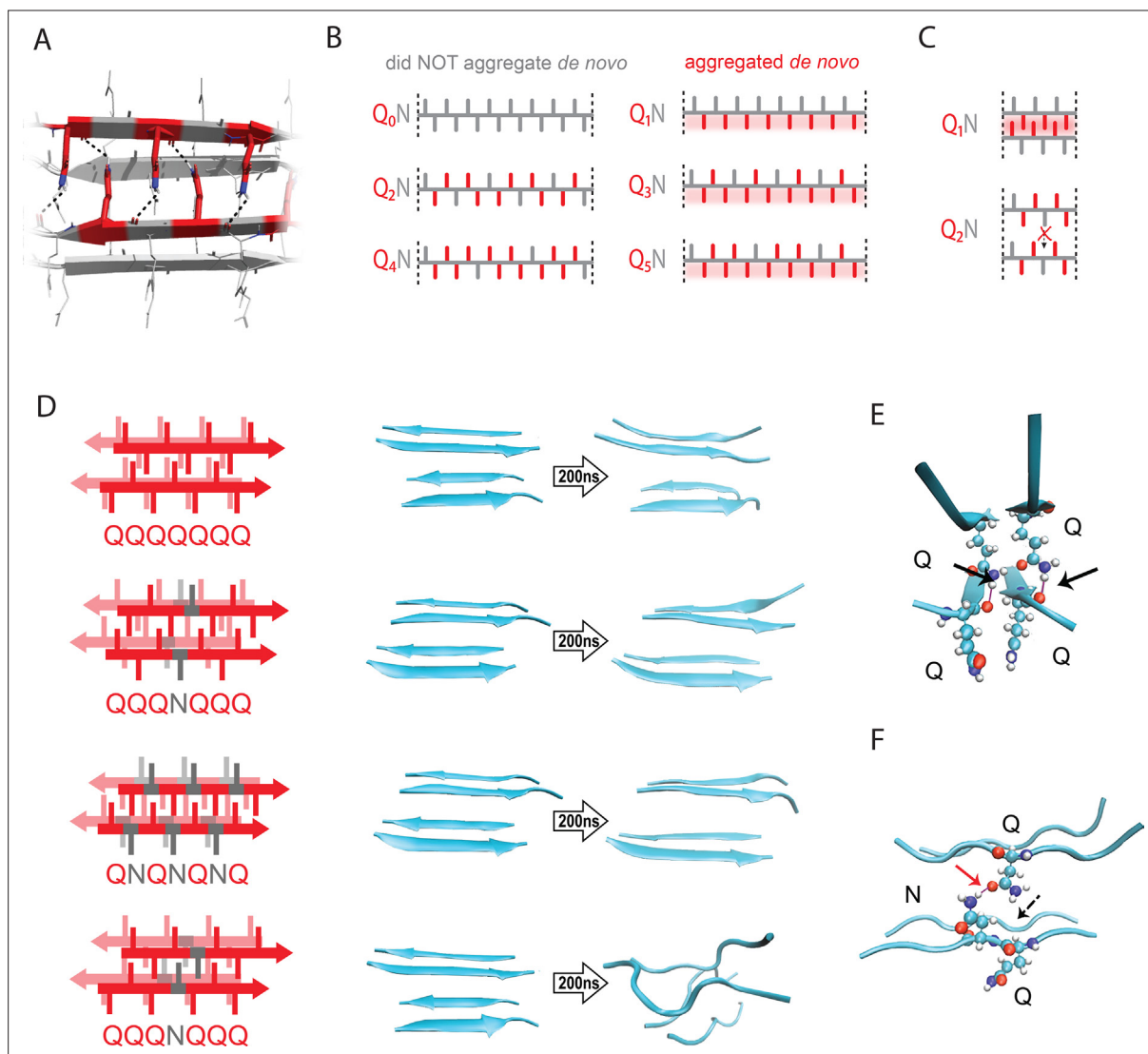


Figure 2. The pattern encodes a single unique steric zipper. **(A)** A view down the axis of a local segment of all-glutamine steric zipper, or “Q zipper”, between two antiparallel two-stranded sheets. Residues with internally facing side-chains on the top layer are colored red to emphasize interdigitation and H-bonding (dashed lines) between the terminal amides and the opposing backbone. **(B)** Schema of the odd-even effect for Q zipper formation, showing side chain arrangements along a continuous β -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 β -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 β -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. Q₁N), but not when it follows an even number of Qs (e.g. Q₂N). **(D)** Molecular simulations of model Q zippers formed by a pair of two-stranded antiparallel β -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.

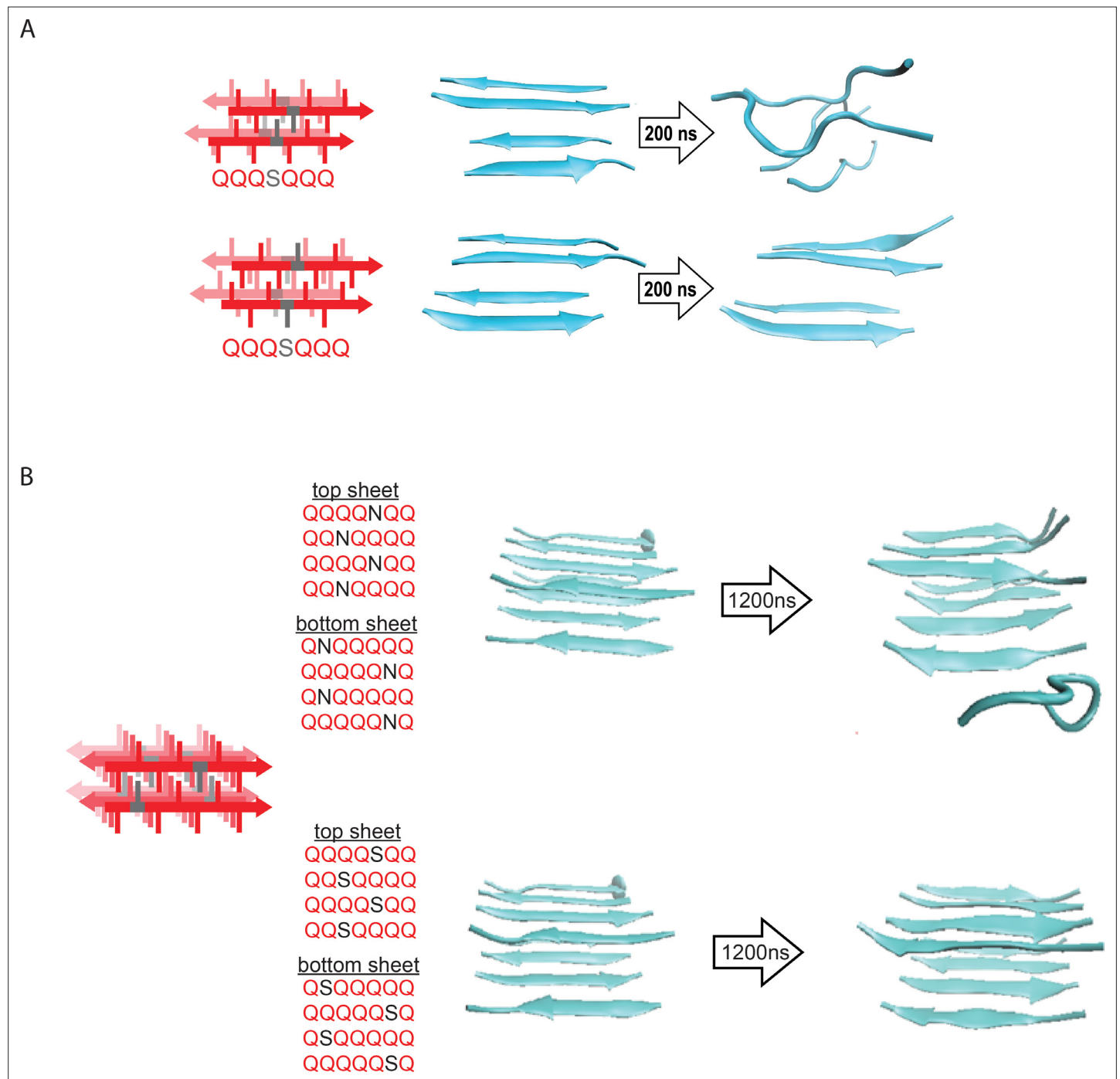


Figure 2—figure supplement 1. Molecular simulations of model steric zippers. **(A)** Molecular simulations of model Q zippers formed by a pair of two-stranded antiparallel β -sheets, containing a single serine residue (QQQSQQQ) per strand. The structure is unstable when the S side chains face inward (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 β -sheets, containing a single asparagine (top) or serine (bottom) residue per strand. The structure proved less stable in the case of asparagine.

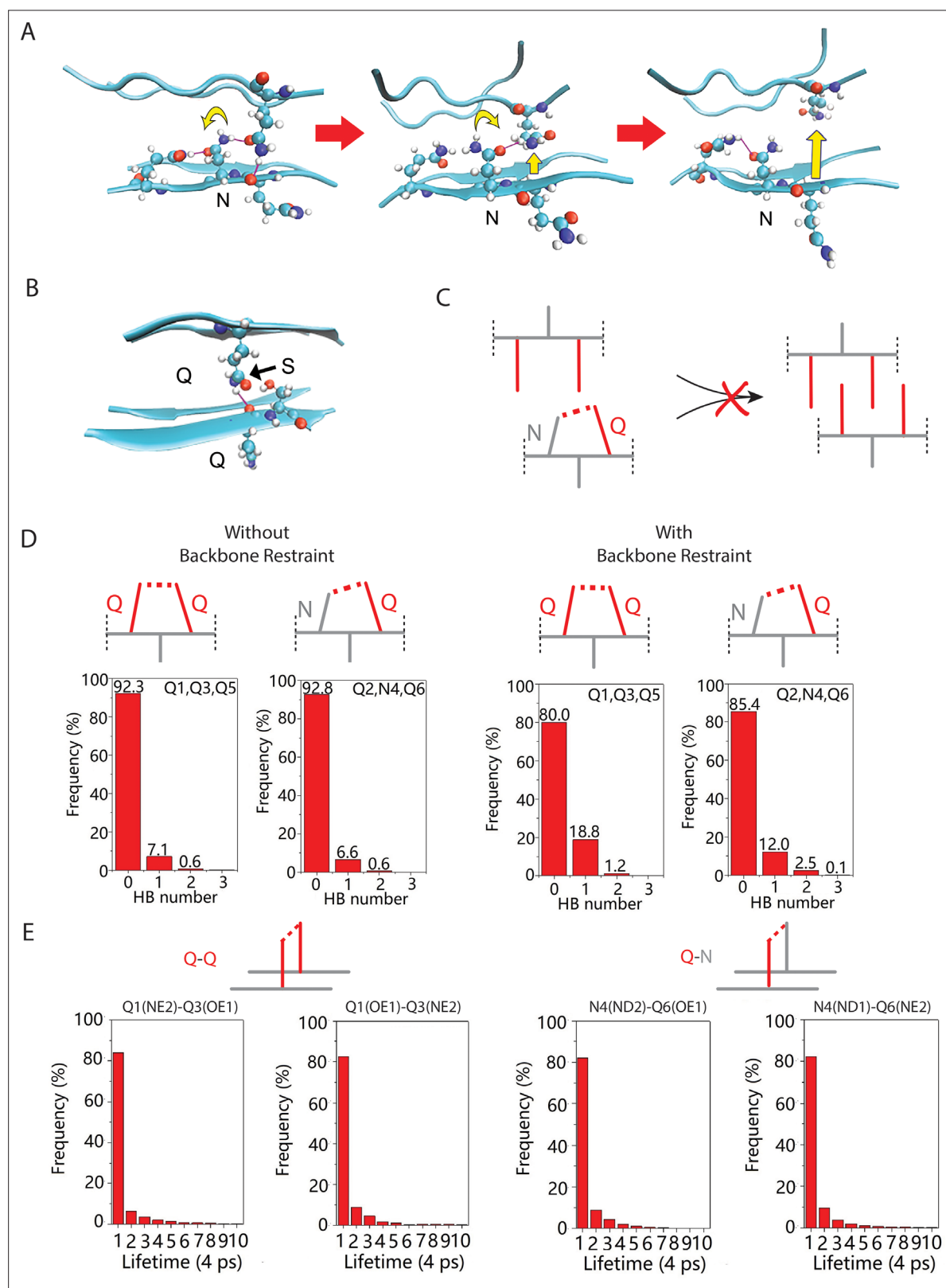


Figure 2—figure supplement 2. Exploration of the mechanism by which N side chains disrupt Q zippers. **(A)** As a consequence of the N side chain's interception of the opposing Q side chain's H-bond, the Q is no longer anchored in the outstretched configuration and sterically interferes with the ordering of adjacent Qs. This effect propagates through the zipper, resulting in its dissolution. **(B)** 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

Figure 2—figure supplement 2 continued on next page

Figure 2—figure supplement 2 continued

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. **(C)** Schematic demonstrating how a polar clasp (red dashed line) would preclude Q zipper formation. **(D)** Schema and frequencies of polar clasps occurring between two unilaterally adjacent Q side chains (left) or a unilaterally adjacent N and Q side chain (right) within a QQNQQQ peptide, simulated either with (top) or without (bottom) the backbone restrained in a β conformation. The bar graphs show that polar clasps between Q and N occur less frequently than between Q and Q, indicating that the mechanism of Q zipper destabilization by N side chains cannot be attributed to polar clasps. **(E)** Schema and lifetimes of H-bonds between exterior stacked (axially adjacent) Q side chains (top) or N and Q side chains (bottom) in the Q zipper simulated in **Figure 2D**, showing no difference in stabilities between axial H-bonds between Q and Q/N side chains.

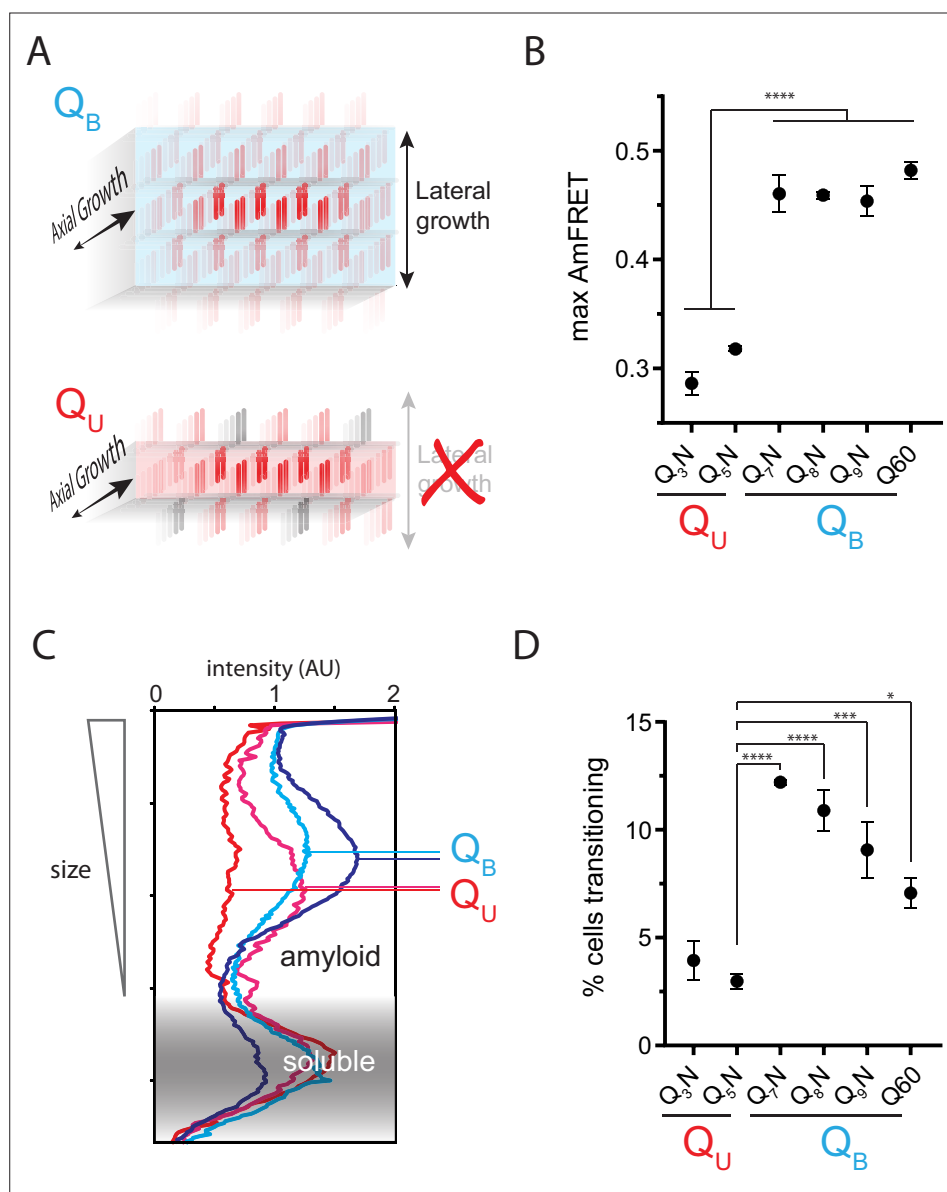


Figure 3. The Q zipper grows in two dimensions. **(A)** Schematic illustrating how sequences with bilaterally contiguous Qs (Q_B) can hypothetically allow for lateral growth (secondary nucleation) of Q zippers giving rise to lamellar amyloid fibers. In contrast, sequences with only unilateral contiguous 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 \pm 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 \pm SEM of the percentage of cells between lower and upper populations, of triplicates. ****, ***, * $p < 0.0001, < 0.001, < 0.05$; ANOVA.

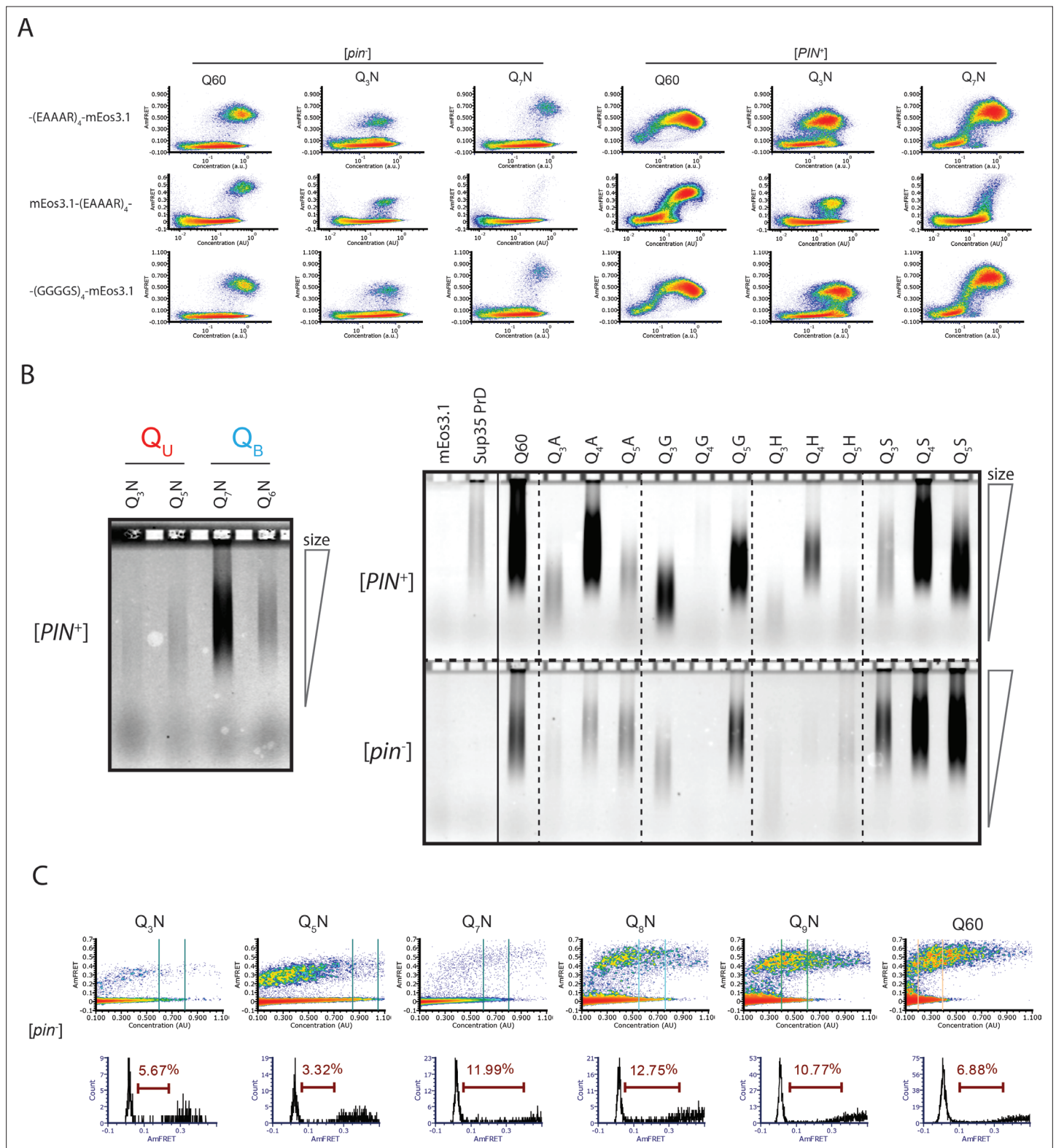


Figure 3—figure supplement 1. Experiments supporting the existence of unilateral and bilateral Q zipper amyloids. **(A)** DAmFRET plots of the indicated sequence variants with either a C-terminal (EAAAR)₄-mEos3.1 fusion, as used throughout this work, an N-terminal mEos3.1-(EAAAR)₄ fusion, or a C-terminal (GGGS)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 **Figure 3C**. Right: Additional Q₃X, Q₄X, Q₅X proteins, **Figure 3—figure supplement 1 continued on next page**

Figure 3—figure supplement 1 continued

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.

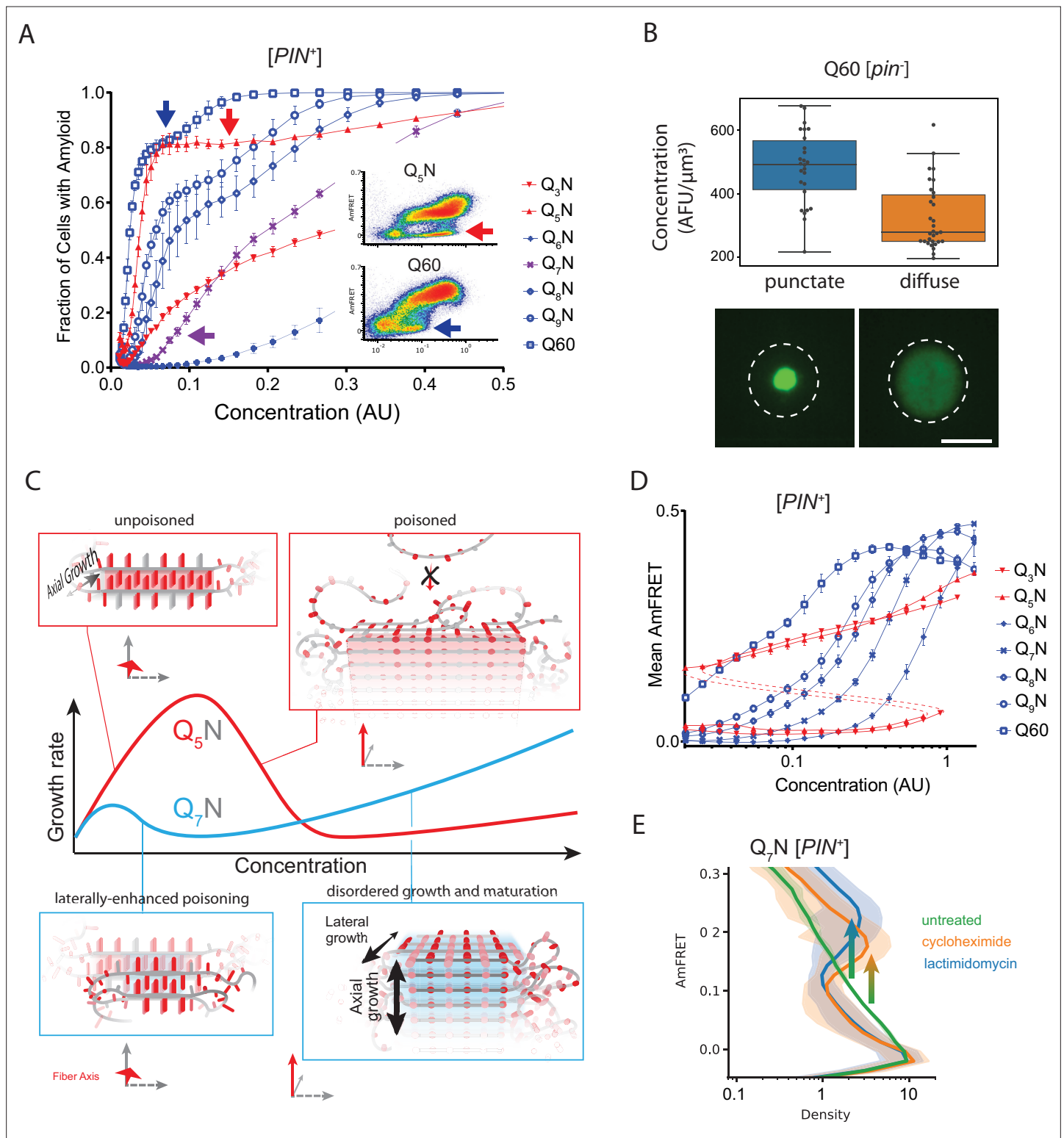


Figure 4. Q zippers poison themselves. **(A)** Fraction of [PIN⁺] cells in the AmFRET-positive population as a function of concentration for the indicated sequences. Arrows denote the population of cells with self-poisoned aggregation (inset), and the corresponding plateaus in the relationship of amyloid formation to concentration. The purple arrow highlights the sharp reduction in aggregation for Q₇N relative to Q₅N, which we attribute to enhanced poisoning as a result of intramolecular Q zipper formation. Shown are means \pm SEM of triplicates. **(B)** Distribution of cytosolic concentrations (AFU/μm³) 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

Figure 4 continued on next page

Figure 4 continued

illustrating self-poisoned growth as a function of concentration for Q zippers of Q₅N and Q₇N. Conformational conversion of Q₅N 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 Q₇N 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 with the addition of successive zippers. **(D)** Graph of spline fits of AmFRET values for the indicated sequences in [PIN⁺] cells. The upper and lower populations of Q₃N and Q₅N 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 Q₆N and then gradually increased with higher q values. Shown are means +/- SEM of triplicates. **(E)** Histogram of AmFRET values for Q₇N -expressing cells transitioning from the low to high populations (boxed region from DAmFRET plot in **Figure 4—figure supplement 1D**) 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).

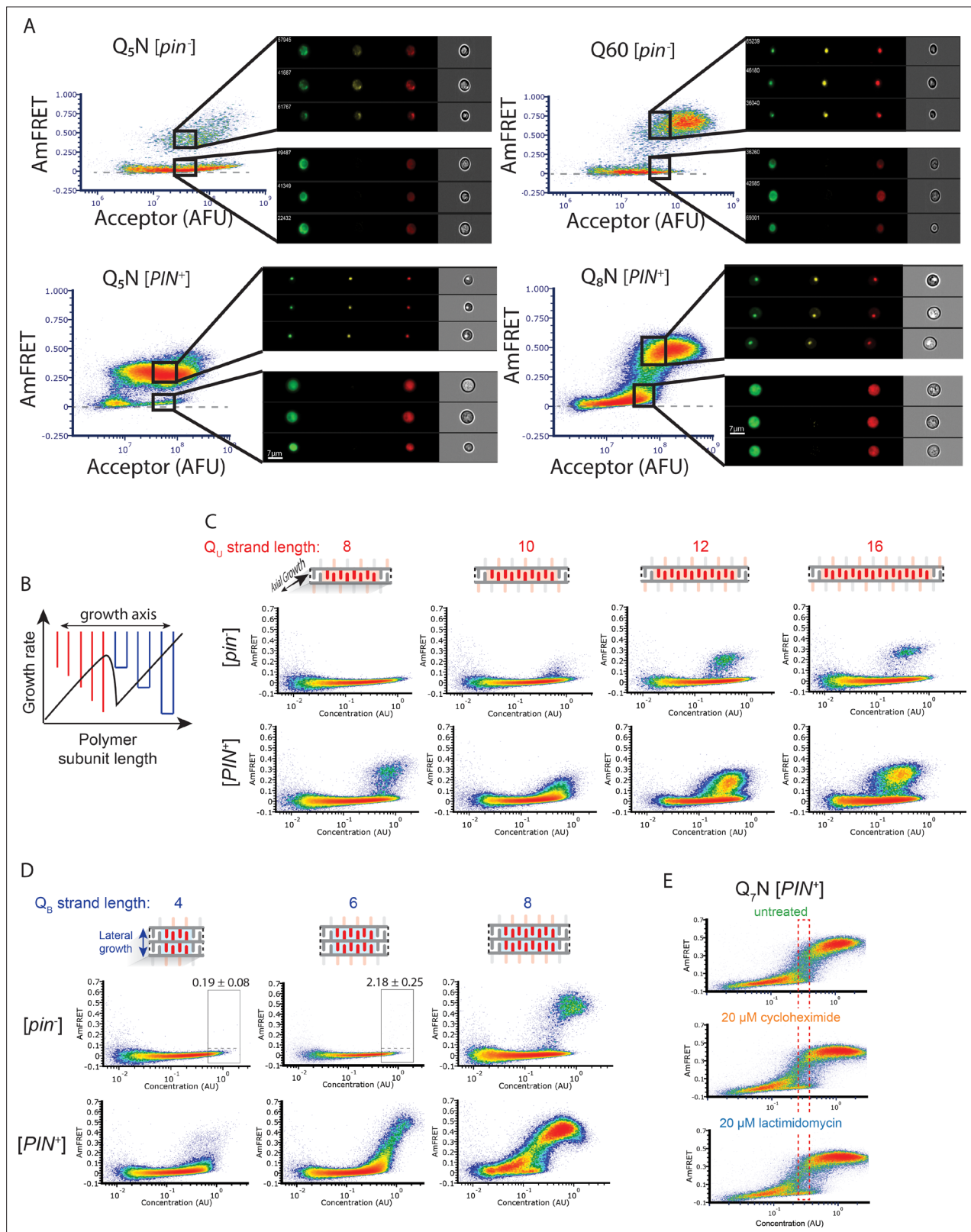


Figure 4—figure supplement 1. Experiments supporting the self-poisoning mechanism. **(A)** DAmFRET plots of Q₅N, Q₆₀, and Q₈N acquired using imaging flow cytometry, showing gates at high expression for both high- or low-AmFRET populations. Insets show from left to right the distribution of donor, FRET, 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, colored red or blue for Figure 4—figure supplement 1 continued on next page

Figure 4—figure supplement 1 continued

single or double stranded subunits, respectively. **(C)** DAmFRET plots of [*pin*⁺] cells expressing unilateral contiguity variants of the Q₃N 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 (Q₄N₂, Q₆N₂, Q₈N₂), 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 Q₆N₂ (p=0.0004, T-test of triplicates). **(E)** DAmFRET plots of Q₇N in [*pin*⁺] cells treated as indicated for 6 hr prior to analysis. The boxed region was used to compute histograms of AmFRET in **Figure 4E**.

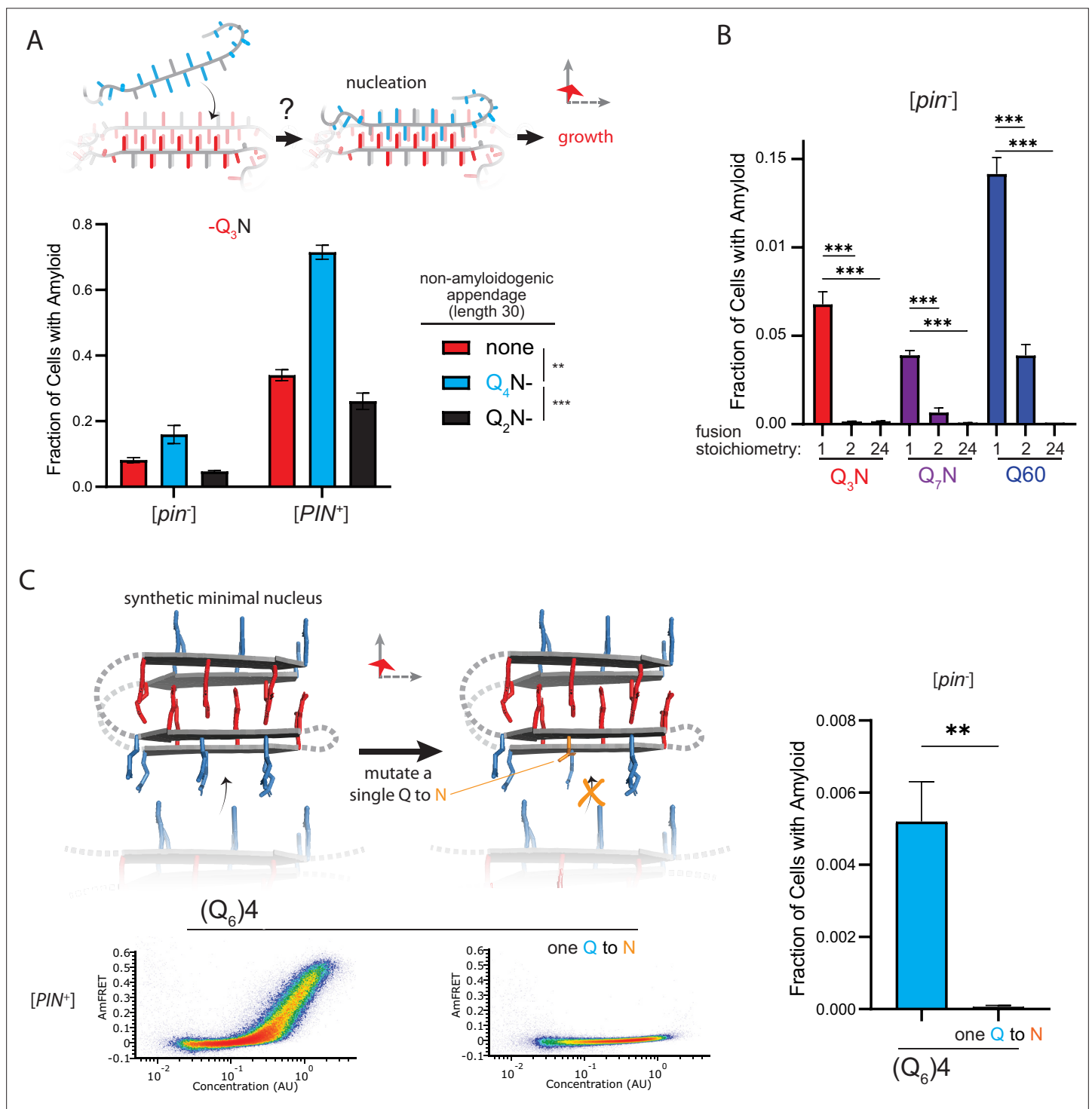


Figure 5. The nucleus forms within a single molecule. **(A)** Fraction of cells in the AmFRET-positive population when expressing Q₃N with the indicated non-amyloidogenic fusions. Shown are means \pm SEM of triplicates. ***, ** $p < 0.001, < 0.01$; ANOVA. **(B)** Fraction of cells in the AmFRET-positive 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 \pm 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 \pm SEM of triplicates. ** $p < 0.01$; t-test.

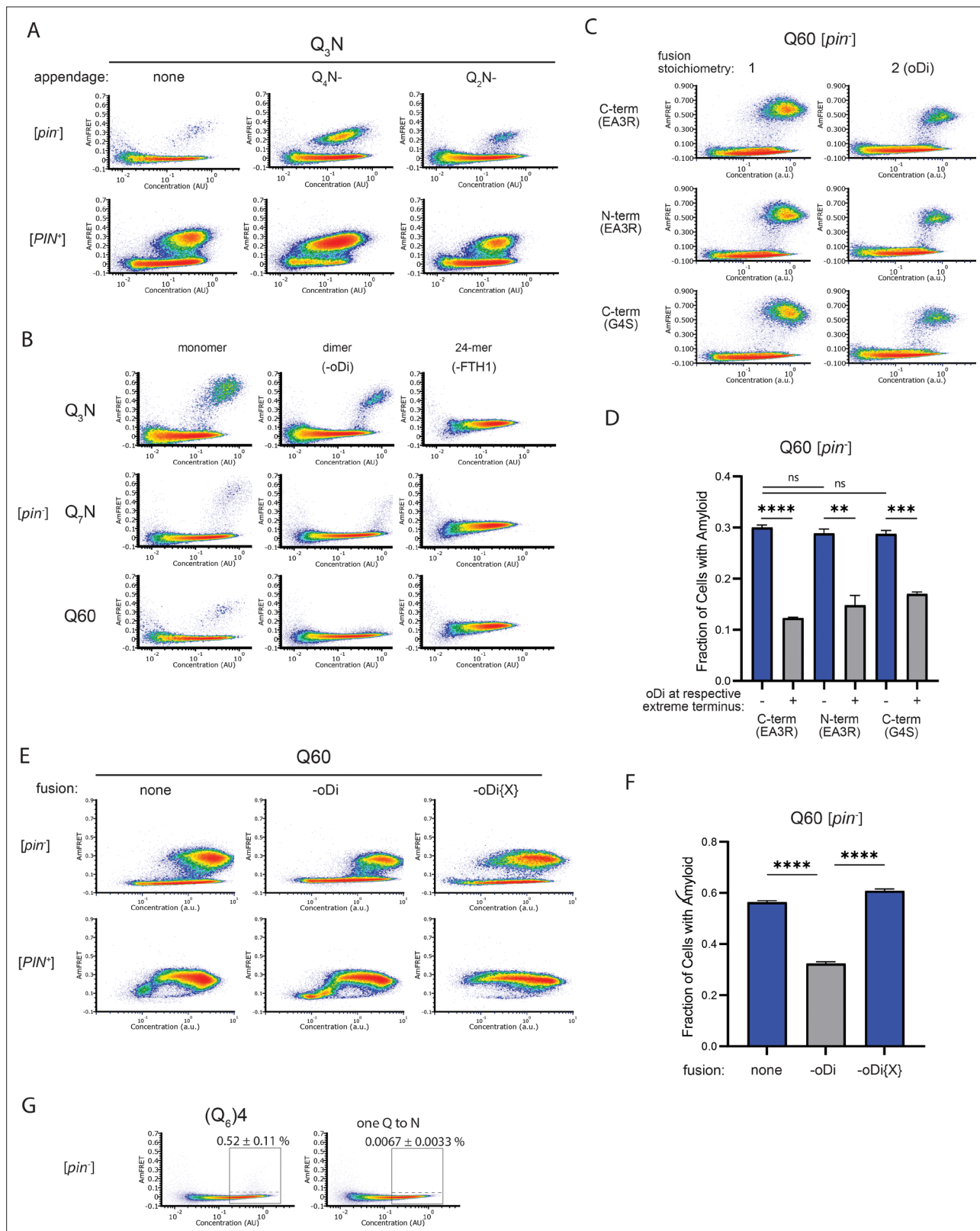


Figure 5—figure supplement 1. Experiments supporting monomeric nucleation. **(A)** DAmFRET plots of cells expressing Q_3N (length 60) with the indicated appendage (length 30). Plots are representative of biological triplicates. **(B)** DAmFRET plots of $[pin^-]$ and $[PIN^+]$ cells expressing the indicated 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)** Bar graph showing the fraction of cells with amyloid for $Q60$ [pin⁻] under different conditions. The y-axis is 'Fraction of Cells with Amyloid' (0.0 to 0.4). The x-axis shows oDi at respective extreme terminus: C-term (EA3R), N-term (EA3R), and C-term (G4S). Conditions are - (no oDi) and + (oDi). Statistical significance is indicated: **** for C-term (EA3R) and C-term (G4S), ** for N-term (EA3R), and ns for C-term (EA3R) and N-term (EA3R) comparisons. **(E)** DAmFRET plots of cells expressing $Q60$ with indicated fusion: none, -oDi, or -oDi{X}. The plots show $\Delta FRET$ vs. Concentration (a.u.) for $[pin^-]$ and $[PIN^+]$ conditions. **(F)** Bar graph showing the fraction of cells with amyloid for $Q60$ [pin⁻] under different fusion conditions. The y-axis is 'Fraction of Cells with Amyloid' (0.0 to 0.8). The x-axis shows fusion: none, -oDi, and -oDi{X}. Statistical significance is indicated: **** for none vs -oDi and -oDi vs -oDi{X} comparisons. **(G)** DAmFRET plots of cells expressing $(Q_6)_4$ and one Q to N. The plots show $\Delta FRET$ vs. Concentration (AU) for $[pin^-]$ condition. The values are $0.52 \pm 0.11\%$ for $(Q_6)_4$ and $0.0067 \pm 0.0033\%$ for one Q to N.

Figure 5—figure supplement 1 continued

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 \pm 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 \pm 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 **Figure 5C**. Plots are representative of biological triplicates.

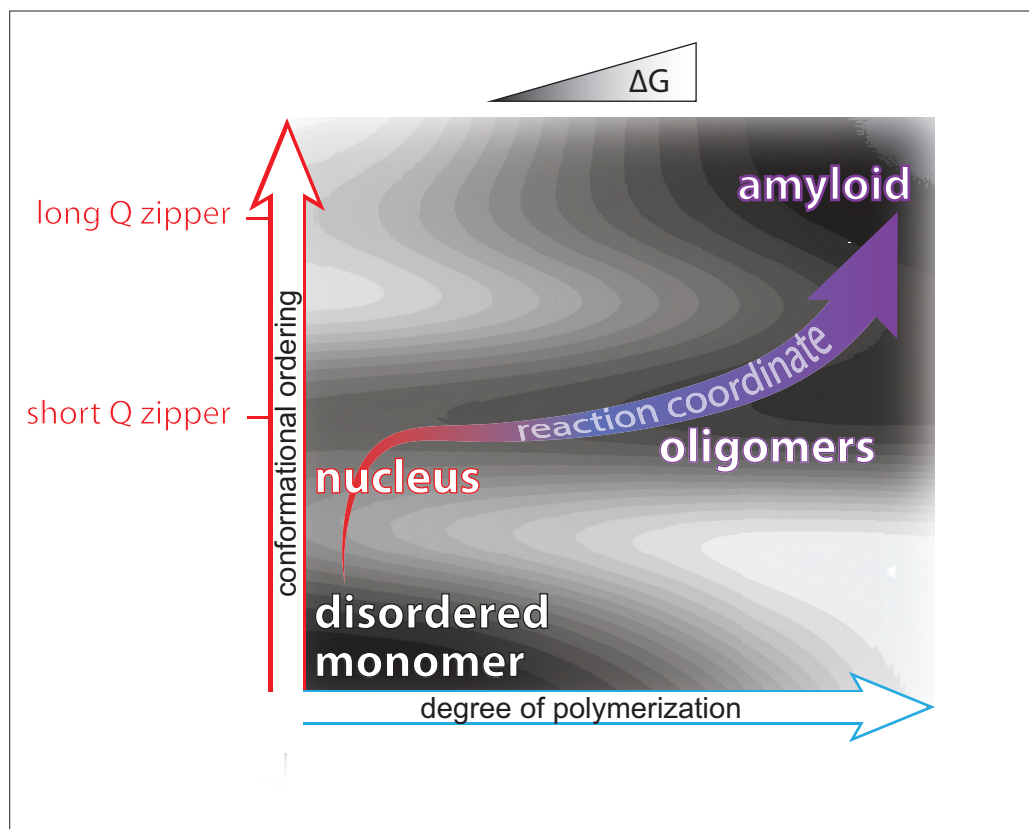


Figure 6. Summary of aggregation mechanism. Schematic of the free energy landscape for amyloid formation by pathologically expanded polyQ at approximately physiological concentrations, showing the reaction pathway as a function of the conformational ordering and degree of polymerization of the species. Qualitative topological features of the landscape, but not absolute heights and positions, are as deduced herein. Naive monomers exist in a local energy minimum at maximum disorder, while mature amyloid exists in a global energy minimum with long lamellar Q zippers. The middle and upper horizontal basins represent Q zippers with short (~6 residue) and long (~11 residue) strands, respectively. Nucleation occurs with the formation of a short intramolecular Q zipper, which then oligomerizes via axial and lateral recruitment of other polypeptides. The Q zipper eventually lengthens, allowing for the growth of mature amyloid.