Systematic Examination of Polymorphism in Amyloid Fibrils by Molecular-Dynamics Simulation

Joshua T. Berryman,[†] Sheena E. Radford,[‡§] and Sarah A. Harris[‡¶*]
[†]Université du Luxembourg, Luxembourg; and [‡]Astbury Centre for Structural Molecular Biology, [§]Institute of Molecular and Cellular Biology, and [¶]School of Physics and Astronomy, University of Leeds, Leeds, United Kingdom

ABSTRACT Amyloid fibrils often exhibit polymorphism. Polymorphs are formed when proteins or peptides with identical sequences self-assemble into fibrils containing substantially different arrangements of the β-strands. We used atomistic molecular-dynamics simulation to examine the thermodynamic stability of a amyloid fibrils in different polymorphic forms by performing a systematic investigation of sequence and symmetry space for a series of peptides with a range of physicochemical properties. We show that the stability of fibrils depends on both sequence and the symmetry because these factors determine the availability of favorable interactions between the peptide strands within a sheet and in intersheet packing. By performing a detailed analysis of these interactions as a function of symmetry, we obtained a series of simple design rules that can be used to determine which polymorphs of a given sequence are most likely to form thermodynamically stable fibrils. These rules can potentially be employed to design peptide sequences that aggregate into a preferred polymorphic form for nanotechnological purposes.

INTRODUCTION

Amyloid fibrils are insoluble fibrous aggregates of proteins or peptides that possess a cross-β architecture (1). Although amyloid deposition in humans is generally associated with degenerative disease (2), fibrils also have biological functions such as melanin biosynthesis, secretory hormone storage, bacterial biofilm formation, and spider silk production (3). They can also confer variant phenotypes that are passed on through cell division, as observed for fungal prions (4). Amyloid fibrils are comprised of long β-sheets that stack together to form protofilaments, which in turn twist around each other to form a fiber (5,6). Many amyloid-forming sequences display polymorphism in that they assemble into fibrils with different β-strand arrangements in response to differences in growth conditions or seeding (7). This can result in different fiber morphologies (8) and biological activities (9). In the case of transmissible prions, these polymorphs are known as strains. The prion strain phenomenon is of particular importance because the transmissibility of the prion across different species has been shown to be strain dependent (10). Various simulation studies have examined the phenomenon of polymorphism in amyloids, including a detailed study of the thermodynamics of parallel versus antiparallel polymorphs of GNNQQNY (11), replica-exchange molecular-dynamics (MD) simulations that showed that the ability of zinc ions to bind to different locations in the N-terminal domain of the Alzheimer's Aβ protein can influence polymorphism (12), an MD investigation of the relative stabilities of different polymorphs of the U-turn conformation of A$\beta_{17\text{-}42}$ (13), and MD calculations that provided insight into how the relative stabilities of polymorphs of Aβ_{42} change with pH (14). MD calculations of the aggregation of A$\beta_{25\text{-}35}$ oligomers have also demonstrated that polymorphism can be determined in the very first steps of fibril formation (15).

Because of the insolubility, filamentous nature, and heterogeneity of amyloid fibrils, it has been difficult to obtain structural information about them at the atomic level. Despite these difficulties, however, a number of 3D crystal structures have been successfully determined from microcrystals of peptides 6–7 residues in length that form β-sheet arrays. These have been shown to adopt one of eight distinct symmetry relations between adjacent peptides (16,17). The symmetries are illustrated with the use of left hands in the key to Fig. 1. To construct these eight symmetries, structures can be assembled from a pair of β-sheets in which the β-strands within each sheet are in either a parallel (classes 1–4) or an antiparallel (classes 5–8) arrangement. However, the relationship between the peptide sequence and its ability to adopt one or more of these different polymorphic forms (i.e., the inherent ability of different sequences to display polymorphism) is unknown. It is difficult to address this issue experimentally because it is not possible to construct and test each of the different polymorphs in a systematic manner in the laboratory.

Here, we used atomistic MD simulations to investigate the thermodynamic stability of amyloid fibrils formed from different peptide sequences over all eight symmetries. By systematically spanning sequence and symmetry space, we were able to obtain a series of simple design rules for determining which of the eight symmetry classes are compatible with a stable fibrillar structure for a particular peptide sequence. The results obtained provide new (to our knowledge) insights into the repertoire of amyloid structures for a given sequence, and reveal why some sequences have the propensity to form a wider range of polymorphs than others.

Submitted October 4, 2010, and accepted for publication February 15, 2011.

*Correspondence: s.a.harris@leeds.ac.uk

Editor: Ruth Nussinov.

0006-3495/11/05/2234/9 $2.00

doi: 10.1016/j.bpj.2011.02.060

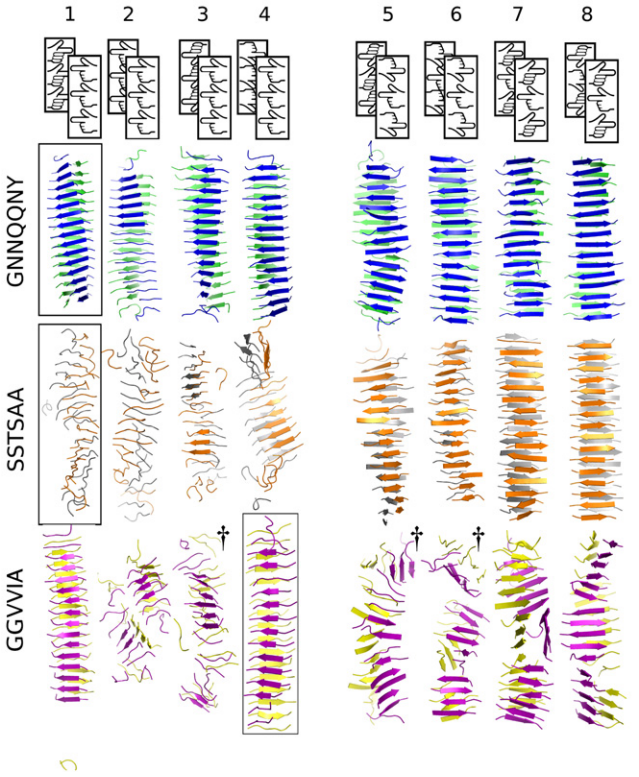


FIGURE 1 (Color available online only). Simulations of GNNQQNY, SSTSAA, and GGVVIA across the eight symmetry classes. Molecular conformations of each peptide aggregate after 10 ns MD simulation. The symmetry relation is illustrated in each case with images of left hands. Assemblies derived from experimental crystal structure data are highlighted with boxes. The structures marked † disintegrated rapidly during the MD and were not continued for the full 10 ns.

MATERIALS AND METHODS

Construction of the initial conformation of the peptide aggregates

The starting structures of the amyloid-like aggregates were constructed from published experimental crystal structures (16) whenever they were available. Crystalline water molecules were retained in cases where there was the possibility of structural significance (SSTSAA class 1, GGVVIA class 4, and VEALYL class 7). When crystal structures were not available, systems were designed rationally with the use of the Nucleic Acid Builder (18). The goal was to maximize backbone and side-chain hydrogen bonding while also obtaining good steric packing within a given symmetry. The glutamic acid (E) present in VEALYL was protonated to have no net charge, in line with the low pH at which the crystals were formed (16). Each amyloid-like aggregate contained two β-sheets, and each β-sheet contained 16 peptides. Initial conformations were rectangular in each case (without a twist). The initial backbone angles for the designed systems were either taken from experimental crystal structures that were compatible with the symmetry or set to typical values for parallel or antiparallel β-sheets (19).

MD simulations

We carried out simulations using the AMBER 9 program (20) with the AMBER99 all-atom force field (21) and the TIP3P explicit water model (22). Systems were solvated in a truncated-octahedral periodic box with a 15 Å cutoff between the solute and the box edge. Because all of the systems simulated carried no net charge, there was no requirement to neutralize with counterions. The particle-mesh Ewald module of AMBER9 was used to calculate long-range electrostatic interactions. We constrained all covalent bonds to hydrogen atoms using the SHAKE algorithm, allowing an integration time step of 2 fs. All MD simulations were performed at constant temperature (300 K) and pressure (1 atm). For simulations based on an experimentally determined crystal structure, the root mean-squared deviations (RMSDs) from the starting structures are provided in Fig. S1 of the Supporting Material. We repeated the simulations based on crystal structure data using the CHARMM22/CMAP (23) force field in conjunction with the NAMD program (24). The AMBER03 all-atom force field was used to rerun all of the 24 simulations of the eight polymorphs of GNNQQNY, SSTSAA, and GGVVIA (Fig. S2 and Fig. S3). Simulations of the 14 polymorphs that maintained sufficient structural order to be classified as stable (order parameter (OP) < 0.07) were extended to 20 ns (as was the unstable SSTSAA class 1, for comparison). Although these polymorphs are ordered, they are not homogeneous and may contain structural defects. To obtain the relative populations of each polymorph that would be expected to be observed in solution in the absence of kinetic effects, it is necessary to minimize the number of these structural defects. To that end, we split the 32-mer aggregates into tetramers and then calculated the average tetramer structure. We discounted tetramers that contained structural defects that were sufficiently disruptive to be observable by visual inspection, and built up a new 32-peptide structure from the average tetramer. We then relaxed these new symmetrized structures by performing a 2 ns MD at 300 K using the generalized Born with surface area (GB/SA) implicit solvent model available within AMBER (25). The enthalpy of each polymorph E_i was calculated from the average force-field energy measured over the final 100 ps of this simulation, which was in turn used to calculate the Boltzmann weighted probability P_i according to:

$$P_i = \frac{e^{-\frac{E_i}{kT}}}{\sum_i e^{-\frac{E_i}{kT}}} \quad (1)$$

The sum was performed over all stable polymorphs, and the Boltzmann weights in Fig. S8 are appropriately normalized so that all probabilities add up to unity for each separate polymorph. To calculate the enthalpies of the structures sampled from the final 1 ns of the MD simulations, the GB/SA method was applied to successive solute structures sampled from the solvated trajectory. All molecular representations in the figures were prepared with the use of Pymol (26).

Assessment of the structural stability of the aggregates

We calculated the shape complementarities as described previously (27,28), and measured the hydrogen-bonding occupancies using the PTRAJ module of the AMBER9 package (20) with an angle cutoff of 135° and a distance cutoff of 3.5 Å. The stability of the aggregates was quantified by defining an OP:

$$\mathrm{OP} = \mathrm{STD}(\cos\theta) \quad (2)$$

where θ is the angle between the end-to-end vectors (Cα-Cα of terminus residues) of adjacent strands in each β-sheet, and STD() is the instantaneous standard deviation measured over the fibril. The parameter OP provides a measure of the deviation of the aggregate from a regular helical structure independently of the particular twist of the helix. The structures with an OP > 0.07 were visibly disordered and thus were defined as unstable. The peptides at the fibril ends (four in total) were discounted in all calculations of stability of the aggregates.

RESULTS

Peptide sequences and symmetries

Simulations of fibril polymorphs

To explore the relationship between the peptide sequence and its ability to form multiple polymorphic forms, we built the peptide sequences GNNQQNY, SSTSAA, and GGVVIA into fibril arrays in all eight symmetries, for a total of 24 polymorph simulations. We chose these three sequences because they are representative of sequences with a range of physicochemical properties. GNNQQNY is taken from the N-terminal region of the (Q/N-rich) yeast prion Sup35, which demonstrates strain behavior and polymorphism (4) in common with many other prions (29). This sequence is strongly polar, and the amide side chains are capable of forming continuous hydrogen-bond ladders within a given β-sheet. Previous resulted obtained by solid-state NMR (30), x-ray crystallography (16,31), x-ray diffraction (32), and MD simulations (11) showed the ability of this sequence to adopt at least two fibril forms. SSTSAA, a fragment taken from bovine RNase A, is also predominantly polar but contains no Q or N residues and therefore no amide side chains. GGVVIA, which comprises residues 37–42 of Aβ42, was chosen as a representative nonpolar sequence. Amyloid-like structures for these systems, each in a single symmetry class (class 1, 1, and 4, respectively) have been determined by x-ray crystallography (16). In addition to the crystal structures (pdb codes 1yjp, 2onw, and 2onv), we used in silico design to construct amyloid-like structures containing 32 β-strands in two β-sheets (defined as a 16$\times$2 array) for each of these three sequences in all of the eight symmetry classes. Assemblies were constructed so as to maximize backbone and side-chain hydrogen bonding and complementary surface packing interactions between the stacked β-sheets. We then performed MD simulations of 10 ns for the 24 structures in explicit solvent. We analyzed the final 1 ns of the MD trajectories to determine the ability of these three unrelated peptides to retain an ordered fibril structure throughout the simulation in each of the eight symmetry classes.

Additional simulations

A series of six simulations were performed to address specific questions raised by the MD simulations of the eight polymorphs described above. A 16$\times$2 fibrillar array constructed from the sequence VEALYL was simulated in symmetry classes 1 (designed) and 7 (taken from the crystal, pdb code 2omq) for comparison with SSTSAA, because the crystal structure of VEALYL has a similarly small interface between the two stacked β-sheets as seen in the crystal structure of SSTSAA (found in symmetry class 1). These two symmetry classes were selected to allow direct comparison between the behavior of VEALYL and SSTSAA, which despite having similar interfacial packing in the crystallographic forms occur in symmetry classes 7 and 1, respectively. The sequence NNQQNY was also examined in classes 1 and 5 to determine whether the additional residue present in GNNQQNY (compared with the other, six-residue peptides considered in the study) plays a role in determining the relative behavior of different polymorphs. A fibrillar assembly containing only two β-sheets (each containing 16 β-strands) taken from the crystal structure of SSTSAA class 1 was unstable during the MD, and therefore a simulation of a larger section cut from the crystalline form of SSTSAA (consisting of 4 β-sheets of 16 peptides each) was performed. Finally, in addition to the crystal structure, we performed a simulation of an in silico designed structure of SSTSAA class 1, which has a shift in register of the two stacked β-sheets compared with the crystal structure, to look for alternative stable conformations. This simulation is referred to as SSTSAA class 1*. For this sequence, we therefore performed simulations across all eight possible symmetries while also investigating the register polymorphism for symmetry class 1 (because this is the crystal structure).

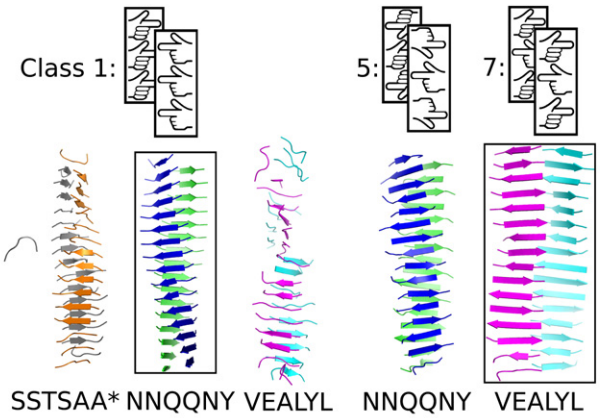
We repeated the simulations based on experimentally determined crystal structures using the CHARMM22-CMAP force field for comparison. We also repeated simulations of all of the 24 polymorphs using the AMBER03 force field. Using AMBER99, we extended the simulations of the sequences that had an OP $<$ 0.07 after 10 ns (and were therefore classified as stable) to 20 ns for GNNQQNY, SSTSAA, and GGVVIA, and one simulation of SSTSAA for which OP $>$ 0.07 after 10 ns (classified as unstable) was also extended. The results from these validation simulations are presented in Fig. S2, Fig. S3, and Fig. S4. We observed only minor differences in the degree of ordering of polymorphs after 10 ns simulation using these alternative MD force fields or when the simulations were extended to 20 ns, as discussed in the Supporting Material.

Aggregate stability is determined by the sequence and symmetry of the polymorph

Fig. 1 shows the structures after 10 ns of MD of the assemblies of the three peptides GNNQQNY, SSTSAA, and GGVVIA simulated across each of the eight symmetry classes. The results show a striking difference in the behavior of these sequences in the various polymorphic forms. Whereas GNNQQNY retains an ordered cross-β architecture characteristic of amyloid in all eight symmetry classes, SSTSAA only retains ordered structures in the antiparallel classes (classes 5–8), and in the parallel classes (classes 1–4) it shows clear evidence for disassembly during the course of the simulations. GGVVIA, by contrast, remains ordered only in polymorphs of classes 1 and 4. The final structures from the 10 ns simulations of NNQQNY and VEALYL (16 β-strands $\times$ 2 β-sheets in classes 1 and 5), the in silico engineered polymorph SSTSAA* (16 β-strands $\times$ 2 β-sheets in class 1) and the

SSTSAA crystal structure (16 β-strands $\times$ 4 β-sheets) are shown in Fig. 2. Although SSTSAA disassembles during the MD simulation in the class 1 polymorph, despite being crystallized in this form, the simulations of SSTSAA* reveal a stable polymorph, emphasizing the sensitivity to the strand register in defining fibril stability. In addition, VEALYL was also sensitive to the polymorphic form, remaining assembled in the antiparallel class 7 but not in class 1. NNQQNY was stable in both architectures, consistent with the known ability of Q/N-rich sequences to display high polymorphism. Because these simulations showed no clear difference between GNNQQNY, we did not simulate the full spectrum of polymorphs.

To quantify the different behaviors of the sequences exhibited in each polymorphic form, we calculated the degree of order remaining in these structures over the final 1 ns of the simulation using the OP (defined in Eq. 2). As shown in Fig. 3, this analysis identifies two populations across the 29 simulations performed: 1), stable aggregates with OP $<$ 0.07 that remain assembled and tightly packed after 10 ns of simulation (e.g., all eight classes of GNNQQNY in Fig. 1); and 2), unstable structures with OP $>$ 0.07 that disintegrate during the 10 ns simulation (e.g., classes 1–4 of SSTSAA in Fig. 1). Based on this separation, we classify polymorphs with an average OP $<$ 0.07 over the final 1 ns of MD as stable structures. These simulations are designed to quantify the minimum molecular-interaction strength that is necessary to maintain a polymorph in an ordered aggregated state. According to this approach, if the simulations of all polymorphs were to be continued for very long timescales, the OP would remain low for those structures classified as stable, whereas for unstable polymorphs the OP would continuously increase as the initial aggregated structure disintegrates into separate peptide monomers, as in Fig. S4.



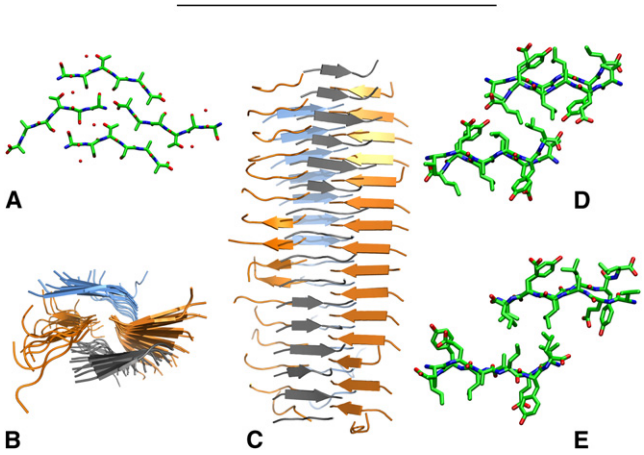


FIGURE 2 (Color available online only). Additional simulations of SSTSAA, NNQQNY, and VEALYL. The upper panel shows molecular conformations after 10 ns simulations. The lower panel shows an axial view of the start structure (*A*) and axial and transverse views of the final structure (*B* and *C*) for the simulation of a four-sheet aggregate of SSTSAA class 1, as well as starting and final axial views of the leucine-rich steric zipper of VEALYL class 7 (*D* and *E*).

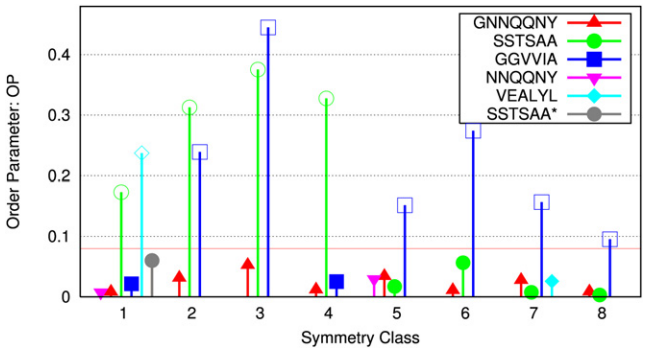


FIGURE 3 (Color available online only). Interstrand orientational order measured over the 10th ns of each simulation. Structures that are classified as stable assemblies have OP $<$ 0.07 (shown by the *red line*), and unstable structures have OP $>$ 0.07. These are denoted by solid and open symbols, respectively.

Simulations of Q/N-rich sequences

Consistent with the visual inspection of the structures of the assemblies remaining after 10 ns of MD (see Fig. 1), all assemblies of the sequence GNNQQNY and NNQQNY have OP $<$ 0.07 over the final 1 ns of the simulation regardless of the symmetry class, indicating that all polymorphs remained in ordered structures during the MD. Previous results based on detailed studies of just two classes of this sequence (classes 1 and 5) suggested an inherent plasticity of Q/N-rich peptide sequences (11) consistent with the prevalence of Q/N residues in prions (2). Our simulations provide further evidence that the presence of Q and N residues in a fibril can increase the number of stable polymorphs that can be adopted, assuming that the growth of these polymorphs is not kinetically suppressed.

Simulations to produce ordered assemblies of SSTSAA

All of the simulations that were initialized from crystallographic coordinates remained stable (OP $<$ 0.07) after 10 ns of MD (e.g., the boxed structures in Figs. 1 and 2), except for that of SSTSAA in symmetry class 1. In the crystal structures reported for GNNQQNY, it is clear that

the repeating unit consists of a pair of β-sheets stacked in a steric zipper conformation. However, this fibril-like structure is less evident in the crystal structure of SSTSAA, and simulations of a single pair of stacked β-sheets (each containing 16 β-strands) formed from this sequence have insufficient surface interactions across the steric zipper to maintain a stable protofilament structure. By contrast, simulations of a protofilament of VEALYL in class 7 (taken from the crystal structure), which has a small intersheet packing interface similar to that in the crystal structure of SSTSAA, remained ordered after 10 ns with an OP $<$ 0.07 (Figs. 2, *D* and *E*, and 3). A detailed analysis of side-chain packing in this structure revealed that VEALYL adopts a knobs-into-holes packing structure between the two β-sheets, similar to that found in the leucine zipper coiled-coil (33). It can be presumed that this packing motif is particularly stabilizing, given that the ordered VEALYL assembly remained intact during the 10 ns simulation even though it had a small hydrophobic contact area between the two β-sheets comparable to that of the unstable SSTSAA simulation.

To confirm that SSTSAA assemblies can remain stable when arranged in the experimentally determined crystalline form, we performed an additional simulation of a larger section of the crystal structure using four β-sheets of 16 peptides each, instead of just two sheets (see Fig. 2). The calculations were performed in 0.07 M NaCl solution to mimic the solution conditions in the crystallization buffer. This larger system remained stable over the 10 ns MD, as shown in Fig. 2, *B* and *C*. We also constructed an alternative structure for SSTSAA in the class 1 symmetry (labeled class 1*) by in silico design while maintaining the organization as symmetry class 1 (Fig. 2, *top*). We constructed this structure de novo to maximize the contact area between the pair of stacked β-sheets. In this case, the fibril-like structure remained stable over 10 ns of MD (OP $<$ 0.07; Fig. 3), confirming that the intersheet surface plays a role in maintaining a stable polymorph for these short peptide sequences. It also suggests that fibrillar forms may exist for this peptide that do not have close identity with the crystal even though they are in the same symmetry class.

The observed variation in stability with respect to symmetry can be explained by hydrogen bonding and steric packing

The stability of an amyloid-like assembly depends on both sequence and symmetry because these factors together determine the availability of favorable interactions between the peptide strands. To examine the role of these factors in determining fibril stability, we performed a detailed analysis of the intermolecular interactions in the last 1 ns of the simulations for each of the 16×2 fibrils of GNNQQNY, SSTSAA, and GGVVIA in all eight symmetry classes (as shown in Fig. 1), and for the 16×2 fibrils of SSTSAA*, NNQQNY, and VEALYL shown in Fig. 2.

Hydrogen-bonding interactions are inherent to β-sheet stability (34). Fig. 4 *a* shows the relationship between the number of solute-solute hydrogen bonds per donor for sequences in each symmetry class and the stability of the different polymorphs. The number of hydrogen bonds was normalized (e.g., per donor) to enable comparison between the polymorphs of different peptide sequences. The hydrogen-bond occupancies were taken as averages over the final 1 ns of each MD trajectory, and stable fibril-like structures were defined as those with OP $<$ 0.07 during the final 1 ns of the simulation, as described previously. The resulting data showed that, averaged over all of the

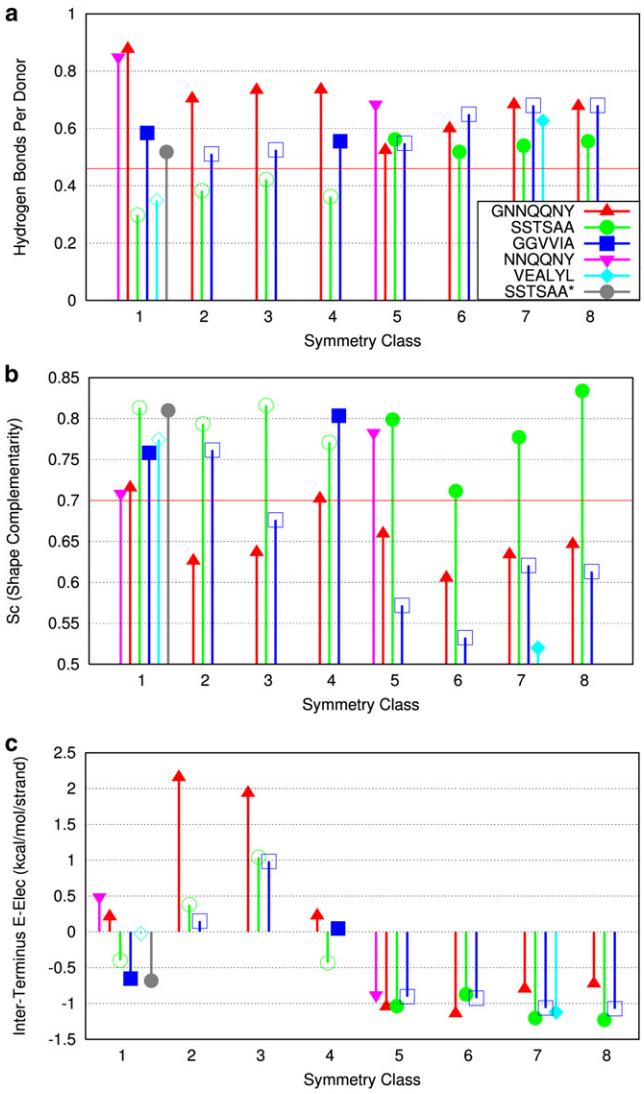


FIGURE 4 (Color available online only). (*a*) Hydrogen-bond occupancy. In each case, averages are over the final 1 ns of each simulation. Solid and open symbols indicate stable (OP $<$ 0.07) and unstable structures, respectively. (*b*) Intersheet shape complementarity S_c. The horizontal lines in *a* and *b* are provided as a guide for the separation of ordered and disordered structures. (*c*) Electrostatic potential energy between charged terminal atoms. Solid and open symbols show structures that are classified as stable or unstable, respectively.

29 sequences/polymorphs studied, the antiparallel β-sheets satisfied a greater proportion of the available backbone hydrogen-bonding interactions than the parallel sheets, as has been observed in other computational studies (35). On average, 0.39 backbone hydrogen bonds per donor were satisfied for parallel structures compared with 0.60 for antiparallel structures at any given time over the final 1 ns of MD. A more detailed discussion of the dependence of the number of hydrogen bonds per donor on the symmetry class for SSTSAA and GNNQQNY is provided in Fig. S5 and Fig. S6. Of the sequences with side chains that contain functional groups capable of forming hydrogen bonds (GNNQQNY, NNQQNY, VEALYL, and SSTSAA), structures with an average of >0.45 hydrogen bonds satisfied per polar hydrogen were stable (OP $<$ 0.07) at the end of the 10 ns simulations, as shown in Fig. 4 *a*. For the sequence containing only nonpolar side chains (GGVVIA), however, the number of satisfied hydrogen bonds does not correlate with the stability of the polymorphs (*blue bar* in Fig. 4 *a*), with all polymorphs containing >0.45 hydrogen bonds satisfied per polar hydrogen, irrespective of whether the fibril structure remained ordered during the MD simulation.

Hydrophobic packing, which is also important for amyloid formation, would be expected to play a more significant role in determining the ability of nonpolar sequences to form stable amyloid-like structures than their polar counterparts. The quality of packing at protein-protein interfaces can be measured using the shape complementarity function S_c (16,27,31,36,37), which is used here to provide a geometric measure of the quality of the packing interactions between the two β-sheets in each amyloid-like polymorph. Fig. 4 *b* shows the relationship between S_c and whether the last 1 ns of the 10 ns simulation was classified as stable (OP $<$ 0.07; *solid symbols*) or unstable (OP $>$ 0.07; *open symbols*) for the eight different polymorphs of each sequence. Fig. 4 *b* shows that for the nonpolar sequence GGVVIA, the S_c of the final structures exceeded 0.75 in the two classes that remained ordered after 10 ns of MD (classes 1 and 4). By contrast, although all of the parallel β-sheet polymorphs of the sequence SSTSAA were unstable, the S_c of these structures exceeded 0.75 throughout. We attribute this instability to the absence of sufficient hydrogen bonds and the unfavorable electrostatics interactions in this system. For the polar sequences (G)NNQQNY and SSTSAA, the assemblies that were still ordered (OP $<$ 0.07) after 10 ns of MD were those with the highest proportion of satisfied hydrogen bonds per donor, whereas for the nonpolar sequence GGVVIA, a high shape complementarity ($S_c > 0.75$) was also required.

Electrostatic interactions make a symmetry-dependent contribution to the energetics of aggregates

In addition to packing and hydrogen bonding, electrostatic forces between charged residues also affect the stability of amyloid fibrils of proteins and peptides. The peptides chosen for this study were charged at the termini primarily because the peptide structures determined crystallographically and used as input to the MD simulations have charged termini, but also to enable us to study the relationship between electrostatic stability and the symmetry of the structure. Symmetries that are composed of antiparallel β-sheets (classes 5–8) and symmetries in which the sheets are arranged antiparallel relative to each other across the steric zipper (classes 1 and 4) have the opposing charges of termini placed close together, yielding a favorable electrostatic interaction. Conversely, the parallel β-sheets that are also parallel across the steric zipper (classes 2 and 3) have unfavorable electrostatics due to the close proximity of like charges at the termini. The distances between the charged termini across the steric zipper are sufficiently similar to those within a given β-sheet for both inter- and intrasheet electrostatics to be important in determining the ability of a polymorph to maintain an ordered fibrillar structure. In parallel symmetries (classes 1–4) of GNNQQNY, for example, the average nearest-neighbor distance between charged termini within a β-sheet is ~5 Å, whereas for the antiparallel symmetries (classes 5– 8) it is between 3.8 Å and 4.9 Å, depending on the precise polymorphic form. The average distance between charged termini across the steric zipper is between 7 Å and 8 Å for symmetry classes 1 and 4 (where the electrostatics are entirely favorable), and between 11 Å and 12 Å for symmetry classes 2 and 3 (where the electrostatics are entirely unfavorable and some structural relaxation occurs to increase this distance). Fig. 4 *c* shows the electrostatic energies for each polymorphic form calculated using unit point-charges at each terminus with a dielectric constant $\varepsilon_w = 80$. This high cost in electrostatic energy for symmetries 2 and 3 may explain their comparatively rare occurrence in the structures of peptide aggregates determined to date (the exception being SNQNNF (16), which has a very large register shift between strands such that the N-termini of each sheet are actually closer to the C-termini of the adjacent sheet). Although shifts in register can modulate the interaction of symmetry and electrostatics, given the measurements presented here, it is unsurprising that the highly charged peptide SIRELEARIRELELRIG prefers classes 5 and 8 (38,39), and that KFFE, QQRQQQQQEQQ, and HQKLVFFAED also prefer antiparallel structures (40,41). Although electrostatic forces do contribute to the stabilities of the model aggregates examined here, they do not dominate. Fig. S7 shows that the van der Waals interaction energies associated with each β-strand are at least 10 times larger in magnitude than the estimate of the electrostatic contribution from the termini shown in Fig. 4 *c*. Fig. 4 *c* also shows that many of the structures with highly unfavorable electrostatic interactions, such as GNNQQNY classes 2 and 3, were stable in the MD simulations, whereas GGVVIA classes 7 and 8, which had very favorable electrostatic interactions, were not.

Minimum criterion for polymorph stability based on hydrogen bonding and surface complementarity

Fig. 5 shows the relationship among the hydrogen-bond occupancy, the shape complementarity S_c, and the OP of the 29 different sequences and polymorphs. The plot shows clustering of stable structures in the region of maximal hydrogen bonding and S_c (with the exception of VEALYL, which has a unique knobs-into-holes intersheet packing motif). All stable structures have >0.5 of possible peptide-peptide hydrogen bonds occupied at all times and S_c-values > 0.6. This observation suggests a minimum criterion that can be used to predict the stability of polymorphs of simple sequences through in silico design: If an assembly can be identified that has an S_c and hydrogen-bonding profile in the stable region, it can be expected to be a stable polymorph when subjected to long MD simulations and to form the structure in vitro or in vivo if the thermodynamically favorable structure is accessible kinetically.

Equilibrium population of polymorphs under thermodynamic control

Our systematic exploration of all eight symmetry classes for GNNQQNY, SSTSAA, and GGVVIA showed that eight, four, and two polymorphs, respectively, remain as ordered fibrillar assemblies (OP < 0.07) when subjected to 10 ns MD simulation. If we can assume that polymorph populations are determined by thermodynamics (rather than kinetics), and that differences in the enthalpy of the polymorphs are more significant than any differences in entropy, we can estimate the equilibrium proportion of each polymorph by Boltzmann weighting the relative enthalpies of the aggregates calculated from the MD force field using Eq. 1. To estimate these populations, we calculated the relative enthalpies of the defect-free structures of the stable polymorphs of GNNQQNY, SSTSAA, and GGVVIA (see Materials and Methods), and for comparison we also calculated the enthalpies of the stable polymorphs of these three sequences over the final 1 ns of the 10 ns MD simulations (Fig. S8). From this analysis, we expect more than one polymorph of these sequences to be present in any given fibril sample. Such inhomogeneity will make it particularly difficult to experimentally determine the structure of polymorphic fibrils. Moreover, we have seen that fibrillar assemblies of peptides can remain in an ordered state even when their structures are not perfectly homogeneous and contain defects, such as regions where the hydrogen-bonding interactions in the β-sheets are locally disrupted. The presence of defects in amyloid will further complicate attempts to ascribe a unique experimental structure to the fibrils of a given sequence.

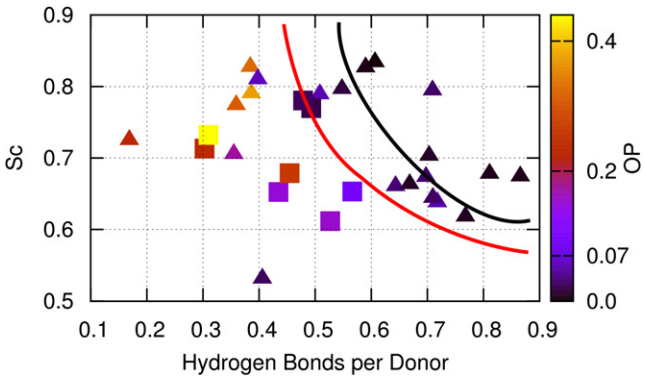


FIGURE 5 (Color available online only). Sequence-dependent importance of contributors to stability. The relationship among hydrogen bonding, S_c, and the OP of the different fibril polymorphs is illustrated. Squares indicate the nonpolar sequence GGVVIA. Triangles represent the other, more polar sequences GNNQQNY, SSTSAA, VEALYL, and NNQQNY. The black curve defines the region where all assemblies were stable. All structures to the left of the red line, aside from the outlier VEALYL (class 7), were classified as unstable.

DISCUSSION

Amyloid was defined as having a cross-β architecture by Geddes et al. (42) in 1968. Recent insights have started to reveal the remarkable array of structures that conform to the definition of a cross-β array (43), as recently reviewed by Miller et al. (44). Polymorphism can arise simply due to differences in the arrangements of the β-strands within a β-sheet, as demonstrated experimentally for the peptide SNNFGAILSS (30). It can also result from more subtle structural differences in longer polypeptide strands, such as the U-turn conformations of Aβ (13), or binding of metal ions, such as Zn^{2+} (12). Changes in pH have also been shown to shift the register of β-strands within an antiparallel β-sheet (39).

One key unresolved question is, how many polymorphs are accessible to different polypeptide sequences? Here, we used a series of 29 atomistic MD simulations to systematically survey the propensity for amyloid-forming sequences of three distinct physicochemical types to adopt different polymorphs. Specifically, the Q/N-rich sequence GNNQQNY, the polar (but non-Q/N-rich) sequence SSTSAA, and the nonpolar sequence GGVVIA were selected for detailed analysis. To be as general as possible, in the MD simulations we aimed to test the thermodynamic stability of the different peptides sequences in a given polymorphic form, and did not consider the nucleation kinetics of fibril formation (which may make a thermodynamically stable polymorph kinetically inaccessible under a particular set of environmental conditions). The results revealed an inherent ability to form polymorphs in all three sequences (GNNQQNY, SSTSAA, and GGVVIA) investigated, although the extent of polymorphism (defined by the ability to adopt one or more of the eight different symmetry classes) varied substantially for these sequences. Of importance, simulations of the Q/N-rich sequences revealed that the structures built in all eight

symmetry classes remained ordered, whereas for other sequences the stable polymorphs depended critically on the precise amino-acid sequence used.

Amyloid-fibril polymorphism is important biologically because it governs the ability of prions to exist as different strains. It can also be responsible for the heterogeneity of fibrils within the same sample, which hinders experimental efforts to determine their structures at atomic resolution. The simulations described here provide an atomistic-level, detailed explanation for the existence of polymorphs and suggest a number of design rules that can be used to predict the polymorphic potential of different short peptide sequences. First, Q/N-rich sequences are inherently polymorphic due to the flexibility and hydrogen-bonding potential of these amino-acid side chains. Second, antiparallel symmetries (classes 5–8) in general are more structurally stable than their parallel counterparts (classes 1–4) because more backbone hydrogen bonds can be formed. Third, the electrostatic interactions in charged sequences favor the antiparallel symmetries (classes 5–8), but if the structure does adopt a parallel symmetry, electrostatics will favor class 1 or 4. Finally, in comparison with hydrophilic sequences, strongly hydrophobic sequences require more favorable hydrophobic packing to remain stably assembled and overcome their inability to gain favorable stabilization energy through side-chain hydrogen bonding or electrostatics.

The fact that clear relationships, with physicochemical justifications, exist between the sequence of a peptide and the stable polymorphs it may form within the amyloid cross-β motif raises the possibility that this variable can be controlled for the development of fibrils as active biomaterials. For example, the effectiveness of decorating fibril surfaces with chromophores (45) or functionalized chemical groups capable of forming intersheet cross-links will be determined by the polymorphic form. Many of the potential nanotechnological and medical applications for amyloid fibrils will demand a unique, self-assembled structure rather than a heterogeneous mix of polymorphs. Therefore, the ability to predict and control amyloid polymorphs is likely to become increasingly important if they are to be engineered for technological use. A number of algorithms have been developed to predict the absolute aggregation propensity of a given peptide sequence (46–53), but the ability of fibril-forming sequences to adopt different polymorphic forms has received less attention. The simulations presented here show that hydrogen bonding, surface packing (S_c), and electrostatics are all important factors in determining the stability of a sequence in a particular symmetry arrangement. In addition, aromatic stacking interactions (which only appear as terminal residues in the sequences studied here) may also preferentially stabilize particular polymorphic forms (54). Because polymorphism results from a balance of these different factors, it is strongly dependent on environmental conditions such the pH and ionic strength, both of which will affect the electrostatics, or the addition of nonaqueous solvents, which may influence hydrogen bonding or intersheet hydrophobic interactions (44). With an improved understanding of these factors, we may be able to design sequences and environmental conditions that are appropriate for controlling the polymorphic form of fibrils, using in silico methods as a design tool.

SUPPORTING MATERIAL

Eight figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(11)00325-0.

We thank Zwe Ndlovu for his assistance with the manuscript.

This work was funded by the Engineering and Physical Sciences Research Council through the award of a doctoral training studentship to J.T.B. Computational resources were supplied by the UK National Grid Service.

REFERENCES

1. Fändrich, M. 2007. On the structural definition of amyloid fibrils and other polypeptide aggregates. *Cell. Mol. Life Sci.* 64:2066–2078.
2. Chiti, F., and C. M. Dobson. 2006. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* 75:333–366.
3. Fowler, D. M., A. V. Koulov, …, J. W. Kelly. 2007. Functional amyloid—from bacteria to humans. *Trends Biochem. Sci.* 32:217–224.
4. Glover, J. R., A. S. Kowal, …, S. Lindquist. 1997. Self-seeded fibers formed by Sup35, the protein determinant of [PSI+], a heritable prion-like factor of *S. cerevisiae*. *Cell.* 89:811–819.
5. Serpell, L. C. 2000. Alzheimer's amyloid fibrils: structure and assembly. *Biochim. Biophys. Acta.* 1502:16–30.
6. Antzutkin, O. N., R. D. Leapman, …, R. Tycko. 2002. Supramolecular structural constraints on Alzheimer's β-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. *Biochemistry.* 41:15436–15450.
7. Toyama, B. H., M. J. Kelly, …, J. S. Weissman. 2007. The structural basis of yeast prion strain variants. *Nature.* 449:233–237.
8. Jiménez, J. L., E. J. Nettleton, …, H. R. Saibil. 2002. The protofilament structure of insulin amyloid fibrils. *Proc. Natl. Acad. Sci. USA.* 99:9196–9201.
9. Krishnan, R., and S. L. Lindquist. 2005. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature.* 435:765–772.
10. Castilla, J., D. Gonzalez-Romero, …, C. Soto. 2008. Crossing the species barrier by PrP(Sc) replication in vitro generates unique infectious prions. *Cell.* 134:757–768.
11. Berryman, J. T., S. E. Radford, and S. A. Harris. 2009. Thermodynamic description of polymorphism in Q- and N-rich peptide aggregates revealed by atomistic simulation. *Biophys. J.* 97:1–11.
12. Miller, Y., B. Ma, and R. Nussinov. 2010. Zinc ions promote Alzheimer Aβ aggregation via population shift of polymorphic states. *Proc. Natl. Acad. Sci. USA.* 107:9490–9495.
13. Miller, Y., B. Ma, and R. Nussinov. 2009. Polymorphism of Alzheimer's Aβ17-42 (p3) oligomers: the importance of the turn location and its conformation. *Biophys. J.* 97:1168–1177.
14. Miller, Y., B. Ma, …, R. Nussinov. 2010. Hollow core of Alzheimer's Aβ42 amyloid observed by cryoEM is relevant at physiological pH. *Proc. Natl. Acad. Sci. USA.* 107:14128–14133.
15. Wei, G. H., A. I. Jewett, and J. E. Shea. 2010. Structural diversity of dimers of the Alzheimer amyloid-β(25-35) peptide and polymorphism of the resulting fibrils. *Phys. Chem. Chem. Phys.* 12:3622–3629.
16. Sawaya, M. R., S. Sambashivan, …, D. Eisenberg. 2007. Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature.* 447:453–457.

17. Wiltzius, J. J., M. Landau, …, D. Eisenberg. 2009. Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* 16:973–978.

18. Macke, T. J., and D. A. Case. 1997. Modeling unusual nucleic acid structures. *In* Molecular Modeling of Nucleic Acids. N. B. Leontes and J. Santa Lucia, Jr., editors. American Chemical Society, Washington, DC. 379–393.

19. Hovmöller, S., T. Zhou, and T. Ohlson. 2002. Conformations of amino acids in proteins. *Acta Crystallogr. D Biol. Crystallogr.* 58:768–776.

20. Case, D. A., T. E. Cheatham, 3rd, …, R. J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.

21. Cornell, W. D., P. Cieplak, …, P. A. Kollman. 1995. A 2nd generation force-field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.

22. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

23. MacKerell, A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

24. Phillips, J. C., R. Braun, …, K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.

25. Tsui, V., and D.A. Case. 2000-2001. Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers 56:275–291.

26. DeLano, W. L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA.

27. Lawrence, M. C., and P. M. Colman. 1993. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* 234:946–950.

28. Collaborative Computational Project, Number 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50:760–763.

29. Collinge, J., and A. R. Clarke. 2007. A general model of prion strains and their pathogenicity. *Science.* 318:930–936.

30. Madine, J., E. Jack, …, D. A. Middleton. 2008. Structural insights into the polymorphism of amyloid-like fibrils formed by region 20-29 of amylin revealed by solid-state NMR and X-ray fiber diffraction. *J. Am. Chem. Soc.* 130:14990–15001.

31. Nelson, R., M. R. Sawaya, …, D. Eisenberg. 2005. Structure of the cross-β spine of amyloid-like fibrils. *Nature.* 435:773–778.

32. Marshall, K. E., M. R. Hicks, …, L. C. Serpell. 2010. Characterizing the assembly of the Sup35 yeast prion fragment, GNNQQNY: structural changes accompany a fiber-to-crystal switch. *Biophys. J.* 98:330–338.

33. Landschulz, W. H., P. F. Johnson, and S. L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science.* 240:1759–1764.

34. Deechongkit, S., H. Nguyen, …, J. W. Kelly. 2004. Context-dependent contributions of backbone hydrogen bonding to β-sheet folding energetics. *Nature.* 430:101–105.

35. Chou, K. C., M. Pottle, …, H. A. Scheraga. 1982. Structure of β-sheets. Origin of the right-handed twist and of the increased stability of antiparallel over parallel sheets. *J. Mol. Biol.* 162:89–112.

36. Thompson, M. J., S. A. Sievers, …, D. Eisenberg. 2006. The 3D profile method for identifying fibril-forming segments of proteins. *Proc. Natl. Acad. Sci. USA.* 103:4074–4078.

37. Zheng, J., B. Ma, and R. Nussinov. 2006. Consensus features in amyloid fibrils: sheet-sheet recognition via a (polar or nonpolar) zipper structure. *Phys. Biol.* 3:1–4.

38. Steinmetz, M. O., Z. Gattin, …, R. A. Kammerer. 2008. Atomic models of de novo designed cc β-Met amyloid-like fibrils. *J. Mol. Biol.* 376:898–912.

39. Verel, R., I. T. Tomka, …, B. H. Meier. 2008. Polymorphism in an amyloid-like fibril-forming model peptide. *Angew. Chem. Int. Ed. Engl.* 47:5842–5845.

40. Tjernberg, L., W. Hosia, …, J. Johansson. 2002. Charge attraction and β propensity are necessary for amyloid fibril formation from tetrapeptides. *J. Biol. Chem.* 277:43243–43246.

41. Fishwick, C. W. G., A. J. Beevers, …, N. Boden. 2003. Structures of helical β-tapes and twisted ribbons: the role of side-chain interactions on twist and bend behavior. *Nano Lett.* 3:1475–1479.

42. Geddes, A. J., K. D. Parker, …, E. Beighton. 1968. "Cross-β" conformation in proteins. *J. Mol. Biol.* 32:343–358.

43. Fändrich, M., J. Meinhardt, and N. Grigorieff. 2009. Structural polymorphism of Alzheimer Aβ and other amyloid fibrils. *Prion.* 3:89–93.

44. Miller, Y., B. Ma, and R. Nussinov. 2010. Polymorphism in Alzheimer Aβ amyloid organization reflects conformational selection in a rugged energy landscape. *Chem. Rev.* 110:4820–4838.

45. Deng, W., A. Cao, and L. Lai. 2008. Distinguishing the cross-β spine arrangements in amyloid fibrils using FRET analysis. *Protein Sci.* 17:1102–1105.

46. Conchillo-Solé, O., N. S. de Groot, …, S. Ventura. 2007. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics.* 8:65–81.

47. Trovato, A., F. Seno, and S. C. Tosatto. 2007. The PASTA server for protein aggregation prediction. *Protein Eng. Des. Sel.* 20:521–523.

48. Tartaglia, G. G., and M. Vendruscolo. 2008. The Zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* 37:1395–1401.

49. Tartaglia, G. G., A. Cavalli, …, A. Caflisch. 2005. Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. *Protein Sci.* 14:2723–2734.

50. Bryan, Jr., A. W., M. Menke, …, B. Berger. 2009. BETASCAN: probable β-amyloids identified by pairwise probabilistic analysis. *PLOS Comput. Biol.* 5:e1000333.

51. Maurer-Stroh, S., M. Debulpaep, …, F. Rousseau. 2010. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods.* 7:237–242.

52. Fernandez-Escamilla, A. M., F. Rousseau, …, L. Serrano. 2004. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* 22:1302–1306.

53. Goldschmidt, L., P. K. Teng, …, D. Eisenberg. 2010. Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. USA.* 107:3487–3492.

54. Goux, W. J., L. Kopplin, …, D. A. Kirschner. 2004. The formation of straight and twisted filaments from short τ peptides. *J. Biol. Chem.* 279:26868–26875.