

SUPPORTING INFORMATION for**Combining Classical MD and QM Calculations to Elucidate
Complex System Nucleation: A Twisted, Three Stranded, Parallel
 β -sheet Seeds Amyloid Fibril Conception**

by Miguel Mompeán^{†,*}, Carlos González[†], Enrique Lomba[†] and Douglas V. Laurents^{†,*}

CONTENTS:**-Methods**

- Nuclear Magnetic Resonance Spectroscopy
- Bimolecular Association Kinetics Calculation
- Molecular Dynamics Simulations
- ONIOM Calculations
- Charged vs. Capped Behavior: two distinct pathways towards the seed formation

-Supporting Figures

- **Figure S1:** RMSD for the **S1·S1** system -5 additional simulations-
- **Figure S2:** RMSD for the **S2** system -5 additional simulations-
- **Figure S3:** RMSD for the **S3** system -5 additional simulations-
- **Figure S4:** RMSD for the **S2·S2** system -5 additional simulations-
- **Figure S5:** RMSD for the **S2·S1** system -3simulations-
- **Figure S6:** RMSF for the **S3** system
- **Figure S7:** RMSD for the **NQYGQNQ** mutant
- **Figure S8:** NMR evidence of the key role of the central Glutamine in aggregation.
- **Figure S9:** RMSD for the **+S1·S1-** system
- **Figure S10:** RMSD for the **S3·S2** system -2 simulations-
- **Figure S11:** RMSD for the **S3** trimers
- **Figure S12:** RMSD for the **S3** system fitted to the averaged structure
- **Figure S13:** RMSD for the **S3·S3** system

-Supporting Tables

- **Table S1:** Main findings in previous manuscripts on GNNQQNY oligomerization
- **Table S2:** Summary of the ONIOM (DFT:PM6) RESULTS
- **Table S3:** Summary of the simulation results
- **Table S4:** H-bonds vs. Electrostatic Repulsion
- **Table S5:** Coordinates of the **S3** & **S3·S3** systems used as Gaussian09 input files for ONIOM (DFT:PM6) interaction energies and frequency calculations.

-Supporting references

METHODS

NMR Spectroscopic Characterization of GNNQQNY and GQNNQNY peptides.

To characterize the initial conformation of GNNQQNY and to corroborate the prediction based on our MD results that the rigidity and interactions of the central residues play key roles in nascent GNNQQNY oligomerization, we studied the conformation of this heptapeptide and a negative control peptide, in which the position of Gly1 and Gln4 are switched, by NMR spectroscopy. By placing Gly, a flexible residue whose side chain is unable to replace the interactions formed by Gln's side chain, we expect that oligomer formation will be blocked. To prevent the formation of pyroglutamate at the N-terminus, we added an additional Gly at the N-terminus. Thus the sequence of this negative control peptide is Gly0-**Gln1**-Asn2-Asn3-**Gly4**-Gln5-Asn6-Tyr7, with the switched residues written in **bold**. This peptide and the Sup35 wild type heptapeptide: Gly1-Asn2-Asn3-Gln4-Gln5-Asn6-Tyr7, were obtained from Genescript Corp. They are over 95% pure (HPLC) and their identities were confirmed by MALDI-TOF mass spectrometry and NMR spectroscopy.

NMR spectra were acquired on a Bruker 600 & 800 MHz (¹H) spectrometers equipped with a triple resonance cryoprobes and Z-gradients at 25.0 °C, in 90% H₂O/ 10% D₂O with 50 μM DSS as the internal chemical shift standard, 10 mM ZnCl₂ and 10 mM deuterated sodium acetate/acetic acid (pH 5.1). These solution conditions are similar to those used for crystallization by Eisenberg and coworkers¹. The peptide concentration was 3.44 mM for the wild type peptide and slightly higher, 3.67 mM, for the negative control peptide. We acquired the following spectra: 1D ¹H (32 transients) and 2D ¹H TOCSY (mixing time 60 ms, 8 transients per increment, 512 increments in the 2nd dimension), 2D ¹H ROESY (spinlock time 300 ms, 64 transients per increment, 512 increments in the 2nd dimension), ¹H-¹³C HSQC (256 transients per increment, 128 increments in the ¹³C dimension) and ¹H-¹⁵N HSQC (512 – 1024 transients per increment, 128 increments in the ¹⁵N dimension).

Bimolecular Association Kinetics Calculation

An estimate for the bimolecular association rate constant can be derived from Fick's first law of diffusion:

$$k = 4 \pi r_o (D_s + D_{s2}) \cdot 10^{-3} N_o$$

where r_o is the distance between the centers during the encounter (assumed to be a 1 nm), D_A is the diffusion constant (about $32.5 \cdot 10^{-7}$ for a 840 Da peptide and about $25.2 \cdot 10^{-7}$ for the 1680 Da dimer at 20 °C in H₂O) and N_o is Avogadro's number (Cantor & Schimmel *Biophysical Chemistry* (1980) WH Freeman & Co. NY page 921) This gives a value of $4.37 \cdot 10^9 \text{ M}^{-1} \cdot \text{s}^{-1}$ for the bimolecular association constant. From this value, it can be calculated that there would be 284 contacts per 65 ns at 1 M concentration, or 1 contact per 65 ns at 3.5 mM of heptapeptide. We think that this estimate is a ballpark figure; all encounters are unlikely to be productive, so a higher concentration of peptide might be needed to drive oligomerization. On the other hand, four of the six simulations of the S2 dimer did not dissociate even after 100 ns. Therefore, it is likely that the monomer would, on average, have longer than 65 ns to encounter S2, and therefore, a lower monomer concentration would be required to further oligomerization.

Molecular dynamics simulations

The models for the peptide structures are based on the X-ray crystal experimental data¹ (PDB code 1YJP), and were generated with PyMOL. Since previous work has shown the importance of electrostatics in the aggregation of Sup35² we have capped the termini residues of each peptide in order to mimic the behaviour of the full peptide sequence. GNNQQNY is just a short heptapeptide segment in a loop of a large 680-residue protein. Thus, in order to represent more accurately the behavior of this aggregating segment, the peptide was amidated at the C-terminus and acetylated in the N-terminus. However, as the peptide with charged termini has been studied extensively by crystallography and computational methods, we have also performed additional simulations with the terminal charges present.

For all the MD simulations, we employed the GROMACS package, version 4.5.5.³ The structures were placed in a cubic box leaving at least 1.0 nm between any atom of the solvent molecules and the box edges. Parameters from the amber99sb-ildn force field⁴ were applied to all the simulated systems, solvated with TIP3P⁵ water molecules. The choice of this force field is based on its appropriateness for amyloid-like systems⁶. Short-range nonbonded interactions were cut off at 1 nm, with long-range electrostatics calculated using the particle mesh Ewald (PME) algorithm⁷ and application of dispersion-correction to account for van der Waals interactions at distances longer than the cut off. Periodic boundary conditions were applied in all directions.

After steepest descents minimization, two equilibration periods were done using the Berendsen⁸ coupling algorithm. Once the temperature and pressure were properly set to 300 K and 1 bar (500 ps and 2 ns runs, respectively) with heavy atom positions restrained with the LINCS⁹ algorithm, the MD simulations were produced with the Nosé-Hoover¹⁰ thermostat and the Parrinello-Rahman¹¹ barostat in absence of restraints, with time constants of 0.5 and 1.0 ps, respectively. A time step of 2 fs was used in both the equilibration stages and the MD run. **Distinct initial velocities were utilized to ensure the independence of the different MD runs starting on the same structure. This is a standard procedure to allow the system, which starts from the same structure, to evolve through different trajectories thereby resulting in different simulations. This procedure is done at the start of a simulation; just after defining the topology of the system is defined and following the energy minimization step, one uses a random seed to generate random initial velocities that will allow the system to reach a Maxwell distribution.** The root mean square displacement was calculated for all of the simulated systems to assess the kinetic stability of all the structures.

To further corroborate these results, we performed a series of additional simulations. Even though the 500 ns simulations are in principle long enough to consider proper sampling, it has been recently shown that in some cases performing multiple simulations for shorter periods of time is a better gauge of the reproducibility¹². By using both approaches, we can ensure that a thorough sampling was done.

These additional runs corroborate that **S2** is clearly more stable than **S1-S1**. The latter exhibits a dual

behavior: it either dissociates or rearranges into an anti-parallel β -sheet, but none of these simulations lasts longer than 10 ns as a “layer”, which is the proposed biological assembly¹. In the layer-like structure (*i.e.*, **S1·S1**) the stabilizing interactions are due to side chain intermeshing and not due to the formation H-bonds as in the S2-like antiparallel structure. The rapid dissociation or alternation of the **S1·S1** structure indicates that these side chain interactions are relatively weak. On the basis of its ephemeral stability, we conclude that the possibility of the **S1·S1** intermediate persisting long enough so that other monomers can join it is negligibly small, when it is uncharged. See below for a description of the behavior when the terminal charges are present (**Sup. Fig. 1 & 2**).

For **S3**, since the RMSD reveals a remarkable stability in multiple 100 ns runs, all of which showed similar behavior, as well as a long 500 ns simulation, we conclude that this structural motif lasts long enough to be the most important intermediate in the formation of the seed for aggregation (**Sup. Fig. 3**).

S2·S2 showed significant structural changes during the course of the 500 ns run; one monomer dissociates and the other three evolve towards an **S3**-like configuration. The results reported in Supplementary Figure 4 show that in only one out of five 100 ns simulations did the motif remain stable without any significant structural rearrangement.

To address whether the dry interface could be formed in some intermediate structures not considered in the above cases, two additional oligomers, named **S2·S1** and **S3·S2** were tested. The first consists of a **S2** dimer with a monomer docked laterally through sidechain packing, whereas the last is a **S3** plus **S2** interacting system through their intermeshed sidechains. Both the two system consider the dry interface formation.

Supporting Figure 5 describes the behavior of **S2·S1** through the RMSD, considering the situations in which the peptides are either charged or with the terminal residues neutralized. Similarly, Sup. Figure 6 explains the results observed of the analogous **S3·S2** system. For the **S2·S1** system, a dry interface cannot be formed, whereas in the case of an **S3·S2** arrangement some effective sidechain intermeshing was found to take place (Sup. Fig. 6). However, its structural stability is considerably lower than that of the **S3·S3** system, hence we conclude that the dry interface is most likely to definitely form when the mating sheets have a minimal length of three strands (**S3**).

We conclude that the **S3** trimer is to be the key intermediate leading to the **S3·S3** nucleus for the following reasons: 1) The dry interface is not prone to form with two or four monomers, **S1·S1** or **S2·S2**, respectively), 2) The **S3** trimer is stable in the long 500 ns simulation and multiple 100 ns simulations, and 3) Two S3 trimers were observed to combine to form **S3·S3** (**Fig. 4 & Sup. Fig. 11**). We consider **S3·S3** to be the “nucleus”, since it is the smallest structure showing the amyloid-like properties derived from experimental data in a solvated context, that is, similar H-bonding pattern, amide stacking and sidechain intermeshing (*i.e.*, dry interface formation) as well as a high stability as evidenced by its lack of dissociation or structural alteration after 100 ns of simulation (**Sup. Fig. 12**).

ONIOM calculations

The ONIOM¹³ methodology permits the division of the system. Here, we made a two-region partition to use two different levels of theory, resulting in the so-called low and high layers. The former was modelled with the PM6¹⁴ semiempirical method, whereas we employed DFT for the latter. For these QM calculations, we optimized the geometry with the B3LYP functional and the 6-311+G(d,p) basis set, accounting for dispersion and polarization effects. This functional was chosen for the geometry optimization procedure since it is well known to yield very good results when applied to organic systems. Recently, a very similar approach to dissect L-asparaginase II's enzymatic mechanism has been described.¹⁵

The high layer consists of all side chain amide groups, since they are involved in the H-bonding and experience hyperpolarization upon monomer addition¹⁶. Thus, a monomer contains 5 amide residues times 5 amide atoms = 25 QM atoms. Accordingly, **S3** contains 75 QM atoms and **S3-S3**, 150.

Since PM6 can deal with β -sheet backbone-backbone hydrogen bonds, the low layer includes the rest of the peptide, *i.e.*, capping atoms CH₃CO- and -NH₂, first and last residues of each strand (Tyr and Gly), and amide backbone atoms.

With respect to the basis set, even though the inclusion of diffuse functions has been reported to substantially raise the computational cost without notably increasing the optimized geometries,¹⁷ we decided to perform our calculations with these functions on heavy atoms. To this end, we performed our calculations in a multi-step way to ensure a smooth convergence. First, we froze the C α atoms to keep the backbone at its experimental position and used the 6-31+G(d) basis set. After that, we employed that optimized structure to be re-optimized with the larger basis set 6-311+G(d), adding 18 extra basis functions to the former. Finally, we performed a 6-311+G(d,p) calculation without any restraints, which adds polarization functions to hydrogen atoms with respect to the previous steps. By this procedure, we obtained the fully relaxed geometry with the appropriate description of the H-bonds between amide groups.

Having seen that the **S3** trimer has a remarkable stability upon twisting (**S3**), one may ask what is the energetic bonus afforded by this structural rearrangement. The hydrogen bonds between the strands within a sheet are stronger than those for water in ice, as determined by Baker¹⁶ *et al.* However, in that manuscript it was advanced that the physiological twisted fibrils may possess higher interaction energy, since those calculations were performed in gas phase and with boundary conditions to account for an infinite system. Moreover, no relaxation/optimization was described, presumably due to the high computational cost. To properly quantify the effect of the twist, we followed a subtractive scheme similar to that of Martín-Pintado *et al.*¹⁸ We first defined the **S3** trimer as State A, which reproduces the crystallographic experimental data. State B is the twisted structure resulting from averaging over the 500 ns trajectory, **S3**. The all-atom contribution to the intra- and intermolecular interaction energies within the strands can be obtained from the fully optimized straight and twisted trimers, named States A & B,

by sequential removal of the strands (**Figure 2**, main text). The overall energy difference between **S3**, **S2**, **S1 & S3**, **S2**, **S1** with respect to the trimers allows for a direct estimation of the energetic contribution of the twist to the global stability (*i.e.*, the gain in interaction energy due to the different hydrogen bonding pattern.)

The choice of the averaged structure as representative of the twisted conformer is based on the low RMSD obtained when fitting the frames of the trajectory to this system, as depicted in **Sup Fig.12**.

The kinetic MD-based study as well as the ONIOM calculation showed that the trimer **S3** is extremely stable and we propose that this is the key kernel structure for amyloidogenesis. The dry interface can only be effectively formed when two of these units mutually interact, since the energy contribution of their sidechain-sidechain contacts is too low when less than three strands per sheet are present. To quantify this value, the interaction energy of two **S3** trimers combining to yield the nucleus **S3-S3** is a direct measure of the energy required for the dry interface formation. To estimate this magnitude, we again performed ONIOM calculations, to keep all the results at the same level of theory. For the **S3** system, the choice of the averaged structure is justified based on the remarkably low RMSD value (**Sup. Fig. 13**). The **S3-S3** structure was also averaged over the last 40 ns for this purpose, and to be sure that we were computing the interaction energy with the corresponding systems in local minima, we also performed frequency calculations that corroborate the absence of transition states in both the two structures. To our knowledge, this is the first time that such time-consuming force constant calculations are performed in a system of this size. The coordinates of the **S3** & **S3-S3** systems used as Gaussian09 input files for ONIOM (DFT:PM6) interaction energies and frequency calculations are listed in **Sup. Table 5**.

With respect to the **S3-S2** system, which exhibits subtle structural distortion although a dry interface might be formed, the interaction energy for these contacts have been computed over different snapshots of the last 10 ns of the trajectory, averaging the corresponding values (See **Sup. Table 2**.)

The gain of interaction energy upon twisting was also calculated at the M06-2X/6-311+G(d,p) level, since this functional¹⁹ was derived for the computation of non-covalent interactions and has been successfully applied to biological systems.¹⁸ This procedure yielded very similar results (**Sup. Table 2**).

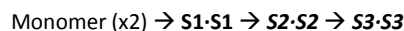
All the calculations were performed using a solvation continuum model (PCM) with the Gaussian09²⁰ package, allowing the effect of solute polarization to be included and avoiding the higher energies derived from gas-phase calculations.

Charged vs. Capped Systems: Two distinct pathways towards the seed formation

We have shown that the smallest stable structural motif that can act as a seed for further nucleation in Sup35 is a *twisted* three-stranded parallel β -sheet (**S3**). The kinetic study shows that the most plausible pathway towards the formation of the **S3·S3** seed of Sup35 can be schematized as:



The alternative pathway:



is shown to be kinetically unstable.

This is the situation for the peptides with neutral termini (*i.e.*, in the context of the full-length protein). However, it seems that for the heptapeptide with charged termini, both are plausible since the electrostatic interactions help hold the two chains together, stabilizing the side chain intermeshed conformation against dissociation.

To explain how the electrostatic repulsion that arises from the proximity of charged termini is overcome, we have simulated the dynamics of +GNY-, +GNNY-, +GNNQY-, +GNNQQY- & +GNNQQNY- peptides in a two-stranded parallel β -sheet. We have run 100 ns of MD on each structure, with the same settings that were listed in the Molecular Dynamics Simulations section. Taking into account that 11 H-bonds are formed between two strands (6 backbone-backbone & 5 sidechain-sidechain), we could determine the minimal number of H-bonds necessary to overcome the electrostatic repulsion. As summarized in **Table S2** at least 7 H-bonds (arising from 2 GNNQY pentapeptides) are needed to achieve stability.

Considering growing β -sheets composed of charged +GNNQQNY- peptides, they can be expected to be more stable due to hyperpolarization of the H-bonding amide groups. Twisting, which manifests itself in dimers, trimers and larger oligomers, will also stabilize β -sheets composed of peptides with charged termini since twisting separates groups carrying like charges, thus reducing their electrostatic repulsion.

SUPPLEMENTARY FIGURES

Figure S1: root mean square deviation (RMSD) for the **S1·S1** dimer. The first 100 ns of the 500 ns run of the simulation presented in the main text (Figure 1) is shown in **brown**. The kinetic stability of an **S1·S1** arrangement is tested, where side chain intermeshing provides the main force holding two peptides together. In two simulations (**yellow** and **brown**), the dimer completely unfolds and dissociates. In two other simulations (**green** and **blue**) the starting structure is largely maintained but the termini undergo significant changes to optimize the formation of a set of stabilizing contacts, which are different from those present in the starting structure. In two other structures (**red** and **black**) the starting **S1·S1** conformation transforms structurally into a pair of H-bonded, anti-parallel β -strands. These results

suggest that the stabilizing contribution afforded by the intermeshing side chains is small. The RMSD values plotted are for both peptides in each dimer (and not to individual peptides).

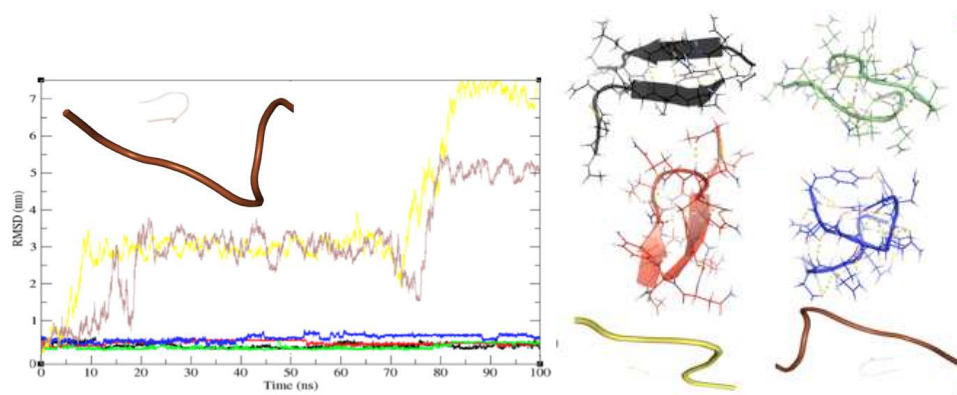


Figure S2: Root mean square deviation (RMSD) for the **S2** system. The **brown** trace represents the first 100 ns of the 500 ns MD run presented in the main text (Figure 1). The other trajectories correspond to repetitions of the simulations to test the convergence of the results. Interestingly, after 100 ns, four out of six **S2** dimers retain conformations closely resemble the starting structure. Two **S2** dimers significantly unfold after 60-80 ns, one of which, in **brown**, dissociates (note that the second **brown** monomer is in the distant background). As in Figure S1, the RMSD values plotted are for both peptides in each dimer, and not to individual peptides.

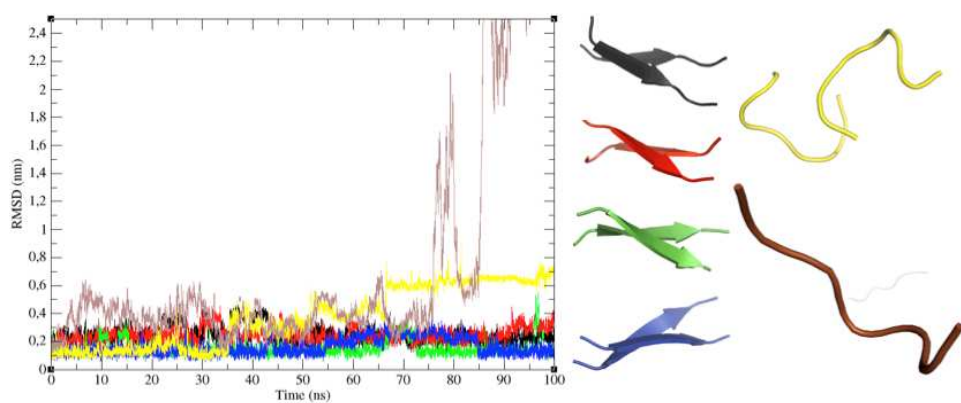


Figure S3: RMSD for the **S3** (which twists to become **S3** in the equilibrium stage). As in other supplementary figures, the **brown** run represents the first 100 ns of the 500 ns run shown in **Figure 1a** in the main text. In all the other independent MD calculations (run for 100 ns and shown in **black, blue, red, yellow and green**), the structures maintain a twisted three stranded β -sheet which minimal structural deviations. This is evidence that twisting provides extraordinary conformational stability. The reproducibility of the results corroborates our proposal that the **S3** structure is highly stable and acts to seed GNNQQNY fibril growth. The RMSD values plotted are for all three peptides in each trimer.

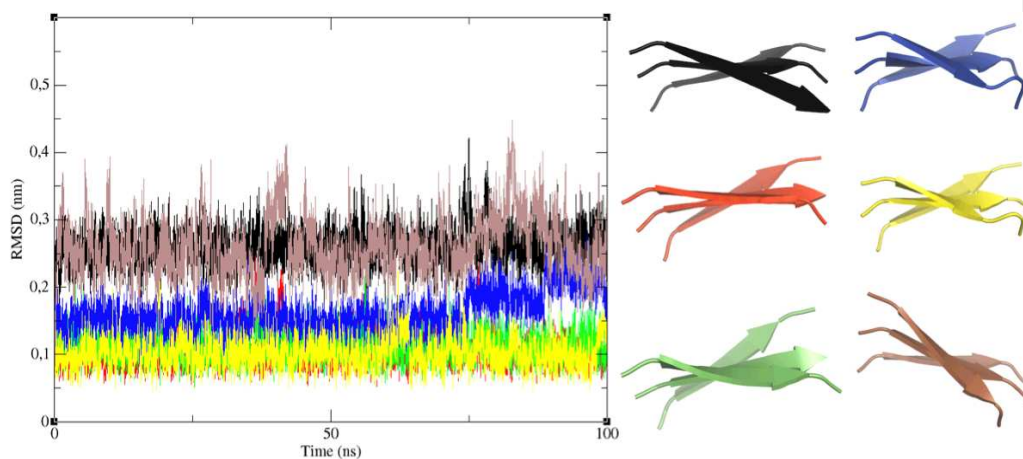


Figure S4: RMSD for the **S2·S2** system. Here the **black** trace corresponds to Figure 1b of the main text (*i.e.*, the first 100 ns of the 500 ns run). This structure is not stable enough to be considered as a key intermediate in Sup35 oligomerization, as can be seen below. All of the snapshots represent the last frame of the trajectory. Only in one case is the starting **S2·S2** configuration retained, although it is highly distorted (**green**). In two simulations (shown in **black** and **red**), one peptide dissociates and the structure evolves towards an **S3**-like system, whereas in the other case (**blue**) it seems to convert into a four stranded β -sheet. The last frame of this four stranded β -sheet was subjected to an additional 100 ns of MD to test its kinetic stability (**yellow**). The sidechains of these two last systems are not shown to highlight their secondary structure, since they differ from the all-parallel β -sheet. These structures might dissociate into two **S2** systems and evolve along our proposed pathway into a seed for amyloid formation (see Figure 3 in the main text). The RMSD values plotted are for correspond to the entire tetrameric assembly.

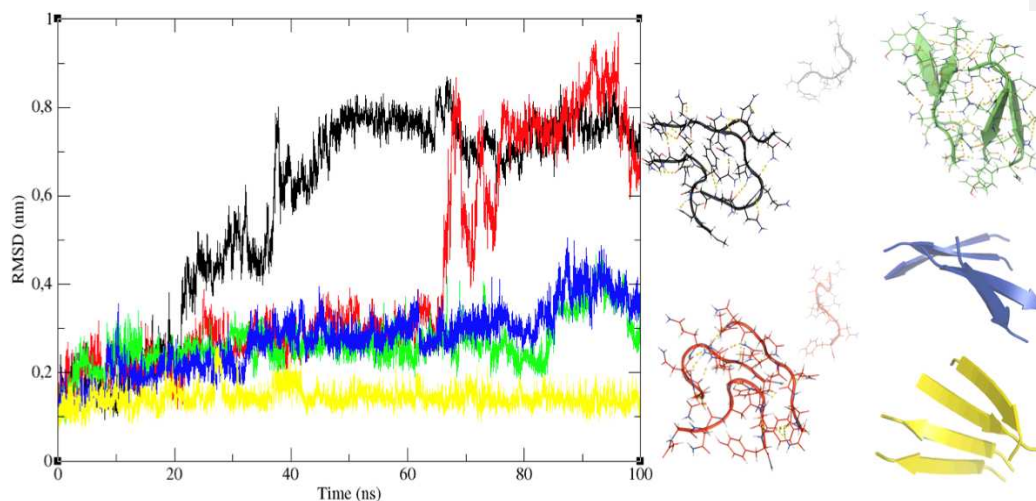


Figure S5: root mean square deviation (RMSD) for the **S2•S1** system. When the system is charged, its behavior is similar to that of **+S1•S1-** and when termini residues are neutral, it resembles **S2** trajectories. The RMSD values shown here, and the remaining figures, correspond to the whole system, not individual peptides.

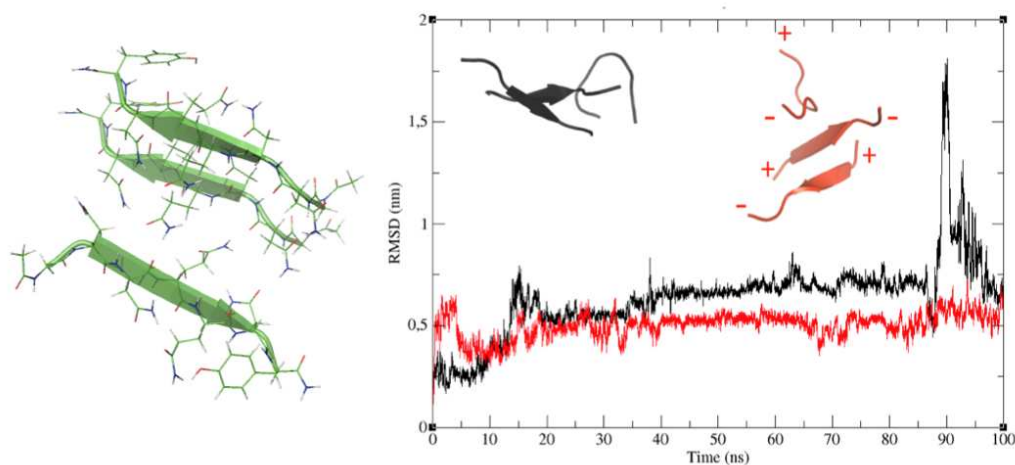


Figure S6: Root mean square fluctuation (RMSF) calculated for all the **S3** structures over the 500 ns trajectory. It is important to note that the lower values correspond to the residues that form the dry interface in the fibril. The fact that this low mobility originates when **S3** twists to yield **S3** suggests that side chain packing to yield the dry interface must proceed after **S3** forms, as can be seen in the proposed mechanism in *Supplementary Figure 11*.

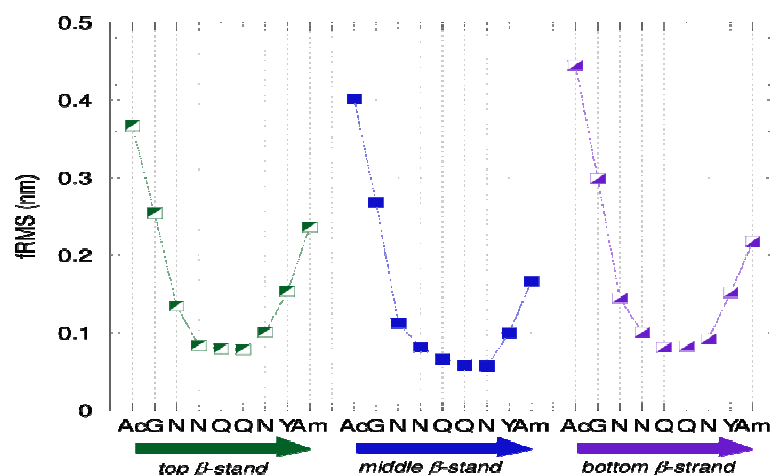


Figure S7: Root mean square deviation (RMSD) of the **NQYGQNQ** mutant. The wt sequence (GNNQQNY) was altered to break up the N-X-Q-X-N motif which had been proposed by Nelson *et al* (Nature, 2005)¹ to be key for dry interface stability. The substitutions N3Y and Q4G impede the formation of stable β -sheet structures, as gauged from three independent 100 ns simulations in which the RMSD value is higher than 0.5 nm. This is evidence that this force field is not biasing these Q/N rich sequences towards any particular conformation, and is therefore appropriate for this system.

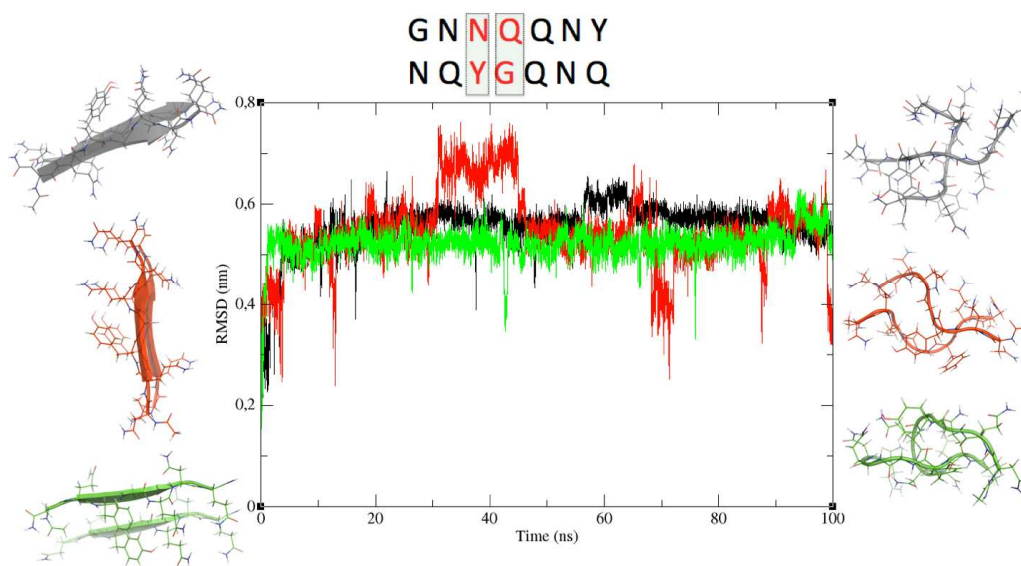


Figure S8: Characterization of GQNNQNY and GNNQQNY peptides by NMR:

Removal of the central Q4 contacts results in non-aggregating systems.

The NMR chemical shift assignments were obtained by a standard approach utilizing TOCSY crosspeaks to assign intra-residue spin systems and ROESY crosspeaks to establish their sequential linkage. These ¹H assignments were corroborated and ¹³C & ¹⁵N assignments were obtained by analysis of the ¹H-¹³C & ¹H-¹⁵N HSQC spectra. Some medium range ROE crosspeaks were seen between the side chain H β and H γ of Gln 5 and the H δ and H ϵ of Tyr 7. In the case of the wild type sequence, we observed that three spin systems, belonging to Gln 4, Gln 5 and Asn 6, share a very similar chemical shift value for their ¹HN and are overlapped in the 1D ¹H spectra. The presence of three ¹HN groups here was confirmed by integrating the peaks in the 1D ¹H spectrum, and the ¹H-¹⁵N HSQC spectrum. The ¹³C α , ¹³C β , ¹⁵N & ¹H α chemical shift values (*vide infra*) measured for homologous ¹H in both peptides similar to each other. Once corrections for local sequence effects are applied, these values are very close to the standard values for expected for short unstructured peptides, and no significant tendency towards partial alpha or beta type backbone structure was detected. Some rather weak sequential ¹HN_(i) – ¹HN_(i+1) ROESY correlations were observed for both the peptides; these signals are consistent with small populations of helix or turn-like conformations. In addition, a few low intensity peaks arising from medium range contacts between the side chains of Gln 5 and Tyr 7 were also detected. These could arise when, for example, the backbone is in an extended conformation so that the Gln 5 and Tyr 7 side chains project in similar directions.

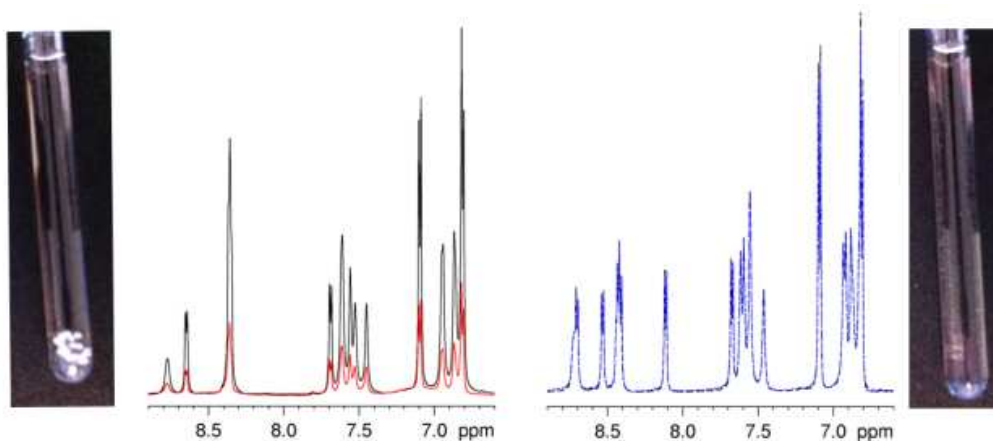
Based on these results, we conclude that these peptides, upon dissolving in aqueous solution, initially adopt a broad ensemble of conformations. This conclusion is consistent with the findings of a ¹H NMR study of GNNQQNY in 50% TFE.³⁸

Sup35: GNNQQNY							Sup35 Negative Control: GQNNQNY						
Residue	¹ H ¹⁵ N	¹ H, ¹³ C α	¹ H, ¹³ C β	¹ H ¹³ C γ	¹ H δ 12 or ¹ H ϵ 12 ¹⁵ N δ or	¹ H δ 22 or ¹ H ϵ 22 ¹⁵ N ϵ	Residue	¹ H ¹⁵ N	¹ H, ¹³ C α	¹ H, ¹³ C β	¹ H ¹³ C γ	¹ H δ 12 - or ¹ H ϵ 12 ¹⁵ N δ or	¹ H δ 22 or ¹ H ϵ 22 ¹⁵ N ϵ
Gly 1		3.86 43.67					Gly 0		3.85 43.6				
							Gln 1*	8.71 119.8	4.39 56.41	2.09, 1.99 29.96	2.35 34.04	7.54 112.3	6.86
Asn 2	8.75 118.7	4.68 53.50	2.85, 2.78 39.1 \pm 0.2		7.60† 112.8	6.94†	Asn 2	8.69 120.1	4.71 53.61	2.85, 2.76 39.13		7.60 113.2	6.92
Asn 3	8.62 119.4	4.68 53.84	2.86, 2.78 39.1 \pm 0.2		7.59† 113.0	6.93†	Asn 3	8.52 119.7	4.72 53.65	2.84 39.05		7.58 112.6	6.90
Gln 4	8.34 120.2	4.29 56.32	2.11, 1.98 29.60	2.33 34.17	7.50 112.5	6.84	Gly 4*	8.42 108.9	3.93 45.81				
Gln 5	8.333 120.1	4.28 56.08	1.97, 1.90 29.79	2.25 34.04	7.43 112.4	6.81	Gln 5	8.11 119.4	4.33 55.82	1.99, 1.90 29.85	2.25 34.00	7.45 112.4	6.82
Asn 6	8.34 120.2	4.70 53.49	2.76, 2.65 39.37		7.53 112.8	6.86	Asn 6	8.40 120.3	4.72 53.59	2.79, 2.69 39.34		7.54 112.4	6.87

Tyr 7	7.76	4.43	3.08, 2.90	7.10	6.81	Tyr 7	7.66	4.40	3.08, 2.92	7.06	6.78 (H _c)
	124.1	59.00	39.44	(H _δ)	(H _c)		125.0	59.35	39.6	(H _δ)	113.0

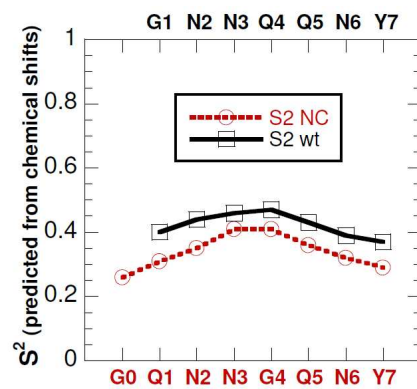
† Inter-residual ambiguity, *swapped relative to the wt sequence

The downfield region of the 1D ¹H NMR spectra of both peptides and these samples' photographs after five days of incubation are shown below:



Sup 35 GNNQQNY (*left panel*) 1 h (**black** spectrum) and 5 days (**red spectrum** & photo) after sample preparation. Sup 35 negative control (*right panel*) 1 h (dotted **black** spectrum) and 5 days (dashed **blue spectrum** and photo) after the sample was dissolved.

After five days of incubation at room temperature, visible aggregates appear in the GNNQQNY sample. Using the signal of DSS, as an internal intensity standard, we determined that 70% of GNNQQNY had aggregated after five days of incubation (red vs. black spectra, *vide supra*). In contrast, the control peptide showed no visible aggregates and no significant alternation in the NMR signal intensity after five days. These results are in line with the somewhat lower backbone order parameters for the negative control peptide with Gly 4 (*vide infra*) and support the conclusions drawn from the MD results on the importance of the central residues, particularly Gln 4 in wild type GNNQQNY oligomerization.



The predicted order parameter (S^2) of wt (black open squares) and the neg. control peptide (red open circles).

Figure S9: Comparing charged versus neutral termini in **S1·S1**. If the terminal residues are charged, the favorable electrostatic interactions increase the stability of this conformer. In contrast, the **S1·S1** system with capped, neutral termini quickly dissociates.

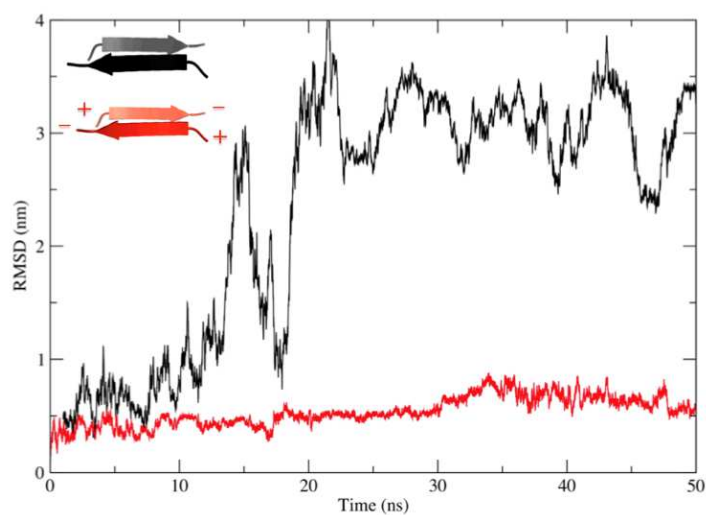


Figure S10: RMSD for the **S3-S2** system. Three different simulations were performed, varying the initial velocities on each atom (**red**) or the initial configuration (**green**), with respect to the initial one (**black**). In all cases, the RMSD was calculated with respect to the corresponding averaged structures.

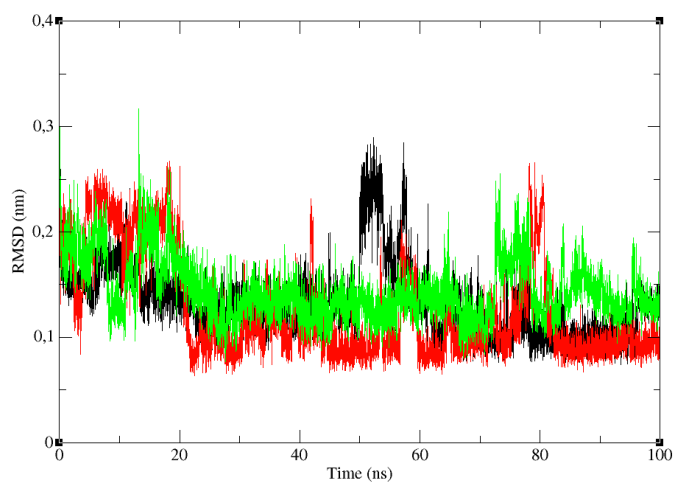


Figure S11: Stability of the trimers. The stability of the **S3** trimers that evolve towards combining to form the **S3-S3** hexamer (**Figure 4**, main text) were analyzed through their RMSD profiles. As expected from the multiple runs on **S3**-like systems, they remained stable throughout the whole simulation period and that twisted conformation is retained when the dry interface is formed. The color code used is that corresponding to each of these trimers in **Figure 4** in the main text.

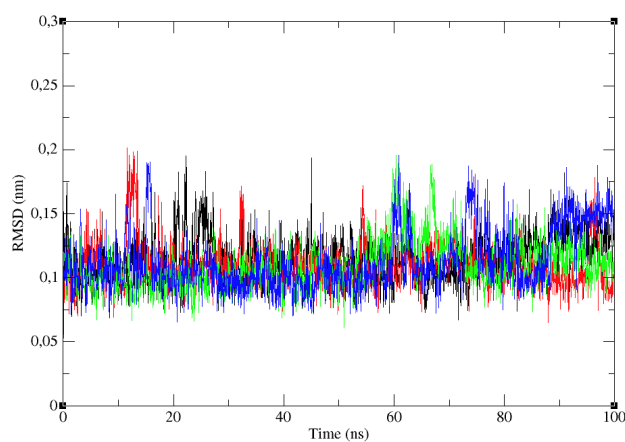


Figure S12: RMSD of the **S3** system fit to the averaged structure (**red**) and initial one (**black**), when the acetyl and amide terminal capping moieties are removed. **S3**'s remarkably stability is the reason for choosing this conformer for the ONIOM calculations: as there are no substantial variations during the whole trajectory, the averaged ensemble conformation is representative of all the frames

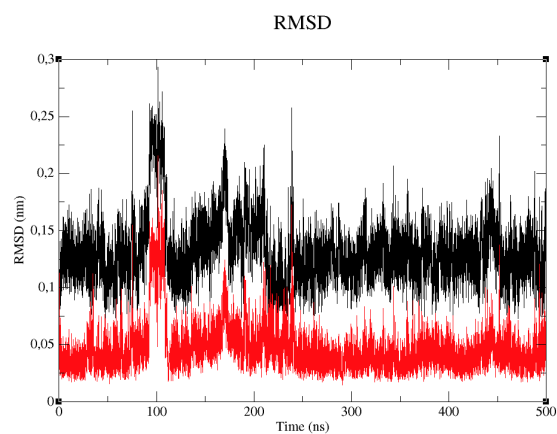
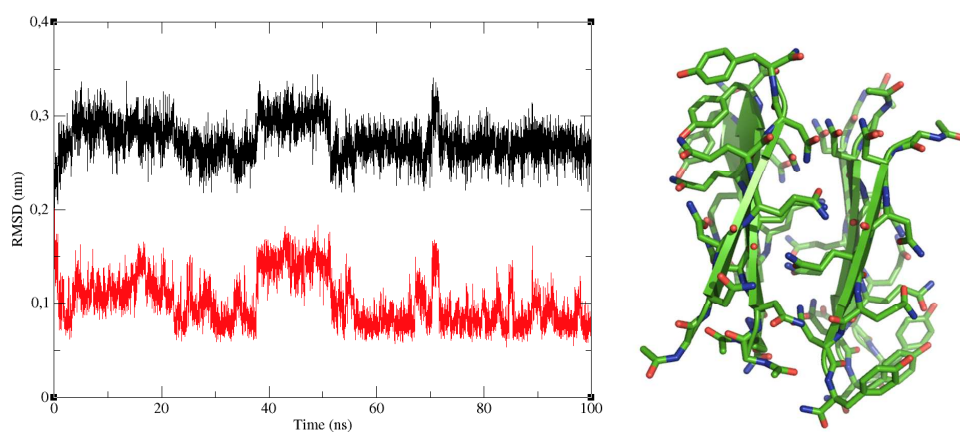


Figure S13: Stability of S3·S3. TOP: The RMSD of the nucleus or seed for aggregation is shown below to illustrate how this system is properly represented by the force field and the set of parameters used in this study. The small fluctuations are mostly due to the acetyl and amide terminal moieties used to cap the termini. The structures of the resulting trajectory were fitted to two reference conformations: the initial one (at time zero, **black** curve) and the time-averaged conformation (**red** curve). BOTTOM: Cartoon and stick representation of the last frame of the 100 ns run, showing the twisted **S3·S3** system. Hydrogen atoms are not shown for the sake of clarity.



SUPPLEMENTARY TABLES

Table S1: Main findings in previous manuscripts analyzing GNNQQNY at atomic resolution.

Reference	Chief Methodology	Main Findings
(Balbirnie et al. 2001) ²¹	Domain mapping / X-ray diffraction	Determined that a heptapeptide of the large Sup35 protein that is able to form long fibrils (as seen by EM) that bind amyloid-specific dyes. By X-ray diffraction, the spacings between the H-bonded (4.8 Å) and van der Waals packed (10 Å) peptides were determined.
(Nelson et al. 2005) ¹	X-ray crystallography	Determined the crystal structure of the Sup35 heptapeptide GNNQQNY with charged termini and a Zn ⁺⁺ coordinated to the C-terminal –COO- and N-terminal amine group. The structure revealed the formation of a complete set of side chain and backbone H-bonds. Each peptide forms a β-strand that is anti-parallel to the fibril axis. There is no twisting. The Sup. Materials of this paper provides ballpark estimates of the factors that contribute to the fibril's conformational stability, emphasizing the importance of H-bonding, close side chain packing, Tyr ring stacking and solvation, and suggest that structural nucleus could be formed by ≤ 6 peptides.
(Sawaya et al. 2007) ²²	X-ray crystallography	Determined the crystal structure of amyloid fibrils formed by several different short peptides. Whereas all the fibrils exhibit very tight sidechain packing and complete H-bond networks, there is a wide variation in the hydrophobicity, hydration and orientation of the strands.
(Tsemekhman et al. 2007) ¹⁶	Energetics by Computation	First attempt to measure the stabilizing bonus that should arise from hyper-polarization of amide groups in the Sup35 amyloid fibril and the smaller oligomers that maintain the same conformation. Did not consider twisting.
(Guo and Eisenberg 2008) ²³	X-ray crystallography	Determined the structure of NNQQNY inserted in a folded protein. This segment adopts a twisted, antiparallel conformation that is incompatible with aggregation.
(Lipfert et al. 2005) ²⁴	MD	Observed that the release of water molecules drives oligomerization. They study the system with charged termini. They see more disorder at the termini when the peptides are aligned in parallel as compared to anti-parallel due to charge repulsion.
(Zheng et al. 2006) ²⁵	MD	Performed short (10 ns) simulations on small oligomers of different sizes based on the X-ray structure of Nelson et al. 2005. Peptides have terminal charges. Simulate point mutations. Recognize the importance of Q4, N6 & Y7 to stability. See that one sheet with 3 or more β-strands is stable. Propose that the nucleus could be as small as 3 or 4 peptides. Did not recognize the importance of twisting.

(Esposito et al. 2006) ²⁶	MD	Performed short simulations of longer oligomers of Sup35 in a fibril like conformation, starting from X-ray structure. Discovered that the fibril, composed of peptides with charged or neutral termini, undergoes left-handed twisting up to 10°. Report that the dry interface, esp. Q4, is very rigid. Twisting leads to further dry interface contacts.
(deSimone et al. 2008) ²⁷	MD	Utilized replica exchange MD to look at the stability of several small GNNQQNY oligomers. See that "S3-S3" and larger oligomers are stable, and that "S3-S2" is borderline. Emphasize that the exposed edge strands are "highly reactive".
(Vitagliano et al. 2008) ²⁸	MD	Utilized replicate exchange MD to determine the free energy surface for GNNQQNY dimers whose strands are in parallel or anti-parallel orientations, for charged, half-charged and neutral peptides. See that the energy differences between the two orientations are subtle. Neutralizing the terminal charges favors the parallel orientation.
(Zhang et al. 2007a) ²⁹	MD	Performed many simulations to follow the formation of GNNQQNY oligomers (dimers, trimers & tetramers). See twisting already in these small oligomers. The peptides seem to be charged. See that parallel b-sheets form faster and dissociate more slowly.
(Periole et al. 2009) ³⁰	MD	Determined that twisting is favored when the termini are neutral, and even more so when they are charged. Likewise, solvent water is not essential for twisting, but does stabilize the twisted form.
(Park et al. 2009) ³¹	MD	Measured the energetics for adding a strand to a fibril, for several different peptides, including GNNQQNY. Find that the conformation adopted is determined by the energetics.
(van der Wel et al. 2007) ³²	ssNMR	Detected, via ssNMR measurements, the existence of polymorphism in Sup35 amyloid fibrils: monoclinic (like Eisenberg <i>et al.</i> 's structure) and orthorhombic. Emphasize the importance of Tyr ring packing.
(van der Wel et al. 2010) ³³	ssNMR	Reported three different types of fibrils; all are parallel, in-register β -sheets.
(Wang et al. 2008) ³⁴	MD	Utilized RT (298K) and higher temperatures (up to 498 K) to follow the dissociation of GNNQQNY duodecamers. Based on this, they propose a pathway for oligomer formation which emphasizes the importance of "S3-S1" and "S2-S2" tetramers as key intermediates/transition states. Peptides seem to have terminal charges and oligomers are twisted. Assigned equal importance to H-bonds and van der Waals interactions.
(Reddy et al. 2009) ³⁵	MD	Examined the process of the addition of GNNQQNY monomers initially in a random coil state onto a preformed amyloid fibril. Emphasized the importance of dehydration of GNNQQNY, a polar peptide, esp. compared to the same growth process for A β , a much more hydrophobic peptide.
(Reddy et al. 2010) ³⁶	MD	Computational studies of GNNQQNY beta sheets docking to form the dry interface. When two β -sheets, each formed of 8 monomers, dock together, water molecules can be trapped inside. These "water nanowires" are meta-stable for GNNQQNY but not seen for A β , which is much more hydrophobic.
(Marshall et al. 2010) ³⁷	Linear Dichroism, X-ray diffraction, Tyr	Monitored structural changes that occur when the Sup35 amyloid fibril crystallizes. Saw that the Tyr side chain undergoes important changes in conformation and exposure to

Field Code Changed

Field Code Changed

fluorescence,
acrylamide quenching

solvent. Affirmed that the crystalline state is more stable than the fibril state.

(Portillo et al. 2012) ²	ThT Fluor. AFM force spectrometry	Measured the effect of variation of pH and ionic strength on the kinetics of amyloid formation (monitored by ThT fluorescence) and the interaction energy of Sup35 peptides in a dimer (using AFM). Find that increasing the ionic strength and pH values close to the isoelectric point (5.3) speed amyloid formation, but that these environment conditions do not strongly affect the interaction energy.
(Chae et al. 2004) ³⁸	Liquid state NMR	¹ H NMR study of GNNQQNY in different concentrations of water&TFE. The obtained ¹ H α chemical shifts in 50% are characteristic of random coil.
(Nasica-Labouze et al. 2011) ³⁹	Coarse-grain implicit solvent MD	Coarse-grain, implicit solvent MD study of the different oligomers formed by GNNQQNY, starting from boxes with 3, 12 or 20 monomers. A large variety of oligomers and structures are observed. The most Representative structures are studied further using all atom MD with explicit solvent.
Nasica-Labouze & Mousseau 2012) ⁴⁰	Coarse-grain, implicit solvent MD	Coarse-grain, implicit solvent MD study of the kinetics of GNNQQNY association, starting from 20 monomers. Observe formation of a variety of beta sheet structures; with small oligomers forming before larger oligomers.

Table S2: Summary of the ONIOM (DFT:PM6) RESULTS. All the values were computed with the PCM method to account for environment effects. In the case of the gain of interaction energy upon twisting, results from gas-phase calculations were included -in parenthesis- to show the agreement between B3LYP and M06-2X functionals.

SYSTEM	ΔE_{int} (kCal/mol)
S3-S3	-30.96
S2-S3 -90ns	-5.68
S2-S3 -92ns	NOT CONSISTENT
S2-S3 -94ns	-7.74
S2-S3 -96ns	-11.85
S2-S3 -98ns	-4.13
S2-S3 -100ns	-12.70
<ΔE_{int} (S3-S2) >	-8.42
FUNCTIONAL	GAIN OF ΔE_{int} UPON TWISTING (kCal/mol)
B3LYP	-8.92 (-12.88)
M06-2X	-9.11 (-12.98)

Table S3: Summary of the Simulation Results Reported Here.

Simulation	# of Simulations / Run time (ns)	Main results	Corresponding Figure
S1-S1	1 x 500 5 x 100 1 x 50 (charged)	Stable for no longer than 10 ns. Dissociation or rearrangement into anti β -sheet	1 (main text), 1 (Sup. Info) & 9 (Sup. Info)
S2	1 x 500 5 x 100	Stable for at least 75 ns. Remarkably stability	1 (main text) & 2 (Sup. Info)
S3	1 x 500 5 x 100	Remarkably stable in all of the cases	1 (main text) & 3 (Sup. Info)
S2-S2	1 x 500 5 x 100	Evolution towards S3 / Structural rearrangement	1 (main text) & 4 (Sup. Info)
S2-S1	1 x 100 (charged) 1 x 100 (capped)	When charged, behaves as +S1-S1- . When capped, behaves as S2 .	5 (Sup. Info)
NQYGQNQ	3 x 100	β -sheet secondary structure not retained	7 (Sup. Info)
S3-S2	3 x 100	Stable, although not as remarkable as S3-S3	10 (Sup. Info)
4 x S3	1 x 100	Evolution towards S3-S3 nucleus	4 (main text) & 11 (Sup. Info)
S3-S3	1 x 100	Stable	12 (Sup. Info)

Total simulation time: 5.05 μ s.

Table S4: Semi-quantitative analysis of the number of H-bonds required to overcome the electrostatic repulsion between charged peptides arranged in a two-stranded parallel β -sheet.

System	Number of H-bonds (Backbone + side chain = total)	Stable?
G-N-Y G-N-Y	$2 + 1 = 3$	NO
G-N-N-Y G-N-N-Y	$3 + 2 = 5$	NO
G-N-N-Q-Y G-N-N-Q-Y	$4 + 3 = 7$	YES
G-N-N-Q-Q-Y G-N-N-Q-Q-Y	$5 + 4 = 9$	YES
G-N-N-Q-Q-N-Y G-N-N-Q-Q-N-Y	$6 + 5 = 11$	YES

23-23 System.

[illegible]

SUPPLEMENTARY REFERENCES

- (1) Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, Grothe R, Eisenberg D (2005) Structure of the cross-beta spine of amyloid-like fibrils *Nature* 435, 773-778.
- (2) Portillo AM, Krasnoslobodtsev AV, Lyubchenko YL (2012) Effect of electrostatics on aggregation of prion protein Sup35 peptide *J. Phys.: Condens. Matter.* 24: 164205.
- (3) Hess B Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4: 435.
- (4) Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Kepleis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78: 1950-8.
- (5) Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79: 926-935.
- (6) Berhanu WM, Hansmann UHE (2012) Side-chain hydrophobicity and the stability of Abeta 16-22 aggregates. *Protein Science* 21: 1837-1848
- (7) Darden T, York D, Pedersen L (1993) Particle mesh ewald: an N-lon(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98: 10089-10092
- (8) Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81: 3684.
- (9) Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18: 1463-1472.
- (10) Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52: 255
- (11) Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52: 7182.
- (12) Lange OF, van der Spoel D, de Groot BL (2010) Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophys. J.* 99: 647-55.
- (13) Dapprich S, Komaromi I, Byun KS, Morokuma K, Frisch MJ (1999) A new ONIOM implementation in Gaussian98. Part 1. The calculation of energetics, gradients, vibrational frequencies and electric field derivatives. *Journal of Molecular Structure: TEOCHEM* 461-462: 1.
- (14) Stewart JJ (2007) Optimization of parameters for semi empirical methods V: modification of approximations and application to 70 elements. *J Mol. Model* 13: 1173-1213.
- (15) Gesto DS, Cerqueira NM, Fernandes PA, Ramos MJ (2013) Unraveling the enigmatic mechanism of L-asparaginase II with QM/QM calculations. *J. Am. Chem. Soc.* 135: 7146-58.
- (16) Tsemekhman K, Goldschmidt L, Eisenberg D, Baker D (2007) Cooperative hydrogen bonding in amyloid formation. *Protein Sci.* 116: 761-4.

- (17) Cerqueira N, Fernandes P, Ramos MJ (2011) Computational mechanistic studies addressed to the transamination reaction present in all pyridoxal 5' phosphate-requiring enzymes. *J. Chem. Theory Comput.* 7: 1356-1368.
- (18) Martín-Pintado N, Yahyaee-Anzahaee M, Deleavey GF, Portella G, Orozco M, Damha MJ, González C (2013) Dramatic effect of furanose C2' substitution on structure and stability: directing the folding of the human telomeric quadruplex with a single fluorine atom. *J. Am. Chem. Soc.* 135: 5344-7.
- (19) Zhao Y, Truhlar D (2008) The M06 suite of density functional for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functional and 12 other functionals. *Theor. Chem. Acc.* 120: 215-241.
- (20) Frisch MJ *et al.* Gaussian 09 Revision A.2, 2009. Gaussian Inc. Wallingford CT.
- (21) Balbirnie M, Grothe R, Eisenberg D (2001) An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proc. Natl. Acad. Sci. (USA)* 98: 2375-2380.
- (22) Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apostol MI, Thompson MJ, Balbirnie M, Wiltzius JJ, McFarlane HT, Madsen AO, Riekel C, Eisenberg D (2007) Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* 447: 453-457.
- (23) Guo Z, Eisenberg D (2008) The structure of a fibril-forming sequence NNQQNY, in the context of a globular fold. *Prot. Sci.* 17: 1617-1623.
- (24) Lipfert J, Franklin J, Wu F, Doniach S. (2005) Protein misfolding and amyloid formation for the peptide GNNQQNY from yeast prion protein Sup35: simulation by reaction path annealing. *J. Mol. Biol.* 349: 648-651.
- (25) Zheng J, Ma B, Tsai CJ, Nussinov R (2006) Structural stability and dynamics of an amyloid-forming peptide GNNQQNY from the yeast prion sup-35. *Biophys J* 91: 824-833.
- (26) Esposito L, Pedone C, Vitagliano L (2006) Molecular dynamics analyses of cross-beta-spine steric zipper models: beta sheet twisting and aggregation. *Proc. Natl. Acad. Sci. (USA)* 103: 11533-11538.
- (27) deSimone A, Esposito L, Pedone C, Vitagliano L. (2008) Insights into the stability and toxicity of amyloid-like oligomers by replica exchange molecular dynamics analyses. *Biophys J.* 95: 1965-1993.
- (28) Vitagliano L, Esposito L, Pedone C, DeSimone A. (2008) Stability of single sheet GNNQQNY aggregates analyzed by replica exchange molecular dynamics: antiparallel versus parallel association. *Biochem Biophys Res Com* 377: 1036-1041.
- (29) Zhang Z, Chen H, Bai H, Lai L. (2007) Molecular dynamics simulations on the oligomer-formation process of the GNNQQNY peptide from yeast prion protein Sup35. *Biophys. J.* 93: 1484-1492.
- (30) Periole X, Rampioni A, Vendruscolo M, Mark AE (2009) Factors that affect the degree of twist in beta-sheet structures: A molecular dynamics simulation study of a cross-beta filament of the GNNQQNY peptide. *J. Phys. Chem.* 113: 1728-1737.

- (31) Park J, Kahng B, Hwang W (2009) Thermodynamic selection of steric zipper patterns in the amyloid cross-beta spine. *PLoS Comput. Biol.* 5: e1000492.
 - (32) van der Wel PCA, Lewandowski JR, Griffin RG (2007) Solid state NMR study of amyloid nanocrystals and fibrils formed by the peptide GNNQQNY from yeast prion protein Sup35p. *J. Am. Chem. Soc.* 129: 5117-5130.
 - (33) van der Wel PCA, Lewandowski JR, Griffin RG (2010) Structural characterization of GNNQQNY amyloid fibrils by magic angle spinning NMR *Biochemistry* 49: 9457-9469.
 - (34) Wang J, Tan C, Chen HF, Luo R (2008) All-atom computer simulations of amyloid fibrils disaggregation. *Biophys. J.* 95: 5037-5047.
 - (35) Reddy G, Straub JE, Thirumalai D (2009) Dynamics of locking of peptides onto growing amyloid fibrils. *Proc. Natl. Acad. Sci. (USA)* 106: 11948-11953.
 - (36) Reddy G, Straub JE, Thirumalai D (2010) Dry amyloid fibril assembly in a yeast prion peptide is mediated by long-lived structures containing water wires. *Proc. Natl. Acad. Sci. (USA)* 107: 21459-21464.
 - (37) Marshall KE, Hicks MR, Williams TL, Hoffmann SV, Rodger A, Dafforn TR, Serpell LC (2010) Characterizing the assembly of the Sup35 yeast prion fragment, GNNQQNY: structural changes accompany a fiber-to-crystal switch. *Biophys. J.* 98: 330-338.
 - (38) Chae YK, Lee K, Kim Y (2004) NMR Studies of the prionogenic peptide derived from Sup35 Protein. *Prot. Pept. Lett.* 11: 23-28.
 - (39) Nasica-Labouze J, Meli M, Derreumaux P, Colombo G, Mousseau N (2011) A multiscale approach to characterize the early steps of the amyloid-forming peptide GNNQQNY from the yeast prion Sup-35. *PLoS Comp. Biol.* 7: e1002051.
 - (40) Nasica-Labouze J, Mousseau N (2012) Kinetics of Amyloid Aggregation: A Study of the GNNQQNY Prion Sequence. *PLoS Comp. Biol.* 8:e1002782
-