Gatekeeper residues in the major curlin subunit modulate bacterial amyloid fiber biogenesis

Xuan Wang[a,1], Yizhou Zhou[a], Juan-Jie Ren[a], Neal D. Hammer[b], and Matthew R. Chapman[a,2]

[a]Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109; and [b]Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109

Edited by Reed B. Wickner, National Institutes of Health, Bethesda, MD, and approved November 6, 2009 (received for review August 3, 2009)

Amyloid fibers are filamentous protein structures commonly associated with neurodegenerative diseases. Unlike disease-associated amyloids, which are the products of protein misfolding, *Escherichia coli* assemble membrane-anchored functional amyloid fibers called curli. Curli fibers are composed of two proteins, CsgA and CsgB. In vivo, the polymerization of the major curli subunit protein, CsgA, is dependent on CsgB-mediated nucleation. The amyloid core of CsgA features five imperfect repeats (R1–R5), and R1 and R5 govern CsgA responsiveness to CsgB nucleation and self-seeding by CsgA fibers. Here, the specificity of bacterial amyloid nucleation was probed, revealing that certain aspartic acid and glycine residues inhibit the intrinsic aggregation propensities and nucleation responsiveness of R2, R3, and R4. These residues function as "gatekeepers" to modulate CsgA polymerization efficiency and potential toxicity. A CsgA molecule lacking gatekeeper residues polymerized in vitro significantly faster than wild-type CsgA and polymerized in vivo in the absence of the nucleation machinery, resulting in mislocalized fibers. This uncontrolled polymerization was associated with cytotoxicity, suggesting that incorrectly regulated CsgA polymerization was detrimental to the cell.

amyloid | curli | protein aggregation | nucleation | *Escherichia coli*

Amyloids are ordered proteinaceous fibers commonly associated with mammalian neurodegenerative diseases and prion-based encephalopathies (1). Amyloid fibers have distinct biochemical and biophysical properties, such as remarkable resistance to chemical and thermal denaturation, and specific tinctorial properties when bound to Congo red and thioflavin T (ThT) (1). The molecular basis of neurodegenerative disease development induced by amyloid propagation remains elusive, partially because of the seemingly erratic and uncontrolled nature of amyloidogenesis. An emerging focus of amyloid biosynthesis has shown that amyloids can also be an integral part of physiology found in different organisms including bacteria, fungi, and mammals (2, 3). How nature coordinates functional amyloid propagation and reduces the associated cytotoxicity is poorly understood.

Curli, a bacterially produced functional amyloid, is an important component of the extracellular matrix and is involved in bacterial community behaviors (4). Because of the amyloid properties of curli fibers (5, 6), the colonies of curli-producing *Escherichia coli* stain red when grown on Congo red indicator plates, which provides a convenient assay to monitor curli assembly in vivo (7). In *E. coli*, at least six proteins are dedicated to directing efficient curli formation. Curli fibers are composed of a major subunit CsgA and a minor subunit CsgB. CsgA remains unpolymerized until it encounters the surface-tethered nucleator CsgB, which initiates CsgA polymerization (8). CsgD is a transcriptional activator for the *csgBA* operon (4). CsgG, CsgE, and CsgF are nonfiber structural accessory proteins involved in secretion and stabilization of the fiber subunits and modulation of fiber assembly (6). CsgG is proposed to be the curli secretion apparatus that directs the secretion of CsgA, CsgB, and CsgF across the outer membrane (9, 10). CsgE and CsgF interact with CsgG at the outer membrane (9). CsgF is required for efficient CsgB-mediated nucleation, and CsgE is critical for CsgA, CsgB, and CsgF stability (6, 9, 10).

The CsgA primary amino acid sequence comprises a Sec-signal peptide (positions 1–20), a 22-residue (positions 21–42) CsgG-specific N-terminal domain, and an amyloid core region composed of five imperfect repeats (positions 43–151), each composed of 19–23 amino acids (Fig. 1*A*) (4). The repeats (R1, R2, R3, R4, and R5) share at least 30% identity at the amino acid level and are distinguished by the consensus sequence Ser-X_5-Gln-X_4-Asn-X_5-Gln (11). We previously found that R1 and R5 direct CsgA to respond to CsgB-mediated nucleation and are critical for fiber elongation (12). Moreover, internally conserved Gln and Asn residues located in R1 and R5 of CsgA are required for CsgB nucleation and curli assembly (13). Because all five repeats have these critical Gln and Asn residues, it has been unclear why R1 and R5 are functionally distinct from R2, R3, and R4. We elucidated these determinants and found that certain Asp and Gly residues in R2, R3, and R4 function as gatekeeper residues, which inhibit the amyloidogenic properties of these repeating units.

Gatekeeper residues are found in many globular proteins; they act to mask aggregation-prone sequences and thereby, promote native protein folding (14–17). Herein, we explore the role of gatekeeper residues in functional amyloidogenesis. Instead of simply inhibiting aggregation, we found that gatekeeper residues actually harness the aggregation of CsgA. A CsgA mutant lacking all gatekeeper residues efficiently polymerized into amyloid fibers independently of the curli assembly proteins CsgB and CsgF, leading to uncontrolled and mislocalized fibers. Moreover, the induced expression of CsgA without gatekeeper residues dramatically decreased the viability of *E. coli* cells.

Results

Seeding Specificity Is Encoded in R1 and R5. Amyloid formation is a self-propagating process (1). Preformed amyloid fibers can accelerate the polymerization of soluble subunits by a mechanism called seeding (18). Our previous results showed that R1 and R5 of CsgA are required for both CsgA–CsgB interactions and CsgA–CsgA interactions (12). Among five repeating units, only R1 and R5 polymerization can be accelerated by CsgB and CsgA (12). In vitro, peptides composed of repeating units of R1, R3, and R5 efficiently polymerize into ThT fluorescence-positive fibrous structures (19). At higher concentrations, R2 and R4 also eventually assemble into a fibrous structure that did not significantly fluoresce in the presence of ThT (19). We used these preformed fibrous materials as seeds to directly test the interaction specificity

Author contributions: X.W. and M.R.C. designed research; X.W., Y.Z., and J.R. performed research; X.W., Y.Z., J.R., and M.R.C. analyzed data; X.W. and M.R.C. wrote the paper; and N.D.H. contributed new reagents/analytic tools.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]Present address: Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611.

[2]To whom correspondence should be addressed at: 830 North University, Ann Arbor, MI 48109. E-mail: chapmanm@umich.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0908714107/DCSupplemental.

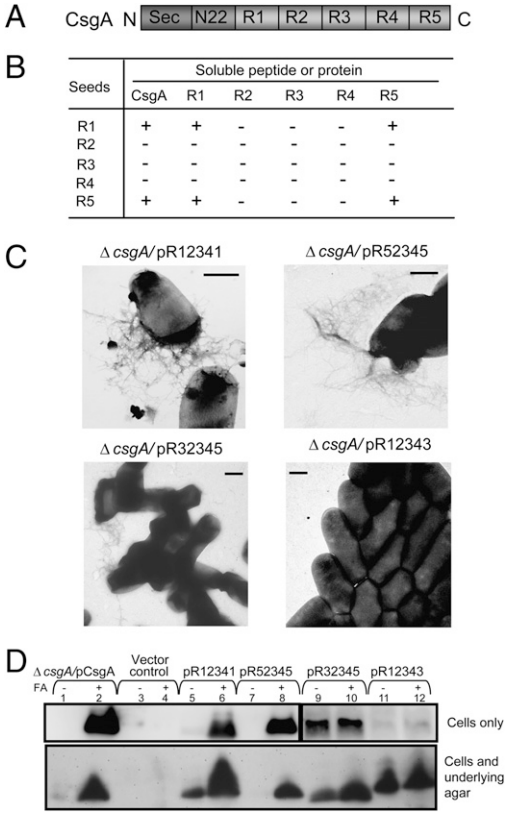

Fig. 1. Seeding specificity of CsgA was shared by R1 and R5, and R1 and R5 can replace each other. (*A*) The schematic of CsgA primary structure. (*B*) Summary of peptides seeding specificity. +, preformed fiber seeds that promote the polymerization of soluble proteins or peptides; –, preformed seeds that did not change polymerization significantly. (*C*) Negative-stain EM micrographs of the *csgA* mutant strain containing plasmids pR12341, pR52345, pR32345, or pR12343. (Scale bar, 500 nm.) (*D*) Western blots of whole-cell lysates *(Top)* or plugs (whole cells and underlying agar; *Bottom*) from *csgA* mutant cells containing plasmids pCsgA (lanes 1 and 2), vector control (lanes 3 and 4), pR12341 (lanes 5 and 6), pR52345 (lanes 7 and 8), pR32345 (lanes 9 and 10), or pR12343 (lanes 11 and 12) grown for 48 h at 26°C on YESCA plates. Cells were treated with (+) or without (–) FA before electrophoresis as indicated. The blots were probed with anti-CsgA antibody.

among the repeating units of CsgA. Cross-seeding and self-seeding occurred for peptides R1 and R5 only (Fig. 1*B*). Preformed R1 or R5 fibers promoted CsgA polymerization (Fig. 1*B*). These results suggest that CsgA seeding specificity is determined and shared by R1 and R5. Recent solid-state NMR and electron microscopy data of CsgA fibers are consistent with the β-helix-like structure predicted by homology modeling where R1 and R5 are positioned at the interface of adjacent fiber subunits (5, 11).

R1 and R5 Are Interchangeable In Vivo. R1 and R5 both can seed CsgA polymerization, and both repeating units constitute the nucleation and seeding responsive domains of CsgA (Fig. 1*B*) (12). In addition, cross-seeding occurred between R1 and R5 (Fig. 1*B*). These results suggest R1 and R5 are functionally redundant. To test this hypothesis in vivo, the residues of R1 of CsgA (R12345) were replaced by the corresponding residues from R5 to form the mutant construct with the following repeating unit order: R52345. The R52345 molecule was able to complement a *csgA* mutant and assembled into fibers on the cell surface (Fig. 1*C* and Table S1). Similarly, a CsgA analog in which R5 was replaced by R1 (pR12341) also assembled into fibers on the cell surface (Fig. 1*C*). Wild-type curli fibers are SDS insoluble, and formic acid (FA) treatment is required to dissociate the monomers from the fibers (20). An abundance of R52345 and R12341 can be resolved with FA treatment, suggesting that both of these chimeras form SDS insoluble fibers similar to those formed by wild-type CsgA (Fig. 1*D*, lanes 1, 2, 5, 6, 7, and 8). The in vivo polymerization of R52345 and R12341 was completely defective in *csgBA* cells, suggesting that their polymerization was still dependent on CsgB (Table S1).

Peptides composed of CsgA R3 can form ThT-positive short fibrous polymers, which suggests that R3 has some amyloidogenic properties (19). In addition, the sequence similarity (70%) and identity (48%) between R3 and R5 are higher than those between R1 and R5 (52% similarity and 39% identity). We asked whether or not R3 could substitute for R1 or R5 in vivo. R32345 was somewhat defective in vivo as observed by transmission electron microscopy (TEM) from *csgA*/pR32345 compared with *csgA*/pR12341 or *csgA*/pR52345 (Fig. 1*C* and Table S1). Furthermore, R32345 remained SDS soluble and did not require FA treatment to migrate into the gel (Fig. 1*D*, lanes 9 and 10). The phenotype of R12343 was even more dramatic. R12343 was not detected by Western blotting of the cell pellets but instead was only detected when both cells and the underlying agar were collected. This suggests that R12343 was stably secreted but defective in curli assembly (Fig. 1*D*, lanes 11 and 12). Very few fibers were produced by *csgA*/pR12343 when analyzed by TEM (Fig. 1*C* and Table S1). Therefore, R3 was not able to substitute for R5 at the *C* terminus and was only partially capable of replacing R1 at the *N* terminus, suggesting that the nucleation/CsgA self-seeding response and specificity was encoded by R1 and R5.

Identification of Gatekeeper Residues of R3 that Inhibit Amyloidogenic Properties. Our results indicated that common residues found in both R1 and R5 are involved in the observed nucleation response and seeding specificity. Alternatively, R3 could contain specific residues that make it inert to the interactions critical for polymerization. Guided by this perspective, we compared R3 with R1/R5, looking for residues in R3 that were distinctive from both R1 and R5. Four residues were targeted as candidates: Asp^{136}, Arg^{140}, Asp^{149}, and Trp^{151} in the molecule R12343 (Fig. 2*A*). If these residues hindered the ability of R3 to phenotypically replace R5, then changing these residues to those found in either R1 or R5 may result in a phenotype more similar to wild type. Constructs $pR12343^{D136N}$, $pR12343^{R140Y}$, $pR12343^{R140V}$, $pR12343^{D149L}$, $pR12343^{D149H}$, $pR12343^{W151T}$, and $pR12343^{W151Y}$ were engineered to test this hypothesis (Fig. 2*A*). Strikingly, $R12343^{D136N}$, $R12343^{D149L}$, and $R12343^{D149H}$ were able to assemble into amyloid-like fibers as measured by TEM (Fig. 2*B*), whereas other mutant proteins such as $R12343^{R140Y}$, $R12343^{R140V}$, $R12343^{W151T}$, and 12343^{W151Y} were as defective as R12343 in fiber formation in vivo as measured by TEM and Congo red binding Table S1). For $R12343^{D136N}$, $R12343^{D149L}$, and $R12343^{D149H}$, there was more SDS insoluble than SDS soluble material detected by whole-cell Western analysis, unlike $R12343^{R140Y}$, $R12343^{R140V}$, $R12343^{W151T}$, and 12343^{W151Y}, which were not associated with cells (Fig. 2*C*). $R12343^{R140Y}$, $R12343^{R140V}$, $R12343^{W151T}$, and 12343^{W151Y} were stably expressed and secreted as measured by Western analysis of the cells and underlying agar (Fig. S1). In vitro polymerization of $R12343^{D149H}$ was dramatically faster than that of R12343 as measured by ThT fluorescence (Fig. 2*D*). In a CsgB-dependent manner, purified-soluble CsgA can be converted into amyloid fibers on *csgA* cells expressing CsgB as evidenced by TEM or Congo red staining (12). When $R12343^{D149H}$ and R12343 were spotted on *csgA* cells at 10 μM, $R12343^{D149H}$ assembled into fibers more readily than R12343 as indicated by stronger Congo red staining (Fig. 2*D Inset*).

There are at least two possibilities for why $R12343^{D136N}$, $R12343^{D149L}$, and $R12343^{D149H}$ function similarly to wild-type CsgA. First, Asp residues at positions 136 and 149 in R12343 could play a negative role in nucleation and fiber elongation. By changing the Asp residues, the amyloidogenic property was improved. Alternatively, Asn and His residues at positions 136 and 149 in wild-type CsgA are critical for CsgA polymerization. To elucidate these

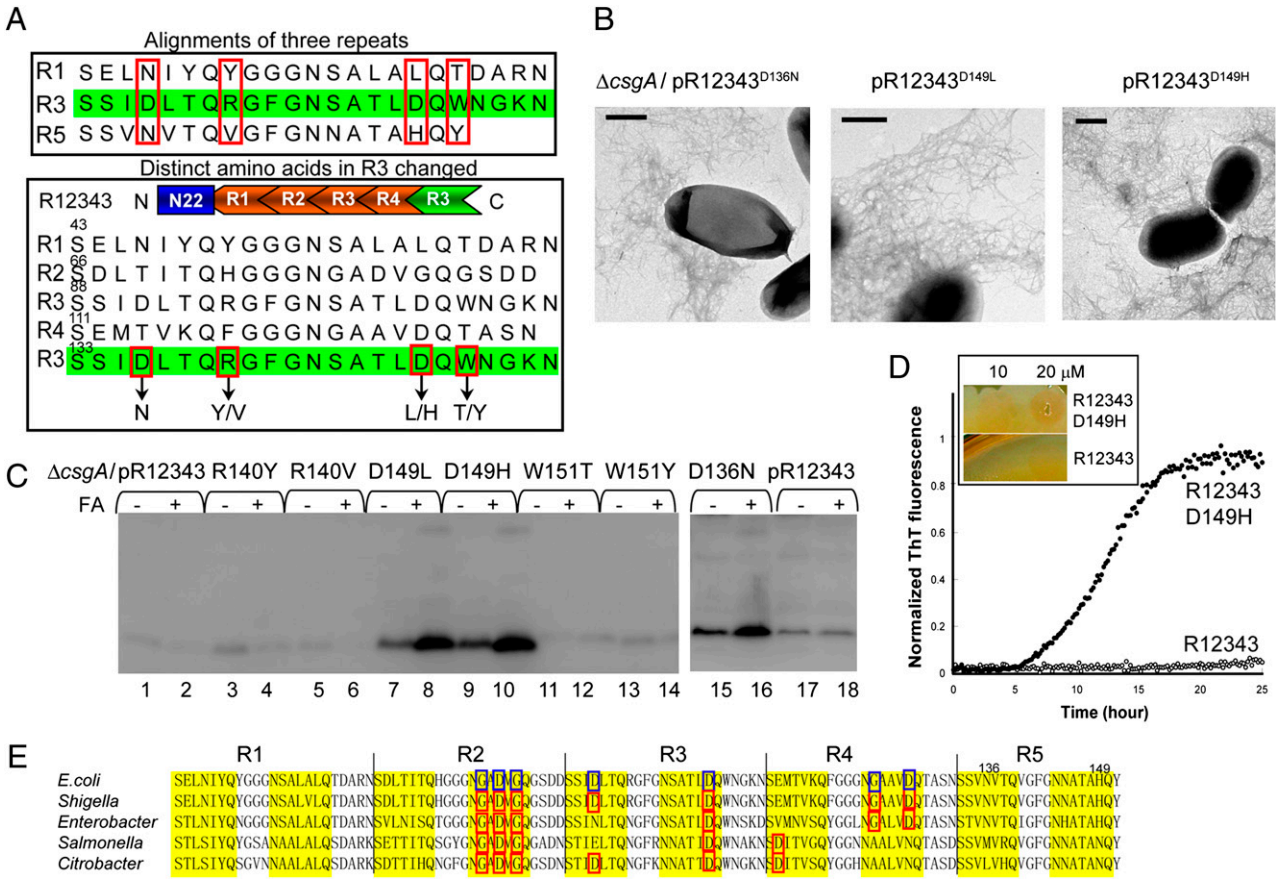

Fig. 2. Mutations of Asp residues of *C*-terminal R3 in R12343 rendered amyloidogenic properties. (*A*) Differences between R3 and R1/R5 amino acid sequences are indicated with red boxes at *Top*. The amino acid changes in *C*-terminal R3 of R12343 are indicated at *Bottom*. (*B*) Negative-stain EM micrographs of *csgA* cells containing plasmids pR12343^{D136N}, pR12343^{D149L}, and pR12343^{D149H}. Cells were grown 48 h at 26°C on YESCA plates. (Scale bar, 500 nm.) (*C*) Western blots of whole-cell lysates from *csgA* cells containing constructs pR12343 (lanes 1, 2, 17, and 18), pR12343^{D136N} (lanes 15 and 16), pR12343^{R140Y} (lanes 3 and 4), pR12343^{R140V} (lanes 5 and 6), pR12343^{D149L} (lanes 7 and 8), pR12343^{D149H} (lanes 9 and 10), pR12343^{W151T} (lanes 11 and 12), and pR12343^{W151Y} (lanes 13 and 14) grown 48 h at 26°C on YESCA plates. Samples were treated with (+) or without (–) FA as indicated. The blots were probed with anti-CsgA antibody. (*D*) The polymerization of 20-μM R12343^{D149H} and R12343 measured by ThT fluorescence. Congo red staining of *csgA* (CsgB^{+}) cells overlaid with10 μM and 20 μM R12343^{D149H} and R12343 is shown (*Inset*). (*E*) The alignment of five repeating units of CsgA from different enteric bacteria. The consensus sequences Ser-X_5-Gln and Asn-X_5-Gln indicated in yellow background were predicted to form β-strands (11). The identified gatekeeper residues of CsgA from *E. coli* are indicated with blue boxes. The putative gatekeeper residues of CsgA from other enteric bacteria are indicated with red boxes.

possibilities, the Asp residue of R12343 at position 149 was changed to Arg or Lys, and curli assembly was measured. Although His, Arg, and Lys have side chains with different properties, R12343^{D149R} and R12343^{D149K} could assemble into fibers in vivo similarly to R12343^{D149H} as detected by Congo red binding and Western blotting (Table S1), suggesting that at least His at position 149 is not absolutely required for fiber assembly. In addition, the Asn^{136} and His^{149} residues of CsgA are not conserved among CsgA homologs from different enteric bacteria (Fig. 2*E*). These observations indicate that the Asn^{136} and His^{149} residues are not essential for fiber formation. Therefore, it is plausible that the Asp residues at positions 136 and 149 in the molecule R12343 specifically impeded the amyloidogenic properties of this mutant protein and that the Asp residues prevented R3 peptide polymerization. To further explore this hypothesis, Asp residues were introduced into R5 of wild-type CsgA at positions 136 and 149. The resulting mutant CsgA$^{N136D/H149D}$ was stably secreted but defective in fiber assembly in vivo as detected by whole-cell Western analysis and TEM (Fig. S2 and Table S1). In addition, an R3 derivative peptide with these Asp residues changed to R1-like residues (D4N/D17L) was synthesized and found to polymerize significantly faster than the R3 peptide, supporting the idea that these Asp residues inhibit the intrinsic aggregation of R3 (Fig. S2). In summary, we designated gatekeeper residues in CsgA as those that negatively influence CsgA assembly and that substitution to other residues results in an amelioration of CsgA polymerization. Therefore, we designated the Asp residues at position 91 and 104 in R3 peptides as gatekeeper residues, because they prevented R3 from participating in critical interactions needed for polymerization.

Identification of Gatekeeper Residues in R2 and R4. Peptides R2 and R4 have poor amyloidogenic properties and cannot efficiently self-assemble into ThT-positive amyloid fibers in vitro (19). The CsgA mutant proteins R2345 (ΔR1) and R1234 (ΔR5) were defective in curli assembly, possibly because R2 and R4 cannot function like R1 or R5 (12). To test the hypothesis that R2 and R4 also contain gatekeeper residues, we compared R4 with R1/R5, looking for residues in R4 that were different from both R1 and R5. Four residues were targeted as candidates: Thr^{114}, Lys^{116}, Gly^{123}, and Asp^{127} in the molecule R1234 (Fig. S3*A*). When Gly^{123} or Asp^{127} were mutated to the corresponding residues from R5 (Asn and His, respectively), in vivo amyloidogenesis was mildly improved as detected by Congo red binding, Western analysis, and TEM (Fig. S3 and Table S1). However, when both Gly^{123} and Asp^{127} were mutated together, curli assembly was significantly improved, suggesting that Gly^{123} and Asp^{127} are gatekeeper residues in R4 (Fig.

S3 and Table S1). All R1234-derivative mutants were stably expressed as shown by Western analysis of the cells and underlying agar (Fig. S3). Similarly, Asp^{67}, Gly^{78}, Asp^{80}, and Gly^{82} (the positions of these amino acids are defined in wild-type CsgA) were targeted as potential gatekeeper residues in R2 in a molecule R2345 (Fig. S4). After Gly^{78}, Asp^{80} and Gly^{82} were mutated to the corresponding residues from R1 (Ser, Leu, and Leu, respectively). R2345 could efficiently assemble into fibers as detected by Western analysis, Congo red binding, and TEM, suggesting that Gly^{78}, Asp^{80}, and Gly^{82} comprise the gatekeeper residues of R2 (Fig. S4). Not all Asp residues in R2 function as gatekeeper residues. When Asp^{67} was mutated to Glu, the resulting mutant $R2345^{D67E}$ was as defective as R2345, indicating that Asp^{67} is not a gatekeeper residue (Fig. S4). To test whether or not these gatekeeper residues impede self-polymerization of R2 or R4, we tested self-polymerization of the R4-derivative peptide with mutations of gatekeeper residues ($R4^{G13N/D17H}$). When the gatekeeper residues were mutated to R5-like residues, $R4^{G13N/D17H}$ polymerized into amyloid fibers much more efficiently compared with the original R4 peptide as measured by ThT fluorescence Fig. S3). Therefore, the residues Gly^{78}, Asp^{80}, Gly^{82}, Gly^{123}, and Asp^{127} in R2 and R4 of wild-type CsgA were designated as gatekeeper residues.

CsgA Lacking All Identified Gatekeeper Residues Polymerized in the Absence of CsgB and CsgF In Vivo. The five repeating units of CsgA have the consensus sequence Ser-X_5-Gln-X_4-Asn-X_5-Gln. The X_5 regions between Ser/Gln or Asn/Gln were predicted to comprise of β-strands in CsgA (11). Seven gatekeeper residues (Gly^{78}, Asp^{80}, Gly^{82}, Asp^{91}, Asp^{104}, Gly^{123}, and Asp^{127} for *E. coli* CsgA) found in R2, R3, and R4 were located in these five-residue stretches (Fig. 2*E*). R1 and R5 of CsgA homologs from different enteric bacteria do not have any Asp or Gly residues in the same regions (Fig. 2*E*). The introduction of Asp residues into R5 of wild-type CsgA at positions of Asp residues in R3 resulted in defective fiber assembly (Fig. S2), suggesting that gatekeeper residues are not tolerated at the critical end repeats. All of the CsgA homologs from different enteric bacteria have putative gatekeeper residues in R2, R3, and R4, indicating that the gatekeeper residues may play an important role in CsgA assembly. To gain some insight into the roles of gatekeeper residues in bacterial amyloid propagation, we constructed a CsgA mutant lacking all gatekeeper residues called CsgA*. In the amino acid sequence of CsgA*, all of the gatekeeper residues were replaced by the corresponding residues found in R1 or R5 ($CsgA^{G78S/D80L/G82L/D91N/D104L/G123N/D127H}$) (Fig. 3*A*). CsgB and CsgF are required for efficient polymerization of wild-type CsgA in vivo (6, 8). Surprisingly, in the absence of CsgB and CsgF, CsgA*-expressing cells contained a deep Congo red staining pattern, which suggests that the fibers had been assembled (Fig. 3*B*). Western blotting of *csgFBA*/pCsgA cells and the underlying agar showed that wild-type CsgA remained SDS soluble in the absence of CsgB and CsgF (Fig. 3*C*, lanes 3 and 4), whereas CsgA* from *csgFBA*/pCsgA* cells was SDS insoluble (Fig. 3*C*, lanes 5 and 6). Many of the CsgA* fibers formed in *csgFBA* genetic background were mislocalized and were not tethered with cells as detected by TEM, whereas no fibers were produced by *csgFBA*/pCsgA cells (Fig. 3*D*). In both *csgBA* and *csgFB* genetic backgrounds, CsgA* formed SDS insoluble fibers and bound Congo red (Fig. 3*B*, Fig. S5, and Table S1). After cells were scraped off the plate, the underlying agar was visibly red, again suggesting that fibers composed of CsgA* were mislocalized (Fig. S6). Like CsgA, the stability of CsgA* was still dependent on CsgG and CsgE. In the absence of CsgG or CsgE, CsgA* was undetectable by Western blotting of cells and underlying agar and could not bind Congo red (Fig. 3*B* and Fig. S5). Collectively, secretion and stability of CsgA* is dependent on CsgG and CsgE. However, in contrast to wild-type CsgA, the polymerization of CsgA* in vivo does not require CsgB and CsgF. It is possible that these gatekeeper residues modulate CsgA fiber assembly by keeping the polymerization of CsgA dependent on CsgB and CsgF.

CsgA* Polymerized Faster than CsgA In Vitro and Induced Expression of CsgA* Caused *E. coli* to Be Less Viable. To understand the biochemical distinction between CsgA* and CsgA, we purified CsgA and CsgA* and compared their polymerization kinetics in vitro using a previously described method (19). Wild-type CsgA was found to be secreted to the medium by co-overexpressing CsgA and CsgG (19). Overexpression of CsgA* resulted in cell lysis, and no CsgA* was secreted to the medium or associated with cells, which suggests that CsgA* is potentially cytotoxic. We took an alternative approach to purify mature CsgA and CsgA*. A vector (pET11d) containing CsgA or CsgA* lacking the Sec-dependent periplasmic signal sequence was transformed into the expression strain NEB C2566. Protein expression was induced by

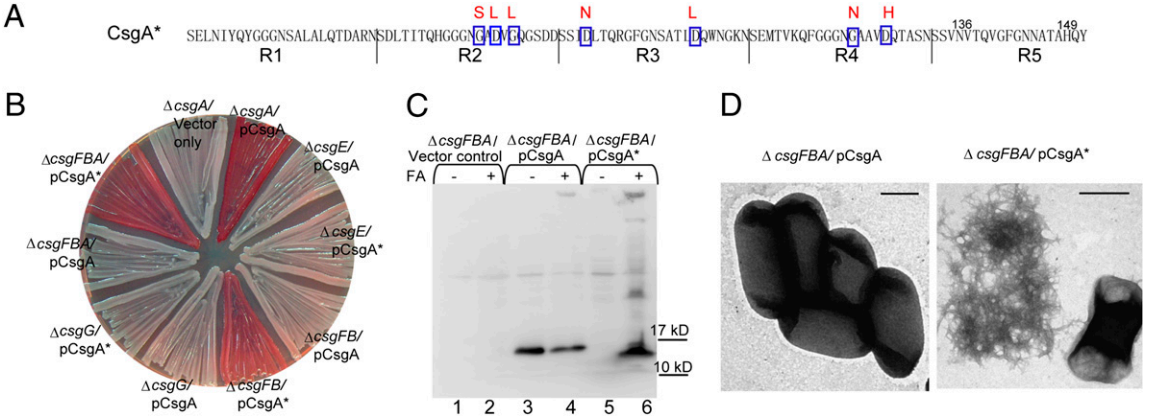

Fig. 3. In vivo polymerization of CsgA* is independent of nucleator CsgB and CsgF. (*A*) Seven gatekeeper residues were changed to R1/R5-like residues (indicated above the CsgA sequence) in CsgA*. (*B*) A Congo red-containing YESCA plate with *csgA*, *csgFBA*, *csgG*, *csgFB*, and *csgE* cells transformed with the plasmids encoding CsgA or CsgA*. Cells were grown 48 h at 26°C on YESCA plates. (*C*) Western blot of whole cells and underlying agar (agar plugs) from *csgFBA* cells containing control vector (lanes 1 and 2), pCsgA (lanes 3 and 4), or pCsgA* (lanes 5 and 6). Samples were treated with (+) or without (–) FA as indicated. The blots were probed with anti-CsgA antibody. The electrophoretic mobility of wild-type CsgA was reported to be slightly slower than its molecular weight (5). CsgA* migrated slightly faster than wild-type CsgA. Similarly, CsgA* purified from the cytosol also migrated faster than CsgA, suggesting that the aberrant migration may be caused by the intrinsic properties of CsgA* compared with wild-type CsgA. (*D*) Negative-stain EM micrographs of *csgFBA* cells containing plasmids pCsgA or pCsgA*. (Scale bar, 500 nm.)

Isopropyl β-D-1-thiogalactopyranoside (IPTG), and CsgA or CsgA* was purified from cell pellets that were resuspended in guanidine hydrochloride (GdnHCl; see *Materials and Methods*). Wild-type CsgA polymerization followed a triphasic process with lag, rapid growth, and stationary phases, similar to previous observations (19). The polymerization lag phase for CsgA at a concentration higher than 5 μM was around ~2 h as detected by the ThT assay (Fig. 4*A*) (19). In contrast to CsgA, CsgA* started polymerizing without a clear lag phase, and the total time interval for its ThT signal to reach stationary phase was 3 h compared with 7 h for wild-type CsgA (Fig. 4*A*). Like CsgA, CsgA* self-assembled aggregates were ordered amyloid-like fibers as evidenced by TEM (Fig. S7).

Although there was no obvious cytotoxicity when CsgA* was expressed under the *csgBA* native promoter, significant cytotoxicity was observed when CsgA* was expressed under the control of an inducible promoter (the *trc* promoter). To carefully compare the potential cytotoxicity of CsgA-like molecules, we measured cell viability when full-length CsgA, CsgA*, or CsgA^{slowgo} (CsgA$^{Q49A/N54A/Q139A/N144A}$) expression was induced. CsgA^{slowgo} was previously reported to be defective in curli assembly, and its self-polymerization had a lag time two orders of magnitude greater than that of CsgA polymerization (13). The viability of *csgA* cells (MC4100 background) did not change significantly with the expression of CsgA^{slowgo} (Fig. 4*B*). At relatively high IPTG concentration (>50 μM), expression of wild-type CsgA resulted in less viable cells compared with either CsgA^{slowgo} or empty vector control (Fig. 4*B*). The induced expression of CsgA* with IPTG at 25 μM caused a marked loss of cell viability compared with wild-type CsgA (Fig. 4*B*). The colony-forming units of cells expressing CsgA* were approximately three orders of magnitude lower than those of cells expressing CsgA when IPTG concentration was 50 μM or higher (Fig. 4*B*). Expression of CsgA* resulted in similar decreased viability of various *E. coli* K12 cells such as C600 and BW25113. These results suggest that the presence of gatekeeper residues of CsgA decreases potential cytotoxicity.

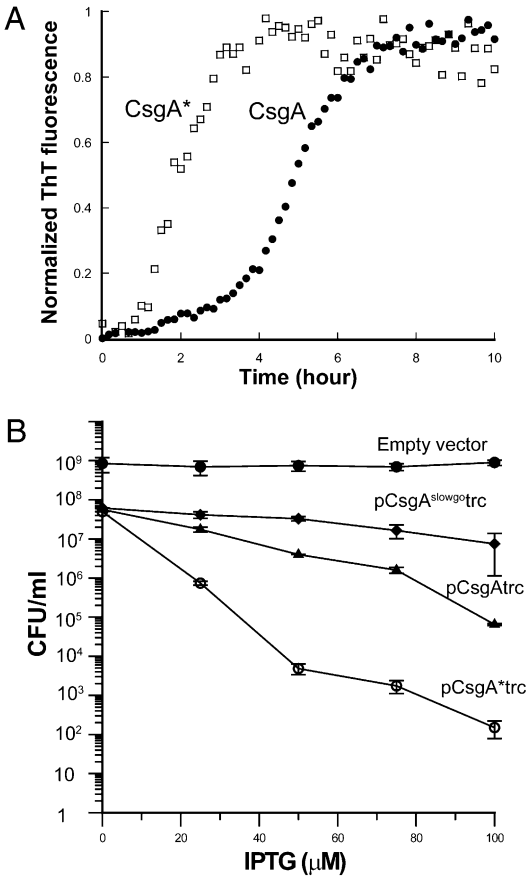

Fig. 4. In vitro polymerization of CsgA* is faster than that of CsgA, and induced expression of CsgA* causes cells less viable. (*A*) The polymerization of CsgA and CsgA* at 7 μM was measured by ThT fluorescence. (*B*) Normalized cell suspensions of *csgA* containing empty vector or plasmids encoding CsgA^{slowgo}, CsgA, or CsgA* under *trc* promoter were spotted on LB chloramphenicol plates in the presence of IPTG at various concentrations. Colony forming units of serial dilutions of cell suspensions were manually counted.

Discussion

It was postulated that charged amino acid residues and other β-breaker residues in globular proteins could function as gatekeepers to prevent amyloid formation (16, 17, 21). Results from several different studies suggest that Pro, Asp, and Gly are among the residues least likely to be incorporated into a β-sheet structure (22, 23). Data presented here identified specific residues (gatekeeper residues) that inhibited the amyloidogenic properties of internal CsgA repeats. Our work showed that the absence of gatekeeper residues resulted in uncontrolled aggregative properties and increased potential cytotoxicity. It is possible that the CsgA amino acid sequence had evolved to balance self-polymerization and still remained under the control of the curli nucleation proteins so that polymerization could occur faithfully at the correct time and place.

Roles of Gatekeeper Residues of CsgA in Curli Biogenesis. During the polymerization of disease-associated amyloids, the rate-limiting step is the formation of metastable protein oligomers or nuclei (18, 24). For curli assembly, amyloid nucleation is proposed to be facilitated by the CsgB protein, which adopts a β-sheet-rich secondary structure and propagates its β-sheet-rich amyloid fold onto CsgA molecules (25). This cross-nucleation mechanism separates nucleation and fiber elongation and controls curli assembly to occur at correct time and location (25). In addition, wild-type nucleation is also dependent of CsgF, an outer membrane-localized chaperone-like protein (10). CsgA polymerization in vivo is a concerted process and is under the control of the assembling factors. The gatekeeper residues of CsgA possibly provide a control mechanism for curli assembly. Self-polymerization of a CsgA mutant lacking all gatekeeper residues (CsgA*) is significantly faster than CsgA, and there is a limited lag phase before rapid polymerization (Fig. 4*A*). The strong aggregation propensity of CsgA* results in efficient polymerization in the extracellular environment even in the absence of CsgB and CsgF (Fig. 3 and Fig. S5). Many of the CsgA* fibers assembled in the absence of CsgB were not associated with cells and thereby, presumably may not help cells to attach to the surface. It is plausible that gatekeeper residues of CsgA modulate CsgA polymerization efficiency so that its polymerization and localization are under the control of CsgB and CsgF. In general, amyloid proteins, including functional amyloids, have less Pro and Gly residues than elastomeric proteins (26). In the *E. coli* proteome, 90% of 26,601 sequences predicted by TANGO (a computer algorithm for prediction of aggregating regions in unfolded polypeptide chains) to be aggregation-prone have at least one charged residue or proline at the first position on either side of the sequence, presumably to oppose aggregation (15, 16). Different gatekeeper residues seem to have varying influences on curli amyloidogenic properties. For instance, the amyloidogenic properties of R12343^{D149R} and R12343^{D149K} were significantly improved compared with R12343 (Table S1), although Asp, Arg, and Lys are all considered potential gatekeeper residues (15). Employing gatekeeper residues seems to be a common strategy to modulate protein aggregation.

Positional Effects of Gatekeeper Residues in CsgA. CsgA amyloid formation is guided by both positive and negative aggregation determinants. The positive determinants are the regularly spaced

Gln and Asn residues in each CsgA repeating unit that are conserved in CsgA homologs in different species. CsgA nucleation by CsgB requires the participation of specific conserved Gln and Asn residues in R1 and R5 of CsgA (13). The middle repeats (R2, R3, and R4) also have Gln and Asn residues at these conserved positions, yet these repeating units do not respond to CsgB nucleation (12). Using rationally designed mutagenesis, we elucidated negative aggregation determinants and found specific Asp and Gly residues of R2, R3, and R4 that prevented CsgB nucleation and self-assembly (Fig. 2 and Figs. S3 and Figs. S4). R2, R3, and R4 were converted to R1/R5 if these Asp and Gly residues were mutated. These Asp and Gly residues were only found in the middle repeats of CsgA homologs from different enteric bacteria, not the *N*-terminal or *C*-terminal repeating units (Fig. 2*E*), and our data also showed that gatekeeper residues in CsgA were not tolerated at end repeats (Fig. 2 and Figs. S2). These results suggest that the distribution of these gatekeeper residues in CsgA is not random and that the positions of *N*- and *C*-termini of CsgA are important for fiber assembly.

Potential Cytotoxicity of Functional Amyloids Was Reduced by Gatekeeper Residues. It has been previously reported that the efficiency of amyloid formation correlates to cytotoxicity and disease development in various models (24, 27–29). Here, different CsgA derivatives with distinct polymerization properties have distinguishable cytotoxicity, which supports this argument (Fig. 4*B*). The induced expression of rapidly polymerizing CsgA* was more cytotoxic compared with equivalent expression of wild-type CsgA (Fig. 4*B*). One possibility is that the existence of gatekeeper residues in CsgA decreases the load of aggregated proteins for cellular molecular chaperones and dedicated quality-control machinery. Alternatively, gain of toxic properties of CsgA* may be distinct from wild-type CsgA. Functional amyloid propagation is a balance to maximize polymerization efficiency while minimizing potential cytotoxicity associated with amyloidogenesis.

Materials and Methods

Bacterial Growth. Bacteria were grown at 26°C on yeast extract casamino acids (YESCA) plates to induce curli production (6). When needed, antibiotics were added to the plates in the following concentrations: kanamycin, 50 μg/mL; chloramphenicol, 25 μg/mL; and ampicillin, 100 μg/mL. Curli production was monitored using Congo red YESCA plates (6).

Bacterial Strains and Plasmids. Curli assembly phenotypes of different strains are summarized in Table S1. Strains, plasmids, and primer sequences used in this study are listed in Table S2 and S3. The details of plasmid construction are included in *SI Materials and Methods*.

Purification of CsgA* and CsgA. C-terminal His_6 tagged CsgA* and CsgA without the Sec periplasmic signal sequence were purified from cell pellets. The details of the purification procedure are included in *SI Materials and Methods*.

Preparation of Peptides. Peptides R3 and R4 and derivative peptides $R3^{D4N/D17L}$ and $R4^{G13N/D17H}$ were synthesized by Proteintech Group Inc. Peptides were prepared similarly as previously described (see *SI Materials and Methods*) (12).

ThT Assay. ThT assays were performed and normalized as previously described (12). Seeds were prepared as previously described (see *SI Materials and Methods*) (19).

TEM. A Philips CM12 transmission electron microscope was used to visualize the cells and fiber aggregates. Samples were placed on formvar-coated copper grids (Ernest F. Fullam, Inc.) for 2 min, washed with deionized water two times, and stained with 2% uranyl acetate for 90 s.

SDS/PAGE and Western Blotting. Bacteria whole-cell lysates or cells and underlying agar (plugs) were prepared and probed for CsgA using previously described methods (6). The Western blotting of whole-cell lysate detects the proteins associated with cells, whereas Western blotting of plugs detects both proteins associated with cells and those secreted to the agar plate.

Overlay Assay. Purified proteins were spotted on the *csgA* cells and stained with 0.5 mg/mL Congo red solution as previously described (12).

Spotting Assay for Cell Viability. Freshly transformed cells with plasmids pCsgAtrc, empty vector pNH3, $pCsgA^{slowgo}$trc, or pCsgA*trc were scraped off the plates and normalized by absorbance at 600 nm. Cells with serial dilutions were spotted on LB chloramphenicol plates with various concentrations of IPTG to induce expression of CsgA, $CsgA^{slowgo}$, or CsgA*. After overnight growth, the colony number was manually counted, and colony forming units per milliliter of cell suspension under different IPTG concentrations were calculated.

ACKNOWLEDGMENTS. We thank the members of the Chapman Laboratory for helpful discussions and review of this manuscript. This work was supported by the National Institutes of Health Grant AI073847-02.

1. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75:333–366.
2. Fowler DM, Koulov AV, Balch WE, Kelly JW (2007) Functional amyloid—from bacteria to humans. *Trends Biochem Sci* 32:217–224.
3. Hammer ND, Wang X, McGuffie BA, Chapman MR (2008) Amyloids: Friend or foe? *J Alzheimers Dis* 13:407–419.
4. Barnhart MM, Chapman MR (2006) Curli biogenesis and function. *Annu Rev Microbiol* 60:131–147.
5. Shewmaker F, et al. (2009) The functional curli amyloid is not based on in-register parallel beta-sheet structure. *J Biol Chem* 284:25065–25076.
6. Chapman MR, et al. (2002) Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* 295:851–855.
7. Collinson SK, et al. (1993) Thin, aggregative fimbriae mediate binding of *Salmonella enteritidis* to fibronectin. *J Bacteriol* 175:12–18.
8. Hammar M, Bian Z, Normark S (1996) Nucleator-dependent intercellular assembly of adhesive curli organelles in *Escherichia coli*. *Proc Natl Acad Sci USA* 93:6562–6566.
9. Robinson LS, Ashman EM, Hultgren SJ, Chapman MR (2006) Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. *Mol Microbiol* 59:870–881.
10. Nenninger AA, Robinson LS, Hultgren SJ (2009) Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. *Proc Natl Acad Sci USA* 106:900–905.
11. Collinson SK, Parker JM, Hodges RS, Kay WW (1999) Structural predictions of AgfA, the insoluble fimbrial subunit of *Salmonella* thin aggregative fimbriae. *J Mol Biol* 290:741–756.
12. Wang X, Hammer ND, Chapman MR (2008) The molecular basis of functional bacterial amyloid polymerization and nucleation. *J Biol Chem* 283:21530–21539.
13. Wang X, Chapman MR (2008) Sequence determinants of bacterial amyloid formation. *J Mol Biol* 380:570–580.
14. Richardson JS, Richardson DC (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci USA* 99:2754–2759.
15. Rousseau F, Serrano L, Schymkowitz JW (2006) How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol* 355:1037–1047.
16. Monsellier E, Chiti F (2007) Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep* 8:737–742.
17. Otzen DE, Kristensen O, Oliveberg M (2000) Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: A structural clue to amyloid assembly. *Proc Natl Acad Sci USA* 97:9907–9912.
18. Jarrett JT, Lansbury PT, Jr. (1993) Seeding "one-dimensional crystallization" of amyloid: A pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* 73:1055–1058.
19. Wang X, Smith DR, Jones JW, Chapman MR (2007) In vitro polymerization of a functional *Escherichia coli* amyloid protein. *J Biol Chem* 282:3713–3719.
20. Collinson SK, Emödy L, Müller KH, Trust TJ, Kay WW (1991) Purification and characterization of thin, aggregative fimbriae from *Salmonella enteritidis*. *J Bacteriol* 173:4773–4781.
21. Thirumalai D, Klimov DK, Dima RI (2003) Emerging ideas on the molecular basis of protein and peptide aggregation. *Curr Opin Struct Biol* 13:146–159.
22. Kallberg Y, Gustafsson M, Persson B, Thyberg J, Johansson J (2001) Prediction of amyloid fibril-forming proteins. *J Biol Chem* 276:12945–12950.
23. Street AG, Mayo SL (1999) Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc Natl Acad Sci USA* 96: 9074–9076.
24. Selkoe DJ (2003) Folding proteins in fatal ways. *Nature* 426:900–904.
25. Hammer ND, Schmidt JC, Chapman MR (2007) The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. *Proc Natl Acad Sci USA* 104:12494–12499.
26. Rauscher S, Baud S, Miao M, Keeley FW, Pomès R (2006) Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure* 14:1667–1676.
27. Luheshi LM, et al. (2007) Systematic in vivo analysis of the intrinsic determinants of amyloid Beta pathogenicity. *PLoS Biol* 5:e290.
28. Van Nostrand WE, Melchor JP, Cho HS, Greenberg SM, Rebeck GW (2001) Pathogenic effects of D23N Iowa mutant amyloid beta -protein. *J Biol Chem* 276:32860–32866.
29. Hardy J, Selkoe DJ (2002) The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* 297:353–356.