Inhibiting Functional Amyloid Formation Using Protein Engineering

by

Anthony Balistreri

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2022

Doctoral Committee:

Professor Matthew R Chapman, Chair
Professor Julie Biteen
Associate Professor Nicole Koropatkin
Assistant Professor Randy Stockbridge

Anthony Balistreri

abalist@umich.edu

ORCID iD: 0000-0002-7409-7031

© Anthony Balistreri 2022

Dedication

To my grandfathers Leo Sciuto and Pietro Balistreri.

Acknowledgements

It would be remiss if I didn't start out by thanking my family, with special consideration towards my mother, Angela. I'm the odd one out, the misfit who dedicated their life to something different than the rest. When I made the decision as a young teenager to pursue a career in science my mom supported me in every way that I could imagine. Neither of us knew what to do. Being the first aspiring scientist in my family history meant walking down unfamiliar roads. And I was never trudging along alone, my family was with me the whole time.

In lieu of their physical presence, I augmented my support system with new people met along the way. Friends who became like family and supported me thusly. We experienced life together, but most importantly they ate my food and said nice things about it. There are too many to mention, which is perhaps a testament to the amount of support I've required. Given the context of this document I need to mention the friends I've made here in Ann Arbor as being especially important. If the adage "a penny for your thoughts" were true, I would be indebted to this band of friends, more so than I already am.

I came to Michigan to enter a PhD program which would train me to be a scientist. By the time I had arrived that training had already started years prior at San Jose State University. Dr. Laura Miller Conrad and the rest of the Chemistry Department at SJSU began the process of molding me into the scientists that I am now. I would not have been nearly as successful at Michigan had I not received Laura's formative guidance.

When I was settled in to the Chapman lab I was given great freedom to explore scientific questions of my own interest. Rather unsurprisingly, diving into the unexplored lead to confusion

and the feeling of being lost in an unfamiliar place. Chapman lab members know who we turn to in times of scientific uncertainty: other Chapman lab members. Simply put, without their support the work outlined in this dissertation would not exist. Special members of the lab include the group of students that I trained, lovingly referred to as "The Ant Farm". Mentoring became a bigger task than I had originally anticipated when I joined the lab and the fruits of that labor extend past the time spent at the benchtop. I hope some of these relationships and friendships last the length of our careers.

Lastly, I need to recognize the outsized effect that Sir Professor Doctor Matthew Chapman, esq. has had on my career and life here at Michigan. The last five years would have been so different had I not told myself, "If I just don't clean out my desk, maybe he'll just keep me around." I left our first meeting together with the strong notion that I wanted to learn how to be a scientist from him. That thought has remained steadfast.

Table of Contents

List of Tables

List of Figures

Abstract

Amyloids are a class of protein assembly known for their extreme stability and ordered, fibrous structure. Functional amyloids are a subclass of amyloids that can be found in all domains of cellular life. Functional amyloids fulfill key biological roles that help the organism that produce them grow and thrive in their environment. Functional amyloids can act as structural scaffolds, storage mechanisms, antibacterial agents, and major components of microbial biofilms. The most well studied functional amyloid is curli, the predominant protein component of the *E. coli* biofilm. The major protein subunit of curli is a fast-aggregating amyloid protein called CsgA. *E. coli* secretes intrinsically disordered CsgA to the outside of the cell where it can then fold into an amyloid-competent state and form fibers. Though CsgA has a high propensity to form amyloid fibers its polymerization can be controlled through different biochemical mechanisms. In this dissertation I focused on two ways that functional amyloid formation is controlled, both *in vivo* and *in vitro*.

E. coli uses an amyloid inhibitor protein called CsgC to control CsgA aggregation within the cell. CsgC is a periplasmic chaperone-like protein that maintains CsgA in in an intrinsically disordered state. In collaboration with the Olofsson (Umeå University, Sweden) and Ruotolo (University of Michigan) labs, I investigated the interaction between CsgC and CsgA. We observed a 1:1 heterodimeric complex of CsgC:CsgA that suggested that CsgC interacts with monomers to keep CsgA unfolded and non-amyloidogenic. We found that after interacting transiently with CsgC, CsgA monomers aggregate much more slowly than compared to a negative control. To learn more about how CsgC interacts with CsgA, I developed an *in vivo*

screening assay that was able to identify CsgC residues that are important for structural stability and amyloid inhibition activity.

Controlling CsgA aggregation *in vitro* is also a useful research endeavor. Upon purification, intrinsically disordered CsgA proteins immediately begin oligomerizing and within a few hours stable, mature amyloid fibers will have formed. To aid the study of amyloid proteins, I created a CsgA variant called $CsgA_{CC}$ that remains monomeric and unfolded while oxidized. When $CsgA_{CC}$ is incubated under reducing conditions, an intramolecular disulfide bond is broken, and the protein adopts the amyloid conformation. Therefore, the aggregation propensity of $CsgA_{CC}$ can be manipulated by the addition of a reducing agent. $CsgA_{CC}$ will enable scientists to control amyloid formation in ways that were never possible before, hopefully leading to insights into how aggregation occurs. In addition, an amyloid protein that responds to a reducing agent to begin polymerization could prove to be a useful building block for constructing complex structures using functional amyloid fibers as the core structural element.

Chapter 1 Functional Amyloids are the Rule Rather Than the Exception in Cellular Biology[1]

1.1 Abstract

Amyloids are a class of protein aggregates that have been historically characterized by their relationship with human disease. Indeed, amyloids can be the result of "misfolded" proteins that self-associate to form insoluble, extracellular plaques in diseased tissue. For the first 150 years of their study, the pathogen-first definition of amyloids was sufficient. However, at the turn of the twenty-first century new observations of amyloids fostered an appreciation for non-pathological roles for amyloids in cellular systems. There is now evidence from all domains of life that amyloids can be non-pathogenic and 'functional', and that their formation can be the result of purposeful and controlled cellular processes. So called "functional amyloids" fulfill an assortment of biological functions including acting as structural scaffolds, regulatory mechanisms, and storage mechanisms. The conceptual convergence of amyloids serving a functional role has been repeatedly confirmed by discoveries of additional functional amyloids. With dozens already known, and with the vigorous rate of discovery, the biology of amyloids is robustly represented by non-pathogenic amyloids.

[1] Portions of this chapter are associated with the following publication: Balistreri, A., Goetzler, E., and Chapman, M. (2020). Functional Amyloids Are the Rule Rather Than the Exception in Cellular Biology. Microorganisms *8*, 1951. AB and EG wrote the manuscript. MC and AB edited the manuscript.

1.2 Introduction

Amyloids are fibril protein aggregates originally identified in 1854 by Rudolph Virchow when he observed iodine-stained plaques in abnormal brain tissues [1–3]. Since the nineteenth century amyloids, and the extracellular bodies they form in human tissue, have been inextricably connected to human disease and neurological dysfunction. At the turn of the twenty-first century, new observations led to an expansion of the definition of amyloids to include a class of non-pathogenic aggregates [1,4]. "Functional amyloids" are created by organisms intentionally, in order to exploit their many useful properties [5]. In the last 20 years the list of functional amyloid proteins has blossomed (See **Table 1-1**). This has shifted views of amyloid biology from singularly pathogenic and cytotoxic structures to a protein fold that contributes positively to cellular biology. Our thesis put forth here is that functional amyloids are the principle class of amyloids found in nature.

1.3 The established perspective of amyloid proteins.

Since their discovery, amyloids have been most closely associated with human disease. The amyloid field was ushered into medical science in 1854 when the German physician Rudolph Virchow observed iodine-stained "copora amylacae" in nervous tissue [6], characterizing the bodies as being starch-like (Latin for starch is amylum). Seven decades later the textile dye Congo Red was identified as a useful compound for specifically staining amyloids in histological samples taken from diseased brain tissues [7]. In the same 1927 publication, the Belgian physician Paul Divry observed that Congo Red stained Alois Alzheimer's plaques that he associated with presenile dementia, thereby connecting these plaques to amyloids [1,7]. Divry's publication sparked an exploration for amyloids in histological samples that continued through the twentieth century, during which scientists found evidence of amyloids related to both

localized and systematic syndromes [1,2]. The Pras extraction method introduced in 1968 allowed for greater biochemical and structural characterization of amyloids [1,8]. Efficient fiber extraction coupled with amino acid sequencing lead researchers to determine that amyloid fibers associated with neuropathic or systemic disorders were composed of different proteins, each with their own clinical manifestations [2]. There are now 36 different proteins that have been identified as forming pathogenic amyloid deposits in human tissues [9]. Amyloids remain a widely discussed topic in scientific and medical literature; in the last five years an average of 5,000 articles per year were written on the topic of "amyloids" [10].

At the turn of the twenty-first century, several groups identified roles for amyloids in biology other than causing disease. In the late 1990's yeast geneticists finally solved the strange issue of non-Mendelian heritable traits propagating in yeast [11–14]. An infective amyloid protein called Sup35 was shown to have epigenetic-like control over protein expression based on the presence of the amyloid or soluble form of the protein [11–14]. In 2000 it was discovered that amyloid rodlets made up of hydrophobins allow fungi to escape an aquatic environment [15]. In the same year, amyloid proteins called chorions were identified as the major protective component surrounding silkmoth eggs [16]. In 2002, curli, the main proteinaceous component of the E. coli biofilm, were revealed to share biophysical characteristics with pathogenic amyloid fibers [17]. However, unlike pathogenic amyloids, these functional amyloids are not the product of stochastic protein misfolding and are often assembled via tightly regulated and controlled pathways. Since the early 2000's many functional amyloids proteins that have been discovered and described. The functional amyloids share biophysical characteristics with their pathogenic counterparts, including tinctorial properties, ability to self-assemble, and their fibrous 3D

structure (**Figure 1-1**). There are now approximately 30 new functional amyloid proteins, and new ones are continuing to be described [18].

1.4 The amyloid fold is intrinsic to polypeptides.

Classically, amyloids have been defined by their biophysical characteristics, including the cross-β structure where β-strands align perpendicular to the fiber axis [3]. Recently, the amyloid structural catalog was expanded to include the cross-α structure of phenol-soluble modulins, a functional amyloid produced by S. aureus [19]. Collectively, all amyloids are highly ordered aggregates that are in a low energy conformation and that are highly stable and resistant to denaturation [20,21].

It has been suggested that the amyloid fold can be achieved by peptides, regardless of their specific amino acid composition. Anfinsen's dogma describes protein folding wherein the 3D structure of a fully folded protein is determined by its primary sequence [22]. Certainly, this is true of globular proteins that adopt their native fold under optimal conditions. However, the native fold is only one of several thermodynamic minima [21,23]. Indeed, even a natively folded protein exists in a metastable state, and with enough energy proteins can fold into the amyloid-specific, cross β-sheet conformation that is the lowest energy state [24] (**Figure 1-2**). Experiments that observed human lysozyme and transthyretin adopting the amyloid fold found that partially unstable or unfolded domains were the focal point of amyloidogenesis [25–27]. However, even stably-folded proteins such as the SH3 domain or acylphosphatase, can form amyloid after partial denaturation [57–59]. Even chains of polyalanine and polyglutamate have been shown to adopt the amyloid fold during in silico modeling experiments [60–62]. Therefore, given the ubiquity of the amyloid conformation in protein structure, it is not surprising that amyloids are part of normal cellular biology.

1.5 Curli and Fap are bacterial amyloids whose assembly is highly orchestrated.

Arguably the most-studied bacterial functional amyloid are curli amyloids that are produced in an exquisitely-controlled process by E. coli [5] (**Figure 1-3a**). The curli specific operon (csg) contains 7 proteins including the master biofilm regulator protein CsgD which regulates curli production by responding to changes in the expression of hundreds of genes and external stimuli [63]. CsgA, the major curli subunit and functional amyloid protein, is translated and translocated directly into the periplasm through the SecYEG complex on the E. coli inner membrane [64]. Inside the periplasm, the nascent and unfolded CsgA is stabilized in an unstructured state by a chaperone-like protein called CsgC [65]. A second chaperone, CsgE, ferries CsgA to an outer membrane pore composed of the homo-nonameric CsgG [66]. CsgA then passes through the pore to be fully secreted to the extracellular space [67]. CsgB, the curli nucleator protein, is also secreted in the same fashion [68], however, it becomes anchored to the cell surface by CsgF [69]. Finally, curli fibers form on the cell surface after extracellular CsgA amyloid formation is templated by CsgB nuclei [70]. It is through the action of all these proteins that E. coli can assemble amyloid fibers at the correct time and space so that cellular fitness is not compromised.

Fap is another bacterial biofilm amyloid that displays a well-controlled mechanism of formation in P. fluorescens and other Pseudomonads [71]. In fact, the mechanism for amyloid formation in P. fluorescens is quite similar to the mechanism responsible for curli production in E. coli [72] (**Figure 1-3b**). Fap production is controlled by a larger operon composed of 6 genes, named fapABCDEF, in which the dominant amyloid forming protein is FapC [73]. FapA acts as a regulator of transcription, which alters the distribution of FapB and FapC in the amyloid product [73]. FapB is a nucleator protein that assists FapC as it assembles on the outer membrane

of the bacterium [71], similar to CsgA-CsgB in E. coli [74]. FapE is also incorporated at the end of the amyloid fibrils, potentially serving as a site for protein-protein interaction [73]. FapF forms a channel that shuttles FapB, FapC, and FapE to the outside of the cell membrane [75], like CsgG in E. coli [76]. The role of FapD is still a little unclear, though it has essential proteolytic activity necessary for FapC secretion [75] and may potentially be involved in cleavage of FapF [75].

Both curli and Fap amyloids are major structural components the biofilm matrix. Biofilms are entrenched colonies of bacteria that secrete a dense matrix of polysaccharides, amyloid fibers, and nucleic acids that collectively make up the extracellular matrix [77]. Within a biofilm, bacteria can continue to grow and survive some of the harshest environments, making biofilms an important factor in bacterial infection and pathogenesis [78]. When it comes to Proteobacteria, curli amyloids are essential biofilm components, illustrated by Reichhardt et al. estimating that curli make up as much as 85% of the total carbon in the E. coli extracellular matrix [79]. Indeed, the use of small molecule inhibitors targeting curli production can have a destructive effect on biofilm formation [80]. Outside of proteobacteria, Firmicutes such as S. aureus and B. subtilis utilize PSM's [81] and TasA [33] amyloids as notable biofilm components. Even in eukaryotes such as fungi [41] and microalgae [42], amyloids are the conduit that give biofilms their adhesive properties.

1.6 Other functional amyloids are also assembled in a controlled manner.

Yeast cells have adapted multiple ways to control the formation of amyloid fibers associated with the yeast prion Sup35 and its commonly observed phenotype [PSI+] [82]. Sup35 prion formation is dependent on another yeast prion called [PIN+], the insoluble amyloid form of Rnq1, a protein of unknown function [83]. The mechanism for the regulation of [PSI+] by

[PIN+] is still unclear as it relies on an inefficient and inconsistent process called “seeding”. Seeding describes the de novo construction of one prion through the interaction with a preexisting prion [83]. Though [PIN+] is required for the de novo formation of [PSI+], [PIN+] is not required for the extensive propagation of [PSI+] [84]. In fact, once the [PSI+] state has been established, [PIN+] is no longer necessary [84]. Once amyloid formation has started, the chaperone protein Hsp104 is required for the maintenance and propagation of [PSI+] [85]. There exists a critical concentration of Hsp104 that is necessary for [PSI+] formation. Too little chaperone prevents prion formation entirely, and too high of a concentration of Hsp104, the chaperone will dissociate from the unfolded prion intermediate and prevent proper aggregation [85]. A lack of Hsp104 has been proven to cure yeast cells of [PSI+] and return them to the [PSI-] state [85]. The manner in which Hsp104 facilitates aggregation is not fully clear, though it is possible Hsp104 cleaves the Sup35 protein into smaller fragments that are necessary for their inheritance and propagation as amyloids [86]. Sup35 aggregation is also controlled by association with Sup45, a binding partner that is essential for translation termination behavior [87]. When Sup45 is overexpressed, [PSI+] formation is inhibited [88].

Sup35 plays an impressive role in yeast biology, acting as a method to quickly increase genetic variation in response to swaying environmental conditions. [PSI+] is a yeast prion that represents the inactivated, aggregated state of a ribosomal elongation factor Sup35 [89]. When Sup35 is soluble and active, the predominant phenotype known as [psi-], the yeast ribosome correctly recognizes stop codons and terminates translation [90]. Cells can undergo a transition to the [PSI+] state, using the controlled mechanism discussed above, which decreases nonsense suppression [91]. When yeast cells are challenged to grow under stressful growth conditions,

[PSI+] cells are capable of creating novel, heritable phenotypes more fit to survive in the new environment [89].

In a recent publication, Yuan et al. showed that bacteria also use functional amyloids to speed up the development of new protein variants [92]. Clostridium botulinum has a transcription factor Rho that was found to contain a well-conserved candidate prion-like domain (cPrD) [92]. Chimeric proteins containing Cb-Rho cPrD produced phenotypes identical to [psi-] and [PSI+] in recombinant E. coil [92]. Interestingly, while [PSI+] decreases translational fidelity in yeast, Rho prions decrease transcriptional fidelity in bacteria, creating genetic variation in distinct yet similar manners (**Figure 1-4**).

Human cells have established a well-controlled mechanism to post-translationally regulate PMel17 amyloid formation in melanosomes. After synthesis, PMel17 associates with intraluminal vesicles (ILVs) of multivesicular bodies, where further processing and amylogenesis take place [93]. The tetraspanin protein CD63 ensures proper association of pre-processed PMel17 with ILVs by protecting it from degradation pathways [94]. Apolipoprotein E (ApoE) is also important for amyloid formation after CD63 has carried out its function [95]. Though its role is not fully understood, it acts downstream of CD63 and likely functions to assist in the sorting of PMel17 as it associates with ILVs [95]. Once associated with ILVs, PMel17 is cleaved into two subunits, Mα and Mβ, in a specified Golgi compartment [96] by a furin-like proprotein convertase [97]. The two subunits remain connected by a disulfide bond [96] until the endosomal sheddase BACE2 catalyzes the release of the Mβ subunit from the membrane-bound Mα complex [98]. The Mβ subunit is subsequently degraded by γ-secretase activity [99]. The larger fragment, Mα, remains membrane-bound to the membranes of ILVs and act as nucleation

sites upon which amyloid formation takes place [96]. Mα is further cleaved into 3 subdomains by lysosomal proteases, fragments which form the core of PMel17 amyloids [100].

While the formation of functional amyloids is tightly regulated and predictable, pathogenic amyloid formation is stochastic and unpredictable. The inappropriate accumulation of amyloid deposits and their associated pathologies are often age-dependent processes [101,102]. Amyloid formation and the resulting protein folding diseases can be coupled to the natural decline in chaperone activity and proteosome capacity in the cell [102]. Amyloidoses most often begin with a spontaneous event during which normal proteins go above a critical concentration and transition into a pathogenic state [24]. In other cases, some amyloidoses are the result of infection. Prusiner's protein-only theory postulated that the infective agent transferred between individuals in prion diseases were misfolded proteins [103]. Regarding sporadic Parkinson Disease, Braak's hypothesis suggests that alpha-synuclein aggregation could be triggered by an outside pathogens that introduce amyloids to distal nervous tissue [104]. Indeed, the microbiome can mitigate neurodegenerative diseases like Parkinson's and Alzheimer's in animal models [105–108]. In the case of dialysis-related amyloidosis, interventional medicine is to blame for the buildup of β2-microglobulin amyloids at needle injection sites [109]. These examples illustrate the sometimes random nature of pathogenic amyloidogenesis, which is in contrast to the controlled and predictable ways that functional amyloids form.

1.7 Conclusions

Historically, amyloids have been conceptually tied to the devastating human diseases that they can cause. However, in the last twenty years there have also been dozens of functional amyloids described that have helped usher in a new appreciation of amyloid biology. Since the amyloid conformation is a structure that is intrinsically available to all polypeptides it is not

surprising that Nature has found many uses for the amyloid state. Indeed, examples of beneficial amyloids can be found all over biology, performing a wide range of tasks. Evidence of the longevity and usefulness of functional amyloids can be seen in their widespread stewardship. Where functional amyloids used to represent the exceptions in amyloid biology, they are now robustly represented and provide a template for understanding how amyloid formation can occur without causing cellular toxicity and death.

1.8 Dissertation Goals

Amyloid formation is a ubiquitous process in cellular biology. Pathogenic amyloids are proteins which have misfolded and aggregated in error, leading to no small amount of human suffering. On the contrary, functional amyloids are proteins that do not misfold, but rather, form amyloid structures for a known purpose. The goal of this dissertation is to better understand how the spontaneous and exergonic aggregation of amyloid proteins can be restrained. In chapter 2 I describe the use of disulfide engineering to create $CsgA_{CC}$, a double cysteine variant of CsgA that stays intrinsically disordered and monomeric under oxidized conditions. $CsgA_{CC}$ aggregation is triggered when the protein is reduced and an intramolecular disulfide bond is broken, allowing for proper folding to take place. Making CsgA aggregation redox sensitive allows researchers to start and stop aggregation whenever they like. In chapter 3 I report our findings about CsgC and its mechanism of action. CsgC is a chaperone-like protein that inhibits CsgA amyloid formation within the *E. coli* periplasm. Amyloid formation consists of several stages as the amyloid protein transition from an intrinsically disordered protein to a beta-sheet rich fiber subunit. It is currently unclear which stage of amyloid formation CsgC is most effective. We found that CsgC makes a heterodimeric complex with CsgA and that the interaction between CsgC and CsgA is transient but affects CsgA monomer aggregation

propensity for at least two days. In chapter 4 I detail additional protein engineering experiments aimed at investigating which residues of CsgC are significant for its activity. We used directed evolution and a rational site-directed mutagenesis strategy to attempt to learn more about CsgC residues of significance. Lastly, I talk about creating three CsgC secreting strains with the goal of using CsgC in a novel way to inhibit amyloid aggregation in the mammalian gut.

Author Contributions: AB and MRC contributed conception of the review. AB and EG wrote the manuscript; AB, EG, and MRC contributed to editing and revising the manuscript.

Funding: This research was funded by the National Institutes of Health R01GM118651 and by the BSF AWD007203 grants.

Acknowledgments: We thank all the members of the Chapman lab for their helpful discussions. We thank Drs. Robert Tycko (Panels b, c, and e) and Neha Jain (Panels a, d, and f) for generously sharing the TEM micrographs used in Figure 1.

1.9 Figures and Tables

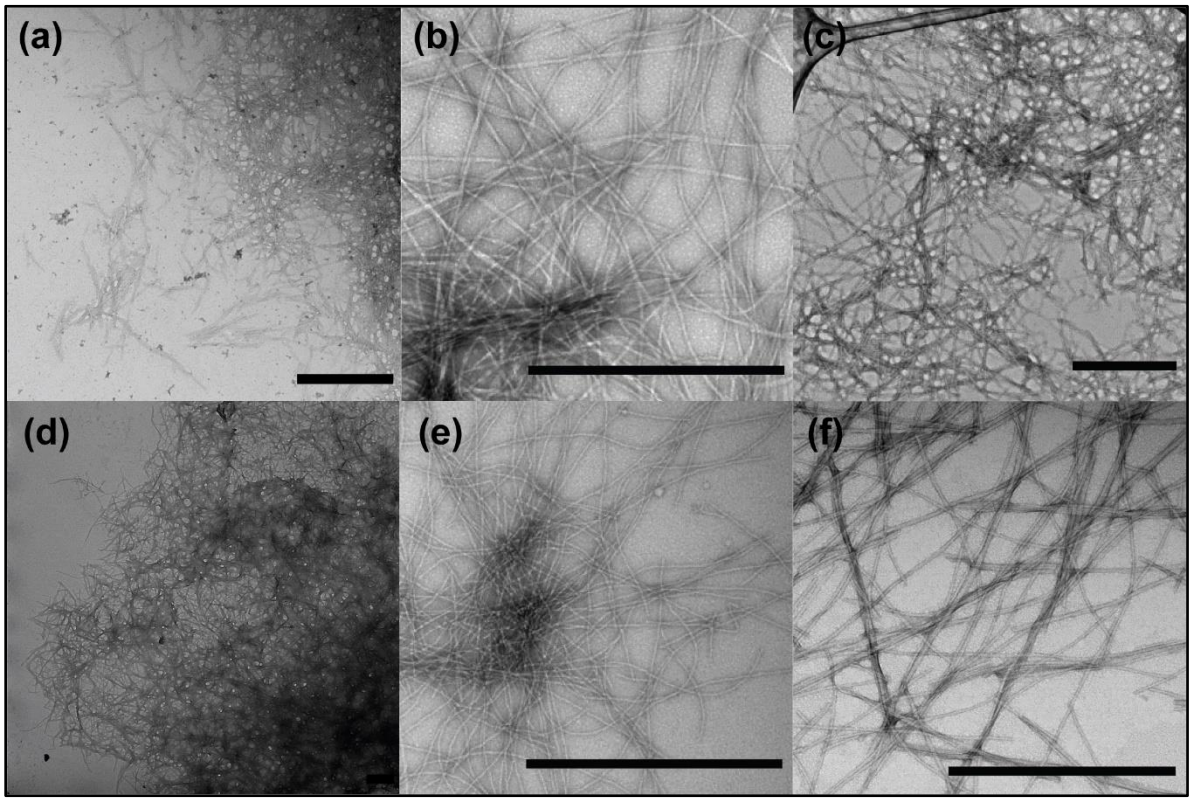

Figure 1-1. Negatively stained transmission electron micrographs of pathogenic and functional amyloid fibrils.
(a) α-Synuclein fibers. **(b)** Aβ1-40 fibrils. **(c)** Amylin fibrils. **(d)** Curli fibrils. **(e)** HET-s218-289 fibrils. **(f)** IIKIIK fibrils associated with PSMα. Scale bars represent 200 nm. Published with permission from Drs. Robert Tycko and Neha Jain.

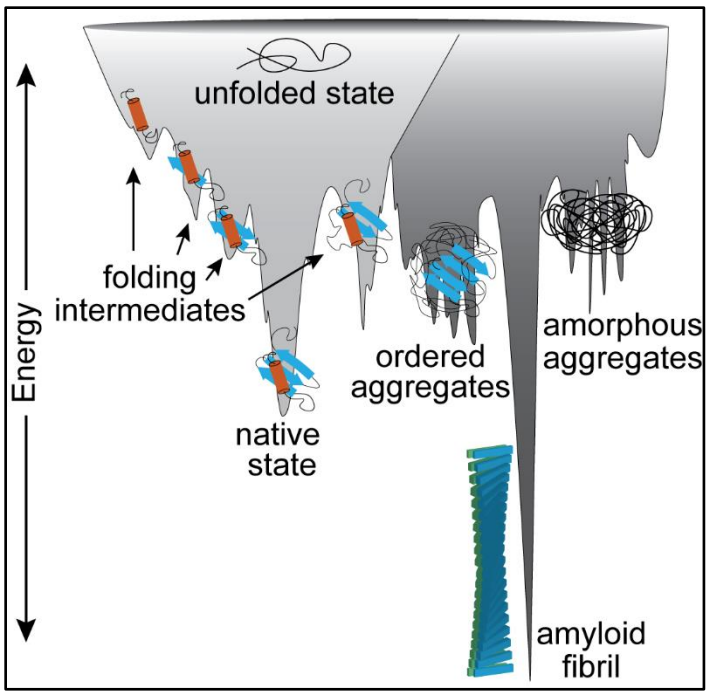

Figure 1-2. Cartoon showing the complex protein folding landscape.
Nascent, unfolded proteins travel down an energy gradient seeking the lowest energy conformation. Proteins can sample various energy minima on this journey including folding intermediates.

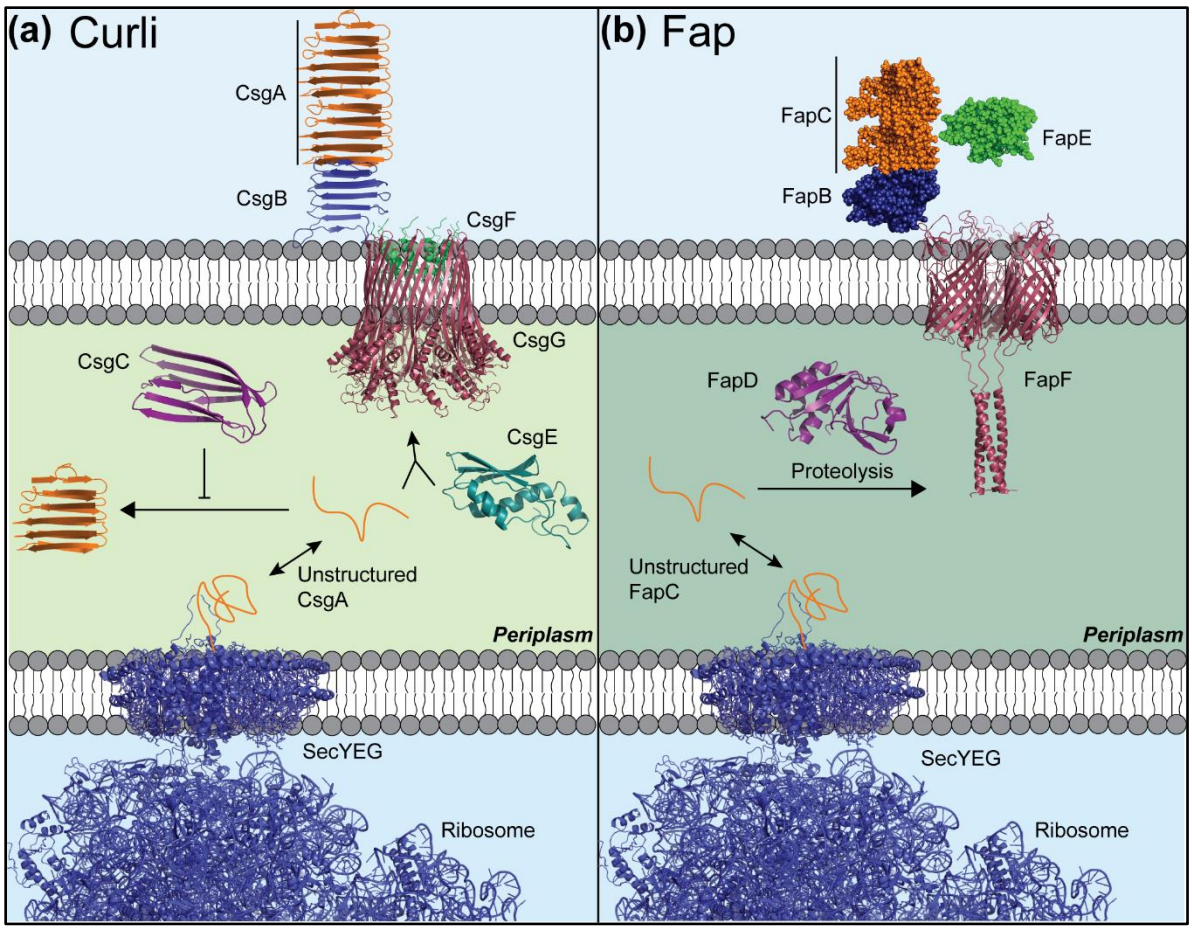

Figure 1-3. The bacterial amyloid systems called curli in E. coli and fap in P. aeruginosa are controlled through complex, multi-protein mechanisms.
(a) The curli operon contains 7 proteins (6 shown here), each playing a necessary role in maintaining spatiotemporal control over polymerization of the major curli subunit CsgA (Tian, 2015). CsgA is translated and translocated directly into the periplasm using the SecYEG secretion pore (PDB: 4V6M). CsgC (PDB: 2y2y) is a chaperone-like protein which inhibits CsgA aggregation within the periplasm. CsgE (PDB: 2NA4) is another periplasmic chaperone which fosters CsgA translocation through the nonomeric curli assembly pore CsgG (PDB: 6L7A). Lastly, CsgF (PDB: 6L7A) and CsgB both help to localize curli formation to the CsgG pore and the outer membrane, respectively. **(b)** In Pseudomonas, the major fap component FapC, is secreted into the periplasm using a SecYEG pore. FapD (modeled after the homologous C39 peptidase domain of ABC transporter PCAT1, PDB: 4RY2) is a peptidase which performs an essential proteolytic modification to one or more of the fap proteins. FapC is passed through the outer membrane using FapF, a trimeric polypeptide transporter (PDB: 5O67). Finally, FapB and FapE are essential minor components of fap amyloids, with FapB potentially playing a nucleator role similar to CsgB. Models shown of FapC, FapB, and FapE are structural predictions produced by the FALCON@home server since there is no putative structural data in the literature.

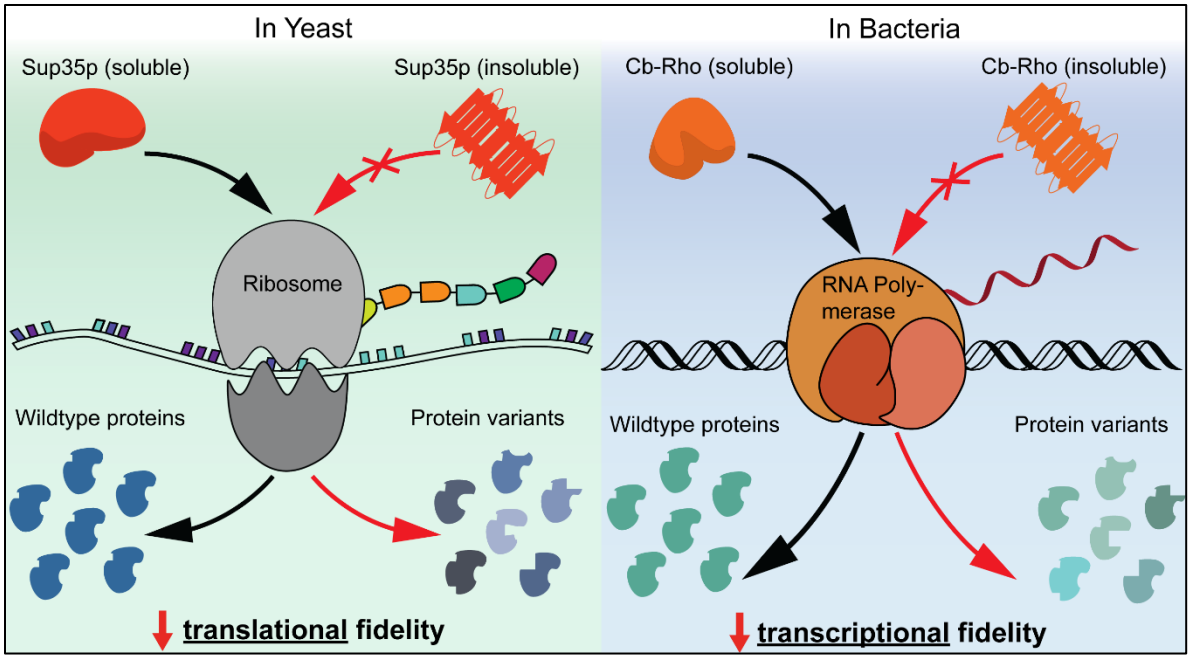

Figure 1-4. The insoluble amyloid form of Sup35p and Cb-Rho allow the creation of protein variants better suited to survive in sudden environmental fluctuations.
In yeast, the loss of an active translation termination factor Sup35p leads to stop codon read-through, giving rise to new phenotypes. In bacteria, the same result is accomplished through the loss of the transcriptional terminator factor Rho and thereby a decrease in transcriptional fidelity.

Table 1-1. Known functional amyloids and their wide range of putative functions.

Amyloid	Gene or Protein	Amyloid Function	Species	Year	Reference
Bacteria					
Curli	CsgA	Biofilm formation	E. coli, Salmonella spp, and other Enterobacteriacae	2002	[17]
Microcin E492	Mcc	Toxin storage	Klebsiella pneumoniae	2005	[28]
Hairpins	HpaG	Virulence Factor	Xanthanomas spp.	2007	[29]
Phenol-soluble modulins	PSM	Biofilm formation and virulence	Staphylococcus aureus	2007	[30]
MTP	mtp	Pili formation	Mycobacterium tuberculosis	2007	[31]
FapC	FapC	Biofilm formation	Pseudomonas fluorescens and other Pseudomonads	2010	[32]
TasA	TasA	Biofilm formation	Bacillus subtilis	2010	[33]
P1	P1	Adhesin	Streptococcus mutans	2012	[34]
Listeriolysin	LLO	Phagolysosome	Listeria monocytogenes	2012	[35]
RepA	RepA	Plasmid replication regulator	Pseudomonas aeruginosa	2016	[36]
Chaplins	ChpA-H	Decreasing surface tension	Streptomyces coelicolor	2017	[37]

Rho	Rho	Transcriptional regulator	Clostridium botulinum	2017	[38]
Protists					
MSP2	msp2	Erythrocyte invasion	Plasmodium falciparum	2009	[39]
Fungi					
Het-s	het-s	Heterokaryon formation	Podospora anserina	1997	[40]
Ure2†	Ure2	Nitrogen catabolism regulator	Saccharomyces cerevisiae	1997	[14]
Sup35†	Sup35	Translation regulator	Saccharomyces cerevisiae	1997	[14]
Hydrophobins	SC3, etc	Breaking water surface tension	Schizophyllum commune, basidiomycetes, etc.	2000	[15]
Adhesions	Als Proteins	Biofilm formation	Candida albicans and other fungi	2004	[41]
Plants					
Adhesive substance	unknown	EPS‡ component	Coccomyxa spp., Glaphyrella trebouxiodes, and other microalgae	2008	[42]
Rubber elongation factor	HevB1	latex biosynthesis	Hevea brasiliensis	2012	[43]
Defensins	RsAFP-19	Antifungal defense	Raphanus sativus	2013	[44]
AMP2	Cn-AMP2	Antimicrobial defense	Cocos nucifera	2016	[45]
Animals					

Chorions	Chorions	Egg protection	Bombyx mori, Papilio xuthus, etc.	2000	[16]
Spidroin	Spidroin	Silk production	Nephila edulis, Araneus diadematus, etc.	2002	[46]
CPEB	CPEB	Translation regulator	Aplysia californica	2003	[47]
Pmel17	Pmel17	Melanin synthesis	Homo sapiens	2005	[48]
Cement	cp-100k	surface adhesion	Megabalanus rosa and other barnacles	2006	[49]
Peptide hormones	GLP-2, VIP, etc.	Storage	Homo sapiens	2009	[50]
Anionic dermaseptin	aDrs	host defense	Pachymedusa dacnicolor	2012	[51]
Orb2	Orb2	Memory persistence	Drosophila melanogaster	2012	[52]
epididymal cystatin	cst8, etc.	Sperm maturation	Mus musculus	2012	[53]
RIP kinases	RIP1/3	Necrosis regulator	Homo sapiens	2012	[54]
Uperin 3.5	uperin 3.5	Antimicrobial defense	Uperoleia mjobergii	2016	[55]
Archaea					
Biofilm amyloid protein	HVO_143	Biofilm formation	Haloferax volcanii	2014	[56]

† Only the two first yeast prions that were identified are mentioned here. This is not an inclusive list; there are many more examples of prions performing important functions in yeast.

‡ Extracellular polymeric substance

1.10 References

1. Tanskanen, M. “Amyloid” — Historical Aspects. In Amyloidosis; InTech, 2013.
2. Sipe, J.D.; Cohen, A.S. Review: History of the Amyloid Fibril. J. Struct. Biol. 2000, 130, 88–98, doi:https://doi.org/10.1006/jsbi.2000.4221.
3. Chiti, F.; Dobson, C.M. Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 2006, 75, 333–366, doi:10.1146/annurev.biochem.75.101304.123901.
4. Pham, C.L.L.; Kwan, A.H.; Sunde, M. Functional amyloid: Widespread in Nature, diverse in purpose. Essays Biochem. 2014, 56, 207–219, doi:10.1042/BSE0560207.
5. Otzen, D. Functional amyloid: turning swords into plowshares. Prion 2010, 4, 256–264, doi:10.4161/PRI.4.4.13676.
6. Virchow, R. Ueber eine im Gehirn und Rückenmark des Menschen aufgefundene Substanz mit der chemischen Reaction der Cellulose. Arch. für Pathol. Anat. und Physiol. und für Klin. Med. 1854, 6, 135–138, doi:10.1007/BF01930815.
7. Divry, P. Étude histo-chimique des plaques seniles. J. Neurol. Psychiatr. 1927, 27, 643–57.
8. Pras, M.; Schubert, M.; Zucker-Franklin, D.; Rimon, A.; Franklin, E.C. The characterization of soluble amyloid prepared in water. J. Clin. Invest. 1968, 47, 924–933, doi:10.1172/jci105784.
9. Benson, M.D.; Buxbaum, J.N.; Eisenberg, D.S.; Merlini, G.; Saraiva, M.J.M.; Sekijima, Y.; Sipe, J.D.; Westermark, P. Amyloid nomenclature 2018: recommendations by the International Society of Amyloidosis (ISA) nomenclature committee. Amyloid 2018, 25, 215–219, doi:10.1080/13506129.2018.1549825.
10. Journal Citation Reports Science Edition (Clarivate Analytics, 2020).
11. Tuite, M.F.; Staniforth, G.L.; Cox, B.S. [PSI+] turns 50. Prion 2015, 9, 318–332, doi:10.1080/19336896.2015.1111508.
12. Patino, M.M.; Liu, J.J.; Glover, J.R.; Lindquist, S. Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. Science (80-.). 1996, 273, 622–626.
13. Paushkin, S. V.; Kushnirov, V. V.; Smirnov, V.N.; Ter-Avanesyan, M.D. Propagation of the yeast prion-like [psi+] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor. EMBO J. 1996, 15, 3127–3134, doi:10.1002/j.1460-2075.1996.tb00675.x.
14. Wickner, R.B. [URE3] as an altered URE2 protein: Evidence for a prion analog in Saccharomyces cerevisiae. Science (80-.). 1994, 264, 566–569, doi:10.1126/science.7909170.
15. Wösten, H.A.B.; De Vocht, M.L. Hydrophobins, the fungal coat unravelled. Biochim. Biophys. Acta - Rev. Biomembr. 2000, 1469, 79–86.
16. Iconomidou, V.A.; Vriend, G.; Hamodrakas, S.J. Amyloids protect the silkmoth oocyte and embryo. FEBS Lett. 2000, 479, 141–145, doi:https://doi.org/10.1016/S0014-5793(00)01888-3.
17. Chapman, M.R.; Robinson, L.S.; Pinkner, J.S.; Roth, R.; Heuser, J.; Hammar, M.; Normark, S.; Hultgren, S.J. Role of Escherichia coli curli operons in directing amyloid fiber formation. Science (80-.). 2002, 295, 851–855, doi:10.1126/science.1067484.
18. Nizhnikov, A.A.; Antonets, K.S.; Inge-Vechtomov, S.G. Amyloids: from pathogenesis to function. Biochem. 2015, 80, 1127–1144, doi:10.1134/S0006297915090047.

19. Tayeb-Fligelman, E.; Salinas, N.; Tabachnikov, O.; Landau, M. Staphylococcus aureus PSMα3 Cross-α Fibril Polymorphism and Determinants of Cytotoxicity. Structure 2020, 28, 301-313.e6, doi:10.1016/j.str.2019.12.006.
20. Eisenberg, D.S.; Sawaya, M.R. Structural Studies of Amyloid Proteins at the Molecular Level. Annu. Rev. Biochem. 2017, 86, 69–95, doi:10.1146/annurev-biochem-061516-045104.
21. Jahn, T.R.; Radford, S.E. Folding versus aggregation: Polypeptide conformations on competing pathways. Arch. Biochem. Biophys. 2008, 469, 100–117.
22. Anfinsen, C.B. Principles that govern the folding of protein chains. Science (80-.). 1973, 181, 223–230.
23. Wolynes, &; Karplus & Shakhno-Vitch ; Karplus, ; Folding funnels, binding funnels, and protein function; Gruebele & Wolynes, 1989;
24. Gazit, E. The "Correctly Folded" State of Proteins: Is It a Metastable State? Angew. Chemie Int. Ed. 2002, 41, 257, doi:10.1002/1521-3773(20020118)41:2<257::AID-ANIE257>3.0.CO;2-M.
25. Kelly, J.W. Amyloid fibril formation and protein misassembly: A structural quest for insights into amyloid and priori diseases. Structure 1997, 5, 595–600.
26. Booth, D.R.; Sundetll, M.; Bellotti, V.; Robinson+, C. V; Hutchinson, W.L.; Fraser, P.E.; Hawkins, P.N.; Dobson+, C.M.; Radford+~, S.E.; Blaket, C.C.F.; et al. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis; 1997;
27. Buxbaum, J.N.; Tagoe, C.; Gallo, G.; Walker, J.R.; Kurian, S.; Salomon, D.R. Why are some amyloidoses systemic? Does hepatic "chaperoning at a distance" prevent cardiac deposition in a transgenic model of human senile systemic (transthyretin) amyloidosis? FASEB J. 2012, 26, 2283–2293, doi:10.1096/fj.11-189571.
28. Bieler, S.; Estrada, L.; Lagos, R.; Baeza, M.; Castilla, J.; Soto, C. Amyloid formation modulates the biological activity of a bacterial protein. J. Biol. Chem. 2005, 280, 26880–26885, doi:10.1074/jbc.M502031200.
29. Oh, J.; Kim, J.G.; Jeon, E.; Yoo, C.H.; Jae, S.M.; Rhee, S.; Hwang, I. Amyloidogenesis of type III-dependent harpins from plant pathogenic bacteria. J. Biol. Chem. 2007, 282, 13601–13609, doi:10.1074/jbc.M602576200.
30. Schwartz, K.; Syed, A.K.; Stephenson, R.E.; Rickard, A.H.; Boles, B.R. Functional Amyloids Composed of Phenol Soluble Modulins Stabilize Staphylococcus aureus Biofilms. PLoS Pathog. 2012, 8, e1002744, doi:10.1371/journal.ppat.1002744.
31. Alteri, C.J.; Xicohténcatl-Cortes, J.; Hess, S.; Caballero-Olin, G.; Girón, J.A.; Friedman, R.L. Mycobacterium tuberculosis produces pili during human infection. Proc. Natl. Acad. Sci. U. S. A. 2007, 104, 5145–5150, doi:10.1073/pnas.0602304104.
32. Dueholm, M.S.; Petersen, S. V; Sonderkaer, M.; Larsen, P.; Christiansen, G.; Hein, K.L.; Enghild, J.J.; Nielsen, J.L.; Nielsen, K.L.; Nielsen, P.H.; et al. Functional amyloid in Pseudomonas. Mol Microbiol 2010, 77, 1009–1020, doi:10.1111/j.1365-2958.2010.07269.x.
33. Romero, D.; Aguilar, C.; Losick, R.; Kolter, R. Amyloid fibers provide structural integrity to Bacillus subtilis biofilms. Proc. Natl. Acad. Sci. 2010, 107, 2230–2234, doi:10.1073/PNAS.0910560107.
34. Oli, M.W.; Otoo, H.N.; Crowley, P.J.; Heim, K.P.; Nascimento, M.M.; Ramsook, C.B.; Lipke, P.N.; Brady, L.J. Functional amyloid formation by Streptococcus mutans. Microbiol. (United Kingdom) 2012, 158, 2903–2916, doi:10.1099/mic.0.060855-0.

35. Bavdek, A.; Kostanjšek, R.; Antonini, V.; Lakey, J.H.; Dalla Serra, M.; Gilbert, R.J.C.; Anderluh, G. pH dependence of listeriolysin O aggregation and pore-forming ability. FEBS J. 2012, 279, 126–141, doi:10.1111/j.1742-4658.2011.08405.x.
36. Molina-García, L.; Gasset-Rosa, F.; Moreno-Del Álamo, M.; Fernández-Tresguerres, M.E.; Moreno-Díaz De La Espina, S.; Lurz, R.; Giraldo, R. Functional amyloids as inhibitors of plasmid DNA replication. Sci. Rep. 2016, 6, 1–8, doi:10.1038/srep25425.
37. Dragoš, A.; Kovács, Á.T.; Claessen, D. The Role of Functional Amyloids in Multicellular Growth and Development of Gram-Positive Bacteria. Biomolecules 2017, 7, 60, doi:10.3390/biom7030060.
38. Yuan, A.H.; Hochschild, A. A bacterial global regulator forms a prion. Science (80-.). 2017, 355, 198–201, doi:10.1126/science.aai7776.
39. Adda, C.G.; Murphy, V.J.; Sunde, M.; Waddington, L.J.; Schloegel, J.; Talbo, G.H.; Vingas, K.; Kienzle, V.; Masciantonio, R.; Howlett, G.J.; et al. Plasmodium falciparum merozoite surface protein 2 is unstructured and forms amyloid-like fibrils. Mol. Biochem. Parasitol. 2009, 166, 159–171, doi:10.1016/j.molbiopara.2009.03.012.
40. Coustou, V.; Deleu, C.; Saupe, S.; Begueret, J. The protein product of the het-s heterokaryon incompatibility gene of the fungus Podospora anserina behaves as a prion analog. Proc. Natl. Acad. Sci. 1997, 94, 9773–9778, doi:10.1073/PNAS.94.18.9773.
41. Rauceo, J.M.; Gaur, N.K.; Lee, K.G.; Edwards, J.E.; Klotz, S.A.; Lipke, P.N. Global cell surface conformational shift mediated by a Candida albicans adhesin. Infect. Immun. 2004, 72, 4948–4955, doi:10.1128/IAI.72.9.4948-4955.2004.
42. Mostaert, A.S.; Higgins, M.J.; Fukuma, T.; Rindi, F.; Jarvis, S.P. Nanoscale mechanical characterisation of amyloid fibrils discovered in a natural adhesive. J. Biol. Phys. 2006, 32, 393–401, doi:10.1007/s10867-006-9023-y.
43. Berthelot, K.; Lecomte, S.; Estevez, Y.; Coulary-Salin, B.; Bentaleb, A.; Cullin, C.; Deffieux, A.; Peruch, F. Rubber Elongation Factor (REF), a Major Allergen Component in Hevea brasiliensis Latex Has Amyloid Properties. PLoS One 2012, 7, e48065, doi:10.1371/journal.pone.0048065.
44. Garvey, M.; Meehan, S.; Gras, S.L.; Schirra, H.J.; Craik, D.J.; Van Der Weerden, N.L.; Anderson, M.A.; Gerrard, J.A.; Carver, J.A. A radish seed antifungal peptide with a high amyloid fibril-forming propensity. Biochim. Biophys. Acta - Proteins Proteomics 2013, 1834, 1615–1623, doi:10.1016/j.bbapap.2013.04.030.
45. Gour, S.; Kaushik, V.; Kumar, V.; Bhat, P.; Yadav, S.C.; Yadav, J.K. Antimicrobial peptide (Cn-AMP2) from liquid endosperm of Cocos nucifera forms amyloid-like fibrillar structure. J. Pept. Sci. 2016, 22, 201–207, doi:10.1002/psc.2860.
46. Kenney, J.M.; Knight, D.; Wise, M.J.; Vollrath, F. Amyloidogenic nature of spider silk. Eur. J. Biochem. 2002, 269, 4159–4163, doi:10.1046/j.1432-1033.2002.03112.x.
47. Si, K.; Lindquist, S.; Kandel, E.R. A Neuronal Isoform of the Aplysia CPEB Has Prion-Like Properties. Cell 2003, 115, 879–891, doi:10.1016/S0092-8674(03)01020-1.
48. Fowler, D.M.; Koulov, A. V; Alory-Jost, C.; Marks, M.S.; Balch, W.E.; Kelly, J.W. Functional Amyloid Formation within Mammalian Tissue. PLoS Biol. 2005, 4, e6, doi:10.1371/journal.pbio.0040006.
49. Kamino, K. Barnacle Underwater Attachment. In Biological Adhesives; Springer Berlin Heidelberg, 2006; pp. 145–166.
50. Maji, S.K.; Perrin, M.H.; Sawaya, M.R.; Jessberger, S.; Vadodaria, K.; Rissman, R.A.; Singru, P.S.; Nilsson, K.P.R.; Simon, R.; Schubert, D.; et al. Functional Amyloids As Natural

Storage of Peptide Hormones in Pituitary Secretory Granules. Science (80-.). 2009, 325, 328–332, doi:10.1126/science.1173155.
51. Gößler-Schöfberger, R.; Hesser, G.; Reif, M.M.; Friedmann, J.; Duscher, B.; Toca-Herrera, J.L.; Oostenbrink, C.; Jilek, A. A stereochemical switch in the aDrs model system, a candidate for a functional amyloid. Arch. Biochem. Biophys. 2012, 522, 100–106, doi:10.1016/j.abb.2012.04.006.
52. Majumdar, A.; Cesario, W.C.; White-Grindley, E.; Jiang, H.; Ren, F.; Khan, M. "Repon"; Li, L.; Choi, E.M.-L.; Kannan, K.; Guo, F.; et al. Critical Role of Amyloid-like Oligomers of Drosophila Orb2 in the Persistence of Memory. Cell 2012, 148, 515–529, doi:https://doi.org/10.1016/j.cell.2012.01.004.
53. Whelly, S.; Johnson, S.; Powell, J.; Borchardt, C.; Hastert, M.C.; Cornwall, G.A. Nonpathological Extracellular Amyloid Is Present during Normal Epididymal Sperm Maturation. PLoS One 2012, 7, e36394, doi:10.1371/journal.pone.0036394.
54. Li, J.; McQuade, T.; Siemer, A.B.; Napetschnig, J.; Moriwaki, K.; Hsiao, Y.S.; Damko, E.; Moquin, D.; Walz, T.; McDermott, A.; et al. The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis. Cell 2012, 150, 339–350, doi:10.1016/j.cell.2012.06.019.
55. Calabrese, A.N.; Liu, Y.; Wang, T.; Musgrave, I.F.; Pukala, T.L.; Tabor, R.F.; Martin, L.L.; Carver, J.A.; Bowie, J.H. The Amyloid Fibril-Forming Properties of the Amphibian Antimicrobial Peptide Uperin 3.5. ChemBioChem 2016, 17, 239–246, doi:10.1002/cbic.201500518.
56. Chimileski, S.; Franklin, M.J.; Papke, R.T. Biofilms formed by the archaeon Haloferax volcaniiexhibit cellular differentiation and social motility, and facilitate horizontal gene transfer. BMC Biol. 2014, 12, 65, doi:10.1186/s12915-014-0065-5.
57. Dobson, C.M. Protein folding and misfolding. Nature 2003, 426, 884–890, doi:10.1038/nature02261.
58. Guijarro, J.I.; Sunde, M.; Jones, J.A.; Campbell, I.D.; Dobson, C.M. Amyloid fibril formation by an SH3 domain. Proc. Natl. Acad. Sci. 1998, 95, 4224–4228, doi:10.1073/PNAS.95.8.4224.
59. Chiti, F.; Webster, P.; Taddei, N.; Clark, A.; Stefani, M.; Ramponi, G.; Dobson, C.M. Designing conditions for *in vitro* formation of amyloid protofilaments and fibrils. Proc. Natl. Acad. Sci. 1999, 96, 3590–3594, doi:10.1073/PNAS.96.7.3590.
60. Nguyen, H.D.; Hall, C.K. Kinetics of fibril formation by polyalanine peptides. J. Biol. Chem. 2005, 280, 9074–9082, doi:10.1074/jbc.M407338200.
61. Marchut, A.J.; Hall, C.K. Side-Chain Interactions Determine Amyloid Formation by Model Polyglutamine Peptides in Molecular Dynamics Simulations. Biophys. J. 2006, 90, 4574–4584, doi:https://doi.org/10.1529/biophysj.105.079269.
62. Chiti, F.; Dobson, C.M. Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 2006, 75, 333–366, doi:10.1146/annurev.biochem.75.101304.123901.
63. Smith, D.; Price, J.; Burby, P.; Blanco, L.; Chamberlain, J.; Chapman, M. The Production of Curli Amyloid Fibers Is Deeply Integrated into the Biology of Escherichia coli. Biomolecules 2017, 7, 75, doi:10.3390/biom7040075.
64. Chapman, M.R.; Robinson, L.S.; Pinkner, J.S.; Roth, R.; Heuser, J.; Hammar, M.; Normark, S.; Hultgren, S.J. Role of Escherichia coli curli operons in directing amyloid fiber formation. Science (80-.). 2002, 295, 851–855, doi:10.1126/science.1067484.

65. Evans, M.L.; Chorell, E.; Taylor, J.D.; Åden, J.; Götheson, A.; Li, F.; Koch, M.; Sefer, L.; Matthews, S.J.; Wittung-Stafshede, P.; et al. The Bacterial Curli System Possesses a Potent and Selective Inhibitor of Amyloid Formation. Mol. Cell 2015, 57, 445–455, doi:10.1016/j.molcel.2014.12.025.
66. Nenninger, A.A.; Robinson, L.S.; Hammer, N.D.; Epstein, E.A.; Badtke, M.P.; Hultgren, S.J.; Chapman, M.R. CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation. Mol Microbiol 2011, 81, 486–499, doi:10.1111/j.1365-2958.2011.07706.x.
67. Goyal, P.; Krasteva, P. V; Van Gerven, N.; Gubellini, F.; Van den Broeck, I.; Troupiotis-Tsailaki, A.; Jonckheere, W.; Pehau-Arnaudet, G.; Pinkner, J.S.; Chapman, M.R.; et al. Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. Nature 2014, 516, 250–253, doi:10.1038/nature13768.
68. Hammer, N.D.; Schmidt, J.C.; Chapman, M.R. The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. Proc Natl Acad Sci U S A 2007, 104, 12494–12499, doi:10.1073/pnas.0703310104.
69. Nenninger, A.A.; Robinson, L.S.; Hultgren, S.J. Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. Proc. Natl. Acad. Sci. U. S. A. 2009, 106, 900–905, doi:10.1073/pnas.0812143106.
70. Wang, X.; Hammer, N.D.; Chapman, M.R. The molecular basis of functional bacterial amyloid polymerization and nucleation. J Biol Chem 2008, 283, 21530–21539, doi:10.1074/jbc.M800466200.
71. Dueholm, M.S.; Petersen, S. V.; Sønderkær, M.; Larsen, P.; Christiansen, G.; Hein, K.L.; Enghild, J.J.; Nielsen, J.L.; Nielsen, K.L.; Nielsen, P.H.; et al. Functional amyloid in pseudomonas. Mol. Microbiol. 2010, 77, 1009–1020, doi:10.1111/j.1365-2958.2010.07269.x.
72. Mann, E.E.; Wozniak, D.J. Pseudomonas biofilm matrix composition and niche biology. FEMS Microbiol. Rev. 2012, 36, 893–916, doi:10.1111/j.1574-6976.2011.00322.x.
73. Dueholm, M.S.; Søndergaard, M.T.; Nilsson, M.; Christiansen, G.; Stensballe, A.; Overgaard, M.T.; Givskov, M.; Tolker-Nielsen, T.; Otzen, D.E.; Nielsen, P.H. Expression of Fap amyloids in Pseudomonas aeruginosa, P. fluorescens, and P. putida results in aggregation and increased biofilm formation. Microbiologyopen 2013, 2, 365–382, doi:10.1002/mbo3.81.
74. Hammar, M.; Bian, Z.; Normark, S. Nucleator-dependent intercellular assembly of adhesive curli organelles in Escherichia coli. PNAS 2007, 93, 6562–6566, doi:10.1073/pnas.93.13.6562.
75. Rouse, S.L.; Hawthorne, W.J.; Berry, J.-L.; Chorev, D.S.; Ionescu, S.A.; Lambert, S.; Stylianou, F.; Ewert, W.; Mackie, U.; Morgan, R.M.L.; et al. A new class of hybrid secretion system is employed in Pseudomonas amyloid biogenesis. Nat. Commun. 2017, 8, 263, doi:10.1038/s41467-017-00361-6.
76. Nenninger, A.A.; Robinson, L.S.; Hammer, N.D.; Epstein, E.A.; Badtke, M.P.; Hultgren, S.J.; Chapman, M.R. CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation. Mol. Microbiol. 2011, 81, 486–499, doi:10.1111/j.1365-2958.2011.07706.x.
77. Hall-Stoodley, L.; Costerton, J.W.; Stoodley, P. Bacterial biofilms: From the natural environment to infectious diseases. Nat. Rev. Microbiol. 2004, 2, 95–108.
78. Costerton, J.W.; Stewart, P.S.; Greenberg, E.P. Bacterial biofilms: A common cause of persistent infections. Science (80-.). 1999, 284, 1318–1322.
79. Reichhardt, C.; Cegelski, L. Solid-state NMR for bacterial biofilms. Mol. Phys. 2014, 112, 887–894, doi:10.1080/00268976.2013.837983.

80. Cegelski, L.; Pinkner, J.S.; Hammer, N.D.; Cusumano, C.K.; Hung, C.S.; Chorell, E.; Åberg, V.; Walker, J.N.; Seed, P.C.; Almqvist, F.; et al. Small-molecule inhibitors target Escherichia coli amyloid biogenesis and biofilm formation. Nat. Chem. Biol. 2009, 5, 913–919, doi:10.1038/nchembio.242.
81. Schwartz, K.; Syed, A.K.; Stephenson, R.E.; Rickard, A.H.; Boles, B.R. Functional amyloids composed of phenol soluble modulins stabilize Staphylococcus aureus biofilms. PLoS Pathog 2012, 8, e1002744, doi:10.1371/journal.ppat.1002744.
82. Cox, B.S. Ψ, A cytoplasmic suppressor of super-suppressor in yeast. Heredity (Edinb). 1965, 20, 505–521, doi:10.1038/hdy.1965.65.
83. Derkatch, I.L.; Bradley, M.E.; Hong, J.Y.; Liebman, S.W. Prions Affect the Appearance of Other Prions: The Story of [PIN] ing of a nonprion molecule into the prion shape or by the chance interaction of two or more nonprion molecules. Either event is more likely when the protein is present; 2001; Vol. 106;.
84. Derkatch, I.L.; Bradley, M.E.; Masse, S. V; Zadorsky, S.P.; Polozkov, G. V; Inge-Vechtomov, S.G.; Liebman, S.W. Dependence and independence of [PSI(+)] and [PIN(+)]: a two-prion system in yeast? EMBO J. 2000, 19, 1942–1952, doi:10.1093/emboj/19.9.1942.
85. Chernoff, Y.O.; Lindquist, S.L.; Ono, B.; Inge-Vechtomov, S.G.; Liebman, S.W. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. Science (80-.). 1995, 268, 880, doi:10.1126/science.7754373.
86. Paushkin, S. V; Kushnirov, V. V; Smirnov, V.N.; Ter-Avanesyan', M.D. Propagation of the yeast prion-like [psi+] determinant is mediated by oligomerization of the SUP35-encoded polypeptide chain release factor; 1996; Vol. 15;.
87. Frolova, L.; Le Goff, X.; Rasmussen, H.H.; Cheperegin, S.; Drugeon, G.; Kress, M.; Arman, I.; Haenni, A.-L.; Celis, J.E.; Phllippe, M.; et al. A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. Nature 1994, 372, 701–703, doi:10.1038/372701a0.
88. Derkatch, I.L.; Bradley, M.E.; Liebman, S.W. Overexpression of the SUP45 gene encoding a Sup35p-binding protein inhibits the induction of the de novo appearance of the [PSI] prion; 1998; Vol. 95;.
89. True, H.L.; Lindquist, S.L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. Nature 2000, 407, 477–483, doi:10.1038/35035005.
90. Tuite, M.F.; Cox, B.S.; McLaughlin, C.S. A ribosome-associated inhibitor of *in vitro* nonsense suppression in [psi-] strains of yeast. FEBS Lett. 1987, 225, 205–208, doi:10.1016/0014-5793(87)81158-4.
91. Tuite, M.F.; Cox, B.S.; McLaughlin, C.S. *In vitro* nonsense suppression in [psi+] and [psi-] cell-free lysates of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. 1983, 80, 2824–2828, doi:10.1073/pnas.80.10.2824.
92. Yuan, A.H.; Hochschild, A. A bacterial global regulator forms a prion. Science (80-.). 2017, 355, 198–201, doi:10.1126/science.aai7776.
93. Hurbain, I.; Geerts, W.J.C.; Boudier, T.; Marco, S.; Verkleij, A.J.; Marks, M.S.; Raposo, G. Electron tomography of early melanosomes: Implications for melanogenesis and the generation of fibrillar amyloid sheets. Proc. Natl. Acad. Sci. 2008, 105, 19726–19731, doi:10.1073/PNAS.0803488105.
94. van Niel, G.; Charrin, S.; Simoes, S.; Romao, M.; Rochin, L.; Saftig, P.; Marks, M.S.; Rubinstein, E.; Raposo, G. The Tetraspanin CD63 Regulates ESCRT-Independent and -

Dependent Endosomal Sorting during Melanogenesis. Dev. Cell 2011, 21, 708–721, doi:10.1016/j.devcel.2011.08.019.
95. van Niel, G.; Bergam, P.; Di Cicco, A.; Hurbain, I.; Lo Cicero, A.; Dingli, F.; Palmulli, R.; Fort, C.; Potier, M.C.; Schurgers, L.J.; et al. Apolipoprotein E Regulates Amyloid Formation within Endosomes of Pigment Cells. Cell Rep. 2015, 13, 43–51, doi:10.1016/j.celrep.2015.08.057.
96. Berson, J.F.; Harper, D.C.; Tenza, D.; Raposo, G.; Marks, M.S. Pmel17 initiates premelanosome morphogenesis within multivesicular bodies. Mol. Biol. Cell 2001, 12, 3451–3464, doi:10.1091/mbc.12.11.3451.
97. Berson, J.F.; Theos, A.C.; Harper, D.C.; Tenza, D.; Raposo, G.; Marks, M.S. Proprotein convertase cleavage liberates a fibrillogenic fragment of a resident glycoprotein to initiate melanosome biogenesis. J. Cell Biol. 2003, 161, 521–533, doi:10.1083/jcb.200302072.
98. Rochin, L.; Hurbain, I.; Serneels, L.; Fort, C.; Watt, B.; Leblanc, P.; Marks, M.S.; Strooper, B. De; Raposo, G.; Niel, G. van BACE2 processes PMEL to form the melanosome amyloid matrix in pigment cells. Proc. Natl. Acad. Sci. 2013, 110, 10658–10663, doi:10.1073/PNAS.1220748110.
99. Kummer, M.P.; Maruyama, H.; Huelsmann, C.; Baches, S.; Weggen, S.; Koo, E.H. Formation of Pmel17 amyloid is regulated by juxtamembrane metalloproteinase cleavage, and the resulting C-terminal fragment is a substrate for γ-secretase. J. Biol. Chem. 2009, 284, 2296–2306, doi:10.1074/jbc.M808904200.
100. Watt, B.; Van Niel, G.; Fowler, D.M.; Hurbain, I.; Luk, K.C.; Stayrook, S.E.; Lemmon, M.A.; Raposo, G.; Shorter, J.; Kelly, J.W.; et al. N-terminal Domains Elicit Formation of Functional Pmel17 Amyloid Fibrils * □ S. 2009, doi:10.1074/jbc.M109.047449.
101. Hartl, F.U. Protein Misfolding Diseases. Annu. Rev. Biochem. 2017, 86, 21–26, doi:10.1146/annurev-biochem-061516-044518.
102. Labbadia, J.; Morimoto, R.I. The Biology of Proteostasis in Aging and Disease. Annu. Rev. Biochem. 2015, 84, 435–464, doi:10.1146/annurev-biochem-060614-033955.
103. Prusiner, S.B. Novel proteinaceous infectious particles cause scrapie. Science (80-.). 1982, 216, 136–144, doi:10.1126/science.6801762.
104. Hawkes, C.H.; Del Tredici, K.; Braak, H. Parkinson's disease: a dual-hit hypothesis. Neuropathol. Appl. Neurobiol. 2007, 33, 599–614, doi:10.1111/j.1365-2990.2007.00874.x.
105. Sampson, T.R.; Challis, C.; Jain, N.; Moiseyenko, A.; Ladinsky, M.S.; Shastri, G.G.; Thron, T.; Needham, B.D.; Horvath, I.; Debelius, J.W.; et al. A gut bacterial amyloid promotes a-synuclein aggregation and motor impairment in mice. Elife 2020, 9, doi:10.7554/eLife.53111.
106. Sampson, T.R.; Debelius, J.W.; Thron, T.; Janssen, S.; Shastri, G.G.; Ilhan, Z.E.; Challis, C.; Schretter, C.E.; Rocha, S.; Gradinaru, V.; et al. Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. Cell 2016, 167, 1469-1480.e12, doi:10.1016/J.CELL.2016.11.018.
107. Challis, C.; Hori, A.; Sampson, T.R.; Yoo, B.B.; Challis, R.C.; Hamilton, A.M.; Mazmanian, S.K.; Volpicelli-Daley, L.A.; Gradinaru, V. Gut-seeded α-synuclein fibrils promote gut dysfunction and brain pathology specifically in aged mice. Nat. Neurosci. 2020, 23, 327–336, doi:10.1038/s41593-020-0589-7.
108. Chen, C.; Ahn, E.H.; Kang, S.S.; Liu, X.; Alam, A.; Ye, K. Gut dysbiosis contributes to amyloid pathology, associated with C/EBPβ/AEP signaling activation in Alzheimer's disease mouse model. Sci. Adv. 2020, 6, eaba0466, doi:10.1126/sciadv.aba0466.

109. Gejyo, F.; Yamada, T.; Odani, S.; Nakagawa, Y.; Arakawa, M.; Kunitomo, T.; Kataoka, H.; Suzuki, M.; Hirasawa, Y.; Shirahama, T.; et al. A new form of amyloid protein associated with chronic hemodialysis was identified as β2-microglobulin. Biochem. Biophys. Res. Commun. 1985, 129, 701–706, doi:10.1016/0006-291X(85)91948-5.

Chapter 2 Tuning Functional Amyloid Formation through Disulfide Engineering[2]

2.1 Abstract

Many organisms produce 'functional' amyloid fibers, which are stable protein polymers that serve many roles in cellular biology. *E. coli* and other Enterobacteriaceae assemble functional amyloid fibers called curli that are the main protein component of the biofilm extracellular matrix. CsgA is the major protein subunit of curli and will rapidly adopt the polymeric amyloid conformation *in vitro*. The rapid and irreversible nature of CsgA amyloid formation makes it challenging to study *in vitro*. Here, we engineered CsgA so that amyloid formation could be tuned to the redox state of the protein. A double cysteine variant of CsgA called $CsgA_{CC}$ was created and characterized for its ability to form amyloid. When kept under oxidizing conditions, $CsgA_{CC}$ did not adopt a β-sheet rich structure or form detectable amyloid-like aggregates. Oxidized $CsgA_{CC}$ remained in a soluble, non-amyloid state for at least 90 days. The addition of reducing agents to $CsgA_{CC}$ resulted in amyloid formation within hours. The amyloid fibers formed by $CsgA_{CC}$ were indistinguishable from the fibers made by CsgA WT. When measured by thioflavin T fluorescence the amyloid formation by $CsgA_{CC}$ in the reduced form displayed the same lag, fast, and plateau phases as CsgA WT. Amyloid formation by $CsgA_{CC}$ could be halted by the addition of oxidizing agents. Therefore, $CsgA_{CC}$ serves as a proof-

[2] This chapter is associated with the following publication: Balistreri, A., Kahana, E., Janakiraman, S., and Chapman, M.R. (2020). Tuning Functional Amyloid Formation Through Disulfide Engineering. Front. Microbiol. *11*. AB and MC contributed conception and design of the study. AB, EK, and SJ performed the experiments. AB wrote the manuscript; AB and MC contributed to editing and revising the manuscript.

of-concept for capitalizing on the convertible nature of disulfide bonds to control the aggregation of amyloidogenic proteins.

2.2 Introduction

Amyloids are fibril aggregates of proteins which have misfolded to form a characteristic cross β-sheet secondary structure (Chiti and Dobson, 2006). The fibers deposit in cells in dense bodies known as plaques and are implicated in many well-known neurodegenerative diseases including Alzheimer, Parkinson, and Huntington Disease (Chiti and Dobson, 2017). For this reason the nature of amyloids and how they form is an active area of research. Though the disease causing amyloids are composed of misfolded proteins, there are now many examples of "functional amyloids" that organisms create for a predetermined purpose (Chapman et al., 2002; Otzen, 2010). These fibers are composed of proteins that adopt the amyloid fold not by misfolding, rather they do so intentionally and efficiently (Deshmukh et al., 2018). Functional amyloids provide researchers with well-defined and adaptable models for studying the basic tenets of amyloid formation.

Curli are functional amyloids produced by *E. coli* and other Enterobacteriaceae (Dueholm et al., 2012). Curli amyloids are the main proteinaceous component of the biofilm extracellular matrix (Chapman et al., 2002; Reichhardt and Cegelski, 2014). The major protein subunit of curli, called CsgA, is secreted in a monomeric, unstructured state (Wang et al., 2007). CsgA undergoes self-assembly into curli fibers on the cell surface with the help of several auxiliary proteins belonging to the *csg* operon (Chapman et al., 2002; Robinson et al., 2006; Hammer et al., 2007; Wang et al., 2008; Nenninger et al., 2009, 2011; Evans et al., 2015; Jain and Chapman, 2019). Purified CsgA readily forms amyloid fibers and there are several methods to monitor *in vitro* amyloid aggregation (Wang et al., 2007; Evans et al., 2018). However, these techniques

require handling protein samples which are in a constant state of aggregation (Evans et al., 2018).

To minimize CsgA aggregation into amyloid during purification, the current methodology relies on chemical denaturants (8M guanidinium hydrochloride) (Zhou et al., 2013). Though this protocol provides useable amounts of purified CsgA, there are still several purification steps that need to take place under non-denaturing conditions. During the final size exclusion and buffer exchange steps there is an unavoidable loss of soluble CsgA which is tolerated in exchange for speediness.

Here, we employ a different strategy to restrict amyloid formation and improve purification. CsgA is an intrinsically disordered protein that transitions into a β-sheet rich conformation upon forming amyloid fibers (Wang et al., 2007). Therefore, a potential strategy for controlling aggregation is targeting this transition. Disulfide bonds are tertiary structural components of many natively folded proteins that aid in protein folding and provide stability (Anfinsen and Scheraga, 1975). They are also by nature convertible; they can be broken and reformed based on the redox state of the environment (Gilbert, 1995). By engineering two cysteine residues into its sequence, we can provide CsgA the option to form an intramolecular disulfide bond under the right conditions. Previous work has shown disulfide engineering can be used to stabilize the native fold state of a protein (Matsumurat et al., 1989; Dombkowski et al., 2014; Schmidt et al., 2015; Liu et al., 2016). Recently, other groups have used this technique to modulate amyloid formation of human amyloid proteins (Hoyer et al., 2008; Carija et al., 2019). In this study we harness disulfide engineering as a method for stabilizing the non-native fold state of a protein with the intention of triggering protein folding.

We created a double cysteine variant of CsgA called $CsgA_{CC}$ with the goal of making CsgA amyloid formation tunable to its redox state. CsgA contains five imperfect repeat sequences (R1-R5) corresponding to five β-strands that exist in the final folded protein (**Figure 2-1A**) (Wang et al., 2007). β-strands R1 and R5 are critical to the ability of CsgA to form amyloid (Wang et al., 2008). We hypothesized that the presence of an unnatural disulfide linkage near R1 and R5 would hinder CsgA from forming the secondary structure required for amyloid formation. In order to allow amyloid formation to eventually occur, regions of the sequence essential to proper folding needed to remain undisturbed. Therefore, the β-turn regions that flank R1 and R5 offered an amenable target region. The sequence of CsgA does not contain any native cysteine residues (Barnhart and Chapman, 2006). We decided to replace two residues of similar size and properties to cysteine in the regions of interest. Alanine-63 and valine-140 fulfilled the criteria and were replaced with cysteine residues (**Figure 2-1B**). As we hypothesized, we found that the ability of $CsgA_{CC}$ to form amyloid could be stimulated by the addition of a reducing agent, and that when kept in an oxidized form, $CsgA_{CC}$ remained in a soluble and non-amyloid conformation.

2.3 Materials and Methods

Bacterial Growth. All overnight cultures were grown in LB supplemented with 100 μg/mL ampicillin or 50 μg/mL of kanamycin at 37 °C with shaking at 220 rpm. When necessary, LB plates were supplemented with ampicillin 100 μg/mL or kanamycin 50 μg/mL. To promote curli formation, liquid cultures were normalized to OD_{600}=1.0 and 4 μL were spotted on YESCA agar (yeast extract casamino acids) plates and grown for 48 hours at 26°C (Evans et al., 2018).

Strains and Plasmids. The full list of strains, plasmids, and primers can be found in the Supplemental Materials. Site directed mutagenesis was performed on pre-existing plasmids using

the Agilent QuikChange II XL kit (Cat No. NC9045762). Primers were designed by Agilent QuikChange Prime Design (https://www.agilent.com/store/primerDesignProgram.jsp) and purchased by IDT (https://www.idtdna.com). Mutagenized plasmids were constructed in the MC1061 cell background. Correct mutations were confirmed using Sanger sequencing provided by UM Biomedical Research Core (https://brcf.medicine.umich.edu/). Plasmids were extracted from transformants using Promega PureYield™ Plasmid Miniprep System (Cat No. PRA1223). Miniprep plasmids were transformed into an expression strain (NEB3016), curli-competent strain (MC4100), or a curli-incompetent strain that lacks *csgA* (LSR10).

Protein Purification. CsgA and its variants were purified as described previously (Zhou et al., 2012). CsgC was purified as described previously (Salgado et al., 2011). Size exclusion chromatography was performed as previously described (Evans et al., 2018). Briefly, Ni-NTA affinity chromatography elution fractions were pooled, concentrated, and passed through a 0.22 μm filter. The sample underwent gel filtration using a Superdex 75 10/300 GL column (Cat No.45-002-903) attached to an Äkta pure protein purification system. Elution fractions were assayed for protein concentration using A220. Samples corresponding to elution peaks were analyzed using reducing (with added β-mercaptoethanol) and non-reducing (without β-mercaptoethanol) SDS-PAGE.

Circular Dichroism (CD). $CsgA_{CC}$ was purified and the final samples were stored at 4°C in 50 mM KP_i, pH 7.3. Secondary structure was analyzed using a Jasco J-1500 CD Spectrometer by scanning from 190 to 260 nm at 25°C in a quartz cell with a 1 cm path. Selected samples were reduced with the addition of tris(2-carboxyethyl)phosphine (TCEP) from a stock solution (200 mM TCEP in MQ water, pH= 7.7) to a final concentration of 8 mM. The spectrum of the appropriate buffer-only control was subtracted from each sample as a baseline.

Denaturing Gel Electrophoresis and Western Blot. Purified protein samples were diluted in 4X SDS loading buffer with or without β-mercaptoethanol and run on a 15% SDS PAGE gel (Evans et al., 2018). The gels were stained with Coomassie blue dye to visualize protein bands or the proteins were transferred to a PVDF membrane for blotting. Western Blot were performed as previously described (Evans et al., 2018). Briefly, blots were probed with a primary antibody against CsgA (1:12,000). Secondary antibodies against rabbit IgG and conjugated with IRDye 800CW (Cat No.NC9401842) were used to image the blots in a Licor Odyessey FC.

Gel Solubility Assay. Freshly purified samples of CsgA WT and $CsgA_{CC}$ were transferred to a microcentrifuge tube, diluted to 20 μM in 50 mM phosphate buffer pH = 7.3, and allowed to incubate at room temperature. Selected samples were reduced with the addition of tris(2-carboxyethyl)phosphine (TCEP) from a stock solution (200 mM TCEP in MQ water, pH= 7.7) to a final concentration of 8 mM. Gel samples were removed at 0, 24, and 48 hours after purification. The gel samples were diluted in 4X SDS loading buffer and run on a 15% SDS PAGE gel (Evans et al., 2018). The gels were stained with Coomassie blue dye. Destained gels were scanned using the 700 nm channel of a Licor Odyessey FC. Band intensity was quantified using ImageJ (https://imagej.nih.gov/ij/) and normalized to the CsgA WT time 0 no TCEP added condition.

Transmission Electron Microscopy. TEM images were produced as previously described (Jain et al., 2017). Briefly, 5 μL of purified protein samples were spotted onto a formvar/carbon 200 mesh copper grid (Cat No.50-260-38). After a 5 min incubation, grids were spotted with 5 μL of DI water followed by 5 μL of 2% uranyl acetate to provide micrograph image contrast. Grids were imaged using a Jeol JEM 1400plus Transmission Electron Microscope.

Thioflavin T Binding Assay. Assays were performed as previously described (Zhou et al.,

2012). Briefly, freshly purified CsgA was diluted with phosphate buffer to 20 μM and combined with an excess of the amyloid-specific dye thioflavin-T (Th-T) (Cat No. AC211760250). Selected samples were reduced with the addition of tris(2-carboxyethyl)phosphine (TCEP) from a stock solution (200 mM TCEP in MQ water, pH= 7.7). Amyloid formation was monitored by measuring an increase in Th-T fluorescence at 495 nm (450 nm excitation). Assays were performed in triplicate at microscale within 96-well plates and measured with Infinite Pro M200 or Infinite Nano^{+} F200 Tecan plate readers. Selected samples include: FN075 diluted from a 50 mM stock solution (DMSO) sourced from the Almqvist Lab (Department of Chemistry, University of Umeå, Sweden), CsgC purified and diluted from a stock solution (phosphate buffer), CsgA WT fibers purified previously and sonicated directly before addition, and hydrogen peroxide (Cat No. H325-100).

Complementation Assay. Congo Red binding assays were performed as previously described (Zhou et al., 2012). Briefly, liquid cultures normalized to OD = 1 were spotted onto a YESCA plate supplemented with 50 μg/mL Congo Red (Cat No. AC110501000). The plates were incubated at 26°C for 48 hours to allow for curli formation. A red colony color indicated a strain capable of forming mature curli fibers.

2.4 Results

Immediately after purification CsgA is intrinsically disordered with a random coil secondary structure. Within hours, CsgA aggregates into stable, β-rich amyloid fibers (Wang et al., 2007). To control amyloid aggregation by CsgA we hypothesized that strategically-placed cysteine residues would allow the generation of a CsgA molecule that could be locked in a non-amyloid state (**Figure 2-1A**). The wildtype sequence of *E. coli* CsgA does not contain native cysteine residues (**Figure 2-1A**). Cysteine residues were engineered into CsgA at positions alanine-63

and valine-140 using site-directed mutagenesis. The resulting protein, called $CsgA_{CC}$, was purified and characterized. From this point forward native CsgA will be referred to as CsgA WT and CsgA that contains cysteine residues at positions 63 and 140 will be referred to as $CsgA_{CC}$.

We found that freshly purified and oxidized $CsgA_{CC}$ remained SDS soluble for at least 48 hours (**Figure 2-2A**). For comparison, freshly purified CsgA WT displays a marked decrease in SDS solubility over a 48 hour period (**Figure 2-2A**). Because $CsgA_{CC}$ remained soluble, size exclusion chromatography could be used to determine what types of protein species were present after purification. Freshly purified $CsgA_{CC}$ was subjected to gel filtration and the protein eluted off the column in two wide peaks at 8.56 (peak 1) and 11.24 mL (peak 2) elution volume (**Figure 2-2B**). Analytical gel filtration chromatography can normally be used to provide an estimate of an analyte's size by comparing column retention times with known globular proteins in a standard curve (Whitaker, 1963). However, because $CsgA_{CC}$ remained unstructured, the globular proteins did not provide comparable retention profiles (Sambi et al., 2010). We therefore used non-reducing SDS-PAGE analysis to view what size protein species were in solution (**Figure 2-2C**). The second and larger peak corresponded to the monomeric species of $CsgA_{CC}$ (**Figure 2-2C**). When samples from peak 1 of the gel filtration were analyzed by non-reducing SDS-PAGE, there were several slowly migrating bands (**Figure 2-2C**). When the same samples from peak 1 were analyzed by SDS-PAGE that included a reducing agent, β-mercaptoethanol, many of the high molecular weight bands disappeared, suggesting that $CsgA_{CC}$ forms homo-oligomers that disassociate when reduced (**Figure 2-3)**. The monomeric species of $CsgA_{CC}$ incubated under oxidized conditions ran faster on the SDS-PAGE gel than CsgA WT prepared in the same manner (**Figure 2-2C**). It is possible that in a non-reducing environment the intramolecular

disulfide bond that holds $CsgA_{CC}$ in a constrained conformation results in the faster migration of $CsgA_{CC}$ relative to CsgA WT.

Oxidized $CsgA_{CC}$ can remain in a non-aggregated and non-amyloid form for much longer than CsgA WT. As measured by circular dichroism (CD), $CsgA_{CC}$ remained in a random coil conformation for at least 88 days when stored under oxidizing conditions at 4°C (**Figure 2-4A**). When the reducing agent TCEP was added to $CsgA_{CC}$ after 4 days of incubation under oxidizing conditions, $CsgA_{CC}$ transitioned from random coil to β-sheet rich secondary structure (**Figure 2-4B**). This suggested that the redox state of $CsgA_{CC}$ is a critical determinant of its secondary structure.

The Thioflavin-T (Th-T) binding assay is the most widely used method for monitoring *in vitro* amyloid formation in real time (Nilsson, 2004). CsgA WT amyloid formation displays a characteristic sigmoidal curve during which Th-T fluorescence begins to increase after a 2-4 hour lag phase (Wang et al., 2007). Importantly, $CsgA_{CC}$ showed no detectable increase in Th-T fluorescence until after the addition of a reducing agent (**Figure 2-5A**). After treatment with reducing agents TCEP or DTT, $CsgA_{CC}$ was able to form amyloid fibers (**Figure 2-5B and 2-6**). Through all experiments there was an observable decrease in overall fluorescence seen in $CsgA_{CC}$ Th-T binding assays when compared to CsgA WT (**Figures 2-5, 5C, and 8**). To verify that intramolecular disulfide bonds played a role in inhibiting $CsgA_{CC}$ amyloid formation, we characterized the two single cysteine variants' ability to form amyloid. The introduction of a single cysteine mutation at either position 63 or 140 was only sufficient to slow down amyloid formation under oxidized conditions. A63C (**Figure 2-7A**) and V140C (**Figure 2-7B**) formed amyloid with and without the treatment of reducing agent, displaying an extended lag phase when compared to wildtype. Though V140C showed a lag phase that was much longer than

A63C (**Figure 2-8**). It should be noted that the addition of a reducing agent has no effect on the rate of CsgA WT amyloid formation (**Figure 2-7 and 2-6B**). In an experiment where $CsgA_{CC}$ amyloid formation had reached a fast growth phase, it could be prematurely stopped after the addition of an oxidizing agent like hydrogen peroxide (**Figure 2-9**). These experiments confirmed that $CsgA_{CC}$ amyloid formation is susceptible to change given the redox state of its environment.

The ability of $CsgA_{CC}$ to form amyloid *in vivo* was assessed. Curli-producing *E. coli* display a red colony phenotype when grown on a plate supplemented with the amyloid-binding dye Congo Red (Hammar et al., 1995; Evans et al., 2018). A Δ*csgA* mutant that can no longer produce curli forms white colonies on Congo Red indicator plates (Hammar et al., 1995; Evans et al., 2018). The red colony phenotype can be rescued in a Δ*csgA* strain when CsgA WT is supplied by a plasmid containing the *csgA* gene downstream of the natural promoter (Hammar et al., 1995; Evans et al., 2018) (**Figure 2-10A**). When Δ*csgA* strains were complemented with a plasmid encoding the variants CsgA A63C, CsgA V140C, and $CsgA_{CC}$ the resulting colonies were white, red, and white respectively (**Figure 2-10A**). Whole cell western blot analysis showed the strain producing V140C was the only strain to harbor protein recognized by a αCsgA antibody (**Figure 2-10B**). Therefore, cysteine residues at the 63rd position or at positions 63 and 140 are more disruptive to curli formation than a single cysteine residue in the valine-140 position. The combination of color phenotype and negative western blot results found in the A63C and $CsgA_{CC}$ samples coincided with previously published work describing Δ*csgG* and Δ*csgE* mutants (Robinson et al., 2006; Nenninger et al., 2011). It is possible that CsgA A63C and $CsgA_{CC}$ are confined to the periplasm and subjected to proteolytic degradation in the same

manner as CsgA WT in Δ*csgG* and Δ*csgE* mutants (Robinson et al., 2006; Nenninger et al., 2011).

We also determined whether reduced $CsgA_{CC,}$ or $CsgA_{CC}$+TCEP, interacted with certain proteins and small molecules in a similar manner to CsgA WT. Amyloid formation by CsgA WT is inhibited by the CsgC chaperone and also by a 2-pyridone called FN075 (Cegelski et al., 2009; Evans et al., 2015). We incubated CsgA WT and $CsgA_{CC}$+TCEP with inhibitory concentrations of FN075 and CsgC and monitored amyloid formation (**Figure 2-11A-B**). Both CsgA WT and $CsgA_{CC}$+TCEP amyloid formation were inhibited *in vitro* with similar efficiencies (**Figure 2-11A-B**). Conversely, amyloid formation can be sped up through a "seeding" effect when a small amount of preformed fibers are introduced to freshly purified CsgA WT (Wang et al., 2007). $CsgA_{CC}$+TCEP amyloid formation was similarly sped up when mixed with CsgA WT fibers (**Figure 2-11C**). Finally, $CsgA_{CC}$ +TCEP formed fibril structures of identical morphology with wildtype fibers as confirmed by transmission electron microscopy (**Figure 9**). Fibers could not be found in oxidized $CsgA_{CC}$ samples (**Figure 9**). This line of evidence suggested that $CsgA_{CC}$, after being reduced, behaves similarly to CsgA WT *in vitro.*

2.5 Discussion

CsgA WT can transition from a mono dispersed, random coil structure to β-rich amyloid fibers within about three hours (Wang et al., 2007). Indeed, the amyloid fiber conformation represents a very favorable energy minimum for amyloidogenic proteins (Eisenberg and Jucker, 2012). $CsgA_{CC}$ incubated under oxidizing conditions remains in a random coil, non-amyloidogenic state for at least 88 days (**Figure 2-4**). The oxidized disulfide bond "kinetically traps" $CsgA_{CC}$ in a fold state of relatively high energy compared to the amyloid conformation into which each protein would naturally deposit (**Figure 2-13**). In a reducing environment, the

transition of $CsgA_{CC}$ into the amyloid conformation required no additional heat or chemical energy at all (**Figure 2-4C**). Moreover $CsgA_{CC}$ exhibits a consistent Th-T lag phase of no shorter than 2-4 hours after reduction (**Figures 2-5, 2-7C, and 2-9**). Therefore, when $CsgA_{CC}$ is "kinetically trapped" in an oxidized environment, it must be trapped in a way that prevents nucleus formation. The introduction of a nucleating species, such as sonicated fibers, to amyloid-competent proteins provides a dramatic decrease in an energy barrier leading to rapid amyloid formation (Gosal et al., 2005). However, providing oxidized $CsgA_{CC}$ with preformed fibers did not spark amyloid formation (**Figure 2-14**). $CsgA_{CC}$ transitions from an amyloid-incompetent species to an amyloid-competent species only after the disulfide bond is reduced (**Figure 2-13**).

Here, we have shown $CsgA_{CC}$ forms amyloid only when incubated in the presence of a reducing agent (**Figures 2-5, 2-7C, and 2-9**), affording a method to tune $CsgA_{CC}$ aggregation depending on the redox state. $CsgA_{CC}$ amyloid formation can be triggered by varying amounts of a reducing agent in a dose-dependent manner (**Figure 2-5**). The ability to tune $CsgA_{CC}$ amyloid formation makes $CsgA_{CC}$ a useful tool for studying amyloid formation because it allow for control over a process that is normally left to occur spontaneously (Lomakin et al., 1996). By remaining soluble for longer, $CsgA_{CC}$ can overcome some of the issues with production scale and storage that plague CsgA WT and other amyloidogenic proteins.

$CsgA_{CC}$ could lend itself to understanding the nucleation phase, which is currently poorly understood for functional amyloids like CsgA WT (Arosio et al., 2015). CsgA WT amyloid formation begins directly after purification (Wang et al., 2007; Zhou et al., 2013). By the time we start to monitor CsgA WT amyloid formation by Th-T fluorescence, aggregation and nuclei formation has already started (Sleutel et al., 2017). On the contrary, $CsgA_{CC}$ amyloid formation can be started and stopped at any desired time by the addition of a reducing or oxidizing agent

(**Figure 2-9**). The ability to halt amyloid formation at different points in the process could allow the characterization of amyloid fiber intermediates. Thus the earliest amyloid processes such as the transition of monomers to oligomers or the formation of nucleating species can be better resolved.

After the addition of an oxidizing agent to $CsgA_{CC}$+TCEP, there is no decrease in Th-T fluorescence observed, but instead the Th-T positive species that had formed seem to persist for at least 12 hours (**Figure 2-9**), suggesting two things. First, during the Th-T growth phase there must be monomers in solution that are still susceptible to forming a disulfide bond that makes them amyloid-incompetent. Second, in a model of amyloid formation wherein the fiber end is in a state of equilibrium with free protein monomers, there would be a k_{on} and k_{off} that described the addition/subtraction of monomers to/from the fiber. If this were the case, then monomers that subtract from the fiber could become susceptible to oxidation. The excess of oxidizing agent would greatly shift the equilibrium in the direction of free, oxidized protein monomers and fibers would begin to unravel at the fiber ends. Because the curve in **Figure 2-9** does not show a time-dependent decrease in Th-T fluorescence, it appears that oxidized $CsgA_{CC}$ amyloid fibers and their fiber ends are stable over at least 12 hours.

2.6 Conclusions.

In this study we have shown the ability to tune aggregation of a functional amyloid protein by engineering a disulfide bond into CsgA from *E. coli*. $CsgA_{CC}$ remains in a soluble, non-amyloidogenic state until the disulfide is reduced (**Figures 2-2A and 2-5)**. Once reduced, amyloid formation can be manipulated depending on the redox state of the environment (**Figure 2-9**). $CsgA_{CC}$ is a tool that makes studying amyloid formation easier and offers researchers extra time and control, two highly valuable resources. Purified $CsgA_{CC}$ samples can now be prepared

ahead of time, concentrated, and stored. This method of using disulfide engineering to control amyloid formation could be translated to other functional and human diseases-causing amyloids. Furthermore, the growing field of bio-inspired nanotechnology is capitalizing on the self-assembly of amyloid proteins. CsgA has been shown to have range of uses such as the passage of electric current, the extraction of rare earth metals, and many others (Nguyen et al., 2014; Seker et al., 2017; Tay et al., 2018). The work presented here will allow material scientists to use amyloid fibers to build systems of higher complexity than is currently possible.

Author Contributions

AB and MC contributed conception and design of the study. AB, EK, and SJ performed the experiments. AB wrote the manuscript; AB and MC contributed to editing and revising the manuscript. We thank members of the Chapman lab for helpful suggestions.

Funding

This work was supported by the National Institutes of Health grant no. R01 GM118651 and R21 AI137535 to M.R.C.

2.7 Figures

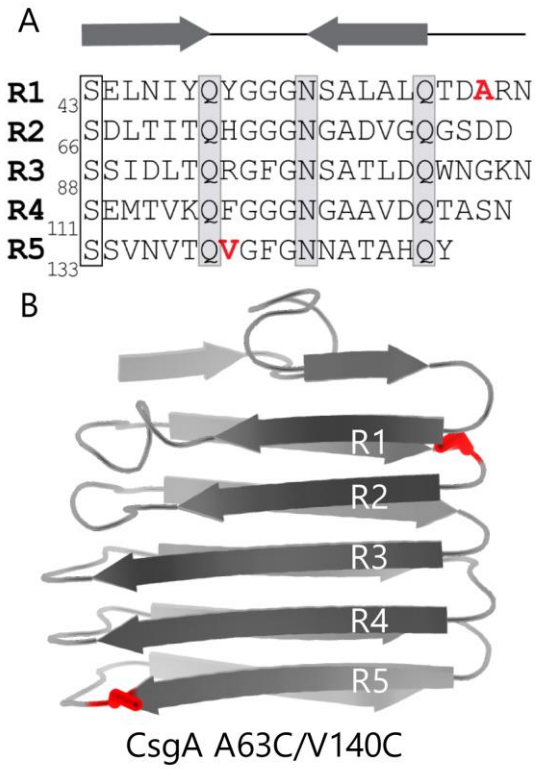

Figure 2-1. CsgACC features two cysteine mutations flanking highly amyloidogenic regions of CsgA WT.
(A) The primary sequence of CsgA contains five imperfect repeats which are labeled R1-R5. Highly conserved serine, glutamine, and asparagine are residues that play a role in amyloid formation are boxed. The location of the two mutated residues are labeled in red. **(B)** Cartoon depiction of a mature and folded CsgA monomer using data from the Lindorff-Larsen lab (Tian et al., 2015). Mutations A63C and V140C are shown in red.

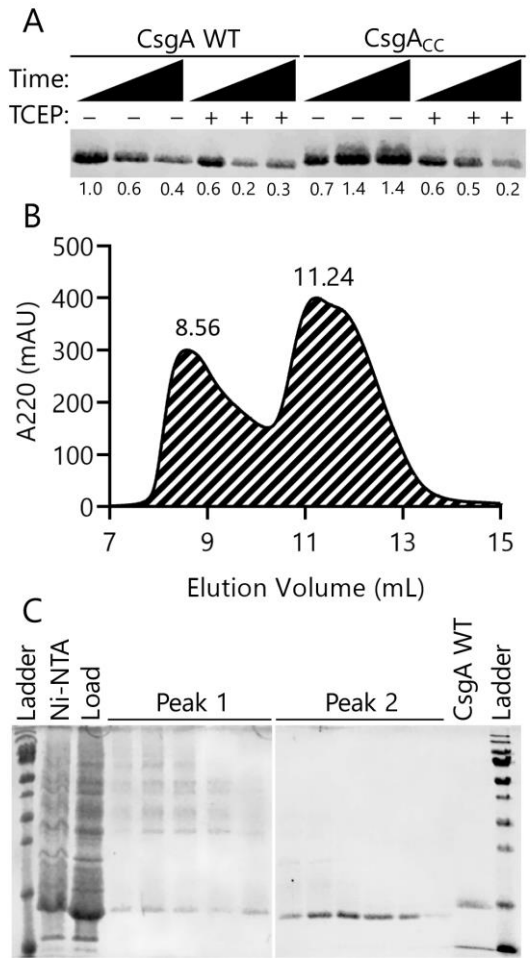

Figure 2-2. Oxidized CsgACC remains SDS-soluble and forms predominantly monomeric species.

(A) Both CsgA WT and CsgA$_{CC}$ were purified and incubated at room temperature in the presence or absence of the reducing agent TCEP. Samples were taken at 0, 24, and 48 hours after purification. An SDS-PAGE gel stained with Coomassie blue shows oxidized CsgA$_{CC}$ remains SDS soluble longer than reduced CsgA$_{CC}$ and CsgA WT. The number below each lane represents a quantification of the band intensity normalized to CsgA WT time 0 hours with no TCEP added. **(B)** A gel filtration chromatogram showing the elution profile of CsgA$_{CC}$. CsgA$_{CC}$ was purified as described in the Materials and Methods without the addition of a reducing agent. There are two broad peak associated with CsgA$_{CC}$ at 8.56 mL (Peak 1) and at 11.24 mL (Peak 2). **(C)** Non-reducing SDS-PAGE gels that contain the eluents from Peaks 1 and 2 of the size exclusion chromatography experiment.

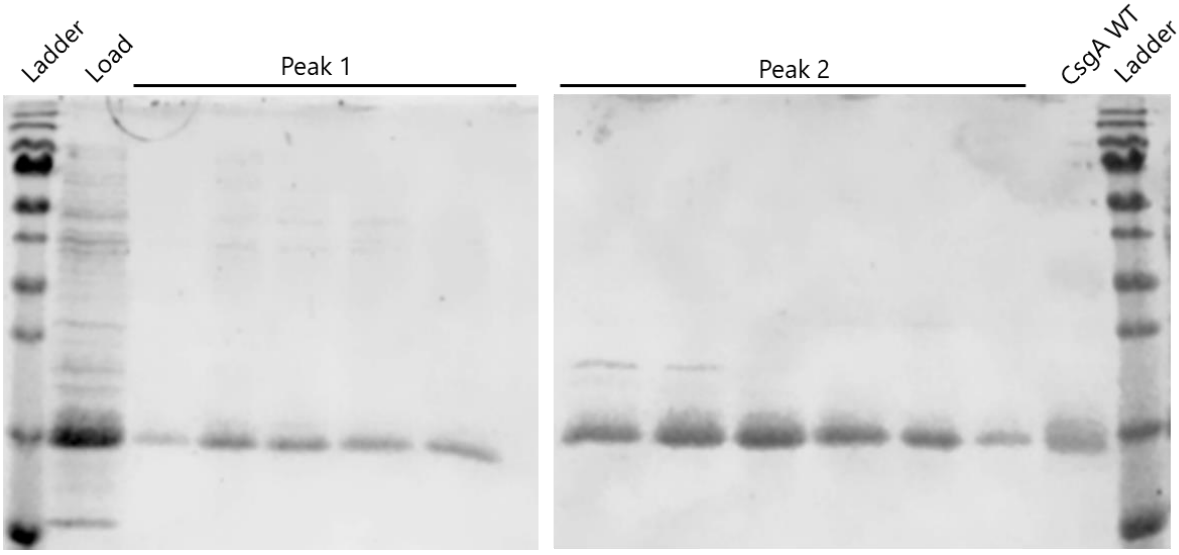

Figure 2-3. High molecular weight bands found in CsgACC gel filtration elution fractions are predominantly CsgA.
Reducing SDS-PAGE gels showing the protein species in SEC elution fractions corresponding to Peaks 1 and 2. In the non-reducing SDS-PAGE gel for the same experiment shown in **Figure 2B**, there is a laddering effect seen in the banding pattern of the peak 1 elution fractions. That pattern is diminished when the samples are reduced with β-mercaptoethanol. In addition, bands associated with monomeric $CsgA_{CC}$ run the same distance as CsgA WT under the reducing conditions shown here.

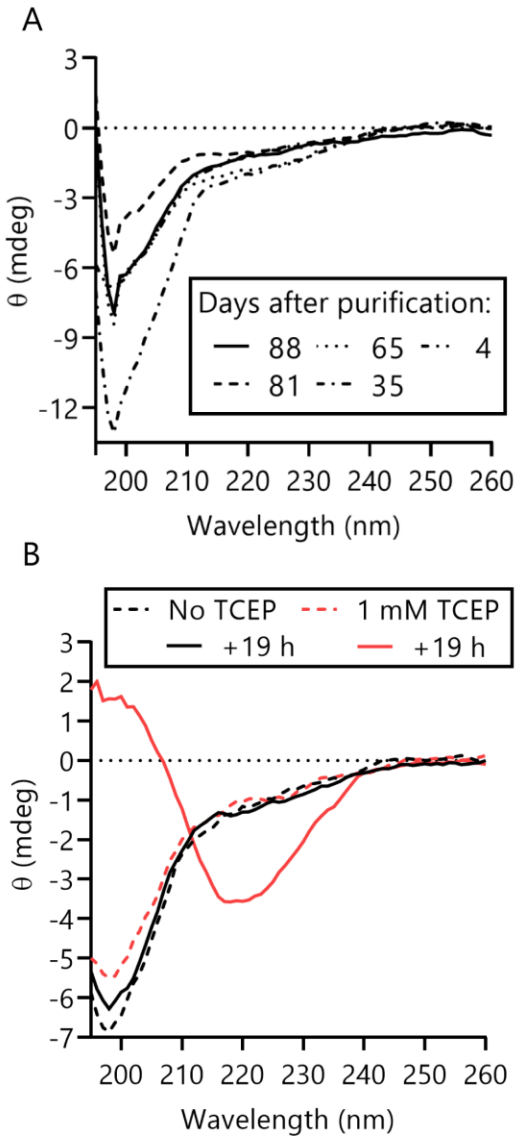

Figure 2-4.CsgACC remains random coil secondary structure until treatment with a reducing agent.

(A) Purified $CsgA_{CC}$ was stored at 4°C and at the indicated times circular dichroism was used to assess secondary structure. At all timepoints assayed, $CsgA_{CC}$ circular dichroism revealed a spectral minima at around 200 nm, which is consistent with random coil secondary structure. Each of the timepoints tested was from a separate purification of $CsgA_{CC}$ and the samples were not normalized to the same concentration, leading to the difference in amplitude of the 200 nm minima. **(B)** A four day old purified protein sample of $CsgA_{CC}$ was split into two and incubated at 26°C with/without reducing agent for 19 hours. $CsgA_{CC}$ shows a CD spectrum characteristic of β-sheet rich protein only when it was incubated in the presence reducing agent.

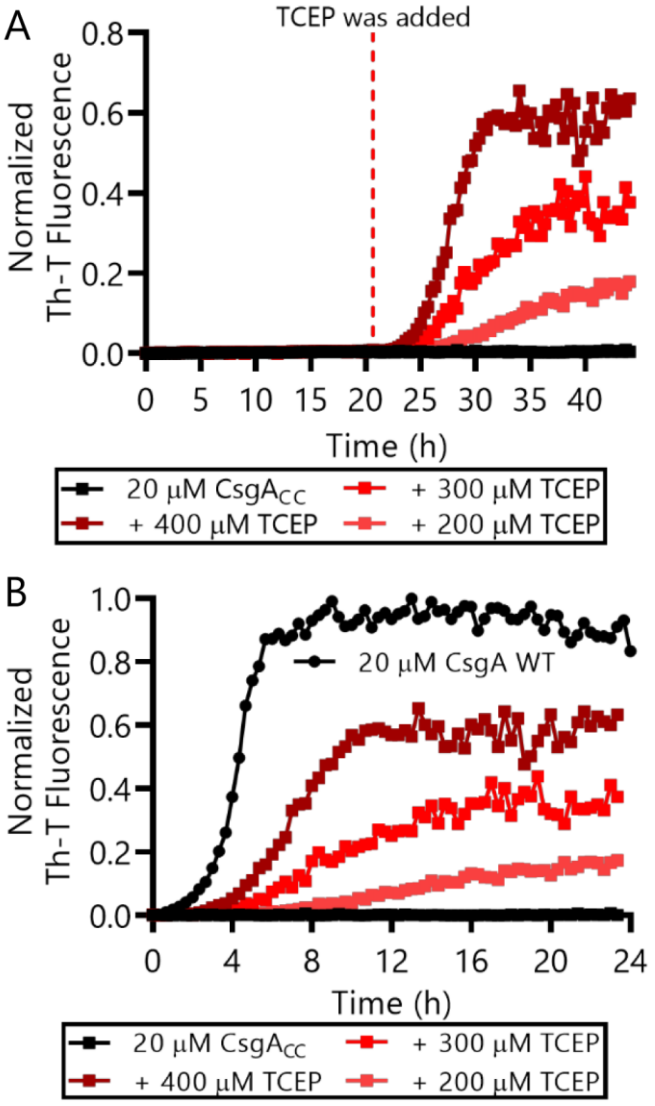

Figure 2-5. Dose-dependent amyloid formation of CsgACC was monitored by Th-T binding.

(A) 20 μM of CsgA$_{CC}$ sample were incubated at room temperature for 20 hours. Afterwards TCEP was added to the solution to a final concentration of 200, 300, and 400 μM. **(B)** The same data is represented where T = 0 hours has been adjusted to reflect the addition of the reducing agent, not the beginning of the experiment, to better observe a change in lag phase. CsgA WT amyloid formation (no reducing agent) was also performed and included for comparison. In all panels, the CsgA$_{CC}$ with no reducing agent data are represented by the curves that are mostly superimposed along the x axis.

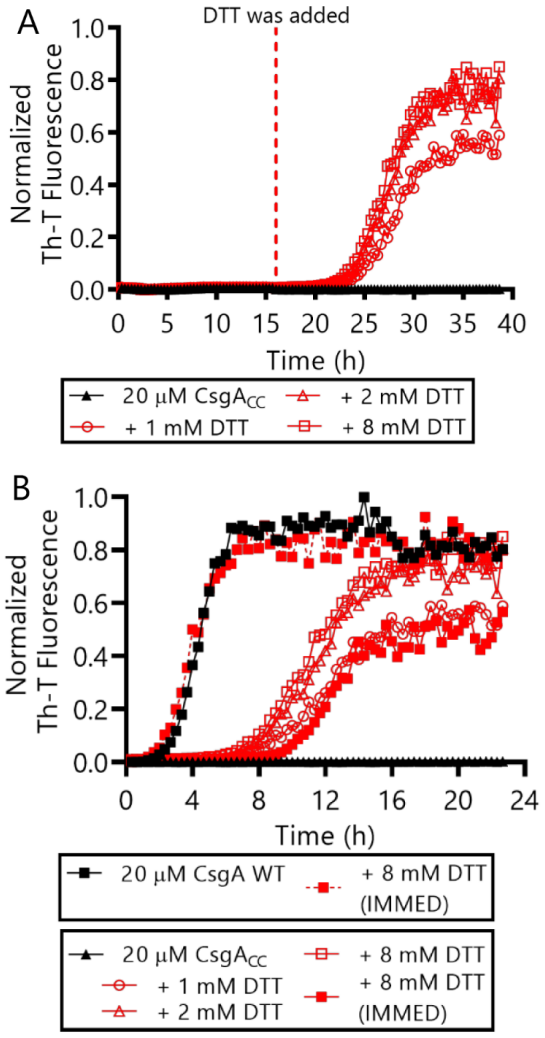

Figure 2-6. Amyloid formation of CsgACC monitored by Th-T binding is triggered upon the addition of reducing agent dithiothreitol (DTT).
(A) Th-T assay showing the formation of amyloids only after a reducing agent is added. In all cases, varying amounts of a reducing agent was added to a 20 µM protein sample after 16 hours. **(B)** The same data is represented where T = 0 hours has been adjusted to reflect the addition of the reducing agent, not the beginning of the experiment. There is one supplementary condition shown wherein the CsgA WT protein is also subjected to the same reducing agent at the highest concentration tested.

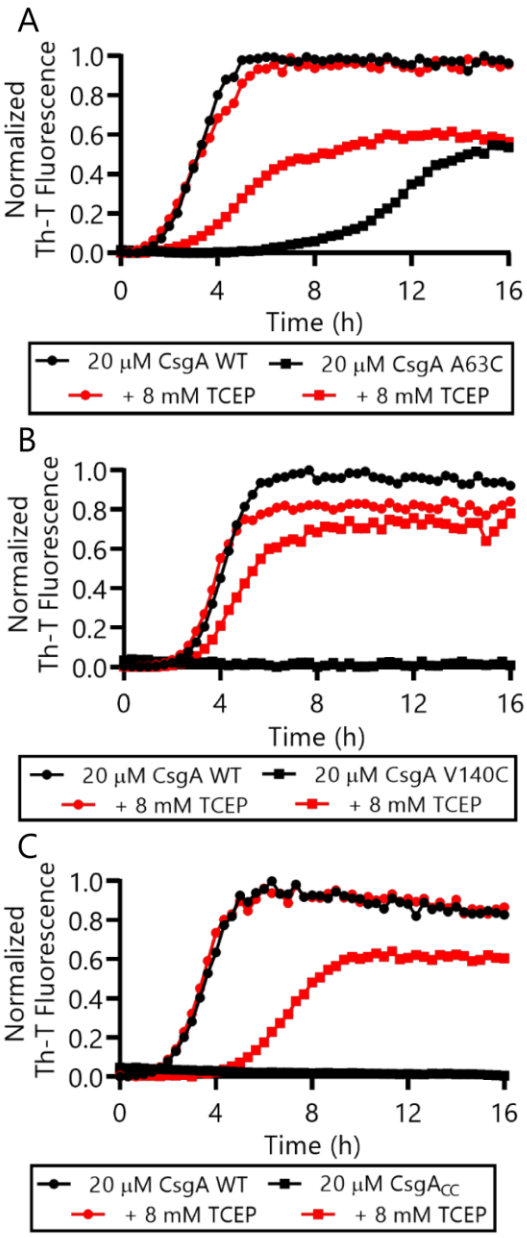

Figure 2-7. CsgA A63C and V140C form amyloid in the absence or presence of a reducing agent while CsgACC requires a reducing agent.
CsgA single cysteine mutant A63C **(A)** and V140C **(B)** form amyloid even without the addition of a reducing agent. The full Th-T curve for V140C can be found in Supp Figure 2. **(C)** In contrast, the double cysteine mutant CsgA$_{CC}$ will only form amyloid after treatment with reducing agent. The CsgA$_{CC}$ with no reducing agent data are represented by the curve that is mostly superimposed along the x-axis.

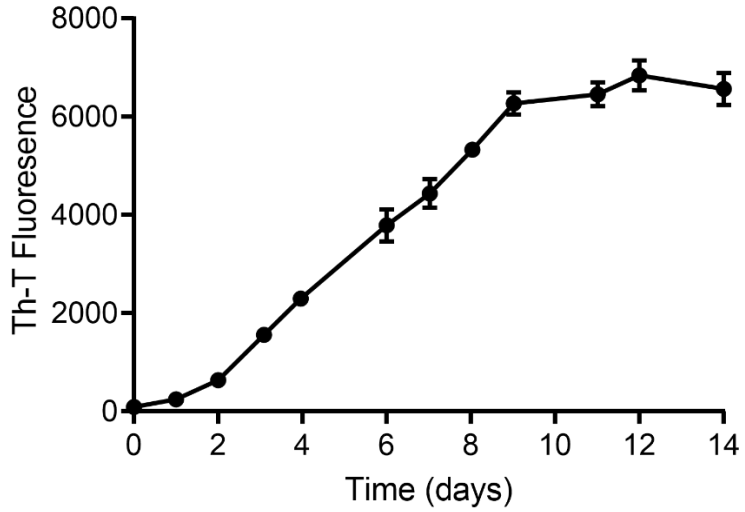

Figure 2-8. Th-T binding assay showing CsgA V140C amyloid formation occur without the addition of reducing agent.

CsgA V140C amyloid formation with and without reducing agent over a 16 hour period can be seen in **Figure 5B**. The oxidized protein forms amyloid much slower than the time frame displayed in **Figure 5B**. Instead, oxidized CsgA V140C requires approximately 9 days to reach the Th-T assay stationary phase. This experiment displays amyloid formation of 20 μM CsgA V140C with a technical replicate value of n = 5.

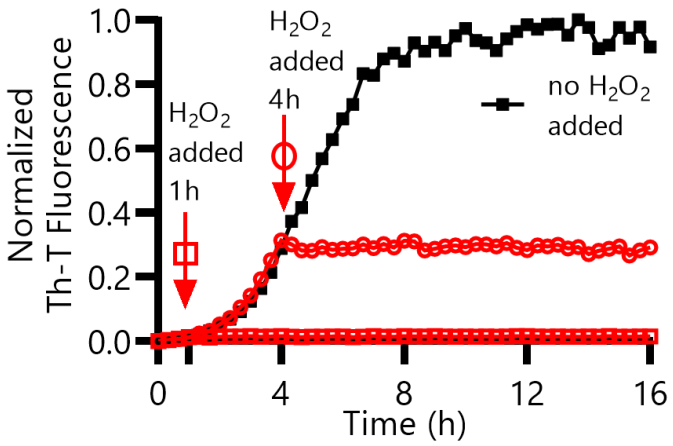

Figure 2-9. CsgACC amyloid formation in the presence of an oxidizing agent, H2O2.

$CsgA_{CC}$ was purified and stored at 4°C for 4 days. After that time had elapsed, TCEP was added to each well to a final concentration of 8 mM. Hydrogen peroxide was added to a final concentration of 2% (v/v) after an additional 1 (square) or 4 (circle) hours had elapsed. This resulted in the arrest of amyloid formation.

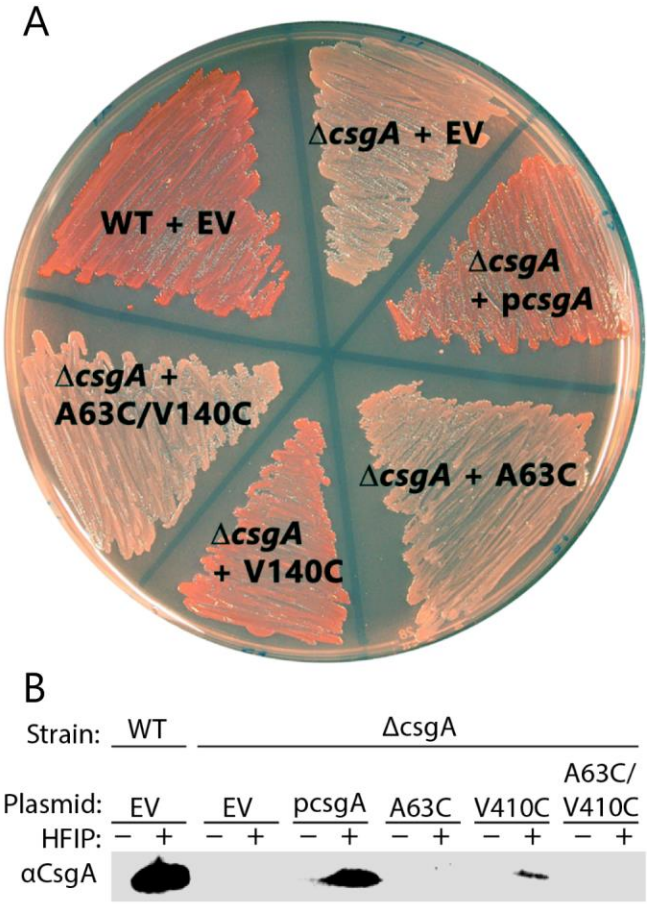

Figure 2-10 *In vivo* complementation assay shows $CsgA_{CC}$ and CsgA A63C expressing bacteria are unable to form curli.
(A) A curli competent strain of *E. coli* (MC4100) forms a red colony phenotype when grown on a YESCA plate supplemented with Congo Red. The curli-deficient strains (Δ*csgA)* form a white colony phenotype under the same conditions. The red phenotype can be rescued when Δ*csgA* strains express CsgA WT from a vector (p*csgA*). $CsgA_{CC}$ and CsgA A63C expressing strains cannot rescue the same red phenotype produced by wildtype bacteria. **(B)** Whole cell western blots were performed on cells scraped from the plate above. After treatment with HFIP, a strong denaturant that solubilizes CsgA, monomeric CsgA is only found in the red colored colonies.

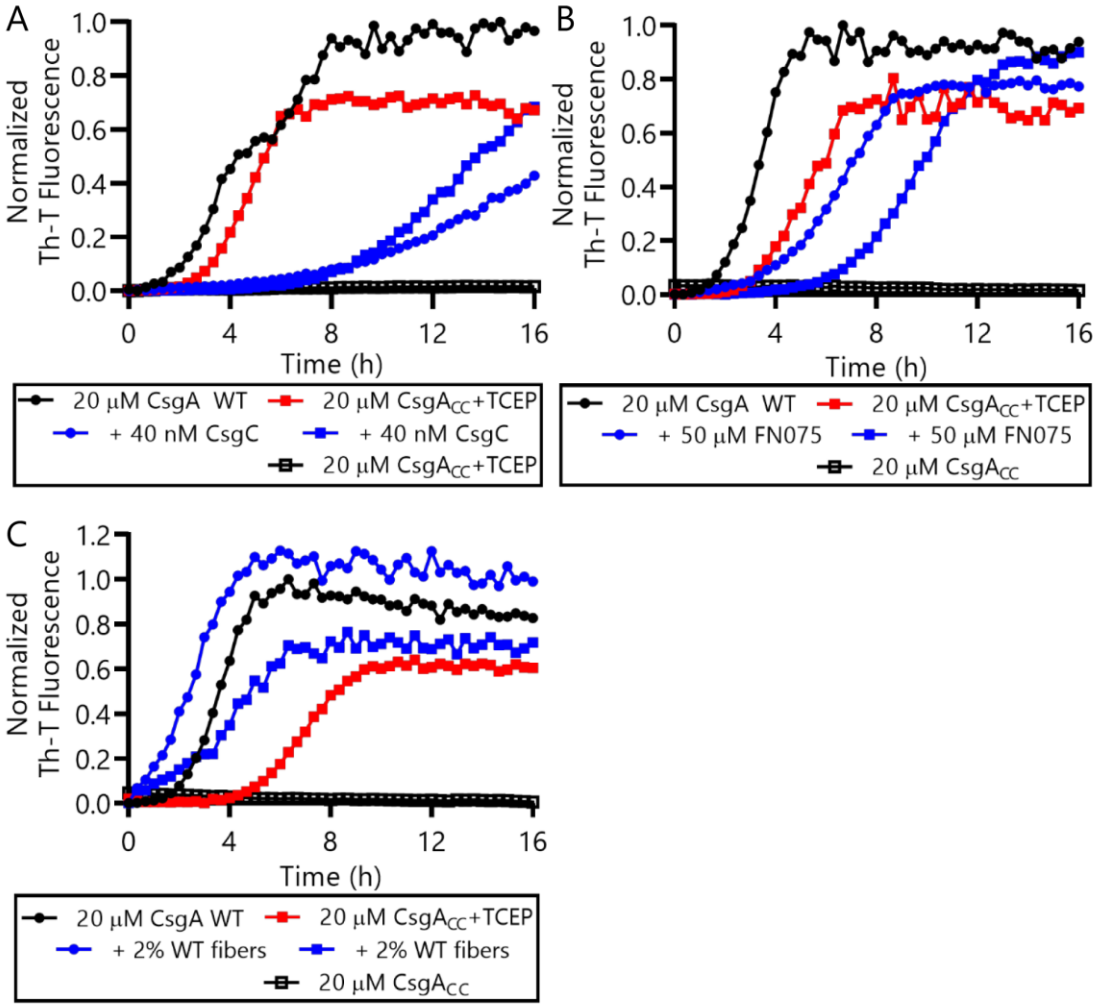

Figure 2-11 CsgA$_{CC}$+TCEP interacts with effectors of amyloid formation similar to the wildtype protein.
Both CsgC **(A)** and FN075 **(B)** are chaperones that selectively inhibit CsgA amyloid formation. CsgA$_{CC}$+TCEP (8 mM TCEP) is similarly inhibited when treated with the same chaperones. **(C)** CsgA amyloid formation is sped up in the presence of preformed fibers in a process called "seeding." CsgA$_{CC}$+TCEP (8 mM TCEP) amyloid formation can also be seeded by preformed fibers. In all panels, the CsgA$_{CC}$ with no reducing agent data are represented by the curves that are mostly superimposed along the x-axis.

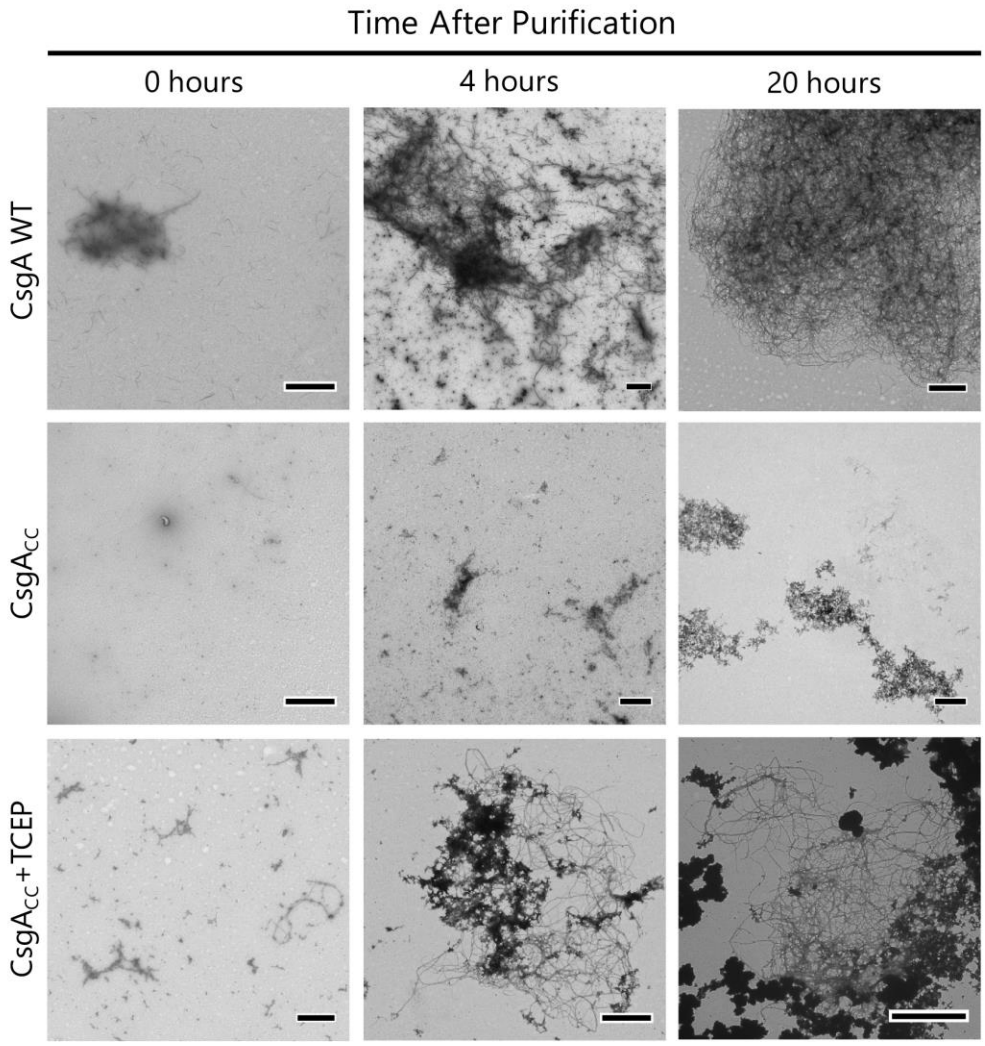

Figure 2-12 CsgA$_{CC}$+TCEP displays fiber morphology similar to wildtype protein. TEM micrograph showing the mature fiber structures made by CsgA$_{CC}$+TCEP. A 5 µL sample was taken from the well of a Th-T binding assay 0, 4, and 20 hours after the addition of a reducing agent. This sample was adsorbed onto a TEM grid and negatively stained with uranyl acetate for contrast. Micrographs shown are representative images of the entire TEM grid. The scale bar is representative of 1 µm.

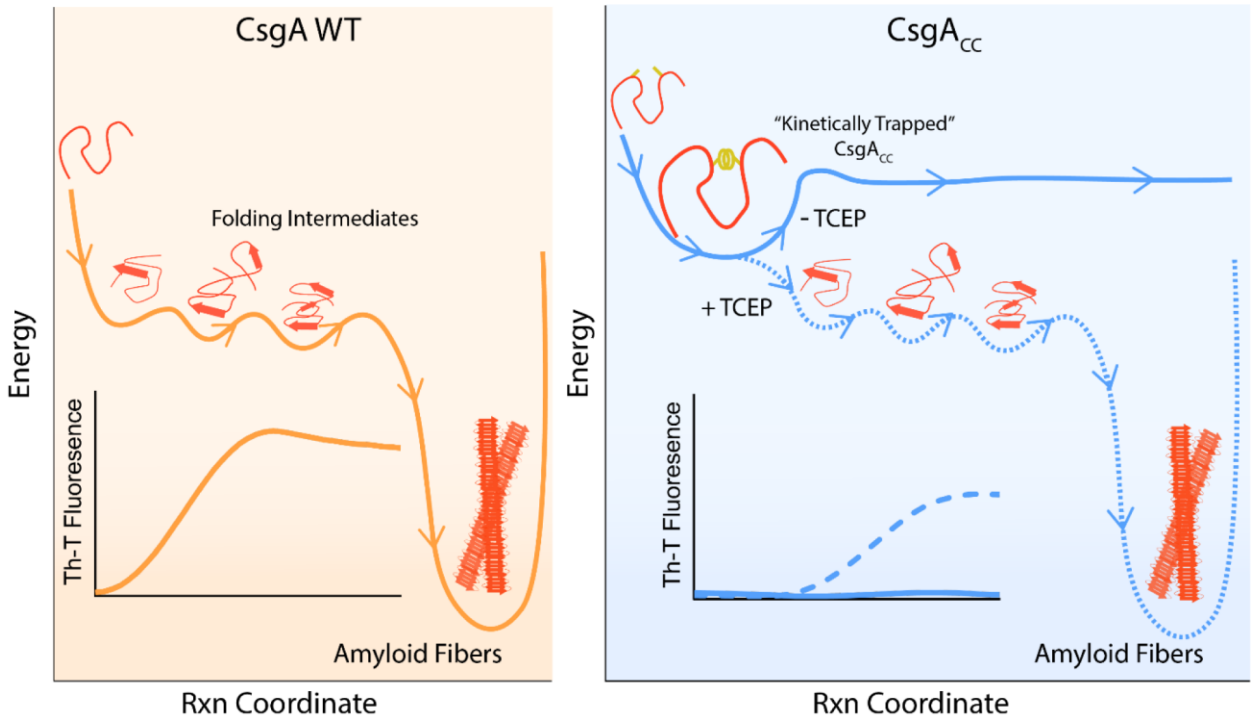

Figure 2-13 A model depicting CsgA$_{CC}$ as kinetically trapped in a non-amyloidogenic state until it is released by a reducing agent.
CsgA WT is an intrinsically disordered protein which initially features random coil secondary structure. CsgA WT eventually folds into a β-sheet rich conformation that is both highly amyloidogenic and energetically favorable. This process can be monitored in ensemble through a Th-T binding assay. On the contrary, CsgA$_{CC}$ remains kinetically trapped in a fold state that is non-amyloidogenic. CsgA$_{CC}$ remains unstructured and monomeric until the disulfide bond is released and the protein is allowed to progress through its normal route towards amyloid formation.

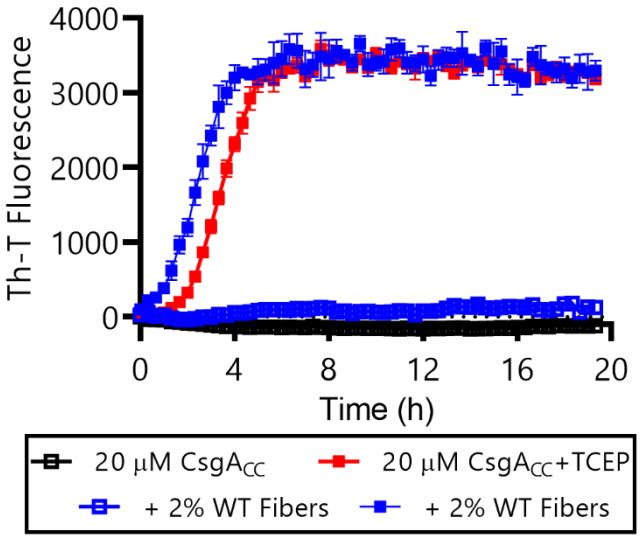

Figure 2-14 The addition of 2% WT fibers does not cause oxidized CsgA$_{CC}$ to form amyloid.

CsgA amyloid formation is sped up in the presence of preformed fibers in a process called "seeding." CsgA$_{CC}$+TCEP amyloid formation can also be seeded by preformed fibers. However, the addition of fibers to oxidized CsgA$_{CC}$ does not cause it to form amyloid before reduction. The CsgA$_{CC}$ with no reducing agent data are represented by the curve that is mostly superimposed along the x-axis.

2.8 Tables

Table 2-1 Strains used in this study.

Strains	Relevant Genotype	References
MC4100	F- *araD*139 Δ(*argF-lac*) U169 *rpsL*150(*strR*) *relA*1 *fibB*5301 *deoC*1 *ptsF*25 *rbsB*	(Casadaban, 1976)
LSR10	MC4100 Δc*sgA*	(Chapman et al., 2002)
MC1061	*F– araD139 Δ(ara-leu)7696 galE15 galK16 Δ(lac)X74 rpsL (StrR) hsdR2 (rK– mK+) mcrA mcrB1*	(Casadaban and Cohen, 1980)
NEB3016	MiniF *lacI*^{q}(Cam^{R}) */ fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10*--Tet^{S}*)2 [dcm] R(zgb-210::Tn10*--Tet^{S}*) endA1 Δ(mcrC-mrr)114::IS10*	New England Biolabs
CsgA	NEB 3016 ΔslyD + pET11d-CsgA -sec C-term 6xHis, amp^{r}	This study
ALB1	NEB 3016 ΔslyD + pET11d-CsgA-sec A63C /V140C C-term 6xHis, amp^{r}	This study
ALB6	NEB 3016 Δ*slyD* + pET11d-CsgA -sec A63C C-term 6xHis, amp^{r}	This study
ALB9	NEB 3016 Δ*slyD* + pET11d-CsgA -sec V140C C-term 6xHis, amp^{r}	This study
ALB13	MC4100 Δc*sgA* + pLR5 (PCsgBAC-CsgA), kan^{r}	This study
ALB14	MC4100 Δc*sgA* + pLR2 (PCsgBAC-EV), kan^{r}	This study
ALB15	MC4100 Δc*sgA* + pLR5 (PCsgBAC-CsgA V140C), kan^{r}	This study

ALB16	MC4100 + pLR2 (PCsgBAC-EV), kan^r	This study
ALB17	MC4100 Δc*sgA* + pLR5 (PCsgBAC-CsgA A63C V140C), kan^r	This study
ALB18	MC4100 Δc*sgA* + pLR5 (PCsgBAC-CsgA A63C), kan^r	This study

Table 2-2 Plasmids used in this study.

Plasmids	Relevant Characteristics	References
pET11d	IPTG inducible expression vector	New England Biolabs
pLR2	Control vector containing CsgBAC promoter	(Robinson et al., 2006)
pLR5	*csgA* sequence in pLR2	(Wang et al., 2008)

Table 2-3 Primers used in this study.

Primers	Primer Sequence (5'→ 3')	Constructs
CsgA_A63C	GCACTTGCTCTGCAAACTGATTGCCGTAACTCTGACTTG	CsgA A63C
CsgA_A63C_AS	CAAGTCAGAGTTACGGCAATCAGTTTGCAGAGCAAGTGC	CsgA A63C
CsgA_V140C	TCCGTCAACGTGACTCAGTGTGGCTTTGGTAACAACGC	CsgA V140C
CsgA_V140C_AS	GCGTTGTTACCAAAGCCACACTGAGTCACGTTGACGGA	CsgA V140C

2.9 References

Anfinsen, C. B., and Scheraga, H. A. (1975). Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.* 29, 205–300. doi:10.1016/S0065-3233(08)60413-1.

Arosio, P., Knowles, T. P. J., and Linse, S. (2015). On the lag phase in amyloid fibril formation. *Phys. Chem. Chem. Phys.* 17, 7606–7618. doi:10.1039/c4cp05563b.

Barnhart, M. M., and Chapman, M. R. (2006). Curli biogenesis and function. *Annu Rev Microbiol* 60, 131–147. doi:10.1146/annurev.micro.60.080805.142106.

Carija, A., Pinheiro, F., Pujols, J., Brás, I. C., Lázaro, D. F., Santambrogio, C., et al. (2019). Biasing the native α-synuclein conformational ensemble towards compact states abolishes aggregation and neurotoxicity. *Redox Biol.* 22, 101135. doi:10.1016/j.redox.2019.101135.

Cegelski, L., Pinkner, J. S., Hammer, N. D., Cusumano, C. K., Hung, C. S., Chorell, E., et al. (2009). Small-molecule inhibitors target Escherichia coli amyloid biogenesis and biofilm formation. *Nat. Chem. Biol.* 5, 913–919. doi:10.1038/nchembio.242.

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 (80-.).* 295, 851–855. doi:10.1126/science.1067484.

Chiti, F., and Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75, 333–366. doi:10.1146/annurev.biochem.75.101304.123901.

Chiti, F., and 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. doi:10.1146/annurev-biochem-061516-045115.

Deshmukh, M., Evans, M. L., and Chapman, M. R. (2018). Amyloid by Design: Intrinsic Regulation of Microbial Amyloid Assembly. *J. Mol. Biol.* 430, 3631–3641. doi:10.1016/j.jmb.2018.07.007.

Dombkowski, A. A., Sultana, K. Z., and Craig, D. B. (2014). Protein disulfide engineering. *FEBS Lett.* 588, 206–212. doi:10.1016/J.FEBSLET.2013.11.024.

Dueholm, M. S., Albertsen, M., Otzen, D., and Nielsen, P. H. (2012). Curli Functional Amyloid Systems Are Phylogenetically Widespread and Display Large Diversity in Operon and Protein Structure. *PLoS One* 7, e51274. doi:10.1371/journal.pone.0051274.

Eisenberg, D., and Jucker, M. (2012). The amyloid state of proteins in human diseases. *Cell* 148, 1188–1203. doi:10.1016/j.cell.2012.02.022.

Evans, M. L., Chorell, E., Taylor, J. D., Åden, J., Götheson, A., Li, F., et al. (2015). The Bacterial Curli System Possesses a Potent and Selective Inhibitor of Amyloid Formation. *Mol. Cell* 57, 445–455. doi:10.1016/J.MOLCEL.2014.12.025.

Evans, M. L., Gichana, E., Zhou, Y., and Chapman, M. R. (2018). "Bacterial amyloids," in *Methods in Molecular Biology* (Humana Press Inc.), 267–288. doi:10.1007/978-1-4939-7816-8_17.

Gilbert, H. F. (1995). Thiol/disulfide exchange equilibria and disulfidebond stability. *Methods Enzymol.* 251, 8–28. doi:10.1016/0076-6879(95)51107-5.

Gosal, W. S., Morten, I. J., Hewitt, E. W., Smith, D. A., Thomson, N. H., and Radford, S. E. (2005). Competing pathways determine fibril morphology in the self-assembly of β2-microglobulin into amyloid. *J. Mol. Biol.* 351, 850–864. doi:10.1016/j.jmb.2005.06.040.

Hammar, M., Arnqvist, A., Bian, Z., Olsén, A., and 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. doi:10.1111/j.1365-2958.1995.mmi_18040661.x.

Hammer, N. D., Schmidt, J. C., and Chapman, M. R. (2007). The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. *Proc Natl Acad Sci U S A* 104, 12494–12499. doi:10.1073/pnas.0703310104.

Hoyer, W., Grönwall, C., Jonsson, A., Ståhl, S., and Härd, T. (2008). Stabilization of a β-hairpin in monomeric Alzheimer's amyloid-β peptide inhibits amyloid formation. *Chemtracts* 20, 499–500. doi:10.1073/pnas.0711731105.

Jain, N., Aden, J., Nagamatsu, K., Evans, M. L., Li, X., McMichael, B., et al. (2017). Inhibition of curli assembly and Escherichia coli biofilm formation by the human systemic amyloid precursor transthyretin. *Proc Natl Acad Sci U S A* 114, 12184–12189. doi:10.1073/pnas.1708805114.

Jain, N., and Chapman, M. R. (2019). Bacterial functional amyloids: Order from disorder. *Biochim. Biophys. Acta - Proteins Proteomics* 1867, 954–960. doi:10.1016/j.bbapap.2019.05.010.

Liu, T., Wang, Y., Luo, X., Li, J., Reed, S. A., Xiao, H., et al. (2016). Enhancing protein stability with extended disulfide bonds. *Proc. Natl. Acad. Sci. U. S. A.* 113, 5910–5. doi:10.1073/pnas.1605363113.

Lomakin, A., Chung, D. S., Benedek, G. B., Kirschner, D. A., and Teplow, D. B. (1996). On the nucleation and growth of amyloid β-protein fibrils: Detection of nuclei and quantitation of rate constants. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1125–1129. doi:10.1073/pnas.93.3.1125.

Matsumurat, M., Becktelt, W. J., Levitr, M., and Matthewstl, B. W. (1989). Stabilization of phage T4 Iysozyme by engineered disulfide bonds (thermostability/lysozyme/protein structure). *Biochemistry* 86, 6562–6566. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC297884/pdf/pnas00284-0130.pdf [Accessed March 8, 2018].

Nenninger, A. A., Robinson, L. S., Hammer, N. D., Epstein, E. A., Badtke, M. P., Hultgren, S. J., et al. (2011). CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation. *Mol Microbiol* 81, 486–499. doi:10.1111/j.1365-2958.2011.07706.x.

Nenninger, A. A., Robinson, L. S., and Hultgren, S. J. (2009). Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. *Proc. Natl. Acad. Sci. U. S. A.* 106, 900–905. doi:10.1073/pnas.0812143106.

Nguyen, P. Q., Botyanszki, Z., Tay, P. K. R., and Joshi, N. S. (2014). Programmable biofilm-based materials from engineered curli nanofibres. *Nat. Commun.* 5. doi:10.1038/ncomms5945.

Nilsson, M. R. (2004). Techniques to study amyloid fibril formation *in vitro*. *Methods* 34, 151–160. doi:10.1016/J.YMETH.2004.03.012.

Otzen, D. (2010). Functional amyloid: turning swords into plowshares. *Prion* 4, 256–64. doi:10.4161/PRI.4.4.13676.

Reichhardt, C., and Cegelski, L. (2014). Solid-state NMR for bacterial biofilms. *Mol. Phys.* 112, 887–894. doi:10.1080/00268976.2013.837983.

Robinson, L. S., Ashman, E. M., Hultgren, S. J., and Chapman, M. R. (2006). Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. *Mol. Microbiol.* 59, 870–881. doi:10.1111/j.1365-2958.2005.04997.x.

Salgado, P. S., Taylor, J. D., Cota, E., and Matthews, S. J. (2011). Extending the usability of the

phasing power of diselenide bonds: SeCys SAD phasing of CsgC using a non-auxotrophic strain. *Acta Crystallogr. D. Biol. Crystallogr.* 67, 8–13.

Sambi, I., Gatti-Lafranconi, P., Longhi, S., and Lotti, M. (2010). How disorder influences order and vice versa - mutual effects in fusion proteins containing an intrinsically disordered and a globular protein. *FEBS J.* 277, 4438–4451. doi:10.1111/j.1742-4658.2010.07825.x.

Schmidt, S., Genz, M., Balke, K., and Bornscheuer, U. T. (2015). The effect of disulfide bond introduction and related Cys/Ser mutations on the stability of a cyclohexanone monooxygenase. *J. Biotechnol.* 214, 199–211. doi:10.1016/J.JBIOTEC.2015.09.026.

Seker, U. O. S., Chen, A. Y., Citorik, R. J., and Lu, T. K. (2017). Synthetic Biogenesis of Bacterial Amyloid Nanomaterials with Tunable Inorganic-Organic Interfaces and Electrical Conductivity. *ACS Synth. Biol.* 6, 266–275. doi:10.1021/acssynbio.6b00166.

Sleutel, M., Van den Broeck, I., Van Gerven, N., Feuillie, C., Jonckheere, W., Valotteau, C., et al. (2017). Nucleation and growth of a bacterial functional amyloid at single-fiber resolution. *Nat. Chem. Biol.* 13, 902–908. doi:10.1038/nchembio.2413.

Tay, P. K. R., Manjula-Basavanna, A., and Joshi, N. S. (2018). Repurposing bacterial extracellular matrix for selective and differential abstraction of rare earth elements. *Green Chem.* 20, 3512–3520. doi:10.1039/c8gc01355a.

Tian, P., Boomsma, W., Wang, Y., Otzen, D. E., Jensen, M. H., and Lindorff-Larsen, K. (2015). Structure of a Functional Amyloid Protein Subunit Computed Using Sequence Variation. *J. Am. Chem. Soc.* 137, 22–25. doi:10.1021/ja5093634.

Wang, X., Hammer, N. D., and Chapman, M. R. (2008). The molecular basis of functional bacterial amyloid polymerization and nucleation. *J Biol Chem* 283, 21530–21539. doi:10.1074/jbc.M800466200.

Wang, X., Smith, D. R., Jones, J. W., and Chapman, M. R. (2007). *In vitro* polymerization of a functional Escherichia coli amyloid protein. *J. Biol. Chem.* 282, 3713–9. doi:10.1074/jbc.M609228200.

Whitaker, J. R. (1963). Determination of Molecular Weights of Proteins by Gel Filtration of Sephadex. *Anal. Chem.* 35, 1950–1953. doi:10.1021/ac60205a048.

Zhou, Y., Smith, D., Leong, B. J., Brannstrom, K., Almqvist, F., and Chapman, M. R. (2012). Promiscuous cross-seeding between bacterial amyloids promotes interspecies biofilms. *J Biol Chem* 287, 35092–35103. doi:10.1074/jbc.M112.383737.

Zhou, Y., Smith, D. R., Hufnagel, D. A., and Chapman, M. R. (2013). Experimental manipulation of the microbial functional amyloid called curli. *Methods Mol Biol* 966, 53–75. doi:10.1007/978-1-62703-245-2_4.

Chapter 3 CsgC Inhibits CsgA Amyloid Formation by Promoting the Intrinsically Disordered Fold State[3]

3.1 Abstract

E. coli secrete a functional amyloid called curli during biofilm formation. The operon that controls curli formation includes CsgC. CsgC can potently inhibit intracellular amyloid formation by CsgA, the major curli amyloid protein. The mechanism through which CsgC inhibits CsgA amyloid formation has been poorly characterized. Here we provide direct observations of CsgC interacting with monomeric and unfolded CsgA using several biophysical techniques. Despite the weak and transient nature of the CsgC-CsgA interaction, CsgA is unable to adopt an aggregation-prone state in the presence of CsgC. Furthermore, an *in vivo* screening assay was designed to identify amino acid residues in CsgC that are required for its ability to inhibit CsgA aggregation. Through the results of this assay, we learned that CsgC activity is partially mediated through of surface exposed charged residues E48 and R84

3.2 Introduction

Amyloids are fibril protein aggregates that are characterized by their morphology, stability, and tinctorial properties (Chiti and Dobson, 2017). The proteins that compose amyloid

[3]The work presented in this chapter is associated with a manuscript in preparation titled "CsgC Inhibits CsgA Amyloid Formation by Promoting the Intrinsically Disordered Fold State". Sanduni Jayakodi contributed to Figure 3-1. Yilin Han contributed writing and data for Figure 3-2 and 3-3. Emily Goetzler, Rachel Alessio, and Lily Kalcec contributed to Figure 3-4, 3-6, 3-8, and 3-9. Anthony Balistreri contributed to the concept and experimental design of all figures.

fibers are often intrinsically disordered, or have intrinsically disordered regions, and go on to adopt a repetitive secondary structure motif (i.e., cross-β sheets) (Chiti and Dobson, 2006). In a solution containing amyloid-competent protein, a species arises from a pool of disordered conformers that acts as a nucleation site for fiber formation (Rochet and Lansbury, 2000). Monomers add on to so called "nuclei" and begin the progressive and thermodynamically-favorable process of amyloid fiber formation (Rochet and Lansbury, 2000). Fiber formation marks the transition of amyloid proteins from the soluble state to the insoluble, between functional and non-functional, and in some cases between benign and pathogenic (Chiti and Dobson, 2006). Functional amyloids are a class of amyloid that are created by an organism through a controlled mechanism to exploit a useful function, the list of which continues to grow with every passing year (Balistreri et al., 2020; Chapman et al., 2002; Otzen, 2010).

The most widely studied functional amyloid is curli, the main protein component of the *E. coli* biofilm (Chapman et al., 2002). The *E. coli csg* operon containing seven curli specific genes functions as the producer and regulator for curli formation, with each protein playing a significant role (Deshmukh et al., 2018). The main amyloid protein called CsgA is translated and translocated into the periplasm by the ribosome and SecYEG pore (Chapman et al., 2002). Periplasmic CsgA is shepherded to the nonameric CsgG outer membrane pore by CsgE, a periplasmic chaperone (Nenninger et al., 2011). Two auxiliary proteins, CsgB and CsgF, are also exported through the CsgG pore and act as a nucleator and anchor respectively to allow for CsgA amyloids fibers to form attached to the cell's outer surface (Hammer et al., 2007; Nenninger et al., 2009). The operon contains a second periplasmic protein, CsgC, and its function was first described as a potent amyloid inhibitor (Evans et al., 2015).

CsgC is a 110 amino acid chaperone-like protein with an immunoglobulin-like β-sandwich structure (Taylor et al., 2011). When assessed *in vitro*, CsgC is an efficient substoichiometric inhibitor of amyloid formation against *E. coli* CsgA, its native client (Evans et al., 2015). Indeed, CsgC is an efficient inhibitor of amyloid formation for a selection of amyloid proteins including CsgA homologs, the bacterial amyloid FapC, and, interestingly, the human pathogenic amyloid alpha-synuclein (Evans et al., 2015). Little is known about the CsgC-client interaction, except i) the interaction is guided by electrostatic interactions (Taylor et al., 2016)

$$[CsgA_{unfolded}] \overset{k_f}{\rightleftharpoons} [CsgA_{folded}] \overset{k_n}{\rightleftharpoons} [Nuclei] + [CsgA_{unfolded}] \overset{k_+}{\rightleftharpoons} [Fibers]$$

Equation 3-1. A simplified mechanism showing the process of CsgA amyloid formation. CsgA is an intrinsically disordered protein that eventually folds to produce an amyloid-competent monomer (described by rate constant k_f). A primary nucleus forms in solution from an unknown number of folded CsgA monomers (described by rate constant k_n). After a critical concentration of nuclei are present, unfolded CsgA monomers use nuclei as folding templates, sparking rapid fiber elongation (described by rate constant k_+)

and ii) there is a weak client sequence determinant (Evans et al., 2015). Many specific details about CsgC-mediated amyloid inhibition remain poorly understood.

Studying CsgC presents many interesting challenges. CsgC is a simple protein with no predicted active sites, domains, or significant amino acids. CsgC requires no hydrolysable substrate, cofactor, or cellular energy to perform its activity. CsgC does not take up a substrate or produce a quantifiable product; the only evidence of CsgC activity is the inhibition of amyloid formation. Amyloid formation is a complex process in a constant state of flux, during which soluble, disordered monomers join an insoluble, ordered amyloid fiber (**Equation 1**). CsgC inhibits amyloid formation, although it is unclear at what point during amyloid formation CsgC is acting. Currently there is conflicting data that suggests CsgC can act during the elongation

phase of amyloid formation (affecting k_+) (Goyal et al., 2014) as well as during primary nucleation (affecting k_n) (Nagaraj et al., 2022).

Here, we use several biophysical techniques to provide direct evidence that CsgC is binding to monomeric and unfolded CsgA. After transiently interacting with CsgC, monomeric CsgA is steered away from the aggregation prone fold state. Using an *in vivo* screening assay, we determined that residues E48 and R84 are potentially significant to CsgC activity.

3.3 Results

3.3.1 Interaction with CsgC Delays Aggregation-Prone State of CsgA Monomer

Our lab found that CsgC can inhibit CsgA amyloid formation at molar ratios as low as 1:500 CsgC:CsgA (Evans et al., 2015). We hypothesize that the effective low molar ratios of CsgC needed to inhibit CsgA amyloid formation can be achieved because CsgC molecules interact only transiently with CsgA.

By coupling a pull-down assay to size exclusion chromatography (SEC) we were able to probe for an interaction between CsgC and CsgA and get information about the oligomeric state of CsgA (**Figure 3-1**). After Ni-NTA purification, the CsgA sample was split into three equal parts. The first aliquot of purified CsgA was passed directly through an SEC column resulting in a chromatogram with one lower (~ 12 mL) molecular weight and one higher (~8 mL) molecular weight peak (**Figure 3-1A**). The remaining two aliquots of purified CsgA were passed through N-hydrosuccinimide (NHS) columns that contained either CsgC or BSA that had been immobilized to the NHS using anime-reactive coupling and affixed in line before the SEC column. CsgA that had passed through the BSA-modified NHS column showed a similar chromatogram to the no-NHS condition, with two peaks at approximately the same column volume of 13 mL (**Figure 3-1B**). CsgA that had passed through the CsgC-modified NHS column

showed a chromatogram with a diminished high molecular weight peak and a much sharper and taller low molecular weight peak (**Figure 3-1C**). SEC elution fractions containing the CsgA monomers from both the no NHS and CsgC-modified NHS separations were placed in a ThT binding assay to check for amyloid formation (**Figure 3-1D and E**). The CsgA monomers that did not pass through an NHS column began aggregating after in less than 2 hours (**Figure 3-1D**). The CsgA monomers that passed by CsgC had a diminished propensity to form amyloid when compared to the no NHS column condition, showing very little ThT fluorescence over the course of the experiment (**Figure 3-1E**). Having observed some interaction between CsgC and CsgA monomers, we decided to look using a more sensitive biophysical technique.

3.3.2 CsgC:CsgA Heterodimer

Native mass spectrometry (native MS) uses gentle electrospray ionization conditions to preserve non-covalent protein-protein and protein-ligand complexes (Tamara et al., 2021). The added dimension of the ion mobility (IM) spectrometry allows ions to be separated based on their size, shape, and charge (May and McLean, 2015). Together, native IM-MS reports on mass to charge (m/z) and rotationally averaged collisional cross section (CCS) values derived from IM arrival time distributions (ATDs) (Gabelica et al., 2019).

We purified both CsgC and CsgA, added them together in a 1:1 mixture, and analyzed the resulting solution for any interactions between the two proteins. 30 minutes after mixing CsgC and CsgA both monomeric and dimeric species of CsgA and CsgC were observed, including heterodimeric complexes of CsgA and CsgC (**Figure 3-2A and B**). CsgA exhibited a broad charge state distribution ranging from 5+ to 13+, as typically observed for intrinsically disordered proteins (IDP) (**Figure 3-2A**) (Santambrogio et al., 2019). The CsgA:CsgC mixture remained soluble throughout the 23 hour time course while apo CsgA, serving as a control

sample, formed micro-scale aggregates that led to the clogging of our nESI emitters and prevented the collection of both 6 hour and 23 hour time point data. An analysis of CCS values of CsgA monomers showed very little difference between time 0 and after 3 hour (**Figure 3-3A**). Similarly, we also saw no changes in CCS for CsgC monomer ions through the 23 hour time course (**Figure 3-3B**).

We then analyzed our recorded IM ATDs so that subtle conformational changes within CsgA and CsgC might be revealed (**Figure 3-2C)**. Our ATD data contains information regarding both ion shape and the number of conformer families present for a single protein ion population (Baumketner et al., 2006). Data shown in **Figure 3-2C** displays three overlaid ATD profiles for CsgA 11+ monomer ions recorded after 0 and 3 hour incubation. In addition, similar IM ATD data is shown for CsgA/CsgC mixtures following a 3 hour incubation (**Figure 3-2C)**. All three datasets yield a bimodal ATD profile, consisting of both a more compacted (shorter IM drift times) and a more extended conformational family (longer IM drift times) (**Figure 3-2C)**. A comparison of CsgA IM data collected at both 0 and 3 hour time points reveals the peak corresponding to the more compact conformation decreased significantly in intensity, resulting in overwhelmingly a more extended protein ion population (**Figure 3-2C)**. Overall, our IM ATD data suggests that the conformational ensemble of CsgA becomes more extended as protein aggregation progresses, an observation common to other amyloid proteins (Hyung et al., 2013; Soper et al., 2013). Importantly, when we compare our IM ATD data recorded for CsgA at 0 hour with ATDs recorded for CsgA/CsgC mixtures following 3 hours of incubation, we detect few significant differences, indicating that the initial, intrinsically disordered state of CsgA is better preserved in the presence of CsgC (**Figure 3-2C)**. The results discussed above were consistent across multiple CsgA charge states (**Figure 3-2C and D**).

3.3.3 CsgC Activity during Elongation Phase

We then asked if CsgC is also an effective inhibitor of CsgA amyloid formation when there are fibers present. When the nucleation-dependent polymerization lag phase has finished, CsgA transitions into the elongation phase during which rapid amyloid formation occurs. The elongation phase of amyloid formation is marked by monomer addition where soluble protein monomers are adding onto fiber ends (Collins et al., 2004; Jahn and Radford, 2008). During the elongation phase there is evidence of unfolded CsgA monomers still being in solution (Dueholm et al., 2011).

CsgC was an effective inhibitor of aggregation when added to CsgA during the fiber elongation phase (**Figure 3-4**). We tested the effect of CsgC during the elongation phase by setting up two similar scenarios. First, CsgC was added to a final concentration of 200 nM (1:100 CsgC:CsgA molar ratio) at different time points during the aggregation of CsgA from 0 to 4 hours (**Figure 3-4A**). When CsgC was added at time 0 hours the ThT signal increase was very small. After waiting 1 hour, which is still within this CsgA purification batch's ThT lag phase, aggregation increased in a linear fashion (**Figure 3-4A**). When CsgC was added after the lag phase had ended the result was a linear increase in ThT signal followed by a gradual decrease reaching a plateau in signal (**Figure 3-4A**). Next, CsgC was added to CsgA along with preformed CsgA fibers (**Figure 3-4B**). During a seeded reaction, CsgA is presented with preformed fibers and the fibers act as a seed for amyloid formation, effectively diminishing the lag phase entirely (**Figure 3-4B,** dark blue squares)(Wang et al., 2007). When CsgC was added to a seeded CsgA reaction the result was a linear increase in ThT signal (**Figure 3-4B**). The addition of CsgC in both similar scenarios where CsgA should be rapidly adding into amyloid

fibers resulted in linear polymerization, which could indicate a decrease in the formation of new nuclei in solution.

3.3.4 Growth and Fluorescence Phenotypes for CsgC Activity In Vivo

When CsgA was expressed intracellularly by removing its secretory signal peptide ($CsgA_{\Delta sec}$) there was an apparent cytotoxicity that was seen during growth assays, beginning approximately an hour after expression (**Figure 3-5A**). If CsgC was coexpressed intracellularly by removing its secretory signal peptide ($CsgC_{\Delta sec}$) the cytotoxic effect was diminished (**Figure 3-5A**). When previously reported, our lab speculated that amyloid fibers formed within the cells after $CsgA_{\Delta sec}$ expression that caused the cells to die and $CsgC_{\Delta sec}$ was playing a protective role when present (Evans et al., 2015). An amyloid-specific and cell-permeable fluorescent probe called AmyTracker 680 was used to visualize the formation of amyloid fibers during expression of $CsgA_{\Delta sec}$ and coexpression with $CsgC_{\Delta sec}$ (**Figure 3-5B**). There was robust fluorescence of Amytracker 680 when $CsgA_{\Delta sec}$ was expressed that was greatly reduced when $CsgC_{\Delta sec}$ was coexpressed, coinciding with the growth phenotype (**Figure 3-5B**). Samples taken during expression tests showed specific cells displaying high fluorescence signal diffuse throughout the cell, consistent with aggregates forming throughout the cytosol, when only $CsgA_{\Delta sec}$ was being expressed (**Figure 3-5C-F**). In contrast, samples taken during coexpression tests showed no diffuse fluorescence (**Figure 3-5G-J**). Instead, high density foci formed consistent with the formation of inclusion bodies. Occasionally after prolonged expression (> 4 hours) the inclusion bodies began to fluoresce, which could indicate aggregation beginning at these locations (**Figure 3-5J**). Taken collectively these phenotypes can be used to indicate whether a functional CsgC is being expressed *in vivo*.

3.3.5 E. coli CsgC and a Closely Related Homolog Have Significantly Different Efficiencies

The sequence or structural determinants that lead to CsgC's high *in vitro* efficiency remains unknown. Though CsgC was first identified in *E. coli, csg* operons from other Gammaproteobacteria contain CsgC homologs (Dueholm et al., 2012). Other members of the Proteobacteria phylum have a structural homolog called CsgH that features low sequence identity but high structural identity (Dueholm et al., 2012; Taylor et al., 2016). To better understand the effect of sequence identity on CsgC activity we cloned and tested several CsgC homologs, including the closely related homolog found in *Citrobacter youngae*, also a Gammaproteobacteria and member of the human microbiome alongside *E. coli*.

A sequence alignment shows CsgC from *E. coli* (CsgC EC) and its homolog from *C. youngae* (CsgC CY) have 17 divergent residues in their primary sequence (an 85% identity) (**Figure 3-6A**). Using AlphaFold 2.0 (Jumper et al., 2021) we obtained a predicted structure of CsgC CY (**Figure 3-6B**). A simple alignment of the CsgC EC structure (PDB:2Y2Y) and the predicted structure of CsgC CY using PyMol (Schrödinger, LLC, 2015) showed a 0.5 angstrom RMSD (**Figure 3-6B**).

Given the high sequence identity and high predicted structural homology between CsgC EC and CsgC CY, we assumed the *in vitro* activity of the two proteins would be similar. When tested under the same conditions, CsgC EC and CsgC CY show different inhibition efficiencies for *E. coli* CsgA aggregation (**Figure 3-6C and D**). For example, the 1:500 molar ratio condition for CsgC EC does not reach a ThT signal plateau within the time frame of the experiment (20 hours) (**Figure 3-6C**). In contrast, the 1:500 molar ratio condition for CsgC CY reaches a ThT signal plateau in less than 12 hours (**Figure 3-6D**). There must be an unknown effector within the 17 divergent residues between CsgC EC and CsgC CY that contributes to the discrepancy in

their *in vitro* efficiencies. We decided to use our growth and fluorescence phenotypes to test a range of CsgC EC variants to determine which of the 17 residue locations were significant.

3.3.6 E48 and R84 Could Play a Significant Role in CsgC Activity

The CsgC EC variants that were significantly different from CsgC EC WT were involved in an interstrand salt bridge between E48 and R84. We tested single amino acid variants of CsgC EC for their ability to perform similar to WT using an vivo screening assay that utilized the growth and fluorescence phenotypes described in **Figure 3-5** (**Figure 3-7**). While examining an alignment of CsgC EC and CsgC CY we noticed that CsgC EC had a glutamate and arginine at position 48 and 84, respectively, while CsgC CY had a glutamine at position 48 and a serine at position 84 (**Figure 3-6A**). The side chains of E48 and R84 likely form a polar contact according to the published crystal structure of CsgC (PDB: 2Y2Y) (**Figure 3-8A**). Among the variants tested were alanine mutants E48A and R84A as well as their counterpart variants E48R and R84E where the new residues represented a charge swap at that location (**Figure 3-7, Figure 3-8A**).

We cloned and purified the E48/R84 variants and tested their ability to inhibit *E. coli* CsgA amyloid formation *in vitro* (**Figure 3-8B-E**). The four CsgC variants were tested under the same conditions and the ThT assay results showed varying efficiencies. For example, the 1:500 molar ratio condition for E48A, E48R and R84E resulted in either no or minimal ThT signal increase within the time frame of the experiment (20 hours) (**Figure 3-8B-C and E**). The 1:500 molar ratio condition for R84A reached a ThT signal plateau in less than 8 hours (**Figure 3-8D**). We determined how all mutations affected the structural stability of CsgC EC by analyzing all CsgC EC variants' secondary structure at different temperatures using circular dichroism (**Figure 3-9**). All CsgC EC variants showed less thermal stability than CsgC EC WT, with CsgC

R84E being the least stable. Interestingly, the thermal stability of the variants did not correlate to their *in vitro* efficiency and the differences in efficiency are not dependent on the instability (**Figure 3-9**).

3.4 Discussion

Inhibiting amyloid formation is a topic of great interest among the protein folding disease community. Designed amyloid inhibitor proteins can come in many styles ranging a large biochemical space, from β-blocking D-amino acid peptides (Sievers et al., 2011) to camelid nanobodies (Chan et al., 2008; Giorgetti et al., 2018). There is a short list of naturally occurring proteins like CsgC that putatively inhibit functional amyloid formation. This list includes molecular chaperones and a diverse set of human and bacterial proteins that have similar 3D structures (Nagaraj et al., 2022). However, it is important to note how CsgC differs from the rest of the proteins in this group: CsgC remains the only example of an amyloid inhibitor protein being in an operon that exists specifically to control amyloid formation. Therefore, studying CsgC provides the unique opportunity to ask questions about how cells resolved to prevent amyloid formation that is otherwise wholly intentional.

3.4.1 CsgC Interaction with CsgA Monomer

Since Evans et al. published a report describing CsgC and its amyloid formation inhibition activity, the mechanism of action for CsgC has remained unclear. There are two potential competing mechanisms that rely primarily on the interaction partner of CsgC. CsgC could be interacting with CsgA monomers or with nascent fibers/oligomers. A solution containing a fast-aggregating protein like CsgA is in a constant state of flux towards amyloid fiber formation and therefore it is difficult to isolate any pre-fibrillar species of CsgA. Previous attempts to measure

the interaction between CsgC and CsgA monomers have been reportedly difficult (Evans et al., 2015; Nagaraj et al., 2022; Sleutel et al., 2017).

The pull-down chromatograms shown in **Figure 3-1A-C** indicate an interaction between CsgC and CsgA monomers, albeit a weak and transient interaction. The peak elution volume of CsgA monomers (~12 mL) after passing through the CsgC-modified NHS column was very similar to that of the BSA control column, a difference of only 0.8% (**Figure 3-1B and C**). This would suggest that immobilized CsgC is not making a lasting interaction with CsgA monomers, as the soluble monomers pass through the column mostly unhindered. Though the elution volume changed very little, the shape of the CsgA monomer peaks were different (**Figure 3-1A-C**). After passing through the CsgC-modified NHS column, the CsgA monomer peak was much taller and sharper than the same peak in the BSA control (**Figure 3-1C**). Additionally, after passing through the CsgC-modified NHS column there was a loss of any high molecular weight or oligomeric CsgA (peak at ~8 mL) (**Figure 3-1C**). Diminished CsgA oligomers suggests that CsgC is either filtering the oligomers out of solution or breaking the oligomers apart. Breaking apart the oligomers could be why the monomer peak is so much taller and sharper than the BSA control experiment (**Figure 3-1C**). The chromatograms are very telling, however, ThT assays results are also striking (**Figure 3-1D and E**). The sample of CsgA monomer that was taken from the CsgC-modified NHS column pull down aggregated much slower than the no NHS column condition. Even after 48 hours, the CsgA monomers that passed by the CsgC-modified NHS column never reach the same level of ThT signal as the no-NHS column condition (**Figure 3-1D and E**). Therefore, the weak interaction that CsgC makes with CsgA monomers causes them to delay aggregation.

We endeavored to learn more about the CsgC and CsgA monomer interaction by probing with IM-MS. When a sample of a 1:1 CsgA:CsgC mixture was analyzed, peaks in the mass spectrum consistent with a 1:1 CsgA:CsgC heterodimer were present in the first reading (**Figure 3-2B**). The heterodimer peaks were in low abundance, consistent with the concept of a transient interaction (**Figure 3-2B**). The heterodimer peaks were our first direct observation of an interaction between CsgC and CsgA monomers in solution.

Just like the pull-down assays, much can be learned from focusing on the IM-MS signals associated with CsgA monomers. Multiple charge states of CsgA monomers can be seen in the mass spectrum, consistent with an intrinsically disordered protein (**Figure 3-2A**). We analyzed how the CsgA monomers shape changed by examining the ATD over time (**Figure 3-2C, Figure 3-3C-D**). The ATD profile suggests that the conformation of CsgA became more extended as the aggregation progresses, a change commonly observed with other amyloid proteins as well (Hyung et al., 2013; Soper et al., 2013) (**Figure 3-2C, Figure 3-3C-D**). Remarkably, the presence of CsgC in solution causes the ATD of CsgA monomers to remain consistent throughout the time course of our experiment (**Figure 3-2C, Figure 3-3C-D**). CsgC must interact with an intrinsically disordered CsgA monomer and maintain that fold state. Our model of CsgC interacting with unstructured CsgA and maintaining an unstructured state fits within greater context of curli biogenesis. CsgC is a periplasmic protein tasked with inhibiting amyloid formation within the intermembrane space, a cellular compartment where only unfolded CsgA is supposed to reside.

CsgC remains an active amyloid inhibitor during the rapid fiber formation phase of aggregation. We show two similar experiments in **Figure 3-4** where CsgC is presented to CsgA during the elongation phase. In the first experiment, CsgC is added to CsgA at different times

throughout aggregation and we monitored the effect of CsgC addition on the increase in ThT fluorescence (**Figure 3-4A**). In the experiment shown, the uninhibited CsgA takes approximately 1.3 hours to transition from the lag to the elongation phase (**Figure 3-4A**). When CsgC is added at time 0, we see little to no increase in ThT fluorescence throughout the time course of the experiment (**Figure 3-4A**). For all conditions tested where CsgC was added after time 0 there was a linear increase in ThT signal, until the signal eventually reached a plateau and flattened out (**Figure 3-4A**). Similarly shaped curves can be seen when CsgC is added to a seeded CsgA aggregation reaction (**Figure 3-4B**). Linear or isodesmic polymerization is consistent with a model of aggregation that is independent of nucleation (Frieden, 2007). Nucleation-dependent polymerization requires the formation of high energy nuclei, a bottleneck for polymerization, followed by the very favorable addition of monomers to a growing fiber (Jain and Udgaonkar, 2011). CsgC-inhibited CsgA could present an isodesmic polymerization profile because the formation of new nuclei is being hindered but CsgA unfolded monomers can still add to fiber ends. CsgC could be fighting the early process of CsgA transitioning between the disordered-denatured state and disordered-collapsed state described by Frieden (Frieden, 2007). When CsgC is added during the elongation phase, any disordered-collapsed CsgA that is able to form is quickly added onto fiber ends instead of forming a new nucleus.

3.4.2 Insights into CsgC Activity

We sought to shed light how the CsgC-CsgA interaction is taking place. Taylor and coworkers discussed the importance of electrostatic interactions between CsgC and its binding partner, remarking on the significance of solvent exposed basic residues on CsgC (Taylor et al., 2016). Otherwise, there is no direct evidence or predictions regarding which CsgC residues are required for its activity.

In vivo phenotypes are useful tools to screen for functional or non-functional variants of a protein. The Congo Red binding assay is the most widely used test for verifying the functionality of *csg* proteins (Evans et al., 2018). During this assay, all csg proteins act in concert in order for the cell to secrete CR+ curli to the cell surface. Attempts to use Congo Red binding as a phenotype for CsgC activity have previously failed; Δ*csgC* cells remain CR+ (Evans et al., 2015). Knowing this, we turned to another method for assessing *in vivo* CsgA amyloid formation and its inhibition.

It is widely known that intracellular amyloid formation is cytotoxic and these observations extend to functional amyloids being expressed within *E. coli* (Marinelli et al., 2016). Indeed, the over expression of $CsgA_{\Delta sec}$ within *E. coli* BL21 leads to an apparent cellular toxicity (**Figure 3-5A**). To confirm the formation of amyloid within cells after expression, we added AmyTracker 680 into the growth media and saw strong fluorescence starting at the same time as the decrease in OD600 (**Figure 3-5B**). Therefore there is a correlation between cytotoxicity and amyloid formation after $CsgA_{\Delta sec}$ is expressed. Coexpression of $CsgC_{\Delta sec}$ rescues from that toxicity and greatly decreases AmyTracker 680 fluorescence (**Figure 3-5A and B**). Samples were taken during the growth assays and visualized using fluorescence microscopy to get additional information at a cellular resolution. In the strain expressing $CsgA_{\Delta sec}$ only we saw diffuse AmyTracker 680 fluorescence throughout certain cells, signaling that aggregation was happening everywhere within their cytoplasms (**Figure 3-5C-F**). This was unexpected since we anticipated that there would be foci of amyloid signal in well-defined plaque-like bodies, consistent with a-syn plaques stained with the same dye (Mahul-Mellier et al., 2020). Instead, foci formed in the phase contrasted images when $CsgC_{\Delta sec}$ was coexpressed (**Figure 3-5G-J**).

This would suggest that the presence of CsgC within the cytoplasm is allowing the cell to process CsgA into protein aggregates consistent with inclusion bodies.

3.4.3 Sequence Determinants of CsgC Activity

CsgC is a semi selective amyloid inhibitor and displays a preference for particular amyloidogenic proteins (Evans et al., 2015). Taylor et al. showed that mutating some of the charged residues on CsgC can decrease its inhibition efficiency for CsgA (Taylor et al., 2016). Therefore, there must be some sequence determinants in CsgC and its client proteins that direct amyloid inhibition. There are no known binding sites, active sites, or locations of interest between CsgC and its clients. We turned to sequence comparisons to guide our efforts to better understand significant CsgC residues.

The CsgC homolog from *Citrobacter youngae* is very similar to CsgC EC in both sequence identity (85%) (**Figure 3-6A**) and predicted structural homology (**Figure 3-6B**). With these similarities in mind, we assumed CsgC EC and CsgC CY would perform with similar efficiencies when tested *in vitro* under the same conditions. However, CsgC CY was much less efficient (**Figure 3-6C and D**). This led us to explore which of the 17 divergent residues in CsgC CY lead to this decrease in efficiency.

We made single amino acid variants of CsgC that corresponded to each of the 17 divergent residues in CsgC CY and tested them in an *in vivo* screening assay to determine which performed similar to CsgC EC. An "optotracing" graph was created to better compare both growth and fluorescence data for each variant tested (**Figure 3-7**)(Butina et al., 2020). The conditions of the *in vivo* screening assay drove the expression of $CsgA_{\Delta sec}$ and $CsgC_{\Delta sec}$ by equal amounts into the cytoplasm since the two expression plasmids were both pET vectors using IPTG as the chemical inducer. Four variants stood out among those tested: E48A, E48R, R84A,

and R84E. The differences between the four variants and the wildtype protein and their ability to suppress toxicity and AmyTracker 680 fluorescence were slight but discernable (**Figure 3-7**). The four variants in question were alanine and charge swap mutants of two CsgC EC residues: E48 and R84. These residues form a predicted salt bridge on the surface of CsgC (**Figure 3-8A**). Therefore, we assumed the salt bridge played a significant role in CsgC activity since the four variants performed worse than CsgC EC in the *in vivo* screening assay.

Remarkably, 3 out of 4 variants tested *in vitro* were even more efficient than CsgC EC WT (**Figure 3-8C**). R84A was the only CsgC variant tested that showed a greatly diminished efficiency. Suggesting that the presence of the arginine side chain plays a significant role in the activity of CsgC. Reversals of charge at E48 and R84, which resulted in two basic or two acidic side chains being adjacent in the fully folded protein, lead to mild increases in efficiency. This would suggest that the proper positioning of the E48/R84 sidechains is important in CsgC activity, an observation consistent with previous previously published work pointed out CsgC's dependence on electrostatic interactions (Taylor et al., 2016). The single amino acid changes only minorly effected the thermal stability of the CsgC variants (**Figure 3-9**). This would suggest that the effect of the changes in residue are more closely associated with loss of activity rather than changing the structure of the CsgC protein. More investigation as to how these residues affect stability and activity of CsgC need to be performed.

3.5 Materials and Methods

Bacterial Growth

All overnight cultures were grown in LB supplemented with 100 μg/mL ampicillin and/or 50 μg/mL of kanamycin at 37 °C with shaking at 220 rpm. When necessary, LB plates were supplemented with ampicillin 100 μg/mL or kanamycin 50 μg/mL.

Strains and Plasmids

The full list of strains, plasmids, and primers can be found in the Supplemental Materials. A library of expression vectors containing *csgC* mutants was purchased and synthesized by GenScript (https://www.genscript.com/). When done in house, site directed mutagenesis was performed on pre-existing plasmids using the Phusion HF polymerase (Cat No. M0530S). Primers were designed using SnapGene (SnapGene® by Dotmatric) and purchased by IDT (https://www.idtdna.com). Mutagenized plasmids were constructed in the MC1061 cell background. Correct mutations were confirmed using Sanger sequencing provided by eurofins (https://www.eurofins.com/genomic-services/our-services/custom-dna-sequencing/). Plasmids were extracted from transformants using Promega PureYield™ Plasmid Miniprep System (Cat No. PRA1223). Miniprep plasmids were transformed into an expression strain (NEB3016) for purification or BL21 (DE3) for *in vivo* screening.

Protein Purification

CsgA was purified as described previously (Zhou et al., 2012). CsgC and its variants were purified as described previously (Salgado et al., 2011). Size exclusion chromatography was performed as previously described (Evans et al., 2018). Briefly, Ni-NTA affinity chromatography elution fractions were pooled, concentrated, and passed through a 0.22 μm filter. The sample underwent gel filtration using a Superdex 75 10/300 GL column (Cat No.45-002-903) attached to an Äkta pure protein purification system. Elution fractions were assayed for protein concentration using A220. Samples corresponding to elution peaks were analyzed using SDS-PAGE.

Denaturing Gel Electrophoresis and Western Blot

Purified protein samples were diluted in 4X SDS loading buffer run on a 15% SDS PAGE gel (Evans et al., 2018). The gels were stained with Coomassie blue dye to visualize protein bands or the proteins were transferred to a PVDF membrane for blotting. Western Blot were performed as previously described (Evans et al., 2018). Briefly, blots were probed with a primary antibody against CsgA (1:12,000) or CsgC (1:4000). Secondary antibodies against rabbit IgG and conjugated with IRDye 800CW (Cat No.NC9401842) were used to image the blots in a Licor Odyssey FC.

Preparation of modified NHS columns

CsgC was purified using a previous protocol (Salgado et al., 2011). 1 mg of purified CsgC in 50 mM phosphate buffer, pH = 7.4 or 1 mg of purified BSA was injected into an NHS activated HP column (Cytiva 17071602) and incubated for 1hr at room temperature. Unreacted NHS was blocked by injecting a blocking buffer (200 mM Tris HCl, pH 7.5) and incubating for 30 mins at room temperature. The modified NHS columns were rinsed with 10 CV of 20mM NaOH and 10 CV of 6M urea to wash away any unbound proteins (eluent tested for protein concentration with Coomassie blue). Modified NHS columns were stored at 4 0C until further use.

Gel Filtration and Aggregation Assay

Frozen affinity purification elution fractions were thawed and injected into an AKTA pure protein purification system to be gel filtered using a Superdex 75 10/300 GL column (Cat No.45-002-903) with an isocratic running buffer (50 mM phosphate buffer, pH = 7.4). Additional separations were performed that included a modified NHS column (CsgC or BSA) attached in line after the SEC column. Elution fractions containing CsgA species of different sizes were assed using UV absorbance, collected, and stored on ice until further analysis. ThT binding assays were performed by standardizing CsgA monomer concentration to 10 μM (Bradford

Assay used to calculate concentration) in phosphate buffer and including an excess of ThT. Fluorescence measurements were performed at 37 °C in 96 microtiter-plate, (black walls and clear bottoms, Corning, Kennebunk, ME, USA) using a FLUOstar Omega microplate reader (BMG Labtech GmbH, Ortenberg, Germany) with an excitation wavelength of 430 nm and an emission wavelength of 480 nm, the samples were shaken for 1s at 300 rpm every 15 mins between readings. All experiments were performed in triplicate or more and each experiment has been verified two times or more.

Sample Preparation for IM-MS

Purified CsgA and CsgC from E. coli were buffer exchanged into 20 mM ammonium acetate (pH 7.4) using Thermo Scientific Zeba™ Spin Desalting Columns 7k MWCO. The protein concentration after buffer exchange was assayed using Thermo Scientific Pierce™ Rapid Gold BCA Protein Assay Kit. CsgA and CSgC are each diluted to20 µM with 20 mM ammonium acetate, mixed at 1:1 ratio and the mixture were incubated at 37°C for 23 hours. Time points were taken at 0 hr, 3 hr, 6hr and 23hr.

IM-MS

IM-MS data was collected on a quadrupole ion-mobility time-of-flight (TOF) mass spectrometer (Synapt G2 HDMS, Waters, Milford, MA, USA) with a nano-electrospray ionization (nESI) source. The source was operated at positive mode with the nESI voltage set at 1.0-1.2 kV, the sampling cone was set to 15 V and the bias was set to 42 V. The source temperature was set to 20°C. The traveling–wave ion mobility separator was operated at a pressure of approximately 3.4 mbar with wave height and wave velocity set at 30V and 500 m/s, respectively. The m/z window was set from 100 – 8000 m/z with a TOF pressure of 1.5e-6 mbar. Mass spectra were analyzed using MassLynx 4.1 and Driftscope 2.0 software (Waters, Milford, MA, USA). CCS (Ω)

measurements were externally calibrated using a database of known values in helium. We reported the standard deviations from replicate measurements of CCS and an additional ±3% to incorporate the errors involves in the calibration process.

Thioflavin T Binding Assay

Assays were performed as previously described (Zhou et al., 2012). Briefly, freshly purified CsgA was diluted with phosphate buffer to 20 μM and combined with an excess of the amyloid-specific dye thioflavin-T (ThT) (Cat No. AC211760250). Amyloid formation was monitored by measuring an increase in ThT fluorescence at 495 nm (450 nm excitation). Assays were performed in triplicate at microscale within 96-well plates and measured with Infinite Pro M200 or Infinite Nano^{+} F200 Tecan plate readers. CsgC proteins were purified, diluted from a stock solution (phosphate buffer), and added to specified assays in the reported stoichiometric ratio. Fiber stimulating "seeds" were produced by obtaining previously purified CsgA WT fibers and sonicated directly before adding to a specified assay.

***In Vivo* Screening Assay**

3 mL overnight cultures were created in LB containing the appropriate selective agent and incubated at 37°C with shaking for 16 hr. The following day, cell density was normalized by diluting overnight cultures to OD600 = 1 using fresh media. The normalized cultures were diluted to OD = 0.01 into the wells of a sterile 96-well microplate to a final volume of 200 uL with sterile LB supplemented with the amyloid specific dye AmyTracker 680 (1:500) (EbbaBiotech). The microplate was placed in a Molecular Devices iD3 microplate reader and absorbance (600 nm) and fluorescence (Ex:550 nm, Em:680 nm) was measured every 15 mins for 16 hr. The plate was incubated at 37 °C with shaking between spectroscopic readings. When

the liquid cultures had grown for 2 hr, ITPG was added to a final concentration of 0.5 mM (diluted from 0.1 M frozen stock) to induce the expression of all plasmids.

Fluorescence Microscopy

Live-cell microscopy was performed at 0, 1, 2, and 4 hour post-induction. 4 µL of cells were dropped onto a piece of 2% UltraPure agarose + LB pad and imaged on a Mantek dish. All fluorescence and phase contrast imaging was performed using a Nikon Ti2-E motorized inverted microscope controlled by NIS Elements software with a SOLA 365 LED light source, a 100X Objective lens (Oil CFI Plan Apochromat DM Lambda Series for Phase Contrast), and a Hamamatsu Orca Flash 4.0 LT + sCMOS camera. AmyTracker 680 was imaged using a "TexasRed" filter set (C-FL Texas Red, Hard Coat, High Signal-to-Noise, Zero Shift, Excitation: 560/40 nm [540-580 nm], Emission: 630/75 nm [593-668 nm], Dichroic Mirror: 585 nm).

Thermostability Assay using Circular Dichroism (CD)

CsgC variants were purified as described above and stored at 4°C in 50 mM KP_i, pH 7.3. Secondary structure was analyzed using a Jasco J-1500 CD Spectrometer by scanning from 190 to 260 nm at 20°C in a quartz cell with a 1 cm path. A second CD spectrum was obtained after the cell holder was heated to 90 °C (5 °C/min).

Author Contributions

AB and MC contributed conception and design of the study. AB, YH, SJ, EG, LK, RA, and LT performed the experiments. AO, BR, and AV provided materials and expertise. AB wrote the manuscript; AB and MC contributed to editing and revising the manuscript. We thank members of the Chapman, Vecchiarelli, Olofsson, and Ruotolo labs for helpful suggestions.

Funding

This work was supported by the National Institutes of Health grant no. R01 GM118651 and R21 AI137535 to M.R.C.

3.6 Figures

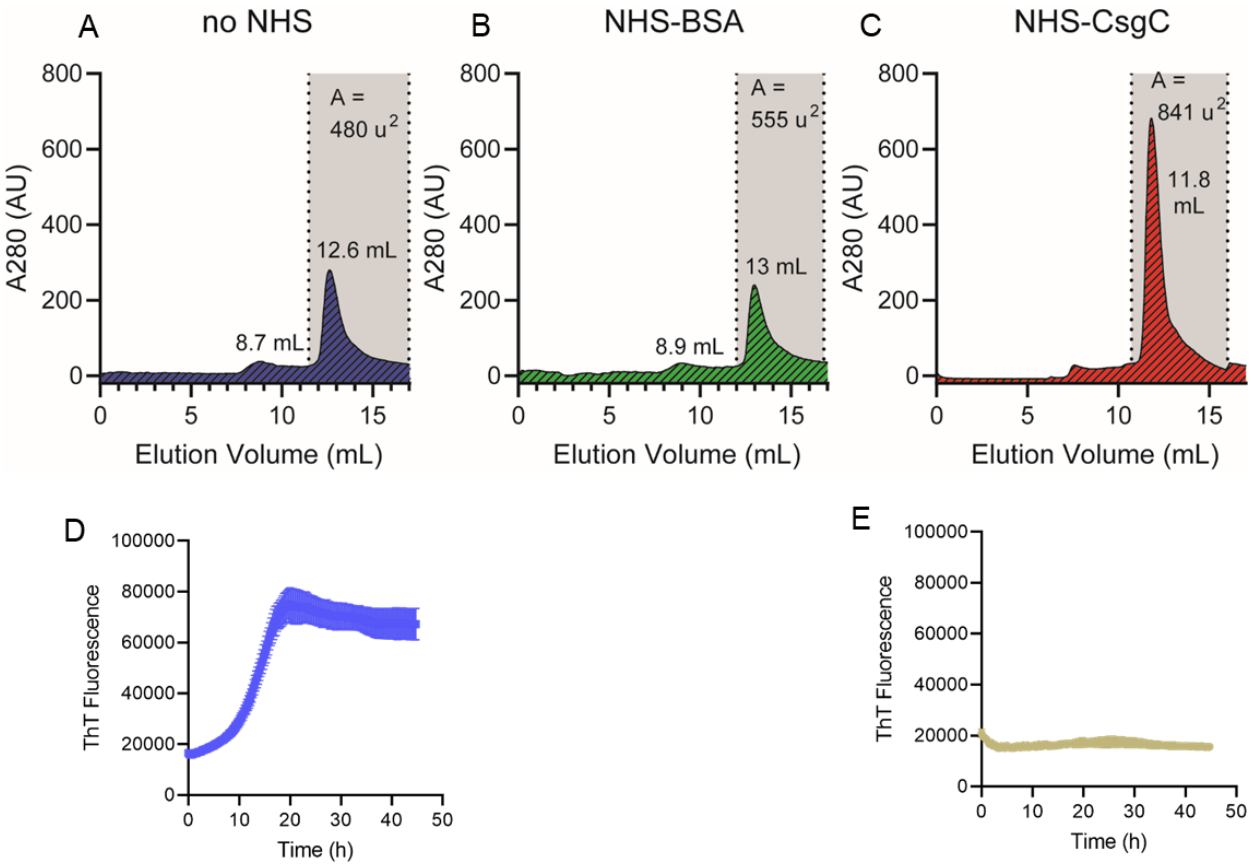

Figure 3-1. CsgC makes a weak, transient interaction with CsgA that delays ThT positive amyloid formation.
Affinity purified CsgA was passed through **A)** an unmodified NHS column, **B)** a column modified with BSA, **C)** or a column modified with CsgC. Afterward the CsgA was separated by gel filtration and the elution of protein was tracked using UV absorbance. Significant peaks were labeled and the area under the peaks defined by dashed lines was calculated. SEC fractions taken from the **D)** unmodified and **E)** NHS-CsgC separations that correspond to monomeric CsgA were moved to a ThT assay to assess the protein's propensity to form amyloid.

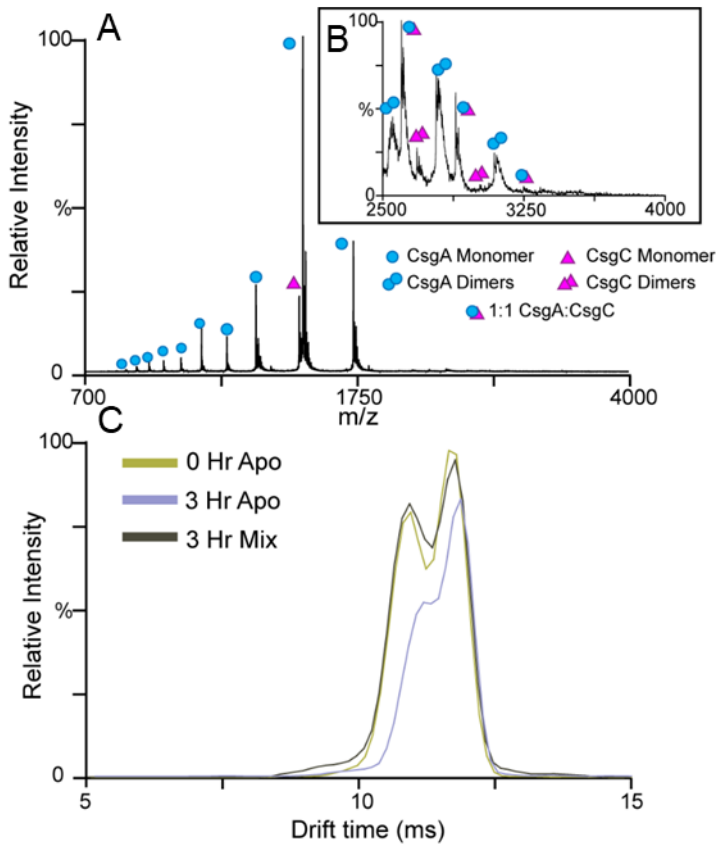

Figure 3-2. IM-MS captures a 1:1 CsgA and CsgC heterodimer.
A) Mass spectra for CsgA incubated with CsgC in a 1:1 molar ratio at 37 °C. Monomeric CsgA (blue circles), dimeric CsgA (blue double circles), monomeric CsgC (purple triangles), dimeric CsgC (double purple triangles) and 1:1 CsgA:CsgC complex (blue circle and purple triangle) is annotated. A magnified MS spectrum **(B)** showed that the complex is seen flanked on either side by dimeric CsgA and dimeric CsgC. **C)** Arrival time distribution of the 11+ monomer of CsgA in three experimental conditions: Apo CsgA in solution at 0 hr (olive trace), apo CsgA in solution after 3 hr (light purple trace), and CsgA in solution with CsgC after 3 hr (dark grey trace). Traces are overlapped to show the changes in ATD after 3 hr of incubation.

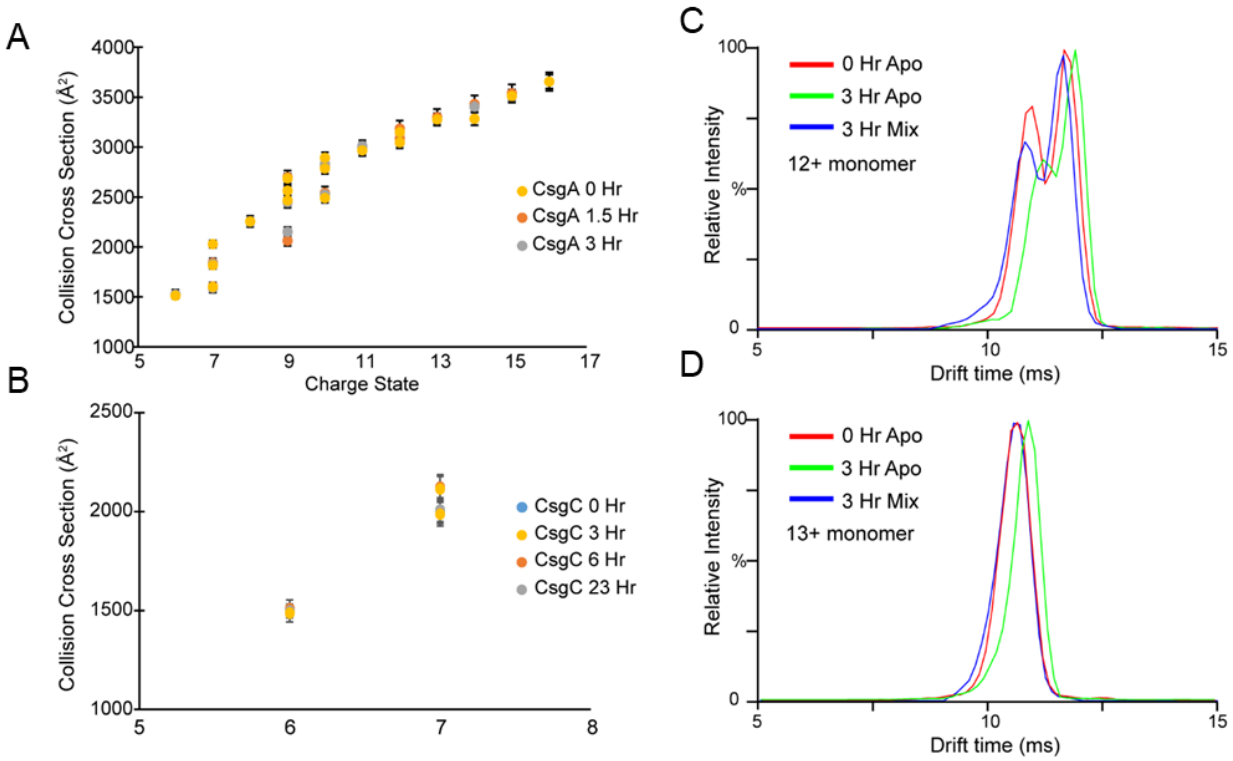

Figure 3-3 Additional IM-MS Data.
CCS values derived from arrival time distribution of **A)** CsgA monomers and **B)** CsgC monomers. **C and D)** Arrival time distribution of the 12+ and 13+ monomer of CsgA in three experimental conditions: Apo CsgA in solution at 0 hr (red trace), apo CsgA in solution after 3 hr (green trace), and CsgA in solution with CsgC after 3 hr (blue trace). Traces are overlapped to show the changes in ATD after 3 hr of incubation.

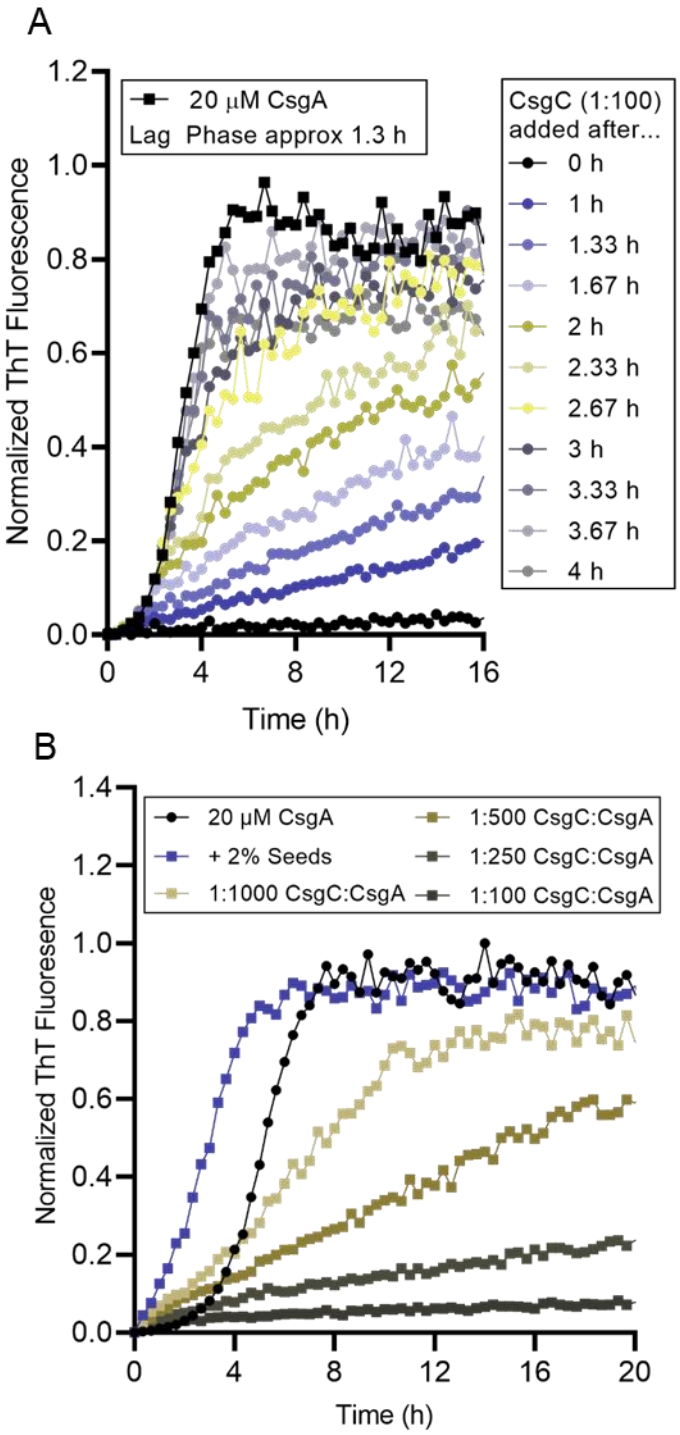

Figure 3-4 CsgC inhibits amyloid formation during the fiber elongation stage.
A) When CsgC is added to actively polymerizing CsgA, fiber formation is weakly inhibited. This scenario can be simulated by "seeding" freshly purified CsgA through the addition of sonicated preformed CsgA fibers. **B)** When CsgC is added to a seeded reaction of CsgA it also inhibits weakly (All olive conditions also contain 2% seeds).

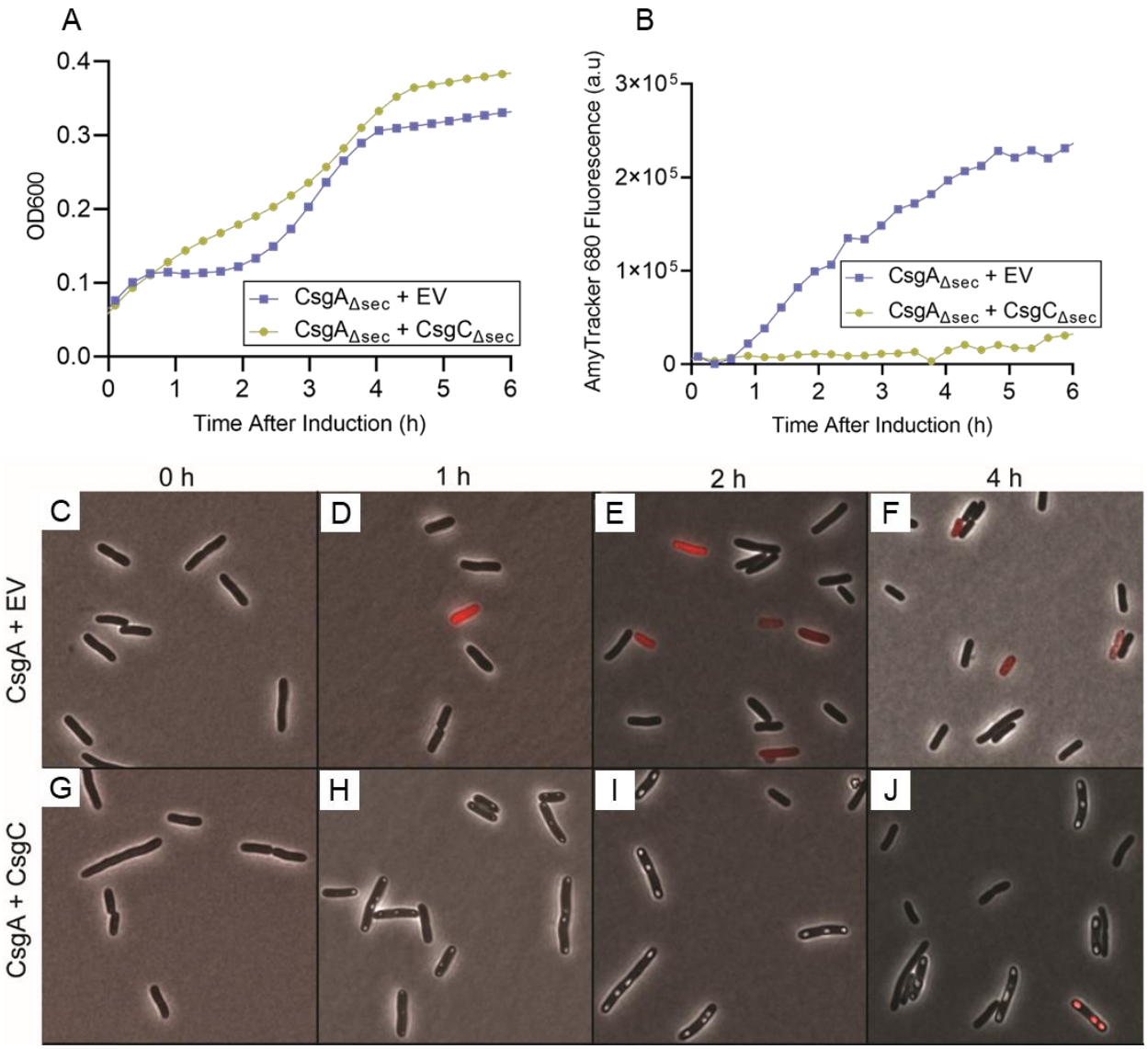

Figure 3-5 CsgC activity can be assessed through at set of *in vivo* phenotypes.
When CsgA is expressed within the cytoplasm of E. coli cells, the cells exhibit a defective growth rate that begins approximately 1 hour after induction. The cells also bind to an amyloid-specific and cell-permeable fluorescent dye called Amytracker when the dye is added to the growth media. The coexpression of CsgC with CsgA leads to **A)** a rescue from the amyloid-mediated toxicity and **B)** a decrease in dye fluorescence within the microplate wells. **C-J)** Samples were taken at hour intervals during the growth experiment and imaged using fluorescence microscopy.

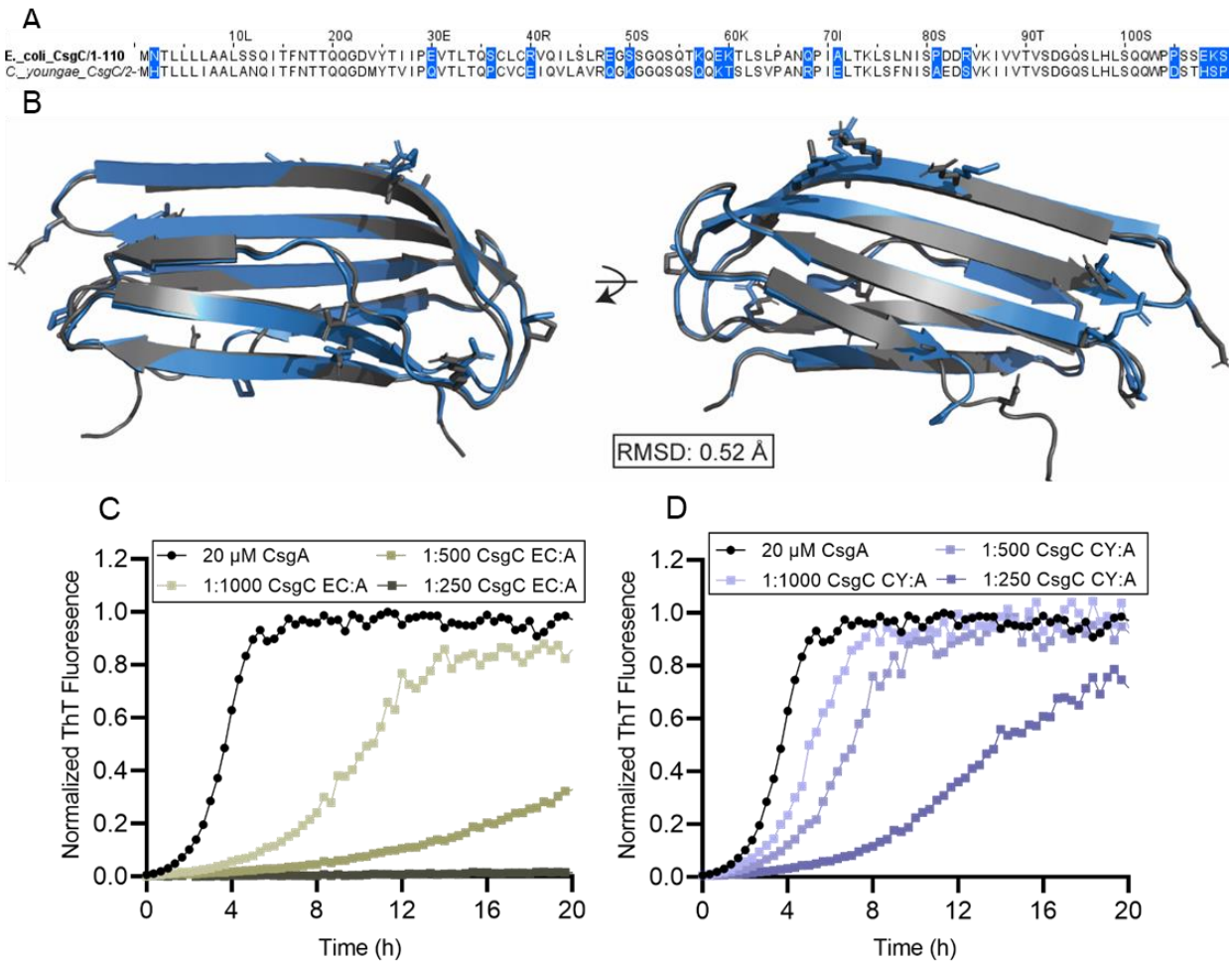

Figure 3-6 E. coli CsgC has closely related homolog which displays a greatly diminished *in vitro* activity.

A) A sequence alignment of E. coli CsgC and C. youngae CsgC. The 17 divergent residues are labeled in blue. **B)** When comparing E. coli CsgC structure (PDB:2y2y) and a predicted structure of C. youngae CsgC produced by AlphaFold 2.0, the two proteins have vary familiar fold. We displayed the side chains of the 17 divergent residues as sticks. **C-D)** The CsgC homologs are substoichiometric inhibitors of amyloid formation of E. coli.

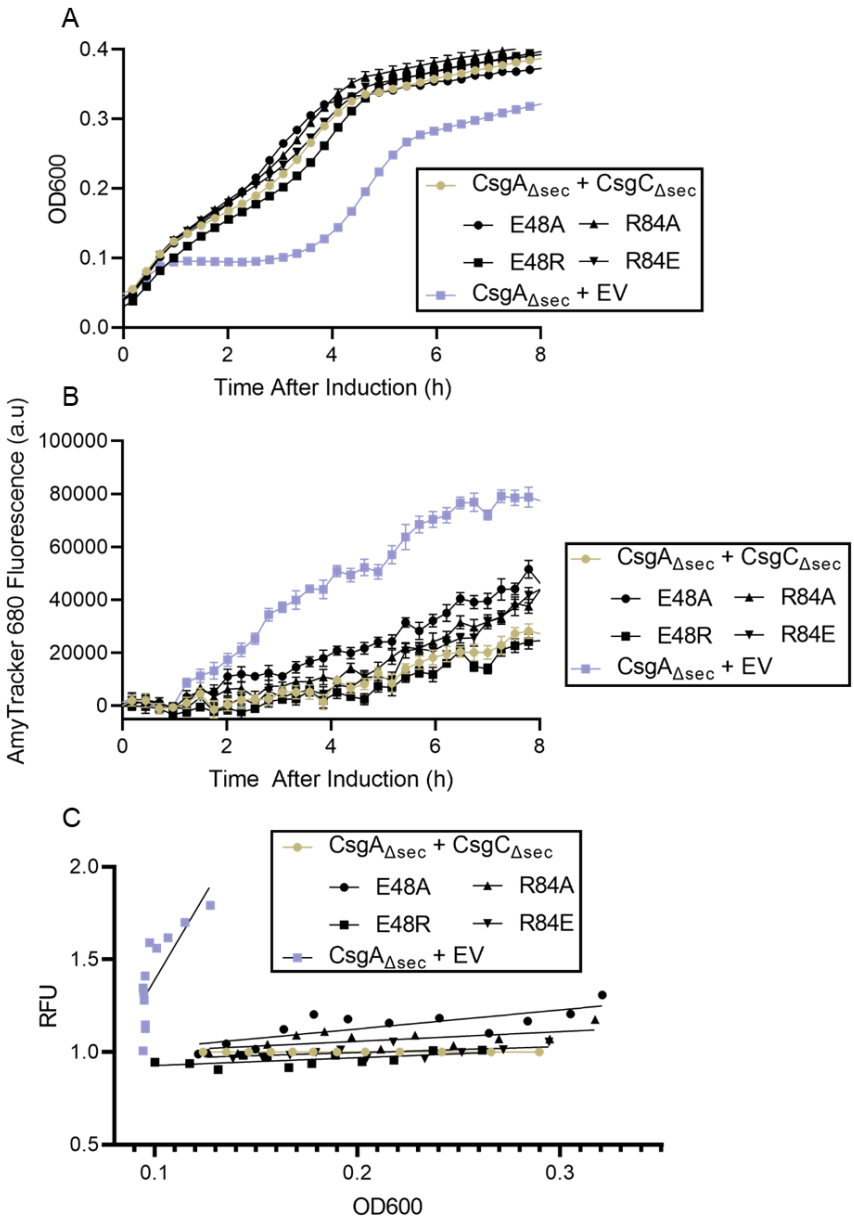

Figure 3-7 *In vivo* screening assay results for CsgC variants E48A, E48R, R84A, and R84E. For each condition an *E. coli* BL21 strain was created harboring two plasmids each encoding for the expression of CsgA$_{\Delta sec}$ and the wildtype or a variant of CsgC$_{\Delta sec}$. The **A)** growth curves and **B)** fluorescence of each culture was tracked after protein expression was induced. **C)** An optotracing plot was created according to (Butina, 2020) to compare the growth and fluorescence of each culture in the same dataset. All fluorescence values are relative to CsgC WT, therefore any strain that displays a positive optotracing slope over this OD range is allowing more amyloid formation than the wildtype protein. Lines shown are simple linear regression lines.

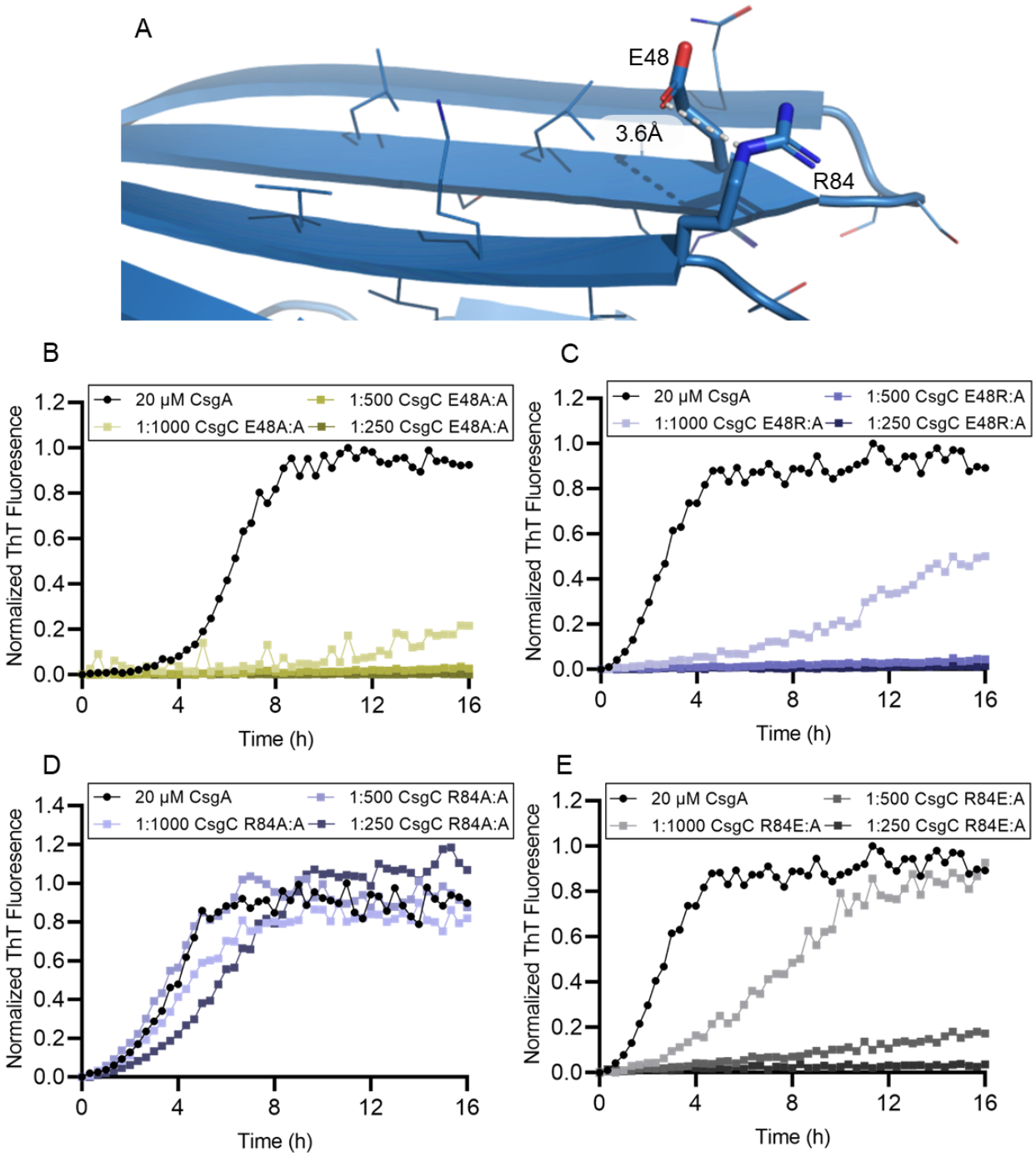

Figure 3-8 CsgC EC single residue variants E48A, E48R, R84A, R84E have varying inhibition efficiencies.

A) Structure of CsgC EC showing E48 and R84 as stick models. The measured distance between side chain atoms was 3.6 Å indicating a potential ionic interaction. ThT assays showing the addition of CsgC **B)** E48A, **C)** E48R, **D)** R84A, and **E)** R84E in varying molar ratios to freshly purified *E. coli* CsgA.

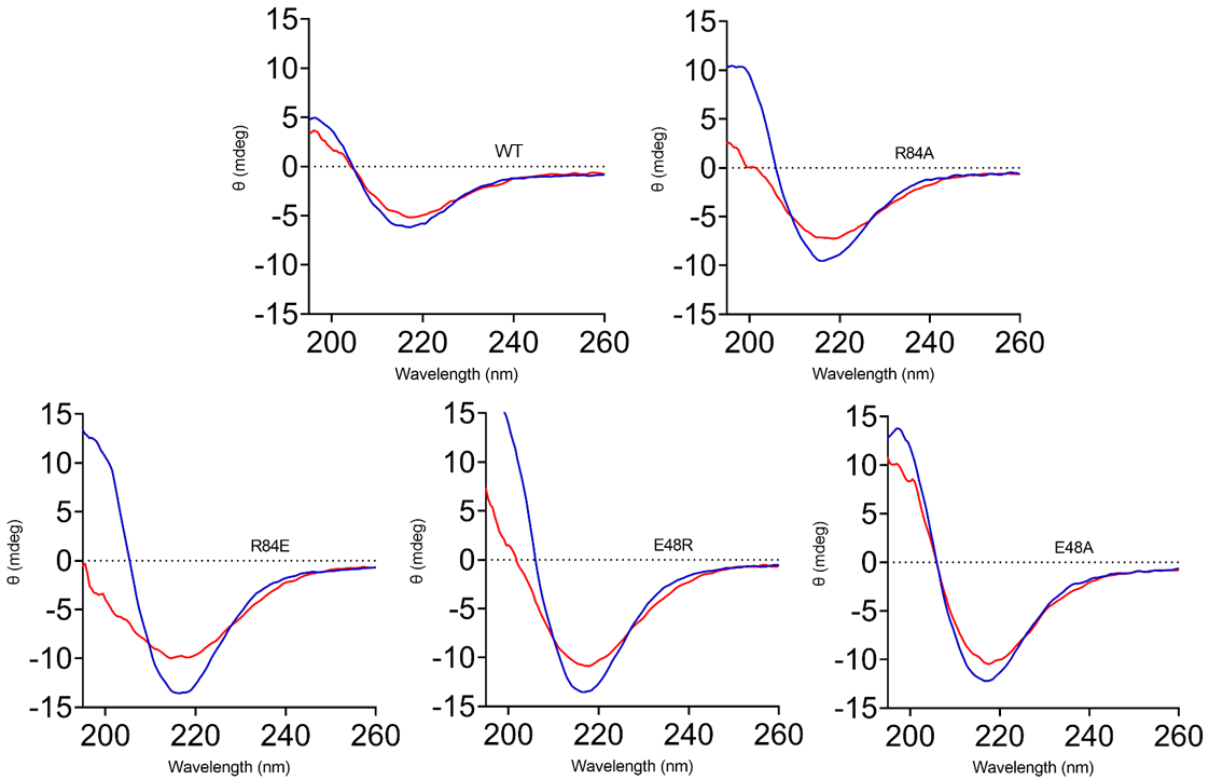

Figure 3-9 Thermostability tests for CsgC WT the four single residue variants E48A, E48R, R84A, R84E.

CsgC WT and its variants were purified in the same manner and stored at 4 °C. The five proteins were standardized to 10 µM concentration in the same buffer and their secondary structure was visualized by scanning far UV absorbance. CD spectra were taken at 20 °C (blue curve), the samples were heated in the cuvette, and second measurements were taken at 90 °C (red curve).

3.7 Tables

Table 3-1 Strains used in this study.

Strains	Relevant Genotype	References
BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS* λ DE3 = *λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	New England Biolabs
MC1061	*F– araD139 Δ(ara-leu)7696 galE15 galK16 Δ(lac)X74 rpsL (StrR) hsdR2 (rK– mK+) mcrA mcrB1*	(Casadaban and Cohen, 1980)
NEB3016	MiniF *lacI*^{q}(Cam^{R}) */ fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--*Tet^{S}*)2 [dcm] R(zgb-210::Tn10--*Tet^{S}*) endA1 Δ(mcrC-mrr)114::IS10*	New England Biolabs
ALB30	BL21(DE3) + pET28a EV + pET11d-CsgA -sec 6xHis	This study
CsgC WT	BL21(DE3) + pET28a-CsgC-sec C-term 6xHis, kan^{r} + pET11d-CsgA -sec 6xHis	This study
ALB30	BL21(DE3) + pET28a EV + pET11d-CsgA -sec 6xHis	This study
E48A	BL21(DE3) + pET28a-CsgC E48A-sec C-term 6xHis, kan^{r} + pET11d-CsgA -sec 6xHis	This study
E48R	BL21(DE3) + pET28a-CsgC E48R-sec C-term 6xHis, kan^{r} + pET11d-CsgA -sec 6xHis	This study
R84A	BL21(DE3) + pET28a-CsgC R84A-sec C-term 6xHis, kan^{r} + pET11d-CsgA -sec 6xHis	This study
R84E	BL21(DE3) + pET28a-CsgC R84E-sec C-term 6xHis, kan^{r} + pET11d-CsgA -sec 6xHis	This study

R84E	BL21(DE3) + pET28a-CsgC R84E-sec C-term 6xHis, kan^r + pET11d-CsgA -sec 6xHis	This study
EMG2	NEB 3016 ΔslyD + pET28a-CsgC E48A C-term 6xHis	This study
EMG4	NEB 3016 ΔslyD + pET28a-CsgC R84A C-term 6xHis	This study
EMG6	NEB 3016 ΔslyD + pET28a-CsgC R84E C-term 6xHis	This study
EMG7	NEB 3016 ΔslyD + pET28a-CsgC E48R C-term 6xHis	This study
RAL1	NEB 3016 ΔslyD + pET28a-*Citrobacter youngae* CsgC-sec C-term 6xHis	This study

Table 3-2 Plasmids used in this study.

Plasmids	Relevant Characteristics	References
pET11d	IPTG inducible expression vector	New England Biolabs
pET28a	IPTG inducible expression vector	New England Biolabs
pCsgA	pET11d-CsgA -sec 6xHis	(Zhou, 2012)
pCsgC	pET28a-CsgC-sec C-term 6xHis, kan^r	(Evans, 2015)
pCsgC E48A	pET28a-CsgC E48A C-term 6xHis	This study
pCsgC R84A	pET28a-CsgC R84A C-term 6xHis	This study
pCsgC R84E	pET28a-CsgC R84E C-term 6xHis	This study

pCsgC E48R	pET28a-CsgC E48R C-term 6xHis	This study
pCsgC CY	pET28a-*Citrobacter youngae* CsgC-sec C-term 6xHis	This study

Table 3-3 Primers used in this study.

Primers	Primer Sequence (5'→ 3')	Constructs
GA CsgC CY Vector_FOR	ATTTGATTTGCCAGCGCCATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAG	CsgC CY
GA CsgC CY Vector_REV	CTGACTCAACCCATAGCCCACACCACCACCACCACCACT	CsgC CY
GA CsgC CY Frag_FOR	CAGTGGTGGTGGTGGTGGTGTGGGCTATGGGTTGAGTCAGGC	CsgC CY
GA CsgC CY Frag_REV	TAAGAAGGAGATATACCATGGCGCTGGCAAATCAAATAACATTCAA	CsgC CY

3.8 References

Balistreri, A., Goetzler, E., and Chapman, M. (2020). Functional Amyloids Are the Rule Rather Than the Exception in Cellular Biology. Microorganisms *8*, 1951. https://doi.org/10.3390/microorganisms8121951.

Butina, K., Tomac, A., Choong, F.X., Shirani, H., Nilsson, K.P.R., Löffler, S., and Richter-Dahlfors, A. (2020). Optotracing for selective fluorescence-based detection, visualization and quantification of live S. aureus in real-time. Npj Biofilms Microbiomes 2020 61 *6*, 1–12. https://doi.org/10.1038/s41522-020-00150-y.

Chan, P.H., Pardon, E., Menzer, L., De Genst, E., Kumita, J.R., Christodoulou, J., Saerens, D., Brans, A., Bouillenne, F., Archer, D.B., et al. (2008). Engineering a camelid antibody fragment that binds to the active site of human lysozyme and inhibits its conversion into amyloid fibrils. Biochemistry *47*, 11041–11054. https://doi.org/10.1021/BI8005797/ASSET/IMAGES/MEDIUM/BI-2008-005797_0010.GIF.

Chapman, M.R., Robinson, L.S., Pinkner, J.S., Roth, R., Heuser, J., Hammar, M., Normark, S., and Hultgren, S.J. (2002). Role of Escherichia coli curli operons in directing amyloid fiber formation. Science (80-.). *295*, 851–855. https://doi.org/10.1126/science.1067484.

Chiti, F., and Dobson, C.M. (2006). Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem *75*, 333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901.

Chiti, F., and 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. https://doi.org/10.1146/annurev-biochem-061516-045115.

Collins, S.R., Douglass, A., Vale, R.D., and Weissman, J.S. (2004). Mechanism of prion propagation: Amyloid growth occurs by monomer addition. PLoS Biol. *2*. https://doi.org/10.1371/JOURNAL.PBIO.0020321.

Deshmukh, M., Evans, M.L., and Chapman, M.R. (2018). Amyloid by Design: Intrinsic Regulation of Microbial Amyloid Assembly. J. Mol. Biol. *430*, 3631–3641. https://doi.org/10.1016/j.jmb.2018.07.007.

Dueholm, M.S., Nielsen, S.B., Hein, K.L., Nissen, P., Chapman, M., Christiansen, G., Nielsen, P.H., and 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. https://doi.org/10.1021/bi200967c.

Dueholm, M.S., Albertsen, M., Otzen, D., and Nielsen, P.H. (2012). Curli functional amyloid systems are phylogenetically widespread and display large diversity in operon and protein structure. PLoS One *7*, e51274. https://doi.org/10.1371/journal.pone.0051274.

Evans, M.L., Chorell, E., Taylor, J.D., Åden, J., Götheson, A., Li, F., Koch, M., Sefer, L., Matthews, S.J., Wittung-Stafshede, P., et al. (2015). The Bacterial Curli System Possesses a

Potent and Selective Inhibitor of Amyloid Formation. Mol. Cell *57*, 445–455. https://doi.org/10.1016/J.MOLCEL.2014.12.025.

Evans, M.L., Gichana, E., Zhou, Y., and Chapman, M.R. (2018). Bacterial amyloids. In Methods in Molecular Biology, (Humana Press Inc.), pp. 267–288.

Frieden, C. (2007). Protein aggregation processes: In search of the mechanism. Protein Sci. *16*, 2334. https://doi.org/10.1110/PS.073164107.

Gabelica, V., Shvartsburg, A.A., Afonso, C., Barran, P., Benesch, J.L.P., Bleiholder, C., Bowers, M.T., Bilbao, A., Bush, M.F., Campbell, J.L., et al. (2019). Recommendations for reporting ion mobility Mass Spectrometry measurements. Mass Spectrom. Rev. *38*, 291–320. https://doi.org/10.1002/MAS.21585.

Giorgetti, S., Greco, C., Tortora, P., and Aprile, F.A. (2018). Targeting amyloid aggregation: An overview of strategies and mechanisms. Int. J. Mol. Sci. *19*. https://doi.org/10.3390/IJMS19092677.

Goyal, P., Krasteva, P. V, Van Gerven, N., Gubellini, F., Van den Broeck, I., Troupiotis-Tsailaki, A., Jonckheere, W., Pehau-Arnaudet, G., Pinkner, J.S., Chapman, M.R., et al. (2014). Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. Nature *516*, 250–253. https://doi.org/10.1038/nature13768.

Hammer, N.D., Schmidt, J.C., and Chapman, M.R. (2007). The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. Proc Natl Acad Sci U S A *104*, 12494–12499. https://doi.org/10.1073/pnas.0703310104.

Hyung, S.J., Detoma, A.S., Brender, J.R., Lee, S., Vivekanandan, S., Kochi, A., Choi, J.S., Ramamoorthy, A., Ruotolo, B.T., and Lim, M.H. (2013). Insights into antiamyloidogenic properties of the green tea extract (-)-epigallocatechin-3-gallate toward metal-associated amyloid-β species. Proc. Natl. Acad. Sci. U. S. A. *110*, 3743–3748. https://doi.org/10.1073/PNAS.1220326110/SUPPL_FILE/PNAS.201220326SI.PDF.

Jahn, T.R., and Radford, S.E. (2008). Folding versus aggregation: Polypeptide conformations on competing pathways. Arch. Biochem. Biophys. *469*, 100–117. https://doi.org/10.1016/j.abb.2007.05.015.

Jain, S., and Udgaonkar, J.B. (2011). Prion protein aggregation. Curr. Sci. *101*. .

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nat. 2021 5967873 *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Mahul-Mellier, A.L., Burtscher, J., Maharjan, N., Weerens, L., Croisier, M., Kuttler, F., Leleu, M., Knott, G.W., and Lashuel, H.A. (2020). The process of Lewy body formation, rather than simply α-synuclein fibrillization, is one of the major drivers of neurodegeneration. Proc. Natl. Acad. Sci. U. S. A. *117*, 4971–4982. https://doi.org/10.1073/PNAS.1913904117/SUPPL_FILE/PNAS.1913904117.SD04.XLSX.

Marinelli, P., Pallares, I., Navarro, S., and Ventura, S. (2016). Dissecting the contribution of Staphylococcus aureus α-phenol-soluble modulins to biofilm amyloid structure. Sci. Reports 2016 61 *6*, 1–13. https://doi.org/10.1038/srep34552.

May, J.C., and McLean, J.A. (2015). Ion mobility-mass spectrometry: Time-dispersive instrumentation. Anal. Chem. *87*, 1422–1436. https://doi.org/10.1021/AC504720M/ASSET/IMAGES/LARGE/AC-2014-04720M_0007.JPEG.

Nagaraj, M., Najarzadeh, Z., Pansieri, J., Biverstål, H., Musteikyte, G., Smirnovas, V., Matthews, S., Emanuelsson, C., Johansson, J., Buxbaum, J.N., et al. (2022). Chaperones mainly suppress primary nucleation during formation of functional amyloid required for bacterial biofilm formation. Chem. Sci. *13*, 536–553. https://doi.org/10.1039/D1SC05790A.

Nenninger, A.A., Robinson, L.S., and Hultgren, S.J. (2009). Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. Proc. Natl. Acad. Sci. U. S. A. *106*, 900–905. https://doi.org/10.1073/pnas.0812143106.

Nenninger, A.A., Robinson, L.S., Hammer, N.D., Epstein, E.A., Badtke, M.P., Hultgren, S.J., and Chapman, M.R. (2011). CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation. Mol. Microbiol. *81*, 486–499. https://doi.org/10.1111/j.1365-2958.2011.07706.x.

Otzen, D. (2010). Functional amyloid: turning swords into plowshares. Prion *4*, 256–264. https://doi.org/10.4161/PRI.4.4.13676.

Rochet, J.C., and Lansbury, P.T. (2000). Amyloid fibrillogenesis: themes and variations. Curr. Opin. Struct. Biol. *10*, 60–68. https://doi.org/10.1016/S0959-440X(99)00049-4.

Salgado, P.S., Taylor, J.D., Cota, E., and Matthews, S.J. (2011). Extending the usability of the phasing power of diselenide bonds: SeCys SAD phasing of CsgC using a non-auxotrophic strain. Acta Crystallogr. D. Biol. Crystallogr. *67*, 8–13. .

Santambrogio, C., Natalello, A., Brocca, S., Ponzini, E., and Grandori, R. (2019). Conformational Characterization and Classification of Intrinsically Disordered Proteins by Native Mass Spectrometry and Charge-State Distribution Analysis. Proteomics *19*, 1800060. https://doi.org/10.1002/PMIC.201800060.

Schrödinger, LLC (2015). The {PyMOL} Molecular Graphics System, Version~1.8.

Sievers, S.A., Karanicolas, J., Chang, H.W., Zhao, A., Jiang, L., Zirafi, O., Stevens, J.T., Münch, J., Baker, D., and Eisenberg, D. (2011). Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. Nat. 2011 4757354 *475*, 96–100. https://doi.org/10.1038/nature10154.

Sleutel, M., Van den Broeck, I., Van Gerven, N., Feuillie, C., Jonckheere, W., Valotteau, C., Dufrêne, Y.F., and Remaut, H. (2017). Nucleation and growth of a bacterial functional amyloid at single-fiber resolution. Nat. Chem. Biol. *13*, 902–908. https://doi.org/10.1038/nchembio.2413.

Soper, M.T., DeToma, A.S., Hyung, S.J., Lim, M.H., and Ruotolo, B.T. (2013). Amyloid-beta-neuropeptide interactions assessed by ion mobility-mass spectrometry. Phys Chem Chem Phys *15*, 8952–8961. https://doi.org/10.1039/c3cp50721a.

Tamara, S., Den Boer, M.A., and Heck, A.J.R. (2021). High-Resolution Native Mass Spectrometry. Chem. Rev. https://doi.org/10.1021/ACS.CHEMREV.1C00212/ASSET/IMAGES/LARGE/CR1C00212_0030.JPEG.

Taylor, J.D., Zhou, Y., Salgado, P.S., Patwardhan, A., McGuffie, M., Pape, T., Grabe, G., Ashman, E., Constable, S.C., Simpson, P.J., et al. (2011). Atomic Resolution Insights into Curli Fiber Biogenesis. Structure *19*, 1307–1316. https://doi.org/10.1016/j.str.2011.05.015.

Taylor, J.D., Hawthorne, W.J., Lo, J., Dear, A., Jain, N., Meisl, G., Andreasen, M., Fletcher, C., Koch, M., Darvill, N., et al. (2016). Electrostatically-guided inhibition of Curli amyloid nucleation by the CsgC-like family of chaperones. Sci. Rep. *6*, 24656. https://doi.org/10.1038/srep24656.

Wang, X., Smith, D.R., Jones, J.W., and Chapman, M.R. (2007). *In vitro* polymerization of a functional Escherichia coli amyloid protein. J Biol Chem *282*, 3713–3719. https://doi.org/10.1074/jbc.M609228200.

Zhou, Y., Smith, D., Leong, B.J., Brannstrom, K., Almqvist, F., and Chapman, M.R. (2012). Promiscuous cross-seeding between bacterial amyloids promotes interspecies biofilms. J Biol Chem *287*, 35092–35103. https://doi.org/10.1074/jbc.M112.383737.

Chapter 4 Discussion and Future Directions[4]

4.1 Introduction

Amyloid fiber formation is known to be a spontaneous and energetically favorable process. Throughout the course of this dissertation, I sought to better understand how amyloid formation could be controlled. I explored an *in vitro* method of controlling amyloid formation using disulfide engineering. I also wanted to investigate how functional amyloid formation was controlled within the context of curli. To that end, I began several projects aimed at better understanding CsgC, a chaperone-like protein in the curli operon that acts as a amyloid formation inhibitor (Evans et al., 2015). CsgC inhibits amyloid formation of several proteins including CsgA homologs, FapC, and alpha-synuclein (Evans et al., 2015). Wildtype CsgC displays varying inhibition efficiency with different client protein that don't share the same primary sequence. Single amino acid variants of CsgC also have varying efficiency when the client protein is kept constant (**Chapter 3**). Therefore, there must be sequence determinants, both in CsgC and its clients, that guide CsgC amyloid formation inhibition. I used two different methods of protein engineering to modify CsgC so that I could learn more about which CsgC residues are significant to its activity. One method uses multiple sequence analysis to make rational changes to individual residues. Another technique employs directed evolution to make random changes and test a broad range of CsgC variants against one another at the same time.

[4] Additional authors contributed work to this chapter. Chia-Yu Kang contributed to Figure 4-5 and 4-7. Jennie Hibma contributed writing and data for Figure 4-6 and 4-8. Anthony Balistreri contributed to the concept and experimental design of all figures.

Here we discuss the future of these projects and what can be gleamed about CsgC using their results.

Lastly, I wanted to apply what was learned about CsgC to investigate a practical use of anti-amyloid activity. We are interested to see how CsgC could disturb processes that depend on amyloid formation in the mammalian gut. Curli fibers play an important role in survival of bacteria in harsh environments (Chapman et al., 2002) such as the gut and in the cross talk between bacteria and a host organism (Hufnagel et al., 2013). I planned to create CsgC secreting bacterial strains to observe the effect of CsgC secretion on the microbiome.

4.2 Protein engineering CsgC: Rational Design and Site Directed Mutagenesis

Rational justification was necessary to decide which residues of CsgC to mutagenize. We turned to a multiple sequence analysis to get clues from natural proteins to see which residues are important for CsgC activity based on evolutionary conservation. Dueholm and coworkers published a list of bacterial species that contain both a CsgA and a CsgC/H homologs (Dueholm et al., 2012). We created a curated list of sequences of both CsgA and CsgC/H proteins from the same bacterial strain using Dueholm's sequences as a guide. We performed a multiple sequence alignment of CsgC/H proteins using MUSCLE (Edgar, 2004) and constructed a phylogenetic tree using the PhyML 3.0 algorithm (Guindon et al., 2010) (**Figure 4-1**). The resulting tree shows a tight relationship between CsgC proteins, with CsgH proteins occupying their own branch of the tree (**Figure 4-1**). Three CsgC homologs that have been previously tested *in vitro* by our lab were highlighted with *C. youngae* CsgC being the most closely related to E. coli CsgC (**Figure 4-1**). To take a closer look at the sequences we created a graphical sequence alignment showing representative CsgC sequences from each branch of the phylogenetic tree (**Figure 4-2**). The CsgC sequence alignment allows for a more direct comparison of residues

between CsgC proteins and highly conserved residues such as the CxC motif, Q53, and L72 that can be easily appreciated (**Figure 4-2**).

Coevolution analyses can be used to investigate evolutionarily conserved residues in a protein-protein interaction (Lovell and Robertson, 2010). Since CsgC has only one known binding partner in the cell, the residues of CsgC that are significant for its activity must remain steady or change to match any mutations that occur in CsgA over evolutionary time. We used MISTIC2 to perform coevolution analysis and provide a measurement called "cumulative mutual information" (cMI) that estimates the amount of mutual coevolution between two residues within a protein (Colell et al., 2018). The CsgA and CsgC sequences were artificially concatenated with a linker to see how residues in both proteins coevolve with one another. The cMI measurements for each residue were visualized using a unique circos graph produced by MISTIC (**Figure 4-3**). The circus graph is a useful representation of cMI and how connected each residue is within the sample of sequences provided (**Figure 4-3**).

Using the sequence analysis and coevolution information we chose residues of interest on CsgC for mutagenesis based on the following criteria: (i) surface exposure, (ii) high conservation, (iii) high cMI, (iv) mutations found in the CY branch, and (v) glutamine residues. Glutamine residues were considered significant because glutamine is over-represented in CsgC sequences compared to other prokaryotic genes. We decided to make the following variations to highly scored residues:

- X > A ; changing any residue to an alanine,
- E, D > R ; changing any negatively charged residue to a positively charged residue,
- K, R > E ; changing any positively charged residue to a negatively charged residue, and
- Changing any aliphatic residue to R, or E.

We chose the first 30 variants of interest and purchased mutagenized plasmids to test the proteins in the *in vivo* screening assay mentioned in **Chapter 3**.

4.2.1 Future Directions

A directed mutational study is a useful method to learn about a protein activity. However, any method that is used requires a good test to provide useful answers. The *in vivo* screening assay that we set up to answer questions about CsgC is lacking in sensitivity. There is a very clear signal difference between cells that contain a CsgC variant and an empty vector control with no CsgC present. Unfortunately, there is currently a small difference in signal between CsgC variants. We were able to identify the potentially significant residues E48 and R84 using the screening assay; the results were discussed in **Chapter 3**. We also identified L72 as a major contributor to the structural stability of CsgC. Any charged residue modifications to L72 caused severe structural instability and a loss of *in vivo* and *in vitro* activity. A modification to protein expression during the assay could increase the sensitivity. The screening assay uses pET expression vectors for both CsgC and CsgA and therefore both proteins are induced by IPTG added to the growth media. It is reasonable to assume that both proteins are being expressed to a similar intracellular concentration at the same time. During *in vitro* studies of CsgC, small differences in CsgC concentration can have large effects on CsgA amyloid formation inhibition. Therefore, a system more tunable control over CsgA and CsgC expression levels could be a useful modification to the screening assay to increase the sensitivity and understand the more subtle differences between CsgC variants.

4.3 Protein engineering CsgC: Directed Evolution

Directed evolution is a protein engineering technique that utilizes random mutagenesis and artificial selection to speed up the evolution of a protein of interest (Farinas et al., 2001). Random mutagenesis through error prone PCR (epPCR) can provide a pool of mutant genes that will produce protein variants. Researchers then choose an appropriate selectable phenotype that can be used to separate variants based on a desired functionality. After selection, DNA sequencing is performed to assess the most successful protein variant. This variant is used as a starting point for the next round of mutagenesis and selection (Packer and Liu, 2015). This directed evolution cycle can be performed many times to continue to produce the optimal protein for the selection mechanism **(Figure 4-4)**. For CsgC, there is an *in vivo* phenotype associated with a rescue from amyloid cytotoxicity that can be used. The CsgC cytotoxicity rescue phenotype is also the basis of the *in vivo* screening assay discussed in **Chapter 3**. When *E. coli* cells are expressing intracellular CsgA they display a decrease in growth that can be seen by tracking OD600 or CFU over a growth period (**Figure 4-5**). When CsgC is coexpressed intracellularly cells do not show the same growth defects (**Figure 4-5**). Therefore, the presence of an effective CsgC protein yields a rescue from intracellular CsgA amyloid associated toxicity. A simple competitive growth assay could be used as a selection technique for CsgC activity. A pool of CsgC variant producing strains are placed in the same liquid culture and grown under conditions that induce the cytotoxic and protective effect. The CsgC variants that allow the cells to outcompete and thrive in the culture represent the most successful CsgC proteins at preventing intracellular CsgA amyloid fibers from forming. Sequencing the survivors should reveal insights into what residues are important for CsgC activity.

Worsdorfer and coworkers published a directed evolution technique that we used as a model to base our CsgC evolution experiments (Worsdorfer et al., 2011). After purifying a sample of our CsgC expression plasmid, we used the GeneMorph II random mutagenesis kit to make a library or mutants with a low mutation rate. The mutant gene library was cloned back into the pET28a expression plasmid and transformed *E. coli* BL21 cells. The surviving transformants were transformed again with the CsgA expression plasmid (**Figure 4-6**). After surviving both transformations, the cells now contain the CsgA expression plasmid and one of the mutant CsgC expression plasmids (**Figure 4-6**). The resulting pool of cells were diluted in fresh selective media and the competitive growth assay was initiated when IPTG induced expression of both proteins (**Figure 4-6**).

The first round of mutagenesis produced the mutant library CsgC*. We chose to plate some of the transformants after the CsgC* plasmid transformation to confirm a low mutagenesis rate. Since we don't know how much variation from the wildtype CsgC will be required to produce an optimal variant, we elected for slow changes to occur. As we had designed, 10 sample transformants had their plasmids purified and sequenced revealing a collection of simple protein variants and the wildtype protein (**Figure 4-7A**).. During the first competitive growth assay the CsgC* culture grew almost identically to the CsgC WT culture (**Figure 4-7B**). This was to be expected since Worsdorfer and coworkers only reported significant differences between their mutant and wildtype proteins after the 5^{th} round of selection (Worsdorfer et al., 2011). Moving forward, a sample of the CsgC* culture was taken after 5 hours of competitive growth. The sample was miniprepped and the plasmids were purified. A second round of mutagenesis was performed to yield the $CsgC^{2}$* library. Interestingly, $CsgC^{2}$* library showed different growth results from CsgC*, signaling that just two rounds through the mutagenesis and

selection process had already provided measurable improvements in growth (**Figure 4-8**). There must be a variant within the CsgC^{2}* library that allows the cells to survive better than the CsgC WT protein.

4.3.1 Future Directions

Two rounds of directed evolution were completed, and promising results already occurred. This would suggest that directed evolution of CsgC using this selection method could yield valuable results. One can assume that CsgC is already at its maximum efficiency acting as an inhibitor of CsgA amyloid formation, especially given the potency that is already observed during *in vitro* CsgC inhibition assays. The results from **Figure 4-8** suggest that even more optimization is possible. Unfortunately, the error prone PCR step which yields the mutagenized library of CsgC genes became increasingly more difficult after CsgC^{2}*. Once the mutagenesis and cloning steps are optimized, that should allow for more successful rounds of directed evolution to take place.

The CsgC-CsgA directed evolution project was always intended to be a preliminary experiment to lay a foundation for the greater goal of tuning CsgC anti-amyloid activity towards alpha-synuclein. Alpha-synuclein is a human pathogenic amyloid that shares a short sequence motif with CsgA (Evans et al., 2015). CsgC is a substoichiometric inhibitor of alpha-synuclein aggregation, however, it is not nearly as effective an inhibitor when compared to CsgA (Evans et al., 2015). Directed evolution could be used to increase CsgC inhibition efficiency with preventing alpha-synuclein amyloid formation. Now that we have evidence that a competitive growth phenotype can be used to select for more active CsgC variants, the next step to confirm that alpha-synuclein can create intracellular cytotoxic amyloid fibers after expression within *E.*

coli. Alpha-synuclein fibers are associated with cell death within the context of human neurons, but cytotoxicity needs to be confirmed in the recombinant system we use for our experiments.

4.4 CsgC Secreting Commensal Strains

Curli is a functional amyloid that is secreted by many species of gut-dwelling bacteria. Within the enteric environment curli could be playing other roles in addition to its part in biofilm formation. Curli has been proven to be immunogenic and can illicit an immune response in mouse models (Hufnagel et al., 2013). Recently our lab showed that curli can act as a promoter of alpha-synuclein aggregation within the gut that can lead to an acceleration of Parkinson disease symptoms in a mouse model (Sampson et al., 2020). Therefore, the inhibition of curli formation within the gut could be an interesting avenue of investigation. CsgC is a semi-selective amyloid formation inhibitor that is extremely potent and selective for bacterial amyloids to a lesser extent alpha-synuclein (Evans et al., 2015). However, CsgC activity has always been sequestered within the periplasm of certain bacterial species. We are interested in exploring the effect of secreting CsgC into the enteric environment. CsgC secretion could have a deleterious effect on the microbiome and thus negatively affect the host organism. Alternatively, CsgC could play a protective role against pathogenic colonization of the gut or CsgC could knock out problematic interaction between curli and alpha-synuclein. Along these lines we decided to create three strains of commensal bacteria that secrete CsgC to test our hypotheses.

Creating an *E. coli* strain that secretes CsgC was a logical first choice. *E. coli* is widely studied enteric bacteria species and the strain *E. coli* Nissle 1917 is often used as a biological platform in synthetic biology for many clinical applications (Ou et al., 2016). CsgC is a native protein to *E. coli* and CsgC is already secreted naturally into the periplasm. We decided to take advantage of the curli export proteins, referred to as the type VIII secretion system (T8SS), to get

CsgC transferred out of the periplasm and to the outside of the cell. Thankfully Sivanathan and Hochschild published a method for researchers to hijack the T8SS, clone heterologous proteins into an expression vector, and secrete the proteins to the outside of the cell (Sivanathan and Hochschild, 2013) (**Figure 4-9**). Using the MC4100 *ΔcsgBAC* strains and plasmids provided by the Hochschild lab, I successfully cloned CsgC into the expression plasmid, yielding an *E. coli* CsgC secreting strain called ALB3. The CsgC that is secreted includes an N-terminal fusion to the N22 CsgA secretion signal peptide that signals the cell to secrete the appended protein through the CsgG pore.

I confirmed that ALB3 secretes CsgC into the supernatant and I have begun testing ALB3 in biofilm inhibition assays. After growing ALB3 in expression and secretion conditions, I performed a TCA precipitation of the supernatant and used a Western blot to analyze for the presence of CsgC (**Figure 4-10**). There is a visible band associated with CsgC within the supernatant after just 3 hours of expression and secretion (**Figure 4-10**). After confirming CsgC secretion, we then asked how effective the secretion strain was at hindering biofilm formation. We performed pellicle assays that included cocultures of the ALB3 and a uropathogenic *E. coli* strain called UTI89, a model organism for studying the *E. coli* biofilm. An increase in the amount of the CsgC secreting strain did show a qualitative effect on UTI89 pellicle formation, decreasing the amount of wrinkled biofilm formed (**Figure 4-11**). However, quantitative methods for measuring biofilm formation were less conclusive.

L. lactis was the second CsgC secreting strain that we decided to create. Lactic acid bacteria are known commensal species and *L. lactis* is a popular species used in synthetic biology to express and secrete heterologous proteins into the gut lumen (Cook et al., 2018; Mays and Nair, 2018). Benbouziane and coworkers published a paper that uses *L. lactis* MG1363 to

secrete proteins in a stress-inducible manner (Benbouziane et al., 2013). Their expression plasmid uses the GroESL promoter to drive expression, a well conserved bacterial transcription regulator most commonly associated with heat shock proteins (Hartke et al., 1997). Thankfully, expression from the GroESL promoter is unrepressed in temperatures above 30 °C, allowing for constitutive protein expression when engineered strains are harbored within the mammalian gut. The plasmid also uses a putative secretory peptide called USP45 that was selected for its ability to efficiently secrete heterologous proteins (van Asseldonk et al., 1990). After receiving the strains and plasmids from the Bermúdez-Humarán lab, all that is required to create the strain is to clone CsgC into the plasmid and begin testing bacterial supernatant for the presence of CsgC (**Figure 4-12**).

4.4.1 Future Directions

There are practical next steps for each CsgC secreting strain that we planned to create. The *E. coli* strain requires the use of arabinose as the chemical inducer for CsgC expression. Although P_{BAD} promoter are commonly used in molecular biology, there was an unforeseen issue when we tested the biofilm inhibition activity of ALB3. When UTI89 is presented with small concentrations of arabinose in its growth media it must use the arabinose as a carbon source and no longer reliably makes a robust biofilm. Therefore, necessary control experiments that contain arabinose will always yield a result that looks as if biofilm formation was inhibited. To solve this problem the experiment would either have to use a different biofilm model organism or swap UTI89 with a *ΔaraC* mutant that is incapable of consuming arabinose. In addition, *in vitro* ThT assays using bacterial supernatant as the source of an amyloid formation inhibitor could be performed using CsgA, alpha-synuclein, or a combination of both proteins.

Regarding the *L. lactis* strain, CsgC needs to be cloned into the plasmid in order for testing to begin. Once the cloning is finished, we can use the same techniques that we are using to test ABL3 to see how effective the *L. lactis* strain is at inhibiting biofilm or amyloid formation.

Lastly, there is a third bacteria species that we planned to use as a CsgC secretion platform. *Bacteriodes thetaiotaomicron* is a third useful probiotic species (Sonnenburg et al., 2005) whose genus has recently been shown to be enriched in the guts of Alzheimer Disease patients (Vogt et al., 2017). A more elegant secretion mechanism was planned since *B. theta* does not readily secrete proteins in the same way *E. coli* and *L. lactis* do. Instead, we planned to take over the mechanism, though which *B. theta* partially degrades environmental starches. When complex carbohydrates like starches are the only carbon source available, *B. theta* releases outer membrane vesicles (OMVs) that have many degratory enzymes, including an essential enzyme called SusG (Valguarnera et al., 2018). By creating a mutant fusion protein wherein the endogenous *susG* gene is replaced with *E. coli csgC* we should be able to create a strain of *B. theta* that can be induced to export CsgC containing OMVs when presented with starch molecules. Regardless of the state of the construction of the three CsgC secreting strains, we continue to believe this line of experimentation has the potential to yield interesting results, especially given the potency of CsgC during *in vitro* amyloid inhibition assays.

4.5 Future Directions for $CsgA_{CC}$ and Disulfide Engineering

$CsgA_{CC}$ should continue to provide useful insights into the study of functional amyloids. The initial goal of modifying CsgA was to make changes to the purification methodology. Many experiments aimed at asking biochemical questions about CsgA are made very difficult given that CsgA is in a constant state of aggregation. The ability to have a stable solution containing CsgA where every individual protein is in the same state should provide the basis for many new

experiments. For example, any biophysical method requiring a certain concentration of monomeric protein is now possible where before the solution would be constantly losing monomers to oligomer formation. Most excitingly, $CsgA_{CC}$ could help us answer questions about the earliest stages of amyloid formation. Upon the reoxidation of reduced $CsgA_{CC}$, the proteins seem to become stuck in the act of aggregation. We can possibly trap CsgA in its various stages down the line of amyloid formation. This amount of control over polymerization for a fast-aggregating protein like CsgA is unprecedented.

There are engineers and scientists that are interested in building bio-inspired and protein-based materials using functional amyloids as the core structural element. In the two years since our publication *Tuning Disulfide Engineering*, researchers have: detected viruses in water sources using functionalized biofilms (Pu et al., 2020) and created new living bio-adhesive (Li et al., 2022). These are two recent examples, but functionalized curli fibers have been utilized in many different and interesting ways. With $CsgA_{CC}$ we have presented the community with a simple method of control CsgA polymerization in a way that was never possible before. I'm excited to see the various ways $CsgA_{CC}$ or the concept of disulfide engineering can be used to move the field forward.

$CsgA_{CC}$ represents the first use of disulfide engineering to control amyloid formation by stabilizing the unfolded state of the protein. Before 2020, all published examples of disulfide engineering had researchers stabilizing a known fold state of their protein of interest, often to allow the protein to function in a wide range of environments. Though $CsgA_{CC}$ is just one example, there are no apparent reasons why using disulfide engineering as a strategy to control protein folding can't be used for other purposes. To remain in the realm of amyloid proteins, many functional amyloids start as intrinsically disordered proteins that eventually adopt an

amyloidogenic fold. Double cysteine variants CsgB, FapC, TasA, and even alpha-synuclein can all be made to control their aggregation. Whatever new information we can learn about CsgA by using a double cysteine variant, the same types of experiments can similarly be done with other amyloid proteins.

4.6 Future Directions for Understanding the CsgC-CsgA Interaction

The amyloid formation inhibition activity of CsgC is selective and extremely potency. Indeed, CsgC is one of the most potent inhibitors of amyloid formation that have been tested. CsgC has the exceptional ability to maintain the unfolded state of a CsgA while requiring no chemical energy and not forming a strong interaction. There are many unanswered questions about how this might be possible. When we do fully understand how CsgC can perform this task we could potentially create a variety of amyloid inhibiting proteins. CsgC is an effective inhibitor of alpha-synuclein, a significant human pathogenic amyloid. However, CsgC is about 50-fold less efficient at inhibiting alpha-synuclein than it is at inhibiting CsgA. Though that may sound discouraging, it signals that there is room to increase CsgC efficiency against alpha-synuclein. It should be possible to tune CsgC activity towards being more functional against alpha-synuclein by changing key residues. Eventually CsgC could be used as an effective protein-based therapeutic to prevent alpha-synuclein aggregation which is a major step in the pathogenesis of Parkinson Disease.

4.7 Final Remarks

There is still so much that we do not know about amyloid biology. How do organisms harness the useful properties of amyloid fibers without reaping the deleterious effects? Studying functional amyloids like curli and their chaperones gives us a system to investigate how

aggregation is controlled. Though other *csg* operon proteins are important, CsgC is a unique protein to have developed a way to efficiently inhibit amyloid formation. With continued use of tools of protein engineering, we should be able to learn more about how CsgC works so efficiently. However, more excitingly, we can begin to tune CsgC activity towards specific amyloid proteins. One might expect CsgC WT to display the theoretical limit of inhibition efficiency with CsgA. Some of the experiments detailed in this thesis show that even small changes in CsgC WT can modulate its ability to inhibit CsgA amyloid formation, even increasing efficiency. Therefore, CsgC anti-amyloid activity should be tunable and can be shaped to become more efficient with non-native clients. An especially exciting target is alpha-synuclein, the human pathogenic amyloid tied to Parkinson Disease. This thesis also showed how making a minor modification to CsgA lead to a variant with tunable aggregation. The original goal was to aid in purification of CsgA by increasing the amount of time CsgA remained soluble in solution. While accomplishing the original goal, $CsgA_{CC}$ will also allow us to study amyloid formation in greater detail. We can now investigate snapshots along the aggregation timeline, a feat that is previously impossible with a fast-aggregating amyloid protein like CsgA. The more exciting future is how $CsgA_{CC}$ can be utilized to build complex structures bio-inspired nanomaterials. Additional engineering of $CsgA_{CC}$ will produce variants with tunable aggregation and a wide variety of other activities to create functionalized fibers.

4.8 Methods and Materials

Bacterial Growth

All overnight cultures were grown in LB supplemented with 100 μg/mL ampicillin and/or 50 μg/mL of kanamycin at 37 °C with shaking at 220 rpm. When necessary, LB plates were supplemented with ampicillin 100 μg/mL or kanamycin 50 μg/mL.

Strains and Plasmids

The full list of strains, plasmids, and primers can be found in the tables below. SnapGene (SnapGene® by Dotmatric) and purchased by IDT (https://www.idtdna.com). Mutagenized plasmids were constructed in the MC1061 cloning cell background. Correct mutations were confirmed using Sanger sequencing provided by eurofins (https://www.eurofins.com/genomic-services/our-services/custom-dna-sequencing/). Plasmids were extracted from transformants using Promega PureYield™ Plasmid Miniprep System (Cat No. PRA1223). A standard heat shock protocol was used for all transformations unless otherwise specified.

Sequence Analysis

The sequences of all CsgC/H and CsgA-like proteins were retrieved by inputting accession numbers into the Batch Entrez from NCBI (https://www.ncbi.nlm.nih.gov/sites/batchentrez). In the resulting FASTA files, all non-alphanumeric characters in the gene names were removed. For all species of interest, the cognate CsgA and CsgC/H-like proteins in each genome were manually concatenated by pasting the sequences together, including a PPPPP linker residues. This resulted in synthetic fusion proteins of CsgA:CsgC. Using SeaView, a MUSCLE alignment of all concatenated sequences was performed and exported using the Clustal format. A tree was created from this alignment using the PhyML option using the standard parameters except changing the # rate categories to 12. The Clustal file was uploaded to the MISTIC webserver and analyzed using standard parameters. UTI-89 CsgC was chosen as the reference sequence to standardize residue numbering. After viewing the "MI data" file from the MISTIC results, significant relationships were identified by: Sorting columns by MI, largest to smallest. Identifying relationships with an MI > 6.5. Sorting columns by residue distance, selecting for relationships between residues that are at least 152

residues apart. Using these criteria, only relationships between original CsgA and original CsgC residues were analyzed.

***In vivo* screening assay**

Overnight cultures were created in sterilized LB containing the appropriate selective agent and incubated the liquid cultures at 37°C with shaking for 16 hours. The following day, cell density was normalized by diluting overnight cultures to 1OD using fresh media. The wells of a sterile 96 well plate (Cat No. CLS3614) were filled to a final volume of 200 µL by diluting the normalized cultures to OD = 0.01 and diluting 1:500 Amytracker® 680 (Ebba Biotech AB). the plate covered using a Breathe-Easy® sealing membrane (Cat No. Z380059). The microplate was placed in a preheated Molecular Devices iD3 microplate reader and absorbance (600nm) and fluorescence (Ex:550nm, Em:680mn) was monitored every 15 mins for 16 hours. The plate was incubated at 37°C with shaking between spectroscopic readings. When the liquid cultures grew for 2 hours, the expression of all plasmids was induced by adding ITPG to a final concentration of 0.5 mM diluted from a 1M frozen stock.

Growth curves

Overnight cultures were created in sterilized LB containing the appropriate selective agent and incubated the liquid cultures at 37°C with shaking for 16 hours. The following day, cell density was normalized by diluting overnight cultures to 1OD using fresh media. Small scale growth cultures of 25 mL were created by diluting the normalized cultures to OD = 0.01 in fresh media with added selective antibiotics. When the liquid cultures reach OD600 = 0.2, induce the expression of all plasmids by adding ITPG to a final concentration of 0.5 mM diluted from a 1 M frozen stock. Immediately upon the addition of ITPG, 500 µL aliquots of the cultures were removed and OD600 was measured. Another method for assessing growth is to monitor the

colony forming units (CFU) in each culture over time. To achieve this, a sterile 96-well plate (Cat No. CLS3614) was prepared for each culture to create serial dilutions, with a 1/10 during each additional step. 10 µL of each dilution was dispensed onto a selective plate, resulting in 6 spots is increasing dilution. The dilutions were dispensed in such a manner 5 times to result in n = 5 technical replicates for each time point and dilution. After an overnight growth, CFU's were counted at the lowest dilution possible and multiplied it by the dilution factor to gain a CFU/mL measurement for each time point.

Directed Evolution Library Construction

Following the Genemorph II Random Mutagenesis Kit (Cat No. #200550) protocol, a library of mutagenized CsgC genes was created using an error prone PCR method and instructions for a low mutation frequency rate. 250 ng of the resulting "megaprimers" was moved to a subsequent EZ Clone reaction that included the wildtype expression plasmid as a template and yielded circular plasmids that contained the mutant genes. 1 µl of Dpn I restriction enzyme (10 U/µl) was added directly to the reaction and incubated at 37°C for 2 hours to digest the parental supercoiled dsDNA.

Competitive Growth Assay

Chemically competent E. coli BL21 cells were transformed with the epPCR libraries (approximately 2.5 µg plasmid) using the heat shock method. The library size was determined by plating serial dilutions of these cells onto LB-agar containing kanamycin (50 µg/mL). Representative clones were sequenced to verify library quality. To the remaining cells, LB medium supplemented with kanamycin (50 µg/mL) was added to a final volume of 50 mL. These precultures were grown overnight at 37 °C and then used to inoculate cultures for the preparation of chemically competent cells. These cells that contain the epPCR library underwent a second

heat shock to transform the $CsgA_{\Delta sec}$ expression plasmid. After heat shock, the surviving transformants were diluted to 50 mL in LB medium with added kanamycin (50 μg/mL) and ampicillin (100 μg/mL) for 1 hr min at 37 °C. CsgA and CsgC expression was induced by the addition of IPTG to a final concentration of 0.5 mM and the incubation continued at 37 °C and 250 rpm. Growth was monitored using the OD600 and CFU method described above. After 5 hours of competitive growth, the cells were harvested, and their plasmids purified to be used for another round of epPCR.

TCA precipitation and Western Blot

Bacterial supernatant samples were passed through 0.22 μm syringe filters (Cat No. SLGP033NS) to completely remove all cells. In a microcentrifuge tube, ice-cold 99% trichloroacetic acid (Cat No. SA433-500) was added to 0.9 mL of the filtrate to a final concentration of 10% (v/v). The tube was transferred to a -20 °C freezer for 20 min. The tube was then centrifuged at top speed (12,000 x g to 16,000 x g) in a microcentrifuge for 30 min at 4 °C. The supernatant was discarded, and the pellet was rinsed with 1 mL ice-cold acetone (Cat No. A18P-4). After centrifuge again (12,000 x g to 16,000 x g, 30 min, 4 °C), the supernatant was removed, and the pellet was left to air dry for 15 min on the bench top. Precipitated protein pellets were diluted in 4X SDS loading buffer and run on a 15% SDS PAGE gel as previously described (Evans et al., 2018). The gels were stained with Coomassie blue dye to visualize protein bands or the proteins were transferred to a PVDF membrane for blotting. Western Blot were performed as previously described (Evans et al., 2018). Briefly, blots were probed with a primary antibody against CsgC (1:4,000). Secondary antibodies against rabbit IgG and conjugated with IR Dye 800CW (Cat No.NC9401842) were used to image the blots in a Licor Odyssey FC.

Pellicle Forming Assay

Pellicle forming assays were performed in the same manner as described previously (Evans, 2018). Briefly, liquid overnight cultures of UTI89-WT and ALB3 were created in 5 mL of sterile LB with any required selective antibiotics and grown at 37 ℃ with shaking at 220 rpm. The next day, the wells of a sterile 24-well plate were filled with 2 mL of sterilized YESCA medium and 2 µL of the overnight culture. Cultures were incubated 26 ºC for 2-3 days before a picture was taken to assess biofilm formation.

4.9 Figures

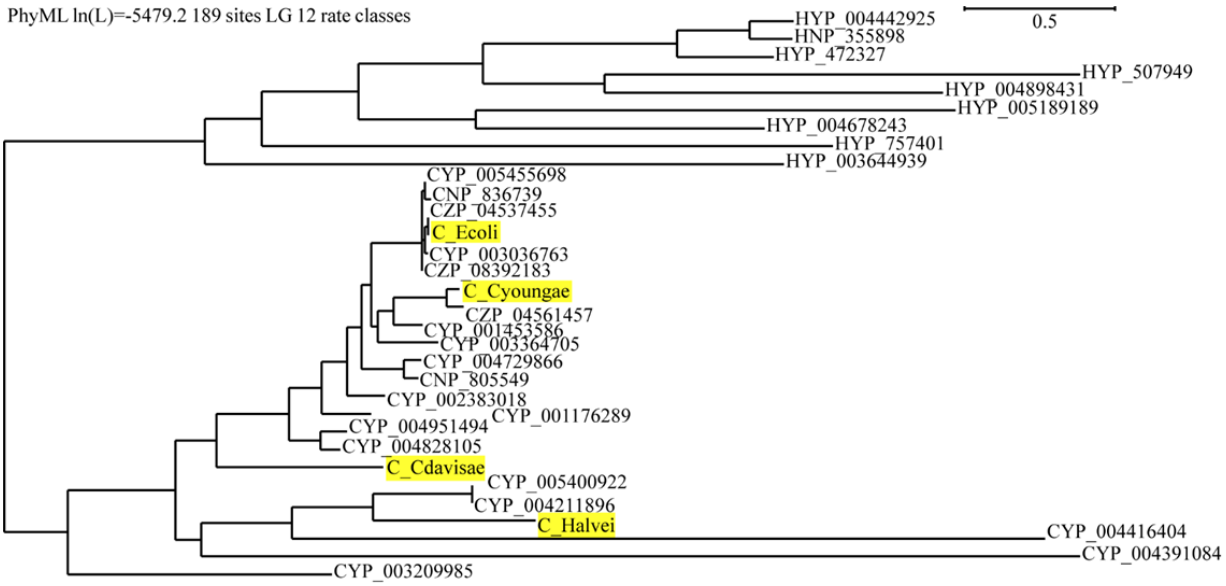

Figure 4-1 Phylogenetic tree showing the similarities between select CsgC homologs. CsgC homologs from the collection published by Dueholm et al. (Dueholm, 2012) were aligned using a MUSCLE alignment tool. A phylogenetic tree was created with the aligned sequences using the PhyML algorithm. All CsgC protein names have been annotated to begin with "C" and all CsgH protein name begin with "H". CsgC homologs for which *in vitro* ThT binding assay data has been gathered are highlighted in yellow.

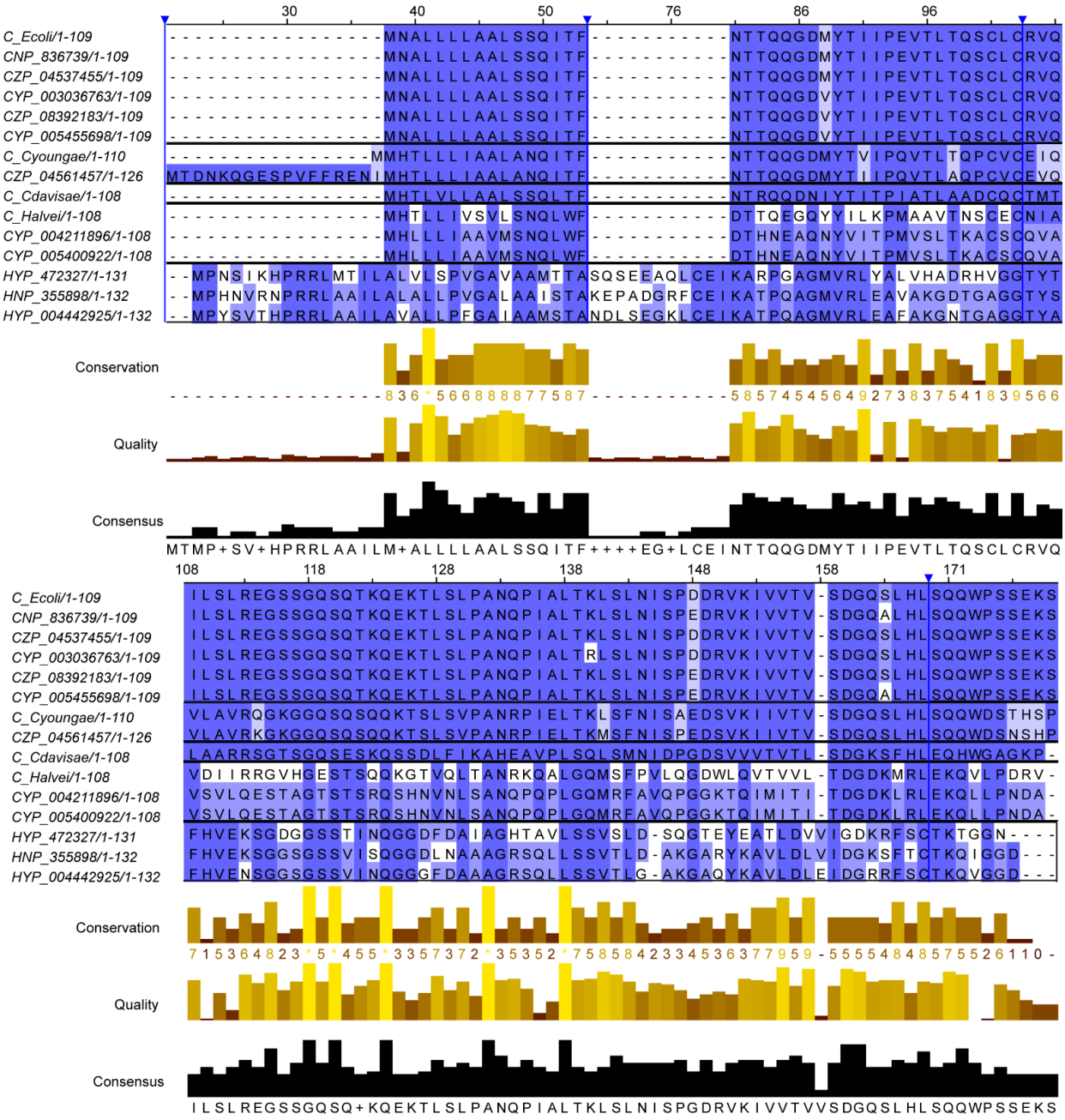

Figure 4-2 Alignment of CsgC homologs

Selected sequences from the phylogenetic tree branches shown in **Figure 4-1** are displayed to indicate the difference between CsgC homologs between branches. Each branch is separated by a thick black border. The background color for each sequence is an indication of the identity of each residue at that location within each branch. Conservation (yellow) is a scored measurement of how well conserved residues are across all CsgC homologs. Quality (yellow) is a likelihood of observing the mutations in a particular residue across all CsgC homologs. Consensus (black) is a consensus sequence across all CsgC homologs with a measurement of the percent identity at each chose residue.

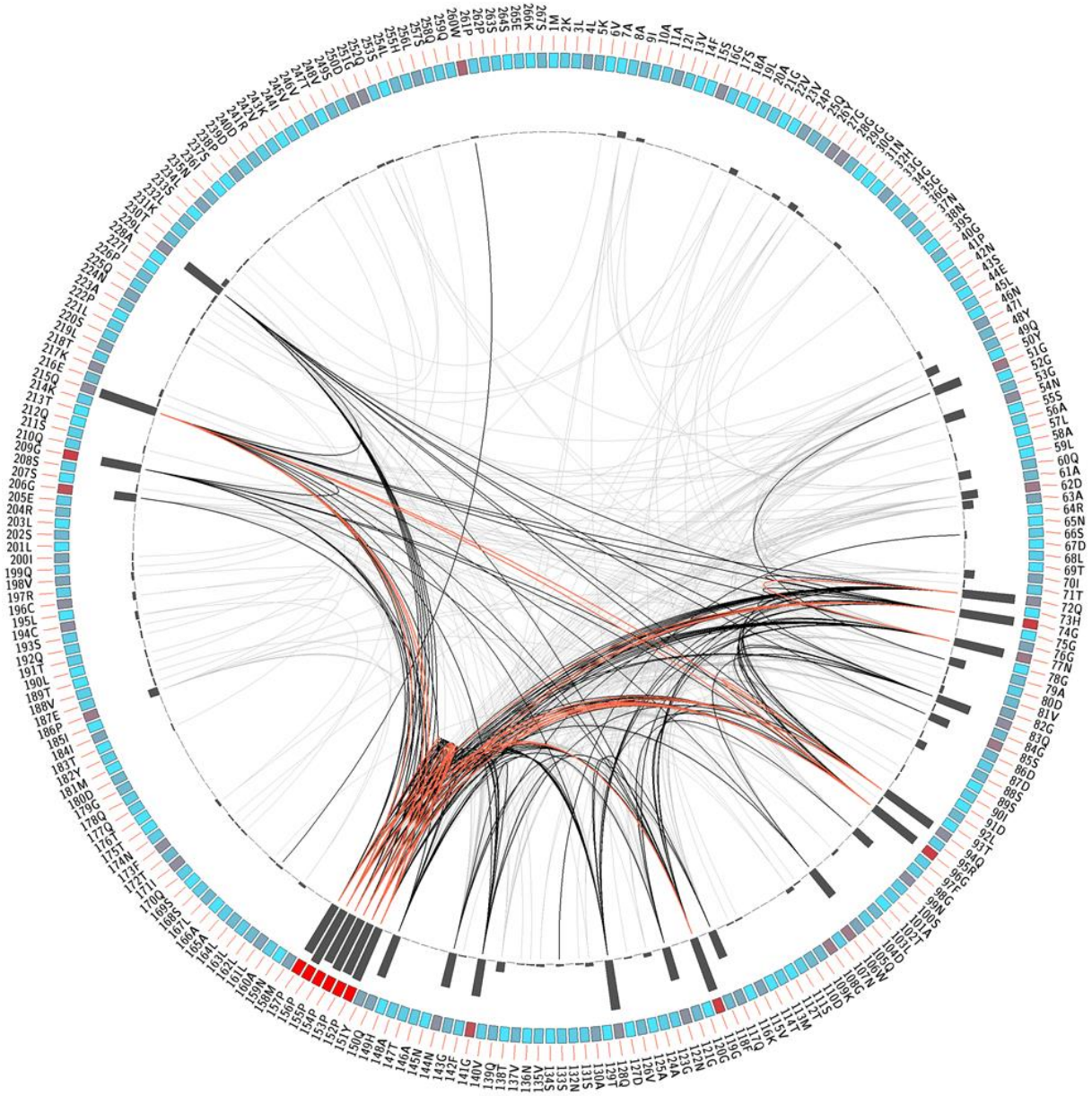

Figure 4-3 Circos graph for all CsgA:CsgC concatenated sequenced used in this analysis. CsgA protein sequences and their cognate CsgC sequences were artificially concatenated to create a synthetic fusion protein. Residue 1 is the first residue of CsgA and residue 158 is the first residue of CsgC (with a 6xPro linker in-between). The bars beneath each residue are representations of the cMI value for each residue. Lines reaching within the circle show evolutionarily conserved relationships between two connected residues (MI value > 6).

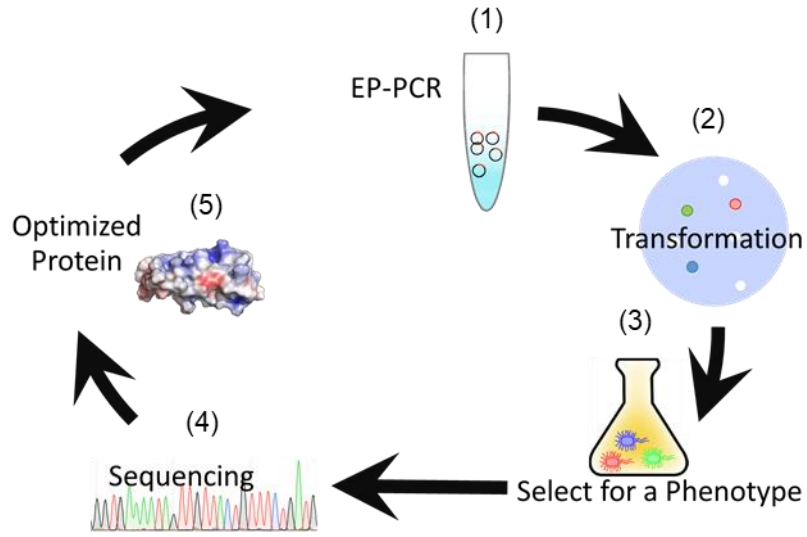

Figure 4-4 Schematic showing the cycles of a directed evolution.
(1) Directed evolution experiments start with the random mutagenesis of a target gene of interest, in this case error prone PCR is used to generate a library of mutant genes. (2) The epPCR library is transformed into a cell to be tested in an *in vivo* assay. (3) A competitive assay is used to produce the mist successful protein variant using artificial selection for a phenotype of choice. (4) Sequencing results reveal which protein variants is the most fit, producing an (5) optimized protein. The optimized protein can be used as the starting point for another round of mutagenesis and selection.

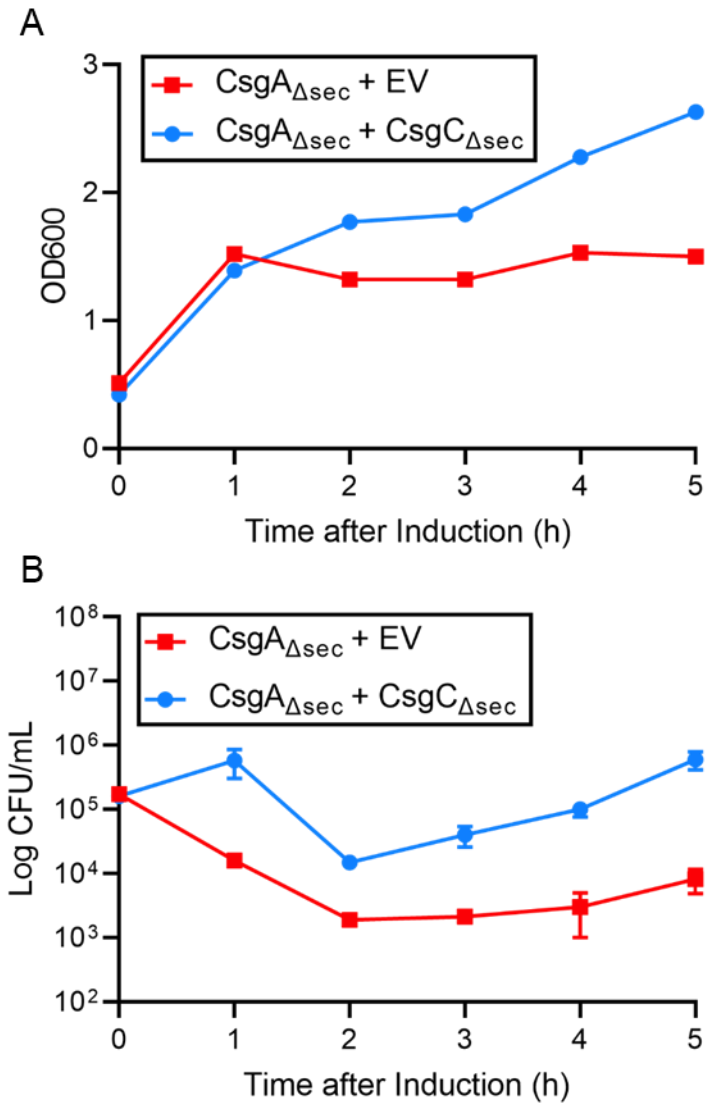

Figure 4-5 Growth curves showing the presence of CsgC rescues from CsgA-associated cytotoxicity.
Two BL21 strains harbored a pair of expression plasmids that expressed both CsgA$_{\Delta sec}$ and CsgC$_{\Delta sec}$ or an empty vector control. After reaching an OD600 = 0.5, the expression of all plasmids was induced with the addition of IPTG to the culture media. Afterward **A)** OD600 and **B)** CFU was measured once per hour.

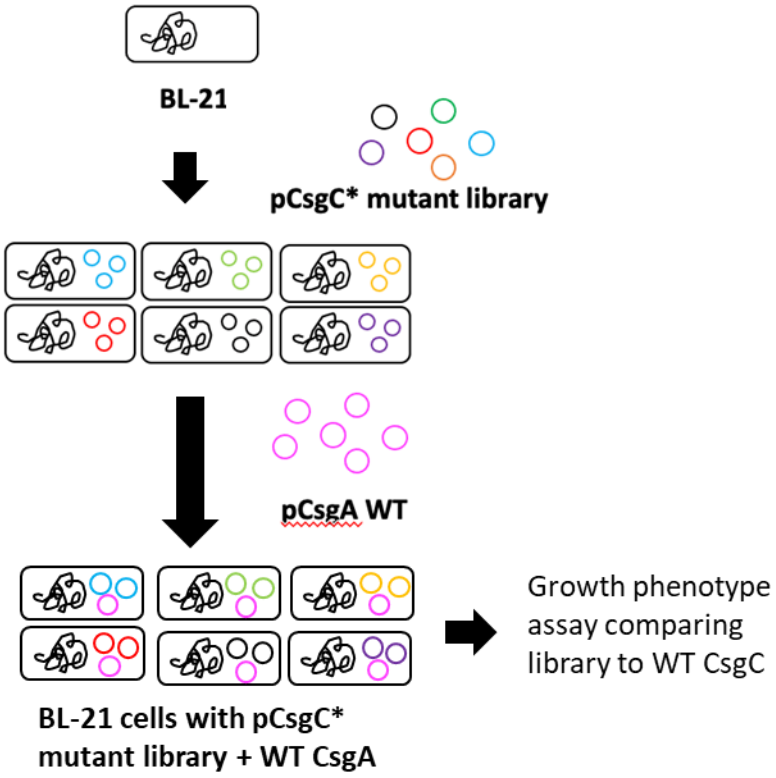

Figure 4-6 Cartoon showing a flow diagram of the steps of the directed evolution selection technique.

First, competent BL21 cells are transformed with the mutant library of $CsgC_{\Delta sec}$ expressing plasmids called "pCsgC". Living transformants are treated with $CaCl_2$ to enable a second transformation. The cells are transformed a second time with the "pCsgA", the $CsgA_{\Delta sec}$ expression plasmid. Living transformants are diluted in media containing two selective antibiotics to maintain both plasmids. They cultures are induced with IPTG to begin the competitive growth assay.

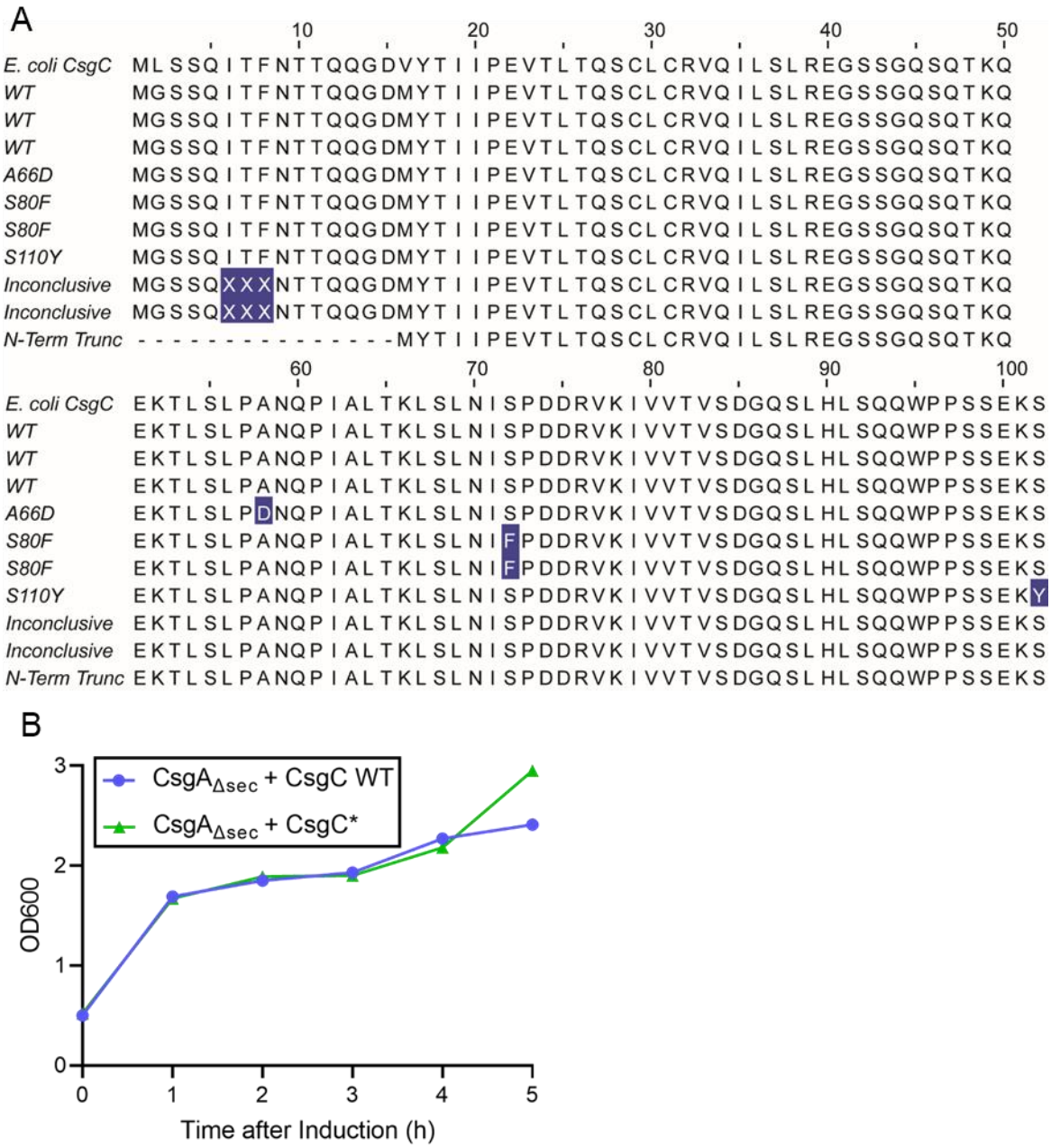

Figure 4-7 Results from the first round of CsgC directed evolution.
A) 10 representative colonies were chosen after BL21 cells were transformed with CsgC*. The sequencing results are shown with the template protein at the top for comparison. There was a mixture of CsgC variants including three WT proteins and a 15 residue N-terminal truncation. **B)** Growth curves tracking OD600 of the CsgC WT expressing strain compared to the whole library of CsgC* expressing strains.

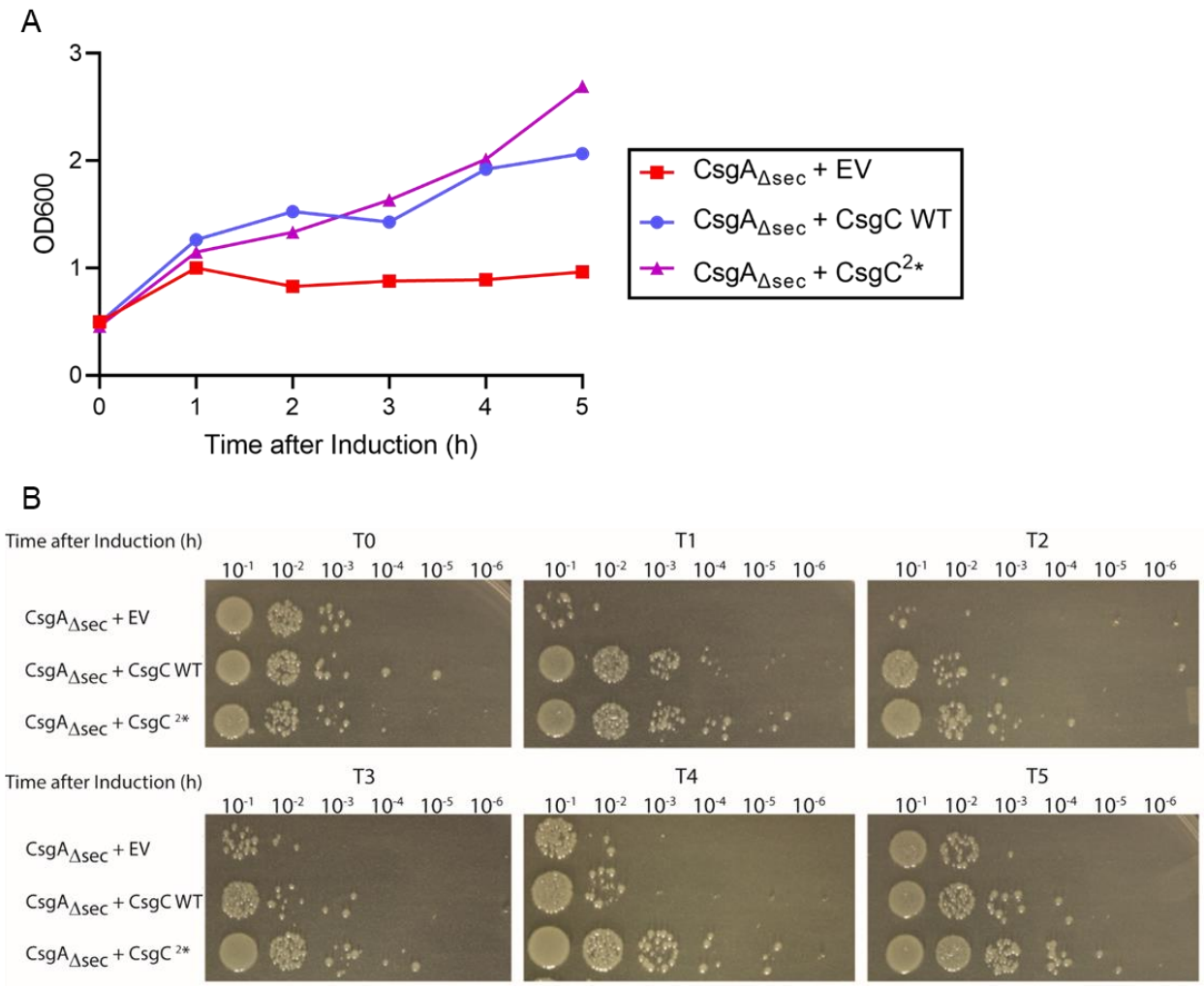

Figure 4-8 Results from the second round of CsgC directed evolution.
A) Growth curves tracking OD600 of the EV control, CsgC WT, and CsgC ^{2*} expressing strains. **B)** The same samples that were used to measure OD600 were serially diluted and plated onto selective LB plates. Pictures show colonies forming across for 5 hours of growth after induction.

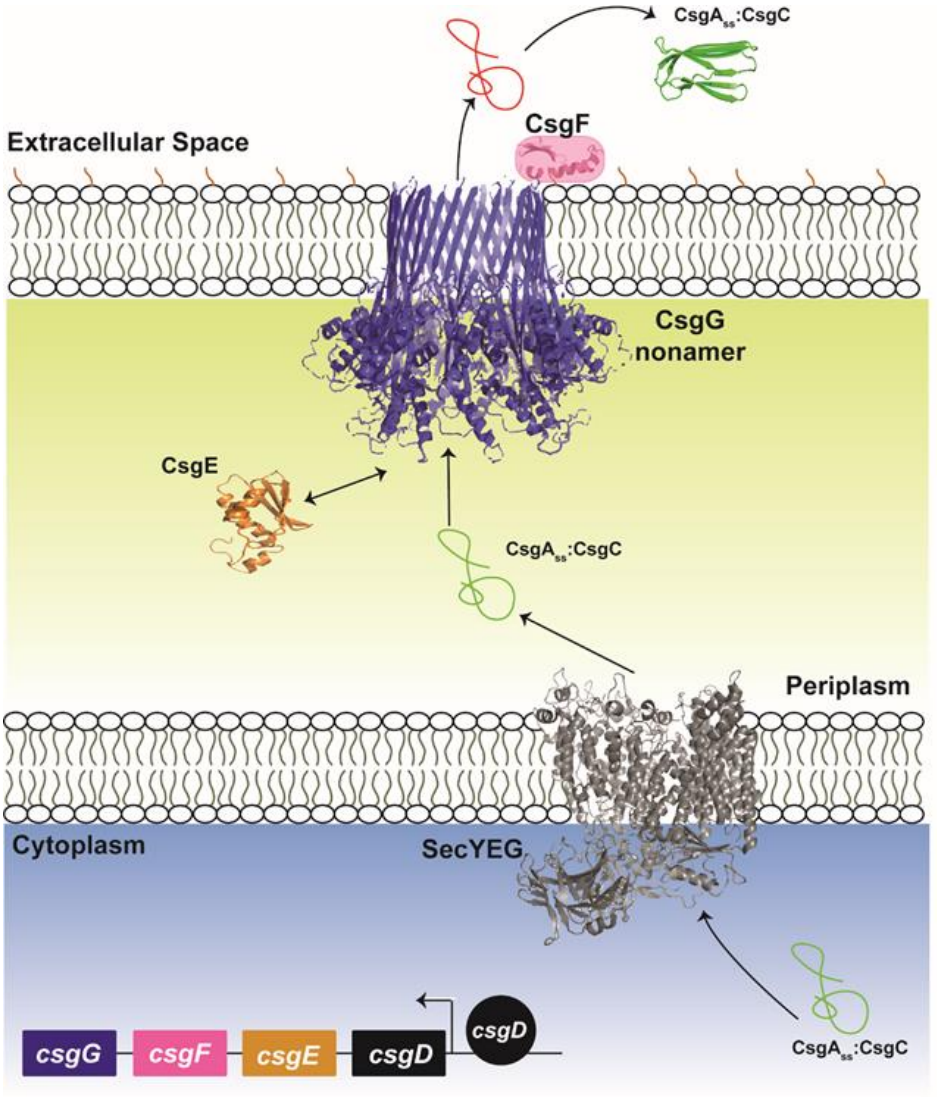

Figure 4-9 Schematic showing the *E. coli* CsgC secretion system.
E. coli MC4100 lacking endogenous curli proteins (*ΔcsgBAC*) contains a plasmid expressing $CsgA_{SS}$:CsgC, a fusion protein featuring an N-terminal fusion of CsgC with the CsgA secretory signal peptides. CsgC is secreted out of the cell through the CsgG pore utilizing the Type VIII secretion system (curli biogenesis machinery).

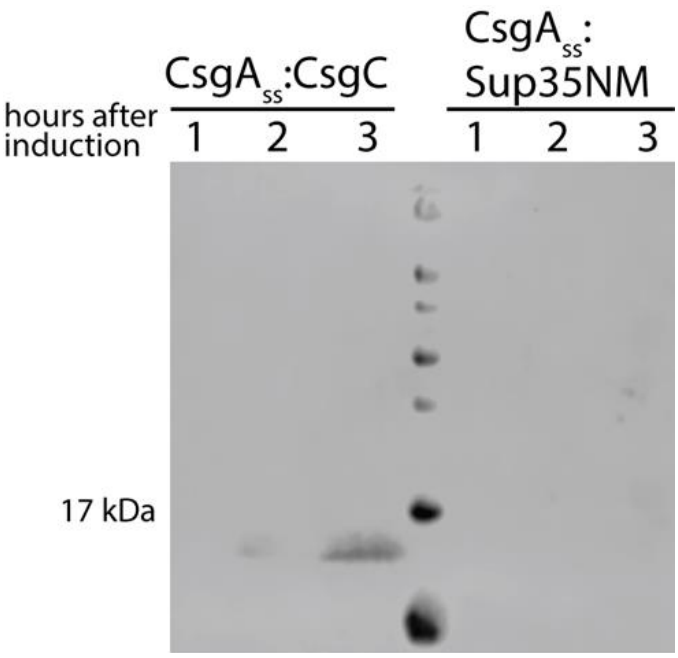

Figure 4-10 ALB3 secreted CsgC into the supernatant after expression.
At three time points, induced culture samples were taken and the bacteria removed. TCA was used to precipitate all proteins in the supernatant. A western blot was performed using a αCsgC antibody. A Sup35NM secreting strain was used as a negative control.

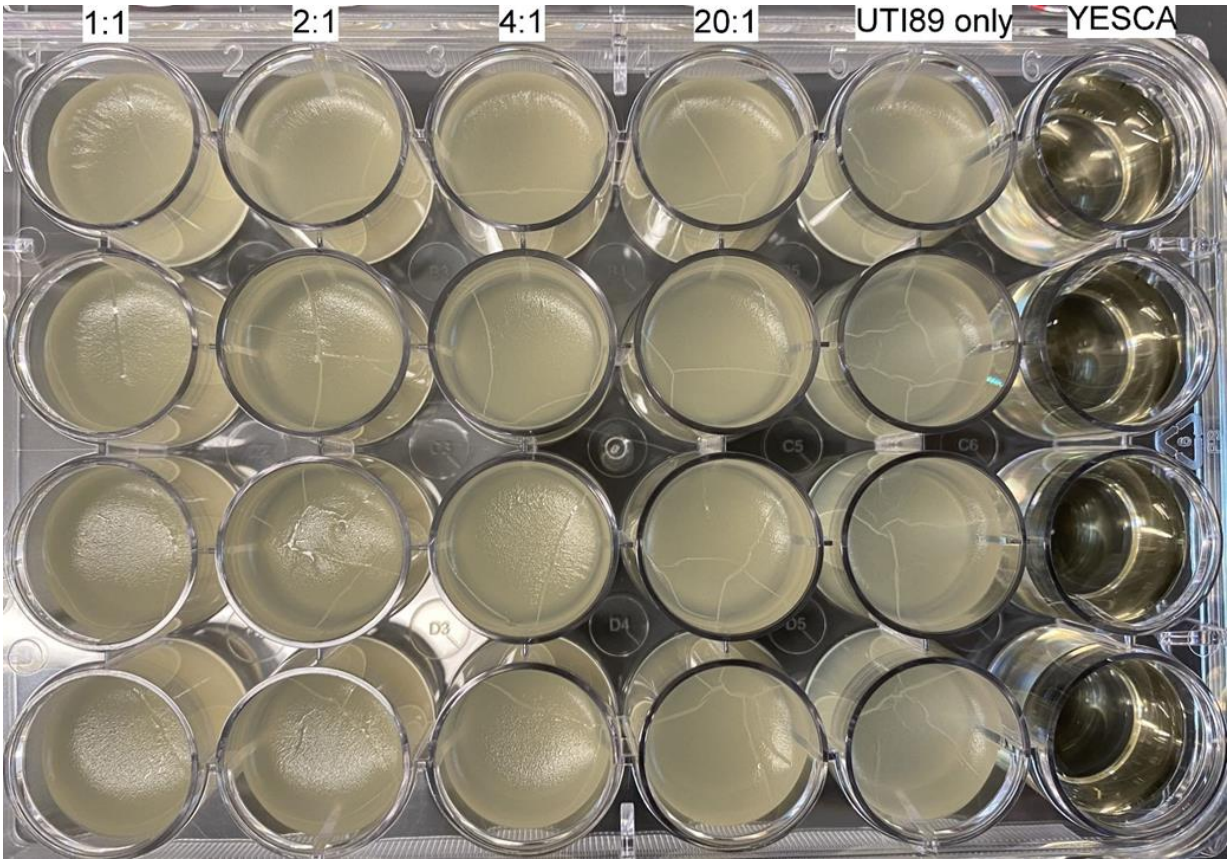

Figure 4-11 An increasing proportion of ALB3 decreases pellicle biofilm formation in a coculture with UTI89.

Figure 4-11. Overnight cultures of ALB3 and the *E. coli* biofilm model strain UTI89 were diluted into fresh YESCA medium and incubated at 26 °C for three days. From right to left: a YESCA only control, UTI89 only, 20:1 UTI89:ALB3, 4:1 UTI89:ALB3, 2:1 UTI89:ALB3, 1:1 UTI89:ALB3. Four replicate wells were plated down each column.

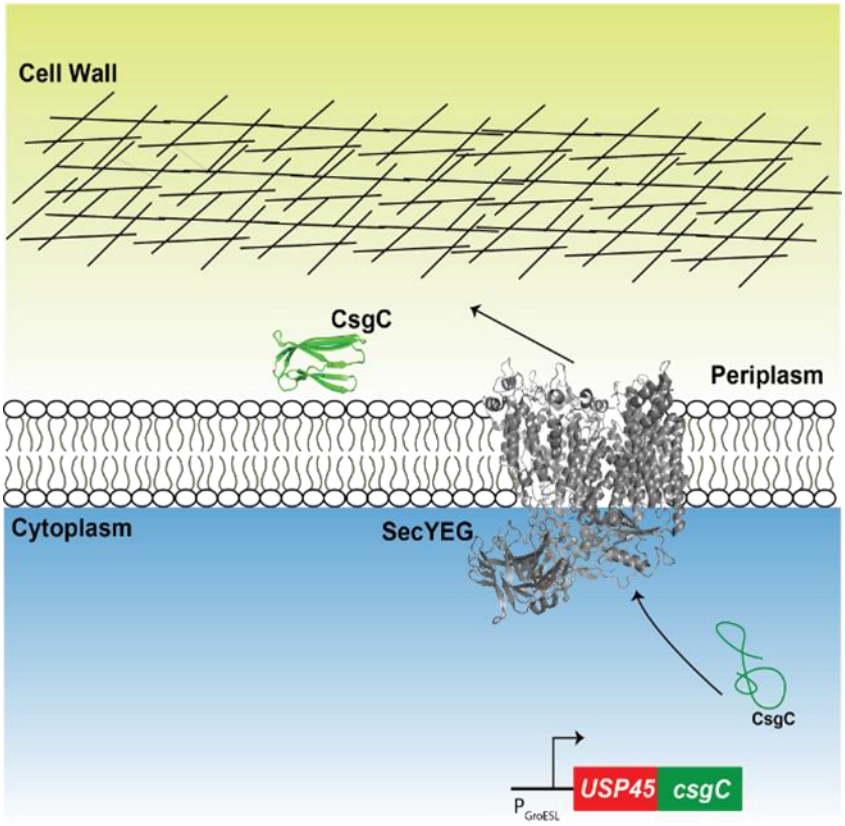

Figure 4-12 Schematic showing the *L. lactis* CsgC secretion system.
L. lactis MG1363 strain contains a plasmid expressing CsgC under control of the GroESL promoter. The plasmid uses the USP45 secretion signal peptide. CsgC is secreted through the endogenous SecYEG pore.

4.10 Tables

Table 4-4-1 Strains used in this study.

Strains	Relevant Genotype	References
BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm]* Δ*hsdS* λ DE3 = λ *sBamHIo* Δ*EcoRI-B int::(lacI::PlacUV5::T7 gene1) i21* Δ*nin5*	New England Biolabs
MC1061	*F– araD139 Δ(ara-leu)7696 galE15 galK16 Δ(lac)X74 rpsL (StrR) hsdR2 (rK– mK+) mcrA mcrB1*	(Casadaban and Cohen, 1980)
NEB3016	MiniF *lacI*^{q}(Cam^{R}) */ fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--*Tet^{S}*)2 [dcm] R(zgb-210::Tn10--*Tet^{S}*) endA1 Δ(mcrC-mrr)114::IS10*	New England Biolabs
ALB30	BL21(DE3) + pET28a EV + pET11d-CsgA -sec 6xHis	This study
CsgC WT	BL21(DE3) + pET28a-CsgC-sec C-term 6xHis, kan^{r} + pET11d-CsgA -sec 6xHis	This study
UTI89	BL21(DE3) + pET28a EV + pET11d-CsgA -sec 6xHis	
ALB3	MC4100 Δc*sgBAC*::(frc)kan(frc), with pVS76 (pAC-PlacUV5-CsgG, cm^{r}) and pVSA1 (pBR322-Pbad-CsgA$_{(1-42)}$-CsgC, cb^{r})	This study
VS39	MC4100 Δc*sgBAC*::(frc)kan(frc) + pVS76 (pAC-PlacUV5-CsgG, cm^{r}) and pVS72 (pBR322-Pbad-CsgA$_{(1-42)}$-Sup35NM 6xHis, cb^{r})	(Sivanathan, 2013)
LB333	*Lactococcus lactis* str. MG1363 + pLB333 (pGK-PgroESL-SP$_{EXP4}$-NucB, cm^{r})	(Benbouziane, 2013)

Table 4-4-2 Plasmids used in this study.

Plasmids	Relevant Characteristics	References
pET11d	IPTG inducible expression vector	New England Biolabs
pET28a	IPTG inducible expression vector	New England Biolabs
pCsgA	pET11d-CsgA -sec 6xHis	(Zhou, 2012)
pCsgC	pET28a-CsgC-sec C-term 6xHis, kan^r	(Evans, 2015)
pVS72	pBR322-Pbad-$CsgA_{(1-42)}$-Sup35NM 6xHis, cb^r	(Sivanathan, 2013
pVSA1	pBR322-Pbad-$CsgA_{(1-42)}$-CsgC, cb^r	This study
pLB333	pGK-PgroESL-SP_{EXP4}-NucB, cm^r	(Benbouziane, 2013)

Table 4-4-3 Primers used in this study.

Primers	Primer Sequence (5'→ 3')	Constructs
CsgC into VS72 Frag_FOR	gcccaaatGCGGCCGCAATGGCACTTTCCAGTCAGATAACCTTTAATACG	ALB3
CsgC into VS72 Frag_REV	gcctgcaggtcgacTctagaTTAAGACTTTTCTGAAGAGGGCGG	ALB3
CsgC into VS72 Vector_FOR	CCTCTTCAGAAAAGTCTTAAtctagAgtcgacctgcaggcat	ALB3
CsgC into VS72 Vector_REV	GTTATCTGACTGGAAAGTGCCATTGCGGCCGCatttgg	ALB3
CsgC-epPCR-F	GTTTAACTTTAAGAAGGAGATATACCATGGGC	CsgC epPCR
CsgC-epPCR-R	CTCAGCTTCCTTTCGGGCTTTGTTAG	CsgC epPCR

4.11 References

van Asseldonk, M., Rutten, G., Oteman, M., Siezen, R.J., de Vos, W.M., and Simons, G. (1990). Cloning of usp45, a gene encoding a secreted protein from Lactococcus lactis subsp. lactis MG1363. Gene *95*, 155–160. https://doi.org/https://doi.org/10.1016/0378-1119(90)90428-T.

Benbouziane, B., Ribelles, P., Aubry, C., Martin, R., Kharrat, P., Riazi, A., Langella, P., and Bermúdez-Humarán, L.G. (2013). Development of a Stress-Inducible Controlled Expression (SICE) system in Lactococcus lactis for the production and delivery of therapeutic molecules at mucosal surfaces. J. Biotechnol. *168*, 120–129. https://doi.org/10.1016/j.jbiotec.2013.04.019.

Chapman, M.R., Robinson, L.S., Pinkner, J.S., Roth, R., Heuser, J., Hammar, M., Normark, S., and Hultgren, S.J. (2002). Role of Escherichia coli curli operons in directing amyloid fiber formation. Science (80-.). *295*, 851–855. https://doi.org/10.1126/science.1067484.

Colell, E.A., Iserte, J.A., Simonetti, F.L., and Marino-Buslje, C. (2018). MISTIC2: comprehensive server to study coevolution in protein families. Nucleic Acids Res. *46*, W323. https://doi.org/10.1093/NAR/GKY419.

Cook, D.P., Gysemans, C., and Mathieu, C. (2018). Lactococcus lactis as a versatile vehicle for tolerogenic immunotherapy. Front. Immunol. *8*, 1961. https://doi.org/10.3389/FIMMU.2017.01961/BIBTEX.

Dueholm, M.S., Albertsen, M., Otzen, D., and Nielsen, P.H. (2012). Curli functional amyloid systems are phylogenetically widespread and display large diversity in operon and protein structure. PLoS One *7*, e51274. https://doi.org/10.1371/journal.pone.0051274.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/NAR/GKH340.

Evans, M.L., Chorell, E., Taylor, J.D., Åden, J., Götheson, A., Li, F., Koch, M., Sefer, L., Matthews, S.J., Wittung-Stafshede, P., et al. (2015). The Bacterial Curli System Possesses a Potent and Selective Inhibitor of Amyloid Formation. Mol. Cell *57*, 445–455. https://doi.org/10.1016/J.MOLCEL.2014.12.025.

Farinas, E.T., Bulter, T., and Arnold, F.H. (2001). Directed enzyme evolution. Curr. Opin. Biotechnol. *12*, 545–551. .

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321. https://doi.org/10.1093/SYSBIO/SYQ010.

Hartke, A., Frère, J., Boutibonnes, P., and Auffray, Y. (1997). Differential Induction of the Chaperonin GroEL and the Co-Chaperonin GroES by Heat, Acid, and UV-Irradiation in Lactococcus lactis subsp. lactis (Springer-Verlag New York Inc).

Hufnagel, D.A., Tükel, Ç., and Chapman, M.R. (2013). Disease to Dirt: The Biology of

Microbial Amyloids. PLoS Pathog. *9*, e1003740. https://doi.org/10.1371/journal.ppat.1003740.

Li, F., Ye, L., Zhang, L., Li, X., Liu, X., Zhu, J., Li, H., Pang, H., Yan, Y., Xu, L., et al. (2022). Design of a genetically programmed barnacle-curli inspired living-cell bioadhesive. Mater. Today Bio *14*, 100256. https://doi.org/10.1016/J.MTBIO.2022.100256.

Lovell, S.C., and Robertson, D.L. (2010). An Integrated View of Molecular Coevolution in Protein–Protein Interactions. Mol. Biol. Evol. *27*, 2567–2575. https://doi.org/10.1093/MOLBEV/MSQ144.

Mays, Z.J., and Nair, N.U. (2018). Synthetic biology in probiotic lactic acid bacteria: At the frontier of living therapeutics. Curr. Opin. Biotechnol. *53*, 224–231. https://doi.org/10.1016/J.COPBIO.2018.01.028.

Ou, B., Yang, Y., Tham, W.L., Chen, L., Guo, J., and Zhu, G. (2016). Genetic engineering of probiotic Escherichia coli Nissle 1917 for clinical application. Appl. Microbiol. Biotechnol. *100*, 8693–8699. https://doi.org/10.1007/S00253-016-7829-5/TABLES/2.

Packer, M.S., and Liu, D.R. (2015). Methods for the directed evolution of proteins. Nat. Rev. Genet. *16*, 379–394. https://doi.org/10.1038/nrg3927.

Pu, J., Liu, Y., Zhang, J., An, B., Li, Y., Wang, X., Din, K., Qin, C., Li, K., Cui, M., et al. (2020). Virus Disinfection from Environmental Water Sources Using Living Engineered Biofilm Materials. Adv. Sci. *7*. https://doi.org/10.1002/ADVS.201903558.

Sampson, T.R., Challis, C., Jain, N., Moiseyenko, A., Ladinsky, M.S., Shastri, G.G., Thron, T., Needham, B.D., Horvath, I., Debelius, J.W., et al. (2020). A gut bacterial amyloid promotes a-synuclein aggregation and motor impairment in mice. Elife *9*. https://doi.org/10.7554/eLife.53111.

Sivanathan, V., and Hochschild, A. (2013). A bacterial export system for generating extracellular amyloid aggregates. Nat. Protoc. *8*, 1381–1390. https://doi.org/10.1038/nprot.2013.081.

Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.-H., Westover, B.P., Weatherford, J., Buhler, J.D., and Gordon, J.I. (2005). Glycan Foraging *in vivo* by an Intestine-Adapted Bacterial Symbiont. Science (80-.). *307*, 1955–1959. .

Valguarnera, E., Scott, N.E., Azimzadeh, P., and Feldman, M.F. (2018). Surface Exposure and Packing of Lipoproteins into Outer Membrane Vesicles Are Coupled Processes in Bacteroides . MSphere *3*. https://doi.org/10.1128/msphere.00559-18.

Vogt, N.M., Kerby, R.L., Dill-McFarland, K.A., Harding, S.J., Merluzzi, A.P., Johnson, S.C., Carlsson, C.M., Asthana, S., Zetterberg, H., Blennow, K., et al. (2017). Gut microbiome alterations in Alzheimer's disease. Sci. Reports 2017 71 *7*, 1–11. https://doi.org/10.1038/s41598-017-13601-y.

Worsdorfer, B., Woycechowsky, K.J., and Hilvert, D. (2011). Directed Evolution of a Protein Container. Science (80-.). *331*, 589–592. https://doi.org/10.1126/science.1199081.