doi:10.1016/j.jmb.2008.05.019 J. Mol. Biol. (2008) 380, 570–580

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Sequence Determinants of Bacterial Amyloid Formation

Xuan Wang and Matthew R. Chapman*

Department of Molecular, Cellular and Developmental Biology, University of Michigan, 830 N. University, Ann Arbor, MI 48109, USA

Received 9 March 2008; received in revised form 30 April 2008; accepted 11 May 2008
Available online 17 May 2008

Amyloids are proteinaceous fibers commonly associated with neurodegenerative diseases and prion-based encephalopathies. Many different polypeptides can form amyloid fibers, leading to the suggestion that amyloid is a primitive main chain-dominated structure. A growing body of evidence suggests that amino acid side chains dramatically influence amyloid formation. The specific role fulfilled by side chains in amyloid formation, especially *in vivo*, remains poorly understood. Here, we determined the role of internally conserved polar and aromatic residues in promoting amyloidogenesis of the functional amyloid protein CsgA, which is the major protein component of curli fibers assembled by enteric bacteria such as *Escherichia coli* and *Salmonella* spp. *In vivo* CsgA polymerization into an amyloid fiber requires the CsgB nucleator protein. The CsgA amyloid core region is composed of five repeating units, defined by regularly spaced Ser, Gln and Asn residues. The results of a comprehensive alanine scan mutagenesis screen showed that Gln and Asn residues at positions 49, 54, 139 and 144 were critical for curli assembly. Alanine substitution of Q49 or N144 impeded the ability of CsgA to respond to CsgB-mediated heteronucleation, and the ability of CsgA to self-polymerize *in vitro*. However, CsgA proteins harboring these mutations were still seeded by preformed wild-type CsgA fibers *in vitro*. This suggests that CsgA-fibril-mediated seeding and CsgB-mediated heteronucleation have distinguishable mechanisms. Remarkably, Gln residues at positions 49 and 139 could not be replaced by Asn residues without interfering with curli assembly, suggesting that the side chain requirements were especially stringent at these positions. This analysis demonstrates that bacterial amyloid formation is driven by specific side chain contacts, and provides a clear illustration of the essential roles of specific side chains in promoting amyloid formation.

Keywords: amyloid; nucleation; polymerization; curli; asparagine and glutamine residues

Edited by S. Radford

Introduction

Amyloid formation is readily associated with neurodegenerative diseases and prion-based encephalopathies.[1] Amyloid fibers are 4–10 nm wide, unbranched proteinaceous filaments.[1] Amyloid fibers possess a characteristic cross-β sheet quaternary structure, in which β strands are perpendicular to the fibril axis.[1] These fiber structures have specific tinctorial properties, binding to dyes such as Congo red and thioflavin T (ThT).[1] Amyloid toxicity is complex, but a growing body of work suggests that pre-fiber aggregates are cytotoxic, while mature fibers are relatively benign.[2] Therefore, understanding the folding cascades involved in amyloid formation is necessary for developing new therapies to target these processes.

A newly described class of "functional" amyloids suggest that amyloid formation can be an integral part of normal cellular physiology.[3,4] Functional amyloids provide a unique perspective on amyloidogenesis because the cell must have co-evolved mechanisms to prevent the toxicity normally associated with amyloid formation. One of the best understood functional amyloids is curli, a bacterially produced extracellular fiber required for biofilm formation and other community behaviors.[5] In *Escherichia coli*, the polymerization of the major

**Corresponding author.* E-mail address: chapmanm@umich.edu.

Abbreviations used: ThT, thioflavin T; FA, formic acid; TEM, transmission electron microscopy.

curli fiber subunit protein CsgA into an amyloid fiber is dependent on the minor curli subunit protein, CsgB.[6] CsgA remains soluble until it encounters outer membrane-localized CsgB,[5] which has been demonstrated to have amyloid-forming properties and apparently serves as a template for CsgA polymerization.[7]

Earlier, we showed that, like many other amyloids, preformed CsgA fibers could seed soluble CsgA polymerization *in vitro*.[8] Therefore, we proposed that the growing CsgA fiber on the cell surface could serves as a template promoting soluble CsgA polymerization in a process akin to seeding.[8] The molecular details of CsgA fiber-mediated seeding and CsgB-mediated heteronucleation are poorly described. Because nucleation underlies the rate-limiting step of amyloid propagation, understanding the nature of this mechanism will shed light on how cells control amyloid formation.

The primary sequence of CsgA can be divided into three functional domains: an N-terminal Sec signal sequence (cleaved after translocation into the periplasmic space); an N-terminal 22 amino acid segment (N22) that directs CsgA secretion across the outer membrane;[9] and an amyloid core region (residues 43 – 151) that contains five imperfect repeating units, each 19–23 amino acid residues in length (Fig. 1a).[5] The five repeating units form a protease-resistant structure,[10] which is proposed to be the amyloid core of CsgA.[8] These repeats are distinguished by the consensus sequence Ser-X5-Gln-X4-Asn-X5-Gln and are linked by four or five residues.[10] These Ser, Gln and Asn residues are conserved among CsgA homologs of many enteric bacteria.[8] This high degree of amino acid sequence conservation suggests that these residues may have an important role in curli assembly.

Many polypeptides, if not all, can assemble into an amyloid fiber *in vitro*, even though amyloid-forming proteins do not necessarily share amino acid similarities.[1] Therefore, it has been proposed that amyloid formation is an inherent property of polypeptide main chains.[1] However, specific residues likely have a role in promoting both disease-associated and functional amyloid formation. Yeast prion protein Sup35p has a Gln/Asn-rich domain at the N-terminus that has been implicated in prion propagation.[11–13] Moreover, the specific sequences in this Gln/Asn-rich domain govern self-recognition and species-specific seeding activity.[14] Aromatic residues in the islet amyloid polypeptide fragment contribute positively to its polymerization into amyloid fibers *in vitro*.[15] Structural analysis of Aβ40 and Aβ42 revealed that two β sheets form a parallel β sandwich stabilized by specific intermolecular side chain contacts and changes of these side chains resulted in defective fiber assembly.[16,17] Therefore, it is clear that side chains can influence amyloid polymerization, but the contribution of side chains to *in vivo* amyloidogenesis and the exact roles of amino acid side chain contacts are poorly understood. Here, we performed a comprehensive mutagenesis study on CsgA and identified the residues that promote CsgA amyloidogenesis. We showed that CsgA amyloidogenesis is driven by the side chain contacts of four Gln and Asn residues in N- and C-terminal repeats. These Gln and Asn residues have essential roles in the response to CsgB-mediated heteronucleation and the initiation of efficient self-assembly *in vitro*. Strikingly, these four Gln and Asn residues are not required for CsgA seeding, suggesting CsgA seeding and CsgB-mediated heteronucleation have distinct mechanisms.

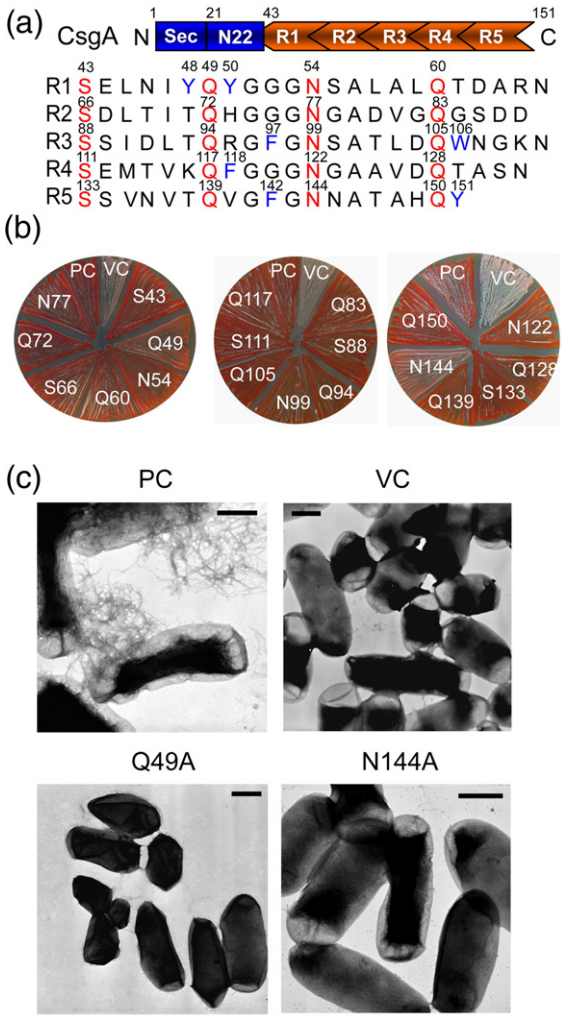

Fig. 1. Effect of Ala substitutions of conserved Ser, Gln and Asn of CsgA on curli assembly. a, Schematic of CsgA including an N-terminal Sec signal sequence and the N-terminal 22 residues that precede the five repeating units. The regularly spaced Ser, Gln and Asn residues are indicated in red and the aromatic residues are indicated in blue. The position of highlighted residues is indicated above the residue. b, CR-YESCA plate with *csgA* cells transformed with the plasmids encoding CsgA with Ala substitution after 48 h of growth at 26 °C. *csgA*/pLR5 and *csgA*/pLR2 were used as the positive control (PC) and vector control (VC), respectively. The specific Ala substitution is indicated. c, Negative-stain TEM micrographs of *csgA* cells containing vector control (pLR2) or plasmids encoding wild-type CsgA (pLR5), $CsgA^{Q49A}$ and $CsgA^{N144A}$. Scale bars represent 500 nm.

Results

Ala scan mutagenesis of internally conserved polar residues

The amyloid core of CsgA is composed of five repeating units, defined by internally conserved and regularly spaced Ser, Gln and Asn residues that are conserved among many enteric bacteria (Fig. 1a).[8] We performed an Ala scan mutagenesis on these 20 polar residues to test their importance in directing bacterial amyloid formation. Amyloid formation of each CsgA mutant was assessed initially by growing bacteria on plates amended with Congo red, as cells expressing wild-type CsgA will stain deep red on this medium (Fig. 1b). Only two mutations, Q49A and N144A, reduced Congo red binding significantly relative to wild-type CsgA (Fig. 1b). Transmission electron microscopy (TEM) was used to validate the Cong red binding phenotypes. *csgA* cells (LSR10) transformed with pLR5 (encoding CsgA) produced curli fibers that were indistinguishable from those assembled by wild-type strain MC4100 by TEM. Cells expressing $CsgA^{Q49A}$ or $CsgA^{N144A}$ assembled fewer fibers than cells expressing wild-type CsgA, as observed by TEM (Fig. 1c).

CsgA polymerization into an amyloid fiber *in vivo* can also be monitored by its ability to migrate as a monomer on SDS-PAGE after dissociation by a strong acid, formic acid (FA).[18] For example, CsgA produced by wild-type cells is whole cell-associated and SDS-insoluble.[19] Brief treatment with FA liberates CsgA monomers from curli fibers produced by wild-type strain MC4100.[6] Similar to the wild-type strain, CsgA produced by *csgA*/pLR5 was SDS-insoluble and required brief pretreatment with FA to enter the gel (Fig. 2a, lanes 1 and 2). All of the Ser mutants (S43A, S66A, S88A, S111A and S133A) were also SDS-insoluble and associated with the whole-cell fraction (Fig. 2a, lanes 5, 6, 13, 14, 21, 22, 29, 30, 37 and 38). However, 13 of 15 Ala substitution mutants of Gln/Asn residues (Q49A, N54A, Q60A, Q72A, N77A, Q94A, N99A, Q105A, Q117A, N122A, Q128A, Q139A and N144A) showed different levels of SDS solubility by whole-cell Western analysis, suggesting these polar residues help stabilize the amyloid structure (Fig. 2a). $CsgA^{Q49A}$ and $CsgA^{N144A}$ were unable to complement Congo red binding to a *csgA* mutant (Fig. 1b), and very little of these mutant proteins could be recovered from whole-cell lysates scraped off YESCA plates (Fig. 2a, lanes 7, 8, 41 and 42).

To test the possibility that $CsgA^{Q49A}$ and $CsgA^{N144A}$ were secreted away from the cell as soluble proteins, cells and the underlying agar were collected and analyzed by Western blotting. In these samples, called plugs, both $CsgA^{Q49A}$ and $CsgA^{N144A}$ were readily detected and SDS-soluble, demonstrating that $CsgA^{Q49A}$ and $CsgA^{N144A}$ were stable, secreted to the cell surface and unpolymerized (Fig. 2b, lanes 2, 3, 11 and 12). $CsgA^{N54A}$ and $CsgA^{Q139A}$ were also significantly different from other mutants in the whole-cell SDS-solubility assays. $CsgA^{N54A}$ was completely SDS-soluble (Fig. 2a, lanes 9 and 10) and $CsgA^{Q139A}$ was not predominately cell-associated (Fig. 2a, lanes 39 and 40). $CsgA^{N54A}$ and $CsgA^{Q139A}$ were SDS-soluble, as detected by Western analysis of cells and the underlying agar (Fig. 2b, lanes 4, 5, 8 and 9), suggesting $CsgA^{N54A}$ and $CsgA^{Q139A}$ were not assembled into wild-type-like fibers *in vivo*. Collectively, Q49A and N144A were the most defective mutants in curli formation of the 20 mutants examined. In addition, the N54A and Q139A mutants were significantly defective in curli assembly as measured by Western analysis.

Ala scan mutagenesis of the aromatic residues

It was reported that aromatic residues may have an important role during amyloidogenesis.[15,20] We tested the contribution of aromatic residues in the CsgA amyloid core region by Ala scan mutagenesis. Except for $CsgA^{Y151A}$, Ala substitutions in the aromatic residues resulted in proteins that were phenotypically identical with wild-type CsgA as detected by Congo red binding and whole-cell Western analysis (Fig. 3a and b). $CsgA^{Y151A}$ was undetectable by Western analysis of cells and the underlying agar, indicating that Y151A was unstable

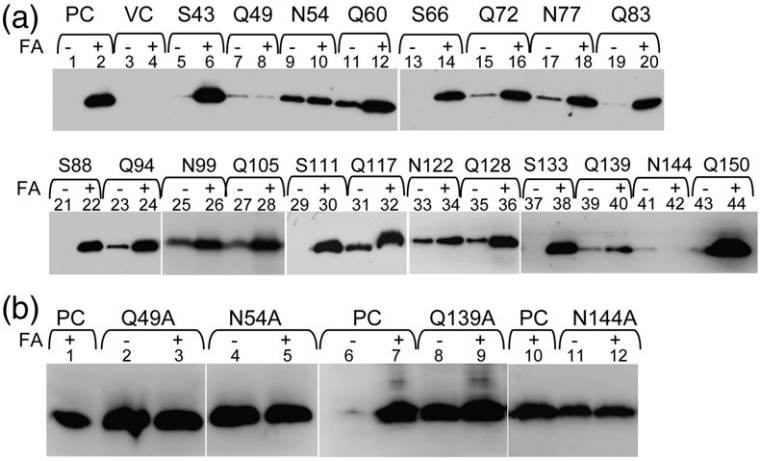

Fig. 2. Western analysis of CsgA mutants with Ala substitutions of internally conserved Ser, Gln and Asn. a, Western blots of whole-cell lysates from *csgA* cells containing vector control or plasmids encoding wild-type CsgA or indicated Ala substituted mutants. b, Western blot of whole cells and underlying agar (agar plugs) from *csgA* strains containing plasmids encoding CsgA, $CsgA^{Q49A}$, $CsgA^{N54A}$, $CsgA^{Q139A}$ or $CsgA^{N144A}$. Samples were treated with (+) or without (–) FA. The blots were probed with anti-CsgA antibody.

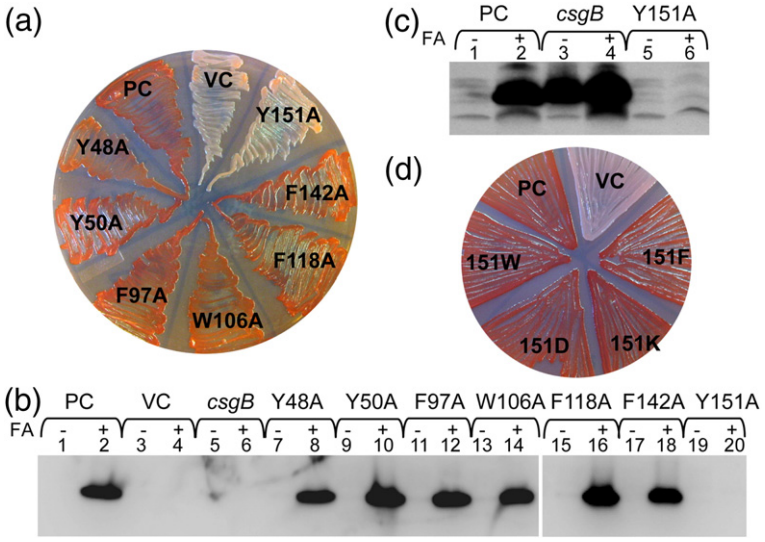

Fig. 3. Ala scan mutagenesis of aromatic residues in CsgA amyloid core regions. a, CR-YESCA plate with *csgA* cells transformed with the plasmids encoding CsgA, $CsgA^{Y48A}$, $CsgA^{Y50A}$, $CsgA^{F97A}$, $CsgA^{W106A}$, $CsgA^{F118A}$, $CsgA^{F142A}$ or $CsgA^{Y151A}$. b, Western blots of whole-cell lysates from *csgB* cells and *csgA* cells containing plasmids encoding wild-type CsgA or the indicated Ala mutants. c, Western blot of whole cells and underlying agar (agar plugs) from *csgB* cells and *csgA* strains containing plasmids encoding CsgA or $CsgA^{Y151A}$. d, The CR-YESCA plate with *csgA* cells transformed with the plasmids encoding CsgA, $CsgA^{Y151F}$, $CsgA^{Y151W}$, $CsgA^{F151D}$ or $CsgA^{Y151K}$.

(Fig. 3c, lanes 5 and 6). When Tyr^{151} was changed to Phe, Trp, Lys or Asp, curli formation was restored as shown by Congo red binding and whole-cell Western analysis, suggesting Tyr residue at position 151 is not absolutely required (Fig. 3d and data not shown).

$CsgA^{Q49A}$ and $CsgA^{N144A}$ are defective in self-polymerization *in vitro*

To further characterize the most defective CsgA mutant proteins, we purified $CsgA^{Q49A}$ and $CsgA^{N144A}$, and compared their polymerization to wild-type CsgA using the ThT assay as described.[8] Wild-type CsgA assembles into an amyloid fiber *in vitro* at concentrations above 2.0 μM in the absence of CsgB.[8] Two parameters were used to compare the polymerization kinetics of CsgA and its mutant analogues. The first kinetic parameter was the time preceding rapid fiber growth, called the lag phase or T_0. The second parameter was the time encompassing the fiber growth phase from initiation of rapid polymerization to its completion, called the conversion time (T_c).[11] At a concentration of 40 μM, the T_0 of $CsgA^{Q49A}$ was similar to that of CsgA, while T_c was much greater than that of CsgA (Fig. 4a). $CsgA^{N144A}$ polymerization had much greater T_0 and T_c than those of CsgA, suggesting the amido group of Asn at position 144 is critical for aggregation (Fig. 4a). After 120 h, both $CsgA^{Q49A}$ and $CsgA^{N144A}$ had assembled into amyloid fibers with fiber morphology similar to that of wild-type CsgA fibers (Fig. 4b, c and d).

$CsgA^{Q49A}$ and $CsgA^{N144A}$ are defective in heteronucleation response

Even though $CsgA^{Q49A}$ and $CsgA^{N144A}$ were defective in self-polymerization, in the presence of wild-type CsgA seeds they polymerized with efficiency similar to that of wild-type CsgA (data not shown). Two different approaches were employed to test of the ability of $CsgA^{Q49A}$ and $CsgA^{N144A}$ to respond to CsgB-mediated heteronucleation. The first was an overlay assay using freshly purified CsgA or CsgA mutant proteins and cells expressing the CsgB nucleator protein.[21] In a CsgB-dependent manner, soluble wild-type CsgA was converted into an amyloid fiber within 1 min of the overlay as evidenced by Congo red staining, or by TEM.[21] Various concentrations of CsgA, $CsgA^{Q49A}$ and $CsgA^{N144A}$ were overlaid on *csgA* cells ($CsgB^+$). At a concentration of 10 μM or higher, CsgA polymerized into an amyloid fiber in a CsgB-dependent fashion as detected by Congo red staining (Fig. 5a). However, 10 μM $CsgA^{Q49A}$ and $CsgA^{N144A}$ did not polymerize into amyloid-like fibers under the same conditions (Fig. 5a). When the concentration of $CsgA^{Q49A}$ and $CsgA^{N144A}$ was increased to 40 μM, a Congo red-binding structure on the surface of CsgB-expressing cells was detected (Fig. 5a).

Curli fibers are proposed to assemble after CsgA is secreted to the extracellular space. This can be illustrated by interbacterial complementation where a *csgA* mutant strain ($CsgB^+$) is grown in close proximity to a *csgB* mutant strain ($CsgA^+$). CsgA molecules secreted by *csgB* donor cells polymerize into amyloid fibers on the *csgA* acceptor cells, as shown in Fig. 5b.[6] We performed an interbacterial complementation assay to test CsgB-mediated heteronucleation responsiveness of CsgA, $CsgA^{Q49A}$ and $CsgA^{N144A}$ *in vivo*. CsgA secreted from *csgBA*/pLR5 (donor) responded to the heteronucleation of CsgB on the *csgA* cells (acceptor), and formed an amyloid fiber on the surface of *csgA* cells as demonstrated by staining with Congo red (Fig. 5b).[6] There was reduced Congo red binding on the surface of *csgA* cells when *csgBA* cells containing plasmid encoding $CsgA^{Q49A}$ or $CsgA^{N144A}$ were grown adjacent to *csgA* cells (Fig. 5b). The protein levels of CsgA, $CsgA^{Q49A}$ and $CsgA^{N144A}$ produced in the *csgBA* genetic background were similar as detected by Western analysis of cells and the underlying agar (data not shown). These data demonstrate collectively that $CsgA^{Q49A}$ and $CsgA^{N144A}$ were defective in responding to the heterogeneous nucleator CsgB.

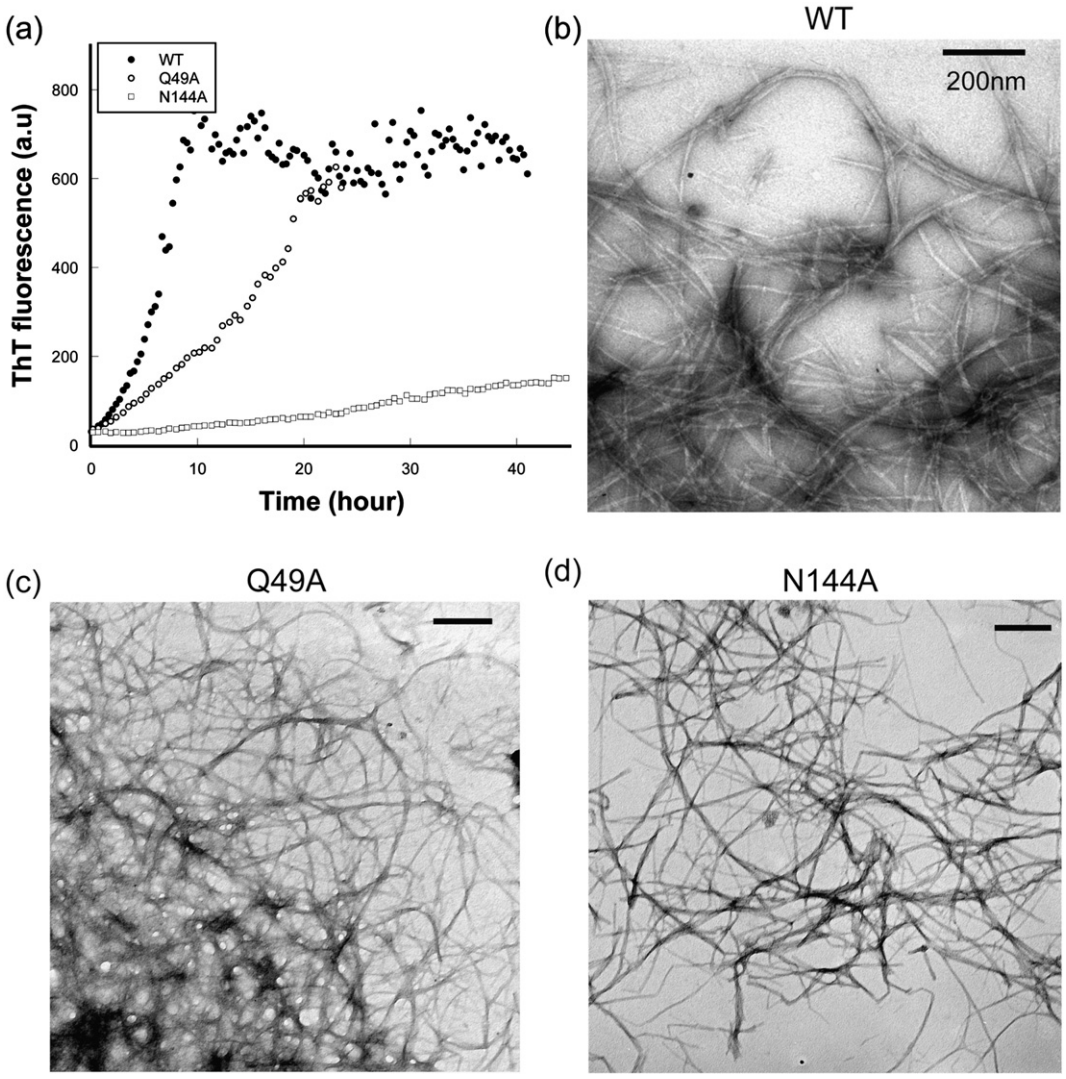

Fig. 4. *In vitro* self-polymerization of $CsgA^{Q49A}$ and $CsgA^{N144A}$ are defective. a, The fluorescence of 40 μM CsgA (filled circle), $CsgA^{Q49A}$ (open circle) and $CsgA^{N144A}$ (open square) mixed with 20 μM ThT was measured in 10 min intervals at 495 nm after excitation at 438 nm. *a.u.*, arbitrary units. b–d, Negative-stain EM micrographs of *in vitro* self-polymerized fibers of CsgA (b), $CsgA^{Q49A}$ (c) and $CsgA^{N144A}$ (d). Scale bars represent 200 nm.

Ala substitutions of Q49, N54, Q139 and N144 result in the dramatic loss of heteronucleation response to CsgB

In order to abolish the four critical Gln/Asn residues identified by our Ala scan mutagenesis, we constructed the mutant Q49A/N54A/Q139A/N144A. *In vivo*, $CsgA^{Q49A/N54A/Q139A/N144A}$ ($CsgA^{slowgo}$) completely lost the ability to assemble into amyloid fibers as detected by Congo red binding, TEM and whole-cell Western analysis (data not shown). We tested the competence of $CsgA^{slowgo}$ to respond to CsgB-mediated heteronucleation by overlay assay. Purified homogeneous $CsgA^{slowgo}$ could not be promoted to form amyloid fibers by cell-associated CsgB as shown by the lack of Congo red binding in the overlay assay, even at a concentration of 100 μM (Fig. 6a). There was no Congo red binding on the surface of *csgA* cells when *csgBA* cells containing plasmid encoding $CsgA^{slowgo}$ were grown adjacent to *csgA* cells, as shown in Fig. 6b, suggesting that $CsgA^{slowgo}$ completely lost the response to CsgB-mediated heteronucleation in the interbacterial complementation. The protein level of $CsgA^{slowgo}$ was similar to that of wild-type CsgA in this genetic background (data not shown). Collectively, these results demonstrate that $CsgA^{slowgo}$ remains unpolymerized even in the presence of the heteronucleator CsgB, suggesting these four polar residues are required in CsgB-mediated heteronucleation.

Q49, N54, Q139 and N144 are required for efficient self-polymerization and $CsgB_{trunc}$ cross-seeding but not for CsgA seeding

$CsgA^{slowgo}$ was purified and its polymerization kinetics were analyzed *in vitro*. The T_0 and T_c of $CsgA^{slowgo}$ were much higher than those of CsgA (Fig. 7a), suggesting that Q49, N54, Q139 and N144 have critical roles in self-polymerization *in vitro*. Self-

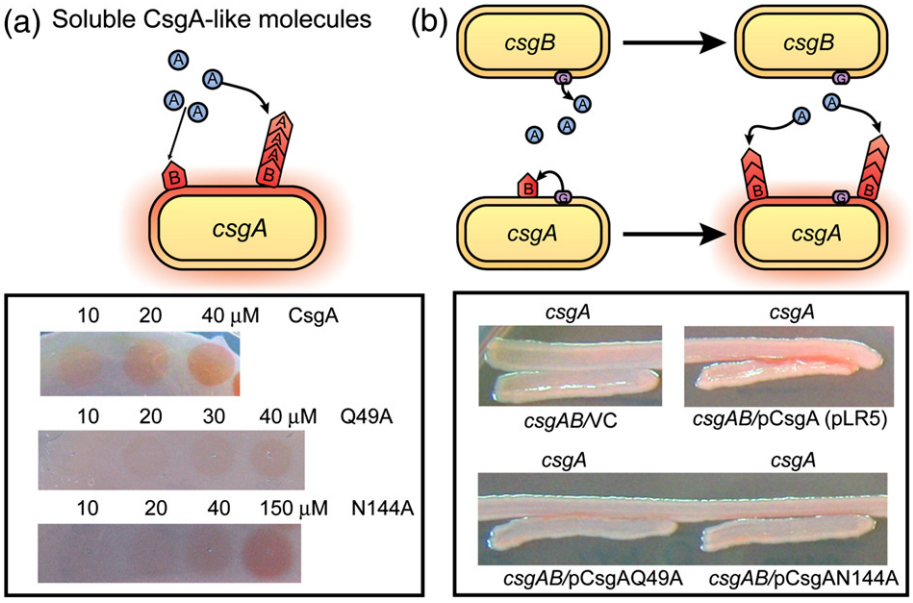

Fig. 5. $CsgA^{Q49A}$ and $CsgA^{N144A}$ are defective in response to CsgB-mediated heteronucleation. a, Schematic of the overlay assay in which soluble CsgA is spotted onto cells (*csgA*) expressing the nucleator protein, CsgB, where CsgA polymerization is catalyzed by CsgB (top). Freshly purified CsgA, $CsgA^{Q49A}$ or $CsgA^{N144A}$ at the concentrations indicated was spotted onto *csgA* cells. After incubation for 10 min, cells were stained with 0.5 mg/ml Congo red solution and washed with potassium phosphate buffer (bottom). b, Schematic of the interbacterial complementation assay in which soluble CsgA molecules secreted from *csgB* cells interact with CsgB on the surface of *csgA* cells to form fibers (top). *csgBA* cells containing plasmids encoding CsgA, $CsgA^{Q49A}$ or $CsgA^{N144A}$ were grown adjacently to *csgA* cells on a CR-YESCA plate for 48 h at 26 °C (bottom).

assembled $CsgA^{slowgo}$ fibers were similar to CsgA fibers, as shown by TEM (Fig. 7b). Interestingly, wild-type CsgA seeds efficiently eliminated the lag phase of the polymerization of $CsgA^{slowgo}$ as detected by ThT assay (Fig. 7c). In the presence of CsgA seeds, the polymerization rate of $CsgA^{slowgo}$ is very similar to that of wild-type CsgA,[8] and the ThT signal reached the stationary phase within 5 h (Fig. 7c). Polymerized

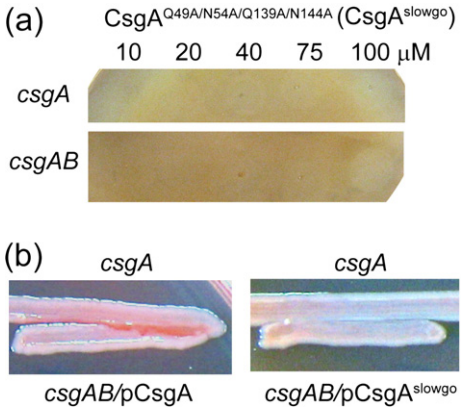

Fig. 6. The heteronucleation responsiveness of $CsgA^{slowgo}$ ($CsgA^{Q49A/N54A/Q139A/N144A}$). (a) Freshly purified homogeneous $CsgA^{slowgo}$ at the concentrations indicated was spotted onto *csgA* or *csgBA* cells. After incubation for 10 min, cells were stained with 0.5 mg/ml Congo red solution and washed with potassium phosphate buffer. b, *csgBA* cells containing plasmids encoding CsgA or $CsgA^{slowgo}$ were grown adjacently to *csgA* cells on a CR-YESCA plate for 48 h at 26 °C.

$CsgA^{slowgo}$ fibers promoted by wild-type CsgA appeared similar to CsgA fibers when examined by TEM (Fig. 7d). A nucleation-competent CsgB truncation mutant (called $CsgB_{trunc}$) was purified as described.[7] $CsgB_{trunc}$ has been demonstrated to seed the polymerization of wild-type CsgA efficiently.[7] Unlike CsgA seeds, $CsgB_{trunc}$ seeds did not promote the polymerization of $CsgA^{slowgo}$ efficiently under the conditions shown in Fig. 7c. Self-assembled $CsgA^{slowgo}$ fibers were able to seed wild-type CsgA polymerization, suggesting $CsgA^{slowgo}$ fibers are similar to wild-type CsgA fibers in terms of seeding specificity (data not shown).

Conservative Q/N substitutions

It was proposed that *in vitro* amyloid formation is strongly influenced by simple physicochemical factors of polypeptides, such as hydrophobicity, charge and secondary structure propensity.[1,22] Ala substitutions at positions 49, 54, 139 and 144 decreased CsgA polymerization *in vivo*. To test the stringency of the amino acid present at these critical positions, we constructed four mutants (Q49N, N54Q, Q139N and N144Q) that contained relatively conservative amino acid changes. Gln and Asn have similar physicochemical features, including the same amido group. Surprisingly, even though the side chain of Asn is only one carbon atom shorter than that of Gln, $CsgA^{Q49N}$ had dramatically reduced Congo red binding and was almost undetectable in lysates of whole cells scraped off YESCA plate (Fig. 8a and b, lanes 7 and 8). $CsgA^{Q49N}$ was

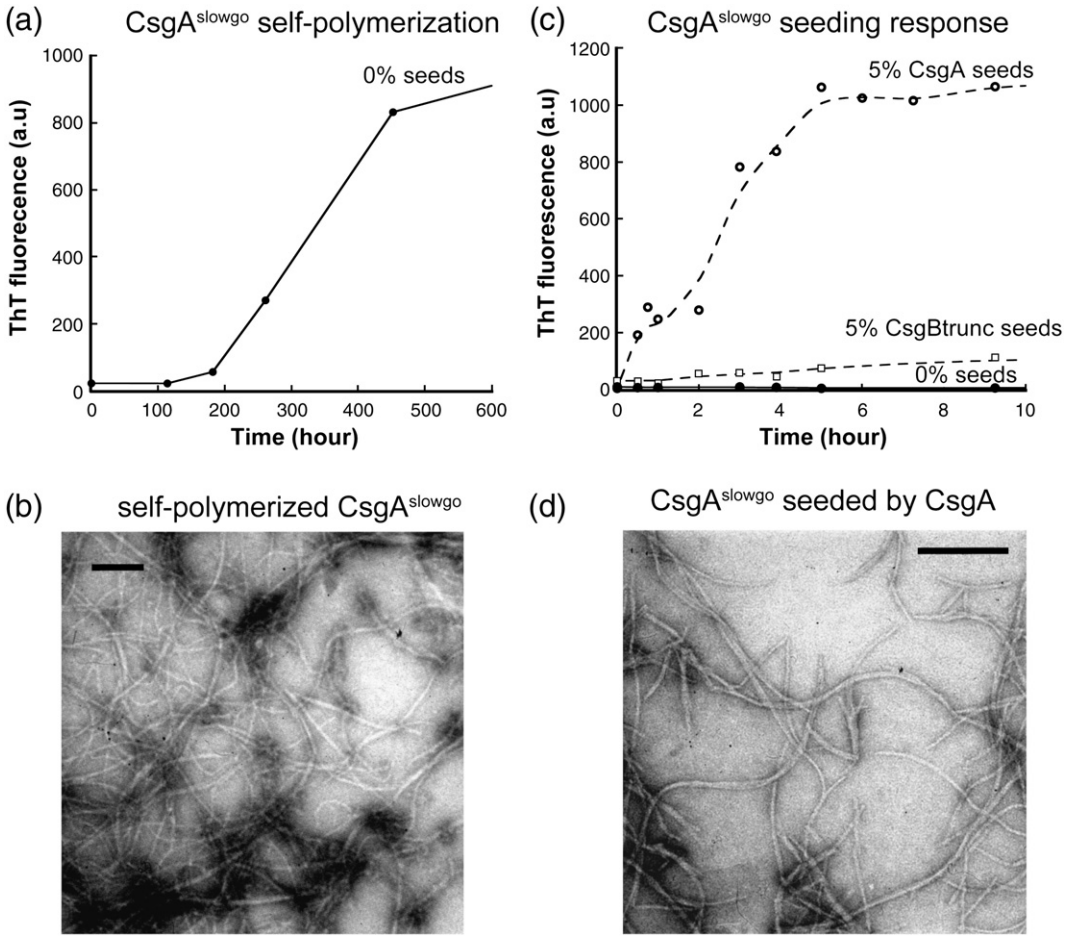

Fig. 7. *In vitro* self-polymerization and seeding responsiveness of $CsgA^{slowgo}$. a, ThT fluorescence intensity of 40 μM $CsgA^{slowgo}$ incubated for the indicated time intervals. *a.u.*, arbitrary units. b, Negative stain TEM micrograph of *in vitro* self-polymerized aggregates of $CsgA^{slowgo}$. Scale bars represent 200 nm. c, CsgA seeds or $CsgB_{trunc}$ seeds (5% w/v) or buffer vehicle were added to 60 μM $CsgA^{slowgo}$ and the polymerization was measured by ThT fluorescence assay. The fluorescence intensity of seeded $CsgA^{slowgo}$ samples were shown at the indicated time intervals. The contribution to fluorescence intensity by seeds was subtracted. d, Negative-stain TEM micrograph of *in vitro* polymerized aggregates of $CsgA^{slowgo}$ promoted by CsgA. Scale bars represent 200 nm.

stable and secreted into the agar, as detected by Western analysis of cells and the underlying agar (Fig. 8c, lanes 7 and 8). In addition, whole-cell Western analysis showed that Q139N (Fig. 8b, lanes 15 and 16) was as defective as Q139A (Fig. 2a, lanes 39 and 40). The sequence requirements at positions 49 and 139 appear extremely stringent during fiber assembly, since either Ala or Asn was not tolerated at these positions. The Congo red binding of $CsgA^{N144Q}$ and $CsgA^{N54Q}$ was similar to that of wild-type CsgA (Fig. 8a). In addition, $CsgA^{N144Q}$ and $CsgA^{N54Q}$ were predominately SDS-insoluble, as shown by whole-cell Western analysis (Fig. 8b, lanes 11, 12, 13 and 14). The SDS-insolubility of $CsgA^{N144Q}$ and $CsgA^{N54Q}$ was similar to that of wild-type CsgA, as detected by Western analysis of cells and the underlying agar (Fig. 8c, lanes 11, 12, 15 and 16). We observed wild-type-like fibers assembled by the *csgA* strain that contained a plasmid encoding $CsgA^{N144Q}$ by TEM, while relatively few fibers were observed on the *csgA* strain expressing $CsgA^{Q49N}$ (Fig. 8d). Collectively, *in vivo* Gln residues at positions 49 and 139 could not be replaced by Asn, while Asn residues at position 54 and 144 could be replaced by Gln and function like wild-type CsgA. *In vitro*, $CsgA^{Q49N}$ was as defective as $CsgA^{Q49A}$ in both heteronucleation responsiveness and self-polymerization. $CsgA^{Q49N}$ polymerization had greater T_c than that of CsgA at the same concentration measured by ThT assay (Fig. 8e). $CsgA^{Q49N}$ did not respond to CsgB nucleation at 10 μM, as detected by overlay assay (Fig. 8e, insert). It has been proposed that amino acid content can be an important factor in promoting amyloid formation.[23,24] To test whether N144Q can suppress the polymerization defect of Q49N, we constructed the mutant Q49N/N144Q. Like Q49N, Q49N/N144Q had significantly reduced Congo red binding relative to that of wild-type CsgA (Fig. 8f). Collectively, these findings suggest that CsgB-mediated

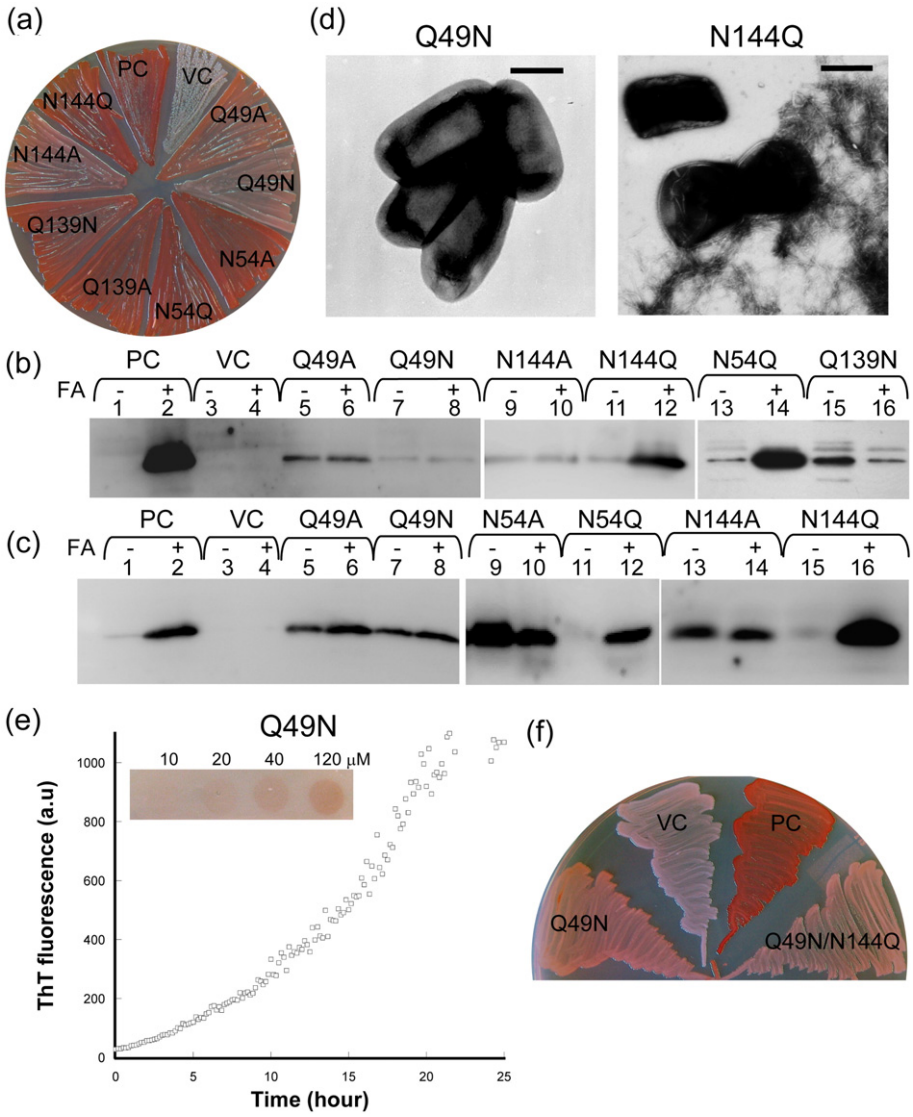

Fig. 8. Conservative Q/N substitutions at positions 49, 54, 139 and 144. a, CR-YESCA plate with *csgA* transformed with the plasmids encoding CsgA, $CsgA^{Q49A}$, $CsgA^{Q49N}$, $CsgA^{N54A}$, $CsgA^{N54Q}$, $CsgA^{Q139A}$, $CsgA^{Q139N}$, $CsgA^{N144A}$ or $CsgA^{N144Q}$. b, Western blots of whole-cell lysates from *csgA* cells containing plasmids encoding wild-type CsgA, $CsgA^{Q49A}$, $CsgA^{Q49N}$, $CsgA^{N54Q}$, $CsgA^{Q139N}$, $CsgA^{N144A}$ or $CsgA^{N144Q}$. c, Western blot of whole cells and underlying agar (agar plugs) from *csgA* strains containing plasmids encoding CsgA, $CsgA^{Q49A}$, $CsgA^{Q49N}$, $CsgA^{N54A}$, $CsgA^{N54Q}$, $CsgA^{N144A}$ or $CsgA^{N144Q}$. Samples were treated with (+) or without (–) FA. The blots were probed with anti-CsgA antibody. d, Negative-stain TEM micrographs of *csgA* cells containing plasmids encoding $CsgA^{Q49N}$ and $CsgA^{N144Q}$. Scale bars represent 500 nm. e, The polymerization of 56 μM $CsgA^{Q49N}$ was measured by ThT fluorescence. Congo red staining of *csgA* cells overlaid with $CsgA^{Q49N}$ at various concentrations is shown in the inset. f, CR-YESCA plate of a *csgA* strain transformed with the control vector pLR2 or the plasmids encoding CsgA, $CsgA^{Q49N}$ or $CsgA^{Q49N/N144Q}$.

heteronucleation and CsgA self-polymerization require spatially constrained contacts between certain amino acid side-chains.

Discussion

Curli functions as part of the extracellular matrix produced by many Gram-negative enteric bacteria. Curli assembly is a precisely coordinated process that is highly amenable to study because of the sophisticated genetic and biochemical tools afforded by *E. coli.* Here, we have elucidated the CsgA sequence determinants that drive amyloid formation. We found that Gln and Asn residues at the N- and C-terminal repeats are critical for curli assembly. These Gln and Asn residues are necessary for CsgA to respond to CsgB-mediated heteronucleation and to initiate efficient CsgA self-assembly *in vitro*.

The contribution of amino acid side-chains to functional amyloid formation

Unlike amorphous aggregates formed by non-specific hydrophobic interactions, amyloid fibers are highly ordered proteinaceous polymers that are stabilized by specific interactions. The side chains can clearly influence amyloidogenesis.[1] But how

specific sequences influence amyloid formation *in vivo* remains poorly understood. We demonstrated that four internally conserved Gln and Asn residues were required to respond to CsgB-mediated heteronucleation. When these four residues are changed to Ala, the resulting protein, $CsgA^{slowgo}$, cannot efficiently initiate self-assembly *in vitro,* as indicated by an extremely long lag phase before polymerization (Fig. 7a). $CsgA^{slowgo}$ also cannot respond efficiently to CsgB-mediated heteronucleation (Fig. 6). These results suggest that the specific side-chain interactions are the driving forces that efficiently initiate bacterial amyloid formation *in vivo* and *in vitro*. Interestingly, we found that these Gln/Asn residues required for CsgB-mediated heteronucleation are not necessary for CsgA fiber-mediated seeding, suggesting that CsgA seeding and CsgB-mediated heteronucleation may have somewhat distinct mechanisms (Fig. 7c). $CsgA^{slowgo}$ is completely defective in amyloid formation *in vivo*, although it can assemble inefficiently into amyloid fibers *in vitro* (Fig. 7a and b). This distinction between behavior *in vitro* and *in vivo* suggests that *in vivo* amyloid formation has highly stringent requirements that limit initiation of amyloid formation to CsgB-mediated heteronucleation. For example, CsgA remains soluble in the extracellular milieu in the absence of nucleator CsgB, even though CsgA is amyloidogenic *in vitro*.[8] This indicates that CsgA may diffuse into the extracellular milieu if it cannot participate in efficient nucleation. Curli also provides an elegant example of how amyloidogenecity can be controlled temporally and spatially. It is possible that an evolved functional amyloid such as curli capitalizes on specific side-chain interactions to tightly regulate and control amyloidogenecity.

The diverse sequence determinants of amyloid formation

The sequence-specific determinants of amyloid formation have been most studied *in vitro*. Clearly, different amyloid proteins employ various interactions such as hydrogen bonding, hydrophobic and electrostatic interactions to promote intra- and intermolecular associations. For example, the hydrophobic stretches in Aβ42, rather than specific side chains, are sufficient to promote aggregation.[25] Yeast prion proteins Sup35p, Ure2p and Rnq1p all have Gln/Asn-rich domains that are essential for prion propagation *in vivo* and *in vitro*.[11,26–28] In Huntington's disease, polyglutamine peptides have an essential role in molecular etiology. In addition, the length of polyglutamine peptides is positively correlated to aggregation *in vitro*.[29] Gln/Asn-rich domains are proposed to have a high potential to be amyloidogenic because they can form intramolecular hydrogen bonds.[26] Here, we have shown that the internally conserved Gln and Asn residues at the N- and C-terminal repeats are critical for CsgA fiber formation. The initiation of CsgA polymerization both *in vitro* and *in vivo* appears to be dependent on hydrogen bonds formed by these Gln and Asn residues.

Moreover, we showed that 13 of 15 Ala substitution mutants of Gln/Asn residues (Q49A, N54A, Q60A, Q72A, N77A, Q94A, N99A, Q105A, Q117A, N122A, Q128A, Q139A and N144A) were somewhat more sensitive to SDS relative to wild-type CsgA (Fig. 2a), suggesting these polar residues help stabilize the amyloid structure. Interestingly, the polar amides of these internally conserved Gln and Asn residues were proposed to form hydrogen bonds with backbones to stabilize its fibril structure in a CsgA-predicted structural model.[10] It is possible that these critical Gln/Asn residues have an important role in the initiation of fiber assembly, and stabilize the fiber structure when fibers are formed.

It was proposed that interactions among aromatic residues by π-stacking have a critical role in the self-assembly of some amyloid fibrils, such as islet amyloid polypeptide.[15] To assess the role of aromatic residues in guiding *in vivo* CsgA amyloidogenesis, we replaced each aromatic residue in the amyloid core of CsgA individually, revealing that aromatic residues do not contribute significantly to curli assembly (Fig. 3).

Bacterial amyloid formation is influenced by the spatial arrangement of Q/N residues

It has been suggested that amyloid formation is determined by simple physicochemical factors of polypeptides, such as hydrophobicity, charge and secondary structure propensity.[22] Furthermore, the amyloidogenecity of yeast prions Ure2p and Sup35p appears to be encoded by a particular amino acid composition, rather than specific sequences.[23,24] In contrast, our work demonstrates that CsgA amyloidogenesis, especially *in vivo*, is highly dependent on its specific sequence rather than on its overall amino acid composition or the physicochemical properties of polypeptide chains. The Gln residue at positions 49 and 139 of CsgA could not be replaced by Asn residues without interfering with curli assembly (Fig. 8), even though Asn shares many physicochemical features with Gln. $CsgA^{Q49N\text{-}N144Q}$ and wild-type CsgA have identical amino acid composition, but $CsgA^{Q49N\text{-}N144Q}$ is dramatically defective in curli assembly (Fig. 8f). Additionally, although $CsgA^{Q49A}$, $CsgA^{Q60A}$, $CsgA^{Q94A}$, $CsgA^{Q105A}$, $CsgA^{Q117A}$, $CsgA^{Q128A}$, and $CsgA^{Q150A}$ have identical amino acid compositions, $CsgA^{Q49A}$ is dramatically more defective than other Ala mutants in fiber assembly (Figs. 1b and 2a). We previously showed that N- and C-terminal CsgA repeats drive curli assembly.[21] Our current study suggests the position of Gln or Asn residues has an equally important role for amyloid assembly.

Globular protein folding depends on specific side chain contacts. Whether amyloid formation, especially *in vivo,* depends on specific side chain contacts is just beginning to be tested. Our work demonstrates that the initiation of CsgA amyloid formation requires specific amino acid side chains. Specific side chain contacts are likely to facilitate the spatial and

temporal control over the polymerization of functional amyloids.

Materials and Methods

Bacterial growth

To induce curli production, bacteria were grown on YESCA plates (per liter; 1.0 g of yeast extract, 10 g of Casamino acids and 20 g of agar) at 26 °C for 48 h.[18] Antibiotics were added to plates at the following concentrations: kanamycin, 50 μg/ml; chloramphenicol, 25 μg/ml; ampicillin, 100 μg/ml. Production of curli was monitored by growth on Congo red-YESCA (CR-YESCA) plates.[18]

Strains and plasmids

Strains LSR10 (MC4100 Δ*csgA*), LSR12 (C600 Δ*csgBAC* and *csgDEFG*), LSR13 (MC4100 Δ*csgBA*) and MHR261 (MC4100 Δ*csgB*) and plasmids pLR5 (encoding CsgA), pLR2 (vector control) and pMC1(encoding CsgG) have been described.[6,7,9,18] Plasmids containing CsgA mutations were constructed by site-specific mutagenesis using standard overlapping PCR extension. PCR products containing relevant mutations and NcoI/BamHI restriction endonuclease sites at 5′/3′ ends were cloned downstream of the *csgBA* promoter into NcoI/BamHI sites of vector pLR2. The expression vector pMC3 was generated earlier.[18] To express and purify CsgA mutant proteins, PCR-amplified mutant sequences including a C-terminal His_6 tag were cloned into pMC3, replacing the sequence encoding CsgA-his. Sequences of constructs were verified by DNA sequencing.

Western analysis

The immunoblotting of CsgA in whole cells and plugs was done as described.[18] Briefly, bacterial cells grown on YESCA plate for 48 h were scraped off and normalized by absorbance at 600 nm. Cell pellets were resuspended in 2× SDS loading buffer either with or without prior treatment with FA as described.[18] Alternatively, 8 mm circular plugs including cells and underlying agar were resuspended in 2× SDS loading buffer either with or without prior FA treatment. Samples were submitted to SDS-PAGE (15% polyacrylamidegel) and transferred onto polyvinylidene difluoride membrane using standard techniques. Western blots were probed and developed as described.[7]

Transmission electron microscopy

A Philips CM10 transmission electron microscope was used to visualize the cell samples and protein fiber aggregates. Samples (10 μl) were placed on Formvar-coated copper grids (Ernest F. Fullam, Inc., Latham, NY) for 2 min, washed with deionized water, and negatively stained with 2% (w/v) uranyl acetate for 90 s.

In vitro polymerization and seeding assay

CsgA or CsgA mutant proteins and CsgG were overexpressed by induction with isopropyl β-D-1-thiogalactopyranoside (IPTG) in LSR12 (C600 Δ*csgBAC* and *csgDEFG*) strain. CsgA or CsgA mutant proteins were secreted to the supernatant and purified as described.[8] After removal of imidazole by passage through a desalting column, purified proteins were passed through a 30 kDa cutoff filter (Amicon® Ultra, MA) to remove possible aggregates and seeds that might alter the polymerization kinetics. Purified homogeneous proteins were loaded onto 96-well opaque plates and ThT was added to a final concentration of 20 μM. Every 10 min, after shaking for 5 s, ThT fluorescence was measured with a Spectramax M2 plate reader (Molecular Devices, Sunnyvale, CA) set to 438 nm excitation and 495 nm emission with a 475 nm cutoff. Alternatively, samples were kept at room temperature in the presence of 0.02% (w/v) NaN_3. At the indicated time-points, aliquots were removed, ThT was added to a final concentration of 20 μM, and fluorescence intensity was measured as described above. For seeding reactions, CsgA fiber seeds, prepared as described,[8] were added to freshly purified CsgA mutant proteins immediately before the start of ThT fluorescence assay.

Overlay assay

A 10 μl sample of freshly purified proteins was spotted onto a lawn of *csgA* (LSR10) or *csgBA* (LSR13) cells grown on YESCA plates at 26 °C for 48 h. The plates were incubated for 10 min at room temperature, stained with a 0.5 mg/ml Congo red in 50 mM potassium phosphate (pH 7.2) for 5 min, then washed with potassium phosphate buffer and photographed.

Interbacterial complementation

Cells that secrete CsgA or CsgA mutant proteins were streaked adjacent to *csgA* cells on a CR-YESCA plate. The plates were photographed after growth for 48 h at 26 °C.

Acknowledgements

We thank the members of the Chapman laboratory and Ryan Frisch for helpful discussions and review of this manuscript. We thank Bryan McGuffie for his technical help in the preparation of figures. This work was supported by NIH award AI073847-01.

References

1. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333–366.
2. Hardy, J. & Selkoe, D. J. (2002). The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, **297**, 353–356.
3. Fowler, D. M., Koulov, A. V., Balch, W. E. & Kelly, J. W. (2007). Functional amyloid–from bacteria to humans. *Trends Biochem. Sci.* **32**, 217–224.
4. Hammer, N. D., Wang, X., McGuffie, B. A. & Chapman, M. R. (2008). Amyloids: friend or foe? *J. Alzheimers Dis.* In the press.
5. Barnhart, M. M. & Chapman, M. R. (2006). Curli biogenesis and function. *Annu. Rev. Microbiol.* **60**, 131–147.

6. 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.
7. Hammer, N. D., Schmidt, J. C. & Chapman, M. R. (2007). The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. *Proc. Natl Acad. Sci. USA*, **104**, 12494–12499.
8. 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.
9. 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**, 870–881.
10. Collinson, S. K., Parker, J. M., Hodges, R. S. & Kay, W. W. (1999). Structural predictions of AgfA, the insoluble fimbrial subunit of Salmonella thin aggregative fimbriae. *J. Mol. Biol.* **290**, 741–756.
11. DePace, A. H., Santoso, A., Hillner, P. & Weissman, J. S. (1998). A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. *Cell*, **93**, 1241–1252.
12. Osherovich, L. Z., Cox, B. S., Tuite, M. F. & Weissman, J. S. (2004). Dissection and design of yeast prions. *PLoS Biol.* **2**, E86.
13. Liu, J. J. & Lindquist, S. (1999). Oligopeptide-repeat expansions modulate 'protein-only' inheritance in yeast. *Nature*, **400**, 573–576.
14. Tessier, P. M. & Lindquist, S. (2007). Prion recognition elements govern nucleation, strain specificity and species barriers. *Nature*, **447**, 556–561.
15. Gazit, E. (2002). A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J.* **16**, 77–83.
16. Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F. & Tycko, R. (2002). A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl Acad. Sci. USA*, **99**, 16742–16747.
17. Luhrs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Dobeli, H. *et al.* (2005). 3D structure of Alzheimer's amyloid-beta(1-42) fibrils. *Proc. Natl Acad. Sci. USA*, **102**, 17342–17347.
18. Chapman, M. R., Robinson, L. S., Pinkner, J. S., Roth, R., Heuser, J., Hammar, M. *et al.* (2002). Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science*, **295**, 851–855.
19. Collinson, S. K., Emody, L., Muller, K. H., Trust, T. J. & Kay, W. W. (1991). Purification and characterization of thin, aggregative fimbriae from Salmonella enteritidis. *J. Bacteriol.* **173**, 4773–4781.
20. Azriel, R. & Gazit, E. (2001). Analysis of the minimal amyloid-forming fragment of the islet amyloid polypeptide. An experimental support for the key role of the phenylalanine residue in amyloid formation. *J. Biol. Chem.* **276**, 34156–34161.
21. Wang, X., Hammer, N. D. & Chapman, M. R. (2008). The molecular basis of bacterial amyloid polymerization and nucleation. *J. Biol. Chem.* In the Press. doi:10.1074/jbc.M800466200.
22. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature*, **424**, 805–808.
23. Ross, E. D., Edskes, H. K., Terry, M. J. & Wickner, R. B. (2005). Primary sequence independence for prion formation. *Proc. Natl Acad. Sci. USA*, **102**, 12825–12830.
24. Ross, E. D., Baxa, U. & Wickner, R. B. (2004). Scrambled prion domains form prions and amyloid. *Mol. Cell Biol.* **24**, 7206–7213.
25. Kim, W. & Hecht, M. H. (2006). Generic hydrophobic residues are sufficient to promote aggregation of the Alzheimer's Abeta42 peptide. *Proc. Natl Acad. Sci. USA*, **103**, 15824–15829.
26. Michelitsch, M. D. & Weissman, J. S. (2000). A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc. Natl Acad. Sci. USA*, **97**, 11910–11915.
27. Sondheimer, N. & Lindquist, S. (2000). Rnq1: an epigenetic modifier of protein function in yeast. *Mol. Cell*, **5**, 163–172.
28. Taylor, K. L., Cheng, N., Williams, R. W., Steven, A. C. & Wickner, R. B. (1999). Prion domain initiation of amyloid formation in vitro from native Ure2p. *Science*, **283**, 1339–1343.
29. Chen, S., Ferrone, F. A. & Wetzel, R. (2002). Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl Acad. Sci. USA*, **99**, 11884–11889.