The route from the folded to the amyloid state: exploring the potential energy surface of a drug-like miniprotein

Nóra Taricska[1], Dániel Horváth[1], Dóra K. Menyhárd[1], Hanna Ákontz-Kiss[1], Masahiro Noji[2], Masatomo So[2], Yuji Goto [2], Toshimichi Fujiwara[2] and András Perczel[1,]*

[1]Laboratory of Structural Chemistry and Biology & MTA-ELTE Protein Modeling Research Group at the Institute of Chemistry, Eötvös Loránd University, 1117, Pázmány Péter st. 1/A, Budapest, Hungary, [2]Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
*E-mail: perczel.andras@ttk.elte.hu

Abstract:

The amyloid formation of the folded segment of a variant of Exenatide (a marketed drug for Type-2 *Diabetes Mellitus*) was studied by ECD and NMR. We found that the optimum temperature for E5 amyloidosis coincides with body temperature and requires well below physiological salt concentration. Decomposition of the ECD spectra and its barycentric representation on the folded-unfolded-amyloid potential energy surface allowed us to monitor the full range of molecular transformation of amyloidogenesis. We identified points of no return (*e.g. T*=37°C, *pH*=4.1, c_{E5}=250µM, c_{NaCl}=50mM, *t*>4-6 h) which will inevitably gravitate into the amyloid-state. The strong B-type FUV-ECD spectra and an unexpectedly strong NUV-ECD signal ($\Theta_{\sim 275\text{-}285nm}$) indicate that the amyloid phase of E5 is built from monomers of *quasi*-elongated backbone structure (φ~-145°, ψ~+145°) with strong interstrand Tyr↔Trp interaction. Misfolded intermediers and the buildup of "toxic" early-stage oligomers leading to self-association were identified and monitored as function of time. Results indicate that the amyloid transition is triggered by subtle misfolding of the α-helix exposing aromatic and hydrophobic side-chains that may provide the first centers for an intermolecular reorganization. These initial clusters provide the spatial closeness and sufficient time for a transition to the β-structured amyloid nucleus thus the process follows a nucleated growth mechanism.

Graphical Abstract:
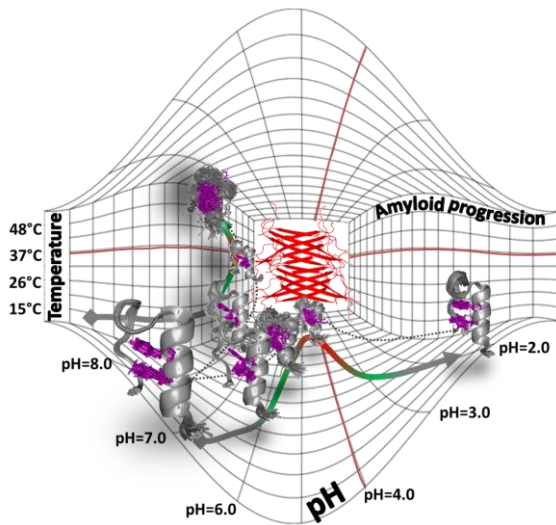
Abbreviations:

AMBER, Assisted Model Building with Energy Refinement, Amy-state, amyloid state; APP, Amyloid precursor protein; ARIA, Ambiguous Restraints for Iterative Assignment; Aβ(1-42), beta-amyloid peptide (1-42); Aβ2M, amyloidogenic β2-microglobulin; CNS, central nervous system; CryoEM, Cryo-electron microscopy; DLS, Dynamic light scattering; ECD, Electronic Circular Dichroism; Ex-4, Exendin-4; F-state, folded state; FT-IR, Fourier-transform infrared spectroscopy; FUV-ECD, far ultraviolet electronic circular dichroism: 185-260 nm; GLP-1R, Glucagon-like peptide-1 receptor; GROMACS, GROningen MAchine for Chemical Simulations, Hanabi, HANdai Amyloid Burst Inducer; HSQC, Heteronuclear single quantum coherence; I-state, intermediate state; MD simulation, Molecular Dynamics; NMR, Nuclear magnetic resonance; NOE, Nuclear Overhauser effect; NOESY, Nuclear Overhauser Effect SpectroscopY; NUV-ECD, near ultraviolet electronic circular dichroism: 260-325 nm; Pro-IAPP, pro-islet amyloid polypeptide RMSD, *backbone all atoms root-mean standard deviation in Ångstroms*, s/n, signal to noise ratio; SAXS, Small-angle X-ray scattering; SS-bond, disulfide bond; ssNMR, Solid-state nuclear magnetic resonance; T2DM, Type-2 *Diabetes Mellitus*; TOCSY, Total Correlated SpectroscopY, Trp-cage, tryptophan- cage; U-state, unfolded state;

Introduction:

Aggregation of proteins and peptide segments into amyloid fibrils have been studied intensively over the past decades since the process was shown to be associated with, or even trigger (1-3) such illnesses as Alzheimer's disease, type II *Diabetes Mellitus*, *Rheumatoid arthritis*, or *Haemodialysis ass. amyloidosis* (4). From the pioneering work on lysozyme (5) and Aβ(1-42) (6-8), the amyloid state of several misfolded proteins (*e.g.* β2-microglobulin (9), crystalline (10), tau protein (11-13), the glucagon peptide hormone (14) and insulin (15) among others) were characterized. The general topology of such aggregates consists of protein segments adopting an extended backbone, interacting through β-edges. The association between the β-sheets thus formed is compact and specific, in most cases it excludes water molecules, leading to the formation of tightly stacked, "dry-zipper" nanostructures (16-18). The state-of-the-art TEM, SAXS, cryo-EM and ssNMR techniques now allow the full characterization of the aggregated end-state (19-20), however, much less is known about the specific molecular species that evolve during the process, especially in the early stages which concern the formation of still soluble but oligomeric assemblies that are the most toxic (1-3, 21-22) and also represent the stage where amyloidosis can still be reversed (23).

The progress of self-association can often be followed by ThT fluorescence and DLS – best in combination (24) - reporting the accumulation of cross-β-backbone (above a minimum size of approximately 10 nm or 4-6 aligned strands (25)) and the size distribution of the species present in the solution, respectively. However, to gain atomistic detail, molecular spectroscopies need to be applied such as CD, IR or NMR (26, 27). In fact, CD spectroscopy can be used to monitor a full range of molecular transformation accompanying amyloidogenesis if the secondary structure contents of the folded, intermediate and amyloid states are distinct. There are notable examples that satisfy that description, namely helical peptide hormones like amylin or glucagon and a considerable number of peptide therapeutics (28) and since over 60% of all protein-protein interfaces - typical targets of drug design – also constitute helices (29), their number will most likely just increase.

Here we present the amyloid formation pathway of a variant of Exenatide, a marketed drug for Type-2 *Diabetes Mellitus* (30) that also contains a well-folded α-helix. We have discovered that this 25 residue long segment (E5: EEEAVRLYIQWLKEGGPSSGRPPPS) (**Figure 1**), comprising of the entire interface needed for GLP-1 receptor binding (31), can be turned into amyloid in a controlled, fully reproducible and tunable manner within a large range of protein concentrations (80μM < $c_{prot.}$< 800μM) at physiologically relevant temperatures. Therefore, understanding the molecular details of the amyloidosis of E5 and mapping its conditions is highly relevant to any optimization efforts targeting Exenatide. In addition, E5 is an ideal model to study the amyloid transition of folded proteins and helix-containing peptides. Beside its helical stretch, E5 contains a β-turn, a polyproline-II helix and a hydrophobic center with a buried Trp, thus has a protein like build-up and also folds quite similarly to a typical globular protein (32). As E5 is small, both its chemical synthesis on a resin and bacterial expression in a fusion system is straightforward (33). Because its folded state is partially helical, its transition toward the amyloid phase results in a significant change in secondary structure content, easy to monitor by CD spectroscopy. Furthermore, E5 has an interacting Trp/Tyr residue pair within its hydrophobic core, enabling to track folding and refolding of the protein by NUV- beside FUV-CD. We also found the process to be quenchable; by dropping the temperature the amyloid formation can be suspended at any time

and restarted by a subsequent temperature rise. Also, the moderate size of E5 allows detailed structure characterization both by NMR and modeling techniques.
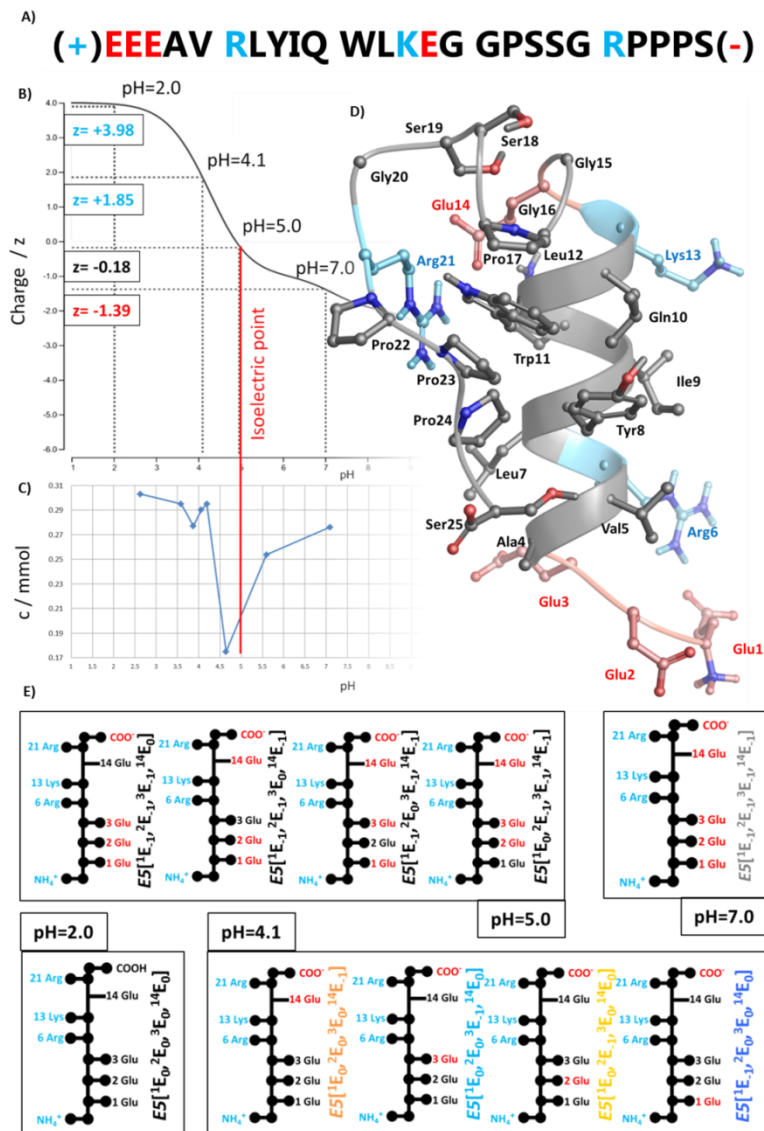

Figure 1. A) *The primary sequence of E5 with residues colored by their charges of pH=7: negatively charged amino acid, positively charged amino acid and neutral amino acid are highlighted with red, blue and dark grey, respectively.* **B)** *The overall charge (z) of E5 as a function of the pH calculated by Prot pi (www.protpi.ch)* **C)** *The solubility of E5 as a function of the pH (at 25°C) shows that near the isoelectric point reversible precipitation occurs.* **D)** *The folded structure of E5 (4°C & pH=7) determined by NMR, with its 4 basic groups (highlighted blue), and 5 acidic groups (highlighted red) with the following pKₐ values: $Ser_{C\text{-}terminal}$: pKa:2.2, $Glu_{sidechain}$:pKa:4.3, $Glu_{N\text{-}terminal}$: pKa:9.1, $Lys_{sidechain}$: pKa:10.8 and $Arg_{sidechain}$: pKa:12.5.* **E)** *The major microstates with their labeling are depicted schematically at 4 different pHs of interest, with side-chain colored by charge: neutral-black, negative-red and positive-blue, respectively.*

We probed various regions of the $f(T, pH, c_{prot}, c_{ion}, t)$ potential energy hypersurface of E5 by acquiring quantitative NUV- and FUV-ECD chiroptical information and NMR data complemented by MD simulations to pinpoint the reaction path that leads from the fully folded- to the amyloid-state. Based on these results we were able to propose a mechanism that though resembles that of well-folded proteins' but relies on special features of the miniprotein. The methodology presented here gains significance as the amyloid state of E5 is ThT silent and thus presents an approach for dealing with such cases as well.

Results and Discussion:

We have shown previously (Farkas Eur J Org Chem 2013) that high concentrations (c~10-30 mM) of E5 trigger self-association, accompanied by an α-helix to antiparallel β-sheet structural transition. Here we set out the way to identify optimum conditions of aggregation structurally characterizing key states along this route at a physiologically more relevant concentration range using NMR, NUV- and FUV-ECD data to pinpoint misfolded structures of the reaction path, and electron-microscopy to confirm the emerging amyloid fibrils.
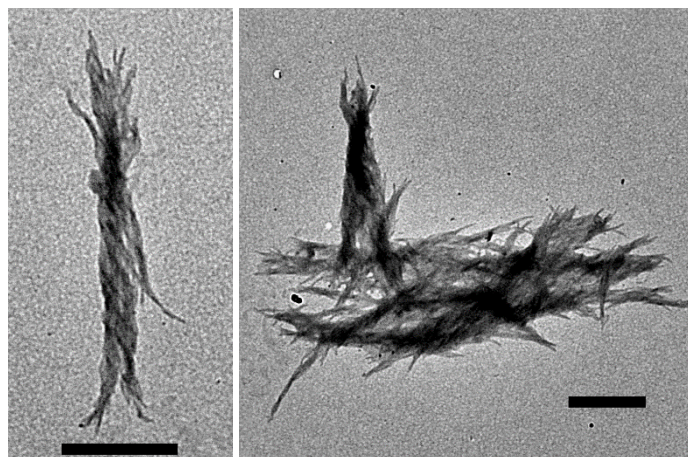

Figure 2. TEM images of E5 amyloid fibrils. Typical twisted amyloid fibrils were observed. Most fibrils were bundled and formed supercoiled structure. Scale bars indicate 500 nm.

To locate roughly and effectively those conditions that enable the amyloid formation, the HANABI system (Umemoto et al., 2014, JBC, 289, 27290-27299) was applied. The 96 well plates were set up with the following boundaries: 2< pH< 7, 80 μM< c_{E5}< 800 μM and 0 mM< c_{NaCl}< 100 mM at T= 37 °C. ECD measurements were carried out after 90 h, with sonication cycle turned „on" for 1 min and repeated every 10 minutes (power level: 700 W and frequency: 25 kHz). FUV-ECD measurements showed that amyloid formation is more effective in the presence than in the absence of salt (c_{NaCl}=100 mM), with the preferred protein concentration ranging from 80 to 160 μM, while the optimal pH is near 4.0. The amyloid thus produced showed a typical twisted fibril structure (**Figure 2**). The formation of these fibrils could not be followed by ThT fluorescence and thus ECD-spectra were recorded for each of the 96 wells. Based on the results additional experiments were designed to identify parameters of amyloid formation separately, to pinpoint their significance and specific molecular consequences.

Optimum conditions located for E5 amyloid formation: the temperature-scan

NMR (2D Homonuclear) measurements were carried out to determine the structure of E5 at each of the following temperatures: 4, 15, 26, 37 and 48°C (@ pH= 7, c_{prot}= ~0.8mM, c_{ion}= ~0 mM) leading to the primary conclusion that as temperature increases partial unfolding occurs without amyloid formation. We found that though NMR-resonance frequencies shift with rising T, line broadening only takes place above 37°C, indicating considerable unfolding only at 48°C (**SFigure 1**). The unsynchronized local backbone fluctuation of the folded F-state enhances as T increases. Furthermore, we have determined the most T-sensitive residues of E5 for which the presence of hidden intermediate(s) (I-state(s)) (Rovo et.al. ChemEur J 2013, Rovo et al. Bichemistry 2014) was revealed. The thermal unfolding of the protein backbone shows a **non-linear** T-dependence for Leu[7], Ile[9], and Lys[13] and to a lesser extent for Ala[4], Val[5] and Tyr[8], indicating an enhanced presence of transient conformers at higher temperatures for the inner helix of E5. (**Figure 1**, **SFigure 2**). 3D-structure elucidations were also completed by acquiring a large number of NOE distance-restrains (**Figure 3** and **STable 1**) at all temperatures of interest. Although the total number of restrains drops as T increases, 666 (4°C) $\rightarrow$ 221 (48°C), the latter number of NOEs are still sufficient to establish the overall 3D-fold of E5 even at 48°C (especially since 16 out of 221 are key long-range restraints) (**Figure 3/E** and **STable 1**). Inter-residue NOEs associated with Y[8] and W[11] (**SFigure 3**) enabled us not only to determine the overall molecular scaffold but also the relative orientation of the two aromatic side chains. At 37°C, 458 distance restraints in total, among which 40 long range ones (i$\rightarrow$(i+<5)) were assigned, enables to get a single time-average and compact 3D-structure for E5. While the central α-helix is tightly folded at this temperature (RMSD of the backbone heavy atoms is 0.64 Å within the 50 best-fit structures), (**STable 1**) as T increases the unfolding of the -P[17]SSGRP[22]- segment was detected. This segment is the least structured bit within the folded scaffold even at 4°C, where only sequential NOEs could be measured **(Figure 3/C)**. Nevertheless, the large number of NOEs associated with Y[8] and W[11] residues and the -P[22]PP[24]- segment ensures their concerted motion, signaling that the hydrophobic core remains well-structured at 37 °C and below (**SFigure 3**).
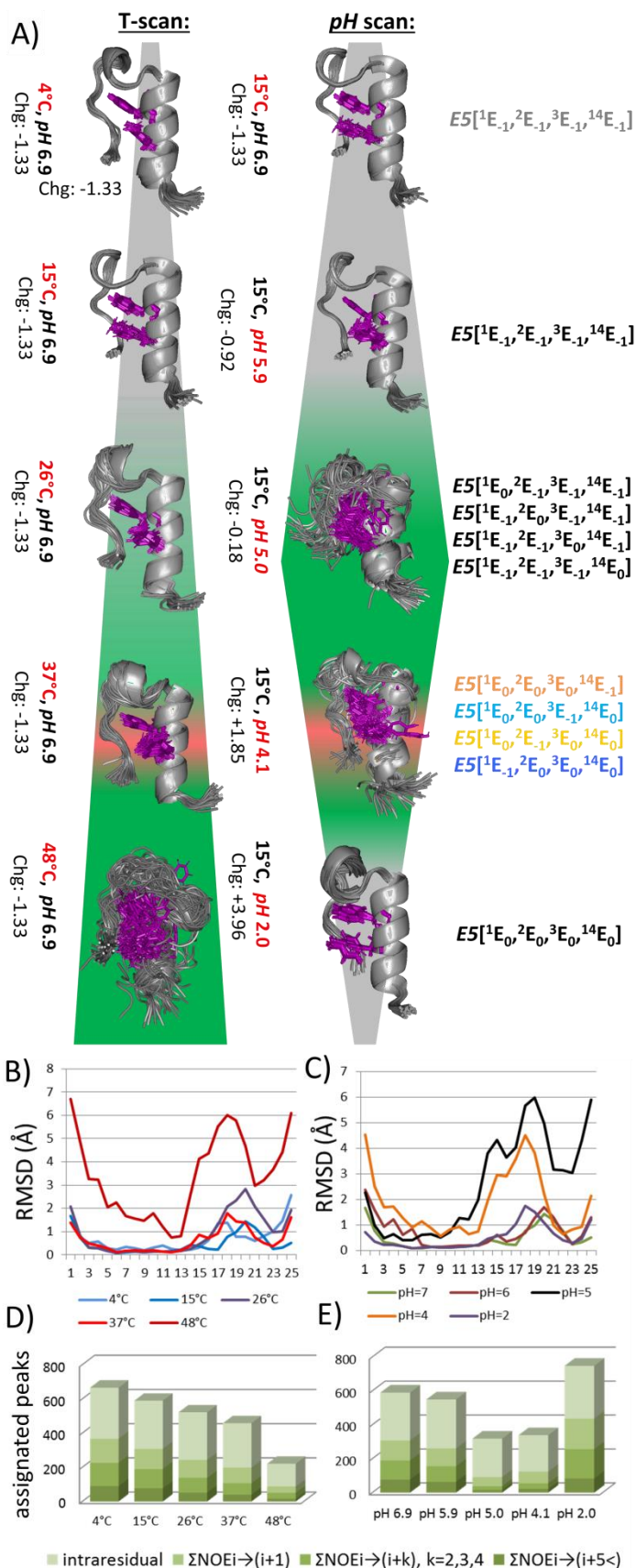
Figure 3 A) *Superimposed best 50 NMR structures of E5 at different pHs and temperatures (conditions and the net charges explicitly shown for each structure ensemble). The background color indicates the fold compactness: green - unfolded, grey - folded, red – misfolded leading to amyloid formation. Aromatic side chains of the key Y^8 and W^{11} are highlighted magenta. The basic (N-term., 6R, 13K, 21R) moieties are protonated, the C-terminus (25S) deprotonated, while the protonation microstates of Glu residues are shown at the right side at each pH. All atoms backbone RMSD (Å) as a function of the primary sequence of E5 at **B)** different temperatures (°C) and **C)** different pHs are plotted. The total number of NOEs assigned for E5: $\Sigma NOE^{i \to (i+1)}$, $\Sigma NOE^{i \to (i+k)}$ where k= 2,3,4, $\Sigma NOE^{i \to (i+5<)}$ are reported as **D)** the temperature was scanned at pH=6.9 and **E)** as pH scanned at T=15°C.*

Due to exchange phenomena, NMR is unable to provide precise structural information at high *T*, therefore complementary FUV-ECD measurements were carried out to follow the transition of the backbone fold up to 85 °C. The recorded spectra were deconvoluted using the CCA+ protocol as mixtures of folded (F-) and unfolded (U-) forms (ref Perczel 1991 ProtEng, Perczel 1992 ProtStructfunc Gen, Jakli 2009, JPeptSci). At lower temperatures, in line with the NMR data, the folded fraction is dominant (F content goes from 100 to 80% as T increases from 5°C to 35 °C), a 50-50% mixture is reached near 60 °C, while at 85°C the spectrum indicates a 70% unfolded content (**SFigure 4**).

7

The pH-scan

2D-NOESY driven 3D-structure elucidation was completed (c_{E5}~ 500-1500 µM, c_{NaCl}< 10 mM) at various *pH*s: 6.9, 5.9, 5.0, 4.1, 2.0 at *T*= 15°C **(Figure 3/B).** (The poorer signal to noise ratio required longer measurement time at *pH*= 5.0 and 4.1.) The solubility of E5 drops close to its isoelectric point (pH=4.8, *Figure* **1**) where unspecific and reversible precipitation was observed. As side chain protonation pattern varies with *pH*, H-bonds and other weak interactions change as well. E5 contains a protonated *N*-terminal (pK$_a$= 9.1) plus 3 basic residues, Arg^6 (pK$_a$= 12.5), Lys^{13} (pK$_a$= 10.8) and Arg^{21} (pK$_a$= 12.5), with an acidic *C*-terminal (pK$_a$= 2.2) and 4 acidic glutamines (Glu^1, Glu^2, Glu^3 and Glu^{14}) (pK$_a$~ 4.25). (The listed pK$_a$-s are nominal values that strongly depend on backbone conformation). The net charge of E5 is predicted to be positive at pHs smaller than 5 (*Figure* **1**). At *pH*=7.0 two salt bridges may contribute to the stabilization the 3D-fold, those of Glu^1(-)$\leftrightarrow$Arg^6(+) and Glu^{14}(-)$\leftrightarrow$Arg^{21}(+). But as *pH* decreases, Glu(s) get partly (or completely) protonated and thus salt bridges weaken and 3D-fold compactness loosens. Accordingly, at *pH*=6.9, 5.9 and 5.0 the measured 3D-structures of E5 are similar **(Figure 3)**, although conformational heterogeneity increases considerably (demonstrated by the reduction of the total number of assigned restrains from 666 to 319 (**Figure 3/F** and **STable 1**)). Thus, the Trp-cage fold holds, though backbone heavy atom RMSD of the 50 best structures increases significantly: RMSD $^{pH=6.9,T=15°C}$= 0.73 Å → RMSD $^{pH=5.0,T=15°C}$= 2.56 Å (**STable 1**). We found that the total number of NOESY cross-peaks between R^{21} and W^{11} residues is a reliable measure of the Trp-cage fold compactness (**SFigure 3**): 11 and 10 such peaks were assigned at *pH*=6.9 and 2.0, but only 3 at 5>*pH*>4 (*T*=15°C). Moreover, RMSD changes show that the α-helix becomes partly unfolded, as the pH gets closer to 4.1 (**Figure 3/D**). The structure loosening effect of the *pH* drop is the most pronounced between residue 13 and 21: the 3$_{10}$-helix (-G^{15}GPSSG^{20}-) tends to unwind, exposing both the Y^8 and W^{11} core residues to external water molecules (**Figure 3**). Interestingly enough, moving beyond the isoelectric point, ordering of the ensembles takes place and the original F-state reappeared at *pH*=2. At *pH*=6.9 77 long-range NOE$^{i\rightarrow(i+5)}$ restraints were assigned, while at pH=2.0 in total 84 NOEs$^{i\rightarrow(i+5)}$ were found. This is rather unexpected since the overall charge, as well as the local charge distribution of E5 is indeed different at the above two pHs. Meanwhile, in between, at pH= 4.9 only 22 NOEs of this kind were recorded (**Figure 3/B**). In conclusion, the pH-scan shows that the basic topological features of the Trp-cage fold of E5 (and other analogs K.A.Graham et al. 2018) are preserved at pH= 7.0 and 2.0, but weakened near the isoelectric point, where I-states of full refolding potential or of amyloidogenic misfolding capacity could be present simultaneously (**SFigure 5**).

The effect of stirring

At pH=4.1 and T=37 °C in the absence of stirring or sonication the gradual decay of the far-UV ECD spectrum intensities of E5 was detected, indicative of self-association leading to the loss of monomeric form or weakly bound low molecular weight associates in the solution (**SFigure 6**). Stirring however greatly speeds up visible precipitation, aiding both the nucleation and fibril maturation stages of amyloidogenesis. The former one is facilitated by the increasing number of collisions occurring between monomers and/or small oligomeric clusters and the later one by removing parts from the longer fibrils and thus initiating new nucleation sites (C-C. Lee Biophys. J. 2007, Serio, T. R. Science 2000).
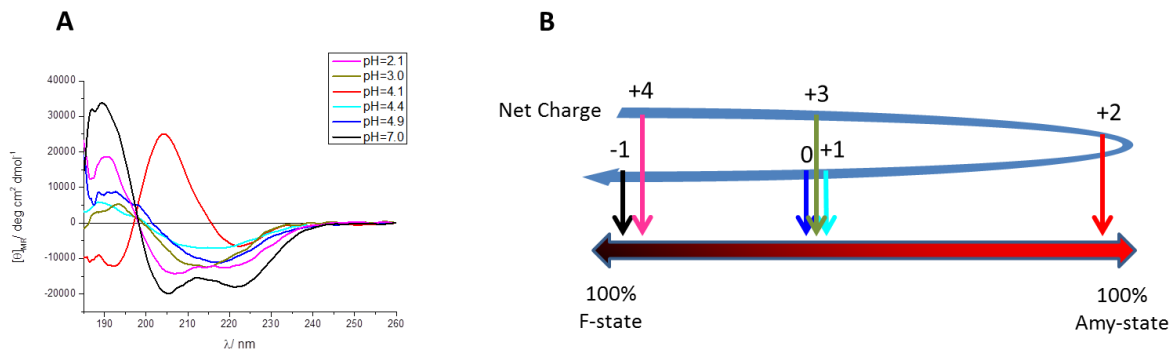

Figure 4: A) *FUV-ECD spectra of E5 recorded after 51 h of incubation time at a constant stirring (cE5=250 μM, cNaCl= 50 mM, T=37 °C). If ΣQE5pH=4.1= +2, then a B-type ECD is detected, which is indicative of a β-sheet backbone structure. However, if pH=7.0 (ΣQE5pH=7.0=-1) or pH=2.1 (ΣQE5pH=2.1=+4.0) E5 remains folded (predominantly α-helical **B)** Spectral deconvolution enables to determine the approximate [F] and [Amyloid] ratios as function of the net molecular charge: ΣQE5.*

Stirring the solution of E5 for 51h (pH=4.1) resulted in strong B-type FUV-ECD spectrum indicating a predominantly β-pleated sheet backbone structure (SREERAMA 1999, Micsonai 2015, PNAS). When setting the pH at 4.9, 4.4 and 3.0 "mixed" ECD spectra were recorded (B/C-type) even after 51 h (**Figure 4**) signaling that in solution both helical and β-stranded structures are present, amyloid formation is less complete. Furthermore, at pH=7 and pH=2 pure C-type ECD spectra were measured (even when stirring) indicating that only α-helical backbone structures are in solution and thus, these pHs prevent amyloid formation. To quantify the extent of the transition from the folded toward the amyloid-like phase, deconvolution of a large collection ECD spectra were carried out (including E5 and E0 *T*-dependent ECD curves). In this way we obtained 3 pure component spectra, which turned out to be the following: a C-type for α-helical conformation, a U-type corresponding to the unstructured and a B-type signaling β-stranded backbone conformation. ~~For E5 these 3 pure ECD curve types stand for *i*) the folded, *ii*) for the unfolded and *iii*) the amyloid states, respectively.~~ More interestingly, a similar NUV-ECD spectrum analysis shows (see below) that this transition is more complex than a simple α- to β-backbone conformational shift. Complemented with dynamic light scattering measurement data we conclude that these NUV-ECD spectral changes are associated with a gradual amyloid formation.

The overall path from the F- to the amyloid state is reported using a barycentric coordinate system (**Figure 4** and **5**) in which the gradual maturation of the amyloid as a function of the time is visualized. At any point along the route, the ratio of the folded, unfolded and amyloid-states can be calculated from the deconvoluted spectral properties. As an example the "mixed state" of [$p_F(t=5h)=0.26$, $p_U(t=5h)=0.50$, $p_{Amy}(t=5h)=0.24$] or simply (0.26, 0.5, 0.24) is shown on Figure 5 (see the "yellow dot" of **Figure 5/D**).

Using this mapping technique three phases of the amyloid formation of E5 could be differentiated as *i*) initially the path runs parallel to the folded→unfolded axis, with no or marginal contribution of the amyloid state, corresponding to a misfolding phase, with the gradual accumulation of the unfolded/misfolded forms. *ii*) During the second phase the route turns and runs parallel to the folded → amyloid axis, corresponding to the nucleation phase. During this phase a critical concentration of misfolded structures is reached: F-state diminishes while Amy-states start to accumulate, while the misfolded content remains nearly constant. *iii*) The third phase runs parallel to the unfolded→amyloid axis, where no further reversible unfolding takes place: the misfolded/unfolded structures get trapped by the growing amyloids: the elongation phase of the E5 amyloidosis.
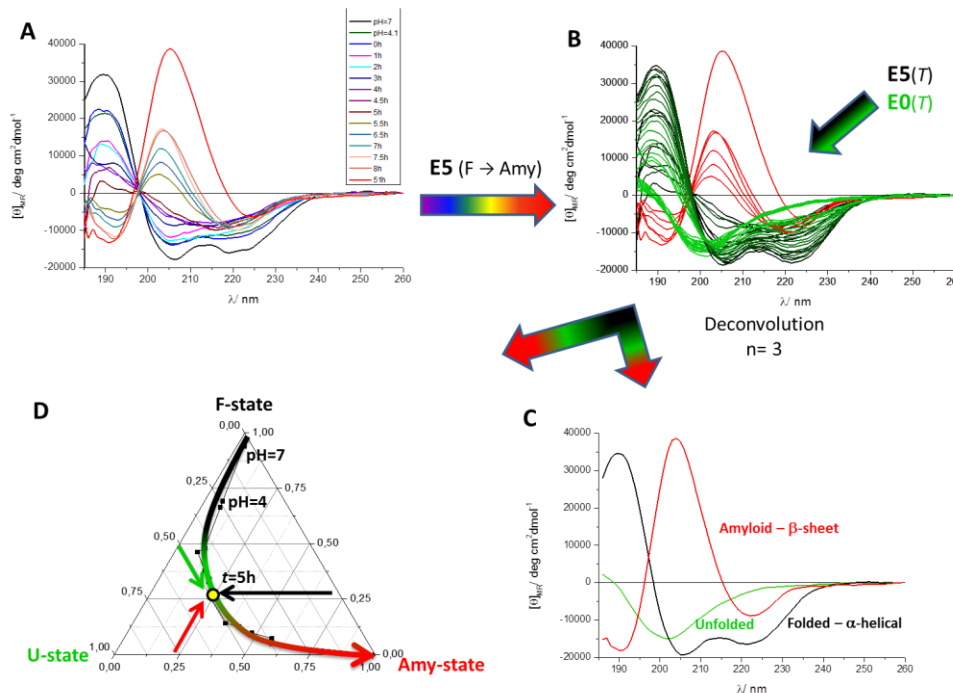

Figure 5: A) *The C-type FUV ECD of E5 (F-state) gradually shifts as function of the time, as amyloid formation progresses (c_{E5}=250μM, c_{NaCl}= 50mM, pH=4.1, stirring at T=37°C). The final B-type spectrum stands for the amyloid state (t= 51 h) composed of β-strands.* ***B)*** *An ensemble of the FUV-ECD spectra of E5 forming amyloid (panel A) was formed with those of the T-dependent E0 (Rovó Biochem 2014) and E5 (5< T(°C)< 85) and deconvoluted into 3 pure component spectra (**C**) standing for the F-(black), U-(green)- and Amyloid-states (red).* ***D)*** *The CCA+ resulted in relative coefficients of these 3 pure ECD spectra depict the amyloid formation path in a Barycentric coordinate system: after 5 h ratio is 26% F-, 50% U- & 24% Amy-state (yellow dot) (c.f.* ***SFigure 7***).

Fine-tuning $c_{protein}$ and c_{salt} affecting the amyloid formation

Although E5 can form amyloid at low salt (c_{NaCl} <1 mM) and high protein concentrations (c_{E5} >5-15 mM) (Farkas 2013), at these extreme conditions amyloid formation is poorly reproducible. To locate physiologically more relevant and controllable amyloidogenic conditions on the $\Delta G= f(T,pH,c_{prot},c_{ion},t,)$ surface, both c_{E5} and c_{NaCl} variables had to be optimized at a sensible concentration range: 250< c_{E5} < 800 µM and 12.5 <c_{NaCl} < 50 mM with T & pH set as: *T =37°C & pH= 4.1*. Fine-tuning was completed with the HANABI system in a concerted manner leading to the identification of several concentration pairs for which amyloid formed and the transition was reproducible, namely *c_{E5}=250 µM & c_{NaCl}=50 mM, c_{E5}=500 µM & c_{NaCl}=25 mM c_{E5}*=800 µM & *c_{NaCl}*=12.5 mM (with the solution constantly stirred). Thus, if the protein concentration is higher, then even at a lower salt concentration amyloid formation was observed **(Figure 6** and **SFigure 8).**

As protein concentration increases, amyloid formation becomes more complete: Amy]$^{c(E5)=400uM}$ (*t*= 8 h)= **0.34**, [Amy]$^{c(E5)=800mM}$(*t*= 8 h)= **0.61** even when salt concentration is low and fixed at c_{NaCl}= 12.5 mM (a condition more suitable for NMR) **(Figure 7/a, SFigure 9**).
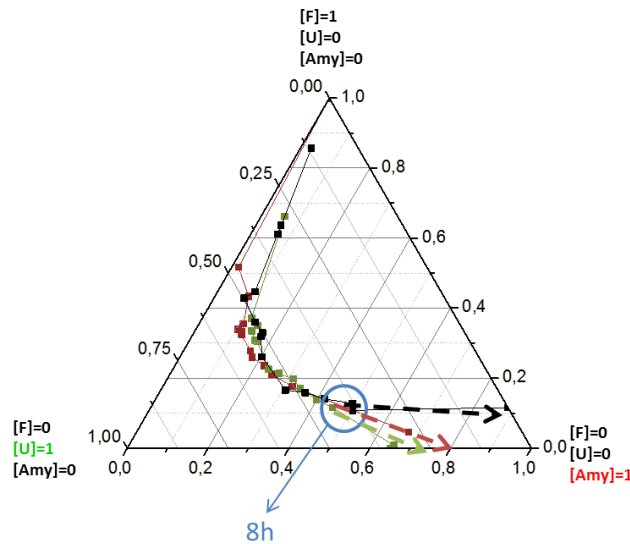

*Figure 6: A*myloid forms at a large concentration range: from *c_{E5}=250 µM & c_{NaCl}=50 mM* up to *c_{E5}*=800 µM & *c_{NaCl}* = 12.5 mM (for ECD spectra see **SFigure 8**). The pinpointed pathways and formation rate looks similar *[c(E5)=250µM & c(NaCl)= 50mM (black) and* c(E5)=500µM & c(NaCl)= 25mM, *(claret) and* c(E5)=800µM & c(NaCl)= 12.5mM, *(green) (*pH=4.1 & *stirring at T=37°C)].*

Furthermore, the rate of amyloid formation was probed at a fixed protein concentration (*e.g.* c$_{E5}$=400 µM) with increasing salt concentration (c_{NaCl}: 0 mM, 12.5mM, 25mM, 50 mM) **(Figure 7/b, SFigure 10).** We found that both misfolding and nucleation occurs at a quasi-similar rate. However, during the elongation phase clear differences were detected as function of c_{NaCl}. Stirring for three days (*t*= 76 h) the amyloid ratio was found higher if salt concentration was lower: [Amy]$^{c(NaCl)=0mM}$ =0.85,

$[Amy]^{c(NaCl)=12.5mM}$ =0.78, $[Amy]^{c(NaCl)=25mM}$ =0.63, $[Amy]^{c(NaCl)=50mM}$ =0.52. This finding, at first glance, could suggest that the easy way to avoid amyloid formation - at least for E5 - is to use a high salt concentration. However, as we only detect amyloids of limited size by ECD (those that remain part of the solution), it is more likely that the larger salt concentration speeds up amyloid maturation and thus, eliminates shorter amyloid fragments from the solution, an assumption more in line with the literature data. Moreover, at the salt concentration used here (< 100 mM) the beside the nature of the anion and cation, specific ion-binding to the polypeptide chain was also shown to contribute to the rate of amyloidosis (Marek Biochemistry 2012, Buell Biophys J. 2013, Klement J. Mol. Biol. 2007). This could well be the case for E5 also: during MD simulations of the monomeric protein (both at pH=7 and pH=4.1 in 0.15M NaCl solution) ~10% of the all ion-protein interactions involved the charged residues of the $Glu^{14}\leftrightarrow Arg^{21}$ salt bridge, which might contribute to the loosening of the hydrophobic core of E5 (**SFigure 11**), influencing the ratio of the misfolded structures present.

Temperature-dependence revisited

The effect of temperature on the kinetics of amyloid formation was revisited by using the fine-tuned conditions of amyloid formation (c_{E5}=250 μM, c_{NaCl}= 50 mM and *pH*=4.1) (**Figure 7/C, SFigure 12**). We found that the process is slower and incomplete both at 23°C and 47°C, compared to that of the physiological temperature (37°C). It seems that at a high temperature the increased thermal motion disfavors self-association: $[Amy]^{47°C \& 51h}$ =0.34 << $[Amy]^{37°C \& 51h}$ =0.86. At temperatures too low, the Brownian motion is significantly slower and thus fewer collisions occur and the misfolding propensity of E5 gets lower (see NMR data above): $[Amy]^{23°C \& 51h}$ =0.12.
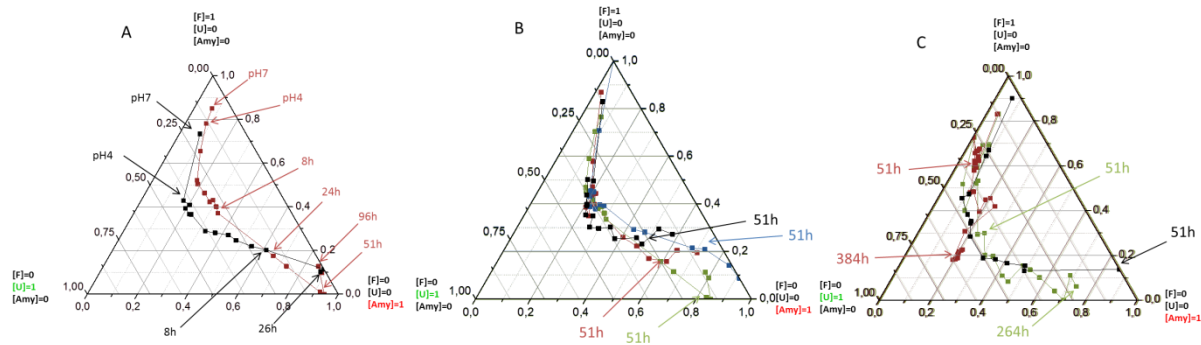
Figure 7.: *Amyloid formation completeness monitored by Far-UV ECD as function of the time **A**) different protein (black: 800 μM, claret: 400 μM) but **same ion** concentration (12.5 mM) **B**) **same protein** (400μM) and different ion concentration (black: 50mM, claret: 25mM green:12.5mM, blue: 0 mM) **C**) **same protein** (250μM) and the same **ion concentration** (50 mM) at different temperatures: T= 23 °C (claret), 37 °C (black) and 47 °C (green).*

Verifying amyloidosis – NUV-CD measurements

The amyloid transformation was monitored from the viewpoint of the $Y^8 \leftrightarrow W^{11}$ aromatic interaction by acquiring NUV-ECD data. At pH=6 the characteristic positive band at 283nm (over the entire tested protein concentration range) is indicative of a shifted face-to-face π-π interaction between the two aromatic rings of the hydrophobic core of E5 **(Figure 8/A)**. However, at pH=4.1 negative bands were detected in this spectral region (270< λ< 290 nm) **(Figure 8/B)** and assigned to an edge-to-face π-π interaction, signaling that misfolding of E5 concerns the relative re-orientation of the two aromatic rings of the core. However, as amyloid transformation proceeds, gradually these negative bands are reverted and positive bands similar to those measured for the folded state but at a slightly shifted position appear. As amyloid formation progresses these bands intensify: after 26 h band intensities are about 10 times higher than those of the initial F-state ($\Theta_{\sim 283nm}$: 4000→ 40000) (**Figure 8/A** and **C**). This side chain restructuring coincides with the backbone changing from an α- to β-state. Thus, we propose, that the enhanced positive NUV-ECD bands between 270< λ< 290 nm arise from no longer a pairwise, intramolecular shifted face-to-face π-π interaction, but rather from the interstrand interaction of Tyr and Trp side-chains packed tighter in the supramolecular assembly of the amyloid phase.
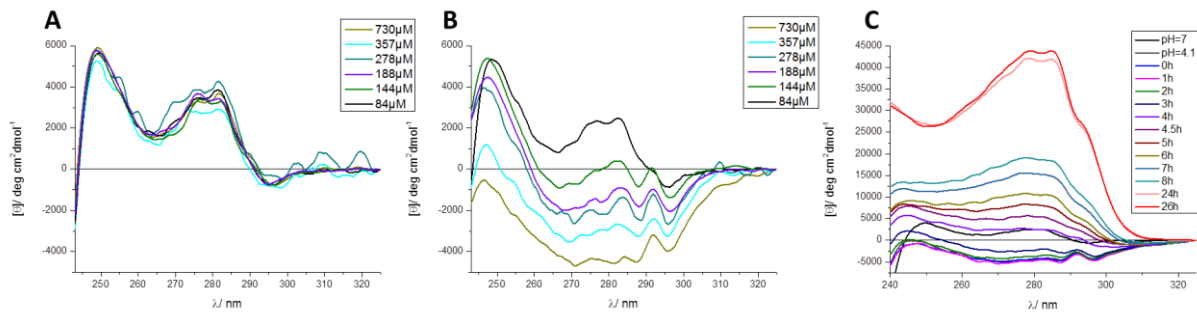

Figure 8. *NUV-ECD (240-325nm) spectra reveals how conformers and interaction modes of $Y^8 \leftrightarrow W^{11}$ aromatic residues change as function of the pH and time. At pH:4.1 (**B**) the π-π interaction profile changes as protein concentration increases (84< c_{E5} < 730 μM @ 25°C, $c_{NaCl} \sim 0mM$), while remains unchanged (**A**) at pH:6.0. (**C**) NUV-ECD spectral shift signals the changing interaction mode of Y^8 & W^{11} as function of the time (0<t <26 h) during amyloid formation ($c_{E5}=800\mu M$, $c_{NaCl}= 12.5mM$, pH=4.1, stirring at T=37°C).*

The nature of the misfolded states

Amyloid formation proceeds at physiological temperature, but it can be quenched if cooled to 4°C. Thus, a series of heteronuclear correlation spectra (^1H-^{15}N HSQC) were recorded at 4°C using samples retrieved at regular time intervals (every 30 minutes) during amyloid formation **(Figure 9).** The $^1H,^{15}N$-chemical shifts of all residues were calculated

$$\Delta(t)= \sqrt{[(\delta H(t))^2+(\delta N(t)/6{,}51)^2]} \qquad\qquad \textbf{equ 2.}$$

as function of the time (0<t<24 h). Residues of larger changes than the average chemical shift ($\Delta(t_{end})$-$\Delta(t_0)$> 0.047 ppm) are those of the α-helix (E^2, E^3, V^5, L^7, I^9, Q^{10}, K^{13}), signaling that amyloid formation affects this region. Residues of the helix were found to be the most temperature sensitive ones as well

(**SFigure 2**), indicating that this region is assailable as soon as the shielding efficacy of the Trp-cage fold against water gets reduced. This region (segment -R[6]LYIQWL[12]- was also predicted to be the most amyloidogenic of the sequence by CamSol (Sormanni 2015). Therefore, we propose that the α-helix itself is the seed of amylogenic nucleation. As its misfolding and transient unwinding allows the reorientation of the aromatic side chains di- and oligomers of the misfolded structures can form interacting through their exposed hydrophobic cores.
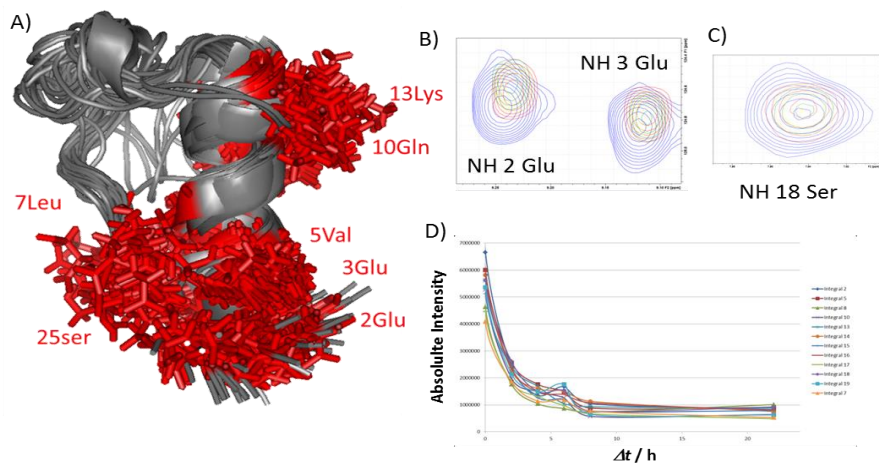

Figure 9. A) *The superimposed best 50 structures of E5. NHs of Glu[2], Glu[3], Val[5], Leu[7], Gln[10], Lys[13] and Ser[25] (highlighted red) shift the most during amyloid forming (**STable 2**). B) Selected ^{1}H- ^{15}N-HSQC resonances of E5 shifting noticeably: Δ(t$_{end}$)- Δ(t$_0$)> 0.047 ppm. C) NH of Ser[18] shifting none was used as a reference. D) The absolute intensity of the ^{1}H- ^{15}N-HSQC resonances decay with time, indicating the loss of monomers and shorter oligomeric assemblies.*

Table.1 *Conformational measures introduced to signal the degree of misfolding of E5. Each measure type a critical value or critical interval was specified to characterize the divergent/misfolded population of the ominous **E5**[^{1}E-$_1$,^{2}E$_0$,^{3}E$_0$,^{14}E$_0$] microstate.*

#	measure type	description of the measure	critical value of the measure
i	$d_{11W\leftrightarrow17P}$	the distance between Trp[11] & Pro[17]	$d_{11W\leftrightarrow17P}$ > 5.7 Å
ii	$d_{11W\leftrightarrow23P}$	the distance between Trp[11] & Pro[23]	$d_{11W\leftrightarrow23P}$ > 4.7 Å
iii	$d_{11W\leftrightarrow21R}$	distance between NH of Trp[11] and CO of Arg[21]	$d_{11W\leftrightarrow21R}$ > 6.0 Å
iv	$d_{14E\leftrightarrow21R}$	distance between donor of Arg[21] and acceptor of Glu[14]	
v	μ (°)	the torsional angle between the main axis of the core α-helix and of the polyPro of the C-terminal	μ> 45°
vi	$d_{Y8\leftrightarrow W11}$	distance between Tyr[8] & Trp[11]	$d_{Y8\leftrightarrow W11}$ > 7.0 Å
vii	α (°)	the angle between the planes of the aromatic rings	α> 70°
viii	ϑ (°)	the elevation of the aromatic rings	ϑ> 6.5° 40°> ϑ> 50° 80°> ϑ> 90°
ix	φ (°)	the azimuthal angle of the aromatic rings	φ> 320°

Amyloid formation is optimal at pH= 4.1, where the overall charge of E5 is near +2 (**Figure 1-3**) and thus, on average 3 out of the 4 Glu side-chains are protonated. To identify that protonation state which is the most likely to trigger misfolding, MD simulations were carried out. At pH=7 only one possible protonation motif, that of **E5**$[^1E_{-1},^2E_{-1},^3E_{-1},^{14}E_{-1}]$, exists, while at pH=4.1 as many as 4 (**E5**$[^1E_0,^2E_0,^3E_0,^{14}E_{-1}]$; **E5**$[^1E_0,^2E_0,^3E_{-1},^{14}E_0]$; **E5**$[^1E_0,^2E_{-1},^3E_0,^{14}E_0]$; **E5**$[^1E_{-1},^2E_0,^3E_0,^{14}E_0]$) different protonation patterns have to be considered, which we will refer to as microstates. Conformational descriptors were selected to measure the extent of misfolding (see Methods, **Table 1** and **Figure 10, SFigure 13**).

The simulations of four microstates provided equilibrium ensembles quite similar to that of the folded structure derived from NMR data: those of, **E5** $_{pH=4.1}[^1E_0,^2E_0,^3E_0,^{14}E_{-1}]$, **E5**$_{pH=4.1}[^1E_0,^2E_0,^3E_{-1},^{14}E_0]$ and **E5**$_{pH=4.1}[^1E_0,^2E_{-1},^3E_0,^{14}E_0]$ belonging to pH=4.1 and the microstate belonging to pH=7.0. On the contrary, a single microstate, that of **E5**$_{pH=4.1}[^1E_{-1},^2E_0,^3E_0,^{14}E_0]$ was significantly different from any of the above when descriptors ***i-ix*** were evaluated, this state shows strong resemblance to the loosened NMR structural ensemble measured at pH=4.1. We found that in case of **E5**$_{pH=4.1}[^1E_{-1},^2E_0,^3E_0,^{14}E_0]$ both $d_{11W\leftrightarrow17P}$, and $d_{11W\leftrightarrow23P}$, distances are lengthened by 0.4 and 0.3 Å, respectively, and in some conformers the $d_{11W\leftrightarrow21R}$ distance shifted from 2.85 Å to 5.85 Å, indicating the appearance of less compact protein folds. Furthermore, as the distribution of μ gets wider, the -PPP^{24}- segment twists more often with respect to the main axes of the α-helix. In parallel, $d_{Y8\leftrightarrow W11}$ increases by 0.9 Å. The relative orientation of the aromatic side-chains changes as well: measured $\vartheta°$ shows a more diverse distribution relative to the other microstates, while the values of $a°$ and $\varphi°$ slightly increase. Summing these changes up, we might say that **E5**$_{pH=4.1}[^1E_0,^2E_0,^3E_0,^{14}E_{-1}]$ protonation microstate has an enhanced backbone conformational freedom and a significantly rearranged hydrophobic core, with respect to all the others. The Trp-cage gets occasionally very unfolded, giving rise to unshielded aromatics and exposed backbone amide groups, ready for self-association and subsequent amyloid formation (**SFigure 13**), though these events are transient, and misfolding is temporary. However, it should be noted that MD simulations carried out here consider isolated molecules. The exposed aromatic and hydrophobic sidechains create a "sticky" interaction center for E5, quite as the free β-edges that appear transiently on the surface of locally misfolded large, globular proteins and become initiators of aggregation. Thus, when collisions among similarly loosened conformers of E5 are also considered, the described changes might become sufficient to create the first di- and oligomer nuclei of aggregates (**Figure 13/A**). These findings also explain why stirring is necessary for amyloid formation of E5 – since only one of the four possible protonation patterns of pH=4.1 produces a "misfolded enough" conformer, a great number of collisions are required for successful nucleation.
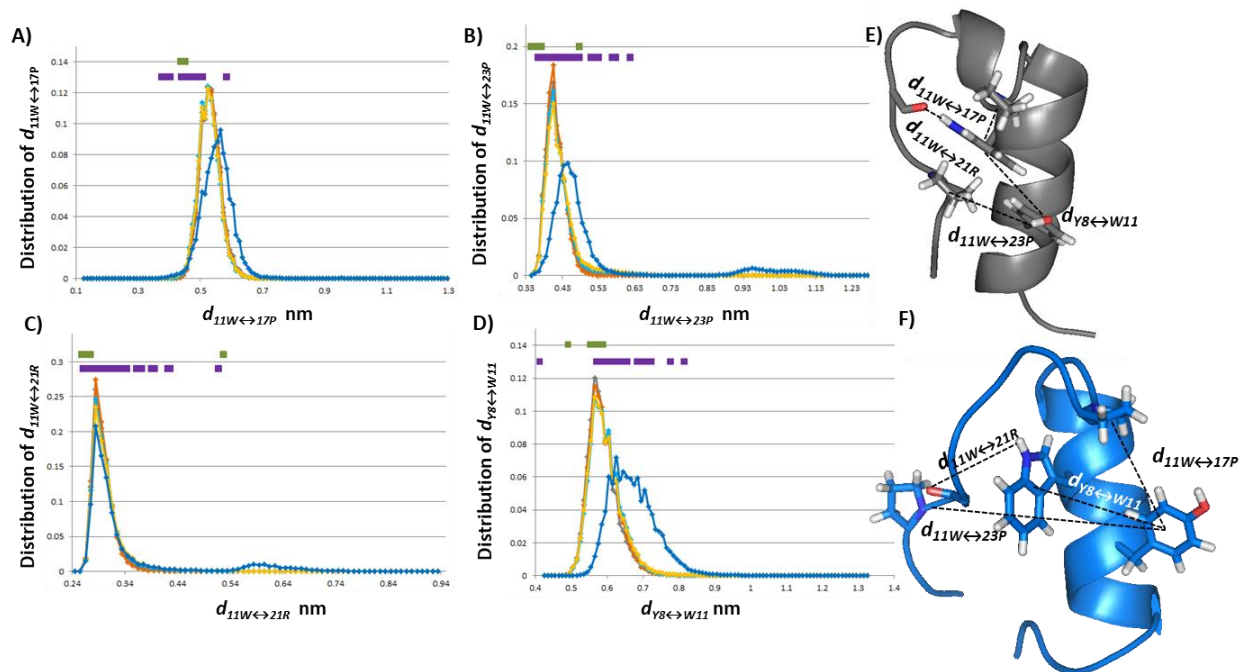
Figure 10. *The distributions of conformational measures introduced to signal the degree of misfolding of E5. A-D The distribution of the selected measures for the MD trajectories (grey – pH7, E5[$^1E_{-1}$,$^2E_{-1}$,$^3E_{-1}$,$^{14}E_{-1}$] ; orange – pH4 E5[$^1E_0$,$^2E_0$,$^3E_0$,$^{14}E_{-1}$] ; light blue – pH4 E5[$^1E_0$,$^2E_0$,$^3E_{-1}$,$^{14}E_0$]; yellow – pH4 E5[$^1E_0$,$^2E_{-1}$,$^3E_0$,$^{14}E_0$]; blue – pH4 E5[$^1E_{-1}$,$^2E_0$,$^3E_0$,$^{14}E_0$] are plotted, while the distribution of the selected measures for NMR structures (green – pH7, purple – pH4) are indicated for every single structures above the MD distribution curves in a linear 1D manner. The measures of E5[$^1E_{-1}$,$^2E_0$,$^3E_0$,$^{14}E_0$] microstate is slightly different from the other microstates, moreover a minor population of E5[$^1E_{-1}$,$^2E_0$,$^3E_0$,$^{14}E_0$] microstate show considerably distinct properties. Table 1 contains a limit for each measure that can be used to distinguish/separate the divergent/misfolded population of E5[$^1E_{-1}$,$^2E_0$,$^3E_0$,$^{14}E_0$] microstate. E) NMR calculated folded structure (grey) and F) one of the MD-based misfolded (blue) structures are shown to visualize the conformational differences. Additional distributions of the selected measures are shown in **SFigure 14**.*

Building a model of the amyloid state

To obtain a picture of the amyloid state, MD simulations were carried out for models of four β-strand-clusters $(E5)_4$ using three different protonation motifs: $(E5_{pH=2}[^1E_0,^2E_0,^3E_0,^{14}E_0, \text{C-term}_0])_4$, $(E5_{pH=4.1}[^1E_{-1},^2E_0,^3E_0,^{14}E_0])_4$ and $(E5_{pH=7}[^1E_{-1},^2E_{-1},^3E_{-1},^{14}E_{-1}])_4$. (Four-stranded models were shown sufficiently large for modeling aggregation nuclei in various different systems ([Weihua Zheng 2017, Ngo 2017, Knowles 2014).) We have probed three different relative β-strand offsets, namely that of "YI", "YK" and "YW" (**Figure 11/A, B** and **C**) for each.
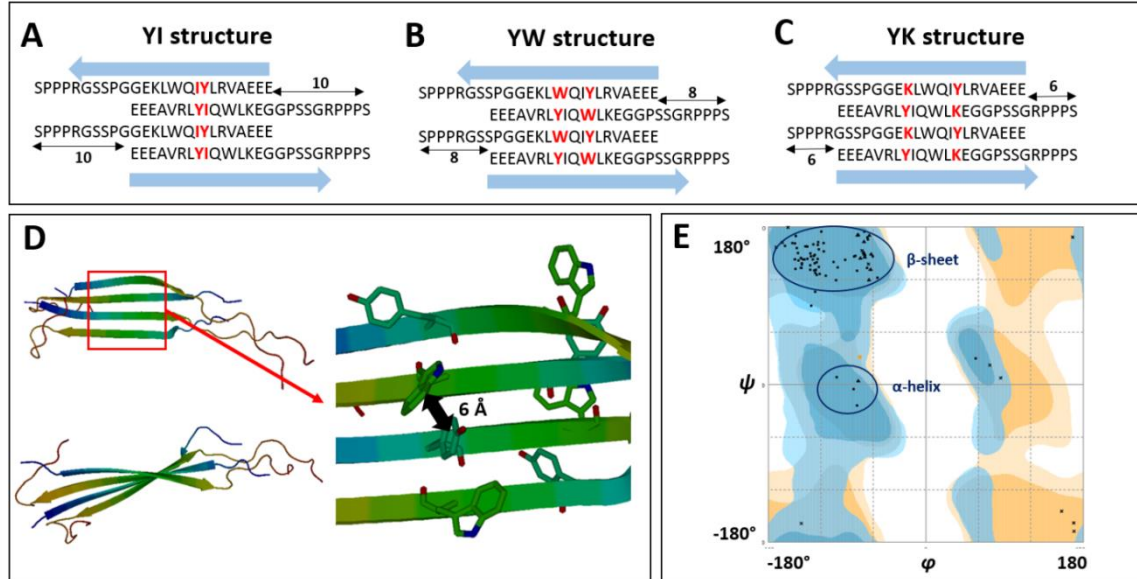

Figure 11. **A-C)** The *three different β-strand offsets of $(E5)_4$ used for MD simulations. In each of the composed models, the central helix was extended and Tyr8 was paired antiparallel either with Ile9, Trp11 or Lys13 residues of the adjacent β-strand. **D)** The most populated amyloid-like tetramer cluster of YW (pH=4.1 between 100-340 ns) have shifted face-to-face π-π aromatic ring orientations (with ring separations of ~6 Å). On the contrary, neither YI, nor YK offsets made stable enough tetramers. **E)** The Ramachandran plot of the most populated cluster of the $(E5)_4$ YW structure with residues mainly located in the β-valley.*

From the three probed offsets YW turned out to be both the most stable at *pH*=4.1 and the most sensitive to pH shift (since both at pH=2 and 7, the folded conformation is most stable and practically no amyloid formation could be detected, we expected the most realistic model to be significantly more stable at *pH*=4.1) **(SFigure 15)**. Considering that in the -PPP^{24}- segment φ^{Pro} must be ~-60° and thus this part of the sequence is unlikely to form an extended β-sheet structure (with φ~-145°, ψ~+145°), in case of the YW offset 18 pairs of interstrand H-bonds could be formed between residue 1-18 in total. In YW$_{pH=4.1}$ on average 10 out of 18 H-bonds are present during the MD simulation between the two middle chains typically involving residue-pairs such as V^5-E^{14}, L^7-L^{12}, I^9-Q^{10}, W^{11}-Y^8, and K^{13}-R^6. These interstrand H-bonds confirm that an aggregation core can indeed form between residues 5-14, as suggested above. Furthermore, in this arrangement of YW$_{pH=4.1}$ an aromatic ladder forms between the Y^8 and W^{11} side-chains of adjacent β-strands, as they come in close proximity ($d_{Y8\leftrightarrow W11}$ <6 Å as measured between the center of the rings) (**Figure 11/D**). In addition, the aromatic ladder enables the formation of the shifted

face-to-face π-π interaction expected based on ECD spectroscopic data. Interestingly, the derived orientation and spacing of the Tyr and Trp sidechains here is also quite reminiscent of the aromatic clusters created by mutations within the central single-layer β-sheet of Borrelia outer surface protein A (OspA) to probe the structure-ordering capacity of such residues [Biancalana, 2015. Prot. Sci.].

Using the mid-structures of six different clusters (see methods) of the MD simulation of ($E5_{pH=4.1}$)$_4$_YW as possible amyloid seeds, we probed different inter-sheet arrangements as well, by using a protein-protein docking algorithm. The near-parallel arrangements thus obtained could be grouped into two basically different topologies (using the nomenclature introduced by Eisenberg and coworkers [Eisenberg 2016 Annu. Rev. Biochem]): those of class 7 (equifacial, antisymmetric, up-up) with translational symmetry and class 8 (antiparallel, equifacial, antisymmetric, up-down) with 2-fold symmetry **(Figure 12)**. In the derived β-sheet-dimers, (($E5_{pH=4.1}$)$_4$)$_2$, the average distance between the β-sheets separated by the zipper interface varies between 10-12 Å, in line with the 8-10 Å experimental values determined for various amyloid structures [Berhanu 2012. PLoS One][Xi 2016. Chin. Phys B.] [Colletier 2011. Proc.Natl.Acad.Sci.USA][Soriaga 2016. J.Phys.Chem.B][Saelices 2018. Protein Sci.]. Furthermore, in this sandwich form, the intersheet aromatic sidechains may approach each other even closer than the intrasheet 6 Å described above (**Figure 13/B**). We found that ($E5_{pH=4.1}$)$_4$ seeds with a slightly twisted backbone find partners more readily to form the sandwich structure than those with fully extended backbones, in accordance with the general view that twisted filaments are more stable and thus more prevalent in nature [Lu 2018. J. Phys. Chem. B]. These findings further support our model for the aggregation of E5 as nucleated by the antiparallel oriented β-sheets formed between the most aggregation-prone segment of the sequence, the -V[5]RLYIQWLK[13]- unit.

Up-down, class8
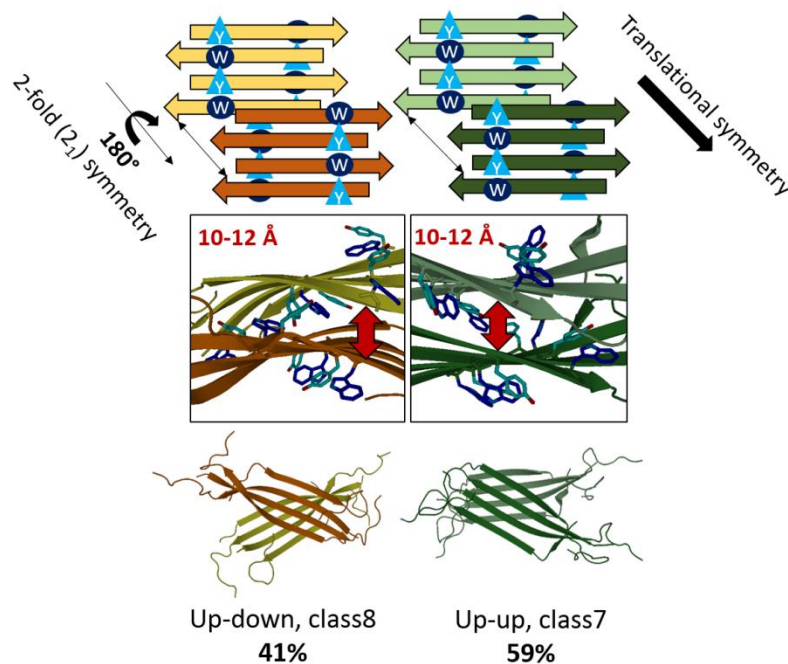**41%**

Up-up, class7
59%

Figure 12. *The two basic topologies found by docking (E5)$_4$ amyloid seeds are the antiparallel equifacial antisymmetric up-down (class8) with a 2 fold symmetry and equifacial antisymmetric up-up (class7) with translational symmetry. The schematic illustrations of the basic amyloid layout shown have the aromatic residues highlighted: above, Tyr: light blue triangle and Trp: dark blue circle. Twisted amyloid β-sheets (below) and 90° rotated plus enlarged view of the aromatic residues are shown in the middle: Tyr (light blue) and Trp (dark blue) with oxygen atoms red and nitrogen middle blue). Red double arrows show the specific distance of the associated β-sheets.*
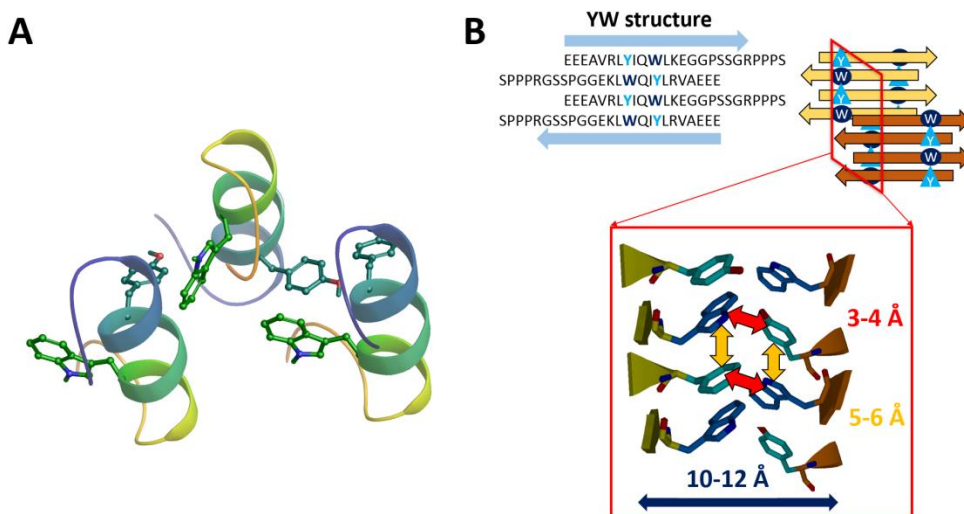
Figure 13. A: *Possible self-aligning mode of a conformer of the MD simulation. The antiparallel orientation of the helices is due to coulombic interactions operative at pH=4.1 in the selected microstate. Aromatic-aromatic (Y-W) interaction driven backbone pairing of the misfolded molecules. **B:** Schematic illustration of a hypothetic late phase intersheet orientation of the aromatic rings (Tyr: light blue, Trp: dark blue). The intrasheet distance between the center of masses of the two aromatic sidechains is 5-6 Å (yellow double arrows), while possible intersheet aromatic-aromatic distance can be 3-4 Å (red double arrows), while the sheet-sheet distance is 10-12 Å (blue double arrow).*

Conclusion

Concerning E5, the Exenatide variant studied here, we found that its amyloid aggregation is most likely triggered by the transiently exposed aromatic and hydrophobic side-chains of loosened – misfolded - conformers, which create a center for intermolecular associations. The clusters thus created provide stabilization (spatial closeness for sufficient time) for the much slower process of α-helix to the coil and then to an extended β-strand transition, which seems to be highly unfavorable for the monomeric forms but eventually leads to the formation of a β-structured amyloid nucleus. The amyloidosis of E5 thus follows the nucleated growth mechanism described elsewhere (C-C. Lee et al. Biophys. J. 2007, Serio, T. R. Science. 2000).

Studying the early phases of amyloid formation is crucially important both for understanding the initialization of various pathophysiological processes and for aiding the design of non-toxic peptide medications that will not become initiators of such processes themselves. We derived a new monitoring technique of amyloid progression that can be applied even if the amyloid in question is ThT silent - using simple ECD measurements. Decomposition of the spectra and its barycentric representation on the folded-unfolded-amyloid potential energy surface pinpointing the pathway of the amyloidic transition can be applied to filter out potentially harmful sequences from development. Generally, it can be concluded that the lowest possible salt concentration, low temperatures and the absence of agitation can prolong the shelf-life of any polypeptide and protein medications, but somewhat disturbingly we also found that the optimum temperature for E5 amyloidosis coincides with our body temperature and requires well below physiological salt concentration. This just underlines how important it is to test the aggregation propensity of any drug candidates.

Acknowledgment

Methods

Protein expression and purification: The E5 miniprotein was produced by bacterial expression using the previously published protocol. ref: Stráner, P., Taricska, N., Szabó, M., Tóth, K. G., Perczel, A. Bacterial expression and/or solid phase peptide synthesis of 20-40 amino acid long polypeptides and miniproteins, the case study of Class B GPCR ligands. Current Protein and Peptide Science 17, 147-155 (2016).

Preparation of amyloid form of E5 miniprotein: The lyophilized E5 samples were dissolved in distilled water. First, the pH of protein solution was adjusted to pH=7 with 0.1 M NaOH solution (Orion Star A211 pH meter (Thermo Scientific™)) then decreased to 4.1 or lower (pH-dependent amyloid formation) with 0.1 M HCl solution. Finally desired NaCl concentration was adjusted with concentrated NaCl solution. The protein solution was stirred (with magnetic stirrer rpm) and incubated at target temperatures. At a given time 10 fold diluted sample was measured with ECD spectroscopy. The concentration of every diluted sample was determined by NanoDrop Lite Spectrophotometer (Thermo Scientific™) at 280 nm.

NMR experiments: Datasets were collected on a 16.4 T Bruker Avance III spectrometer equipped with a 5 mm inverse TCI probe-head with z-gradient. Each sample contained 700-900 µM of E5 peptide, 10% D_2O, 0.1 mM NaN_3 and DSS as an internal proton reference standard. The concentrations were determined by Nanodrop. The sample pH was set between pH 7.1 and 2.0 by using 0.1M NaCl or HCl and the measurements were recorded from 277.0 K to 321K by equal steps of 11 K. Spinlock (d9) for 1H–1H TOCSY was 80ms, while the mixing time (d8) of 150 ms was taken for 1H–1H NOESY spectra. Raw spectra were processed by TopSpin 3.2 and the homonuclear proton NMR assignations were made using CcpNmr Analysis 2.4.1. (Vranken; 2005; Proteins) The NMR structure calculations were performed and refined by cooperative use of CcpNmr Analysis 2.4.1., Aria 2.0 (Rieping; 2007; Bioinformatic) and CNS Solve 1.2. (Brunger,2007;Nature Protocols)   The NOE cross-peak intensities were converted into distance. The peak volume scaling factor is defined such that the peak list's average volume corresponds to the default 3.2 Angstroms. Each ensemble was refined in water and contains 50 structures.

Molecular dynamics simulations: MD simulation were carried out as implemented in GROMACS59, using the AMBER-ff99SBildnp* forcefield. The systems were solvated with TIP3P water molecules in dodecahedral boxes with a size allowing 10Å between any protein atom and the box. The total charge was neutralized and physiological salt concentration was set using Na^+ and Cl^- ions. Energy minimization of starting structures was followed by sequential relaxation of constraints on protein atoms in three steps and an additional NVT step (100 ps) to stabilize pressure. Trajectories of 600 ((E5)$_4$ models) -1000ns (E5 monomers) NPT simulations with a 2fs time step at 310K and 328K and 1bar were collected (with snapshots at every 4ps). The last 500ns of the trajectories were clustered fitting the main-chain atoms using a 1Å cutoff in case of the monomers, while the last 400ns of the trajectories of the tetramer sheets were clustered using a 1.7 Å cutoff fitting the main-chain atoms of the seed segment (residues 6-13). 112 clusters were found (95% of the snapshots are in 32 clusters). For docking 4-stranded seed models to one-another, we choose six different 4-stranded sheet conformer (middle structures of six different clusters from the MD simulation of the (***E5***$_{pH=4.1}$[^{1}E$_{-1}$,^{2}E$_0$,^{3}E$_0$,^{14}E$_0$])$_4$ microstate with an YW offset) which were docked to another identical 4-stranded sheet, using Piper [Kozakov, 2006, Proteins]During docking the sidechains were rigid, 10000 ligand rotations were probed and the 30 largest clusters were chosen for further analysis. We selected manually the most relevant amyloid-like structures from the clusters (less than 45° angle between the strands, the plane of the sheets close to parallel): these were 10-24% of the total 30 clusters. The distance of the β-sheets was measured as the backbone distance the aromatic

residues of the two middle chains between the two opposite sheets of the amyloid seed region (residues 6-13).

Electronic circular dichroism spectroscopy: Far and near-UV ECD measurements were carried out on a Jasco J810 spectrophotometer in 1.0 and 10 mm quartz cuvettes. Typically we used spectral scanning speed of 50 nm/min with 1 nm bandwidth and the 0.2 nm step resolution over wavelength range 185-260 nm (far-UV) and 240-325 nm (near-UV) with three and four scans averaged for each spectrum. The temperature at the cuvette was controlled by Peltier-type heating system (temperature dependent measurements were made from 5°C to 85°C in 5°C steps and the amyloid samples were measured at 25°C). The raw ellipticity data were converted into mean residue molar ellipticity units ($[\theta]_{MR}$/ deg*cm^2*dmol^{-1}) for the far-UV region and molar ellipticity ($[\theta]$/ deg*cm^2*dmol^{-1}) for the near-UV region.

CCA+ deconvolution: During deconvolution of amyloid formation we used the temperature dependent ECD spectra of E5 (as folded) and E0 (as unfolded) miniprotein. (These spectra were measured from 5°C to 85°C for 5°C step (17 spectra/ miniprotein)). This resulting matrix was deconvoluted into 3 components and then reported in a barycentric coordinate system.

TEM measurement: Sample solution (5 µL) was spotted onto a collodion-coated copper grid (Nisshin EM Co., Tokyo, Japan). After 1 min, the remaining solution was removed with filter paper and 5 µL of 2 % (w/v) ammonium molybdate was spotted onto the grids. After 1 min, the remaining solution was removed in the same manner. TEM images were obtained with a Hitachi H-7650 transmission electron microscope (Tokyo, Japan) with a voltage of 80 kV.

References

[1] Kim, T.; Vidal, G. S.; Djurisic, M.; William, C. M.; Birnbaum, M. E.; Garcia, K. C.; Hyman, B. T.; Shatz, C. J., Human LilrB2 is a β-amyloid receptor and its murine homolog PirB regulates synaptic plasticity in an Alzheimer's model. *Science* **2013**, 341(6152), 1399-404.

[2] Shankar, G. M.; Bloodgood, B. L.; Townsend, M.; Walsh, D. M.; Selkoe, D. J.; Sabatini, B. L., Natural oligomers of the Alzheimer amyloid-beta protein induce reversible synapse loss by modulating an NMDA-type glutamate receptor-dependent signaling pathway. *J Neurosci*. **2007**, 27(11), 2866-75.

[3] Laurén, J.; Gimbel, D. A.; Nygaard, H. B.; Gilbert, J. W.; Strittmatter, S. M., Cellular prion protein mediates impairment of synaptic plasticity by amyloid-beta oligomers. *Nature* **2009**, 457(7233), 1128-32.

[4] Bussi, C.; Peralta Ramos, J. M.; Arroyo, D. S.; Gallea, J. I.; Ronchi, P.; Kolovou, A.; Wang, J. M.; Florey, O.; Celej, M. S.; Schwab, Y.; Ktistakis, N. T.; Iribarren, P., Alpha-synuclein fibrils recruit TBK1 and OPTN to lysosomal damage sites and induce autophagy in microglial cells. *J Cell Sci* **2018,** 131, jcs226241

[5] Booth, D. R.; Sunde, M.; Bellotti, V.; Robinson, C. V.; Hutchinson, W. L.; Fraser, P. E.; Hawkins, P. N.; Dobson, C. M.; Radford, S. E.; Blake, C. C. F.; Pepys, M. B., Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **1997,** 385, 787–793.

[6] Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R., A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A*. **2002**, 99(26), 16742-16747.

[7] Gremer, L.; Scholzel, D.; Schenk, C.; Reinartz, E.; Labahn, J.; Ravelli, R. B. G.; Tusche, M.; Lopez-Iglesias, C.; Hoyer, W.; Heise, H.; Willbold, D.; Schröder, G. F., Fibril structure of amyloid-beta(1-42) by cryo-electron microscopy. *Science* **2017,** 358(6359), 116-119.

[8] Wälti, M. A.; Ravotti, F.; Arai, H.; Glabe, C. G.; Wall, J. S.; Böckmann, A.; Güntert, P.; Meier, B. H.; Riek, R., Atomic-resolution structure of a disease-relevant Aβ(1–42) amyloid fibril. *PNAS* **2016,** 113(34), E4976-E4984.

[9] Iwata, K.; Fujiwara, T.; Matsuki, Y.; Akutsu, H.; Satoshi Takahashi, S.; Naiki, H.; Goto, Y., 3D structure of amyloid protofilaments of β2-microglobulin fragment probed by solid-state NMR. *PNAS* **2006,** 103(48), 18119-18124.

[10] Meehan, S.; Berry, Y.; Luisi, B.; Dobson, C. M.; Carver, J. A.; MacPhee C. E., Amyloid fibril formation by lens crystallin proteins and its implications for cataract formation. *J Biol Chem*. **2004**, 279(5), 3413-9.

[11] Weingarten, M. D.; Lockwood, A. H.; Hwo, S. Y.; Kirschner, M. W., A protein factor essential for microtubule assembly. *Proc Natl Acad Sci U S A*. **1975**, 72(5), 1858–1862.

[12] Fitzpatrick, A. W. P.; Falcon, B.; He, S.; Murzin, A. G.; Murshudov, G.; Garringer, H. J.; Crowther, R. A.; Ghetti, B.; Goedert, M.; Scheres, S. H. W., Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **2017,** *547*(7662), 185-190.

[13] Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, A. Ø.; Riekel, C.; Eisenberg. D., Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **2007,** 447, 453–457.

[14] Oliveira, C. L.; Behrens, M. A.; Pedersen, J. S.; Erlacher, K.; Otzen, D.; Pedersen, J. S., A SAXS Study of Glucagon Fibrillation, *Journal of Molecular Biology* **2009**, 387(1), 147-161.

[15] Jiménez, J. L.; Nettleton, E. J.; Bouchard, M.; Robinson, C. V.; Dobson, C. M.; Saibil,H. R., The protofilament structure of insulin amyloid fibrils. *PNAS* **2002**, 99(14), 9196-9201.

[16] Rodriguez, J. A.; Ivanova, M. I.; Sawaya, M.R.; Cascio, D.; Reyes, F. E.; Shi, D.; Sangwan, S.; Guenther, E. L.; Johnson, L. M.; Zhang, M.; Jiang, L.; Arbing, M. A.; Nannenga, B. L.; Hattne, J.; Whitelegge, J.; Brewster, A. S.; Messerschmidt, M.; Boutet, S.; Sauter, N. K.; Gonen, T.; Eisenberg D., Structure of the toxic core of α-synuclein from invisible crystals. *Nature* **2015,** 525(7570), 486-90.

[17] Wiltzius, J. J. W.; Sievers, S. A.; Sawaya, M. R.; Cascio, D.; Popov, D.; Riekel, C.; Eisenberg, D., Atomic structure of the cross-β spine of islet amyloid polypeptide (amylin). *Protein Science* **2008,** 17(9), 1467-1474.

[18] Nielsen, J. T.; Bjerring, M.; Jeppesen, M. D.; Pedersen, R. O.; Pedersen, J. M.; Hein, K. L.; Vosegaard, T.; Skrydstrup, T.; Otzen, D. E.; . Nielsen, N. C., Unique identification of supramolecular structures in amyloid fibrils by solid-State NMR spectroscopy. *Angewandte* **2009,** 48(12), 2118-2121.

[19] Fernandez-Leiro, R.; Scheres, S. H. W., Unravelling biological macromolecules with cryo-electron microscopy. *Nature* **2016,** *537* (7620), 339-346.

[20] Ashbrook, S. E.;, Griffin, J. M.; Johnston, K. E., Recent Advances in Solid-State Nuclear Magnetic Resonance Spectroscopy. *Annu Rev Anal Chem* (Palo Alto Calif). **2018**, 11(1), 485-508.

[21] Fusco, G.; Chen, S. W.; Williamson, P. T. F.; Cascella, R.; Perni, M.; Jarvis, J. A.; Cecchi, C.; Vendruscolo, M.; Chiti. F.; Cremades, N.; Ying, L.; Dobson, C. M.; De Simone, A., Structural basis of membrane disruption and cellular toxicity by α-synuclein oligomers. *Science* **2017**, 358(6369), 1440-1443.

[22] Conway, K. A.; Lee, S. J.; Rochet, J.C.; Ding, T. T.; Williamson, R. E.; Lansbury, P. T. Jr., Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: implications for pathogenesis and therapy**.** *Proc Natl Acad Sci U S A* **2000**, 97(2), 571–576.

[23] Makin, S., The amyloid hypothesis on trial. *Nature* **2018**, 559(7715), S4-S7.

[24] Streets, A. M.; Sourigues, Y.; Kopito, R. R.; Melki, R.; Quake, S. R., Simultaneous Measurement of Amyloid Fibril Formation by Dynamic Light Scattering and Fluorescence Reveals Complex Aggregation Kinetics. *PLoS One* **2013**, 8(1), e54541.

[25] Biancalana, M.; Makabe, K.; Koide, A.; Koide, S., Molecular mechanism of thioflavin-T binding to the surface of beta-rich peptide self-assemblies**.** *J Mol Biol.* **2009**, 385(4), 1052-63.

[26] Skamris, T.; Marasini, C.; Madsen, K. L.; Foderà, V.; Vestergaard, B., Early Stage Alpha-Synuclein Amyloid Fibrils are Reservoirs of Membrane-Binding Species. *Sci Rep.* **2019**, 9, 1733.

[27] Gobeaux, F.; Wien, F., Reversible Assembly of a Drug Peptide into Amyloid Fibrils: A Dynamic Circular Dichroism Study. *Langmuir*, **2018**, 34(24), 7180–7191.

[28] Lau, J. L.; Dunn, M. K., Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg Med Chem*. **2018**, 26(10), 2700-2707**.**

[29] Bullock, B. N.; Jochim, A. L.; Arora, P. S., Assessing Helical Protein Interfaces for Inhibitor Design. *J. Am. Chem. Soc.* **2011**, 133(36), 14220–14223.

[30] Uccellatore, A.; Genovese, S.; Dicembrini, I.; Mannucci, E.; Ceriello, A., September Comparison Review of Short-Acting and Long-Acting Glucagon-like Peptide-1 Receptor Agonists. *Diabetes Therapy* **2015**, 6(3), 239–256

[31] Runge, S.; Thøgersen, H.; Madsen, K.; Lau, J.; Rudolph, R., Crystal Structure of the Ligand-bound Glucagon-like Peptide-1 Receptor Extracellular Domain. *J. Biol. Chem.* **2008**, 283, 11340-11347.

[32] Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J., Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 μs. *J. Am. Chem. Soc.* **2002**, 124(44), 12952 – 12953.

[33] Stráner, P.; Taricska, N.; Szabó, M.; Tóth, K. G.; Perczel, A., Bacterial expression and/or solid phase peptide synthesis of 20-40 amino acid long polypeptides and miniproteins, the case study of Class B GPCR ligands. *Current Protein and Peptide Science* **2016**, 17, 147-155.

Results

Farkas, V.; Csordás, B.; Hegyi, O.; Tóth, G. K.; Perczel, A., Foldamer Stability Coupled to Aggregation Propensity of Elongated Trp-Cage Miniproteins. *European Journal of Organic Chemistry* **2013**, 13, 3513–3522.

Chatani, E.; Lee, Y-H.; Yagi, H.; Yoshimura, Y.; Naiki, H.; Goto, Y., Ultrasonication-dependent production and breakdown lead to minimum-sized amyloid fibrils. *Proc. Natl. Acad. Sci.* **2009** 106(27) 11119-11124.

Rovó, P.; Stráner, P.; Láng, A.; Bartha, I.; Huszár, K.; Nyitray, L.; Perczel, A., Structural insights into the Trp-cage folding intermediate formation. *Chemistry A European Juornal* **2013,** 19(8), 2628-2640.

Rovó, P.; Farkas, V.; Stráner, P.; Szabó, M.; Jermendy, Á.; Hegyi, O.; Tóth G. K.; Perczel, A., Rational Design of α-Helix-Stabilized Exendin-4 Analogues. *Biochemistry* **2014,** 53(22), 3540-3552.

Perczel, A.; Hollósi, M.; Tusnády, G.; Fasman, G. D., Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins. *Protein Eng.* **1991**, 4(6), 669-79.

Perczel, A.; Park, K.; Fasman, G. D., Deconvolution of the circular dichroism spectra of proteins: The circular dichroism spectra of the antiparallel β-sheet in proteins. *Proteins: Structure, Function, and Genetics* **1992** 13(1), 57–69.

Jákli, I.; Perczel, A., The inherent flexibility of peptides and protein fragments quantitized by CD in conjunction with CCA+. *J. Pept. Sci.* **2009**, 15, 738-752.

Graham, K. A.; Byrne, A.; Ruth Son, R.; Niels H. Andersen, N. H., Reversing the typical pH stability profile of the Trp-cage. *Biopolymers* **2019**, 110(3), e23260.

Lee C. C.; Nayak, A.; Sethuraman, A.; Belfort, G.; McRae, G. J., A three-stage kinetic model of amyloid fibrillation. *Biophys J.* **2007**, 92(10), 3448-58.

Serio, T. R.; Cashikar, A. G.; Kowal, A. S.; Sawicki, G. J.; Moslehi, J. J.; Serpell, L.; Arnsdorf, M. F.; Lindquist, S. L., Nucleated conformational conversion and the replication of conformational information by a prion determinant. Science **2000**, 289(5483), 1317-21.

Sreerama, N.; Venyaminov, S.Y.; Woody, R.W., Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci.* **1999**, 8(2), 370-80.

Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y. H.; Goto, Y.; Réfrégiers, M.; & Kardos, J., Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences* **2015**, 112(24), E3095–E3103.

Marek, P. J.; Patsalo, V.; Green, D. F.; Raleigh, D. P., Ionic Strength Effects on Amyloid Formation by Amylin Are a Complicated Interplay among Debye Screening, Ion Selectivity, and Hofmeister Effects. *Biochemistry*, **2012**, 51(43), 8478–8490.

Buell, A. K.; Hung, P.; Salvatella, X.; Welland, M. E.; Dobson, C. M.; Knowles, T. P. J., Electrostatic Effects in Filamentous Protein Aggregation. *Biophys J.* **2013**, 104(5), 1116-1126.

Klement, K.; Wieligmann, K.; Meinhardt, J.; Hortschansky, P.; Richter, W.; Fändrich, M., Effect of different salt ions on the propensity of aggregation and on the structure of Alzheimer's abeta(1-40) amyloid fibrils. *J Mol Biol.* **2007**, 373(5), 1321-33.

Sormanni, P.; Aprile, F. A.; Vendruscolo, M., The CamSol method of rational design of protein mutants with enhanced solubility. *J Mol Biol.* **2015**, 427(2), 478-90.

Zheng, W.; Tsai M-Y.; Wolynes, P. G., Comparing the Aggregation Free Energy Landscapes of Amyloid Beta(1–42) and Amyloid Beta(1–40). *J. Am. Chem. Soc.* **2017**, 139(46), 16666–16676.

Ngo, S. T.; Luu, X.-C.; Nguyen, M. T.; Le, C. N.; Vu, V. V., In silico studies of solvated F19W amyloid β (11–40) trimer. *RSC Advances* **2017**, 7(67), 42379–42386.

Knowles, T. P. J.; Vendruscolo, M.; Dobson, C. M., The amyloid state and its association with protein misfolding diseases. *Nature Reviews Molecular Cell Biology* **2014**, 15(6), 384-396.

Biancalana, M., Makabe, K., Yan, S., Koide, S., Aromatic cluster mutations produce focal modulations of beta-sheet structure. *Protein Sci.* **2015**, 24, 841-849.

Eisenberg, D.S.; Sawaya, M. R., Structural Studies of Amyloid Proteins at The Molecular Level. *Annu. Rev Biochem* **2017**, 86, 3.1-3.27.

Berhanu, W. M.; Hansmann, U. H. E., Structure and Dynamics of Amyloid-β Segmental Polymorphisms. *PLoS One* **2012**, 7(7), e41479.

Xi, W. H.; Wei, G. H., Amyloid-β peptide aggregation and the influence of carbon nanoparticles. *Chin. Phys. B.* **2016**, 25(1), 018704.

Colletier, J.P.; Laganowsky, A.; Landau, M.; Zhao, M.; Soriaga, A.B.; Goldschmidt, L.; Flot, D.; Cascio, D.; Sawaya, M.R.; Eisenberg, D., Molecular basis for amyloid-beta polymorphism. *Proc.Natl.Acad.Sci.USA* **2011**, 108, 16938-16943.

Soriaga, A.B.; Sangwan, S.; Macdonald, R.; Sawaya, M.R.; Eisenberg, D., Crystal Structures of IAPP Amyloidogenic Segments Reveal a Novel Packing Motif of Out-of-Register Beta Sheets. *J.Phys.Chem.B* **2016**, 120(26), 5810-5816.

Saelices, L.; Sievers, S.A.; Sawaya, M.R.; Eisenberg, D.S., Crystal structures of amyloidogenic segments of human transthyretin. *Protein Sci.* **2018,** 27(7), 1295-1303.

Lu, L.; Deng, Y.; Li, X.; Li, H.; Karniadakis G.E. Understanding the Twisted Structure of Amyloid Fibrils via Molecular Simulations. *J. Phys. Chem. B* **2018**, 122 (49), 11302–11310.

Methods

Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E.D., The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **2005**, 59(4), 687-96.

Rieping, W.; Habeck, M.; Bardiaux, B.; Bernard, A.; Malliavin, T. E.; Nilges, M., ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics **2007**, 23(3), 381-2.

Brunger, A. T., Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2007**, 2, 2728–2733.

Kozakov, D.; Brenke, R.; Comeau, S.R.; Vajda, S. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Structure, Function, and Bioinformatics*. **2006**, 65(2):392-406.

References will perhaps be eliminated:

Paolo Arosio,a Tuomas P. J. Knowlesa and Sara Linse*b , **On the lag phase in amyloid fibril formation, Phys Chem Chem Phys. 2015 Mar 28;17(12):7606-18.** Phys. Chem. Chem. Phys, 2015, 17, 7606

Antzutkin, O. N.; Leapman, R. D.; Balbach, J. J.; Tycko, R., Supramolecular structural constraints on Alzheimer's beta-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. *Biochemistry* **2002,** *41*(51), 15436-15450.

Sipe, J. D.; Benson, M. D.; Buxbaum, J. N.; Ikeda, S.; Merlini, G.; Saraiva, M. J. M.; Westermark, P., Nomenclature 2014: Amyloid fibril proteins and clinical classification of the amyloidosis. *Amyloid-Journal of Protein Folding Disorders* **2014,** *21* (4), 221-224.

Paravastu, A.K.; Leapman, R.D.; Yau, W.M.; Tycko, R., Molecular structural basis for polymorphism in Alzheimer's beta-amyloid fibrils. *Proc. Natl. Acad. Sci. USA* **2008,** 105, 18349-18354.

Colvin, M.T.; Silvers, R.; Ni, Q.Z.; Can, T.V.; Sergeyev, I.; Rosay, M.; Donovan, K.J.; Michael, B.; Wall, J.; Linse, S.; Griffin, R.G., Atomic Resolution Structure of Monomorphic A beta 42 Amyloid Fibrils. *J.Am.Chem.Soc.* **2016,** 138, 9663-9674.

Papp, D.; Rovo, P.; Jakli, I.; Csaszar, A. G.; Perczel, A., Four Faces of the Interaction between Ions and Aromatic Rings. *Journal of Computational Chemistry* **2017,** *38* (20), 1762-1773.

Ivanova, M. I.; Sievers, S. A.; Sawaya, M. R.; Wall, J. S.; Eisenberg, D., Molecular basis for insulin fibril assembly. *Proceedings of the National Academy of Sciences of the United States of America* **2009,** *106* (45), 18990-18995.

De La Paz, M. L.; Goldie, K.; Zurdo, J.; Lacroix, E.; Dobson, C. M.; Hoenger, A.; Serrano, L., De novo designed peptide-based amyloid fibrils. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99* (25), 16052-16057.

Evans, J. M.; Newton, R. W.; Ruta, D.A.; MacDonald, T. M.; Morris, A. D., Socio-economic status, obesity and prevalence of Type 1 and Type 2 diabetes mellitus. *Diabet Med J Br Diabet Assoc* **2000**, 17, 478–80.

Holman, N.; Young, B.; Gadsby, R., Current prevalence of Type 1 and Type 2 diabetes in adults and children in the UK. *Diabet Med J Br Diabet Assoc* **2015**, 32, 1119–20.

IDF Diabetes Atlas, Eight Edition, 2017 ISBN: 978-2-930229-87-4

Jahn, T. R.; Radford, S. E., : Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch Biochem Biophys.* **2008**, 469(1), 100–117.

Farkas, V.; Jákli, I.; Tóth, G. K.; Perczel, A., aromatic cluster sensor of protein folding: Near-UV electronic circular dichroism bands assigned to fold compactness. *Chemistry A European Juornal* **2016,** 22:(39), 13871-13883.

ref. Dobson 2000 körül

Greenfield, N.; Fasman, G. D., Computed Circular Dichroism Spectra for the Evaluation of Protein Conformation. *Biochemistry*, **1969**, 8(10), 4108–4116.

Umemoto, A.; Yagi, H.; So, M.; Goto, Y., High-throughput Analysis of Ultrasonication-forced Amyloid Fibrillation Reveals the Mechanism Underlying the Large Fluctuation in the Lag Time. *J Biol Chem*. **2014**, 289(39), 27290–27299.

Tartaglia, G.G.; Vendruscolo, M., The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev*. **2008**, 37(7), 1395-401.

Chi, E. Y.; Krishnan, S.; Randolph, T. W.; Carpenter, J. F., Physical Stability of Proteins in Aqueous Solution: Mechanism and Driving Forces in Nonnative Protein Aggregation. *Pharmaceutical Research* **2003**, 20(9), 1325-1336.

Munishkina, L. A.; Henriques, J.; Uversky, V. N.; Fink, A. L., Role of protein-water interactions and electrostatics in alpha-synuclein fibril formation. *Biochemistry* **2004**, 43(11), 3289-300.

Perczel, A.; Hudáky, P.; Pálfi, V.,Dead-end street of protein folding: thermodynamic rationale of amyloid fibril formation. *J. Am. Chem. Soc*. **2008**, 129(48), 14959-14965.

Lovell, S.C.; Davis, I.W.; Arendall III, W.B.; de Bakker, P.I.W.; Word, J.M.; Prisant, M.G.; Richardson J.S. and Richardson D.C., Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins: Structure, Function & Genetics*, **2002**, 50, 437-450.