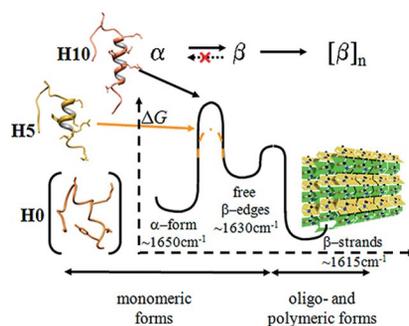


Optical spectroscopic studies of the small, well-folded peptide foldamer, Trp-cage miniprotein. Investigation of stability-associated aggregation properties of three Trp-cage foldamers.

**Foldamer Aggregation Propensity**

V. Farkas, B. Csordás, O. Hegyi,

G. K. Tóth, A. Perczel\* ..... 1–11

Foldamer Stability Coupled to Aggregation Propensity of Elongated Trp-Cage Miniproteins 

**Keywords:** Protein aggregation / Foldamers / Amyloid formation / Trp-cage miniprotein / VCD spectroscopy

DOI: 10.1002/ejoc.201300071

# Foldamer Stability Coupled to Aggregation Propensity of Elongated Trp-Cage Miniproteins

Viktor Farkas,<sup>[a]</sup> Barbara Csordás,<sup>[a]</sup> Orsolya Hegyi,<sup>[b]</sup> Gábor K. Tóth,<sup>[b]</sup> and  
András Perczel<sup>\*[a]</sup>

*Dedicated to Professor Miklós Hollósi on the occasion of his 70th birthday*

**Keywords:** Protein aggregation / Foldamers / Amyloid formation / Trp-cage miniprotein / VCD spectroscopy

Here we present folding-associated aggregation propensity of three Trp-cage foldamers: E0 (20 aa), E5 (25 aa) and E10 (30 aa), models of different sizes but comparable molecular properties. Electronic circular dichroism (ECD), vibrational circular dichroism (VCD) and FT-IR spectroscopic measurements were used to monitor their concentration-dependent, heat-induced (5 °C → 65 °C) "α→β" fold transition. The ECD curves of E0 display an ensemble of highly dynamic structures. ECD of both E5 and E10 foldamers show the expected Trp-cage fold, dominated by their α-helical properties. No sign of β-structures was revealed by ECD at any conditions (5 °C < T < 65 °C, 5 < pH < 7, c ≈ 30 μM) for any of these miniproteins. However, at higher concentration (c ≈ 1–30 mM) both VCD and FT-IR spectral features of E5 as well as E10 resemble that of a β-strand (ca. 1615 cm<sup>-1</sup>), accompanied with "free β-edges", or native β-sheets (ca. 1635 cm<sup>-1</sup>). E5 at lower concentrations (c ≈ 1–3 mM), and E10 at higher concen-

tration (c ≈ 30 mM) display the α→native-β→β-sheet folding transitions, monitored by the characteristic C=O vibrational normal mode frequency shift as follows: ca. 1650 cm<sup>-1</sup> → ca. 1635 cm<sup>-1</sup> → ca. 1615 cm<sup>-1</sup>, respectively. The latter folding path is irreversible. The shortest polypeptide E0 has an "un-ordered" fold, while E10 presents the most tightly packed Trp-cage 3D-structure. We have found that both high dynamicity and/or tight molecular core packing are different in nature, but common in efficacy in preventing the polypeptide backbone chain against self-aggregation. However, E5 is intermediate in size and stability, and thus among these three polypeptides it is the quickest to aggregate. The present molecular triad, E0, E5 and E10, serves as a good example of larger globular proteins for which aggregation and amyloid fiber-like nanoparticle formations are often associated with Alzheimer's, Creutzfeldt–Jakob, or prion diseases.

## Introduction

The aggregation propensity of peptides and proteins is a known and feared fact, resulting in filaments and plaques both in vitro and in vivo. Self-aggregation and fibril formation of biopolymers and biocompatible nanomaterials is therefore intensively studied. For most globular proteins, the amino acid sequence encodes their 3D structures – the forms responsible for their cellular activities.<sup>[1]</sup> A handful of examples support, however, that besides the bioactive fold,

additional well-organized, alternative 3D structures can also exist, which manifest during their aggregation.<sup>[2]</sup> Thus, proteins do not just simply misfold and get precipitated via random interactions, but rather, as they adopt their bioactive forms (within a few μs or ms), they can also develop a second type of well-organized structural form promoted by just slightly abnormal cellular conditions. The latter conversion from globular fold to amyloid fibrils takes typically hours, days or even longer. Transition of the native supramolecular structure into the nonfunctional and highly pathogenic form is intimately linked to the overstabilization of some folding intermediate states. Point mutations, pH or temperature changes could promote intermediate formation/stabilization and thus, the system would gravitate toward its aggregate state. These amyloid-fibre-like nanoparticles are either causes or symptoms of conformational diseases (e.g. Alzheimer's disease, Creutzfeldt–Jakob disease, prion disease).<sup>[3]</sup> Thus, understanding the molecular basis and the mechanism of transitions between native, intermediate and aggregated states is crucial to maintaining the hope of in vivo diagnosis and cure of these diseases. Re-

[a] Protein Modeling Group HAS-ELTE and Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, Pázmány P. sétány 1/A, Budapest 1117, Hungary  
Fax: +36-1-3722-620  
E-mail: perczel@chem.elte.hu  
Homepage: <http://prot.chem.elte.hu/en/>

[b] Department of Medical Chemistry, University of Szeged, Dóm tér 8, Szeged 6720, Hungary

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/ejoc.201300071>.

search targeting molecular details of aggregation requires, however, two conditions to match, namely *i*) finding a relevant model system of tractable size which provides atomic details, and *ii*) selecting an adequate spectroscopic method to monitor essential transitions over a broad range of particle size and concentration. Here we show how vibrational spectroscopic methods (e.g. FTIR and VCD) applied to selected mini-proteins provide a fruitful combination for monitoring foldamer stability and associated self-aggregation.

Most proteins of biological and pathological relevance (e.g. APP,  $\alpha$ -synuclein, huntingtin, transthyretin,  $\beta$ 2-microglobulin etc.) are often too large to work with and too complex to pinpoint crucial changes at an atomic level.<sup>[4]</sup> Aggregation is intimately associated with a series of micro-events coupled to intermolecular interactions of unshielded and free  $\beta$ -edges of the protein's backbone. Although aggregation-prone residues facilitate and stabilize intermolecular interaction, it seems that with time, the thermodynamically driven interstrand association of the polypeptide chains will occur. Aggregation seems to be coupled to successful formation of  $\beta$ -sheets from any local backbone fold, stabilized in aggregates by interchain H-bonds.<sup>[5]</sup> Along with theoretical studies, experimental investigations also provide atomic-level information on backbone fold transitions leading to protein aggregation (e.g. insulin, betanovosin, Tc5b).<sup>[6]</sup> Therefore, mini-proteins are indeed relevant and good model systems for studying self-aggregation at a tractable size, bearing inherent features of larger proteins.

In looking for the adequate physico-chemical method to monitor early aggregate formation at atomic detail, diffraction methods (e.g. X-ray) are not the first technique to select. Nevertheless, X-ray-determined packing information on the differently oriented  $\beta$ -sheets revealed recently by the Eisenberg group is of great value.<sup>[7]</sup>  $^1\text{H}$  NMR spectroscopy is also unsuitable for studying large supramolecular aggregates, due to line-broadening effects; which are a manifestation of inherent  $^1\text{H}$  relaxation. Dynamic light scattering and electron (cryo)microscopy provide valuable data on size distribution and forms of aggregates and nanoparticles (e.g. 10–100 nm); however, atomic details remain hidden. Optical and vibrational spectroscopy are often used to monitor such changes. Among these commonly used techniques, infrared (IR) and vibrational circular dichroism (VCD) spectroscopy are easy and quick techniques to use even for a broad range of concentrations and temperatures (5–95 °C). Thus, monitoring the conformational changes and aggregate formation of small proteins (e.g. mini-proteins and foldamers<sup>[8–12]</sup>) by FT-IR and VCD methods and then extrapolating results to proteins seems to be a good compromise.

Aggregation and fibrillogenesis of some polypeptides [insulin,<sup>[13–15]</sup> A $\beta$  heptapeptide (16–22), VP1 decapeptide of calpain<sup>[16]</sup> etc.] and smaller proteins (e.g. lysozyme, RNase Sa,  $\beta$ 2-microglobulin<sup>[17]</sup>) have been monitored by VCD and FT-IR. The effect can be accelerated with the increase of concentration in water (up to 100 mg/mL), but more importantly, by lowering the pH (ca. 2)<sup>[14,15]</sup> and/or increasing the temperature (ca. 65 °C) for a given length of incubation

time (e.g. from hours to days). Cosolvent(s) (e.g. TFE in water) can not only initiate  $\alpha$ -helix formation, but can also help the development of fibril-like aggregates.<sup>[18]</sup> Time-resolved VCD spectroscopy has been used to determine some kinetic properties, lag time and apparent first-order rate constant of fibril formation.<sup>[19]</sup> It was pointed out that for maturation of amyloid fibrils, the solubility of the pre-formed smaller aggregates is a key factor.<sup>[20]</sup> By using various conditions (pH cosolvent,  $T$ , etc.), FT-IR and VCD spectroscopic methods are straightforward ways to show that most aggregates are composed of  $\beta$ -strands. However, details of folding and characterization of the possible intermediates involved in aggregation remain unrevealed and uncharacterized. The in vitro aggregation study of the 30-amino-acid-long glucagon-like peptide 1, GLP-1 for short, shows spontaneous fibril formation, resulting in  $\beta$ -sheet-rich amyloid-like fibrils.<sup>[21]</sup>

Poon et al. determined that in such a supramolecular complex, interaction was mediated by 12–13 residue-long antiparallel  $\beta$ -strands. The structure of the GLP-1 aggregate was elucidated by ECD, X-ray fibre diffraction, fluorescence, IR and TEM measurements.<sup>[21]</sup> On the other hand, NMR (DOSY) experiments of GLP-1 showed that TFE-induced self-aggregation resulted in dimers in which a parallel coiled-coil-type interaction was operative between molecules, with a few hydrophobic residues on surface, leading to further aggregation even in pure water.<sup>[22]</sup> This was one of the few studies that revealed information on crucial intermediates which drive the system to further aggregation and to a folding “dead end”.<sup>[5]</sup> Prompted by this work we have designed and synthesized 3 different peptides, all related to Exendin-4, Ex-4 for short, an agonist of the incretin GLP-1. The 39-residue long Ex-4 – a modern drug for treating type II diabetes<sup>[23]</sup> – was a starting molecule for designing shorter foldamers and mini-proteins known as Trp-cage folds.<sup>[24]</sup> “Oligomers that display well-defined and predictable conformations”, also called as foldamers, often manifest as helices, extended structures or sheets.<sup>[25]</sup> Tc5b, a Trp-cage foldamer, was shown to be an autonomously folding polypeptide chain presenting a single 3D structure in a time-averaged manner and thus, it is in line with the above flexible definition of foldamers. In this paper we are elaborating foldamer transitions occurring between the native state of the Tc5b foldamer and some of its aggregates. The latter foldamer, Tc5b, is composed of only 20 amino acids (NLYIQ WLKDG GPSSG RPPPS), optimized from the C-terminus of Ex-4. Although Tc5b is to date the smallest self-folding polypeptide chain,<sup>[24]</sup> it is considered a realistic model of larger proteins. Its temperature- and pH-induced folding and unfolding are as complex as those of larger proteins.<sup>[26–28]</sup> VCD spectral studies of Tc5b have identified a negative-positive-negative couplet (in  $\text{D}_2\text{O}$ ) attributed to the  $\alpha$ -helical part of the mini-protein.<sup>[29]</sup> In addition, the PPII segment could also be detected, supporting the viability of VCD for such a study. It was found, however, that with the increase of  $T$  ( $> 35$  °C) the Trp-cage fold was lost through a “ $\alpha \rightarrow \text{RND}$ ” transition. No aggregation of it was described. Recently Tc5b folding and unfolding was experi-

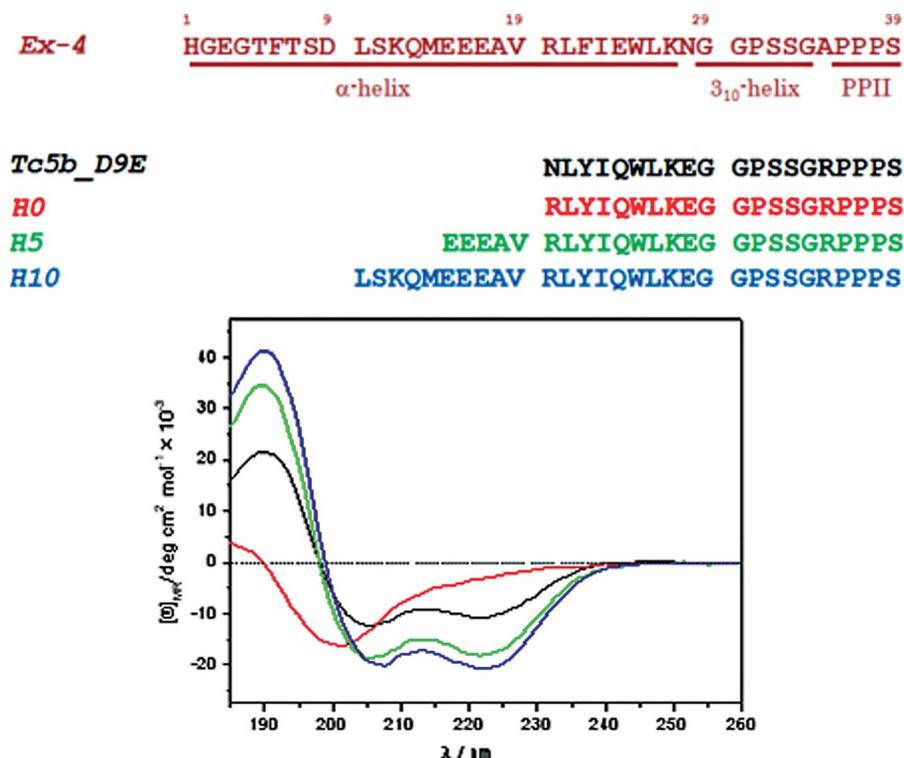


Figure 1. Primary sequence and the far-UV ECD spectra in H<sub>2</sub>O of Tc5b\_D9E, E0, E5 and E10 mini-proteins ( $c \approx 30 \mu\text{M}$ ,  $T = 5^\circ\text{C}$ ).

mentally studied in solution at atomic resolution by NMR spectroscopy.<sup>[30,31]</sup>

Here we present an insight into the aggregation propensity of three carefully selected Tc5b foldamers of complementary nature; E0, E5 and E10 (E stands for their gradually elongated sequence). To some extent, all three polypeptides adopt the Trp-cage fold. However, they are sensitive to external conditions (pH  $T$ , concentration, ionic strength, etc.) in different ways, and thus ready to provide subtle but important structural changes to understand more on the details of protein aggregation. The Trp-cage fold of Tc5b\_D9E was characterised previously<sup>[32,34,35]</sup> (Figure 1) and used here to build up E0, E5 and E10 models, which are designed to have different conformational properties. Even though all three foldamers share the 20 residue long C-terminal part (RLYIQWLKEG GPSSGRPPPS), E0 is very dynamic (highly unstructured), while E5 (EEEAV RLYIQWLKEG GPSSGRPPPS) and especially E10 (LSKQMEEEAV RLYIQWLKEG GPSSGRPPPS) contain well-matured  $\alpha$ -helical segments at lower concentration (ca. 0.1–1 mM) and at neutral pH.<sup>[32]</sup> Lowering pH and/or increasing concentration however, could initiate the appearance of intermediate structure(s) and mediate the escalation of the foldamers to aggregate. Dobson and coworkers have predicted that enhancing intermediate stability could result in the acceleration of aggregate formation.<sup>[33]</sup> On the pathway of stabilizing the  $\alpha$ -helical part, however, intermediates of enhanced half-life could appear, triggering self-aggregation.<sup>[7,36–38]</sup> In this paper we give solid experimental evidence on how enhanced intermediate stability leads to aggregate formation.

## Results and Discussion

The structure of the 39-residue Ex-4 measured in 30%TFE (70%H<sub>2</sub>O) is composed of three secondary structural elements at 25 °C: a long  $\alpha$ -helix, a short  $3_{10}$ -helix, and a PPII segment at its C-terminus (Figure 1).<sup>[39]</sup> Tc5b (NLYIQWLKDG GPSSGRPPPS) was further modified to give Tc5b\_D9E (NLYIQWLKEG GPSSGRPPPS), a salt-bridge optimized variant.<sup>[34,35]</sup> The Asp→Glu change enhances the salt bridge (Asp<sup>-</sup>: Arg<sup>+</sup>) stability but also shifts gently the Trp-cage fold.<sup>[35]</sup> The back substitution of the original Arg20 of Ex-4 into Tc5b\_D9E, still keeping the Trp-cage numbering results in E0: Tc5b\_N1R\_D9E or RLYIQWLKEG GPSSGRPPPS (Figure 1).

The introduction of a positively charged side-chain residue (Arg<sup>+</sup>) at the N-terminus of an  $\alpha$ -helix in E0 destabilized the fold and jeopardized the Trp-cage structure.<sup>[32]</sup> The addition of the parent pentapeptide, EEEAV, of Ex-4 to the N-terminus of E0 results in E5. E10 was obtained by further adding the LSKQM pentapeptide of Ex-4 to E5 (Figure 1).

## ECD Spectroscopy

As foreseen, the Asn1→Arg mutation deeply influenced the overall fold of E0 with respect to Tc5b\_D9E, which could be nicely monitored by the far-UV ECD-spectra. The C-type CD-spectrum of Tc5b\_D9E changed to a U-type curve, reflecting a highly dynamic and thus disordered structure (red curve, Figure 1). The extension of E0 to E5 and E5 to E10 gradually restored the original highly  $\alpha$ -heli-

cal fold, resulting in well-matured, C-type ECD spectra (green and cyan curves of E5 and E10, respectively Figure 1).

At lower peptide concentration in water (ca. 30  $\mu\text{M}$ ), none of the three model systems showed dramatic conformational shifts induced by lowering the pH (data not shown) or increasing the temperature (5  $\rightarrow$  65  $^{\circ}\text{C}$ ). The far-UV ECD spectra of E0 did not change with  $T$ , indicating

that even at lower  $T$  (5  $^{\circ}\text{C}$ ), E0 was very dynamic, often referred to as an unfolded structure. On the contrary, the helix content of both E5 and E10 dropped only gently as temperature increased (5  $\rightarrow$  65  $^{\circ}\text{C}$ ), as indicated by the  $n \rightarrow \pi^*$  (222 nm) band intensity changes (black to cyan curves, Figure 2). Nevertheless, even at 65  $^{\circ}\text{C}$ , the C-type ECD curve dominated the spectra, indicating the overwhelming presence of an  $\alpha$ - and/or  $3_{10}$ -helix determined Tc5b fold, both for E5 and E10. In this low concentration regime (10–50  $\mu\text{M}$ ) acidifying the solution (7 > pH > 3) made insignificant effect on the overall fold: no sign of aggregation was seen (data not shown).

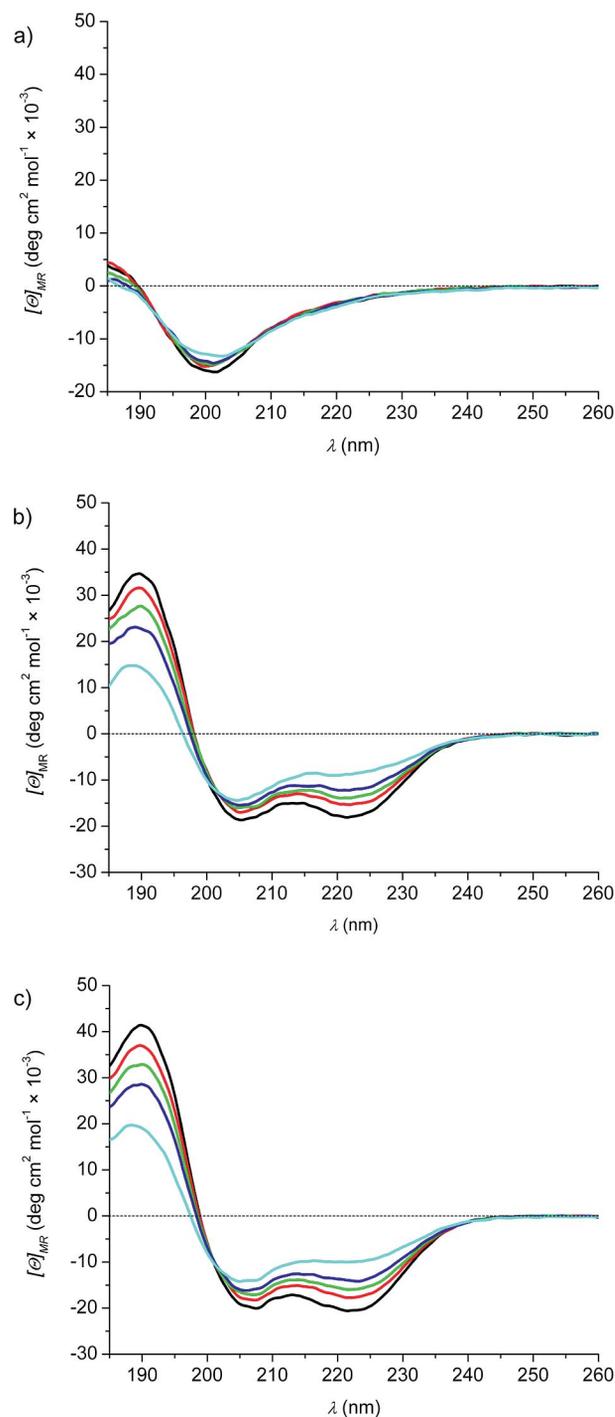


Figure 2. The far-UV ECD spectra of a) E0, ( $c = 33.7 \mu\text{M}$ ), b) E5, ( $c = 26 \mu\text{M}$ ) and c) E10, ( $c = 31 \mu\text{M}$ ) between, 5  $\leq T \leq 65 \text{ }^{\circ}\text{C}$ , 5  $^{\circ}\text{C}$  (black), 25  $^{\circ}\text{C}$  (red), 35  $^{\circ}\text{C}$  (green), 45  $^{\circ}\text{C}$  (blue) and 65  $^{\circ}\text{C}$  (cyan).

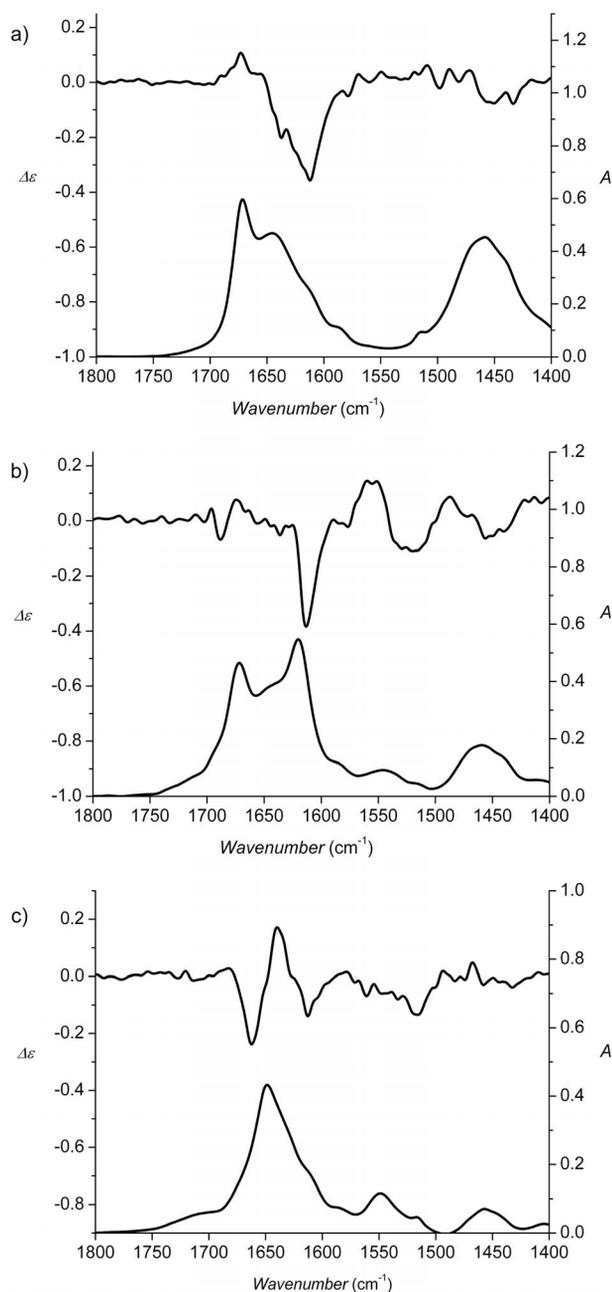


Figure 3. The amide I' and II' regions of the VCD (top) and IR (bottom) spectra of a) E0 ( $c = 18 \text{ mM}$ ), b) E5, ( $c = 14 \text{ mM}$ ) and c) E10 ( $c = 8 \text{ mM}$ ) at 5  $^{\circ}\text{C}$ .

## VCD Spectroscopy

The VCD spectrum of E0 showed a broad negative band at  $1612\text{ cm}^{-1}$  with a weaker positive maximum at  $1673\text{ cm}^{-1}$  (Figure 3). The latter spectral feature reflected a mixture of conformers<sup>[40]</sup> including  $\beta$ -sheet, random coil and PPII helices. The amide II' band was marked by a negative band. The overall shape and the broad (ca.  $80\text{ cm}^{-1}$ ) amide I' region made this spectrum hard to assign. The VCD spectrum of E5 was marked by a strong negative band at  $1613\text{ cm}^{-1}$  with weak positive-negative-positive bands above  $1675\text{ cm}^{-1}$  (Figure 3), a pattern typically assigned to antiparallel  $\beta$ -pleated sheet.<sup>[40]</sup> As expected for  $\beta$ -stranded structures, both the amide II and II' bands appeared at ca.  $1550\text{ cm}^{-1}$  and ca.  $1450\text{ cm}^{-1}$ , respectively. No traces of a helical fold could be depicted for E5. Finally, E10 presented a typical  $\alpha$ -helix VCD spectrum with a definite negative-positive-negative triplet in the amide I' region. This shape highly resembled the VCD spectrum of Tc5b in  $\text{D}_2\text{O}$ ,<sup>[29]</sup> which also includes a longer  $\alpha$ -helical secondary structural element (Figure 3).<sup>[29]</sup> Thus, VCD spectra of E0, E5 and E10 were all very different: the first one featured a RND-structure, the second one a  $\beta$ -pleated sheet, while the third one an  $\alpha$ -helix dominated foldamer. The conformational property associated with E5 by using VCD is incompatible with what was said on the basis of its ECD spectra (Figure 2). The first one refers to a  $\beta$ -sheet, the second one to an  $\alpha$ -helix.

## Comprehensive Analysis by Using VCD and FT-IR Spectroscopy

Before trying to explain the above contradiction recorded for E5 by ECD and VCD measurements, the temperature dependence of the VCD curves needs to be analyzed. By using a suitable curve-fitting algorithm, FT-IR bands of the  $1580\text{--}1710\text{ cm}^{-1}$  range were assigned (Table 1), according to literature.<sup>[41,29]</sup>

Increasing the temperature did not affect the amide I' vibrational mode ( $1600\text{--}1700\text{ cm}^{-1}$ ) of E0 [the amide II' band ( $1450\text{ cm}^{-1}$ ) changed slightly Figure 4, a]. The VCD spectrum of E5 showed at higher  $T$  some intensity changes, for example, the intensity dropped at  $1613\text{ cm}^{-1}$  with increasing temperature. Furthermore, the negative band intensity increased at  $1688\text{ cm}^{-1}$  with that of the positive one (ca.  $1695\text{ cm}^{-1}$ ; Figure 4, b). The FT-IR spectrum of E5 changed just slightly at the amide I' band, but more significantly at the amide II/amide II' regions due to H $\rightarrow$ D exchange (Figure 4, b). The appearance of the strong amide II' band (negative couplet at  $1450\text{ cm}^{-1}$ ) indicated that even at lower  $T$  ( $5\text{ }^\circ\text{C}$ ), E5 featured a  $\beta$ -strand structure, but with increasing temperature, the  $\beta$ -structure became even more stable.

The increase of temperature had at first a marginal effect on the VCD spectra of E10 – changes were minor between  $5$  and  $45\text{ }^\circ\text{C}$ , but more pronounced upon reaching  $65\text{ }^\circ\text{C}$ . At high temperature ( $65\text{ }^\circ\text{C}$ , cyan on Figure 4, c) the VCD spectrum showed a strong negative couplet in the amide II' region (ca.  $1450\text{ cm}^{-1}$ ) and a negative couplet near

Table 1. IR-band assignment of the three foldamers: a) E0, b) E5 and c) E10 at  $T = 5\text{ }^\circ\text{C}$ .

a) Wavenumber [ $\text{cm}^{-1}$ ]	Assignment
1585	aromatic side-chain (Trp, Tyr)
1611	$\beta$ -aggregation/anti-parallel $\beta$ -sheet
1635	native $\beta$ -sheet/free $\beta$ -edges
1647	random coil
1672	TFA
b) Wavenumber [ $\text{cm}^{-1}$ ]	Assignment
1588	aromatic side-chain (Trp, Tyr)
1618	$\beta$ -aggregation/anti-parallel $\beta$ -sheet
1644	random coil
1673	TFA
1691	$\beta$ -aggregation/anti-parallel $\beta$ -sheet
1705	Glu side-chain [ $\nu(\text{C}=\text{O})$ ]
c) Wavenumber [ $\text{cm}^{-1}$ ]	Assignment
1581	aromatic side-chain (Trp, Tyr)
1613	$\beta$ -aggregation/anti-parallel $\beta$ -sheet/or PPII <sup>[29]</sup>
1638	native $\beta$ -sheet/free $\beta$ -edges
1651	$\alpha$ -helix
1670	$\beta$ -turn
1709	Glu side-chain [ $\nu(\text{C}=\text{O})$ ]

$1680\text{ cm}^{-1}$ , both indicative of an antiparallel  $\beta$ -sheet structure. In addition, at higher temperature, a clear sign of the decrease in helicity was monitored (Figure S1–S9). In the FT-IR spectrum, the band at  $1651\text{ cm}^{-1}$  decreased, and those at  $1616\text{ cm}^{-1}$  and  $1681\text{ cm}^{-1}$  appeared and increased (Figure 4, c). These changes indicated an  $\alpha$ -helix to antiparallel  $\beta$ -sheet transition:  $\alpha\rightarrow\beta$ , where the band at  $1616\text{ cm}^{-1}$  signaled the  $\beta$ -sheet and the band at  $1681\text{ cm}^{-1}$  indicated its antiparallel form. Thus, even the tightest Trp-cage fold E10 presented dominant  $\beta$ -sheet structural properties at higher  $T$ , unlike E5, which featured the same characteristic  $\beta$ -sheet already at  $5\text{ }^\circ\text{C}$ .

The inconsistency between the structural properties revealed by the ECD and VCD measurements was especially pronounced for E5. The conformational difference is most likely to rise from the different conditions used to record spectra. For example, ECD was measured on a relatively diluted sample (ca.  $10\text{--}30\text{ }\mu\text{M}$ ), and its VCD counterpart was recorded at higher concentration (ca.  $10\text{--}20\text{ mM}$ ). The three orders-of-magnitude-higher concentration used for VCD could clearly promote self-aggregation even at neutral pH and low temperature. Aggregation via  $\beta$ -sheet formation as witnessed by VCD must arise from a key intermediate structure – probably a free  $\beta$ -edge containing foldamer<sup>[42]</sup> present at low relative concentration with respect to the folded or unfolded forms of the polypeptide chain. The plasticity of the Trp-cage fold allowed monitoring of  $\beta$ -sheet formation even for the most stable E10 foldamer. Its fold comprises a well-matured and five residue longer  $\alpha$ -helical segment with respect to E5 (Figure 1), which seemed to protect the globular structure against aggregation.

At  $25\text{ }^\circ\text{C}$ , E10 was resistant against aggregation even at a high concentration ( $> 20\text{ mM}$ ); no aggregation was seen – only signs of the native  $\beta$ -sheet (free  $\beta$ -edges) were ob-

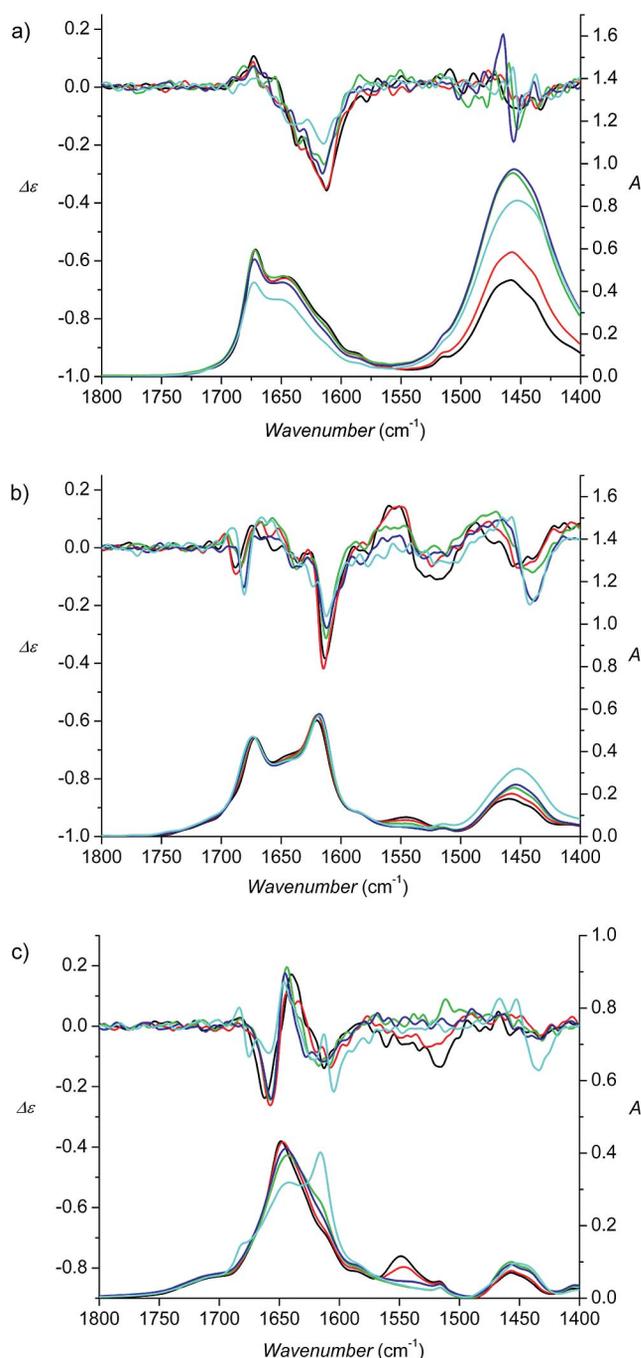


Figure 4. The amide I' and II' regions of the VCD (top) and IR (bottom) spectra of a) E0 ( $c = 18 \text{ mM}$ ), b) E5, ( $c = 14 \text{ mM}$ ) and c) E10 ( $c = 8 \text{ mM}$ ) at 5 °C (black), 25 °C (red), 35 °C (green), 45 °C (blue) and 65 °C (cyan).

served. However, the free  $\beta$ -edge-containing intermediate structure could be present even for E10. When concentration and temperature are simultaneously increased ( $T$  up to 65 °C), E10 started to present a considerable amount of  $\beta$ -sheet-containing fold (green curve Figure 5). We were keen to see whether such an  $\alpha \rightarrow \beta$  backbone-fold transition was reversible. Thus, we cooled the solution of E10 from 65 °C to 5 °C (red curve Figure 5) and recorded its VCD spectrum again. The two 5 °C VCD spectra, the one recorded initially

and the one after heating and cooling (black and red curves Figure 5) were very different in nature. In fact, the latter one resembled that recorded at 65 °C, which is a clear indication that once a  $\beta$ -structure is formed, it self-stabilizes and no  $\beta \rightarrow \alpha$  transition becomes possible even if the system is cooled down again. Thus, the latter backbone change was irreversible, and the second half of an  $\alpha \rightarrow \beta \rightarrow \alpha$  transition was inaccessible, because the intermediate characterised as a native  $\beta$ -fold serves as templates for further aggregation. (Note that the physical appearance of the solution containing  $\beta$ -forms remained unchanged. Gel formation was seen, for example, as the aggregated particles remained soluble at this phase of the transition.) Thus,  $\beta$ -sheet formation followed by multimerization via aggregation presented a folding “dead-end street” for these molecules as well.<sup>[5]</sup>

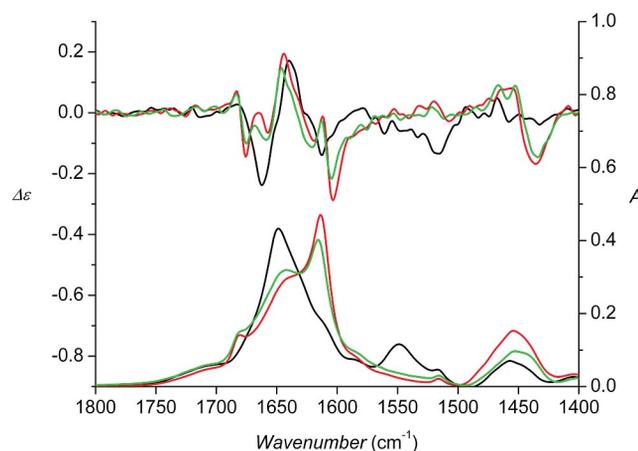


Figure 5. The amide I' and II' regions of the VCD (top) and IR (bottom) spectra of E10 (ca. 8 mM). Black curve: 5 °C (initial state), green curve: 65 °C (elevated  $T$  state), red curve: cooled to 5 °C (after heating).

Even though free  $\beta$ -edges form larger  $\beta$ -strands and subsequently aggregates, the process takes time. Progression can be speeded up by lowering the pH. It took about 8–10 days for the otherwise stable E10 to aggregate at elevated temperature, higher concentration and acidic conditions:  $2 < \text{pH} < 3$ ,  $T = 65 \text{ °C}$ ,  $c \approx 8 \text{ mM}$  (Figure 6). The constant changes of the spectral features associated with the  $\alpha$  to antiparallel  $\beta$ -sheet transition were monitored both by VCD and FT-IR for the first half of the experiment (4 d), where  $T$  was increased from 5 to 65 °C (Figure 5). Upon reaching  $T = 65 \text{ °C}$ , the antiparallel  $\beta$ -sheet structure was already formed (1681 and 1616  $\text{cm}^{-1}$ ), but the aggregation process was not yet fully completed. During the second half of the measurement (6 d), there were slow but significant VCD band intensity changes. While the IR band intensity at 1616  $\text{cm}^{-1}$  did not change, that of the associated VCD band (1619  $\text{cm}^{-1}$ ) increased significantly (green and blue curves; Figure 6). As only the chiroptical VCD spectra changed, this should be attributed to some “chiral” property adjustment: the (re)ordering of the  $\beta$ -strands and layers. Due to their reorientation, the  $\beta$ -strands became more tightly packed and thus less hydrated. As water was “pumped out” from the compartments between strands, more, tighter H-

## FULL PAPER

bonds were formed. Finally, the 1604/1608  $\text{cm}^{-1}$  band/shoulder (green/blue Figure 6) of the VCD spectra reflected a C=O vibrational mode participating in bifurcated H-bonds,<sup>[43]</sup> as described for shifted  $\beta$ -sheets measured on polyGlu by Itoh et al. and Fulara et al.<sup>[44–46]</sup>

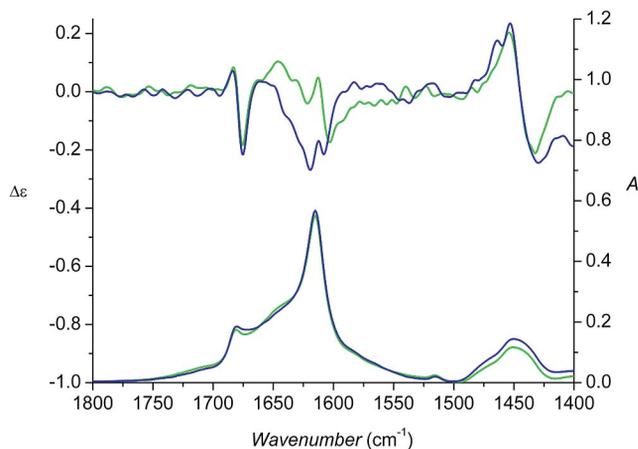


Figure 6. The amide I' and II' regions of the VCD (top) and IR (bottom) spectra of E10 incubated at 65 °C and monitored over 6 d (ca. 8 mM, green: 1st day and blue: 6th day).

So far we have seen for the  $\alpha$ -helical part of the Trp-cage fold that both concentration and temperature increases induced  $\beta$ -sheet formation (monitored by VCD and IR spectroscopy), and the efficacy of the  $\alpha \rightarrow \beta$  transition was a function of the primary sequence of the model. This suggests that the relative abundance of the intermediate structure(s) was higher for E5 than it was for E10, promoting  $\beta$ -sheet formation. In addition, the observed  $\beta$ -sheets were irreversibly formed via self-aggregation. The  $\alpha \rightarrow \beta$  fold change could not be reversed either by cooling/heating or waiting. The next task was to narrow down the concentration range (and conditions) where E5 and E10 start to form  $\beta$ -sheets – in other words, to fill the gap of two to three orders of magnitude in concentration (ca. 26  $\mu\text{M} \rightarrow$  ca. 14 mM) arising from the ECD and VCD measurement conditions. Note that E5 at lower concentration (ca. 26  $\mu\text{M}$ ) presented a “Trp-cage” fold dominated by a longer  $\alpha$ -helical segment (as seen by ECD), while the very same molecule at higher concentration (ca. 14 mM) showed  $\beta$ -sheet like structure (seen by VCD). Due to instrumental limitations, only FT-IR measurements could offer to bridge the above large concentration range. Besides concentration, the pH dependence of the  $\alpha \rightarrow \beta$  transformation will also be fine-tuned.

By measuring the amide I regions of E5 by FT-IR between 0.36 and 3.6 mM, the  $\alpha$ -helix-to- $\beta$ -sheet transition was nicely observed (Figure 7). The transition occurred at a concentration between 0.36 and 0.9 mM. The TFA contamination originating from the HPLC eluent was unmasked (left panel) and eliminated by subsequent lyophilization (Figure 7, b). We have further narrowed down this concentration range (0.72 mM), subject to the limitations of curve-fitting for a semi-quantitative secondary structure analysis. The curve-fitted spectra show that the  $\beta$ -sheet

started to form above 0.36 mM (1619  $\text{cm}^{-1}$ ) and became dominant at concentrations above 1.8 mM. This signaled that for E5 at 0.36 mM and above (at  $T = 25$  °C, pH ca. 3), the hidden intermediate form comprising free  $\beta$ -edges reached the relative concentration where aggregation could become effective and irreversible (Figure 8).

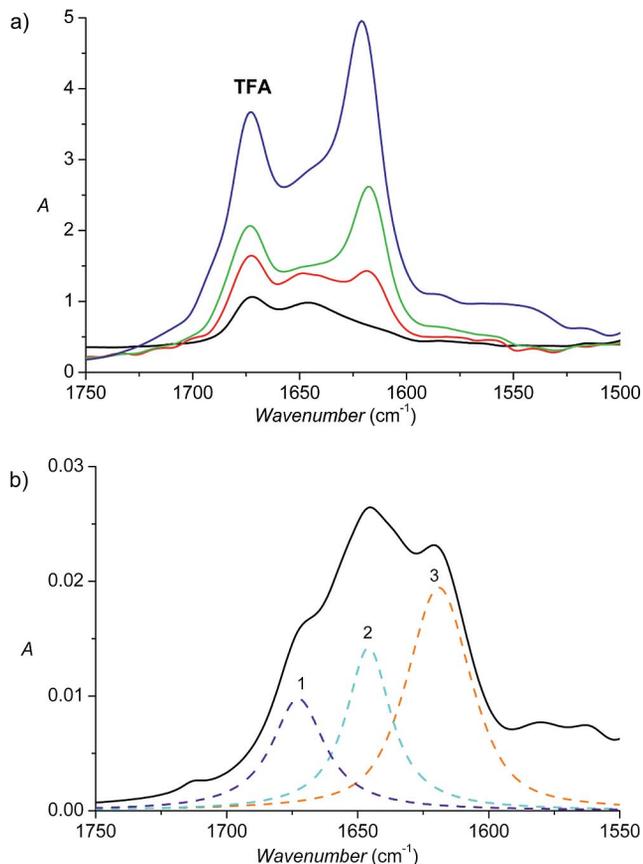


Figure 7. Superimposed FT-IR spectra of E5 at  $T = 25$  °C and pH  $\approx 3$ : a) black curve:  $c \approx 0.36$  mM, red curve:  $c \approx 0.9$  mM, green curve:  $c \approx 1.8$  mM, blue curve:  $c \approx 3.6$  mM. b) spectra deconvoluted into 3 components at  $c \approx 0.72$  mM (1: 1672  $\text{cm}^{-1}$ , 2: 1655  $\text{cm}^{-1}$ , 3: 1619  $\text{cm}^{-1}$ ).

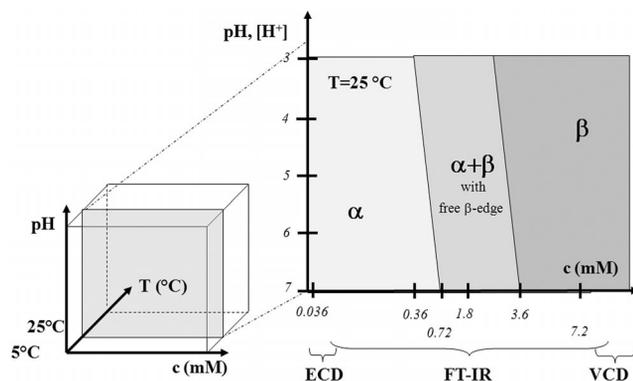


Figure 8. Structural properties of E5 as a function of pH (7 to 3), temperature (5 to 65 °C) and concentration ( $0.03 < c < 14$  mM) investigated by ECD, FT-IR or VCD spectroscopy. The room temperature (25 °C) “slide” is highlighted;  $\alpha$ -form ( $< \approx 0.72$  mM),  $\alpha + \beta$  with free  $\beta$ -edges intermediate state (ca.  $0.72 < c < 3.6$  mM) and  $\beta$ - or aggregated form ( $> \text{ca. } 3.6$  mM).

## Conclusions

We have investigated stability-associated aggregation properties of 3 Trp-cage foldamers and found the following: *i)* E0, the shortest foldamer at 20 aa, was the least stable system investigated here, however, even E0 presented signs of a Trp-cage fold.<sup>[32]</sup> *ii)* E10, the longest foldamer at 30 aa, contained the whole  $\alpha$ -helix-stabilizing decapeptide (LSKQMEEEEAV), and thus it had the most stable Trp-cage fold with the tightest core and the most matured  $\alpha$ -helical segment, as investigated both with ECD and <sup>1</sup>H-NMR spectroscopy.<sup>[32]</sup> *iii)* E5, intermediate in size at 25 aa, presented a less-stable fold because it lacked half of the *N*-terminal decapeptide. It was considerably more stable than E0, due to its *N*-terminal EEEAV pentapeptide, where the negatively charged Glu residues especially stabilized the *N*-terminus of the  $\alpha$ -helix. *iv)* E5 presented a dynamic ensemble of conformers compared to E10, among which transient fold(s) with free  $\beta$ -edges could also be present, facilitating  $\beta$ -sheet formation and thus leading to aggregation. The latter phenomenon could be accelerated when intermediate misfolded structures were present at a higher concentration.<sup>[33]</sup> *v)* E10 did not form a considerable amount of  $\beta$ -sheet below 6 mM (data not shown), while E5 did at 10 times lower concentration (ca. 0.72 mM). This showed that in E5, the intermediate with free  $\beta$ -edges (ca. 1635 cm<sup>-1</sup>) was either more stable, or somehow was formed more easily than in the more tightly packed E10. This observation was in line with the “Dobson-concept”,<sup>[33]</sup> namely that intermediate stability is a key issue for opening a path toward aggregation (Figure 9). *vi)* E10 was the most resistant to increases in temperature and concentration, or decreases in pH; factors all promoting aggregation. E10 preserved its Trp-cage fold, which comprises an elongated  $\alpha$ -helical part up to very high practical concentration (ca. 6 mM). However, no globular fold can fully resist aggregation, one just has to find the suitable conditions. Indeed at pH lower than 2, with the gradual increase of *T* from 25 to 65 °C,  $\beta$ -sheet formation emerged even for the very stable E10 foldamer. *vii)* The firmly folded E10 and the rather unfolded E0 resisted aggregation by “choosing” two very different strategies. Indeed, both globular proteins and intrinsically disordered proteins (IDPs) can successfully resist against aggregation. The most vulnerable construct here was E5, a semi-stable foldamer (a molten globule), for which the free- $\beta$ -edge-containing intermediate state was more easily accessible and thus, aggregated the most rapidly, even at the mildest conditions. In conclusion, previous studies have showed that unfolded polypeptide models (insulin, albumin or additional globular proteins) could get into the aggregated state by increasing the critical concentration or temperature, or by lowering the pH. However, these model systems were all different and often did not reveal information on their intermediate structures leading to aggregation. Here we presented a set of model systems belonging to the same Trp-cage family. They had different, however comparable, molecular properties with respect to globular fold and aggregation. The present comprehensive analysis of E0, E5 and E10

foldamers supports that a loosened globular fold easily becomes defenseless against aggregation. With the appearance of the free  $\beta$ -edges, the molecular system quickly gravitates toward its aggregated state. These mini-protein foldamers could serve as a good example of larger globular proteins for which aggregation and amyloid fiber-like nanoparticle formations are often associated with Alzheimer's, Creutzfeldt–Jakob, prion<sup>[3]</sup> and other, so-called conformational diseases.

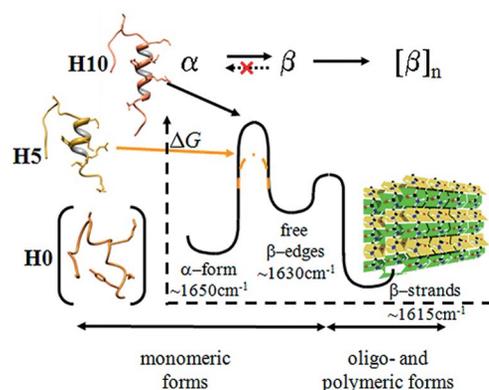


Figure 9. Schematic aggregation energy landscape of folded mini-proteins via intermediate(s). The  $\alpha$ -form stands for the folded structure, the “free  $\beta$ -edges” for the intermediate(s), and  $[\beta]_n$  for the aggregated structure (corresponding amide I' band marked).

## Experimental Section

**Sample Preparation and Purification:** E0, E5 and E10 mini-proteins were synthesized manually by using the standard Boc solid-phase peptide synthesis method. MBHA resin was used as a solid support, and the side-chain protecting groups were the following: Boc-Ser(Bzl), Boc-Arg(Tos), Boc-Glu(OcHex), Boc-Tyr(2BrZ), and Boc-Lys(2ClZ). Couplings were performed with DCC and HOBt. Amino acid incorporation was monitored by the quantitative ninhydrin test. The completed peptide resins were treated with liquid HF/dimethyl sulfide/*p*-cresol/*p*-thiocresol (86:6:4:2, vol/vol), at 0 °C for 45 min. HF was removed, and the resulting free peptides were solubilized in 10% aqueous acetic acid, filtered and lyophilized. The crude peptides were purified by reverse-phase HPLC on a Phenomenex Jupiter C-18 column (21.2 × 250 mm), using water/acetonitrile mixtures of 0.1% TFA in water (A) and 0.1% TFA, 80% acetonitrile in water (B), with a gradient of 0% → 40% B over

Table 2. Selected analytical properties of the three different mini-proteins.

Mini-protein	Yield [mg]	MS calculated	MS measured	HPLC data ( <i>t<sub>R</sub></i> )
E0	12	2224.54	2224.6	7.60 min <sup>[a]</sup>
E5	13.8	2782.10	2782.8	14.60 min <sup>[b]</sup>
E10	7.5	3369.83	3370.0	12.24 min <sup>[c]</sup>

[a] Column: Lichrosorb 10  $\mu$  C-18 100 $\Delta$  (4.6 × 250 mm), flow: 1.2 mL/min, gradient 30 → 45%B in 15 min. [b] Column: Phenomenex Jupiter 10  $\mu$  C-18 300 $\Delta$  (4.6 × 250 mm), flow: 1.2 mL/min, gradient 25 → 45%B in 20 min. [c] Column: Phenomenex Jupiter 10  $\mu$  C-18 300 $\Delta$  (4.6 × 250 mm), flow: 1.2 mL/min, gradient 30 → 50%B in 20 min.

## FULL PAPER

80 min, at a flow 3 mL/min, and UV detection at 220 nm. Collected fractions were lyophilized, and the identities of the resulting peptides were confirmed by using a Finnigan TSQ 7000 tandem quadrupole mass spectrometer equipped with an electrospray ion source. The analytical data of the peptides are summarized in Table 2.

**FT-IR Spectroscopy:** Spectra of E0, E5 and E10 were recorded with a Bruker-Equinox 55 spectrometer equipped with a liquid N<sub>2</sub>-cooled MCT detector. Measurements were carried out in the 0.05 mm fixed pathlength cell containing BaF<sub>2</sub> windows, from 4000 to 750 cm<sup>-1</sup>, at 4 cm<sup>-1</sup> resolution, in D<sub>2</sub>O. All IR absorption spectra were solvent-subtracted. Spectra of E5 were measured at concentrations from 0.36 to 7.2 mM, at pH 3.0 and 25 °C. In addition, pH-dependent measurements (pH 4.0, pH 5.4, pH 7.0) were completed at 25 °C over a concentrations range of 1.0 to 2.5 mg/mL (0.36 to 0.72 mM). Spectra of E10 were recorded over a concentration range of 1.0 to 2.5 mg/mL (0.3 to 0.6 mM) at pH 3.0 and pH 7.0, at 25 °C. Deconvolution of the spectra was carried out using the Curve Fit procedure of the OPUS 6.0 software.

**ECD Spectroscopy:** Far-UV ECD measurements were completed with a Jasco J810 dichrograph in 0.1 cm quartz cells. The temperature at the cell was controlled by a Peltier-type heating system. For each measurement, the sample in the cell was allowed to equilibrate for 5 min at the adjusted temperature prior to data acquisition. Each spectrum was obtained by averaging a total of 4 scans. The solvent reference spectra were used as baselines, automatically subtracted from the peptide ECD spectra. ECD band intensities were expressed in mean residue ellipticity ( $[\theta]_{MR}$ , deg × cm<sup>2</sup>/dmol).

**VCD Spectroscopy:** All VCD spectra of E0, E5 and E10 were recorded in D<sub>2</sub>O with a Bruker PMA 37 VCD/PM-IRRAS module connected to an Equinox 55 FT-IR spectrometer at a resolution of 4 cm<sup>-1</sup>. The ZnSe photoelastic modulator of the instrument was set to 1600 cm<sup>-1</sup>, and an optical filter with a transmission range of 1960–1250 cm<sup>-1</sup> was used in order to increase the sensitivity in the amide I–II spectral region. The instrument was calibrated for VCD intensity with a CdS multiple-wave plate. A CaF<sub>2</sub> cell of 0.055 mm path length was used. In order to improve the S/N ratio, spectra were averaged for 7 h (corresponding to ca. 24500 accumulated interferograms). Baseline correction was achieved by subtracting the spectrum of the solvent obtained under the same conditions. IR spectra were calculated from the single-channel DC spectra of the sample and solvent, respectively.

**Supporting Information** (see footnote on the first page of this article): Additional FT-IR spectra, as well as HPLC, ESI-MS, <sup>1</sup>H NMR characterizations of the purified compounds.

## Acknowledgments

The authors thank to Dr. Elemér Vass, for help in VCD measurements and Petra Rovó for NMR spectroscopic data and scientific consultation. This work was supported by grants from the Hungarian Scientific Research Fund (grant numbers OTKA K72973, NK101072 and TÁMOP-4.2.1.B-09/1/KMR).

- [1] C. B. Anfinsen, *Nature* **1973**, *181*, 223–230.
- [2] C. M. Dobson, *Nature Struct. Mol. Biol.* **2006**, *13*, 295–297.
- [3] C. M. Dobson, *Nature* **2005**, *435*, 747–749.
- [4] C. F. Wright, S. A. Teichmann, J. Clarke, C. M. Dobson, *Nature* **2005**, *438*, 878–881.
- [5] A. Perczel, P. Hudáky, V. K. Pálfi, *J. Am. Chem. Soc.* **2007**, *129*, 14959–14965.

- [6] J. Kardos, B. Kiss, A. Micsónai, J. Kovács, P. Rovó, Gy. Váradi, G. K. Tóth, A. Perczel, *Chem. Eur. J.* **2013**, manuscript submitted.
- [7] R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe, D. Eisenberg, *Nature* **2005**, *435*, 773–778.
- [8] T. A. Martinek, I. M. Mándity, L. Fülöp, G. K. Tóth, E. Vass, M. Hollósi, E. Forró, F. Fülöp, *J. Am. Chem. Soc.* **2006**, *128*, 13539–13544.
- [9] T. Buffeteau, L. Ducasse, L. Poniman, N. Delsuc, I. Huc, *Chem. Commun.* **2006**, 2714–2716.
- [10] L. Ducasse, F. Castet, A. Fritsch, I. Huc, T. Buffeteau, *J. Phys. Chem. A* **2007**, *111*, 5092–5098.
- [11] A. Hetényi, G. K. Tóth, Cs. Somlai, E. Vass, T. A. Martinek, F. Fülöp, *Chem. Eur. J.* **2009**, *15*, 10736–10741.
- [12] G. Longhi, S. Abbate, F. Lebon, N. Castellucci, P. Sabatino, C. Tomasini, *J. Org. Chem.* **2012**, *77*, 6033–6042.
- [13] S. Ma, X. Cao, M. Mak, A. Sadik, C. Walkner, T. B. Freedman, I. K. Lednev, R. K. Dukor, L. A. Nafie, *J. Am. Chem. Soc.* **2007**, *129*, 12364–12365.
- [14] D. Kuruski, R. A. Lombardi, R. K. Dukor, I. K. Lednev, L. A. Nafie, *Chem. Commun.* **2010**, *46*, 7154–7156.
- [15] L. N. Arnaudov, R. de Vries, *Biophys. J.* **2005**, *88*, 515–526.
- [16] G. Shanmugam, N. Phambu, P. L. Polavarapu, *Biophys. Chem.* **2011**, *155*, 104–108.
- [17] J. Kardos, D. Okuno, T. Kawai, Y. Hagihara, N. Yumoto, T. Kitagawa, P. Závodszy, H. Naiki, Y. Goto, *Biochim. Biophys. Acta Proteomics Proteomics Biochim. Biophys. Acta* **2005**, *1753*, 108–120.
- [18] G. Shanmugam, P. L. Polavarapu, *J. Struct. Biol.* **2011**, *176*, 212–219.
- [19] T. J. Measey, R. Schweitzer-Stenner, *J. Am. Chem. Soc.* **2011**, *133*, 1066–1076.
- [20] J. P. Schmittschmitt, J. M. Scholtz, *Prot. Sci.* **2003**, *12*, 2374–2378.
- [21] S. Poon, N. R. Birkett, S. B. Fowler, B. F. Luisi, C. M. Dobson, *J. Zurdo, Prot. Pep. Lett.* **2009**, *16*, 1548–1556.
- [22] X. Chang, D. Keller, S. I. O'Donoghue, J. J. Led, *FEBS Lett.* **2002**, *515*, 165–170.
- [23] S. R. Drab, *Pharmacotherapy* **2010**, *30*, 609–624.
- [24] J. W. Neidigh, R. M. Fesinmeyer, N. H. Andersen, *Nat. Struct. Biol.* **2002**, *9*, 425–430.
- [25] X. Wang, J. F. Espinosa, S. H. Gellman, *J. Am. Chem. Soc.* **2000**, *122*, 4821–4822.
- [26] S. Chowdhury, M. C. Lee, G. Xiong, Y. Duan, *J. Mol. Biol.* **2003**, *327*, 711–717.
- [27] R. Zhou, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13280–13285.
- [28] J. Juraszek, P. G. Bolhuis, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15859–15864.
- [29] J. Copps, R. F. Murphy, S. Lovas, *Biopolymers* **2007**, *88*, 427–437.
- [30] P. Rogne, P. Ozdowy, C. Richter, K. Saxena, H. Schwalbe, L. T. Kuhn, *PLoS ONE* **2012**, *7*, e41301.
- [31] P. Rovó, P. Stráner, A. Láng, I. Bartha, K. Huszár, L. Nyitray, A. Perczel, *Chem. Eur. J.* **2013**, *19*, 2628–2640.
- [32] P. Rovó, V. Farkas, P. Stráner, G. K. Tóth, A. Perczel, **2013**, in preparation.
- [33] F. Chiti, C. M. Dobson, *Nat. Chem. Biol.* **2009**, *5*, 15–22.
- [34] P. Hudáky, P. Stráner, V. Farkas, Gy. Váradi, G. K. Tóth, A. Perczel, *Biochemistry* **2008**, *47*, 1007–1016.
- [35] P. Rovó, V. Farkas, O. Hegyi, O. Szolomajer-Csikós, G. K. Tóth, A. Perczel, *J. Pept. Sci.* **2011**, *17*, 610–619.
- [36] M. Sunde, C. Blake, *Adv. Protein Chem.* **1997**, *50*, 123–159.
- [37] J. C. Chan, N. A. Oyler, W. M. Yau, R. Tycko, *Biochemistry* **2005**, *44*, 10669–10680.
- [38] R. Nelson, D. Eisenberg, *Curr. Opin. Struct. Biol.* **2006**, *16*, 1–6.
- [39] J. W. Neidigh, R. M. Fesinmeyer, K. S. Prickett, N. H. Andersen, *Biochemistry* **2001**, *40*, 13188–13200.
- [40] T. A. Keiderling, *Curr. Opin. Chem. Biol.* **2002**, *6*, 682–688.
- [41] A. Barth, *Quart. Rev. Biophys.* **2002**, *35*, 369–430.

## Stability and Aggregation Propensity of Mini-protein Foldamers

- [42] J. S. Richardson, D. C. Richardson, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2754–2759.
- [43] A. Perczel, I. Lengyel, H. H. Mantsch, G. D. Fasman, *J. Mol. Struct.* **1993**, *297*, 115–126.
- [44] K. Itoh, B. M. Foxman, G. D. Fasman, *Biopolymers* **1976**, *15*, 419–455.
- [45] A. Fulara, W. Dzwolak, *J. Phys. Chem. B* **2010**, *114*, 8278–8283.
- [46] A. Fulara, A. Lakhani, S. Wójcik, H. Nieznanska, T. A. Keiderling, W. Dzwolak, *J. Phys. Chem. B* **2011**, *115*, 11010–11016.

Received: January 15, 2013

Published Online: ■