

## Protein Folding | Very Important Paper |

## VIP The Route from the Folded to the Amyloid State: Exploring the Potential Energy Surface of a Drug-Like Miniprotein

Nóra Taricska,<sup>[a]</sup> Dániel Horváth,<sup>[a]</sup> Dóra K. Menyhárd,<sup>[a]</sup> Hanna Ákontz-Kiss,<sup>[a]</sup> Masahiro Noji,<sup>[b]</sup> Masatomo So,<sup>[b]</sup> Yuji Goto,<sup>[b]</sup> Toshimichi Fujiwara,<sup>[b]</sup> and András Perczel\*<sup>[a]</sup>

**Abstract:** The amyloid formation of the folded segment of a variant of Exenatide (a marketed drug for type-2 diabetes mellitus) was studied by electronic circular dichroism (ECD) and NMR spectroscopy. We found that the optimum temperature for E5 protein 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.;  $T=37^\circ\text{C}$ ,  $\text{pH } 4.1$ ,  $c_{\text{E5}}=250\ \mu\text{M}$ ,  $c_{\text{NaCl}}=50\ \text{mM}$ ,  $t>4\text{--}6\ \text{h}$ ) that will inevitably gravitate into the amyloid state. The strong B-type far ultraviolet (FUV)-ECD spectra and an unexpectedly strong near ultraviolet (NUV)-

ECD signal ( $\Theta_{\approx 275\text{--}285\ \text{nm}}$ ) indicate that the amyloid phase of E5 is built from monomers of quasi-elongated backbone structure ( $\phi\approx -145^\circ$ ,  $\psi\approx +145^\circ$ ) with strong interstrand Tyr $\leftrightarrow$ Trp interaction. Misfolded intermediates and the build-up of “toxic” early-stage oligomers leading to self-association were identified and monitored as a function of time. Results indicate that the amyloid transition is triggered by subtle misfolding of the  $\alpha$ -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  $\beta$ -structured amyloid nucleus, thus the process follows a nucleated growth mechanism.

## 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,<sup>[1–3]</sup> such illnesses as Alzheimer’s disease, type-2 diabetes mellitus, rheumatoid arthritis, or haemodialysis ass. amyloidosis.<sup>[4]</sup> From the pioneering work on lysozyme<sup>[5]</sup> and A $\beta$ (1–42),<sup>[6–8]</sup> the amyloid state of several misfolded proteins (e.g.,  $\beta$ 2-microglobulin,<sup>[9]</sup> crystallin,<sup>[10]</sup> tau protein,<sup>[11–13]</sup> the glucagon peptide hormone,<sup>[14]</sup> and insulin<sup>[15]</sup> among others) were charac-

terized. The general topology of such aggregates consists of protein segments adopting an extended backbone, interacting through  $\beta$ -edges. The association between the  $\beta$ -sheets thus formed is compact and specific; in most cases it excludes water molecules, leading to the formation of tightly stacked, “dry-zipper” nanostructures.<sup>[16–18]</sup> The state-of-the-art TEM, SAXS, cryo-EM and ssNMR techniques now allow full characterization of the aggregated end-state,<sup>[19,20]</sup> 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<sup>[1–3,21,22]</sup> and also represent the stage at which amyloidosis can still be reversed.<sup>[23]</sup>

The progress of self-association can often be followed by ThT fluorescence and DLS—best in combination<sup>[24]</sup>—reporting the accumulation of cross- $\beta$ -backbone (above a minimum size of ca. 10 nm or 4–6 aligned strands<sup>[25]</sup>) and the size distribution of the species present in the solution, respectively. However, to gain atomistic detail, molecular spectroscopy needs to be applied such as CD, IR or NMR spectroscopy.<sup>[26,27]</sup> In fact, CD spectroscopy can be used to monitor a full range of molecular transformations accompanying amyloidogenesis if the secondary structure content of the folded, intermediate, and amyloid states are distinct. There are notable examples that satisfy this description; namely, helical peptide hormones such as amylin or glucagon and a considerable number of peptide therapeutics<sup>[28]</sup> and since over 60% of all protein-protein interfaces—

[a] N. Taricska, D. Horváth, Dr. D. K. Menyhárd, H. Ákontz-Kiss, Prof. Dr. A. Perczel  
Laboratory of Structural Chemistry and Biology &  
MTA-ELTE Protein Modeling Research Group  
Eötvös Loránd University  
Pázmány Péter sétány 1A  
1117 Budapest (Hungary)  
E-mail: perczel.andras@ttk.elte.hu

[b] M. Noji, Dr. M. So, Prof. Dr. Y. Goto, Prof. Dr. T. Fujiwara  
Institute for Protein Research  
Osaka University  
3-2 Yamadaoka, Suita, Osaka 565-0871 (Japan)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:  
<https://doi.org/10.1002/chem.201903826>.

© 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

typical targets of drug design—also constitute helices,<sup>[29]</sup> 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<sup>[30]</sup> that also contains a well-folded  $\alpha$ -helix. We have discovered that this 25 residue long segment (E5: EEEAVRLYIQWLKEGGPSSGRPPPS) (Figure 1), comprising the entire interface needed for GLP-1 receptor binding,<sup>[31]</sup> can be turned into amyloid in a controlled, fully reproducible and tunable manner within a large range of protein concentrations ( $80 \mu\text{M} < c_{\text{protein}} < 800 \mu\text{M}$ ) 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 pro-

teins and helix-containing peptides. Beside its helical stretch, E5 contains a  $\beta$ -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.<sup>[32]</sup>

As E5 is small, both its chemical synthesis on a resin and bacterial expression in a fusion system is straightforward.<sup>[33]</sup> Given that its folded state is partially helical, its transition toward the amyloid phase results in a significant change in secondary structure content, which is easy to monitor by electronic circular dichroism (ECD) spectroscopy. Furthermore, E5 has an interacting Trp/Tyr residue pair within its hydrophobic core, enabling folding and refolding of the protein to be tracked by near ultraviolet (NUV)- in addition to far ultraviolet (FUV)-ECD. We also found the process to be quenchable; the amyloid formation can be suspended at any time by dropping the temperature, and restarted by a subsequent temperature rise. Furthermore, the moderate size of E5 allows detailed structure characterization both by NMR and modeling techniques.

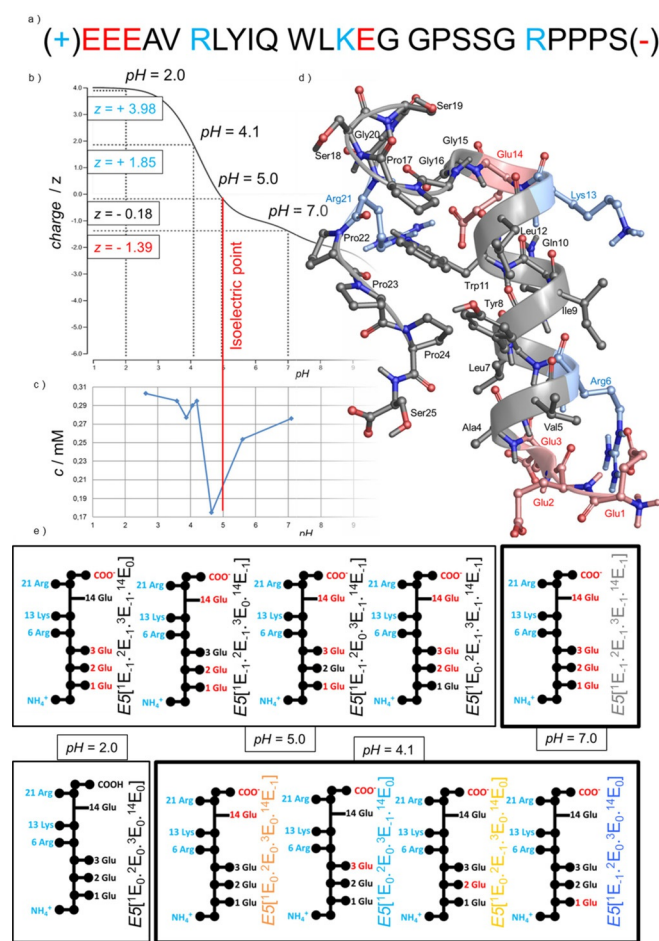
We probed various regions of the  $f(T, \text{pH}, c_{\text{protein}}, c_{\text{ion}}, t)$  potential energy hypersurface of E5 by acquiring quantitative NUV- and FUV-ECD chiroptical information and NMR data complemented by MD simulations to pinpoint the reaction path that leads from the fully folded- to the amyloid-state. Based on these results, we were able to propose a mechanism that resembles that of well-folded proteins but relies on special features of the miniprotein. The methodology presented here gains significance because the amyloid state of E5 is ThT silent and thus presents an approach for dealing with such cases as well.

## Results and Discussion

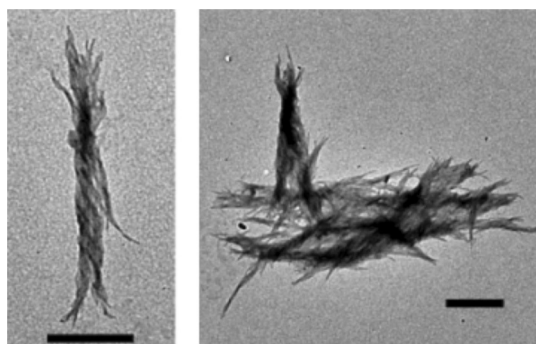
We have shown previously<sup>[34]</sup> that high concentrations ( $c \approx 10$ – $30 \text{ mM}$ ) of E5 trigger self-association, accompanied by an  $\alpha$ -helix to antiparallel  $\beta$ -sheet structural transition. Here, we set out to identify optimum conditions of aggregation, structurally 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.

To locate roughly and effectively those conditions that enable the amyloid formation, the HANABI system<sup>[35]</sup> was applied. The 96 well plates were set up with the following boundaries:  $2 < \text{pH} < 7$ ,  $80 \mu\text{M} < c_{\text{E5}} < 800 \mu\text{M}$  and  $0 \text{ mM} < c_{\text{NaCl}} < 100 \text{ mM}$  at  $T = 37^\circ\text{C}$ . ECD measurements were carried out after 90 h, with sonication cycle turned “on” for 1 min and repeated every 10 min (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 ( $c_{\text{NaCl}} = 100 \text{ mM}$ ), with the preferred protein concentration ranging from 80 to 160  $\mu\text{M}$ , 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



**Figure 1.** a) Primary sequence of E5 with residues colored by their charges of pH 7.0: negatively charged amino acid, positively charged amino acid, and neutral amino acid are highlighted with red, blue, and black, respectively. b) Overall charge ( $z$ ) of E5 as a function of the pH calculated by Prot pi (<http://www.protpi.ch>). c) The solubility of E5 as a function of the pH (at  $25^\circ\text{C}$ ) shows that near the isoelectric point reversible precipitation occurs. d) Folded structure of E5 ( $4^\circ\text{C}$  & pH 7.0) determined by NMR analysis, with its four basic groups (highlighted blue), and five acidic groups (highlighted red) with the following  $\text{pK}_a$  values: Ser<sub>C-terminal</sub>: 2.2, Glu<sub>side chain</sub>: 4.3, Glu<sub>N-terminal</sub>: 9.1, Lys<sub>side chain</sub>: 10.8, and Arg<sub>side chain</sub>: 12.5. e) The major microstates with their labeling are depicted schematically at four different pH values of interest, with side chains colored by charge: neutral (black), negative (red) and positive (blue).

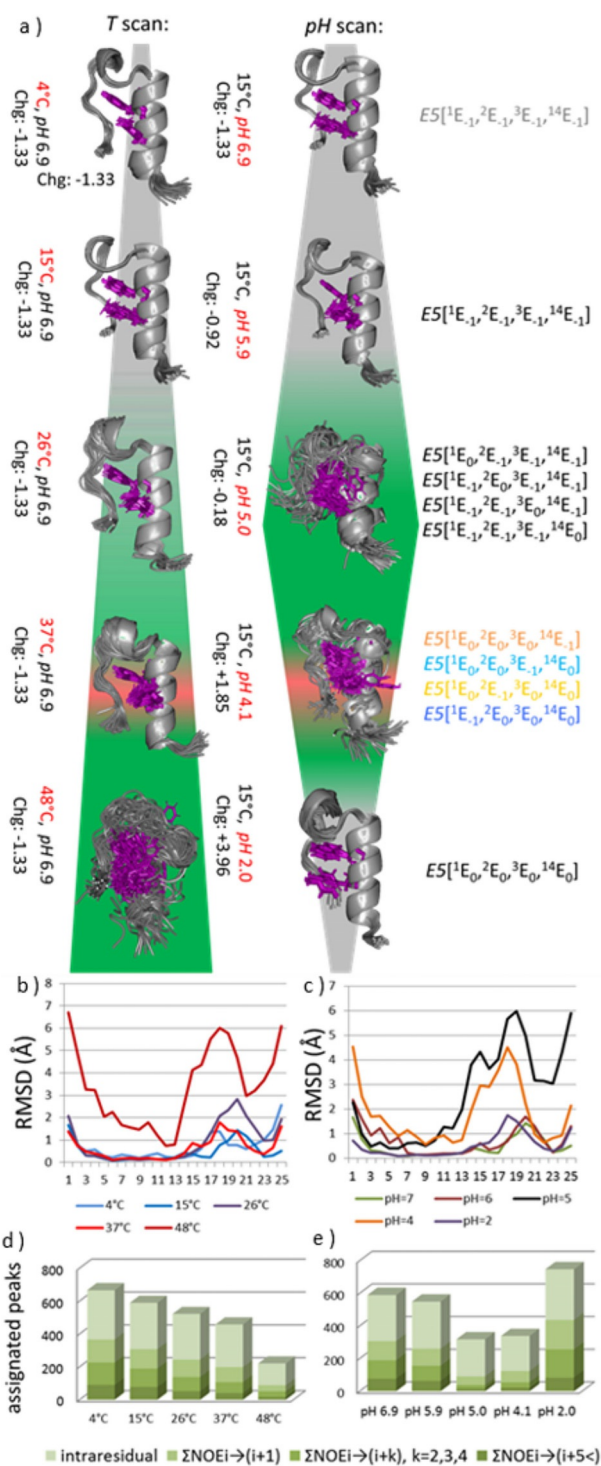


**Figure 2.** TEM images of E5 amyloid fibrils. Typical twisted amyloid fibrils were observed. Most fibrils were bundled and formed supercoiled structure. Scale bars: 500 nm.

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 varying the temperatures between 4 and 48 °C (at pH 7,  $c_{\text{protein}} \approx 0.8$  mM,  $c_{\text{ion}} \approx 0$  mM) leading to the primary conclusion that as temperature increases, partial unfolding occurs without amyloid formation. We found that although NMR frequencies shift with rising  $T$ , line broadening only takes place above 37 °C, indicating that considerable unfolding occurs only at 48 °C (Figure S1 in the Supporting Information). 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))<sup>[36,37]</sup> was revealed. The thermal unfolding of the protein backbone shows a nonlinear  $T$  dependence for Leu<sup>7</sup>, Ile<sup>9</sup>, and Lys<sup>13</sup> and, to a lesser extent, for Ala<sup>4</sup>, Val<sup>5</sup>, and Tyr<sup>8</sup>, indicating an enhanced presence of transient conformers at higher temperatures for the inner helix of E5 (Figure 1, Figure S2). 3D structure elucidations were also completed by acquiring a large number of NOE distance-restraints (Figure 3 and Table S1). Although the total number of restraints drops as  $T$  increases, 666 (4 °C)→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 3e and Table S1). Inter-residue NOEs associated with Y<sup>8</sup> and W<sup>11</sup> (Figure S3) 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, affording a single time-average and compact 3D structure for E5. While the central  $\alpha$ -helix is tightly folded at this temperature (RMSD of the backbone heavy atoms is 0.64 Å within the 50 best-fit structures) (Table S1), as  $T$  increases the unfolding of the -P<sup>17</sup>SSGRP<sup>22</sup>- segment was detected. This segment is the least structured area within the folded scaffold even at 4 °C,



**Figure 3.** a) Superimposed best 50 NMR structures of E5 at different pH values and temperatures (conditions and the net charges explicitly shown for each structure ensemble). The background color indicates the fold compactness: green—unfolded, gray—folded, red—misfolded leading to amyloid formation. Aromatic side chains of the key Y<sup>8</sup> and W<sup>11</sup> are highlighted in magenta. The basic (N-terminal; R<sup>9</sup>, K<sup>13</sup>, R<sup>21</sup>) moieties are protonated, the C-terminal (S<sup>25</sup>) deprotonated, whereas the protonation microstates of the 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 pH values are plotted. The total number of NOEs assigned for E5:  $\Sigma \text{NOE}^{i \rightarrow (i+1)}$ ,  $\Sigma \text{NOE}^{i \rightarrow (i+k)}$  where  $k = 2, 3, 4$ ,  $\Sigma \text{NOE}^{i \rightarrow (i+5 <)}$  are reported as d) the temperature was scanned at pH 6.9 and e) as pH was scanned at  $T = 15$  °C.



where only sequential NOEs could be measured (Figure 3c). Nevertheless, the large number of NOEs associated with Y<sup>8</sup> and W<sup>11</sup> residues and the -P<sup>22</sup>PP<sup>24</sup>- segment ensures their concerted motion, signaling that the hydrophobic core remains well-structured at 37 °C and below (Figure S3).

Due to exchange phenomena, NMR spectroscopy cannot be used 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.<sup>[38–40]</sup> 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 to 35 °C), a 50:50% mixture is reached near 60 °C, while at 85 °C the spectrum indicates a 70% unfolded content (Figure S4).

### The pH-scan

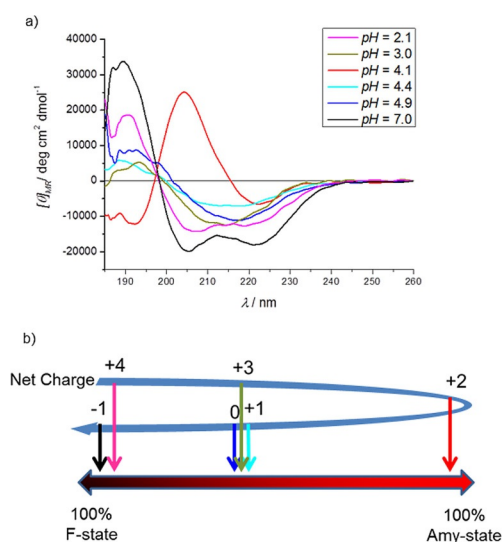
2D-NOESY driven 3D structure elucidation was completed (*c*<sub>E5</sub> ≈ 500–1500 μM, *c*<sub>NaCl</sub> < 10 mM) at various pH values (6.9, 5.9, 5.0, 4.1, 2.0) at *T* = 15 °C (Figure 3b) (The poorer signal to noise ratio meant that longer measurement time was required at pH 5.0 and 4.1). The solubility 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 also change. E5 contains a protonated N-terminal (p*K*<sub>a</sub> = 9.1) plus three basic residues, Arg<sup>6</sup> (p*K*<sub>a</sub> = 12.5), Lys<sup>13</sup> (p*K*<sub>a</sub> = 10.8) and Arg<sup>21</sup> (p*K*<sub>a</sub> = 12.5), with an acidic C-terminal (p*K*<sub>a</sub> = 2.2) and four acidic glutamines (Glu<sup>1</sup>, Glu<sup>2</sup>, Glu<sup>3</sup> and Glu<sup>14</sup>) (p*K*<sub>a</sub> ≈ 4.25) (the listed p*K*<sub>a</sub> values are nominal values that strongly depend on backbone conformation). The net charge of E5 is predicted to be positive at pH values smaller than 5 (Figure 1). At pH 7.0, two salt bridges may contribute to the stabilization of the 3D-fold, those of Glu<sup>1</sup>(-) ↔ Arg<sup>6</sup>(+) and Glu<sup>14</sup>(-) ↔ Arg<sup>21</sup>(+). 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 restraints from 666 to 319; Figure 3f and Table S1). 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 (Table S1). We found that the total number of NOESY cross-peaks between R<sup>21</sup> and W<sup>11</sup> residues is a reliable measure of the Trp-cage fold compactness (Figure S3): 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 3d). The structure loosening effect of the pH drop is the most pronounced between residues 13 and 21: the <sub>3</sub><sub>10</sub>-helix (-G<sup>15</sup>GPSSG<sup>20</sup>-) tends to unwind, exposing both the Y<sup>8</sup> and W<sup>11</sup> core residues to external water molecules (Figure 3). Interestingly, moving beyond the isoelectric point, ordering of the ensembles takes place and the original F-state

reappeared at pH 2.0. At pH 6.9, 77 long-range NOE<sup>*i*→(*i*+5<)</sup> restraints were assigned, while at pH 2.0 in total 84 NOEs<sup>*i*→(*i*+5<)</sup> 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 pH values. Meanwhile, in between, at pH 4.9 only 22 NOEs of this kind were recorded (Figure 3b). In conclusion, the pH-scan shows that the basic topological features of the Trp-cage fold of E5 (and other analogues<sup>[41]</sup>) 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 (Figure S5).

### 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, which is indicative of self-association leading to the loss of monomeric form or weakly bound low molecular weight associates in the solution (Figure S6). Stirring, however, greatly speeds up both the nucleation and fibril maturation stages of amyloidogenesis. This affects aggregation by increasing fragmentation and hence the number of free ends to support further fibril growth and possibly by increasing the number of collisions occurring between monomers and/or small oligomeric clusters.<sup>[42,43]</sup>

Stirring the solution of E5 for 51 h (pH 4.1) resulted in a strong B-type FUV-ECD spectrum, indicating a predominantly β-pleated backbone structure.<sup>[44,45]</sup> When setting the pH to 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, and amyloid forma-



**Figure 4.** a) FUV-ECD spectra of E5 recorded after 51 h of incubation time at constant stirring (*c*<sub>E5</sub> = 250 μM, *c*<sub>NaCl</sub> = 50 mM, *T* = 37 °C). If Σ*Q*<sub>E5</sub><sup>pH4.1</sup> = +2, then a B-type ECD is detected, which is indicative of a β-sheet backbone structure. However, if pH 7.0 (Σ*Q*<sub>E5</sub><sup>pH7.0</sup> = -1) or pH 2.1 (Σ*Q*<sub>E5</sub><sup>pH2.1</sup> = +4.0) E5 remains folded (predominantly α-helical). b) Spectral deconvolution enables the approximate [F] and [Amyloid] ratios to be determined as a function of the net molecular charge: Σ*Q*<sub>E5</sub>.

tion is less complete. At pH 7 and pH 2, pure C-type ECD spectra were measured (even when stirring), indicating that only  $\alpha$ -helical backbone structures are in solution and thus, these pH values prevent amyloid formation. To quantify the extent of the transition from the folded toward the amyloid-like phase, deconvolution of a large collection of ECD spectra were carried out (including  $T$ -dependent ECD curves of E5 and E0 as unfolded model mini-protein<sup>[37]</sup>). In this way, we obtained three pure component spectra: A C-type for  $\alpha$ -helical conformation, a U-type corresponding to the unstructured and a B-type signaling  $\beta$ -stranded backbone conformation. More interestingly, a similar NUV-ECD spectrum analysis (see below) shows that this transition is more complex than a simple  $\alpha$ - to  $\beta$ -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.

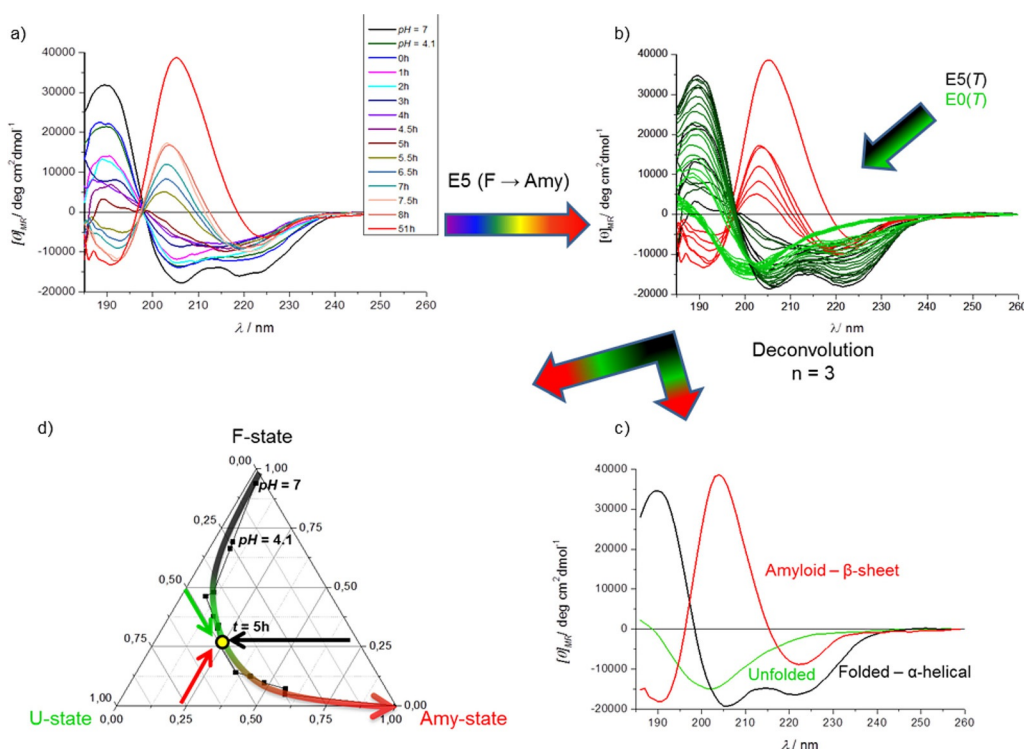
The overall path from the F- to the amyloid state is reported using a barycentric coordinate system (Figure 5) in which the gradual maturation of the amyloid as a function of time is visualized. At any point along the route, the ratio of the folded, unfolded, and amyloid states can be calculated from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

By 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  $\rightarrow$  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 runs parallel to the folded  $\rightarrow$  amyloid axis, corresponding to the nucleation phase. During this phase a critical concentration of misfolded structures is reached, while the misfolded content remains nearly constant. F-state diminishes while Amy-states start to accumulate, iii) The third phase runs parallel to the unfolded  $\rightarrow$  amyloid axis, where no further reversible unfolding takes place: the misfolded/unfolded structures are trapped by the growing amyloids, which is the elongation phase of the E5 amyloidosis.

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

Although E5 can form amyloid at low salt ( $c_{\text{NaCl}} < 1\text{ mM}$ ) and high protein concentrations ( $c_{\text{E5}} > 5\text{--}15\text{ mM}$ ),<sup>[34]</sup> under these extreme conditions amyloid formation is poorly reproducible. To locate physiologically more relevant and controllable amyloidogenic conditions on the  $\Delta G = f(T, \text{pH}, c_{\text{protein}}, c_{\text{ion}}, t)$  surface, both  $c_{\text{E5}}$  and  $c_{\text{NaCl}}$  variables had to be optimized at a sensible concentration range:  $250 < c_{\text{E5}} < 800\ \mu\text{M}$  and  $12.5 < c_{\text{NaCl}} < 50\text{ mM}$  with  $T$  & pH set as:  $T = 37^\circ\text{C}$  & 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,  $c_{\text{E5}} = 250\ \mu\text{M}$  and  $c_{\text{NaCl}} = 50\text{ mM}$ ,  $c_{\text{E5}} = 500\ \mu\text{M}$  and  $c_{\text{NaCl}} =$



**Figure 5.** a) The C-type FUV-ECD of E5 (F-state) gradually shifts as a function of time, as amyloid formation progresses ( $c_{\text{E5}} = 250\ \mu\text{M}$ ,  $c_{\text{NaCl}} = 50\text{ mM}$ , pH 4.1, stirring at  $T = 37^\circ\text{C}$ ). The final B-type spectrum stands for the amyloid state ( $t = 51\text{ h}$ ) composed of  $\beta$ -strands. b) An ensemble of the FUV-ECD spectra of E5 forming amyloid (Panel A) was formed with those of the  $T$ -dependent E0<sup>[37]</sup> and E5 ( $5 < T(^{\circ}\text{C}) < 85$ ) and deconvoluted into three pure component spectra c) standing for the F (black), U (green), and amyloid (red) states. d) The CCA+ resulted in the relative coefficients of these three pure ECD spectra and depicts the amyloid formation path in a Barycentric coordinate system: for example after 5 h the ratio is 26% F, 50% U, and 24% Amy state (yellow dot) (c.f. Figure S7).

25 mM,  $c_{E5} = 800 \mu\text{M}$  and  $c_{\text{NaCl}} = 12.5 \text{ mM}$  (with the solution constantly stirred) (Figure S8).

As protein concentration increases, amyloid formation becomes more complete:  $[\text{Amy}]^{c_{E5}=400 \mu\text{M}} (t=8 \text{ h}) = 0.34$ ,  $[\text{Amy}]^{c_{E5}=800 \mu\text{M}} (t=8 \text{ h}) = 0.61$  even when salt concentration is low and fixed at  $c_{\text{NaCl}} = 12.5 \text{ mM}$  (a condition more suitable for NMR) (Figure 6a, Figure S9).

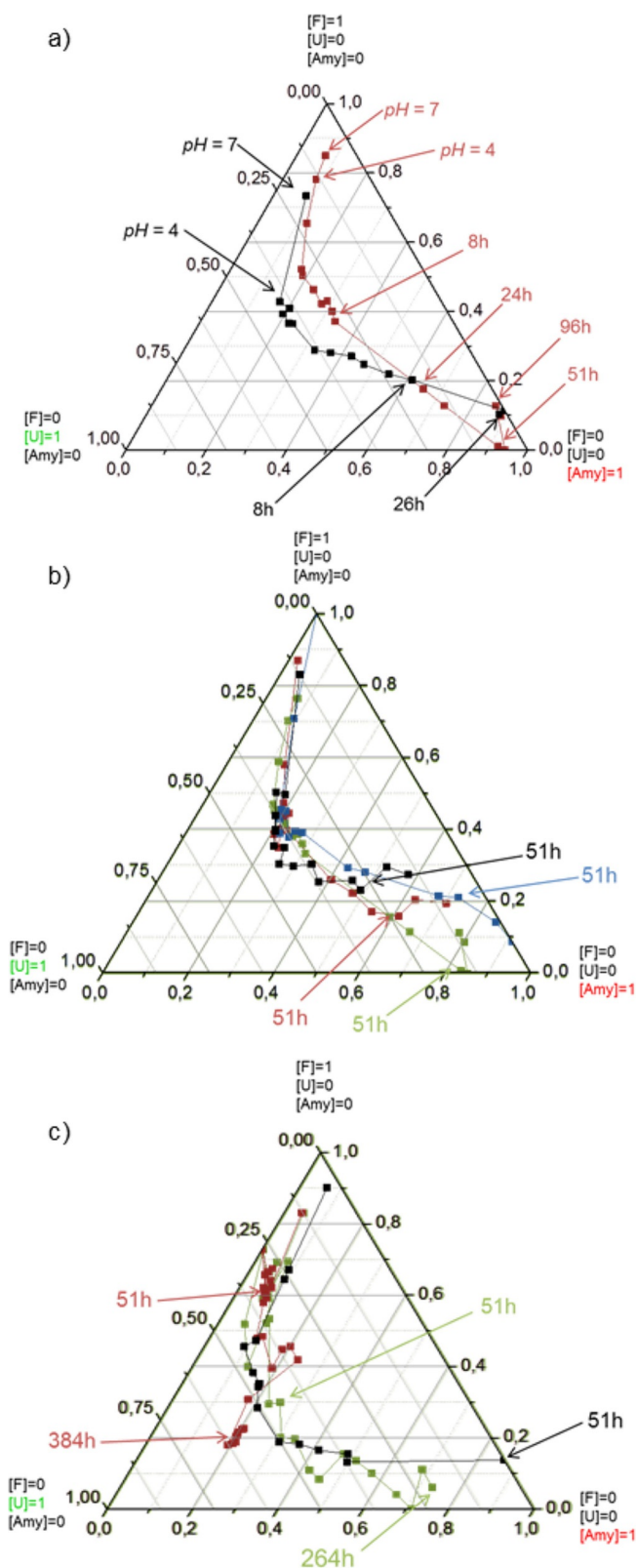
The rate of amyloid formation was also probed at a fixed protein concentration (e.g.,  $c_{E5} = 400 \mu\text{M}$ ) with increasing salt concentration ( $c_{\text{NaCl}}$ : 0 mM, 12.5 mM, 25 mM, 50 mM) (Figure 6b, Figure S10). We found that both misfolding and nucleation occurs at a similar rate. However, during the elongation phase clear differences were detected as a function of  $c_{\text{NaCl}}$ . Upon stirring for three days ( $t = 76 \text{ h}$ ) the amyloid ratio was found to be higher if salt concentration was lower:  $[\text{Amy}]^{c_{\text{NaCl}}=0 \text{ mM}} = 0.85$ ,  $[\text{Amy}]^{c_{\text{NaCl}}=12.5 \text{ mM}} = 0.78$ ,  $[\text{Amy}]^{c_{\text{NaCl}}=25 \text{ mM}} = 0.63$ ,  $[\text{Amy}]^{c_{\text{NaCl}}=50 \text{ mM}} = 0.52$ . 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 ( $< 100 \text{ mM}$ ), 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.<sup>[46–48]</sup> 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 0.15 M NaCl solution) ca. 10% of the all ion-protein interactions involved the charged residues of the  $\text{Glu}^{14} \leftrightarrow \text{Arg}^{21}$  salt bridge, which might contribute to the loosening of the hydrophobic core of E5 (Figure S11), 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 ( $c_{E5} = 250 \mu\text{M}$ ,  $c_{\text{NaCl}} = 50 \text{ mM}$  and pH 4.1) (Figure 6c, Figure S12). We found that the process is slower and incomplete both at  $23^\circ\text{C}$  and  $47^\circ\text{C}$ , compared to that of the physiological temperature ( $37^\circ\text{C}$ ). It seems that at a high temperature the increased thermal motion disfavors self-association:  $[\text{Amy}]^{47^\circ\text{C}, 51 \text{ h}} = 0.34 \ll [\text{Amy}]^{37^\circ\text{C}, 51 \text{ h}} = 0.86$ . At temperatures too low, the Brownian motion is significantly slower and thus fewer collisions occur and the misfolding propensity of E5 is lowered (see NMR data above):  $[\text{Amy}]^{23^\circ\text{C}, 51 \text{ h}} = 0.12$ .

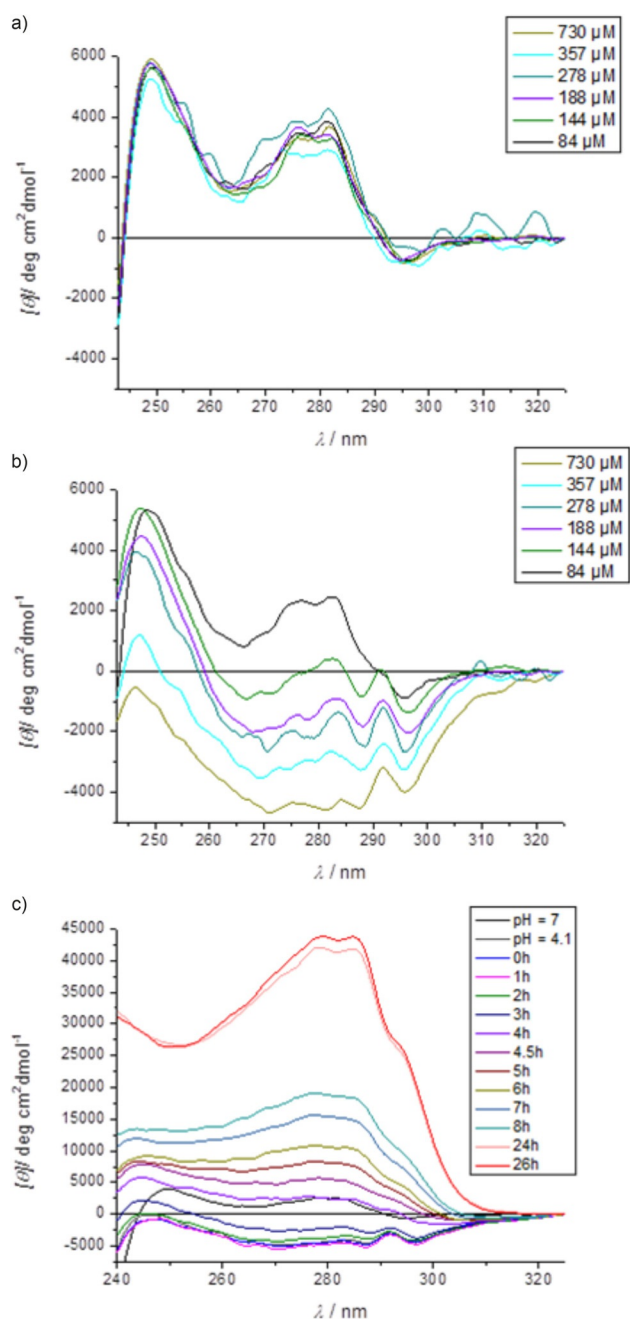
### Verifying amyloidosis: NUV-ECD measurements

Amyloid transformation was monitored from the viewpoint of the  $\text{Y}^8 \leftrightarrow \text{W}^{11}$  aromatic interaction by acquiring NUV-ECD data. At pH 6 the characteristic positive band at 283 nm (over the entire tested protein concentration range) is indicative of a shifted face-to-face  $\pi$ - $\pi$  interaction between the two aromatic rings of the hydrophobic core of E5 (Figure 7a). However, at



**Figure 6.** Amyloid formation completeness monitored by far-UV ECD as a function of time with a) different protein (black: 800, claret: 400  $\mu\text{M}$ ) but same ion concentration (12.5 mM); b) same protein (400  $\mu\text{M}$ ) and different ion concentration (black: 50, claret: 25, green: 12.5, blue: 0 mM); c) same protein (250  $\mu\text{M}$ ) and the same ion concentration (50 mM) at different temperatures:  $T = 23$  (claret),  $37$  (black), and  $47^\circ\text{C}$  (green).





**Figure 7.** NUV-ECD (240–325 nm) spectra reveal 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  $\pi$ - $\pi$  interaction profile changes as protein concentration increases ( $84 < c_{E5} < 730 \mu\text{M}$  at  $25^\circ\text{C}$ ,  $c_{\text{NaCl}} \approx 0 \text{ mM}$ ), whereas it remains unchanged a) at pH 6.0. c) NUV-ECD spectral shift signals indicating the changing interaction mode of  $Y^8$  &  $W^{11}$  as function of the time ( $0 < t < 26 \text{ h}$ ) during amyloid formation ( $c_{E5} = 800 \mu\text{M}$ ,  $c_{\text{NaCl}} = 12.5 \text{ mM}$ , pH 4.1, stirring at  $T = 37^\circ\text{C}$ ).

pH 4.1 negative bands were detected in this spectral region ( $270 < \lambda < 290 \text{ nm}$ ) (Figure 7b) and assigned to an edge-to-face  $\pi$ - $\pi$  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 at a slightly shifted posi-

tion 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 ( $\theta_{\approx 283 \text{ nm}}$ :  $4000 \rightarrow 40000$ ) (Figure 7a and c). This side chain restructuring coincides with the backbone changing from an  $\alpha$ - to  $\beta$ -state. Thus, we propose that the enhanced positive NUV-ECD bands between  $270 < \lambda < 290 \text{ nm}$  no longer arise from a pairwise, intramolecular shifted face-to-face  $\pi$ - $\pi$  interaction, but rather from the interstrand interaction of Tyr and Trp side chains packed tighter in the supramolecular assembly of the amyloid phase.

### The nature of the misfolded states

Amyloid formation proceeds at physiological temperature, but it can be quenched if cooled to  $4^\circ\text{C}$ . Thus, a series of heteronuclear correlation spectra ( $^1\text{H}$ - $^{15}\text{N}$ -HSQC) were recorded at  $4^\circ\text{C}$  using samples retrieved at regular time intervals (every 30 min) during amyloid formation (Figure S13).

The  $^1\text{H}$ ,  $^{15}\text{N}$ -chemical shifts of all residues were calculated by using Equation (1):

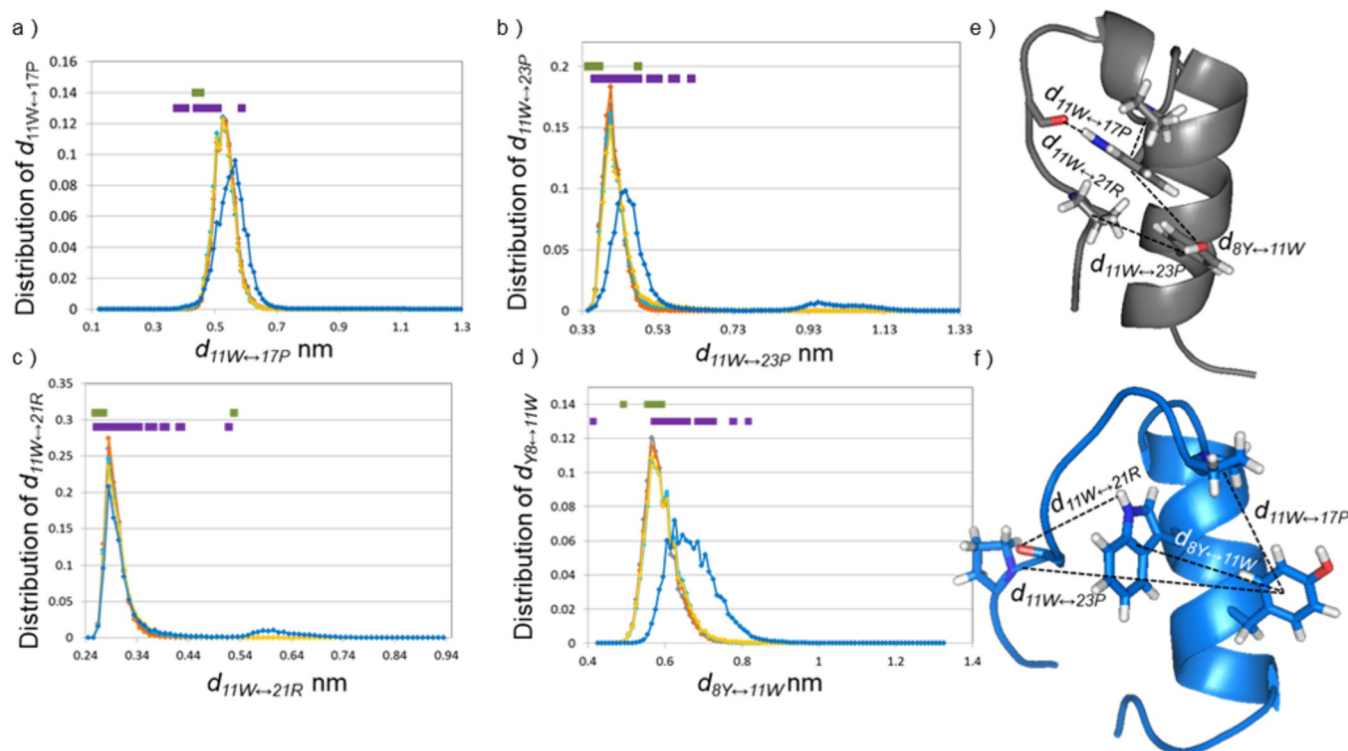
$$\Delta(t) = \sqrt{[(\delta\text{H}(t))^2 + (\delta\text{N}(t)/6.51)^2]} \quad (1)$$

as function of the time ( $0 < t < 24 \text{ h}$ ). Residues of larger changes than the average chemical shift ( $t_{\text{end}} - t_0 > 0.047 \text{ ppm}$ ) are those of the  $\alpha$ -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 as well (Figure S2), 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<sup>6</sup>LYIQWL<sup>12</sup>-) was also predicted to be the most amyloidogenic of the sequence by CamSol.<sup>[49]</sup> Therefore, we propose that the  $\alpha$ -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 monomers can interact via their exposed hydrophobic cores.

Amyloid formation is optimal at pH 4.1, where the overall charge of E5 is near +2 (Figures 1, 2, and 3) and thus, on average three out of the four Glu side chains are protonated. To identify the most likely protonation state 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 four ( $E5[{}^1E_0{}^2E_0{}^3E_0{}^{14}E_0]$ ;  $E5[{}^1E_0{}^2E_0{}^3E_0{}^{14}E_0]$ ;  $E5[{}^1E_0{}^2E_{-1}{}^3E_0{}^{14}E_0]$ ;  $E5[{}^1E_{-1}{}^2E_0{}^3E_0{}^{14}E_0]$ ) 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 8, Figure S14).

The simulations of four microstates provided equilibrium ensembles quite similar to that of the folded structure derived from NMR data: those of  $E5_{\text{pH } 4.1}[{}^1E_0{}^2E_0{}^3E_0{}^{14}E_{-1}]$ ,  $E5_{\text{pH } 4.1}[{}^1E_0{}^2E_0{}^3E_{-1}{}^{14}E_0]$  and  $E5_{\text{pH } 4.1}[{}^1E_0{}^2E_{-1}{}^3E_0{}^{14}E_0]$ . In contrast, a single microstate, that of  $E5_{\text{pH } 4.1}[{}^1E_{-1}{}^2E_0{}^3E_0{}^{14}E_0]$  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

	Measure type	Description of the measure	Critical value of the measure
i	$d_{11W \rightarrow 17P}$	distance between Trp <sup>11</sup> & Pro <sup>17</sup>	$d_{11W \rightarrow 17P} > 5.7 \text{ \AA}$
ii	$d_{11W \rightarrow 23P}$	distance between Trp <sup>11</sup> & Pro <sup>23</sup>	$d_{11W \rightarrow 23P} > 4.7 \text{ \AA}$
iii	$d_{11W \rightarrow 21R}$	distance between NH of Trp <sup>11</sup> and CO of Arg <sup>21</sup>	$d_{11W \rightarrow 21R} > 6.0 \text{ \AA}$
iv	$d_{14E \rightarrow 21R}$	distance between donor of Arg <sup>21</sup> and acceptor of Glu <sup>14</sup>	
v	$\mu$	torsion angle between the main axis of the core $\alpha$ -helix and of the polyPro of the C terminus	$\mu > 45^\circ$
vi	$d_{8Y \rightarrow 11W}$	distance between Tyr <sup>8</sup> & Trp <sup>11</sup>	$d_{8Y \rightarrow 11W} > 7.0 \text{ \AA}$
vii	$\alpha$	angle between the planes of the aromatic rings	$\alpha > 70^\circ$
			$\theta > 65^\circ$
viii	$\theta$	elevation of the aromatic rings	$40 > \theta > 50^\circ$
			$80 > \theta > 90^\circ$
ix	$\phi$	azimuthal angle of the aromatic rings	$\phi > 320^\circ$



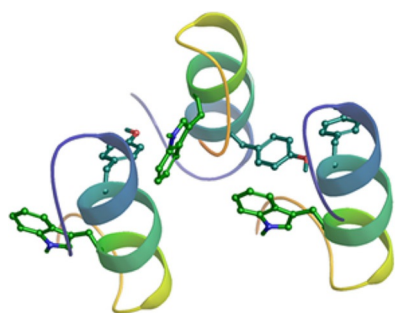
**Figure 8.** The distributions of conformational measures introduced to signal the degree of misfolding of E5. a–d) The distribution of the selected measures for the MD trajectories (gray: pH 7, E5[<sup>1</sup>E<sub>-1</sub>,<sup>2</sup>E<sub>-1</sub>,<sup>3</sup>E<sub>-1</sub>,<sup>14</sup>E<sub>-1</sub>]; orange: pH 4.1 E5[<sup>1</sup>E<sub>0</sub>,<sup>2</sup>E<sub>0</sub>,<sup>3</sup>E<sub>0</sub>,<sup>14</sup>E<sub>-1</sub>]; light blue: pH 4.1 E5[<sup>1</sup>E<sub>0</sub>,<sup>2</sup>E<sub>0</sub>,<sup>3</sup>E<sub>0</sub>,<sup>14</sup>E<sub>0</sub>]; yellow: pH 4.1 E5[<sup>1</sup>E<sub>0</sub>,<sup>2</sup>E<sub>-1</sub>,<sup>3</sup>E<sub>0</sub>,<sup>14</sup>E<sub>0</sub>]; blue: pH 4.1 E5[<sup>1</sup>E<sub>-1</sub>,<sup>2</sup>E<sub>0</sub>,<sup>3</sup>E<sub>0</sub>,<sup>14</sup>E<sub>0</sub>]) are plotted, whereas the distribution of the selected measures for NMR structures (green: pH 7, purple: pH 4.1) are indicated for every structure above the MD distribution curves in a linear 1D manner. The measure of the E5[<sup>1</sup>E<sub>-1</sub>,<sup>2</sup>E<sub>0</sub>,<sup>3</sup>E<sub>0</sub>,<sup>14</sup>E<sub>0</sub>] microstate is slightly different from the other microstates; moreover, a minor population of the E5[<sup>1</sup>E<sub>-1</sub>,<sup>2</sup>E<sub>0</sub>,<sup>3</sup>E<sub>0</sub>,<sup>14</sup>E<sub>0</sub>] microstate shows considerably distinct properties. Table 1 contains a limit for each measure that can be used to distinguish/separate the divergent/misfolded population of the E5[<sup>1</sup>E<sub>-1</sub>,<sup>2</sup>E<sub>0</sub>,<sup>3</sup>E<sub>0</sub>,<sup>14</sup>E<sub>0</sub>] microstate. e) NMR calculated folded structure (gray), 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 Figure S15.

that in case of E5<sub>pH 4.1</sub>[<sup>1</sup>E<sub>-1</sub>,<sup>2</sup>E<sub>0</sub>,<sup>3</sup>E<sub>0</sub>,<sup>14</sup>E<sub>0</sub>] both  $d_{11W \rightarrow 17P}$  and  $d_{11W \rightarrow 23P}$  distances are lengthened by 0.4 and 0.3 Å, respectively, and in some conformers the  $d_{11W \rightarrow 21R}$  distance shifted from 2.85 to 5.85 Å, indicating the appearance of less compact protein folds. Furthermore, as the distribution of  $\mu$  gets wider, the -PPP<sup>24</sup>- segment twists more often with respect to the main axes of the  $\alpha$ -helix. In parallel,  $d_{8Y \rightarrow 11W}$  increases by 0.9 Å. The relative orientation of the aromatic side chains changes as well: measured  $\theta$  shows a more diverse distribution relative to

the other microstates, while the values of  $\alpha$  and  $\phi$  increase slightly. Summing up, we might say that E5<sub>pH 4.1</sub>[<sup>1</sup>E<sub>0</sub>,<sup>2</sup>E<sub>0</sub>,<sup>3</sup>E<sub>0</sub>,<sup>14</sup>E<sub>-1</sub>] 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 unfolded, giving rise to unshielded aromatics and exposed backbone amide groups, ready for self-association and subsequent amyloid formation (Figure S14), 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 side chains create a “sticky” interaction center for E5, quite as the free  $\beta$ -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 9). These findings also explain why stirring is necessary for amyloid formation of E5; given that 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 9.** 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.

### Building a model of the amyloid state

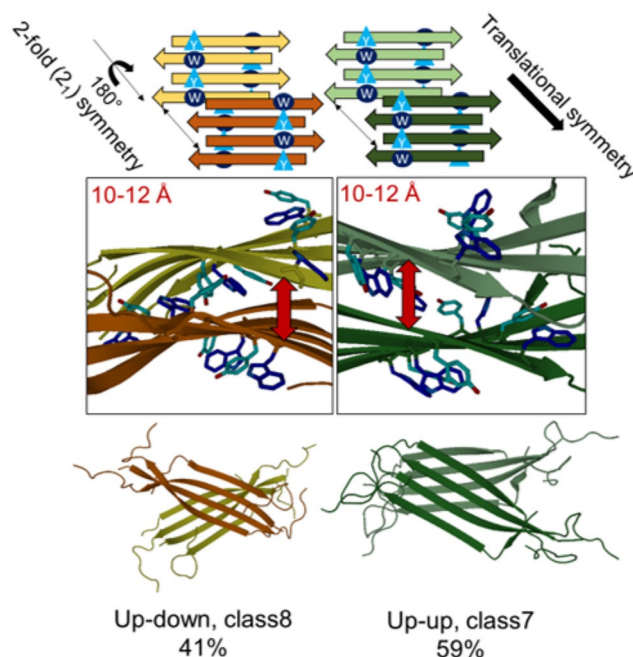
To obtain a picture of the amyloid state, MD simulations were carried out for models of four- $\beta$ -strand clusters ( $E5$ )<sub>4</sub> using three different protonation motifs: ( $E5_{pH 2}[{}^1E_0, {}^2E_0, {}^3E_0, {}^{14}E_0, C\text{-term}_0]_4$ ), ( $E5_{pH 4.1}[{}^1E_{-1}, {}^2E_0, {}^3E_0, {}^{14}E_0]_4$ ) and ( $E5_{pH 7}[{}^1E_{-1}, {}^2E_{-1}, {}^3E_{-1}, {}^{14}E_{-1}]_4$ ) (four-stranded models were shown to be sufficiently large for modeling aggregation nuclei in various different systems<sup>[50–52]</sup>). We have probed three different relative offsets, namely that of “YI”, “YK”, and “YW” (Figure S16A, B, and C) for each.

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) (Figure S17). Considering that in the -PPP<sup>24</sup>-segment  $\phi^{pro}$  must be ca.  $-60^\circ$  and thus this part of the sequence is unlikely to form an extended  $\beta$ -sheet structure (with  $\phi \approx -145^\circ$ ,  $\psi \approx +145^\circ$ ), 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 adja-

cent  $\beta$ -strands, as they come in close proximity ( $d_{8Y-11W} < 6 \text{ \AA}$  as measured between the center of the rings) (Figure S16D). The aromatic ladder enables the formation of the shifted face-to-face  $\pi$ - $\pi$  interaction expected based on ECD spectroscopic data. Interestingly, the derived orientation and spacing of the Tyr and Trp side chains here is also quite reminiscent of the aromatic clusters created by mutations within the central single-layer  $\beta$ -sheet of *Borrelia* outer surface protein A (OspA) to probe the structure-ordering capacity of such residues.<sup>[53]</sup>

Using the mid-structures of six different clusters (see Methods) of the MD simulation of ( $E5_{pH 4.1}$ )<sub>4</sub> YW as possible amyloid seeds, we also probed different inter-sheet arrangements by using a protein-protein docking algorithm. The near-parallel arrangements thus obtained could be grouped into two fundamentally different topologies (using the nomenclature introduced by Eisenberg and co-workers<sup>[54]</sup>): those of class 7 (equifacial, antisymmetric, up-up) with translational symmetry, and class 8 (antiparallel, equifacial, antisymmetric, up-down) with twofold symmetry (Figure 10).

In the derived  $\beta$ -sheet-dimers,  $((E5_{pH 4.1})_4)_2$ , the average distance between the  $\beta$ -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.<sup>[55–59]</sup> Furthermore, in this sandwich form, the intersheet aromatic side chains may approach each other even closer than the intrasheet 6 Å described above (Figure S18). We found that



**Figure 10.** The two basic topologies found by docking ( $E5$ )<sub>4</sub> amyloid seeds are the antiparallel equifacial antisymmetric up-down (class 8) with a two-fold symmetry and equifacial antisymmetric up-up (class 7) 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  $\beta$ -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  $\beta$ -sheets.

(E5<sub>pH 4.1</sub>)<sub>4</sub> 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.<sup>[60]</sup> These findings further support our model for the aggregation of E5 as nucleated by the antiparallel oriented  $\beta$ -sheets formed between the most aggregation-prone segment of the sequence, the -V<sup>5</sup>RLYIQWLK<sup>13</sup>- unit.

## Conclusions

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  $\alpha$ -helix to coil and then to extended  $\beta$ -strand transition, which seems to be highly unfavorable for the monomeric forms but eventually leads to the formation of a  $\beta$ -structured amyloid nucleus. The amyloidosis of E5 thus follows the nucleated growth mechanism.<sup>[42,43]</sup>

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 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.

## Experimental Section

**Protein expression and purification:** The E5 miniprotein was produced by bacterial expression using the previously published protocol.<sup>[33]</sup>

**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) and incubated at target temperatures. At a given time, 10-fold diluted sample was measured with ECD spectroscopy. The concentration of the diluted sample was determined with a NanoDrop Lite Spectrophotometer (Thermo Scientific™) at 280 nm.

**NMR experiments:** Datasets were collected with a 16.4 T Bruker Avance III spectrometer equipped with a 5 mm inverse TCI probehead with z-gradient. Spinlock (d9) for <sup>1</sup>H-<sup>1</sup>H TOCSY was 80 ms, while the mixing time (d8) of 150 ms was taken for <sup>1</sup>H-<sup>1</sup>H NOESY spectra. NMR structure calculations were performed and refined by cooperative use of CcpNmr Analysis 2.4.1., Aria 2.0<sup>[61]</sup> and CNS Solve 1.2.<sup>[62]</sup>

**Molecular dynamics simulations:** MD simulation were carried out as implemented in GROMACS59, using the AMBER-ff99SBildnp\* force field. The systems were solvated with TIP3P water molecules in dodecahedral boxes with a size allowing 10 Å between any protein atom and the box. Trajectories of 600 ns ((E5)<sub>4</sub> models) to 1000 ns (E5 monomers) NPT simulations with a 2 fs time step at 310 K and 328 K and 1 bar were collected (with snapshots at every 4 ps).

**Electronic circular dichroism spectroscopy:** Far- and near-UV ECD measurements were carried out with a Jasco J810 spectrophotometer. The temperature at the cuvette was controlled with a Peltier-type heating system.

## Acknowledgements

The authors gratefully acknowledge Viktor Farkas for his contribution. NMR spectrometer measurement time (700 MHz Bruker) was courtesy of MedInProt Grant Facilitating Access to Instruments from the Hungarian Academy of Sciences. Calculations were carried out at the NIIF Supercomputing Center of KIFU (Hungary). This research project was supported by the European Union and the State of Hungary and co-financed by the European Regional Development Fund (VEKOP-2.3.2-16-2017-00014), and the K116305 OTKA grant of the NKFIH of the Hungarian Academy of Sciences. This work was in part performed under the International Collaborative Research Program of Institute for Protein Research, Osaka Univ. (ICR-17-02). We also acknowledge the generous financial support by Sanofi (Sanofi Doctoral Foundation Program).

## Conflict of interest

The authors declare no conflict of interest.

**Keywords:** aggregation · amyloid beta-peptides · protein folding · protein modeling

- [1] T. Kim, G. S. Vidal, M. Djuricic, C. M. William, M. E. Birnbaum, K. C. Garcia, B. T. Hyman, C. J. Shatz, *Science* **2013**, *341*, 1399–404.
- [2] G. M. Shankar, B. L. Bloodgood, M. Townsend, D. M. Walsh, D. J. Selkoe, B. L. Sabatini, *J. Neurosci.* **2007**, *27*, 2866–75.
- [3] J. Laurén, D. A. Gimbel, H. B. Nygaard, J. W. Gilbert, S. M. Strittmatter, *Nature* **2009**, *457*, 1128–32.
- [4] C. Bussi, J. M. Peralta Ramos, D. S. Arroyo, J. I. Gallea, P. Ronchi, A. Kolovou, J. M. Wang, O. Florey, M. S. Celej, Y. Schwab, N. T. Ktistakis, P. Iribarren, *J. Cell Sci.* **2018**, *131*, jcs226241.
- [5] D. R. Booth, M. Sunde, V. Bellotti, C. V. Robinson, W. L. Hutchinson, P. E. Fraser, P. N. Hawkins, C. M. Dobson, S. E. Radford, C. C. F. Blake, M. B. Pepys, *Nature* **1997**, *385*, 787–793.
- [6] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, R. Tycko, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16742–16747.

- [7] L. Gremer, D. Scholzel, C. Schenk, E. Reinartz, J. Labahn, R. B. G. Ravelli, M. Tusche, C. Lopez-Iglesias, W. Hoyer, H. Heise, D. Willbold, G. F. Schröder, *Science* **2017**, *358*, 116–119.
- [8] M. A. Wälti, F. Ravotti, H. Arai, C. G. Glabe, J. S. Wall, A. Böckmann, P. Güntert, B. H. Meier, R. Riek, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E4976–E4984.
- [9] K. Iwata, T. Fujiwara, Y. Matsuki, H. Akutsu, S. Satoshi Takahashi, H. Naiki, Y. Goto, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18119–18124.
- [10] S. Meehan, Y. Berry, B. Luisi, C. M. Dobson, J. A. Carver, C. E. MacPhee, *J. Biol. Chem.* **2004**, *279*, 3413–9.
- [11] M. D. Weingarten, A. H. Lockwood, S. Y. Hwo, M. W. Kirschner, *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 1858–1862.
- [12] A. W. P. Fitzpatrick, B. Falcon, S. He, A. G. Murzin, G. Murshudov, H. J. Garringer, R. A. Crowther, B. Ghetti, M. Goedert, S. H. W. Scheres, *Nature* **2017**, *547*, 185–190.
- [13] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. Ø. Madsen, C. Riekel, D. Eisenberg, *Nature* **2007**, *447*, 453–457.
- [14] C. L. Oliveira, M. A. Behrens, J. S. Pedersen, K. Erlacher, D. Otzen, J. S. Pedersen, *J. Mol. Biol.* **2009**, *387*, 147–161.
- [15] J. L. Jiménez, E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson, H. R. Saibil, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9196–9201.
- [16] J. A. Rodriguez, M. I. Ivanova, M. R. Sawaya, D. Cascio, F. E. Reyes, D. Shi, S. Sangwan, E. L. Guenther, L. M. Johnson, M. Zhang, L. Jiang, M. A. Arbing, B. L. Nannenga, J. Hattne, J. Whitelegge, A. S. Brewster, M. Messerschmidt, S. Boutet, N. K. Sauter, T. Gonen, D. Eisenberg, *Nature* **2015**, *525*, 486–90.
- [17] J. J. W. Wiltzius, S. A. Sievers, M. R. Sawaya, D. Cascio, D. Popov, C. Riekel, D. Eisenberg, *Protein Sci.* **2008**, *17*, 1467–1474.
- [18] J. T. Nielsen, M. Bjerring, M. D. Jeppesen, R. O. Pedersen, J. M. Pedersen, K. L. Hein, T. Vosegaard, T. Skrydstrup, D. E. Otzen, N. C. Nielsen, *Angew. Chem. Int. Ed.* **2009**, *48*, 2118–2121; *Angew. Chem.* **2009**, *121*, 2152–2155.
- [19] R. Fernandez-Leiro, S. H. W. Scheres, *Nature* **2016**, *537*, 339–346.
- [20] S. E. Ashbrook, J. M. Griffin, K. E. Johnston, *Annu. Rev. Anal. Chem.* **2018**, *11*, 485–508.
- [21] G. Fusco, S. W. Chen, P. T. F. Williamson, R. Cascella, M. Perni, J. A. Jarvis, C. Cecchi, M. Vendruscolo, F. Chiti, N. Cremades, L. Ying, C. M. Dobson, A. De Simone, *Science* **2017**, *358*, 1440–1443.
- [22] K. A. Conway, S. J. Lee, J. C. Rochet, T. T. Ding, R. E. Williamson, P. T. Lansbury, Jr., *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 571–576.
- [23] S. Makin, *Nature* **2018**, *559*, S4–S7.
- [24] A. M. Streets, Y. Sourigues, R. R. Kopito, R. Melki, S. R. Quake, *PLoS One* **2013**, *8*, e54541.
- [25] M. Biancalana, K. Makabe, A. Koide, S. Koide, *J. Mol. Biol.* **2009**, *385*, 1052–63.
- [26] T. Skamris, C. Marasini, K. L. Madsen, V. Foderà, B. Vestergaard, *Sci. Rep.* **2019**, *9*, 1733.
- [27] F. Gobeaux, F. Wien, *Langmuir* **2018**, *34*, 7180–7191.
- [28] J. L. Lau, M. K. Dunn, *Bioorg. Med. Chem.* **2018**, *26*, 2700–2707.
- [29] B. N. Bullock, A. L. Jochim, P. S. Arora, *J. Am. Chem. Soc.* **2011**, *133*, 14220–14223.
- [30] A. Uccellatore, S. Genovese, I. Dicembrini, E. Mannucci, A. Ceriello, *Diabetes Ther.* **2015**, *6*, 239–256.
- [31] S. Runge, H. Thøgersen, K. Madsen, J. Lau, R. Rudolph, *J. Biol. Chem.* **2008**, *283*, 11340–11347.
- [32] L. Qiu, S. A. Pabit, A. E. Roitberg, S. J. Hagen, *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953.
- [33] P. Stráner, N. Taricska, M. Szabó, K. G. Tóth, A. Perczel, *Curr. Protein Pept. Sci.* **2016**, *17*, 147–155.
- [34] V. Farkas, B. Csordás, O. Hegyi, G. K. Tóth, A. Perczel, *Eur. J. Org. Chem.* **2013**, 3513–3522.
- [35] A. Umamoto, H. Yagi, M. So, Y. Goto, *J. Biol. Chem.* **2014**, *289*, 27290–27299.
- [36] P. Rovó, P. Stráner, A. Láng, I. Bartha, K. Huszár, L. Nyitray, A. Perczel, *Chem. Eur. J.* **2013**, *19*, 2628–2640.
- [37] P. Rovó, V. Farkas, P. Stráner, M. Szabó, Á. Jermendy, O. Hegyi, G. K. Tóth, A. Perczel, *Biochemistry* **2014**, *53*, 3540–3552.
- [38] A. Perczel, M. Hollósi, G. Tusnády, G. D. Fasman, *Protein Eng.* **1991**, *4*, 669–79.
- [39] A. Perczel, K. Park, G. D. Fasman, *Proteins Struct. Funct. Genet.* **1992**, *13*, 57–69.
- [40] I. Jáklí, A. Perczel, *J. Pept. Sci.* **2009**, *15*, 738–752.
- [41] K. A. Graham, A. Byrne, R. Son, N. H. Andersen, *Biopolymers* **2019**, *110*, e23260.
- [42] C. C. Lee, A. Nayak, A. Sethuraman, G. Belfort, G. J. McRae, *Biophys. J.* **2007**, *92*, 3448–58.
- [43] T. R. Serio, A. G. Cashikar, A. S. Kowal, G. J. Sawicki, J. J. Moslehi, L. Serpell, M. F. Arnsdorf, S. L. Lindquist, *Science* **2000**, *289*, 1317–21.
- [44] N. Sreerama, S. Y. Venyaminov, R. W. Woody, *Protein Sci.* **1999**, *8*, 370–80.
- [45] A. Micsonai, F. Wien, L. Kernya, Y. H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3095–E3103.
- [46] P. J. Marek, V. Patsalo, D. F. Green, D. P. Raleigh, *Biochemistry* **2012**, *51*, 8478–8490.
- [47] A. K. Buell, P. Hung, X. Salvatella, M. E. Welland, C. M. Dobson, T. P. J. Knowles, *Biophys. J.* **2013**, *104*, 1116–1126.
- [48] K. Klement, K. Wieligmann, J. Meinhardt, P. Hortschansky, W. Richter, M. Fändrich, *J. Mol. Biol.* **2007**, *373*, 1321–33.
- [49] P. Sormanni, F. A. Aprile, M. Vendruscolo, *J. Mol. Biol.* **2015**, *427*, 478–90.
- [50] W. Zheng, M.-Y. Tsai, P. G. Wolynes, *J. Am. Chem. Soc.* **2017**, *139*, 16666–16676.
- [51] S. T. Ngo, X.-C. Luu, M. T. Nguyen, C. N. Le, V. V. Vu, *RSC Adv.* **2017**, *7*, 42379–42386.
- [52] T. P. J. Knowles, M. Vendruscolo, C. M. Dobson, *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 384–396.
- [53] M. Biancalana, K. Makabe, S. Yan, S. Koide, *Protein Sci.* **2015**, *24*, 841–849.
- [54] D. S. Eisenberg, M. R. Sawaya, *Annu. Rev. Biochem.* **2017**, *86*, 69.
- [55] W. M. Berhanu, U. H. E. Hansmann, *PLoS One* **2012**, *7*, e41479.
- [56] W. H. Xi, G. H. Wei, *Chin. Phys. B* **2016**, *25*, 018704.
- [57] J. P. Colletier, A. Laganowsky, M. Landau, M. Zhao, A. B. Soriaga, L. Goldschmidt, D. Flot, D. Cascio, M. R. Sawaya, D. Eisenberg, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 16938–16943.
- [58] A. B. Soriaga, S. Sangwan, R. Macdonald, M. R. Sawaya, D. Eisenberg, *J. Phys. Chem. B* **2016**, *120*, 5810–5816.
- [59] L. Saelices, S. A. Sievers, M. R. Sawaya, D. S. Eisenberg, *Protein Sci.* **2018**, *27*, 1295–1303.
- [60] L. Lu, Y. Deng, X. Li, H. Li, G. E. Karniadakis, *J. Phys. Chem. B* **2018**, *122*, 11302–11310.
- [61] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin, M. Nilges, *Bioinformatics* **2007**, *23*, 381–2.
- [62] A. T. Brunger, *Nature Protocols* **2007**, *2*, 2728–2733.

---

 Manuscript received: August 21, 2019

Accepted manuscript online: October 24, 2019

Version of record online: December 27, 2019