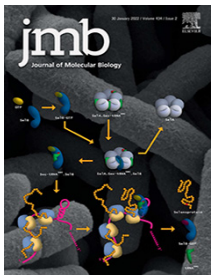

Folding Steps in the Fibrillation of Functional Amyloid: Denaturant Sensitivity Reveals Common Features in Nucleation and Elongation

Thorbjørn V. Sønderby [1,2], Helena Ø. Rasmussen [1,3], Signe A. Frank [1], Jan Skov Pedersen [1,3] and Daniel E. Otzen [1,4*]

*1 - **Interdisciplinary Nanoscience Center (iNANO),** Aarhus University, Gustav Wieds Vej 14, 8000 Aarhus C, Denmark*
*2 - **Sino-Danish Center (SDC),** Eastern Yanqihu campus, University of Chinese Academy of Sciences, 380 Huaibeizhuang, Huairou district, Beijing*
*3 - **Department of Chemistry,** Aarhus University, Langelandsgade 140, 8000 Aarhus C, Denmark*
*4 - **Department of Molecular Biology and Genetics,** Aarhus University, Gustav Wieds Vej 10C, 8000 Aarhus C, Denmark*

Correspondence to Daniel E. Otzen: Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Gustav Wieds Vej 14, 8000 Aarhus C, Denmark. *dao@inano.au.dk (D.E. Otzen), @janskovp (J. Skov Pedersen), @daniel_otzen (D.E. Otzen)*
https://doi.org/10.1016/j.jmb.2021.167337
Edited by Ronald Wetzel

Abstract

Functional bacterial amyloids (FuBA) are intrinsically disordered proteins (IDPs) which rapidly and efficiently aggregate, forming extremely stable fibrils. The conversion from IDP to amyloid is evolutionarily optimized and likely couples folding to association. Many FuBA contain several imperfect repeat sequences which contribute to the stability of mature FuBA fibrils. Aggregation can be considered an intermolecular extension of the process of intramolecular protein folding which has traditionally been studied using chemical denaturants. Here we employ denaturants to investigate folding steps during fibrillation of CsgA and FapC. We quantify protein compactification (*i.e.* the extent of burial of otherwise exposed surface area upon association of proteins) during different stages of fibrillation based on the dependence of fibrillation rate constants on the denaturant concentration (*m*-values) determined from fibrillation curves. For both proteins, urea mainly affects nucleation and elongation (not fragmentation), consistent with the fact that these steps involve both intra- and intermolecular association. The two steps have similar *m*-values, indicating that activation steps in nucleation and elongation involve the same level of folding. Surprisingly, deletion of two or three repeats from FapC leads to larger *m*-values (*i.e.* higher compactification) during the activation step of fibril growth. This observation is extended by SAXS analysis of the fibrils which indicates that weakening of the amyloidogenic core caused by repeat deletions causes a larger portion of normally unstructured regions of the protein to be included into the amyloid backbone. We conclude that the sensitivity of fibrillation to denaturants can provide useful insight into molecular mechanisms of aggregation.

Introduction

Amyloids are stable protein aggregates with a highly ordered, cross-β-sheet structure.[1] Amyloid fibrils can form from proteins with a wide variety of native conformations, including intrinsically disordered proteins (IDPs). Diseases associated with amyloidogenic IDPs include Parkinson's (α-synuclein)[2] and Alzheimer's (amyloid-β and tau).[3] However, a growing number of examples

illustrates that amyloid can be beneficial for the host organism. Functional amyloid is found in many organisms including humans[4] and fungi,[5] but are especially abundant in bacteria.[6–8] Bacterial functional amyloids (FuBA) form structural components of biofilm where they provide strength, stiffness and adhesive properties.[8,9] Examples include curli fibrils in *Escherichia coli* (*E. coli*)[10] and Functional amyloids in *Pseudomonas* (Fap).[6,7]

FuBA biogenesis is tightly regulated and requires numerous dedicated accessory proteins in addition to the general Sec machinery which transports them across the inner membrane to the periplasmic space.[11] The seven proteins involved in curli biogenesis are encoded in two linked operons, *csgBAC* and *csgDEFG*.[12] CsgA is the major structural component of curli fibrils whereas CsgB is a nucleator protein essential for CsgA aggregation *in vivo* but not *in vitro*.[13,14] CsgC, CsgE and CsgF are chaperones[15–17] while CsgG is a channel-forming outer membrane protein.[18] Fap is evolutionarily unrelated to curli but the two systems share many similarities.[19] Besides the major amyloid component FapC,[7] the Fap system includes a nucleator (FapB)[7] and the outer membrane pore protein (FapF).[20] The roles of FapA and FapD are not clear but they are likely chaperones.[20,21] Fap fibrils also contain a small amount of FapE protein whose function remains unclear.[21]

The complex and well-regulated machineries of curli and Fap indicate evolutionary optimization. This is also revealed at the sequence level. CsgA and FapC (in *Pseudomonas* sp. UK4) contain 5 and 3 imperfect repeat sequences, respectively, each of which is believed to form a β-hairpin which stack on top of one another.[11,22–24] In CsgA, the repeat sequences are each about 22 residues long and linked by β-turns of 4–5 residues in length.[11,22,25] The 3 repeat sequences of FapC are each about 37 residues long and are connected by linkers of variable length (31 and 39 residues in UK4). Mature CsgA and FapC (in *Pseudomonas* sp. UK4) contain 131 (13.1 kDa) and 226 (22.6 kDa) residues, respectively. The repeat sequences of FuBA impact on both the fibrillation mechanism and the stability of the fibrils. Removal of repeats makes FapC vulnerable to fragmentation[26] and reduces the stability of mature fibrils.[27]

Monomeric CsgA and FapC are both IDPs *in vitro* and only gain structure upon association.[7,28–30] *In vivo*, both proteins remain unfolded until they reach the outer membrane where they interact with their respective nucleator proteins, CsgB and FapB. However, *in vitro* CsgA and FapC fibrillate readily without the presence of their nucleator proteins after a short lag phase of a few hours,[7,28] forming SDS-resistant fibrils which are only dissociated in formic acid. For this reason, they are purified and stored in high concentrations of denaturants such as urea or guanidinium chloride (GdmCl) and are typically only desalted immediately prior to starting aggregation studies.[28,29,31]

Denaturants have been extensively used to study the stability and folding kinetics of natively folded proteins,[32] thanks to their ability to stabilize the unfolded state of proteins through preferential binding.[33,34] Empirically, the logarithm of protein folding and unfolding rate constants in denaturant (k^{den}(as well as equilibrium constants of unfolding) depend linearly on denaturation concentration [den][32] via constants denoted *m*:

$$logk^{den} = logk^{water} + m * [den] \quad (1)$$

Each molecular step has a specific *m*-value associated with it (*e.g.* m_f for folding kinetics or m_u for unfolding kinetics). *m*-values correlate with the change in solvent accessible surface area (SASA) during the associated step in protein folding/unfolding[35,36] and therefore inform about the change in compaction in that step. While equilibrium *m*-values report on changes between the two involved equilibrium states (*e.g.* unfolded and folded states), kinetic *m*-values focus on the activation step, i.e. the SASA change between the ground state and the transition state TS that is rate-determining for this step (*e.g.* the TS between the unfolded and folded state in a simple two-state system). If available, *m*-values from forward and reverse activation steps (*e.g.* folding and unfolding) may be combined to yield the corresponding equilibrium *m*-value.

Here we use urea to elucidate the changes in compaction associated with different steps involved in FuBA fibrillation *in vitro*. We extend our analysis to variants lacking some or all of the three imperfect repeats of FapC to understand how FuBA repeat sequences contribute to compaction. We monitor the percentage of aggregated FuBA protein over time in increasing concentrations of urea, using a combination of Thioflavin T (ThT) fluorescence and Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE). The kinetic parameters of the fibrillation curves are analyzed using a linear regression model as well as the AmyloFit fitting webserver,[37] allowing us to obtain kinetic *m*-values for individual steps in the aggregation process, namely primary nucleation (the activation step of formation of the nucleus from which fibrils can grow) and elongation (the activation step of extension of the fibrils from the growing ends). Note that we can only deduce the changes associated with the forward direction of fibrillation (moving to the final fibrillar state) and not the reverse direction (dissolution of fibrils and formation of monomers) due to the resistance of the fibrils to dissolution, and we must therefore limit firm conclusions to these forward parts of the process. Furthermore, since the kinetic rate constants report on intermolecular processes involving at least two

molecules associating to form nuclei or extend fibril ends, the *m*-values most likely reflect the average change in SASA per protein molecule involved in these encounters (caused *e.g.* by burying interfaces between two monomers as well as folding of the incoming protein molecule in the elongation step). This means that our kinetic *m*-values can only be taken as general indicators of changes in compaction. Moreover, we use Small Angle X-ray Scattering (SAXS) as a complementary technique to determine the compactness of WT FapC and the deletion mutant of FapC lacking all three repeats. We find that both CsgA and FapC WT fibrillate up to 8 M urea and that rate constants associated with both nucleation and elongation decrease log-linearly with increasing concentration of urea, similar to folding of globular proteins. Moreover, urea primarily affects primary processes (elongation and nucleation) as opposed to the secondary process of fragmentation. Remarkably, deletion of two or three repeats increases the *m*-values associated with activation of elongation. All other things being equal, this implies that deletion mutants form more compact fibrils. This was confirmed by SAXS which showed that deletion of all repeats from FapC results in more compact (but destabilized) fibrils, indicating a significant rearrangement of fibrillar structure.

Results

Urea and GdmCl increase lag-times and lower the ThT signal of CsgA fibrillation

To study the impact of denaturant on CsgA fibrillation, we followed ThT fluorescence over time for CsgA incubated with increasing concentrations of urea or GdmCl (Figure 1(A) and (B)). Urea increased the lag-time of CsgA and decreased the ThT fluorescence signal, leading to a 5-fold reduction in end-point ThT fluorescence at 6 M urea (Figure 1(A)). GdmCl had a much stronger effect, reducing fluorescence 8-fold already at 0.5 M GdmCl and completely suppressing ThT signal increase around 2–3 M GdmCl (Figure 1(B)). We therefore concentrated our efforts on urea whose weaker effects on fibrillation allowed us to explore its influence over a broad concentration range, providing more robust *m*-values as we will show below.

High concentrations of urea only modestly reduce the level of CsgA aggregation

ThT fluorescence is sensitive not only to the total amount of fibrils but also to changes in solvent and solutes, *e.g.* small molecules competing for binding to fibrils. To ascertain whether the decrease in ThT fluorescence reflected a genuine decrease in insoluble material, we used SDS-PAGE to estimate the amount of monomeric CsgA left in solution after the stationary phase of fibrillation was reached (Figure 1(C)). We have found the extent of the fibril formation to be most accurately determined by quantifying the concentration of remaining monomeric protein. The wide concentration range of urea complicated the use of BCA or Bradford assays while the low levels of monomer remaining in solution required higher sensitivity than could be provided by OD_{280} measurements. We have previously shown that the fibrillation process of CsgA correlates well with decrease of monomer concentration and is easily followed using SDS-PAGE.[31] We therefore decided to use a SDS-PAGE assay for quantifying the remaining monomer concentration. No detectable monomeric CsgA was left at 0–3 M urea, indicating that all CsgA protein in these samples had been incorporated into insoluble aggregates. Between 3.5–6 M urea, the amount of monomeric CsgA increased to a modest ~8%, much less than the 80% decrease in ThT fluorescence. We therefore examined if the decrease in ThT signal reflected changes in aggregate structure. Transmission electron microscopy (TEM) analysis confirmed that fibrils were the dominant type of aggregate at all [urea]s (Figure 1(D)). Fourier transform infrared spectroscopy (FTIR) analysis (data not shown) confirmed that the aggregates formed in 0 to 8 M urea all contained a dominant peak around 1620 cm^{-1} characteristic of fibrils.[38] This indicated that the loss in ThT signal was caused by other phenomena. To address this, we incubated preformed ThT-bound fibrils to increasing amounts of urea and followed the ThT fluorescence. Urea significantly reduced ThT fluorescence when preformed fibrils formed in the absence of urea were resuspended in increasing concentration of urea (Figure 2(A)). However, no soluble monomers were detectable after 28 days of incubation of preformed fibrils in urea (SDS-PAGE data not shown). This suggested that the reduction of ThT signal could be caused by either (1) displacement of ThT from fibrils by urea, (2) difference in the internal structure of fibrils formed at different urea concentrations, leading to a different binding of ThT or (3) quenching of ThT fluorescence by urea. We could rule out the first possibility as follows: when 100 μM CsgA fibrils were incubated with 60 μM ThT in 0 and 8 M urea, we observed the same amount of ThT in the supernatant, namely 54.2 ± 1.1 and 53.3 ± 1.6 μM (absorbance was converted to concentration using ThT calibration curves recorded in 0 and 8 M urea, respectively). To investigate the second possibility would require structural analysis methods beyond the scope of this investigation such as cryoEM, HDX-MS or solid state NMR (more accessible methods such as limited proteolysis are challenged by the high stability of the fibrils). However, we indirectly investigated this possibility by resuspending mature fibrils in 0–8 M urea and immediately recorded ThT fluorescence spectra (Figure 2(A)). Here, we observed a drastic

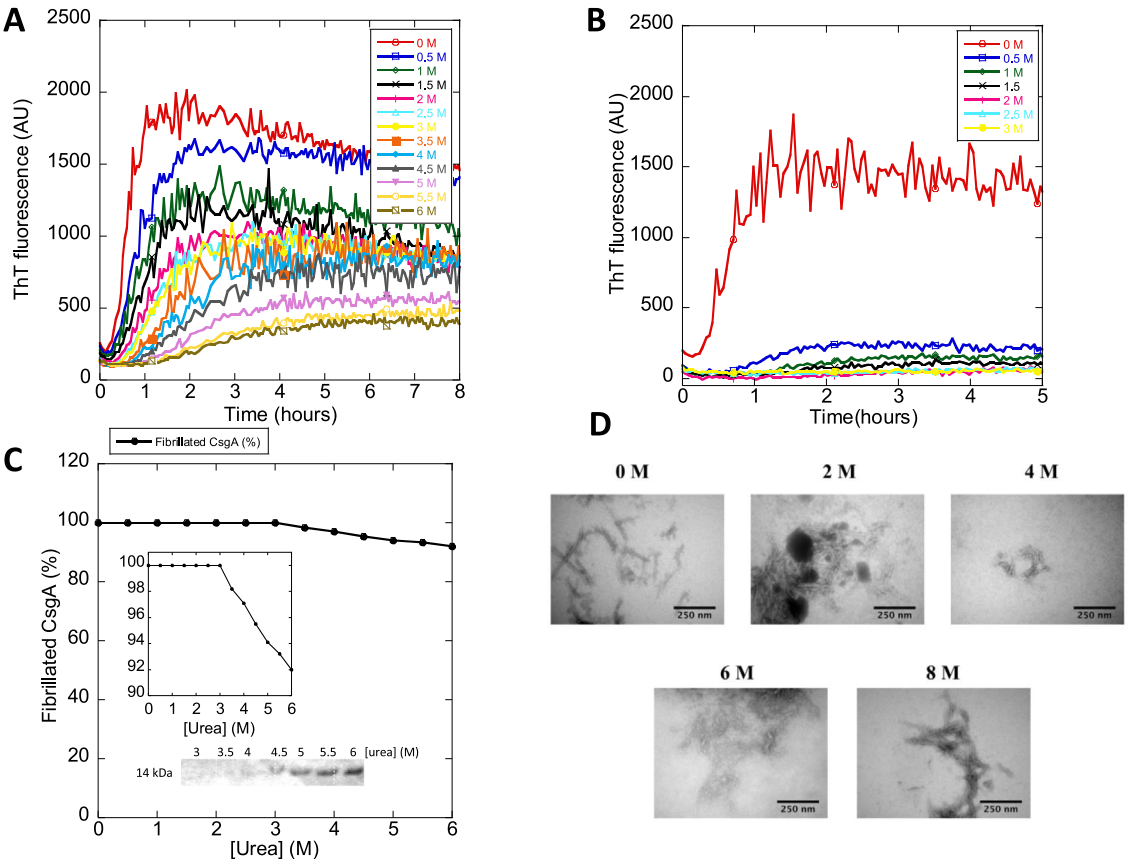

Figure 1. Thioflavin T assay of CsgA WT (40 μM) fibrillation in presence of 0–6 M urea (A) or 0–3 M GdmCl (B). Increasing concentrations of urea or GdmCl increased the lag-time of CsgA WT and lowered the fluorescence signal intensity of the stationary phase. (C) SDS-PAGE analysis of the supernatant from CsgA WT samples fibrillated in the presence of 0–6 M urea. At 0–3 M urea, no soluble CsgA was observable. At 3.5–6 M urea increasing amount of soluble CsgA was observed at ~14 kDa. The CsgA protein bands were quantified using ImageJ and the percentage of aggregated CsgA was plotted as a function of [urea]. Insert: Zoom-in at 90–100% fibrillated CsgA. (D) TEM images of amyloids formed in 0 M, 2 M, 4 M, 6 M and 8 M urea after overnight incubation with shaking: Mature fibrils were formed at 0 M, 2 M and 8 M with length ranging from <30 nm to several hundred nm. The fibrils formed at 4 M and 6 M were relatively short. Relatively few fibrils were found in all samples, as they were predominated by large amounts of clumps (not shown). The black clumps seen in the 2 M sample are probably precipitated urea crystals.

decrease in the ThT fluorescence at increasing urea concentration. This result makes structural rearrangements of the fibrils unlikely on such a short time scale, taking the high stability of FuBA into account.[7,31] This led us to the third possibility, which we addressed by recording ThT absorption spectra in 0–8 M urea in the absence of fibrils (the low ThT fluorescence in the absence of fibrils makes it unfeasible to record fluorescence spectra). We observe a gradual redshift and an increase in absorbance (Figure 2(B)). We therefore conclude that urea modulates ThT spectral properties but does not displace ThT from fibrils.

The rate of CsgA fibrillation decreases systematically with urea

Although urea only has a modest effect on the final yield of CsgA fibrils, the rate of fibrillation is clearly slowed down over a wide concentration range. We therefore decided to elucidate the mechanism of fibrillation in urea. The raw fibrillation curves show the overall time profile of the fibrillation containing a lag phase, growth phase and plateau phase. However, to compare the kinetics of the fibrillation at different urea concentrations and avoid artifacts from changes in ThT fluorescence at different urea concentrations, we have to convert our ThT fluorescence profiles to fraction aggregated protein. This was done by accurately determining the percentage of fibril formation at the fibrillation plateau phase at each urea concentration using SDS-PAGE analysis (Figure 1(C))). This allowed us to correct for the presence of non-aggregated protein, so the final graphs showed percentage of fibrillated protein as a function of time (Supplementary Figure S1). The normalized and corrected curves now show the

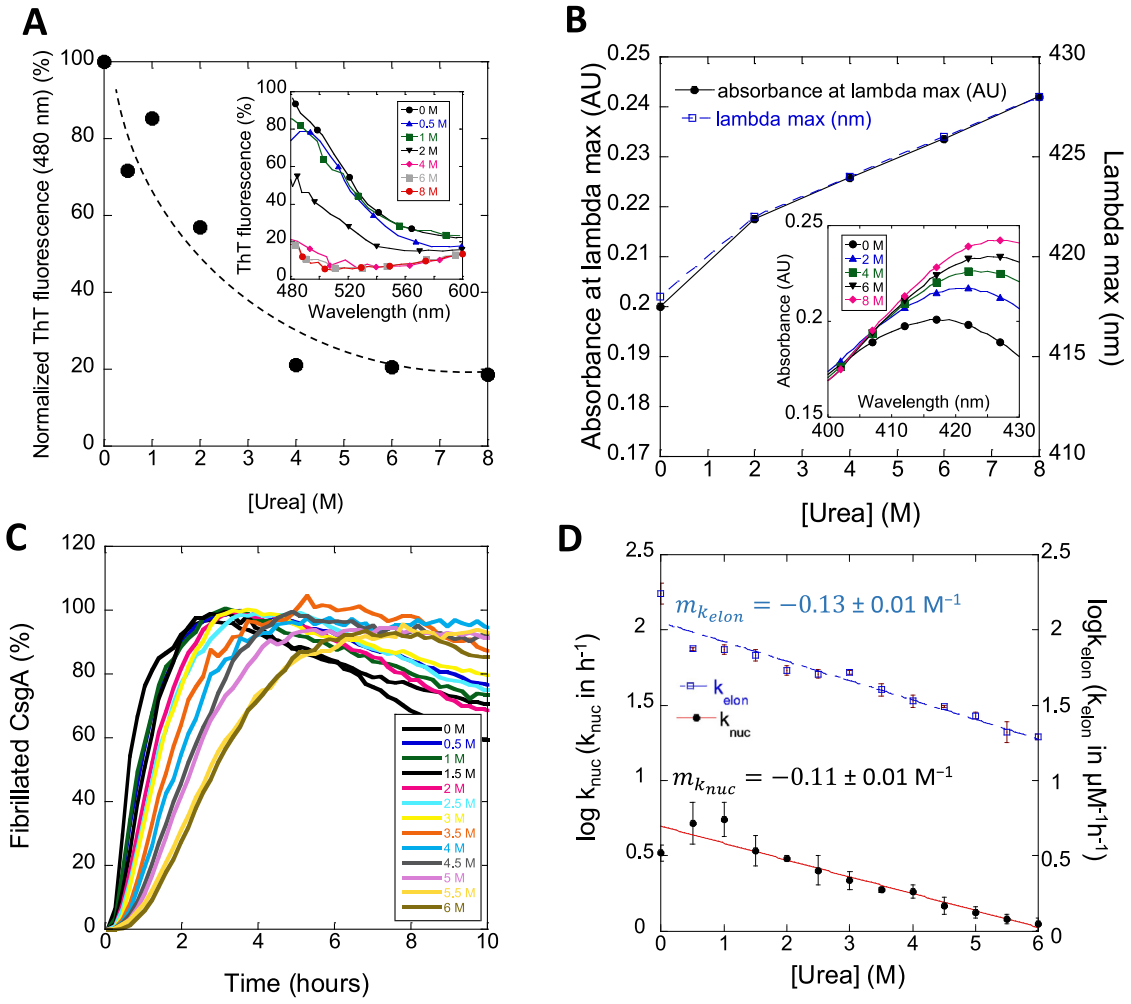

Figure 2. (A) Normalized ThT fluorescence at 480 nm as a function of [urea]. A drastic decrease in the ThT signal is caused by increasing the concentration of urea. Inset: ThT spectra of mature CsgA fibrils added to 0–8 M urea. (B) Absorbance (arbitrary units) and λmax(nm) of ThT in 0–8 M urea. Inset: Absorbance spectrum of ThT in 0, 2, 4, 6 and 8 M urea. (C) Normalized fibrillations curves of CsgA in 0–6 M urea showing percentage of aggregated CsgA protein as a function of time. (D) Logarithm of nucleation rate constant, k_{nuc} and growth rate constant, k_{elon} plotted as a function of [urea]. A linear function was fitted to the plots to obtain the *m*-values, $m_{k_{nuc}}$ and $m_{k_{elon}}$. Error bars are standard deviation (n = 3).

percentage of aggregated CsgA as a function of time (Figure 2(C)) revealing retardation of both the lag phase and the elongation phase. As an initial analysis, curves were analysed with a linear regression method, which provides an apparent nucleation rate constant, k_{nuc} and a growth rate constant, k_{elon}. The log of both k_{nuc} and k_{elon} decreased in a reasonably linear fashion with [urea], providing both values of log k^{water} and their sensitivity to [urea], i.e. kinetic *m*-values. The values of $m_{k_{nuc}}$ and $m_{k_{elon}}$ (0.11 ± 0.01 and 0.13 ± 0.01 M^{-1}) (Figure 2(D) and Table 1) were identical within error, indicating that the activation steps of nucleation and growth of CsgA fibrils were affected to a similar extent by changes in [urea]. Given the association of *m*-values with changes in SASA, our results suggest that a similar amount of surface area is buried during the activation steps of nucleation and growth. A more detailed understanding requires us to obtain *m*-values for

Table 1 *m*-values determined using linear regression analysis of fibrillation curves using data shown in Figures 3, 4 and S1. Values for FapC ΔR2R3 and FapC ΔR1R2R3 omitted due to poor fibrillation kinetics.

Protein	$m_{k_{elon}}$ (M^{-1})	$m_{k_{nuc}}$ (M^{-1})
CsgA	0.13 ± 0.01	0.11 ± 0.01
FapC WT	0.12 ± 0.02	0.15 ± 0.02
ΔR1	0.08 ± 0.02	0.09 ± 0.03
ΔR2	0.16 ± 0.01	0.16 ± 0.01
ΔR3	0.00 ± 0.02	0.06 ± 0.02
ΔR1R2	0.13 ± 0.01	0.13 ± 0.00
ΔR1R3	0.14 ± 0.02	0.10 ± 0.02

the individual molecular steps involved in aggregation (see below). Note that in these experiments we are examining processes that are intermolecular and consequently sensitive to changes in viscosity. Between 0 and 8 M urea the viscosity of urea changes ca. 1.6 fold[39] with a linear decrease in a semi-log plot (log viscosity *versus* [urea]) corresponding to a slope of 0.029 M^{-1}. It could be argued that we should subtract this value from all our kinetic *m*-values. However, a general subtraction implies that all the investigated processes are equally sensitive to diffusion which is not necessarily the case. Furthermore, we focus on comparative rather than absolute values and this is not affected by the absence or presence of this subtraction, so for simplicity we have avoided this step.

Removal of repeats can regulate the urea-susceptibility of FapC

To address if urea's effect on CsgA fibrillation applied to other FuBA, we turned to the FuBA FapC from *Pseudomonas sp. UK4.* Similar to CsgA, the normalized fibrillation curves of wildtype FapC show an increase in lag times and lowered elongation rates as [urea] increases to 8 M (Figure 3). The fibrillation yield remains around 100% up to 6 M urea but decreased slightly from 7 M. Lag times and rate constants estimated in the same way as for CsgA showed a linear decline. The *m*-values were identical to CsgA within error, possibly reflecting a similar degree of surface area burial per protein incorporated during the activation step. To elucidate the contributions of the individual repeats of FapC to this aspect of fibrillation, we turned to FapC constructs where the imperfect repeats had been removed, either one, two or three at a time and analyzed them in the same way as WT FapC. Overall, urea slowed down the fibrillation of all FapC variants (representative examples in Figure 3; full collection in Supplementary Figure S2. With the exception of the ΔR1 variant, deletion of repeats made the proteins more sensitive to urea and gradually reduced aggregation yields at higher [urea]. Thus FapC ΔR2R3 and FapC ΔR1R2R3 only fibrillated up to 1 M urea and 0.5 M urea, respectively, consistent with their dramatically increased sensitivity to formic acid and SDS.[27]

Most FapC variants showed a linear decrease of both $\log k_{nuc}$ and $\log k_{elon}$ with increasing [urea], with the exception of FapC ΔR3 which had a reduced m_{nuc} and an essentially urea-insensitive elongation rate. Just as observed for CsgA, $m_{k_{elon}}$ and $m_{k_{nuc}}$ values for individual variants were similar (Table 1; values for FapC ΔR2R3 and FapC ΔR1R2R3 omitted due to the paucity of reliable data), again indicating that the activation steps of nucleation and elongation involve essentially the same degree of surface burial. It is less clear how to interpret the differences in *m*-values between FapC variants as there is no clear trend in the development of these values as repeats are removed. *A priori* this suggests variations in the fibrillation mechanism between different mutants, leading to different levels of folding compactness in the TS. We therefore turned to a closer examination of the effect of urea on aggregation. We recently showed that FapC fibrillation in the absence of urea is best described by a model dominated by fragmentation.[26] We investigated if urea alters the overall fibrillation mechanism of FapC. We fibrillated 8–92 μM FapC WT in 0 and 4 M urea. After normalizing the data (Figure 4(A) and (B)), we obtained straight lines at both 0 and 4 M urea when we plotted the half-time (time to reach 50% of maximal ThT fluorescence) *versus* monomer concentration in a log–log plot (Figure 4(C) and (D)). This suggests that the kinetic data can be described by a single fibrillation mechanism in this concentration range. Further evidence for this single mechanism is provided when we normalize time points by the midpoint of fibrillation,[40] leading to a collapse of curves around a single time profile (Figure S3). This indicates that the different reactions scale, *i.e.* they can be described by the same single function. Using the program AmyloFit,[37] we established that the kinetic data is best fitted by a fragmentation dominated model in both presence and absence of urea (Figure 4(A) and (B)). Thus, we conclude that the overall fibrillation mechanism is not affected by urea.

Urea affects protein folding and association during fibrillation

In the absence of urea, as more and more repeats are removed from FapC, fragmentation becomes dominant at earlier stages of the reaction.[26] We therefore decided to elucidate which microscopic steps in the fibrillation of CsgA and the different FapC variants were most affected by urea. Using the fragmentation dominated model, we performed global curve fitting to the normalized fibrillation curves of each protein while allowing the rate constant of either elongation (k_+), primary nucleation (k_n) or fragmentation (k_-) to float (FapC WT shown in Figures 4(E); full collection in Figures S4 and S5). For CsgA and all six FapC variants, best fits were obtained by varying k_+, with the caveat that for the 2 most destabilized FapC variants (ΔR1R3 and ΔR2R3), variation in k_- fitted equally well. The fibrillation curves could also be fitted by changes in the primary nucleation rate constant (k_n), but the fits obtained in this way give slightly higher Mean Squared Residual Errors (MRE) (Table 2; note that MRE values for different fits to the same data are normalized relative to the fit with the lowest MRE value). Overall, the quality of the different fits was not dramatically different between the different variants, except for the poor fits with variable k_- seen for the four most stable mutants.

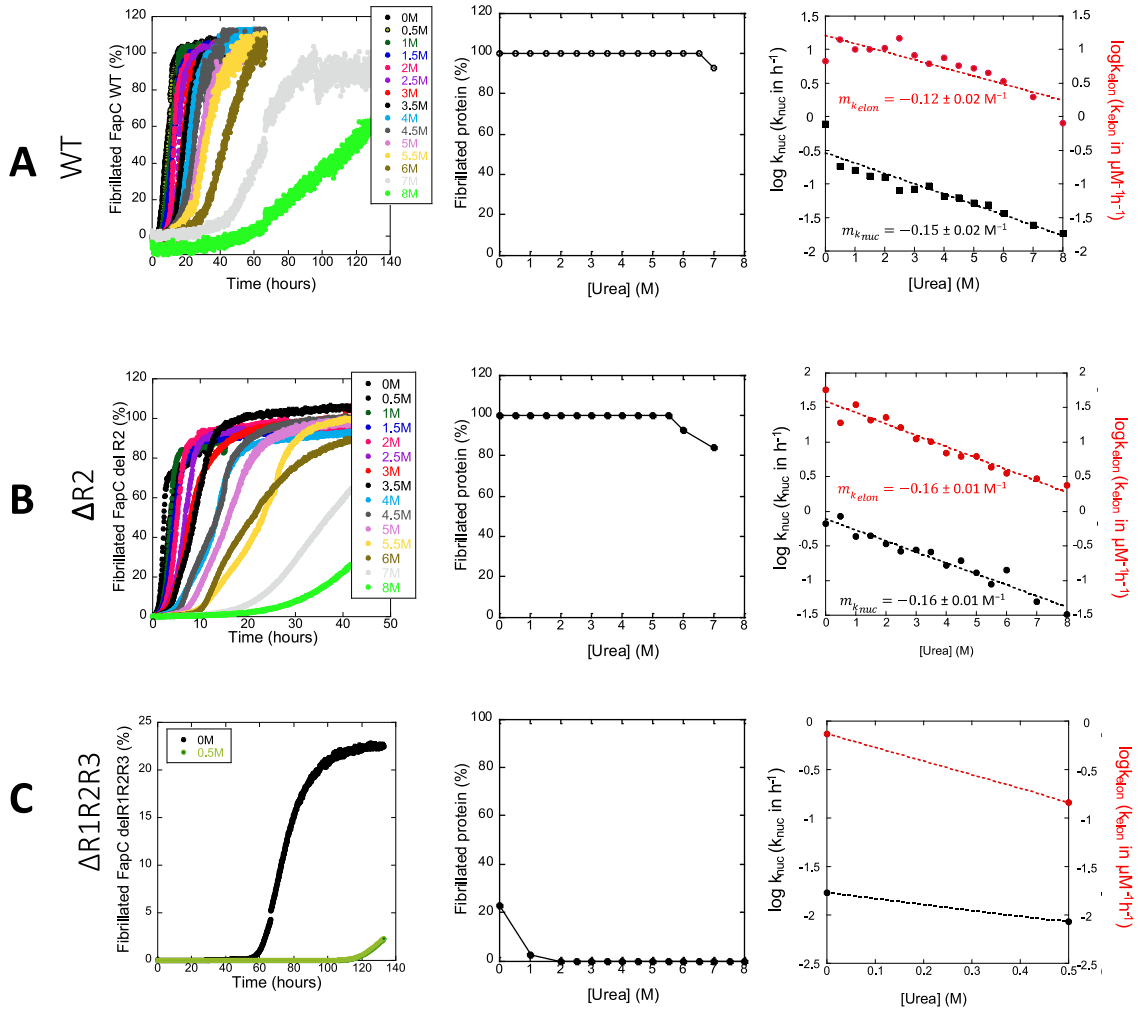

Figure 3. Left column: Normalized fibrillation curves of FapC WT (A) FapC ΔR2 (B) and FapC ΔR1R2R3 (C) (35 μM) in 0–8 M urea. The curves are based on ThT fluorescence. Middle column: The percentage of aggregated FapC determined by SDS-PAGE analysis and plotted as a function of [urea]. Right column: The kinetic parameters, k_{nuc} (determined from lag-time^{-1}) and k_{elon} (determined from the slope of elongation phase) were log-transformed and plotted as a function of urea. Linear regression was performed on the plots to obtain *m*-values (slopes of the regressions, $m_{k_{elon}}$ and $m_{k_{nuc}}$), which describe the urea-sensitivity of each kinetic parameter.

Although the data in principle could be explained by variations in a single parameter (which in most cases was either k_+ or k_n), the modest differences in fit quality and the comparable values of m_{k+} and m_{kn} suggest that urea affects more than one parameter of the fibrillation mechanism. As an additional analysis, we therefore fitted the data to the unseeded fibrillation model in AmyloFit where we allowed variation in two of three parameters, k_+, k_- and k_n. When more than one parameter is allowed to vary, AmyloFit outputs composite rate constants, k_+k_n and k_+k_-, as it is not possible to determine k_+, k_- or k_n individually from unseeded fibrillation curves. Similar to the analysis of individual parameters, variation in primary processes (primary elongation and nucleation) gave lower MREs for the stable protein variants whereas the most unstable variants showed low MREs for variation in both primary elongation and fragmentation. We plotted the log of both the individual and the composite rate constants as a function of [urea] and evaluated the plots (Figures S4–S6). As expected, the rate constants belonging to primary processes (k_+, k_n and k_+k_n) showed the best log-linear correlation with increasing concentration of urea (Table 3). We conclude that urea-induced changes in the fibrillation rate are best described by modulation of primary processes (primary nucleation and elongation) rather than secondary processes such as fragmentation. From plots of the log rate constants *versus* [urea], the corresponding *m*-values were determined (Table 2 and 3 and Figure S6. The composite *m*-values in Table 3

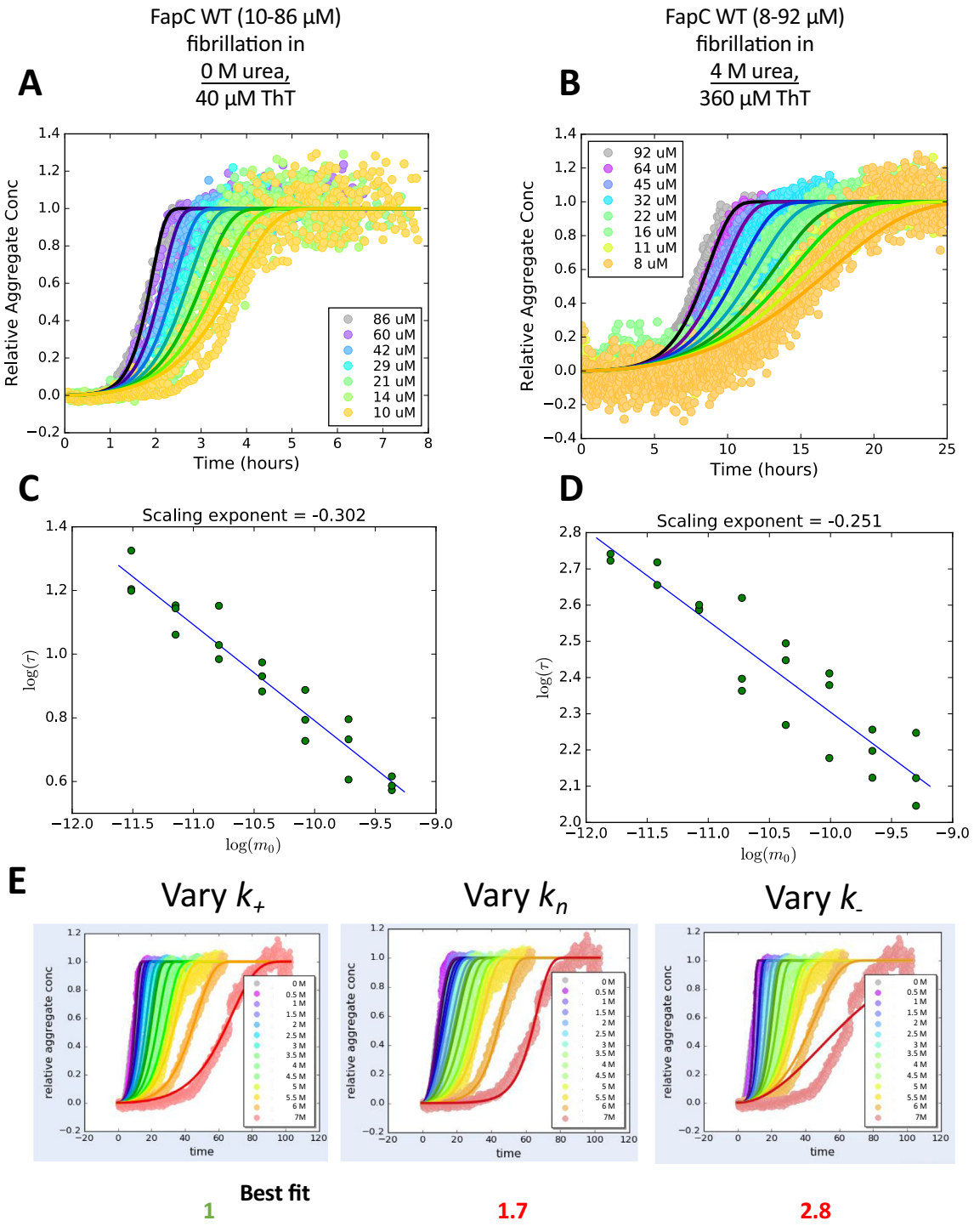

Figure 4. The fibrillation curves of FapC WT fitted well to a fragmentation dominated model in both 0 M urea (A) and 4 M urea (B) using AmyloFit.[37] The natural log transformed plots of half-time *versus* initial monomer concentration show that a single linear correlation describes the entire monomer concentration range tested in both 0 M urea (C) and 4 M urea (D). Normalized fibrillation curves of FapC WT (35 µM) in 0–8 M urea were fitted to a fragmentation dominated model using global fitting (E) where only one kinetic parameter was allowed to vary for each curve, namely k_+ (left panel), k_n (middle panel) or k_- (right panel). MRE was normalized relative to the best fit for each FapC variant which was given the value of 1.

were all larger than the individual *m*-values shown in Table 2, consistent with the fact that *m*-values of a composite rate constant should be the sum of the individual *m*-values. However, there was no simple correlation between the removal of repeats from FapC and the magnitude of the *m*-values

Table 2 *m*-values associated with elongation rate constants (m_{k+}) and nucleation rate constants (m_{kn}) determined using AmyloFit. The Pearson coefficient (R) of the correlation is given in brackets. Regressions with R < 0.65 are colored grey. The three right-most columns show MRE of separate fits where either k_+, k_n or k_- was allowed to vary one at a time while keeping the other two rate constants globally constrained. Due to the generally poor fit with k_-, *m*-values for this fit are not included.

Protein	m_{k+} (M^{-1})	m_{kn} (M^{-1})	MRE (k_+)	MRE (k_n)	MRE (k_-)
CsgA	0.21±0.01 (0.99)	0.21±0.01 (0.99)	1	1	2
FapC WT	0.22±0.01 (0.98)	0.43±0.04 (0.88)	1	1.7	2.8
ΔR1	0.15±0.02 (0.80)	0.19±0.03 (0.77)	1	1.3	2.3
ΔR2	0.35±0.01 (0.99)	0.42±0.02 (0.98)	1	1.8	3.8
ΔR1R2	0.34±0.03 (0.90)	0.34±0.09 (0.55)	1	3.2	15
ΔR1R3	0.25±0.06 (0.68)	0.30±0.08 (0.70)	1	1.4	1
ΔR2R3	1.35±0.40 (0.92)	1.90±0.67 (0.89)	1	2.8	1.1

Table 3 *m*-values associated with composite kinetic parameters of elongation and nucleation ($m_{k_+k_n}$) and elongation and fragmentation ($m_{k_+k_-}$). The Pearson coefficient (R) is given in brackets. Regressions with R < 0.65 are colored grey. The two right-most columns show MRE of fits with variation in the composite rate constants.

Protein	$m_{k_+k_n}$ (M^{-1})	$m_{k_+k_-}$ (M^{-1})	MRE (k_+k_n)	MRE (k_+k_-)
CsgA	0.23±0.01 (0.99)	0.44±0.03 (0.98)	1.0	1.8
FapC WT	0.42±0.04 (0.89)	0.52±0.09 (0.75)	1.0	2.5
ΔR1	0.19±0.03 (0.76)	0.75±0.25 (0.39)	1.0	1.6
ΔR2	0.42±0.02 (0.98)	0.96±0.24 (0.62)	1.0	2.0
ΔR1R2	0.48±0.04 (0.94)	0.90±0.25 (0.51)	1.0	1.5
ΔR1R3	0.30±0.08 (0.70)	0.58±0.14 (0.70)	1.4	1.0
ΔR2R3	1.89±0.67 (0.89)	Not linear	1.0	1.4

despite relatively large differences between protein constructs. This indicated that the deletion mutants may be affected by urea in a complex manner during each step of the fibrillation mechanism. In addition, the analysis of unseeded fibrillation curves is complicated by the fact that the fibrillation kinetics is dependent on multiple rate constants which need to be separated.

Seeding of FapC fibrillation in urea

Having determined that urea primarily inhibits primary fibrillation processes (nucleation and elongation) but presumably to almost comparable extents, we turned to seeding of fibrillation in order to bypass nucleation and focus directly on elongation. This allows a more accurate determination of the elongation rates and their associated *m*-values and also allows us to more quantitively compare the *m*-values between FapC variants and thus estimate their change in compactness or SASA during the activation step of elongation. The fibrillation of FapC (35 μM monomer) was monitored after the addition of seeds (19.5 μM monomer equivalents) (Figures 5 and S7). The fibrillation curves were corrected for

Table 4 *m*-values associated with the elongation rate estimated from the first phase of the seeded fibrillation curves

Protein	$m_{\text{seeded elongation}}$ (M^{-1})
FapC WT	0.15 ± 0.01
ΔR1	0.24 ± 0.02
ΔR2	0.14 ± 0.01
ΔR1R3	0.37 ± 0.02
ΔR2R3	0.34 ± 0.02
ΔR1R2R3	0.40 ± 0.02

the ThT-quenching effect of urea by the same procedure as described for the unseeded fibrillation curves (Figure S1). This correction included quantification of the remaining monomer concentration to determine the percentage of fibrillated protein at the time of termination of the experiment (Figures 5, Middle panel and S7 Middle Column). Note that the seeded fibrillation curves may not have reached equilibrium by the time of termination, since we were only interested in the elongation kinetics immediately after seeding, and the graphs should not be compared directly to percentage of aggregated protein of the unseeded curves (Figure 3, Middle column). We monitored the initial rate of elongation (*i.e.* the slope around time zero) of the seeded fibrillation curves. This value declined in a log linear fashion with the concentration of urea (Figures 5 and S7). Interestingly, the addition of seeds rescued the fibrillation of the deletion mutants, ΔR2R3 and ΔR1R2R3 which was otherwise inhibited already at low concentrations of urea in the unseeded fibrillation experiments. This allowed us to analyse the fibrillation kinetics of these mutants over a wider [urea] range to obtain *m*-values associated with the initial elongation rate of the FapC variants (summarized in Table 4). FapC WT and ΔR2 showed the lowest *m*-values followed by ΔR1 while deletion of two or three repeats resulted in dramatically larger *m*-values. Thus, deletion of repeats resulted in higher kinetic *m*-values. Of note, the elongation rate of ΔR3 appeared to be largely unaffected up to 4 M of urea, whereafter the elongation rate suddenly plummets as the [urea] increases (Figure S7). This indicates a change in fibrillation mechanism above 4 M.

Small-angle X-ray scattering confirms more compact fibrils in a destabilized mutant

Finally, we used small-angle X-ray scattering to investigate if the larger *m*-values obtained by deletion of repeats reflected a larger compactification during fibril elongation. For this analysis, we compared WT FapC with the mutant lacking all three repeats, ΔR1R2R3. We started out by investigating the initial states of both proteins, where an Indirect Fourier Transformation routine was applied to obtain the forward scattering *I*(0) which can be used to calculate the mass of the scattering species. The obtained mass values are expressed in number of monomers (# mon) in Table 5 by dividing with the monomer mass for WT FapC (23.6 kDa including the His-tail terminal extension) and ΔR1R2R3 (13.9 kDa). This showed that neither samples display a pure monomer state; rather, they both form complexes consisting of 6–7 monomers on average.

The initial state of FapC WT was fitted on arbitrary scale with a linear Gaussian polymer model containing the radius of gyration and a constant background (Table 5). It was not possible to obtain a good fit of the initial state of FapC ΔR1R2R3 using this model, as it could only be optimized to a χ^2 value of 12.2 which is far from the ideal value of unity. Data were instead fitted with a polymer star model on arbitrary scale, which allows for the number of random coil arms to be fitted (for a linear Gaussian polymer model this is per definition fixed at two) along with the radius of gyration for the arms and a constant background (Table 5), where the number of arms was also optimized. The star model does not fit perfectly (χ^2 = 5.51). This is probably due to polydispersity in the length of the arms and between the particles, which the model does not take into account. A dimensionless Kratky plot (Figure S8) shows that both samples are unfolded as the signal does not have a maximum at qR_g = 1.7 as for folded proteins and stays constant at larger qR_g values.[41]

As both data sets can be described as a random coil, they are mainly unfolded (see sketch in Figure 6). The need for the star model for FapC ΔR1R2R3 indicates that the initial state is slightly

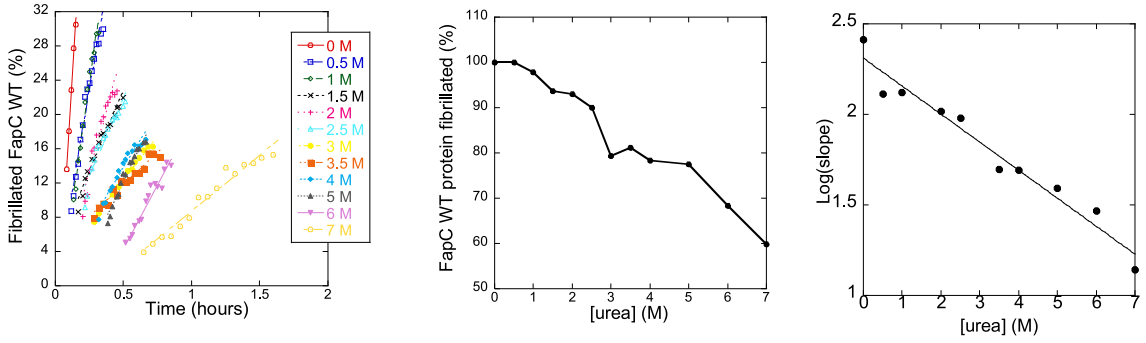

Figure 5. Left panel: Normalized fibrillation curves of FapC WT (35 μM) in 0–8 M urea after addition of seeds. The curves are based on ThT fluorescence. Middle column: The percentage of aggregated FapC WT was determined by SDS-PAGE analysis and plotted as a function of [urea]. Right panel: The rate constant of elongation, k_{elon} (determined from the slope of elongation phase) was log-transformed and plotted as a function of urea. Linear regression was performed on the plot to obtain the *m*-value for k_{elon}.

Table 5 Modelling parameters from fitting of SAXS data

Sample	Initial state					Final state				
	# mon[a]	χ^2	R_g (Å)[b]	$R_{g,arm}$ (Å)[c]	# arms[d]	C_{Ini} ($\frac{kDa}{nm^3}$)[e]	χ^2	R (Å)[f]	AR[g]	MPL ($\frac{kDa}{nm}$)[h]	C_{Fibril} ($\frac{kDa}{nm^3}$)[i]
FapC WT	6.2 ± 0.3	0.815	155 ± 12	110 ± 12	2 (f)[j]	0.009	1.08	34 ± 0.8	3.4 ± 0.1	31	0.24
FapC ΔR1R2R3	6.7 ± 0.1	5.51	55.6 ± 1.1	37 ± 0.7	2.7 ± 0.1	0.13	1.24	42 ± 0.5	2.41 ± 0.05	61	0.46

[a] Number of monomers calculated from I(0).
[b] Radius of gyration for the whole structure.
[c] Radius of gyration for each individual arm.
[d] Number of polymer arms extending from the center.
[e] Estimated compactness of the initial state calculated assuming a sphere with radius of R_g.
[f] Radius of the short axis of the fibril cross section.
[g] Aspect ratio between the short and long axis of the fibrillar cross section.
[h] Mass per length of the fibrils.
[i] Estimated compactness of the fibrils calculated by dividing MPL with area of ellipsoidal cross section.
[j] The simple model with one polymer chain is equivalent to a star model with two arms.

more compact. We have tried to quantify the compactness by dividing the mass of the average particle with the volume by assuming a sphere with radius R_g. This confirms that the initial state of FapC ΔR1R2R3 is more compact (0.13 $\frac{kDa}{nm^3}$) compared to FapC WT (0.009 $\frac{kDa}{nm^3}$).

Next, we measured the fibril states of both proteins using SAXS. The fibril states were measured after 25 hours for FapC WT and 168 hours (7 days) for FapC ΔR1R2R3 where fibrillation plateaued according to ThT fluorescence (Figure S9(A) and (B)). There are no FapC WT monomers left in solution at the end of fibrillation (Figure S9(C)). On the other hand, FapC ΔR1R2R3 still showed a considerable amount of monomer left in solution at this stage. This was quantified by constructing a standard curve using two dilution series of a freshly desalted monomer on a SDS-PAGE (control dilution series 1 and 2, Figure S9(D)). Based on this, we calculate that 58.1 +/- 7.3% SD of the soluble protein was incorporated into fibrils. The mass of the fibrillar state cannot be determined as $I(0)$ cannot be estimated from the low-q behavior since it does not levels off at a plateau value for q approaching zero. Instead the mass per length (MPL) can be calculated by determining the $I(0)$ of the cross section ($I(0)_{CS}$). This is possible because a long one-dimensional fibril will give a contribution following a q^{-1} slope (see blue line in Figure 6) that takes into account the scattering from the large dimension, leaving the signal from the cross section. To calculate the MPL, we also need to know the concentration of fibrils in the sample. For FapC WT (which is 100% incorporated into fibrils) this corresponds to the concentration of monomer (1.0 mg/mL), whereas for FapC ΔR1R2R3 this is 58% of the initial monomer concentration (0.58 mg/mL). However, in our calculation of MPL, we can neglect monomeric FapC ΔR1R2R3 because the initial state of FapC ΔR1R2R3 has a much lower signal (notice the y axis is $\log(I(q))$ compared to the fibrils due to the large mass difference. We obtain a MPL of 31 and 61 kDa/nm for FapC WT and ΔR1R2R3, respectively (Table 5), demonstrating clearly that FapC ΔR1R2R3 has a significantly higher MPL than WT.

The fibril states were fitted on arbitrary scale as an infinitely long cylinder with an ellipsoidal cross section. Furthermore, a random coil scattering contribution with R_g = 20 Å with a fitted overall scale was added to the model to describe the data at mid to high q ($q > 0.05$ Å^{-1}). In this q region the random coil scattering has a high signal compared to the form factor for a cylinder. Therefore, the contribution from random coil is evident in this q region despite the lower concentration and size. As the contribution is only evident in a narrow q region, the size of the random coil cannot be fitted. Instead, a reasonable value was estimated

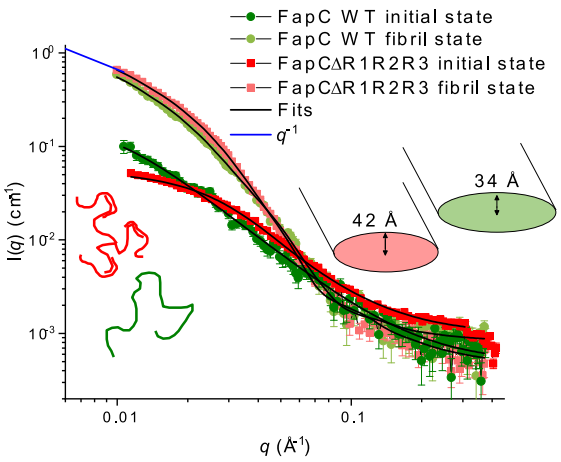

Figure 6. SAXS data and fits for initial and final states for FapC WT and FapC ΔR1R2R3. Sketches show the models used for each of the samples. For the initial state of FapC ΔR1R2R3 is shown with three arms that each contains around two monomers, giving ca. 6 monomers. The cylinder cross sections are shown with the fitted aspect ratios and the ratio between the short axis of the two models is calculated from fitted values. Data for model fits are listed in Table 5. The blue line of q^{-1} shows how the scattering of an infinitely long one-dimensional cylinder will behave at low q.

for R_g based on the predicted loop regions of FapC WT fibril (30–40 amino acids[42]and the whole length of the FapC ΔR1R2R3 monomer left in solution (132 amino acids). We used the empirical relation $R_g = 1.927\ N^{0.598}$, where N is the number of amino acids.[43] For loop regions this lies between 15 and 17 Å and for FapC ΔR1R2R3 monomer it is 36 Å. To compare fit values more easily, we set the contribution to 20 Å for both samples.

For FapC WT, the random coil contribution is thought to originate from the loop regions that are of a considerable size and predicted to be disordered even in the fibril state (see Supplementary Information and Figures S10–12 for a bioinformatic discussion of the identification of loop and repeat regions in the FapC sequence). For FapC ΔR1R2R3, this random coil contribution could also be due to loop regions, but is most likely an effect of the considerable amount of monomer still in solution. We find that the radius of the short axis is larger for FapC ΔR1R2R3 (42 compared to 34 Å), whereas the cross section of FapC WT is more elongated (aspect ratio 3.4 compared to 2.4) (Table 5). Sketches of the cross sections with relative scales in accordance with fitting parameters are shown as insets in Figure 6. By calculating the area of each of the cross sections, the MPL can be converted to a compaction with units mass per volume. This shows that the compactness of FapC ΔR1R2R3 (0.46 kDa/nm^3) is significantly larger than for FapC WT (0.24 kDa/nm^3). Also, the increase in compaction is greater for FapC ΔR1R2R3 (0.33 kDa/nm^3) than FapC WT (0.23 kDa/nm^3), commensurate with the increased *m*-value for FapC ΔR1R2R3. Note that when these fibril compactions are converted to g/cm^3, it is clear that both fibril structures are less compact (0.41 and 0.76 g/cm^3) than the expected 1.4 g/cm^3 for a globular compactly folded protein.

Discussion

Fibrillation sensitivity to urea reflects changes in surface area burial and predominantly affects nucleation and elongation

Our work demonstrates that fibrillation of both CsgA and FapC is sensitive to urea, though urea does not change the overall fibrillation mechanism for FapC WT (and by inference not CsgA either). Removal of repeats from FapC clearly increased the protein's sensitivity to urea, particularly the strongly destabilized variants FapC ΔR2R3 and FapC ΔR1R2R3, whose fibrillation was quenched at relatively low concentrations of urea, except when adding preformed seeds (which bypasses the nucleation step). Seeded fibrillation experiments provided elongation *m*-values, which were significantly higher for deletion mutants lacking two or three repeats. Using SAXS, we showed that the compactness of the deletion mutant, ΔR1R2R3 was larger compared to WT FapC in both the initial state (pre-fibril) and in the final fibril state. In addition, the change in compactness from initial to fibril state was larger for ΔR1R2R3 compared to WT FapC, consistent with the kinetic *m*-values (although the latter only reports on the activation step).

The basis for these *m*-values, *i.e.* a log-linear relationship between rate constants and denaturant concentration, echoes the folding and unfolding of globular proteins.[32] The ability to measure these *m*-values is aided by the unfolded nature of these proteins in the (largely) monomeric initial state, which by all accounts compels them to undergo a folding step in order to form nuclei or to attach and fold onto the growing ends of fibrils. Such a folding step involves a large reduction in SASA and can therefore be counteracted by urea.

By extracting molecular rates relating to nucleation, $k_{nuc}$and elongation, k_{elon} using linear regression we were able to obtain kinetic *m*-values for these individual steps in the fibrillation process. These data indicated that the activation steps of both primary nucleation and elongation were affected by urea, seemingly to the same extent. However, care must be taken when analyzing the lag time of unseeded fibrillation curves, since the duration of the lag time is not only dependent on the primary nucleation rate but also on the elongation rate and secondary processes such as secondary nucleation and fragmentation rates. When using the linear

regression method to extract rate constants from the fibrillation kinetics, it must be kept in mind that k_{nuc} and k_{elon} are pseudo-elementary rate constants which are composites of multiple elementary steps similar to the parameters obtained by the Finke-Watzky analysis.[44] There may therefore be microscopic processes hidden under the averaged rate constants, which are also affected by urea. The webserver programme AmyloFit can provide more precise information about how these processes are affected by urea. FapC fibrillation is best explained by a mechanism including primary nucleation, elongation and fragmentation.[26] Using a global curve-fitting analysis, we found that urea mainly reduces fibrillation-related folding and association (i.e. nucleation and growth), rather than fragmentation. This makes sense from the perspective of burial of surface area, since fragmentation leads to a smaller change in SASA on a monomer-unit level than the two other processes.

Seeding of FapC fibrillation simplifies fibrillation kinetics and reveals increased compaction in destabilized mutants

From our unseeded fibrillations curves, we did not observe a simple correlation between the *m*-values and the number of repeats in FapC. Considering that *m*-values are known to correlate with the amount of buried surface area during folding,[35,36] one would expect that a lower number of repeats would lead to a lower SASA (also during the activation step) and therefore a lower *m*-value. However, as explained above, the analysis of unseeded fibrillation curves is challenged by the complexity of the fibrillation mechanism which leads to composite rate constants. This phenomenon is also apparent from the higher magnitude of the apparent elongation rate constant, k_{elon} in the seeded reactions *versus* the unseeded reactions. The rate of elongation is dependent on both the elongation rate constant, the monomer concentration and the fibril number concentration.[37] By adding a high concentration of sonicated seeds, the fibril number concentration is dramatically increased, thereby increasing the overall rate of elongation. Since k_{elon} is determined from the slope of the elongation phase, k_{elon} is higher in seeded reactions. In contrast to seeded reactions, the fibril number concentration is a limiting factor for unseeded reactions, which makes k_{elon} strongly dependent on nucleation rates. The elongation rate constant of seeded reactions can therefore be considered a "cleaner" parameter to compare between protein constructs since it only depends on the elongation rate of the reaction whereas the elongation rate constants of unseeded reactions are also affected by the nucleation rates. The *m*-values based on rate constants extracted from unseeded fibrillation curves may therefore be composite values reporting on the sum of multiple molecular steps. To simplify this, we therefore used seeded fibrillation to bypass the nucleation step.

A high concentration of seeds was needed to effectively eliminate the lag phase of FapC fibrillation. This can be explained by the low degree of secondary nucleation existing in the system[29] compared to e.g. Aβ[45] and α-synuclein.[46] The *m*-values calculated from the seeded fibrillations curves were solely associated with the activation step of elongation. We were surprised to find that deletion of repeats resulted in higher *m*-values as this possibly reflected that otherwise solvent exposed parts of the protein were incorporated into the fibril during elongation. This would require a relatively large change in the overall fibril construction. Considering that the repeat-less mutant ΔR1R2R3 is able to fibrillate, it is perhaps not surprising that the weakening of the amyloidogenic core caused by removal of repeat sequences can be somewhat compensated for by including a larger portion of otherwise exposed protein regions into the amyloid backbone, thus creating a more compact fibril. Rosetta analysis (which predicts amyloid propensity at the hexapeptide level[47]indicates that regions outside the imperfect repeats also have significant aggregation propensity (Figure S13). We used SAXS analysis to confirm that fibrils formed by ΔR1R2R3 were more compact than WT FapC fibrils. However, given the low stability of these fibrils according to both the sensitivity of their formation to urea and extreme lack of resistance to formic acid[27], this compaction is likely to be quite ineffective and suggests an inferior packing arrangement, simply because the imperfect repeats form a much more optimized cross-β structure. Interestingly, the initial (unfolded) state of ΔR1R2R3 was also more compact than WT FapC but the change in compactness from the initial state to the fibril state remained larger for ΔR1R2R3, which was consistent with our observation of a larger *m*-value for ΔR1R2R3 compared to WT FapC.

In the seeded fibrillation experiments, larger *m*-values compared to WT FapC were not only seen in the extreme example where all three repeats were deleted. Deletion of two repeats (ΔR1R3 and ΔR2R3) or even a single repeat (ΔR2) also resulted in larger *m*-values compared to WT FapC. In view of the likely relation between *m*-values and compactness of the fibril TS (which again provides an indication of the compactness of the resulting fibril), this suggest that the removal of repeats from FapC does not simply lead to modular removal of the deleted repeat sequences from the mature fibril core. On the contrary, there seems to exist a certain degree of flexibility in terms of which parts of the sequences are incorporated into the fibril. This flexibility allowed FapC mutants to form fibrils of normally non-amyloid regions. This may be an important fact to consider in the development of rationally designed inhibitors aimed to block the interaction of specific

amyloidogenic regions of a protein. Based on our results which suggested significant architectural chances in the mature fibril of deletion mutants, it is tempting to speculate that FuBA can also circumvent blocking interactions of inhibitory compounds by changes in the fibril architecture. Although FapC deletion mutants are able to forms fibrils *in vitro*, it remains to be seen if these fibrils are functional *in vivo*. Indeed, previous work showed that deletion of repeats weakened the stability of fibrils and led to increased fragmentation during fibril formation which would most likely attenuate the function of these fibrils *in vivo*.[26,27]

Increased fragmentation rates may compensate for decreased nucleation and elongation rates

In our experiments, the ΔR3 mutant showed a different behavior than the remaining FapC variants. In the unseeded fibrillation curves, the slope of the elongation curve was essentially unaffected by increasing [urea] and the lag time was only moderately affected compared to other FapC variants. In the seeded fibrillation data for ΔR3, we observed a sharp transition at 4 M urea in the log elongation rate *versus* [urea] plot. This is indicative of a transition of the elongation mechanism as urea increases above 4 M. It has previously been shown that removal of R3 makes FapC particularly vulnerable to fragmentation during fibrillation.[26] Considering this increased tendency for fragmentation, the presence of urea may increase fragmentation even further of the ΔR3 mutant. Although the nucleation and elongation rates of ΔR3 are most likely reduced by urea, we speculate that the overall fibrillation rate is not dramatically reduced because of a compensating effect caused by an urea-induced increase of the fragmentation rate. Above a certain concentration of urea, the reduction in the primary fibrillation processes may eclipse the compensating acceleration of the fragmentation rate, leading to a sharp drop in overall elongation around 4 M.

The core of the FuBA fibril contains an evolutionarily optimized number of residues

Repeat sequences are commonly found in FuBA and are known from several proteins besides CsgA and FapC.[48,49] The number of repeat sequences is 4–5 in CsgA homologs in enterobacteria but variation in the number of repeats sequences is seen in CsgA homologs in other bacteria.[50] The number of repeat sequences in FapC homologs is 3 in *Pseudomonas* strains and in the majority of other genera, although a few exceptions with either fewer or more repeat sequences are known.[19] We found similar kinetic *m*-values for the two WT proteins, CsgA and FapC in our linear regression analysis and for the *m*-value associated with elongation rate. This may reflect a similar amount of SASA between the two proteins during the activation step of elongation (and by inference during the whole elongation step). Interestingly, CsgA and FapC contain a similar number of residues in repeat sequences (∼109 and ∼111 respectively) despite the large difference in size and overall composition. This indicates that the protein packing seen in WT CsgA and FapC provides optimal fibrillation kinetics and fibril stability.

Materials and Methods

Expression and purification of CsgA WT and FapC variants

Recombinant FapC variants were expressed and purified as described previously.[26] CsgA WT was expressed and purified according to the same protocol as FapC with the main difference that the CsgA WT construct was expressed in a pET11d plasmid encoding His-tagged CsgA WT with ampicillin resistance. Colonies and cultures expressing CsgA were therefore grown on agar plates and in Lysogeny Broth media containing 100 μg/mL ampicillin.

FuBA fibrillation assays monitored by ThT fluorescence

Fibrillation of CsgA and FapC variants in 0–8 M urea was monitored using ThT fluorescence as follows: CsgA or FapC in elution buffer (8 M GdmCl, 50 mM Tris-HCl, 500 mM imidazole, pH 8) was buffer exchanged into either buffer 1 (50 mM Tris-HCl, pH 7.4) or buffer 2 (buffer 1 with 8 M urea) using PD-10 desalting columns (GE Healthcare) according to the manufacturer's protocol. The protein concentration was estimated by measuring the absorbance at 280 nm using a NanoDrop 1000 (ND-1000 Spectrophotometer, Scientific, U.S.A.) (See Supplementary Table S1 for extinction coefficients and molecular weights). Protein in buffer 1 was mixed in different ratios with protein in buffer 2 to obtain 40 μM CsgA or 35 μM FapC in 0–8 M urea in steps of 0.5 M denaturant. 1.2 mM ThT in buffer 1 was added to all samples to a final concentration of 40 μM. 100 μL was added to a black clear-bottomed 96 well plate (Costar, Corning). Fibrillation was monitored at 37 °C in a CLARIOstar platereader (BMG Labtech) with excitation at 448 nm and emission at 485 nm at a gain of 1550. Measurements were made every 2 min and double orbital shaking at 200 RPM was applied between reads. The data was processed in KaleidaGraph (Synergy Software). Urea buffers were always stored frozen and thawed immediately before use to avoid decomposition of urea to isocyanic acid and carbamylation of protein samples.[51]

SDS-PAGE quantitation of aggregation

Samples at the stationary stage were transferred to Eppendorf tubes and centrifuged at 13,500 rpm for 15 min to pellet the fibrils. Then 25 μL supernatant from each sample was mixed with 5 μL 6x SDS loading buffer, heated at 95 °C for 5 min, and analyzed by SDS-PAGE on a 15% poly-acrylamide Bis-Tris gel. The protein bands of the SDS-PAGE gel were quantified using ImageJ and normalized using control samples of monomeric CsgA and FapC protein of a known concentration.

Normalization and correction of kinetic data

All fibrillation curves were corrected for the quenching effect of urea on the ThT signal and normalized to show percentage of aggregated protein as a function of time (Figure S1). This was done before any kinetic analysis was performed on the fibrillation curves. First, the percentage of aggregated protein was determined immediately after the experiment was ended using SDS-PAGE analysis as described above. Next, the raw fibrillation curves were baseline corrected by subtracting values obtained from ThT blanks without protein and the end-level plateau of the fibrillation was adjusted to the percentage determined by SDS-PAGE analysis. The corrected and normalized fibrillation curves now showed the true percentage of aggregated protein as a function of time.

Analysis of kinetic parameters from CsgA and FapC fibrillation curves

To obtain model-free values for lag times and elongation rates, linear regression was performed on the linear area of the normalized fibrillation curves. The slope of the elongation phase gives us the apparent elongation rate k_{elon}. The nucleation rate constant k_{nuc} is defined as $1/t_{lag}$, where t_{lag} is the time point for the intercept of the ThT fluorescence baseline with the elongation phase.

Determination of the fibrillation mechanism of FapC WT 0 M or 4 M urea

Fibrillation of FapC WT was monitored at 8–92 μM protein in 0 and 4 M urea as described in the previous section. Due to the significant reduction in ThT fluorescence by urea, we screened different ThT concentrations and found that the best signal was obtained at 40 μM ThT in 0 M urea and at 360 μM ThT in 4 M urea. In accord with our previous analyses of FapC fibrillation,[26] time curves were fitted with a fragmentation dominated model with the programme AmyloFit.[37] We used global curve fitting where only one kinetic parameter was allowed to vary for each curve at increasing [urea]. The curves were fitted by the model with variation in either elongation rate constant (k_+), nucleation rate constant (k_n) or fragmentation rate constant (k_-). The same experimental procedure and analysis was performed for CsgA WT.

Seeded FapC fibrillation assays monitored by ThT fluorescence: Fibrillation of FapC variants (35 μM monomer) in 0–7 M urea was monitored after addition of 19.5 μM seeds. The procedure was identical to the one for the unseeded fibrillation assays except that ThT was added to 100 μM and the volume of each sample was increased to 180 μL. Measurements were made every 1 min and double orbital shaking at 200 RPM was applied for 7 s between reads.

Small angle X-ray scattering (SAXS)

SAXS data were obtained for FapC ΔR1R2R3 in the initial state using the in-house laboratory instrument HyperSAXS[52] (NanoSTAR from Bruker AXS) that uses an Excillum liquid Ga metal jet source (λ_{Ga} = 1.34) and scatterless slits.[53–55] SAXS data for FapC WT in the initial state and fibril states for both samples were measured on another in-house Bruker AXS Nanostar instrument using a rotating Cu anode (λ_{Cu} = 1.54 Å), also with scatterless slits.[56] All data are plotted as a function of q defined as $q = (4\pi\sin(\theta))/\lambda$, where 2θ is the scattering angle. FapC ΔR1R2R3 samples were measured with a concentration of 1 mg/mL for 0.5 h at 20 °C. FapC WT samples were measured at 1 mg/mL for 10 min at 20 °C. Buffer was subtracted and data converted to absolute scale using the SUPERSAXS package (Olivera & Pedersen, unpublished) and the scattering from a MilliQ water sample measured at 20 °C. The overall forward scattering, $I(0)$, and the forward scattering of the fibrillary cross settion, $I(0)_{CS}$, were determined by the home-made program WIFT[57,58] that performs an indirect Fourier transformation, where the contribution of the long fibrils dimension were taken into account when calculating $I(0)_{CS}$ as described in[59] $I(0)$ was used for calculating the mass for the initial states as $M = I(0)/(c\times\rho_m^2)$, where c is the concentration and ρ_m is the excess scattering length density of the protein per unit mass, which is set to 2.0×10^{10} cm/g. The initial states were fitted with either a linear Gaussian polymer model[60] (WT) or a Gaussian star polymer model,[61] where the number of arms is fitted (ΔR1R2R3). Conversion between the radius of gyration R_g of the whole structure and that of an arm $R_{g,arm}$ for the star model was done using $R_g^2 = \left(3 - \frac{2}{n}\right) \times R_{g,arm}^2$, where n is the average number of arms. The fibril state is fitted with a model of an infinitely long cylinder with an ellipsoidal cross section.[42,62] Furthermore, a scaled contribution of polymers with a radius of gyration of of 20 Å is added, as this was found to be necessary to adequately describe the data at high q. For FapC WT, this contribution most likely stems from the large loop regions and for FapC ΔR1R2R3 it proba-

bly stems from the considerable amount of monomer left in solution.

Accession Numbers

CsgA: UniProtKB - P28307.
FapC: UniProtKB - C4IN70.

CRediT authorship contribution statement

Thorbjørn V. Sønderby: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Helena Ø. Rasmussen:** Formal analysis, Investigation, Writing – original draft. **Signe A. Frank:** Investigation. **Jan Skov Pedersen:** Formal analysis, Supervision. **Daniel E. Otzen:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Acknowledgements

We thank Dr. Georg Meisl for from the Department of Chemistry at University of Cambridge for valuable discussion of results. D.E. O. and T.V.S. gratefully acknowledge support from the Independent Research Foundation Denmark | Natural Sciences (grant no. 8021-00208B and 8021-00133B) and from the Sino-Danish Center.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2021.167337.

Received 15 June 2021;
Accepted 27 October 2021;
Available online 5 November 2021

Keywords:
bacterial amyloid;
aggregation mechanisms;
chemical denaturation;
m-values;
protein compaction

Abbreviations used:
E. coli, *Escherichia coli* (*E. coli*); Fap, Functional amyloids in Pseudomonas; FTIR, Fourier transform infrared spectroscopy; FuBA, Functional Amyloids in Bacteria; MPL, Mass per length; MRE, Mean residual error; SDS-PAGE, Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis; SASA, solvent accessible surface area; SAXS, Small angle X-ray scattering; TEM, Transmission electron microscopy; ThT, Thioflavin T

References

1. Chiti, F., Dobson, C.M., (2017). Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86**, 27–68.
2. Fink, A. L. (2006). Factors affecting the fibrillation of α-synuclein, a natively unfolded protein, in: Misbehaving Proteins Protein (Mis)Folding, Aggregation, Stab., 2006: pp. 265–285.
3. Chiti, F., Dobson, C.M., (2006). Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **75**, 333–366.
4. Fowler, D.M., Koulov, A.V., Alory-Jost, C., Marks, M.S., Balch, W.E., Kelly, J.W., (2006). Functional amyloid formation within mammalian tissue. *PLoS Biol.* **4**, 0100–0107.
5. Saupe, S.J., (2020). Amyloid Signaling in Filamentous Fungi and Bacteria. *Annu. Rev. Microbiol.* **74**, 673–691.
6. Chapman, M.R., Robinson, L.S., Pinkner, J.S., Roth, R., Heuser, J., Hammar, M., Normark, S., Hultgren, S.J., (2002). Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science (New York, N.Y.)* **295**, 851–855.
7. Dueholm, M.S., Petersen, S.V., Sønderkær, M., Larsen, P., Christiansen, G., Hein, K.L., Enghild, J.J., Nielsen, J.L., et al., (2010). Functional amyloid in pseudomonas. *Mol. Microbiol.* **77**, 1009–1020.
8. Dueholm, M.S., Nielsen, P.H., Chapman, M., Otzen, D., Romero, D., Kolter, R., Dueholm, M.S., Nielsen, P.H., et al., (2013). Functional Amyloids in Bacteria. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 411–438.
9. Zeng, G., Vad, B.S., Dueholm, M.S., Christiansen, G., Nilsson, M., Tolker-Nielsen, T., Nielsen, P.H., Meyer, R.L., et al., (2015). Functional bacterial amyloid increases Pseudomonas biofilm hydrophobicity and stiffness. *Front. Microbiol.* **6**, 1–14.
10. Chapman, M.R., Robinson, L.S., Pinkner, J.S., Roth, R., Heuser, J., Hammar, M., Normark, S., Hultgren, S.J., (2002). Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* **295**, 851–855.
11. Barnhart, M.M., Chapman, M.R., (2006). Curli biogenesis and function. *Annu. Rev. Microbiol.* **60**, 131–147.
12. Hammar, M., Arnqvist, A., Bian, Z., Olsén, A., Normark, S., (1995). Expression of two csg operons is required for production of fibronectin- and Congo red-binding curli polymers in Escherichia coli K-12. *Mol. Microbiol.* **18**, 661–670.
13. White, A.P., Collinson, S.K., Banser, P.A., Gibson, D.L., Paetzel, M., Strynadka, N.C.J., Kay, W.W., (2001). Structure and characterization of AgfB from Salmonella

enteritidis thin aggregative fimbriae. *J. Mol. Biol.* **311**, 735–749.
14. Hammer, N.D., Schmidt, J.C., Chapman, M.R., (2007). The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. *PNAS* **104**, 12494–12499.
15. Nenninger, A.A., Robinson, L.S., Hultgren, S.J., (2009). Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. *Proc. Natl. Acad. Sci.* **106**, 900–905.
16. Evans, M. L., Chorell, E., Almqvist, F., Chapman, M. R., Go, A., Li, F., Koch, M., Sefer, L., et al. (2015). The Bacterial Curli System Possesses a Potent and Selective Inhibitor of Amyloid Formation, Mol. Cell. 57 (2015) 445–455.
17. Taylor, J. D., Hawthorne, W. J., Lo, J., Dear, A., Jain, N., Meisl, G., Andreasen, M., Fletcher, C., et al. (2016). Electrostatically-guided inhibition of Curli amyloid nucleation by the CsgC-like family of chaperones, Sci. Rep. 6 (2016) 24656.
18. Robinson, L. S., Ashman, E. M., Hultgren, S. J. & Chapman, M. R. (2006). Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein, Mol. Microbiol. 59 (2006) 870–881.
19. Dueholm, M. S., Otzen, D. & Nielsen, P. H. (2013). Evolutionary Insight into the Functional Amyloids of the Pseudomonads, PLoS One. 8 (2013) e76630.
20. Rouse, S.L., Hawthorne, W.J., Berry, J.L., Chorev, D.S., Ionescu, S.A., Lambert, S., Stylianou, F., Ewert, W., et al., (2017). A new class of hybrid secretion system is employed in Pseudomonas amyloid biogenesis. *Nature Commun.* **8**
21. Dueholm, M.S., Sondergaard, M.T., Nilsson, M., Christiansen, G., Stensballe, A., Overgaard, M.T., Givskov, M., Tolker-Nielsen, T., et al., (2013). Expression of Fap amyloids in Pseudomonas aeruginosa, P. fluorescens, and P. putida results in aggregation and increased biofilm formation. *MicrobiologyOpen* **2**, 365–382.
22. Tian, P., Boomsma, W., Wang, Y., Otzen, D.E., Jensen, M. H., Lindorff-Larsen, K., Lindor, K., (2015). Structure of a functional amyloid protein subunit computed using sequence variation. *J. Am. Chem. Soc.*, 19–22.
23. Wang, X., Chapman, M.R., (2008). Curli provide the template for understanding controlled amyloid propagation. *Prion* **2**, 57–60.
24. Rouse, S.L., Matthews, S.J., Dueholm, M.S., (2018). Ecology and biogenesis of functional amyloids in pseudomonas. *J. Mol. Biol.* **430**, 3685–3695.
25. Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B., Blake, C.C., Wang, X., Chapman, M.R., (2008). Curli provide the template for understanding controlled amyloid propagation. *Prion* **2**, 57–60.
26. Rasmussen, C.B., Christiansen, G., Vad, B.S., Lynggaard, C., Enghild, J.J., Andreasen, M., Otzen, D., (2019). Imperfect repeats in the functional amyloid protein FapC reduce the tendency to fragment during fibrillation. *Protein Sci.* **28**, 633–642.
27. Christensen, L. F. B., Nowak, J. S., Sønderby, T. V., Frank, S. A. & Otzen, D. E. (2020). Quantitating denaturation by formic acid: Imperfect repeats are essential to the stability of the functional amyloid protein FapC, J. Biol. Chem. 295 (2020) 13031–13046.
28. Wang, X., Smith, D.R., Jones, J.W., Chapman, M.R., (2007). In vitro polymerization of a functional Escherichia coli amyloid protein. *J. Biol. Chem.* **282**, 3713–3719.
29. Andreasen, M., Meisl, G., Taylor, J.D., Michaels, T.C.T., Levin, A., Otzen, D.E., Chapman, M.R., Dobson, C.M., et al., (2019). Physical Determinants of Amyloid Assembly in Biofilm Formation. *mBio* **10**, 1–12.
30. Wang, H., Shu, Q., Rempel, D. L., Frieden, C. & Gross, M. L. (2017). Understanding curli amyloid-protein aggregation by hydrogen – deuterium exchange and mass spectrometry., Int. J. Mass Spectrom. 420 (2017) 16–23.
31. Dueholm, M.S., Nielsen, S.B., Hein, K.L., Nissen, P., Chapman, M., Christiansen, G., Nielsen, P.H., Otzen, D.E., (2011). Fibrillation of the major curli subunit CsgA under a wide range of conditions implies a robust design of aggregation. *Biochemistry* **50**, 8281–8290.
32. Parker, M.J., Spencer, J., Clarke, A.R., (1995). An integrated kinetic analysis of intermediates and transition states in protein folding reactions. *J. Mol. Biol.* **253**, 771–786.
33. Timasheff, S.N., Xie, G., (2003). Preferential interactions of urea with lysozyme and their linkage to protein denaturation. *Biophys. Chem.* **105**, 421–448.
34. Pace, C. N., Grimsley, G. R. & Scholtz, J. M. (2008). Denaturation of Proteins by Urea and Guanidine Hydrochloride, Vol. 1, pp. 45–69.
35. Fersht, A.R., (1999). Structure and mechanism in protein science. A guide to enzyme catalysis and protein folding. Freeman & Co., New York.
36. Myers, J.K., Pace, C.N., Scholtz, J.M., (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding [published erratum appears in Protein Sci 1996 May; 5(5):981]. *Protein Sci.* **4**, 2138–2148.
37. Meisl, G., Kirkegaard, J.B., Arosio, P., Michaels, T.C., Vendruscolo, M., Dobson, C.M., Linse, S., Knowles, T.P., (2016). Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nat Protoc.* **11**, 252–272.
38. Zandomeneghi, G., Krebs, M.R., McCammon, M.G., Fandrich, M., (2004). FTIR reveals structural differences between native beta-sheet proteins and amyloid fibrils. *Protein Sci.* **13**, 3314–3321.
39. Kawahara, K., Tanford, C., (1966). Viscosity and density of aqueous solutions of urea and guanidine hydrochloride. *J. Biol. Chem.* **241**, 3228–3232.
40. Flyvbjerg, H., Jobs, E., Leibler, S., (1996). Kinetics of self-assembling microtubules: an “inverse problem” in biochemistry. *Proc. Natl. Acad. Sci. USA* **93**, 5975–5979.
41. Rambo, R.P., Tainer, J.A., (2011). Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559–571.
42. Nagaraj, M., Ahmed, M., Lyngsø, J., Vad, B.S., Bøggild, A., Fillipsen, A., Pedersen, J.S., Otzen, D.E., et al., (2020). Predicted Loop Regions Promote Aggregation: A Study of Amyloidogenic Domains in the Functional Amyloid FapC. *J. Mol. Biol.*.
43. Kohn, J.E., Millett, I.S., Jacob, J., Zagrovic, B., Dillon, T.M., Cingel, N., Dothager, R.S., Seifert, S., et al., (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA* **101**, 12491–12496.
44. Morris, A.M., Watzky, M.A., Agar, J.N., Finke, R.G., (2008). Fitting neurological protein aggregation kinetic data via a 2-step, minimal/“Ockham’s razor” model: The Finke-Watzky mechanism of nucleation followed by autocatalytic surface growth. *Biochemistry* **47**, 2413–2427.

45. Cohen, S.I., Linse, S., Luheshi, L.M., Hellstrand, E., White, D.A., Rajah, L., Otzen, D.E., Vendruscolo, M., et al., (2013). Proliferation of amyloid-beta42 aggregates occurs through a secondary nucleation mechanism. *Proc. Natl. Acad. Sci. USA* **110**, 9758–9763.
46. Gaspar, R., Meisl, G., Buell, A.K., Young, L., Kaminski, C. F., Knowles, T.P.J., Sparr, E., Linse, S., (2017). Secondary nucleation of monomers on fibril surface dominates alpha-synuclein aggregation and provides autocatalytic amyloid amplification. *Q. Rev. Biophys.* **50**, e6
47. Goldschmidt, L., Teng, P.K., Riek, R., Eisenberg, D., (2010). Identifying the amylome, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci.* **107**, 3487–3492.
48. Jordal, P.B., Dueholm, M.S., Larsen, P., Petersen, S.V., Enghild, J.J., Christiansen, G., Højrup, P., Nielsen, P.H., et al., (2009). Widespread abundance of functional bacterial amyloid in mycolata and other gram-positive bacteria. *Appl. Environ. Microbiol.* **75**, 4101–4110.
49. Oh, J., Kim, J.G., Jeon, E., Yoo, C.H., Moon, J.S., Rhee, S., Hwang, I., (2007). Amyloidogenesis of type III-dependent harpins from plant pathogenic bacteria. *J. Biol. Chem.* **282**, 13601–13609.
50. Dueholm, M. S., Albertsen, M., Otzen, D. & Nielsen, P. H. (2012). Curli Functional Amyloid Systems Are Phylogenetically Widespread and Display Large Diversity in Operon and Protein Structure, PLoS One. 7 (2012) e51274.
51. Sun, S., Zhou, J.Y., Yang, W., Zhang, H., (2014). Inhibition of protein carbamylation in urea solution using ammonium-containing buffers. *Anal. Biochem.* **446**, 76–81.
52. Lyngso, J., Pedersen, J.S., (2021). A high-flux automated laboratory small-angle X-ray scattering instrument optimized for solution scattering. *J. Appl. Crystallogr.* **54**, 295–305.
53. Li, Y., Beck, R., Huang, T., Choi, M.C., Divinagracia, M., (2008). Scatterless Hybrid Metal-Single Crystal Slit for Small Angle X-ray Scattering and High Resolution X-ray Diffraction. *Appl. Cryst.* **41**, 1134–1139.
54. Pedersen, J. S. (2018). X-ray analyzing system for x-ray scattering analysis. US Patent US9958404B2 vols.
55. Pedersen, J. S. (2019). X-ray analyzing system for x-ray scattering analysis. European Patent EP13159569.6.
56. Pedersen, J.S., (2004). A flux- and background-optimized version of the NanoSTAR small-angle X-ray scattering camera for solution scattering. *J. Appl. Crystall.* **37**, 369–380.
57. Pedersen, J.S., Hansen, S., Bauer, R., (1994). The Aggregation Behavior of Zinc-free Insulin Studied by Small-angle Neutron Scattering. *Eur. Biophys. J.* **23**, 379–389.
58. Glatter, O., (1977). A new method for the evaluation of small-angle scattering data. *J. Appl. Crystall.* **10**, 415–421.
59. Pedersen, J.S., Schurtenberger, P., (1996). Cross-Section Structure of Cylindrical and Polymer-Like Micelles from Small-Angle Scattering Data. I. Test of Analysis Methods. *J. Appl. Crystallogr.* **29**, 646–661.
60. Debye, P., (1947). Molecular-Weight Determination by Light Scattering. *J. Phys. Coll. Chem.* **51**, 18–32.
61. Benoit, H., (1953). On the effect of branching and polydispersity on the angular distribution of the light scattered by gaussian coils. *J. Polym. Sci.* **11**, 507–510.
62. Rasmussen, H., Otzen, D.E., Pedersen, J.S., (2021). A multimethod approach for analyzing FapC fibrillation and determining mass per length. *Biophys. J.* **120**, 2262–2275.