Dimorphic aggregation behavior of a fusion polypeptide incorporating a stable protein domain (EGFP) with an amyloidogenic sequence (retroCspA)

Swati Sharma, Purnananda Guptasarma*

Protein Science and Engineering Division, Institute of Microbial Technology (IMTECH), Chandigarh 160 036, India
Council of Scientific and Industrial Research, New Delhi 110 001, India

Abstract We describe the behavior of a polypeptide consisting of the genetic fusion of a structurally stable single-domain protein, EGFP (an analog of the green fluorescent protein) with an amyloidogenic sequence, retroCspA (known to readily form amyloid fibrils). Refolding of the fusion protein through single-step removal of denaturant and salt results in precipitation into amyloid aggregates displaying fibrillar morphology, thioflavin T binding as well as green fluorescence. Refolding through step-wise reduction of denaturant concentration in the presence of salt yields a soluble aggregate containing a folded, thermally-stable, non-fluorescent EGFP domain. Together, these results indicate that retroCspA forces the fusion protein to aggregate; however, the EGFP domain remains folded in a native-like structural format in both soluble aggregates and precipitates.
© 2008 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: Beta sheet propagation; Amyloid formation; Stable protein domains; Fusion proteins; Conformational change

1. Introduction

Polypeptide chains within amyloid fibrils tend to be organized into 'cross-beta' structures that have β-strands running perpendicular to the long axis of the fibril [1]. Many different proteins with different secondary structural compositions are reported to form amyloid fibrils, following prolonged incubation in structurally-destabilizing (e.g., acidic) environments [2–4]. Formation of amyloids can thus involve profound changes in protein conformation, suggesting that the ease with which a given protein transforms into amyloid fibrils may be correlated with the extent to which its native structure is amenable to structural reorganization, i.e., with the degree to which its structure is thermodynamically and kinetically stable. Support for this view comes from reports indicating that, for naturally-occurring proteins, the presence of disordered polypeptide chain segments and/or conditions favorable to unfolding could be required for deposition into amyloid forms [5]. A detailed comparison of the kinetics, and quantum, of amyloid formation shown by different proteins of different (known) structural contents, and stabilities, could be expected to provide critical insights in this regard. However, in this paper, we have taken a different approach to explore how protein structural stability relates to amyloid-forming potential.

We genetically fused a known amyloid-forming polypeptide sequence to the C-terminus of a single-domain protein of high structural stability, to examine whether the amyloidogenic sequence pulls the stable domain down with it into a precipitated, amyloid-like state and, if so, whether the stable domain is deposited in a native-like, or reorganized, structure. Since a short amyloidogenic sequence fused to a large stable protein domain could fail to engage in intermolecular interactions for purely steric reasons, we decided arbitrarily that the length of the amyloidogenic sequence would need to be about one-half or at least one-third the length of the stable domain, to keep the latter from sterically occluding the former.

We chose the 238 amino acids-long enhanced analog (EGFP) of the well-known green fluorescent protein (GFP) of *Aquorea Victoria* [6], as the stable protein domain. The EGFP domain is an ideal choice as a stable domain for three reasons: (1) it refolds quantitatively within minutes of removal of denaturant without any attendant aggregation [7]; (2) it requires either an extreme of pH (pH < 4.0 or >12.0), or a combination of high denaturant concentration (6 M guanidinium hydrochloride) and high temperature (90 °C), to undergo any detectable unfolding [8]; and (3) it shows the characteristic GFP fluorescence that serves as a spectroscopic signature of 'native' conformational status, since GFP fluorescence is known to be critically dependent on the maintenance of the correct three-dimensional structure, with dramatic quenching accompanying even the slightest structural deformation [9]. To the C-terminus of the EGFP domain, we fused an amyloidogenic polypeptide encoding a backbone-reversed form of the cold shock protein A of *Escherichia coli* (retroCspA). This 82 residues-long polypeptide is an ideal choice as an amyloidogen because it is known to readily and quantitatively form amyloid fibrils in aqueous solution over a timescale of hours at room temperature, following purification and refolding, without requiring any specific conditions for amyloid formation [10].

Below, we describe our findings which suggest that the EGFP-retroCspA fusion protein forms soluble or insoluble (amyloid) aggregates, depending on the manner of refolding,

*Corresponding author. Address: Protein Science and Engineering Division, Institute of Microbial Technology (IMTECH), Chandigarh 160 036, India. Fax: +91 172 2690585.
E-mail address:* pg@imtech.res.in (P. Guptasarma).

URL: http://webmail.imtech.res.in/~pg/.

and the solvent conditions maintained during refolding. We also show that retroCspA pulls the EGFP domain down into insoluble aggregates within which the EGFP domain remains quantitatively folded, displaying green fluorescence.

2. Materials and methods

2.1. Creation of the EGFP-T-retroCspA fusion construct

The DNA construct (see supplementary Fig. S1) encoding the EGFP-T-retroCspA fusion protein encodes the EGFP sequence at its 5′ end and the retroCspA sequence at its 3′ end, separated by a thrombin cleavage sequence (also functioning as a flexible linker), with a total length of 335 amino acids and a calculated molecular weight of 37.4 kDa. The thrombin cleavage site was introduced to separate the fusion partners if necessary, at any stage of the studies, and also to provide the two proteins with the flexibility to be somewhat independent of each other and thus display their own characteristic behavioral features. Several PCR reactions were performed to create the construct, as described below. The first PCR reaction was carried out with primers that amplified the EGFP sequence from the pEGFP vector (Clontech) to incorporate a 5xHis sequence for affinity purification at the 5′ end of the EGFP domain, along with a thrombin cleavage sequence at the 3′ end of the EGFP domain. The forward and reverse primers specified below were used for this reaction. Forward Primer 1: **5′ TAT ACT CCA AGC TTG CAT CAT CAT CAC CAC ATG GTG AGC AAG GGC GAG 3′**. Reverse Primer 1: **5′ GTG ATG CGG GGA TCC ACG TGG AAC CAG CTT GTA CAG CTC GTC CAT GCC 3′**. For the second PCR reaction (preparing retroCspA for genetic fusion), it may be noted that a gene encoding retroCspA was already available in the lab, cloned into the Qiagen pQE-30 vector [10]. We amplified the retroCspA sequence from the pQE-30 vector using the following primers, which modified the sequence at its 5′ end to incorporate the necessary overlapping sequences for splicing by overlap-extension (SOE) PCR with the 3′ end of the EGFP-encoding construct: Forward Primer 2: **5′ CCA CGT GGA TCC CCG CAT CAC CAT CAC CAT CAC CTG AGC ACC 3′**. Reverse Primer 2: **5′ ACT GTC GCG GCC GCT TTA CAT GGA ACC TTT CAT AGT ACC 3′**. The forward primer 2 also incorporates a 6xHis encoding sequence that become located at the 5′ end of the retroCspA gene, and the reverse primer 2 incorporates a NotI restriction site at the 3′ end of the retroCspA gene, for cloning into the pEGFP vector. In a third PCR reaction, the products of the first and the second reactions were cleaned up and joined through SOE PCR using the forward primer 1 and reverse primer 2 specified above, to create a full-length gene encoding the fusion, EGFP-T-retroCspA, with 'T' denoting a thrombin cleavage site. The final PCR product and the pEGFP vector were double digested with HindIII and NotI enzymes, ligated, and transformed into BL21(DE3)pLysS E. coli. Protein was produced through overexpression induced by IPTG.

2.2. Expression of recombinant proteins

Overnight grown cultures of bacteria containing pEGFP clones encoding EGFP-T-retroCspA were inoculated at a final concentration of 1% of total volume into LB medium containing appropriate antibiotics and grown to an OD_{600} of ∼0.6 at 37 °C with shaking at 220 rpm. Cells were induced with 0.5 mM IPTG and grown for a further period of 5 h at 30 °C after induction. Appropriate amounts of control (uninduced) and induced cultures (after 2 h and 4 h of induction) were each loaded on SDS–PAGE after harvesting and boiling of cells in SDS–PAGE sample buffer, to check for expression of the recombinant protein EGFP-T-retroCspA protein.

2.3. Denaturing purification of EGFP-T-retroCspA

The EGFP-T-retroCspA fusion polypeptide deposited within inclusion bodies upon production (see Fig. 1A) and could not be recovered in soluble form after bacterial cell lysis by either of the two standard procedures for affinity purification suggested by the Ni-NTA resin manufacturer (Qiagen), i.e., using either denaturing (8 M urea- or 6 M GdmCl-containing) buffers, or non-denaturing buffers. Therefore, we tested combinations of different denaturing influences to solubilize the fusion from the cell lysis pellet, and finally managed to solubilize it by using a combination of heat, salt and 6 M GdmCl, as described below. First we suspended the cell pellet in extraction buffer containing 6 M GdmCl, 1 M NaCl, 0.1 M NaH_2PO_4, 0.01 M Tris, pH8.0 at 5 ml per gram wet weight, with shaking at 37 °C for 30 min. This was followed by heating of the suspension at 70 °C for 30 min, after which the suspension was sonicated for 10–15 min. The lysate was then spun at $15\,000 \times g$ for 1 h and the supernatant loaded onto a Ni-NTA column pre-equilibrated with the same buffer. After collecting the column's flow-through, the column was re-equilibrated with a urea-containing buffer (8 M urea, 10 mM Tris, pH 8.0, 1 M NaCl, 0.1 M NaH_2PO_4) that effectively replaced the GdmCl-containing buffer used at earlier stages. Hereafter, we proceeded with the standard protocol for Ni-NTA purification under denaturing conditions suggested by the resin manufacturer (Qiagen). It may be noted that we also attempted to extract the fusion protein by individually using heat, salt and denaturant, and combinations of two of these factors, to try and dissect out which of the three factor(s) were predominantly responsible for the extraction. We discovered that extraction occurred only by using a combination of all three factors, i.e., denaturant (6 M
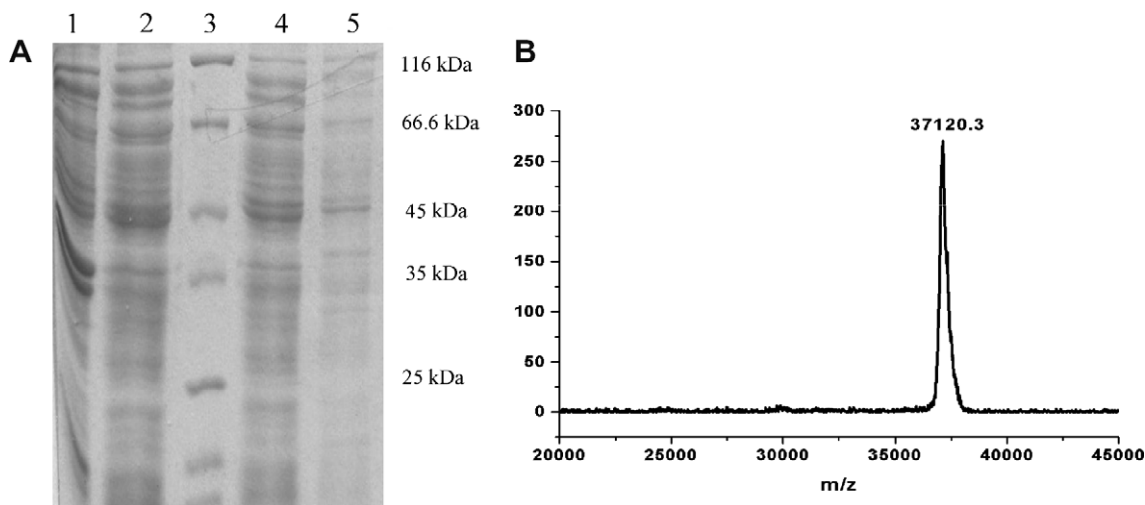

Fig. 1. Panel A: Insolubility of the fusion protein and lack of its presence in lysed cell supernatants, as evidenced by SDS–PAGE data. Lane 1: Pellets of lysed cells expressing EGFP-T-retroCspA; lane 2: supernatant of lysed cells; lane 3: molecular weight markers; lane 4: Ni-NTA column flow through (lane 4); lane 5: Ni-NTA column wash. Panel B: MALDI-TOF mass spectrum of the EGFP-T-retroCspA fusion, following extraction by heat, salt and denaturant, purification by Ni-NTA chromatography, and precipitation through refolding.

GdmCl), salt (1 M NaCl) and heating (70 °C). No individual factor, or combination of two factors, achieved solubilization of the fusion (data not shown).

2.4. Refolding of EGFP-T-retroCspA through step-wise removal of denaturant while retaining salt

The urea in the purified protein fractions eluted from the Ni-NTA column was removed through step-wise dialysis against a 100-fold higher volume of 50 mM Tris pH 8.5 buffer containing 1 M NaCl and progressively lower concentrations of urea (differing by 1 M urea), down from 8 M urea, with the final dialysis performed to remove 1 M urea against a buffer containing no urea, and employing multiple buffer changes. Under these conditions, very little visible aggregation was seen. The solution was centrifuged at $12000 \times g$ for 30 min and the supernatant was subjected to further studies.

2.5. Refolding of EGFP-T-retroCspA through single-step removal of denaturant and salt

A second method was used to remove the urea from purified protein fractions eluted from the Ni-NTA column, as follows. The 8 M urea-containing solution was directly dialyzed against 20 mM Tris pH 8.0 buffer containing no NaCl. This direct dialysis against a solution lacking salt led to very heavy aggregation and precipitation of the fusion protein, unlike what was seen during step-wise dialysis done with 1 M NaCl retained on both sides of the dialysis membrane (see Section 2.4). The obtained precipitate was repeatedly washed with the same buffer to remove any soluble protein, and subjected to further studies.

2.6. Protein concentration estimation and UV/visible absorption spectroscopy

Concentration of purified EGFP-T-retroCspA protein was estimated through measurement of absorbance at 280 nm ($A_{280 \, nm}$). The value obtained was compared to the theoretically predicted value based on amino acid composition, determined using Vector NTI software. An absorbance of 1.0 at 280 nm was taken to correspond to a protein concentration of 1.39 mg/ml. UV/visible absorption studies were performed on a Cary 50-Bio spectrophotometer.

2.7. Gel filtration chromatography

The gel filtration elution profile of EGFP-T-retroCspA was acquired on a Pharmacia SMART system fitted with Superdex-75 or Superdex-200 micro-columns, each of 2.4 ml bed volume. Samples of 50 µl volume were loaded for micro-purification or analysis of changes in hydrodynamic volumes. Elution volumes were compared with a calibration profile collected using protein standards. Fusion protein elutions were simultaneously monitored using light of 220 (peptide absorption), 280 (protein aromatic absorption), and 488 nm (EGFP chromophore absorption), as appropriate.

2.8. Fluorescence spectroscopy

Fluorescence spectra were acquired on a Perkin–Elmer LS-50B spectrofluorimeter, or a Jasco J-810 spectropolarimeter fitted with a fluorimetric (FMO-427) monochromator attachment. For GFP emission, 478 nm excitation was used, with emission being recorded between 497 and 550 nm. Excitation and emission bandpasses were set at 5 and 10 nm, respectively, with scanning done at 60 nm/min, and with all spectra averaged over 10 scans. For emission studies of thioflavin T (ThT, which is known to become strongly fluorescent upon binding to amyloid fibrils), protein samples were incubated for 1 h with ThT (using a final ThT concentration of 6 µM with protein suspended in 10 mM Tris, pH 8.0) and emission was monitored between 455 and 550 nm, using 444 nm excitation, excitation and emission bandpasses of 7 nm each, and scanning at 100 nm per minute.

2.9. Circular dichroism (CD) spectroscopy

Far-UV CD spectra were collected on a Jasco J810 CD spectrometer in cuvettes of appropriate path lengths (1 cm/0.5 cm/0.2 cm) with flushing of Iolar grade I nitrogen gas at 6–12 l/min. Ellipticity data was collected in the range of 250–190 or 250–200 nm. Raw ellipticity was converted to mean residue ellipticity [θ] using the formula shown below, in which MRW stand for mean residue weight (total molecular weight of the protein/ total number of amino acids), [θ] stands for mean residue ellipticity, expressed in degrees cm² dmol⁻¹. θ_{obs} stands for observed (raw) ellipticity in millidegrees:

$$[\theta] = \{\theta_{obs} \times 100 \times MRW\}/\{1000 \times \text{concentration (mg/ml)} \times \text{path length (cm)}\}$$

2.10. MALDI-TOF mass spectrometry

MALDI-TOF mass spectral measurements were performed on an ABI Voyager DE-STR spectrometer (Applied Biosystems). Protein samples were prepared for analysis by mixing with an equal volume of solution containing the matrix (Sinapinic acid), and loading a total volume of 2 µl per well of the sample plate. Samples were setup in duplicates, or triplicates. Matrix solution stocks (10 mg/ml) were prepared in 0.3% TFA (trifluoroacetic acid), 50% acetonitrile.

2.11. Detection of fluorescence and imaging of insoluble EGFP-T-retroCspA aggregates

Precipitated aggregates were visualized through excitation by the 488nm laser on a Zeiss Confocal LSM 510 Meta microscope. Fluorescence spectra were recorded on the Jasco J-810 spectropolarimeter, following suspension of aggregates. For the EGFP emission spectrum, a suspension of aggregated fusion protein was excited by 478 nm wavelength, with excitation and emission bandpasses set at 7 nm each.

2.12. Electron microscopy

Transmission electron microscopy (TEM) of the protein aggregates was done in a JEOL 1200 EX2 microscope using negative staining with phosphotungstic acid (PTA) and uranyl acetate, using standard protocols.

3. Results and discussion

3.1. Fusion protein behavior during expression

Within the E. coli BL21(DE3)pLysS cells used as expression host, EGFP-T-retroCspA was found to have been expressed to high levels. The cells were brightly fluorescent, showing the characteristic green fluorescence of GFP or EGFP, upon excitation with ultraviolet light from a standard laboratory UV torch. However, all of the expressed protein was in insoluble aggregates, with no soluble protein detectable in the supernatant upon centrifugation of bacterial cell lysates (Fig. 1A). No solubilization/extraction of the insoluble fusion protein could be achieved even through use of 8 M urea, or 6 M GdmCl. Only by using a combination of heat (70 °C), salt (1 M NaCl) and denaturant (6 M GdnCl) could EGFP-T-retroCspA be extracted from the deposits and solubilized. GFP, or EGFP, domains are themselves highly soluble, and also amongst the most fusion-friendly proteins known to man; fusions of these domains with various other proteins are commonly used as cytoplamically-soluble fluorescent reporters of protein cellular localization. Prima facie, therefore, it would seem that it is the presence of the retroCspA chain in fusion with EGFP which causes EGFP-T-retroCspA to precipitate into urea- and GdnCl-insoluble aggregates during biosynthesis.

We proceeded to try and obtain the fusion protein in a soluble and unfolded form, as described above, by using a combination of salt, heat and denaturant (see also Section 2.3), to examine its ability to refold to soluble folded form and/or aggregate and precipitate. We solubilized EGFP-T-retroCspA from pellets of lysed bacterial cells, cooled the extracts, confirmed by SDS-PAGE whether the protein had been solubilized and extracted, performed IMAC purification of the

fusion protein in the presence of denaturant and salt on a Ni-NTA column (see supplementary Fig. S2). We then dialyzed samples to remove the denaturant and salt used for solubilization/extraction through a single-step dialysis procedure (see Section 2.5), and found that the fusion protein precipitates into solid aggregates. The characteristics of these aggregates are described in the following sections.

3.2. EGFP-T-retroCspA precipitates contain full-length fusion protein

Insoluble aggregates of the EGFP-T-retroCspA were obtained through single-step dialysis of the purified denatured protein into 20 mM Tris, pH 8.0 buffer lacking NaCl. These were repeatedly washed through cycles of resuspension in buffer, and centrifugation, to ensure the absence of trace amount of any soluble protein. The washed aggregate was then solubilized in acetonitrile containing 0.1% trifluoroacetic acid and sinapinic acid and subjected to MALDI-TOF mass spectrometry, to examine whether the aggregates contain only the fusion polypeptide construct, or also some naturally proteolysis-derived populations of free EGFP domain, or retroCspA. Fig. 1B establishes that the aggregated sample consists entirely of the full-length fusion protein, characterized by a single sharp peak of the correct expected mass.

3.3. EGFP-T-retroCspA precipitates show EGFP fluorescence

Washed EGFP-T-retroCspA aggregates showed a visible green color under ordinary conditions of room lighting, akin to the color associated with a solution of purified concentrated

GFP, owing presumably to the capability of natural lighting to excite the green fluorescence. The aggregates were suspended in buffer, and examined for GFP fluorescence on a spectrofluorimeter, with light of 478 nm used for excitation instead of light of 488 nm, in order to reduce the Rayleigh scatter component from the short-wavelength end of the EGFP emission spectrum (this being necessitated by the closeness of the excitation and emission maximal wavelengths of EGFP, and the heightened Rayleigh scattering expected from suspended aggregates). As Fig. 2A shows, a very clear emission from the EGFP green fluorophore is seen in the suspended aggregate with a peak at ~508 nm, indicating the presence of folded EGFP protein chains in native conformation in the population of resuspended insoluble aggregates. Further, we also examined the aggregates on a confocal microscope through excitation by 488 nm light from an Argon laser. Fig. 2A–C, which represent different clumps of aggregates within the field of the confocal microscope, show that the aggregates display a uniform green EGFP fluorescence. This confirms that folded, and fluorescent, EGFP is indeed incorporated into the solid precipitates, because the EGFP fluorophore is well known to be protected and stabilized by the tertiary structure of the green fluorescent protein, with the slightest perturbations of tertiary structure severely quenching fluorescence.

3.4. EGFP-T-retroCspA precipitates contain amyloid-like nanofibril structures

Fig. 3A and B show representative TEM photomicrographs of two fields of fusion protein aggregates containing isolated
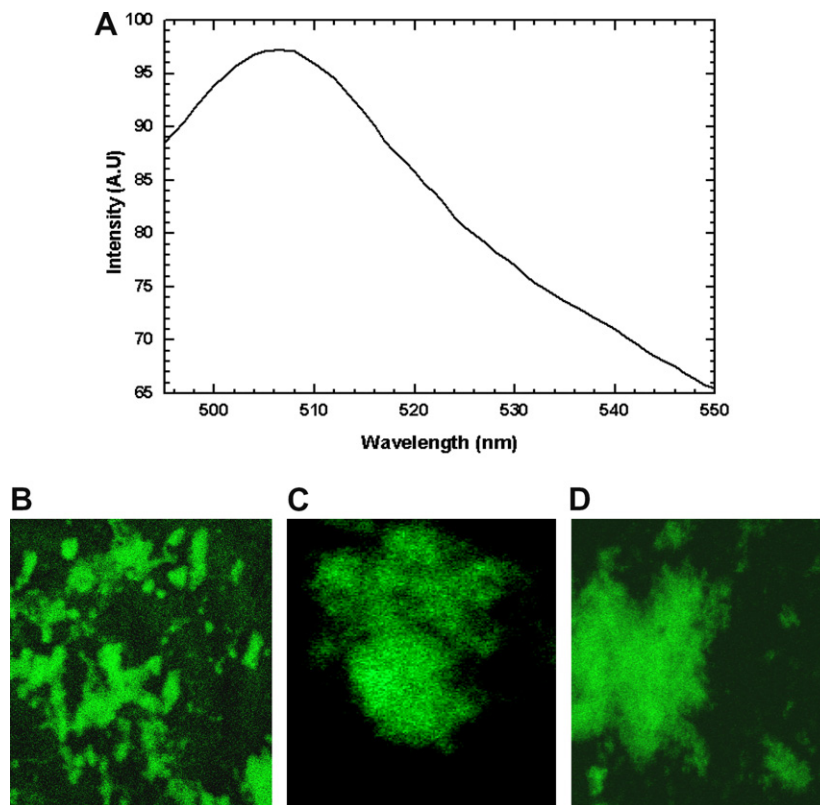

Fig. 2. Green fluorescence in EGFP-T-retroCspA precipitates. Panel A: Fluorescence emission spectrum of washed and resuspended EGFP-T-retroCspA precipitates excited by 478 nm light. Panel B: Confocal microscopic fluorescence images of EGFP fluorescence from within precipitated material, acquired through excitation of solid aggregates by 488 nm laser light.
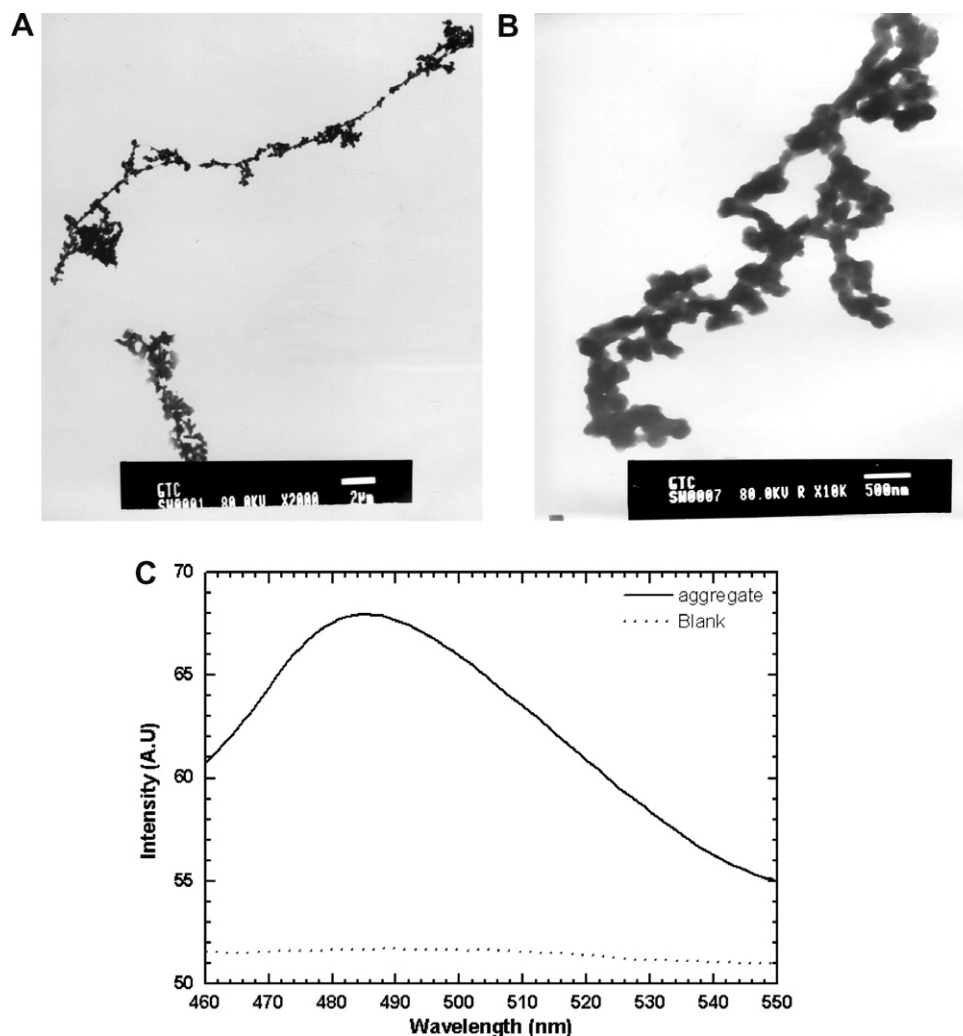
Fig. 3. Amyloid formation by EGFP-T-retroCspA. Panels A and B: Transmission electron micrographs of EGFP-T-retroCspA precipitates at two different magnifications, showing a filamentous-beaded morphology, with interspersing of electron-dense and non-dense regions. Panel C: Fluorescence emission from thioflavin T (ThT) bound to resuspended EGFP-T-retroCspA precipitated; the dotted line represents the emission from the control ThT solution (6 μM) in 10 mM Tris, pH 8.0 buffer, and the solid line represents the emission from the same concentration of ThT in the presence of resuspended EGFP-T-retroCspA precipitates.

fibril-like structures, at different magnifications. These fibers arise, in all likelihood, from the amyloid-forming propensity of the retroCspA component known to readily form amyloid fibers [10]. However, the fibers seen in the micrograph differ significantly in morphology from the fibrils made by retroC-spA on its own; also, an interspersing of electron-dense and non-dense structures, as well as branching, can be seen all along the fibers in the magnified photomicrograph in Fig. 3B.

3.5. Thioflavin T binds to EGFP-T-retroCspA precipitates

To examine whether the fiber-like fusion protein precipitates also contain amyloid-like microstructure, we carried out the well-known ThT binding assay. This assay is commonly used to detect cross-beta structures, because ThT intercalates into cross-beta structures formed by beta strands in amyloid fibrils and displays fluorescence upon excitation with 444 nm light, with peak emission at ~482 nm. Fig. 3C shows the emission of ThT bound to suspended precipitates of EGFP-T-retroC-spA, displaying the characteristic ~482 nm emission peak, confirming the presence of amyloid microstructures.

3.6. High molecular weight soluble aggregates of EGFP-T-retroCspA

To examine whether EGFP-T-retroCspA can be kept from precipitating during removal of denaturant, we attempted various combinations of procedures. We found that use of a step-wise dialysis procedure effecting gradual removal of the urea in steps of 1 M, combined with retention of 1 M NaCl on both sides of the dialysis membrane (see Section 2.4) caused EGFP-T-retroCspA to remain soluble, showing no precipitation even upon complete removal of urea through extensive dialysis of the final 1 M urea-containing solutions in the last step of this step-wise dialysis. The secondary, tertiary and quaternary structural features of this soluble EGFP-T-retroCspA fusion, as well as its thermal stability characteristics, were studied. Purified, soluble EGFP-T-retroCspA was chromatographed on a superdex-200 gel filtration column in 20 mM Tris, pH 8.0 buffer (SMART system, Pharmacia) to examine its hydrodynamic volume and quaternary structural state. Fig. 4A and B shows the elution profiles of the purified EGFP-T-retroCspA and wild-type EGFP. Both panels show
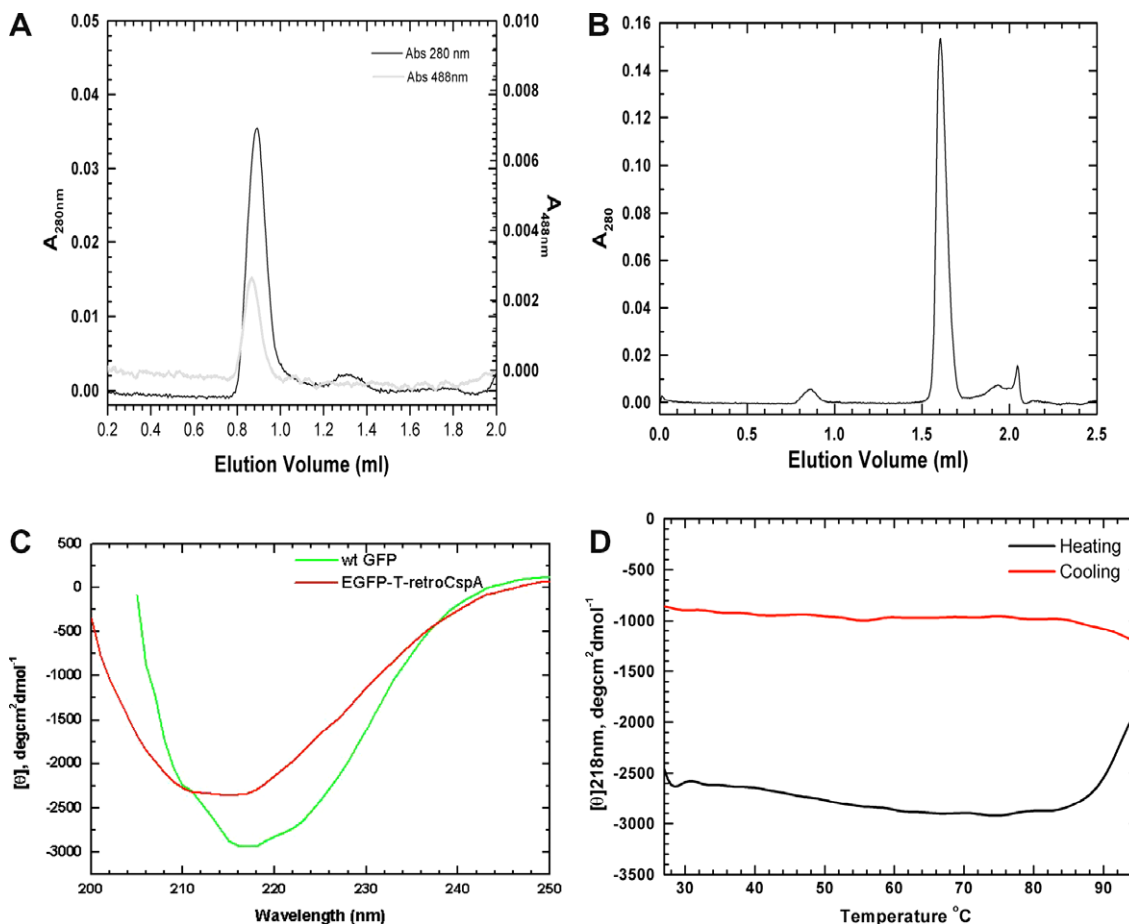
Fig. 4. Panel A: Gel filtration elution profiles of purified soluble EGFP-T-retroCspA on a Superdex 200 SMART column, showing 280 nm (dark line) and 488 nm absorbance (light line). Panel B: Gel filtration elution profiles of control green fluorescent protein on the same column, showing 280 nm (dark line) and 488 nm absorbance (light line). Panel C: Far-UV CD spectra of soluble EGFP-T-retroCspA aggregates (red) and green fluorescent protein (green) at 25 °C. Panel D: Change in the 218 nm MRE signal of soluble EGFP-T-retroCspA as a function of increasing (black) and decreasing (red) temperature.

double-Y axis plots of the respective elution profiles monitored in terms of 280 nm absorbance owing to aromatic residues (left Y-axis) as well as 488 nm absorbance owing to the EGFP chromophore (right Y-axis). Monitoring of both absorptions allows specific assessment of which eluting components contain the EGFP chromophore. As Fig. 4A shows, the form of the protein displaying absorption at both 280 and 488 nm elutes at ~0.9 ml, close to the void volume of the column (the column's bed volume is 2.4 ml). Elution at the column's void volume indicates that the protein exists entirely as soluble aggregates, larger in size than 600 000 Da, since the Superdex-200 column optimally fractionates proteins between 10 000 and 600 000 Da, and has an exclusion limit of 1.3×10^6 Da. In contrast, Fig. 4B shows that the purified wild-type EGFP control lacking the C-terminal retroCspA extension elutes at 1.68 ml, corresponding to the expected elution volume of a monomeric protein of 26.8 kDa (the size of the EGFP domain). From this data, it may be inferred that if the EGFP domain in the fusion protein remains folded into a native-like structural format, the multimeric soluble aggregates must form because of intermolecular associations of the aggregation-prone retroCspA sequences in the fusion construct. In the following sections, we establish that the EGFP domains in these soluble HMW samples are indeed fully folded and native-like, although with some differences in stability.

3.7. Soluble EGFP-T-retroCspA secondary structure: Folded EGFP and retroCspA

Fig. 4C shows the far-UV CD spectrum of the purified soluble high molecular weight fusion protein aggregates at 25 °C, together with the spectrum of the control EGFP (lacking the C-terminal retroCspA extension). It can be seen that despite its high molecular weight status, the soluble EGFP-T-retroCspA fusion is highly structured, adopting a typical β-sheet rich structure characterized by a ~218 nm mean residue ellipticity (MRE) minimum. Comparison with the far-UV CD spectrum of control EGFP protein indicates that there is considerable similarity between the spectra of the two proteins in the region of wavelengths ranging from 210 to 250 nm, as would be anticipated if EGFP exists in a native-like fold within the high molecular weight soluble aggregates. Notably, the presence of the retroCspA sequence in the fusion could be expected to shift the zero crossover point of the fusion protein's spectrum towards lower wavelengths on account of its greater contribution to the negative mean residue ellipticity below 210 (because retroCspA is less folded [10]), and indeed this is observed. The

control EGFP spectrum has a zero crossover point of 205 nm, while the fusion protein has a crossover below 200 nm. It appears, therefore, that EGFP in the fusion is native in this soluble aggregated form.

3.8. Soluble EGFP-T-retroCspA structural stability; Akin to native EGFP

Thermal melting studies of the soluble aggregated form of the fusion protein, and the native protein, were carried out by measuring changes in the MRE at 218 nm as a function of temperature. For EGFP-T-retroCspA, the data is shown in Fig. 4D which plots the change in the MRE value at 218 nm as a function of temperature, during heating from 25 °C to 95 °C at a rate of change of temperature of 3 °C per minute. At lower temperatures, there is some consolidation of structure ostensibly on account of the retroCspA component, since retroCspA is known to show some structural consolidation with heating due to the heat-induced transformation of polyproline type II (PPII) structures into beta sheets. What is important to note, however, is the fact that the EGFP-T-retroCspA fusion is able to withstand temperatures up to approx 85 °C without any loss of secondary structure. At higher temperatures, it starts losing structure through a melting transition that we find to be irreversible. As is well known, native GFP and its enhanced form, EGFP, are very stable to heat, requiring a combination of 90 °C and 6 M

GdnCl to undergo structural unfolding. The various panels in Fig. 5 show the far-UV CD spectra of the EGFP-T-retroCspA fusion (Fig. 5A), the control wild-type GFP (Fig. 5B), and retroCspA (Fig. 5C), at room temperature, upon heating to 90 °C and upon cooling back to room temperature, to illustrate how these different proteins respond to thermal destabilization/unfolding and cooling. From the data in Figs. 4 and 5, it is evident that but for the fact that the fusion protein loses structure substantially upon heating to 90 °C (Fig. 5A), while the control (Fig. 5B) gains structure, the fusion protein is clearly extremely thermostable, surviving structurally almost until 80–85 °C (Fig. 4D), comparing favorably with the known thermostability of GFP/EGFP that is evident also in Fig. 5B. This thermostability of the fusion protein between 25 °C and 85 °C is evidence of the native-like structural status of the EGFP domain within the fusion, since it is clear from the retroCspA spectra in Fig. 5C that neither the far-UV CD spectra of fusion protein (Fig. 4C), nor its thermostability (Fig. 4D) can owe mainly to retroCspA. The question, however, arises about why the folded EGFP domain within the soluble EGFP-T-retroCspA fusion fails to show the stability displayed by GFP at 90 °C, or the structural consolidation displayed by GFP upon heating. The answer would seem to be tied in with the lack of a completely native-like structure in this soluble aggregated state, which is evident also in the fluorescence data presented in the very next section. Perhaps, the inability of the
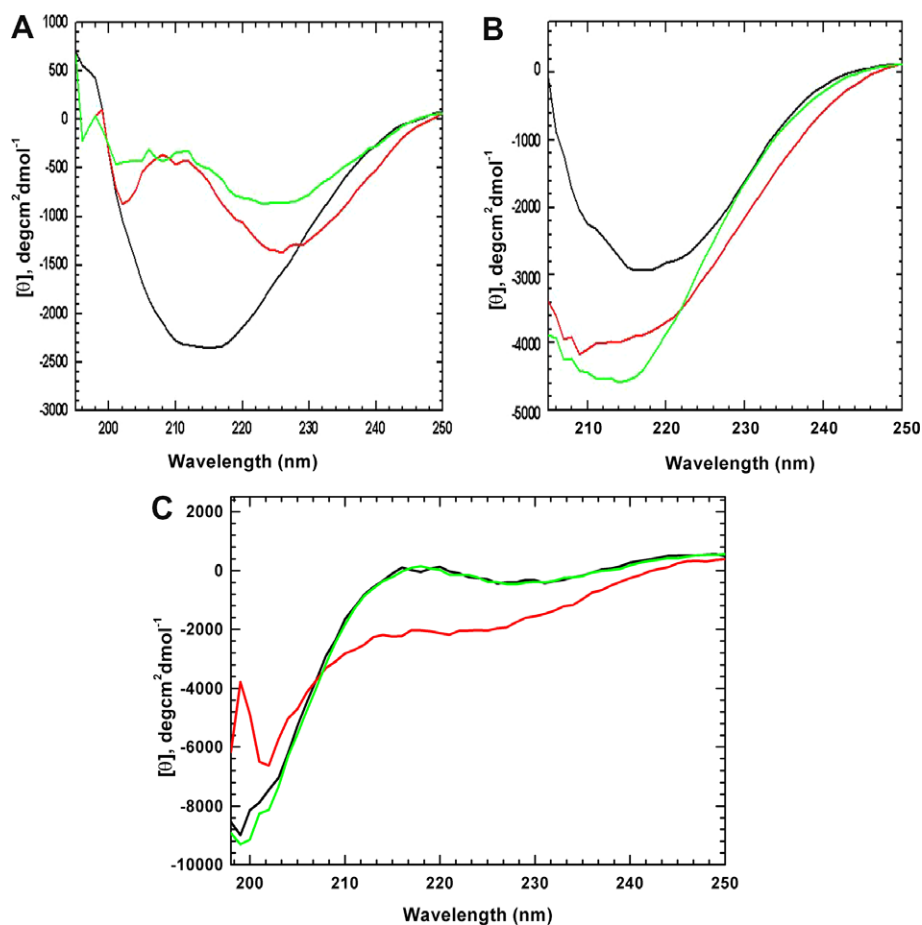

Fig. 5. Far-UV CD spectra of different proteins acquired at 25 °C (black), after heating to 90 °C (red) and upon cooling to 25 °C (green). Panel A: EGFP-T-retroCspA. Panel B: green fluorescent protein. Panel C: retroCspA.

EGFP domain within the fusion to shift to a more structured form like the control GFP (owing to steric reasons) causes it to undergo somewhat more facile structural melting, under the influence of the retroCspA sequence.

3.9. Soluble EGFP-T-retroCspA tertiary structure: quenched green fluorescence

Unlike the precipitated fusion protein which showed the typical GFP fluorescence, the soluble high molecular weight aggregates of EGFP-T-retroCspA showed no detectable green fluorescence whatsoever in course of spectrofluorimetry (data not shown), despite the obvious presence of a folded EGFP domain. This suggests that there is deformation of the native structure of the EGFP domain in the soluble aggregated state, which also explains the somewhat poorer thermostability reported in Section 3.8. Further, this soluble aggregated form also did not show any ThT binding or fluorescence, suggesting that retroCspA-based aggregation could not proceed to a stage where the amyloid-like microstructure required for ThT binding forms (data not shown).

3.10. Implications for the beta-zipper propagation of amyloid structures

It is known that disordered polypeptide segments are necessary for fibril formation, e.g., studies with apomyoglobin have suggested that fibril formation depends on the presence of disordered polypeptide segments as well as the prevalence of conditions selectively unfavorable to folding [5]. When such disordered segments associate during fibril formation, what happens to the ordered segments? Do stable structural domains resist conformational change? What brings about the conformational change? One possibility is that inter-strand formation of beta sheets begins with the mutual approach of unstructured regions or strands in extended conformation, in neighboring molecules, through the formation of a few initial hydrogen bonds. With every hydrogen bond formed that manages to form between the carbonyl group on one strand and the amide group on a proximal strand, another set of amide and carbonyl groups located further along either strand become brought into close proximity (and appropriate geometry) in the correct register for the formation of another hydrogen bond. This can lead to the propagation of a sheet in either direction through a zipper-like mechanism, with successive residues becoming co-opted into these intermolecular sheets from either an unstructured state, or even from a structured state.

Assuming that intermolecular sheets may be formed and extended by such zipper-like mechanisms, what could be anticipated to happen when an advancing sheet approaches the boundary of a stable structured domain? In principle, the domain could either halt further progression of the sheet because of its intrinsically high stability, and its inability to undergo profound conformational change or, alternatively, surrender to the advancing sheet and change conformation to become a part of the sheet. In theory, a very stable protein domain could halt, at its boundaries, the propagation of a beta sheet advancing towards it through a zipper-like mechanism, whereas a less stable domain would be more likely to allow the propagating sheet to advance past its boundaries, destabilize it and thus destroy its folded structure.

The CD data with the soluble aggregates and the fluorescence data with the precipitates indicate that the extremely stable

EGFP domain successfully resists structural reorganization by the advances of the retroCspA sequence present in fusion, although this does not prevent the aggregation and precipitation, or amyloid formation caused by retroCspA per se. Whether the 'native-like' or even 'native' state of the EGFP domain can be claimed for all molecules in the aggregated population, or only for a fraction which contributes to the spectroscopic data, of course, cannot be answered. Still, we emphasize that our data shows that native or native-like EGFP can participate in soluble aggregates and amyloids without necessarily undergoing structural destabilization and unfolding.

4. Conclusions and perspectives

We examined the structural-biochemical behavior of a fusion construct physically joining enhanced green fluorescent protein (EGFP), a highly thermally (and thermodynamically) stable protein with a beta barrel structure, to retroCspA, a non-naturally-occurring amyloidogenic protein derived by reversing the sequence of cold shock protein A, a model beta sheet protein. Our intention was to check whether, or not, the EGFP domain is at all amenable to aggregation on account of the amyloidogenic sequence present in fusion at its C-terminus. Our data seems to demonstrate conclusively that retroCspA causes the fusion protein to aggregate both intracellularly, upon expression, and following purification. Step-wise removal of denaturant and maintenance of NaCl helps to keep aggregates soluble, while single-step dialysis into NaCl-lacking buffer causes formation of precipitated, insoluble amyloid-like forms displaying ThT binding and fluorescence, as well as EGFP fluorescence. The insoluble form is associated with green fluorescence, both when produced in the cell and subsequently, upon dialysis. However, when the fusion construct adopts a soluble high molecular weight aggregated form through step-wise folding into 1 M NaCl-containing buffer, these soluble aggregates show no green fluorescence.

These data are made especially interesting by the earlier observation that fusions of the Alzheimer's peptide Aβ 1–42 with GFP produce no fluorescence when such fusion constructs undergo aggregation (because the Aβ 1–42 component of the fusion prevents GFP folding and fluorescence), whereas certain mutations which stop aggregation of the Aβ 1–42 segment allow GFP to fold and fluoresce within such fusion constructs [11]. We suggest that the reason that the Aβ 1–42 segment prevents fluorescence of GFP could have to do with the presence of the amyloidogenic peptide at the N-terminus of GFP. In data that has not been presented here, we also found that when retroCspA is fused to the N-terminus of EGFP, rather than to its C-terminus (with or without a linker sequence separating it from EGFP), the fusion proteins thus obtained, although easily extracted from inclusion bodies by urea, were non-fluorescent under all conditions – both within the cell after overexpression, and after refolding, unlike the C-terminal fusion reported here – as well as extremely precipitation-prone. We could not obtain a soluble form of such a fusion, except by maintenance in the denaturant-containing solutions used to extract the protein from inclusion bodies.

We suggest that the explanation for this could be that when GFP (or EGFP) is at the N-terminus of the fusion, it rapidly folds co-translationally to a structure that is not perturbed,

or prevented from folding to a mature protein by the amyloidogenic sequence added later during translation. On the other hand, when the amyloidogenic sequence is at the N-terminus, the intermolecular interactions that it engages in before the synthesis/folding of the GFP/EGFP component can be completed, causes the lack of folding and fluorescence of the GFP/EGFP domain. The data presented in this paper suggests that if EGFP is given the opportunity to fold first by being present at the N-terminus of the fusion, it remains folded even within the soluble aggregate, or precipitate, indicating that folded stable protein domains can participate in amyloid formation while resisting structural reorganization.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2008.05.008.

References

[1] Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B. and Blake, C.C. (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. J. Mol. Biol. 273, 729–739.

[2] Dobson, C.M. (2002) Getting out of shape. Nature 418, 729–730.

[3] Fandrich, M. and Dobson, C.M. (2002) The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. EMBO J. 21, 5682–5690.

[4] Dobson, C.M. (2004) Experimental investigation of protein folding and misfolding. Methods 34, 4–14.

[5] Fandrich, M., Forge, V., Buder, K., Kittler, M., Dobson, C.M. and Diekmann, S. (2003) Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments. Proc. Natl. Acad. Sci. USA 100, 15463–15468.

[6] Cormack, B.P., Valdivia, R.H. and Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). Gene 173, 33–38.

[7] Ward, W.W. and Bokman, S.H. (1982) Reversible denaturation of Aequorea green-fluorescent protein: physical separation and characterization of the renatured protein. Biochemistry 21, 4535–4540.

[8] Yang, F., Moss, L.G. and Phillips Jr., G.N. (1996) The molecular structure of green fluorescent protein. Nat. Biotechnol. 14, 1246–1251.

[9] Cody, C.W., Prasher, D.C., Westler, W.M., Prendergast, F.G. and Ward, W.W. (1993) Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein. Biochemistry 32, 1212–1218.

[10] Shukla, A., Raje, M. and Guptasarma, P. (2003) A backbone-reversed all-beta polypeptide (retroCspA) folds and assembles into amyloid nanofibres. Protein Eng. 16, 875–879.

[11] Kim, W. and Hecht, M.H. (2005) Sequence determinants of enhanced amyloidogenicity of Alzheimer Aβ42 peptide relative to Aβ40. J. Biol. Chem. 280, 35069–35076.