Formation of functional, non-amyloidogenic fibres by recombinant *Bacillus subtilis* TasA

Elliot Erskine[1*], Ryan J Morris [2*], Marieke Schor [2*], Chris Earl [1], Rachel M. C. Gillespie [1], Keith Bromley [2] Tetyana Sukhodub[1], Lauren Clark[2], Paul K. Fyfe[1], Louise C. Serpell [3], Nicola R. Stanley-Wall [1#], Cait E. MacPhee[2#]

[1] Division of Molecular Microbiology, School of Life Sciences, University of Dundee, Dundee, UK DD1 5EH

[2] James Clerk Maxwell Building, School of Physics, University of Edinburgh, The Kings Buildings, Mayfield Road, Edinburgh EH9 3JZ

[3] School of Life Sciences, University of Sussex, Falmer, BN1 9QG

*These authors contributed equally and are listed alphabetically

# These authors contributed equally

Running Title: Functional recombinant non-amyloidogenic TasA fibres

To whom correspondence should be addressed:

Prof. Nicola R. Stanley-Wall

Division of Molecular Microbiology, School of Life Sciences, University of Dundee, Dundee DD1 5EH.

Email: n.r.stanleywall@dundee.ac.uk

OR

Prof. Cait MacPhee

James Clerk Maxwell Building, School of Physics and Astronomy, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom

Email: cait.macphee@ed.ac.uk

Keywords: **TasA fibres, *Bacillus subtilis*, biofilm matrix, functional amyloid, amyloid-like fibres**

Contributions. Conceived and designed the experiments: CE, EE, RG, CEM, RJM, MS, NSW; Performed the experiments: KB, LC, CE, EE, PKF, RG, CEM, RJM, MS, TS; Contributed new analytical tools: CE, EE, RG, TS; Analysed the data: CE, EE, CEM, RJM, MS, LCS, NSW; Wrote the paper: EE, RJM,CEM, MS, NSW.

Abstract max

Bacterial biofilms are communities of microbial cells encased within a self-produced protective polymeric matrix. The extracellular, fibrous *Bacillus subtilis* protein TasA is essential for structural integrity of the biofilm. We have developed a recombinant system that allows us to interrogate TasA form and function, and reveal that in both monomeric and fibre forms the secondary structure is comparable. We show that recombinant TasA fibres are biologically active and form without the aid of a partner protein, whereas a TasA variant restricted to an inactive monomeric form is subjected to extracellular proteolysis. The biophysical properties of the TasA fibres indicate that they are not functional amyloid fibres. We show that a gel formed by TasA fibres can recover after physical shear force, suggesting that the biofilm matrix is not static and that these properties may enable *B. subtilis* to remodel its local environment in response to external cues. Finally through the use of recombinant fibres formed by TasA orthologues, we uncover species variability in the ability of heterologous fibres to cross-complement the *B. subtilis tasA* deletion strain. These findings are indicative of specificity in the biophysical requirements of the TasA fibres across different species and/or reflect the precise molecular interactions that are needed to allow the biofilm matrix to assemble.

Significance

The protein TasA is an abundant component of the *Bacillus subtilis* biofilm matrix and has previously been classified as a “functional amyloid”. Functional amyloid-like fibres have been identified in bacteria and eukaryotes, and are viewed of interest due to their potential structural relationship to the disease-related amyloid fibres implicated in Alzheimer’s and Parkinson’s disease. Here we show that although TasA is a functional fibre-forming protein, it possesses few of the characteristics of canonical amyloid-like fibres and is likely made up of a helical assembly of globular protein units. This knowledge has implications for how *B. subtilis* biofilms may respond to their local environment when in a biofilm.

Key Words: TasA fibres, Bacillus subtilis, biofilm matrix, functional amyloid, amyloid-like fibres

Introduction

Biofilms are communities of microbial cells that underpin diverse processes including sewage bioremediation, plant growth promotion, chronic infections and industrial biofouling (1). The microbial cells resident in the biofilm are encased within a self-produced extracellular polymeric matrix that commonly comprises lipids, proteins, extracellular DNA, and exopolysaccharides (2, 3). This matrix fulfils a variety of functions for the community, from providing structural rigidity and protection from the external environment, to supporting signal transduction and nutrient adsorption (3, 4). *Bacillus subtilis* is a soil dwelling bacterium that is a model for biofilm formation by Gram-positive bacteria; beyond this it is of commercial interest due to its biocontrol and plant growth promoting properties that highlight its potential to substitute for petrochemical derived pesticides and fertilizers (5–7). Biofilm formation is subjected to complex regulatory pathways (8) and it is known that the *B. subtilis* biofilm matrix predominantly comprises three specific components. The first is an exopolysaccharide that serves to retain moisture within the biofilm and functions as a signalling molecule (9, 10) . The composition of the exopolysaccharide remains unclear due to three different monosaccharide composition analyses being detailed thus far (11–13). The second component is the protein BslA that is responsible for the non-wetting nature of the biofilm (14, 15). We have previously determined the structure of BslA (14), identified a unique structural metamorphosis that enables BslA to self-assemble at an interface in an environmentally-responsive fashion (15), and discovered that BslA is also important for determining biofilm architecture, independently of its ability to render the surface of the biofilm water-repellent (16). The third component of the biofilm matrix is the protein TasA (together with accessory protein TapA) that is needed for biofilm structure including attachment to plant roots (17–19).

TasA is a product of the *tapA-sipW-tasA* locus (20). It is post-translationally modified by SipW (21), a specialized signal peptidase that releases the mature 261-amino acid TasA into the extracellular environment where it forms long protein fibres that contribute to the superstructure of the biofilm matrix and are needed for biofilm integrity (22, 23). TasA fibres can be extracted from *B. subtilis* biofilms, and exogenous provision to a *tasA* null strain has previously been shown to reinstate structure to floating pellicles (23). Due to the reported ability of TasA fibres to bind the dyes Congo Red and thioflavin T (ThT), *ex vivo* purified TasA fibres have been classified as functional bacterial amyloid fibres (23), placing them in a class with the curli fibres of *E. coli* (24). In addition to functions involved in the process of biofilm formation, TasA is also linked with sliding motility (25) and spore coat formation (21, 26).

Amyloid-like fibres are well-known for their association with diseases like Alzheimer's and Parkinson's (27). In these conditions, highly stable fibrillar protein deposits are found in the brain, and are associated with cell damage (28). The amyloid fibres in these deposits are characterised by several properties: i) β-sheet-rich structures that are assembled into the canonical "cross-β" structure; ii) the ability of the fibres to bind the dye Congo Red and exhibit green birefringence under polarised light; iii) kinetics of formation that indicate a nucleated self-assembly process; and iv) a fibre structure that is unbranched, 6-12 nm in diameter, and often microns in length (29, 30). Once formed, these protein aggregates are highly stable, and in many cases are thought to be the lowest energy structural form shorter polypeptide chains can adopt (31). 'Functional' amyloid fibres refer to fibrillar protein deposits that share the characteristic structural properties of amyloid fibres, but are beneficial to the organism rather than being associated with disease (32). Significant caution is required in identifying functional amyloid-like fibres from predominantly *in vitro* data however, as many proteins and peptides can be induced to adopt the canonical amyloid fibre cross-β fold through appropriate manipulation of solution conditions such as changes in pH, temperature, cosolvent, salt or the presence of an interface (e.g. 34–38). Indeed, the ability of proteins to assemble into the cross-β architecture appears to be a 'generic' property of the polypeptide chain, independent of the amino acid sequence or the native structure of the precursor (38, 39).

Here we show that, although TasA is a fibre-forming protein, it is not amyloid-like in character. In contrast to canonical amyloid-like fibres, TasA fibres do not demonstrate an extensively β-sheet character (23, 40), and self-assembly of TasA oligomers into fibres *in vitro* has only previously been induced on hydrophobic surfaces, or after acid treatment (40), conditions that are routinely used to induce non-specific amyloid-like self-assembly in proteins including, for example, insulin (41). We have produced recombinant TasA in both fibre and monomeric forms, and show that the secondary structures of these are both identical to each other and to those reported previously for the exogenous purified TasA fibres, and appear predominantly helical in character. The recombinant form of the fibrous TasA protein is able to reinstate biofilm structure to a Δ*tasA sinR* deletion strain, and is thus biologically active. X-ray fibre diffraction of the recombinant TasA fibres shows that they are assembled from a helical repeat of globular protein units arranged approximately 45 Å apart, and the data are not consistent with the canonical "cross-β" diffraction pattern associated with amyloid-like fibres. Neither monomeric nor fibrous forms of recombinant TasA bind the dyes Congo Red or ThT, and although heat treatment induces a structural change into a different fibrous form that binds both ThT and Congo Red, X-ray fibre diffraction shows that this form is also unlikely to be amyloid-like in character. This ThT-positive fibre form of TasA is also no longer biologically

functional. Thus, TasA does not fall into the class of “functional amyloid-like fibres”; nonetheless it plays a critical role in biofilm structure.

Results

TasA forms non-amyloid fibres and is rendered monomeric by placement of a single N-terminal amino acid.

We predicted the identity of the N-terminus of mature TasA protein *in silico* using SignalP v4.2 (42) and subsequently confirmed this *in vivo* using mass spectrometry. Based on this information we designed an expression construct to allow purification of recombinant *B. subtilis* TasA (Fig. SI 1A), corresponding to the mature TasA sequence covering amino acids 28-261, after production in *E. coli* (Fig. SI 1B). The purified protein displayed obvious viscosity, not flowing upon inversion of the tube, and bead tracking microrheology confirmed the gel-like nature of the solution (Fig. SI 1C). This viscosity results from the formation of a fibrous aggregate that, when viewed by transmission electron microscopy (TEM) (Fig. 1A-B), shares some of the characteristics of the fibres extracted from native *B. subtilis* biofilms (23). Within these fibres we observed a subunit repeat along the fibre axis, repeating at approximately 4 nm. Hereafter we refer to this protein as fTasA. Circular dichroism (CD) spectroscopy shows a minimum at 208 nm, a shoulder at ~222 nm, and a maximum below 200 nm. The overall shape and the position of the minima are consistent with a predominantly (>50%) α-helical conformation, however the ratio of the two minima suggests there are likely to be contributions from other secondary structural elements (Fig. 1C). The spectrum resembles that reported for TasA oligomers (40) or fibres purified directly from *B. subtilis* biofilms (23), and does not display the high β-sheet content (represented by a single minimum at ~216-218 nm) typical for proteins in amyloid-like fibres; nor do the measured minima/maximum correspond to those predicted or measured for highly twisted β-sheets (43). X-ray fibre diffraction from partially aligned TasA fibres (Fig. 1D) showed a series of layer lines on the meridian. The lowest resolution visible was 41-45 Å/4.1-4.5 nm and further strong layer lines were measured at 21 Å and 14.8 Å. A weak, high resolution meridional diffraction signal was observed at 4.15 Å. These spacings are consistent with a helical or globular repeat distance of 40-45 Å. On the equator of the pattern, strong diffraction signals were measured at approx. 40 Å, 28.1 Å and 15.5 Å. [Accurate measurement of the reflections close to the backstop is not possible]. These spacings are consistent with packing of a fibre with globular units of dimensions 28 x 45 Å. The diffraction signals expected for a cross- β amyloid-like structure (i.e. 4.7 Å (the inter-strand distance) and ~10-12 Å (the intersheet distance) (44)) were not observed.

Through addition of a single amino acid to the N-terminus (Fig. SI 1D), we discovered that it was possible to block fibre formation *in vitro.* Inhibition of TasA fibre formation was found not to depend on the chemical properties of the single added amino acid, and was prevented by the addition of a lysine, phenylalanine, glutamic acid, alanine, or serine to the N-terminal end of mature TasA (Fig. SI

1E-G). Having compared the behaviour at this level we focussed our subsequent analysis on the purified monomeric serine-tagged TasA (Fig. SI 1F), where size-exclusion chromatography (SEC) showed a single peak of around 30 kDa (Fig. SI 1G). The molecular mass was confirmed by LC-MS, ensuring the specificity of the protein sequence. Monomeric TasA (mTasA) was not viscous, with bead tracking microrheology confirming the liquid-like nature of the purified protein solution (Fig. SI 1C); moreover, no fibres were apparent by TEM. CD spectroscopy indicated that the secondary structure of this monomeric mTasA was indistinguishable from fTasA (Fig. 1C) indicating that addition of a single amino acid to the N-terminus did not affect the secondary structure, and that the fibrous form is likely constructed from a linear assembly of these monomeric units. Thus, examination of TasA form and function in both fibrous and monomeric states was possible.

Recombinant TasA fibres are biologically active

A Δ*tasA* deletion mutant (NRS5267) was constructed (Fig. 2A and SI 2A) to test the functionality of the recombinant TasA fibres. The deletion could be genetically complemented by expression of *tasA* under the control of an inducible promoter at the ectopic *amyE* locus (NRS5313) (Fig. SI 2B-C). In contrast, exogenously added recombinant fTasA (Fig. SI 3A) was unable to reinstate biofilm architecture to the Δ*tasA* mutant (Fig. 2B and SI 3B). Likewise, no change in biofilm phenotype was detected when monomeric TasA was added (Fig. 2B and SI 3B). We considered two possibilities: 1) the protein was unstable in the presence of cells when added exogenously, or 2) the protein was stable, but not functional. Immunoblot analysis of whole biofilm protein extracts using anti-TasA antibodies revealed that while fTasA does not influence biofilm morphology when cultured with the *tasA* deletion strain, the exogenously added protein is still detectable after 48 hours incubation (Fig. 2C). In contrast, immunoblot analysis revealed that mTasA reached undetectable levels after incubation with the *tasA* strain (Fig. 2C). Subsequent analysis of the stability of recombinant fTasA and mTasA incubated in cell-free spent culture supernatant revealed that assembly of TasA into the fibre form confers protection from degradation by extracellular proteases (Fig. 2D and SI 3C-D)

SinR is a major repressor of biofilm formation that functions, in part, by directly inhibiting transcription from the operons needed for the production of the exopolysaccharide and TasA components of the biofilm matrix (45). Deletion of *sinR* results in a biofilm that is densely wrinkled and highly adherent to a surface when compared to the parental strain due to increased production of the biofilm macromolecules (Fig. 2A). While constructing the Δ*tasA* strain we inadvertently isolated a Δ*tasA sinR* deletion strain (Fig. SI 2D), which displayed a flat, featureless biofilm by comparison with a *sinIR* deletion strain (Fig. 2A). When the *tasA* coding region was expressed from the ectopic *amyE* locus (Fig. SI 2E) in the Δ*tasA sinR* deletion strain (NRS5255) the densely wrinkled

biofilm structure was recovered (Fig. SI 2F). Serendipitously, we found that upon co-culture with 10 µg of recombinant fTasA, the flat, featureless morphology of the Δ*tasA sinR* double mutant could be broadly returned to the wrinkled *sinR* mutant-like phenotype (Fig. 2E and SI 3B). Thus, the recombinant form of fTasA is biologically functional. In contrast, monomeric TasA protein did not reinstate rugosity to the Δ*tasA sinR* deletion strain (Fig. 2E and SI 3B) suggesting that *in vivo* templating of mTasA into a functional fibrous form does not occur. Moreover, as was observed when protein was supplied exogenously to the single *tasA* deletion strain, mTasA protein was not detectable by anti-TasA immunoblot analysis after co-incubation with the Δ*tasA sinR* deletion strain, but fTasA was detected (Fig. 2C), indicated the susceptibility of mTasA to proteolysis.

Recombinant TasA fibres require the biofilm exopolysaccharide for activity

TasA protein fibres are reported to be anchored to the cell wall *via* an interaction with a partner protein called TapA (18). Moreover, deletion of *tapA* is associated with a reduction in the level of TasA (46). To test if TapA is needed for wrinkling of the Δ*tasA sinR* deletion strain upon addition of preassembled TasA fibres, we constructed a Δ*tapA* Δ*tasA sinR* triple deletion strain (Fig. 3A and SI 4A). This strain could be returned to the *sinR* morphology upon genetic complementation with the *tapA-sipW-tasA* gene cluster at an ectopic location in the chromosome (Fig. SI 2H-I). When fTasA was co-cultured with the Δ*tapA* Δ*tasA sinR* strain, we visually observed similar levels of *ex vivo* complementation by fTasA in the Δ*tapA* Δ*tasA sinR* triple mutant compared with the Δ*tasA sinR* deletion strain (Fig. 2D and SI 3A), thus suggesting that TapA is not required to reinstate biofilm architecture when fully formed TasA fibres are supplied.

In light of the findings above we explored if the rugosity displayed by the Δ*tasA sinR* in the presence of *ex vivo* recombinant fTasA was due to a specific interaction with the matrix components, or whether the presence of sufficient fTasA fibres is enough to confer rugosity simply due to the gelatinous nature of the concentrated protein solution. To test this we took two approaches. First we tested if an entirely unrelated fibrous protein could substitute for fTasA, simply by provision of a fibrous protein scaffold. Therefore, we provided amyloid-like fibres assembled from the well-characterised transthyretin peptide 105-115 ($TTR_{105-115}$)(47) exogenously to the Δ*tasA sinR* strain followed by incubation under biofilm formation conditions. However, despite the obvious viscosity of the $TTR_{105-115}$ gel, it did not reinstate biofilm rugosity (Fig. 3B and SI 4B). Therefore, a biochemically distinct fibre cannot substitute for fTasA. Next, we assessed whether the biofilm exopolysaccharide was needed for rugosity under these conditions. This experiment was based on the premise that if the wrinkle formation after addition of exogenous fTasA was derived from the gelling properties of fTasA, the exopolysaccharide would not be needed. To determine this we used

a strain where the entire *eps*A-O operon was placed under the control of an inducible promoter at the native location on the chromosome (48). We then added fTasA in the presence and absence of induction of the *epsA-O* operon. Analysis of the biofilm phenotypes revealed that we were able to induce rugosity with fTasA only in the presence of IPTG (Fig. 3C and SI 4C), although not to the same extent as seen in the parent strain - most likely as production of the exopolysaccharide is uncoupled from its native regulation circuity, impacting the level of polymer produced. Therefore we can conclude that both the biofilm exopolysaccharide and fTasA are required to return rugosity to the biofilm, and that the recombinant fTasA fibres are biologically active *in vivo*.

Increased temperature induces a transition to another fibre form

Binding of the dyes Congo Red (CR) and Thioflavin T (ThT) to β-sheet-rich secondary structural elements is often used to report on amyloid fibril growth (49), although both dyes are only semi-specific (see *Discussion*). At room temperature, both mTasA and fTasA showed little evidence of binding to ThT, with a fluorescence emission approximately 2.5-fold higher than fluorophore alone (~123 *vs* 325 arbitrary units). Likewise, we found no evidence that recombinant fTasA bound the dye Congo Red since we observed no significant redshift in absorbance (Fig. SI 5A-B).

It has previously been reported that fibrous TasA can be isolated from native *B. subtilis* biofilms (23). This yielded large fibrous aggregates that required treatment with 10% formic acid for 5 minutes to yield monomeric protein. A later report, however, indicated that TasA purified in this way from native biofilms yielded small oligomers (40). In the former case, the depolymerised fibres acquired the ability to bind ThT following resuspension at pH7; in the latter case the oligomers acquired the ability to bind ThT following addition of formic acid to a pH of ~2.5. We therefore assessed if changes in environmental conditions (temperature or solution pH) would induce mTasA or fTasA to convert into a fibre form that bound ThT. At elevated temperatures, both mTasA and fTasA convert to structures that bind ThT, eliciting a ~8-fold increase in fluorescence intensity (Fig. 4A). This fluorescence enhancement can be compared to amorphous (~7-fold enhancement) and fibrillar (~35-fold enhancement) aggregates of lysozyme (50). Additionally, the temperature-converted form of fTasA caused a redshift in Congo Red absorbance (Fig. SI 5A-B), which is suggestive of binding to cross-β architectures. This conversion takes place over a period of 10-20 min (mTasA) or ~2 hrs (fTasA) at 60 degrees at pH 7 (Fig. 4A). Conversion to a ThT-positive form was reproducible for both mTasA and fTasA at pH 7 at 45°C, with fTasA showing enhanced stability compared to mTasA (Fig. SI 5C). Acidic pH had little effect on the kinetics of conversion to a ThT-positive form at temperatures above 45°C (Fig. SI 5D) but enhanced the rate of conversion at 37°C.

The nature of the heat converted, ThT-binding protein structures was further characterised using CD, TEM, and XRD. The CD spectra show an apparent increase in random coil for both fTasA and mTasA, although the change was more significant for the monomeric protein (Fig. 4B). Subsequent TEM imaging shows that upon incubation at 60°C both the previously monomeric mTasA, as well as fTasA, form short, worm-like fibrils (Fig. 4C). Binding of ThT and Congo Red can indicate a structural conversion to an amyloid-like conformation, with an increased level of non-native β-sheet structure and assembly into a cross-β architecture, however our x-ray diffraction data (Fig. 4D) are not consistent with the canonical "cross-β" or β-sheet diffraction patterns. We reproducibly observe a series of diffuse reflections at 3.9, 5.8 and 8.3 Å, and although we have not been able to assign these to a specific packing or structure (Fig. SI 5F), they do not correspond to the 4.7 Å and ~10-12 Å characteristic of amyloid-like fibres. Thus, although heating and/or acid treatment do indeed convert both mTasA and fTasA into a form that binds ThT and Congo Red, this form is also not amyloid-like in structure.

The temperature-induced fibre forms of TasA have no bioactivity

We have shown that recombinant, non-amyloidogenic TasA fibres are functional in a biofilm context (Fig. 2E); this raised the question of whether conversion into worm-like fibres changed the functionality of fTasA. We attempted to complement the Δ*tasA sinR* deletion strain by exogenous provision of fTasA after pre-incubation at 60°C for 24 hours. SDS-PAGE analysis demonstrated that the protein was not degraded after the treatment regime (Fig. 5A). However, co-culture of the Δ*tasA sinR* deletion strain with heat-treated fTasA showed that it was unable to reinstate structure to the biofilm (Fig. 5B). Consistent with the lack of bioactivity the heat treated fTasA exhibited limited viscosity as measured by bead tracking microrheology (Fig. SI 5E). Immunoblot analysis demonstrated that no TasA was detectable following incubation with the cells, indicating the temperature-treated fTasA wormlike fibrils became susceptible to degradation (Fig. 5C); correspondingly, the heat treated fTasA protein was sensitive to exoprotease degradation when it was incubated with cell-free spent culture supernatant (Fig. 5D). Therefore heat treatment of fTasA to yield ThT binding properties renders the fibres non-functional.

Biophysical properties of recombinant orthologous TasA

Using the *B. subtilis* TasA protein sequence we identified orthologous proteins from a range of *Bacillus* species. The sequences were aligned using Clustal Omega (51) (Fig. SI 6) and used to generate a phylogenetic tree (Fig. 6A). Further analysis of gene synteny within the *tapA* operon revealed two distinct sub-classes based on the presence or absence of *tapA*. This had previously been noted for *B. cereus* which contains two TasA paralogues (52). Highlighted on the phylogenetic

tree are *B. amyloliquefaciens* TasA, *B. licheniformis* TasA and TasA and CalY from *B. cereus* (Fig. 6A) that were chosen for further analysis. Each of these proteins are predicted to encode an N-terminal signal sequence and were used to establish if *1)* orthologous TasA fibres assembled *in vitro* after purification of the predicted mature protein; and *2)* if any fibres formed could cross-complement the *B. subtilis tasA* deletion strain.

The quality and identity of the recombinant TasA orthologous proteins was confirmed by SDS-PAGE (Fig. SI 7A) and mass spectrometry (Fig. SI 7B). CD spectroscopy indicated that the secondary structures of *B. licheniformis* and *B. amyloliquefaciens* TasA, and *B. cereus* CalY, were broadly similar to that of *B. subtilis* TasA, with a primary minimum at 208 nm, a shoulder at ~222 nm, and a maximum below 200 nm (Fig. 6B). In contrast, *B. cereus* TasA has a single broad minimum centred on ~216 nm (Fig. 6B), suggesting this protein may contain increased β-sheet content, although both the breadth and the intensity of the minimum suggest significant remaining contribution from helical elements. TEM imaging revealed that all of the orthologous proteins spontaneously self-assembled into fibres (Fig. 6C and SI 7C), and all showed evidence of a regular subunit repeat along the fibre axis of approximately 4-5 nm. The finding that proteins corresponding to the mature region of *B. cereus* TasA and CalY form fibres *in vitro* is consistent with previous data from *B. cereus*, which revealed the presence *in vivo* of extracellular fibres dependent on *tasA* and *calY* (52) and with our findings that TapA is dispensable for TasA fibre formation *in vitro.* Through the TEM imaging, we observed qualitative differences between the ability of the different proteins to form fibre bundles, with the *B. cereus* proteins TasA and CalY forming thick fibre bundles, *B. licheniformis* and *B. subtilis* TasA forming intermediate-diameter fibre bundles, and *B. amyloliquefaciens* forming a distributed mesh of thin fibres.

Functionality of orthologous protein fibres in *B. subtilis*

To test the ability of the orthologous proteins to function in place of *B. subtilis* TasA fibres, 10 µg of each recombinant fibrous protein was exogenously added to the Δ*tasA sinR* mutant. The cells were then incubated under biofilm formation conditions. We determined that rugosity of the biofilm community could be recovered when the more closely related *B. amyloliquefaciens* and *B. licheniformis* TasA proteins were provided but not when either of the more divergent *B. cereus* proteins were supplied (Fig. 6D and SI 8A). This is in contrast to previously published data where expression of both *B. cereus calY* and *tasA*, alongside the signal peptidase *sipW*, was reported to recover biofilm formation to a *B. subtilis tasA* mutant (52). Analysis of the stability of the protein fibres after incubation with spent cell-free culture supernatant revealed that *B. cereus* TasA fibres, like the *B. amyloliquefaciens* and *B. licheniformis* TasA fibres, were resistant to exoprotease

degradation while CalY fibres were susceptible (Fig. SI 8B-C). From our analyses of protein function we can conclude that either the interaction of TasA fibres with the *B. subtilis* matrix is dependent on the exact identity of the TasA fibres, suggesting specific molecular interactions with other matrix molecules, or that the subtle differences in the physiochemical properties of the TasA fibres may be influential in establishing rugosity in the bacterial biofilm. For example, after serially diluting the recombinant protein, and therefore shearing of the samples, recombinant *B. cereus* TasA was significantly less viscous than the equivalent samples of *B. licheniformis* and *B. subtilis* TasA (Fig. SI 7D-F); it may be that shearing of the samples breaks the thicker bundles observed in TEM. However, after allowing all samples to recover for 3 days, both *B. cereus* TasA and *B. licheniformis* TasA formed a gel at a lower concentration of protein than *B. subtilis* fTasA (Fig. SI 7D-F). The origins of the variability in the properties of the gels formed by the fibrous TasA orthologues, and its possible link to the bundling observed by TEM, may have implications for the mechanical properties of *in vivo* biofilms and the ability of one orthologue of TasA to substitute for another.

Discussion

We have demonstrated that recombinant fibrous TasA can return rugosity to a *B. subtilis* Δ*tasA sinR* deletion strain and is thus biologically functional. Biophysical analysis indicates that these fibres are assembled as a helical arrangement of globular units that lack the characteristic "cross-β" architecture of canonical amyloid-like fibres. The CD spectrum of the recombinant protein resembles that published previously for TasA isolated directly from native *B. subtilis* biofilms (23, 40) and is suggestive of a predominantly helical secondary structure. Moreover, we have demonstrated that recombinant TasA can be rendered monomeric by the addition of a single amino acid to the N-terminus, and that this monomeric protein shares the same secondary structure as the fibrous form. This strongly suggests that the fibres comprise a linear assembly of these monomeric units, with no large structural rearrangement, although domain-swapping between monomers cannot be ruled out. Indeed, a repeating unit is visible along the length of the fibre axis, most clearly in the TEM images of recombinant fibres of the orthologous TasA protein from *B. cereus* where the protein subunits appear horizontally aligned across a fibre bundle, but also visible in all orthologous forms of TasA we have examined. Such a structure is not consistent with current structural models of amyloid-like fibrils, which comprise a single continuous hydrogen-bonded array along the long axis of the fibril. Taken together, our data indicate that TasA is unlikely to fall into the class of functional amyloid-like fibres. Moreover, although we were able to generate a form of fibre that binds Congo red and also enhances ThT fluorescence, this form was also not amyloid-like in nature as assessed by CD and by XRD, and moreover had no measureable biological function.

Caution should be taken when inferring the formation of amyloid-like fibres from enhanced fluorescence following binding of ThT, which exhibits enhanced fluorescence in the presence of globular proteins including bovine serum albumin (53), human serum albumin (54) and acetylcholinesterase (55); in the presence of amorphous aggregates of lysozyme and bovine serum albumin (50), and amorphous aggregates formed by a thrombin-derived C-terminal peptide (56); as well as in the presence of non-amyloid wormlike aggregates of an artificial dimer of an Aβ peptide (57). Conversely, ThT does not exhibit enhanced fluorescence in the presence of, for example, cross-β fibrils formed by poly-L-lysine (58, 59). Thus, ThT enhanced fluorescence should only be considered suggestive, but not indicative, of amyloid-like fibre formation.

The widespread nature of functional amyloid fibres in bacterial biofilms has been hypothesized, and a well-characterised example is the curli fibres of *E.coli*, *Enterobacter cloacae*, and *Salmonella* spp (60). These show a CD spectrum, dye-binding behaviour and proteolytic insensitivity that are consistent with an amyloid-like β-sheet structure, but solid-state NMR data suggests an architecture

comprising stacked β-helical subunits (61), a structural motif commonly employed by bacteria (62). Many amyloid-like fibres formed *in vitro* from proteins associated with disease show an in-register parallel cross-β arrangement (63); recently however native Tau filaments from the human brain of an Alzheimer's Disease patient have been demonstrated to form an elaborate mixed β-helix/ cross-β structure formed of in-register, parallel β-strands (64). Thus, both cross-β and β-helix structures appear to be characteristic of canonical amyloid fibres.

Making the correct distinction between amyloid-like and non-amyloid fibrous proteins is more than a semantic argument: a number of papers have drawn a link between functional amyloid-like fibres formed by bacteria and their relevance to human disease (60, 65, 66), for example in the determination of the mechanistic details of self-assembly, or in the possible discovery of new therapeutics. As the amyloid-like fibre macrostructure is thought to be a 'generic' property deriving from the chemical structure of the polypeptide backbone that is common to all proteins and peptides – and thus to a large extent independent of primary sequence, although this will influence overall fibre morphology - small drug molecules that target the generic amyloid fold (as Congo Red and ThT to some extent do) may have widespread applicability in a number of devastating human diseases. Thus it is important to make the distinction between non-amyloid fibrous assemblies and amyloid-like fibres appropriately.

The fibrous nature of TasA likely imparts mechanical rigidity to the biofilm, thereby restoring the highly wrinkled architecture characteristic of the *sinR* deletion strain. It is unclear why the fTasA protein cannot recover biofilm architecture to the single *tasA* deletion, however since SinR has pleiotropic roles in biofilm formation (8, 67) it may be that overproduction of the biofilm polysaccharide compensates for the loss of native regulation that intricately controls native TasA production in space and time (68). Our results also indicate that when in a fibrous form, TasA does not require the TapA protein to fulfil this function, which contradicts previous reports suggesting that TapA is an accessory protein required for correct TasA assembly and localisation (18). Therefore the role played by TapA in biofilm formation, while evidently essential (45), is unclear. It may be that while TapA is not essential for TasA fibre formation *in vitro*, it functions as a chaperone *in vivo* to aid the transition of monomeric TasA into a fibrous state. This hypothesis is consistent with the overall reduction in the level of TasA and the corresponding reduction in the number of TasA fibres observed in the *tapA* mutant (18). Moreover it is consistent with the demonstration that monomeric TasA, but not fibrous TasA, is susceptible to degradation by the extracellular proteases.

A non-amyloid-like structure for TasA is possibly beneficial in the context of the *B. subtilis* biofilm; amyloid-like self-assembled fibres are very stable, with curli fibres, for example, requiring treatment

with concentrated acid solutions to drive disassembly (24). Curli fibres also appear to form a brittle matrix which, once fractured, does not recover (69). In contrast, we have shown that the gelation properties of fibrous TasA solutions recover after shear (Fig. SI 7D-F), suggesting that *in vivo* the biofilm matrix could be remodelled in response to mechanical environmental perturbations. The TasA fibres may also be in equilibrium with the monomeric form of the protein, which would allow dynamic restructuring of the biofilm in response to environmental changes. As the fibrous form of the protein confers protection against degradation by extracellular proteases whereas the monomeric protein is degraded, an appropriate secretion of monomeric protein and/or proteases could provide dynamic control of biofilm elasticity and structure.

Materials and Methods

Growth conditions

E. coli and *B. subtilis* were routinely grown in Lysogeny Broth (LB) media (10 g NaCl, 5 g yeast extract and 10 g tryptone per litre) or plates supplemented with the addition of 1.5% (w/v) select agar (Invitrogen). Samples were grown at 37°C unless stated otherwise. When required, antibiotics were used at the following concentrations: ampicillin (100 µg ml^{-1}), spectinomycin (100 µg ml^{-1}) and chloramphenicol (5 µg ml^{-1}). For biofilm assays MSgg minimal media was used (5 mM KH_2PO_4 and 100 mM MOPS at pH 7 supplemented with 2 mM $MgCl_2$, 700 µM $CaCl_2$, 50 µM $MnCl_2$, 50 µM $FeCl_3$, 1 µM $ZnCl_2$, 2 µM thiamine, 0.5% glycerol, 0.5% glutamate). When appropriate isopropyl β-D-1-thiogalactopyranoside (IPTG) was added at the indicated concentration. For protein production auto-induction media (6 g Na_2HPO_4, 3 g KH_2PO_4, 20 g Tryptone, 5 g yeast extract, 5 g NaCl, 10 ml 60% v/v glycerol, 5 ml 10% w/v glucose and 25 ml 8% w/v lactose per litre at a 1:1000 volume ratio (supplemented with 100 µg/ml ampicillin)) was used (70).

Strain construction

A complete list of *E. coli* and *B. subtilis* strains used in this study can be found in Table S1. Plasmids and primers are detailed in Tables S2 and S3 respectively. All *B. subtilis* strains used for physiological assays were derived from the wild-type laboratory isolate NCIB3610 and constructed using standard protocols. SSP1 phage transductions for DNA transfer into *B. subtilis* NCIB3610 were carried out as previously described (71).

Plasmid construction and mutagenesis

Construction of an in-frame *tasA* deletion in NCIB3610 was achieved using the pMiniMAD (72) temperature sensitive allelic replacement vector pNW1448 (pMiniMAD-Δ*tasA*). The plasmid was constructed by PCR amplification of two fragments: the 511 bp upstream of *tasA* including the first 6 bp of *tasA* coding sequence and a second fragment covering the last 3bp of the *tasA* coding sequence, the stop codon and the 512 bp downstream using primer pairs NSW2005/NSW2006 and NSW2007/NSW2008 respectively. The PCR fragments were each digested with SalI/EcoRI and simultaneously ligated into the pMiniMAD plasmid that was digested with the same restriction sites to yield plasmid pNW1448. Plasmid pNW1448 was introduced into 168 and then transferred to NCIB3610 using phage transduction. The *tasA* deletion was introduced into the *B. subtilis* chromosome using the method described previously (73). After homologous recombination and selection for loss of the pMiniMad plasmid, two morphologically distinct isolates carrying the desired deletion in *tasA* were identified. Whole genome sequencing (see below) was used to genotype the isolates in an unbiased manner. Analysis of single nucleotide polymorphisms (Table S4) revealed one

strain carried a short duplication of the *sinR* coding region effectively yielding a *ΔtasA sinR* double mutant (NRS5248) while the other was a single Δ*tasA* strain (NRS5267).

The *tapA* in frame deletion was generated via the pMAD protocol as above, with amplification of the 395 bp upstream fragment using primers NSW1308 and NSW1332 and 641 bp downstream fragment using primers NSW1333 and NSW1334. The two PCR fragments were digested BamHI/SalI and EcoRI/SalI respectively and ligated into the intermediate plasmid pUC19 yielding pNW686, and was subsequently moved into pMAD to generate pNW685. Plasmid pNW685 was introduced into *B. subtilis* 168, generating strain NRS3789, and transferred to NCIB3610 using phage transduction. The same 168 strain was used to generate the Δ*tapA* Δ*tasA* and Δ*tapA* Δ*tasA sinR* strains by transferring via phage to Δ*tasA* (NRS5267) and Δ*tasA sinR* (NRS5248) respectively.

Genetic complementation of Δ*tasA* and Δ*tasA sinR* was achieved by PCR amplification of the *sipW-tasA* (using primers NSW2218 and 2219) and *tasA* (using primers NSW1857 and NSW1858) regions respectively from NCIB3610. Both were cut using SalI/SphI restriction enzymes and ligated into the pDR110 vector. Plasmids pNW1619 and pNW1432 were introduced into 168 and transferred to Δ*tasA* (NRS5267) and Δ*tasA sinR* (NRS5248) respectively using phage transduction.

Genetic complementation of the double Δ*tapA* Δ*tasA* (NRS5748) and triple Δ*tapA* Δ*tasA sinR* (NRS5749) mutants was performed using the whole *tapA-sipW-tasA* operon which was amplified from NCIB3610 using primers NSW1896 and NSW2219, cut SalI/SphI and ligated into pDR110 to generate pNW1804 which was introduced to both strains by phage transduction via 168 at the ectopic amyE location on the chromosome.

Protein purification was achieved using GST fusion constructs. The *tasA* overexpression plasmid pNW1437 (pGex-6-P-1-TEV-$tasA_{(28\text{-}261)}$) was generated by amplifying the $tasA_{(28\text{-}261)}$ coding region from *B. subtilis* NCIB3610 genomic DNA using primers NSW660 and NSW661 and insertion into the vector pGEX-6P-1 cleaved BamHI/XhoI yielding the vector pNW543. The TEV protease cleavage site was next introduced by site-directed mutagenesis using primers NSW1892 and NSW1893 to give pNW1437. Amino acids were introduced at the N-terminal end of *tasA* also by site-directed mutagenesis; primer pairs are indicated in Table S2. The constructs used to purify the TasA orthologue proteins were generated in a similar manner from genomic DNA isolated from *B. cereus* ATCC14579, *B. licheniformis* ATCC14580 and *B. amyloliquefaciens* FZB42 and likewise primers used

for amplification are detailed. The plasmids were used to transform BL21 (DE3) *E. coli* strain for protein production.

Genome sequencing

Whole genome sequencing and bioinformatics analysis of strains NCIB3610, NRS5248 and NRS5267 was conducted by MicrobesNG (http://microbesng.uk) which is supported by the BBSRC (grant number BB/L024209/1). Three beads were washed with extraction buffer containing lysozyme and RNase A, incubated for 25 min at 37°C. Proteinase K and RNaseA were added and incubated for 5 minutes at 65°C. Genomic DNA was purified using an equal volume of SPRI beads and resuspended in EB buffer. DNA was quantified in triplicates with the Quantit dsDNA HS assay in an Ependorff AF2200 plate reader. Genomic DNA libraries were prepared using Nextera XT Library Prep Kit (Illumina™, San Diego, USA) following the manufacturer's protocol with the following modifications: 2 ng of DNA instead of one were used as input, and PCR elongation time was increased to 1 min from 30 seconds. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system. Pooled libraries were quantified using the Kapa Biosystems Library Quantification Kit for Illumina on a Roche light cycler 96 qPCR machine. Libraries were sequenced on the Illumina HiSeq using a 250 bp paired end protocol. Reads were adapter trimmed using Trimmomatic 0.30 with a sliding window quality cut-off of Q15 (74). De novo assembly was performed on samples using SPAdes version 3.7 (75) and contigs were ordered using Abacas (76) and annotated using Prokka 1.11 (77). Reads were aligned to the reference 168 genome (accession number: NZ_CM000487.1) using BWA-Mem 0.7.5 and processed using SAMtools 1.2 (78). Variants were called using VarScan 2.3.9 with two thresholds, sensitive and specific, where the variant allele frequency is greater than 90% and 10% respectively. The effects of the variants were predicted and annotated using SnpEff 4.2 (79) (Table S4).

Protein production and purification

The pGEX-6-P-1 GST-gene fusion system was used in the *E. coli* BL21 (DE3) strain for protein production (GE Healthcare™). After the required plasmid was introduced into BL21(DE3), a 5 ml LB culture (supplemented with 100 µg/ml ampicillin) was grown overnight at 37°C and used to inoculate 1 L of auto-induction media at 1/1000 dilution. The cultures were incubated at 37°C with 130 rpm shaking until optical density at 600 nm was approximately 0.9 at which point the temperature was lowered to 18°C and cultures were grown overnight. Cells were harvested by centrifugation at 4000 g for 45 minutes and the cell pellet was suspended in 25 ml of purification buffer (25 mM Tris-HCl, 250 mM NaCl, pH 7.5) supplemented with Complete EDTA-free Protease inhibitor (Roche) then lysed using an Emulsiflex cell disrupter (Avestin™) with 3 passes made at

~15000 psi or sonication at 25% for 6 minutes. Cell debris was removed by centrifugation at 27000 g for 35 minutes. The supernatant was removed and added to 450 µL of Glutathione Sepharose 4B beads (GE Healthcare™) and incubated on a roller for 3 hours at 4°C. The protein-bead mix was loaded onto disposable gravity flow columns (Bio-Rad™) and washed 3 times with 25ml of purification buffer. Beads were collected from the column and suspended in 25 ml of purification buffer supplemented with 1 mM DTT and 0.5 mg TEV protease and incubated on roller at 4°C overnight. The flow-through was then added to 300 µL GST beads and 250 µL Ni-NTA beads () and incubated for 2 hours at 4°C. Final pass through column removes beads and flow-through is concentrated using 10kDa Vivaspin™. For biophysical experiments performed at Edinburgh University, buffer exchange into 25 mM phosphate buffer (pH 7) was performed using same concentrators. Purity was determined by SDS-PAGE and molecular mass determined by loading 80 µg onto a qTOF liquid chromatography mass spectrometry performed by the FingerPrints Proteomics Facility at the University of Dundee.

Transthyretin fibre preparation

TTR fibres were prepared as previously described (80). Briefly, 0.8 mg of the peptide was dissolved in 200 µL 25mM phosphate buffer (pH 7) for *ex vivo* complementation and 20% (w/v) acetonitrile (pH 5) for X-ray diffraction pattern collection. Sample was sonicated for 10 minutes and combined with 10 µL of TTR seeds and incubated at 60°C for 5 hours.

Biofilm phenotypes, ex vivo complementation and protein collection

To characterise biofilm phenotype samples were set up as detailed previously (81) briefly, 10 µL of LB culture grown to mid-exponential phase was spotted onto solidified MSgg media and incubated for 2 days at 30°C. The resultant colonies were imaged using a Leica MZ16 stereoscope. For *ex vivo* complementation, 10 µg of recombinant protein (or 10 µL TTR where indicated) was pipetted with cells immediately prior to spotting. To release all biofilm proteins for subsequent analysis, the biofilm was resuspended in 500 µL BugBuster Master Mix (Novagen) followed by sonication and agitation for 20 mins at room temperature. Insoluble debris was removed by centrifugation at 17,000 g for 10 mins at 4°C.

SDS-PAGE

SDS-polyacrylamide gel electrophoresis (PAGE) was performed using 10 µg of purified TasA protein and 4X loading buffer (6.2 g SDS, 40 ml 0.5 M Tris pH 6.8, 6.4 ml 0.1 M EDTA, 32 ml 100% glycerol, 1 mg Bromophenol blue). Samples were heated at 99°C for 5 minutes prior to loading on the gel and

were run on a standard 14 % polyacrylamide SDS-PAGE at 200 V for 60 min, before staining with InstantBlue (Expedeon™).

Immunoblot Analysis

Samples were separated by SDS-PAGE and transferred onto a PVDF membrane (Millipore™) by electroblotting at 100 mA for 75 minutes. The membranes were blocked with 3% (w/v) milk in 1xTBS overnight at 4°C with shaking followed by 1 hr incubation with primary antibody (TasA (1:25000 v/v) as indicated) diluted in 3% (w/v) milk in 1x TBS. This was followed by 3 washes of 10 minutes each with 1x TBS and 2% (v/v) Tween20 and subsequent 45 minute incubation with goat anti-rabbit conjugated secondary antibody (1:5000 v/v) (Pierce™). Membrane was washed 3 times for 10 mins with TBST then developed by ECL incubation and exposing to X-ray film (Konica™) using the Medical Film Processor SRX-101A (Konica™). This is with the exception of the data shown in Figure 4C which was developed as detailed above and visualised using GeneGnomeXRQ (Synegene™).

Size-Exclusion Chromatography

Monomeric TasA was examined by size-exclusion chromatography using either Superdex 5/150 or 10/300 GL increase column as indicated (GE Healthcare) on an ÄKTA FPLC system using 25 mM Tris-HCl, 250 mM NaCl, pH 7 buffer. Column was calibrated using conalbumin (75000 Da), ovalbumin (44000 Da), carbonic anhydrase (29000 Da), ribonuclease A (13700 Da) and aprotinin (6500 Da) and void volume was calculated using blue dextran 2000 (GE Healthcare).

Exoprotease stability

PY79, PY79 KO6 and PY79 KO7 and/or NCIB3610 strains were grown to OD600 ~2.5 in 25 ml MSgg growth media at 37°C with 130 RPM shaking overnight. The cultures were normalised to same OD_{600} and 5 ml was centrifugation at 3750 *g* for 15 mins at 4 °C to pellet cells. The culture supernatant was collected and filtered through a 0.22 µM filter (Milipore) to remove residual cells. Aliquots of the culture supernatant generated by NCIB3610 and PY79 were heated inactivated at 100°C for 10 minutes. 15 µL of each cell-free culture supernatant was incubated with 30 µg recombinant protein alongside a media only-control at 37°C for 24 hours. The integrity of the protein was analysed by SDS-PAGE alongside a non-incubated sample of recombinant protein as a loading control.

Protein Sequence Alignment

TasA orthologues were identified by BLASTP (82, 83) using the protein sequence of TasA from *B. subtilis* as the query. TasA protein sequences were aligned using Clustal Omega with the default settings (51). The aligned sequences were imported and manually coloured for homology as

indicated in the legend in Microsoft Word. The signal sequences were predicted using the SignalP v4.1 server and are indicated by underline (42). A maximum likelihood tree was calculated from the Clustal Omega alignment using the phylogeny.fr platform (84), Gblocks was used to eliminate divergent and poorly aligned segments for tree construction (85). The tree was estimated using the PhyML algorithm (86) with mid-point rooting, using a WAG substitution model (87) and bootstrapping procedure set to 100 replicates. The outputted tree was visualised using TreeDyn (88).

Protein Precipitation

A strain carrying an IPTG inducible copy of the *tasA* gene (NRS5313) in a Δ*tasA* background was grown in 200 ml Msgg at 37°C until OD_{600} of 1 in the presence of 1 mM IPTG. The culture supernatant was collected and separated from cell fraction by centrifugation at 5000 *g* for 30 minutes at 4°C with iterative removal of the supernatant into a fresh tube for 4 rounds. 40 ml of the clarified supernatant was precipitated overnight with 6.25% (w/v) trichloroacetic acid (Sigma™) at 4°C and the precipitated proteins were recovered by centrifugation as before. The protein pellet was washed 5 times with 1 ml ice-cold dH_20 and air dried for 1 hour (protocol modified from Cianfanelli *et al*. 2016). The protein pellet was suspended in 50 µL 2x laemmli buffer and separated by SDS-PAGE on a 14% gel alongside *in vitro* purified fTasA protein as a size marker. The section of the lane at the position expected to contain mature TasA was excised and analysed by mass spectrometry.

Mass Spectrometry

Samples were processed prior to overnight (16 h) trypsin digestion (Modified Sequencing Grade, Pierce). Peptides extracted from gel and dried in SpeedVac (Thermo Scientific™). Peptides re-suspended 50 µl 1% formic acid, centrifuged and transferred to HPLC vial. 15 µl of this was typically analysed on the system. The peptides from each fraction were separated using a mix of buffer A (0.1% formic acid in MS grade water) and B (0.08% formic acid in 80% MS grade CH_3CN). The peptides from each fraction were eluted from the column using a flow rate of 300 nl/min and a linear gradient from 5% to 40% buffer B over 68 min. The column temperature was set at 50 °C. The Q Exactive HF Hybrid Quadrupole-Orbitrap Mass Spectrometer was operated in data dependent mode with a single MS survey scan from 335-1,800 m/z followed by 20 sequential *m/z* dependent MS2 scans. The 20 most intense precursor ions were sequentially fragmented by higher energy collision dissociation (HCD). The MS1 isolation window was set to 2.0 Da and the resolution set at 60,000. MS2 resolution was set at 15,000. The AGC targets for MS1 and MS2 were set at $3e^6$ ions and $5e^5$ ions, respectively. The normalized collision energy was set at 27%. The maximum ion injection times for MS1 and MS2 were set at 50 ms and 100 ms, respectively. Exactive HF Hybrid Quadropole .RAW data files were extracted and converted to mascot generic files (.mgf) using MSC

Convert. Extracted data then searched against the Local peptide database containing the relevant TasA sequence using the Mascot Search Engine (Mascot Daemon Version 2.3.2).

Fibre formation and X-ray Diffraction

To prepare samples for X-ray diffraction, 5 µL of recombinant at ~ 5 mg/ml fTasA or heat treated fTasA (60°C for 24 hours) was suspended between 2 borosilicate glass capillaries (Harvard Apparatus) and allowed to dry (Makin & Serpell, 2005). The dried fibres were mounted onto a Rigaku M007HF X-ray generator equipped with a Saturn 944HG+ CCD detector, and images collected with 60s exposures at room temperature. Diffraction patterns were inspected using Ipmosflm CCP4} and then converted to TIFF format. CLEARER (91) was used to measure the diffraction signal positions.

Circular Dichroism (CD) Spectroscopy

All CD measurements were performed using a Jasco J-810 spectropolarimeter. Solution-state samples were measured at a protein concentration of 0.2 mg/ml (in 25 mM phosphate buffer) in a 0.1 cm quartz cuvette. A scan rate of 50 nm/s was used, with a data pitch of 0.1 nm and a digital integration time of 1 s. Twenty scans were accumulated and averaged to produce the final curve.

Transmission Electron Microscopy (TEM) imaging

A 5 µl droplet of 0.02 mg/ml protein solution was pipetted onto a carbon-coated copper grid (TAAB Laboratories Equipment Ltd) and left for 4 minutes before being wicked away from the side with filter paper. Subsequently, a 5 µl droplet of 2% (w/v) uranyl acetate was pipetted onto the grid and left for 3 minutes before being wicked away from the side with filter paper. The stained grids were imaged using a Philips/FEI CM120 BioTwin transmission electron microscope and ImageJ software was used for image analysis.

Thioflavin T binding kinetics

Protein samples were diluted to 3 mg/ml in 25 mM phosphate buffer at either pH3 or pH7. 200 µL of protein was added into the wells of a Corning 96-well plate. ThT was added to a final concentration of 20 µM. The plates were sealed with a transparent film and put into a BMG Fluorstar plate reader at 37°C, 45°C or 60°C as indicated. Measurements of ThT fluorescence were taken every 5 minutes for a period of 8 hours (60°C) or 24 hours (37°C and 45°C).

Mean square displacement via bead tracking

A 1 μL aliquot of carboxylate-modified polystyrene, fluorescent yellow-green latex beads with a mean particle size of 1 μm (Sigma-Aldrich, L4655) was diluted into 1 mL of phosphate buffer. 5 μL of this stock solution was added to the protein solution and gently mixed to disperse the particles. 80 μL of the bead and protein solution was placed on a cavity slide (Brand GmBH, 0.6-0.8 mm depth) and sealed with a coverslip using nail varnish. Movies of the motion of the particles were taken using a Nikon Eclipe Ti microscope equipped with a Hamamatsu Orca-Flash 4.0 CCD camera. Images were acquired using μ-manager software at a framerate of 10 fps (92). Movies were then analysed using TrackPy (available from github.com/soft-matter/trackpy).

Congo-Red Binding Assay

A stock solution of 2 mg/ml Congo Red (Sigma-Aldrich 75768) was prepared in phosphate buffer and filtered three times using a 0.22 μm syringe filter (Millipore). 2 mg/ml bovine insulin (Sigma-Aldrich I5500) was prepared in MilliQ water adjusted to pH 1.6 using concentrated HCL. The insulin sample and 2 mg/ml of fTasA (pH 7 in phosphate buffer) were incubated overnight at 60°C. Additionally, a 2 mg/ml sample of fTasa that had been kept refrigerated at 4°C was also prepared. 60 μL of each protein sample was added to a cuvette containing 1 mL of buffer and 10 μL of the Congo Red stock solution. The samples were then allowed to incubate at room temperature for 30 minutes. A control spectrum containing only Congo Red was measured where 10 μL of the Congo Red stock solution was added to 1 mL of buffer plus an additional 60 μL of buffer (to match the amount of protein added to each cuvette). Spectra were acquired using a Cary 1E spectrophotometer over a wavelength range of 400-700 nm.

Acknowledgments

Work was supported by the Biotechnology and Biological Sciences Research Council [BB/L006804/1; BB/L006979/1; BB/M013774/1; BB/N022254/1]. EE and CE are supported by the Wellcome Institutional Strategic Support Fund (Award no. 097818/Z/11). We would like to acknowledge the FingerPrints Proteomics Facility and the X-ray Crystallography Facility at the University of Dundee (which is supported by Wellcome (Award no. 094090)). We are grateful to Ms. Ho, Dr. Hobley and Dr. Ostrowski for contributing plasmids and a strain used in this study.

References

1. Costerton JW, et al. (1987) Bacterial Biofilms in Nature and Disease. *Annu Rev Microbiol* 41(1):435–464.

2. Hobley L, Harkins C, MacPhee CE, Stanley-Wall NR (2015) Giving structure to the biofilm matrix: an overview of individual strategies and emerging common themes. *FEMS Microbiol Rev*, 015.

3. Flemming H-C, Wingender J (2010) The biofilm matrix. *Nat Rev Microbiol* 8(9):623.

4. Dragoš A, Kovács ÁT (2017) The Peculiar Functions of the Bacterial Extracellular Matrix. *Trends Microbiol* 25(4):257–266.

5. Bais HP, Fall R, Vivanco JM (2004) Biocontrol of *Bacillus subtilis* against infection of Arabidopsis roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production. *Plant Physiol* 134(1):307–19.

6. Chen Y, et al. (2012) A *Bacillus subtilis* sensor kinase involved in triggering biofilm formation on the roots of tomato plants. *Mol Microbiol* 85(3):418–30.

7. Chen Y, et al. (2013) Biocontrol of tomato wilt disease by *Bacillus subtilis* isolates from natural environments depends on conserved genes mediating biofilm formation. *Environ Microbiol* 15(3):848–64.

8. Cairns LS, Hobley L, Stanley-Wall NR (2014) Biofilm formation by *Bacillus subtilis*: new insights into regulatory strategies and assembly mechanisms. *Mol Microbiol* 93(4):587–98.

9. Elsholz AKW, Wacker SA, Losick R (2014) Self-regulation of exopolysaccharide production in *Bacillus subtilis* by a tyrosine kinase. *Genes Dev* 28(15):1710–1720.

10. Seminara A, et al. (2012) Osmotic spreading of *Bacillus subtilis* biofilms driven by an extracellular matrix. *Proc Natl Acad Sci* 109(4):1116–1121.

11. Chai Y, Beauregard PB, Vlamakis H, Losick R, Kolter R (2012) Galactose Metabolism Plays a Crucial Role in Biofilm Formation by *Bacillus subtilis*. *MBio* 3(4):e00184-12-e00184-12.

12. Roux D, et al. (2015) Identification of Poly- *N*- acetylglucosamine as a Major Polysaccharide Component of the *Bacillus subtilis* Biofilm Matrix. *J Biol Chem* 290(31):19261–19272.

13. Jones SE, Paynich ML, Kearns DB, Knight KL (2014) Protection from Intestinal Inflammation by

Bacterial Exopolysaccharides. *J Immunol* 192(10):4813–4820.

14. Hobley L, et al. (2013) BslA is a self-assembling bacterial hydrophobin that coats the *Bacillus subtilis* biofilm. *Proc Natl Acad Sci USA* 110(33):13600–5.

15. Bromley KM, et al. (2015) Interfacial self-assembly of a bacterial hydrophobin. *Proc Natl Acad Sci USA* 112(17):5419–24.

16. Arnaouteli S, et al. (2017) Bifunctionality of a biofilm matrix protein controlled by redox state. *Proc Natl Acad Sci USA* 114(30):E6184–E6191.

17. Branda SS, et al. (2004) Genes Involved in Formation of Structured Multicellular Communities by *Bacillus subtilis*. *J Bacteriol* 186(12):3970–3979.

18. Romero D, Vlamakis H, Losick R, Kolter R (2011) An accessory protein required for anchoring and assembly of amyloid fibres in *B. subtilis* biofilms. *Mol Microbiol* 80(5):1155–68.

19. Beauregard PB, Chai Y, Vlamakis H, Losick R, Kolter R (2013) *Bacillus subtilis* biofilm induction by plant polysaccharides. *Proc Natl Acad Sci USA* 110(17):E1621-30.

20. Michna RH, Zhu B, Mäder U, Stülke J (2016) SubtiWiki 2.0--an integrated database for the model organism *Bacillus subtilis*. *Nucleic Acids Res* 44(D1):D654-62.

21. Stöver AG, Driks A (1999) Secretion, localization, and antibacterial activity of TasA, a *Bacillus subtilis* spore-associated protein. *J Bacteriol* 181(5):1664–72.

22. Branda SS, Chu F, Kearns DB, Losick R, Kolter R (2006) A major protein component of the *Bacillus subtilis* biofilm matrix. *Mol Microbiol* 59(4):1229–38.

23. Romero D, Aguilar C, Losick R, Kolter R (2010) Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms. *Proc Natl Acad Sci USA* 107(5):2230–4.

24. Chapman MR, et al. (2002) Role of *Escherichia coli* Curli Operons in Directing Amyloid Fiber Formation. *Science (80)* 295(5556):851–855.

25. van Gestel J, Vlamakis H, Kolter R (2015) From Cell Differentiation to Cell Collectives: *Bacillus subtilis* Uses Division of Labor to Migrate. *PLOS Biol* 13(4):e1002141.

26. Serrano M, et al. (1999) A *Bacillus subtilis* secreted protein with a role in endospore coat assembly and function. *J Bacteriol* 181(12):3632–43.

27. Eisenberg D, Jucker M (2012) The amyloid state of proteins in human diseases. *Cell*

148(6):1188–203.

28. Hardy J, Selkoe DJ (2002) The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science (80)* 297(5580):353–356.

29. Sunde M, Blake C (1997) The Structure of Amyloid Fibrils by Electron Microscopy and X-Ray Diffraction. *Advances in Protein Chemistry Volume 50* (Elsevier), pp 123–159.

30. Sipe JD, et al. (2016) Amyloid fibril proteins and amyloidosis: chemical identification and clinical classification International Society of Amyloidosis 2016 Nomenclature Guidelines. *Amyloid* 23(4):209–213.

31. Baldwin AJ, et al. (2011) Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *J Am Chem Soc* 133(36):14160–14163.

32. Fowler DM, Koulov A V., Balch WE, Kelly JW (2007) Functional amyloid – from bacteria to humans. *Trends Biochem Sci* 32(5):217–224.

33. Uversky VN, Li J, Fink AL (2001) Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J Biol Chem* 276(14):10737–44.

34. Kayed R, et al. (1999) Conformational transitions of islet amyloid polypeptide (IAPP) in amyloid formation in Vitro. *J Mol Biol* 287(4):781–796.

35. Ferrão-Gonzales AD, Souto SO, Silva JL, Foguel D (2000) The preaggregated state of an amyloidogenic protein: hydrostatic pressure converts native transthyretin into the amyloidogenic state. *Proc Natl Acad Sci USA* 97(12):6445–50.

36. Hong D-P, Ahmad A, Fink AL (2006) Fibrillation of Human Insulin A and B Chains. *Biochemistry* 45(30):9342–9353.

37. Kalapothakis JMD, et al. (2015) A kinetic study of ovalbumin fibril formation: the importance of fragmentation and end-joining. *Biophys J* 108(9):2300–11.

38. Dobson CM (1999) Protein misfolding, evolution and disease. *Trends Biochem Sci* 24(9):329–332.

39. MacPhee CE, Dobson CM (2000) Formation of Mixed Fibrils Demonstrates the Generic Nature and Potential Utility of Amyloid Nanostructures. doi:10.1021/JA0029580.

40. Chai L, et al. (2013) Isolation, characterization, and aggregation of a structured bacterial

matrix precursor. *J Biol Chem* 288(24):17559–68.

41. Morris RJ, et al. (2013) Mechanistic and environmental control of the prevalence and lifetime of amyloid oligomers. *Nat Commun* 4:1891.

42. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–786.

43. Micsonai A, et al. (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci* 112(24):E3095–E3103.

44. Makin OS, Serpell LC (2005) Structures for amyloid fibrils. *FEBS J* 272(23):5950–61.

45. Chu F, Kearns DB, Branda SS, Kolter R, Losick R (2006) Targets of the master regulator of biofilm formation in *Bacillus subtilis*. *Mol Microbiol* 59(4):1216–1228.

46. Romero D, Vlamakis H, Losick R, Koltera R (2014) Functional analysis of the accessory protein TapA in *Bacillus subtilis* amyloid fiber assembly. *J Bacteriol* 196(8):1505–1513.

47. Fitzpatrick AWP, et al. (2013) Atomic structure and hierarchical assembly of a cross-β amyloid fibril. *Proc Natl Acad Sci USA* 110(14):5468–73.

48. Terra R, Stanley-Wall NR, Cao G, Lazazzera BA (2012) Identification of *Bacillus subtilis* SipW as a bifunctional signal peptidase that controls surface-adhered biofilm formation. *J Bacteriol* 194(11):2781–90.

49. Groenning M (2010) Binding mode of Thioflavin T and other molecular probes in the context of amyloid fibrils-current status. *J Chem Biol* 3(1):1–18.

50. Yang M, Dutta C, Tiwari A (2015) Disulfide-Bond Scrambling Promotes Amorphous Aggregates in Lysozyme and Bovine Serum Albumin. *J Phys Chem B* 119(10):3969–3981.

51. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.

52. Caro-Astorga J, Pacrez-Garcia A, de Vicente A, Romero D (2015) A genomic region involved in the formation of adhesin fibers in Bacillus cereus biofilms. *Front Microbiol* 5(January):1–11.

53. Freire S, de Araujo MH, Al-Soufi W, Novo M (2014) Photophysical study of Thioflavin T as fluorescence marker of amyloid fibrils. *Dye Pigment* 110:97–105.

54. Sen P, Fatima S, Ahmad B, Khan RH (2009) Interactions of thioflavin T with serum albumins:

Spectroscopic analyses. *Spectrochim Acta Part A Mol Biomol Spectrosc* 74(1):94–99.

55. De Ferrari G V, Mallender WD, Inestrosa NC, Rosenberry TL (2001) Thioflavin T is a fluorescent probe of the acetylcholinesterase peripheral site that reveals conformational interactions between the peripheral and acylation sites. *J Biol Chem* 276(26):23282–7.

56. Petrlova J, et al. (2017) Aggregation of thrombin-derived C-terminal fragments as a previously undisclosed host defense mechanism. *Proc Natl Acad Sci USA* 114(21):E4213–E4222.

57. Yamaguchi T, Yagi H, Goto Y, Matsuzaki K, Hoshino M (2010) A Disulfide-Linked Amyloid-β Peptide Dimer Forms a Protofibril-like Oligomer through a Distinct Pathway from Amyloid Fibril Formation. *Biochemistry* 49(33):7100–7107.

58. LeVine H (1999) Quantification of β-sheet amyloid fibril structures with thioflavin T. *Methods Enzymol* 309:274–284.

59. Benditt EP (1986) Amyloid Protein AA and its Precursor, The Acute Phase Protein(s) ApoSAA: A Perspective. *Amyloidosis*, eds Marrink J, M.H. van Rijswijk (Springer Netherlands, Dordrecht), pp 101–106. 2012th Ed.

60. Evans ML, Chapman MR (2014) Curli biogenesis: order out of disorder. *Biochim Biophys Acta* 1843(8):1551–8.

61. Shewmaker F, et al. (2009) The Functional Curli Amyloid Is Not Based on In-register Parallel β-Sheet Structure. *J Biol Chem* 284(37):25065–25076.

62. Kajava A V., Steven AC (2006) β-Rolls, β-Helices, and Other β-Solenoid Proteins. *Advances in Protein Chemistry*, pp 55–96.

63. Margittai M, Langen R (2008) Fibrils with parallel in-register structure constitute a major class of amyloid fibrils: molecular insights from electron paramagnetic resonance spectroscopy. *Q Rev Biophys* 41(3–4):265.

64. Fitzpatrick AWP, et al. (2017) Cryo-EM structures of tau filaments from Alzheimer's disease. *Nat Publ Gr* 547. doi:10.1038/nature23002.

65. Chai L, et al. (2013) Isolation, characterization, and aggregation of a structured bacterial matrix precursor. *J Biol Chem* 288(24):17559–68.

66. Epstein EA, Chapman MR (2008) Polymerizing the fibre between bacteria and host cells: the biogenesis of functional amyloid fibres. *Cell Microbiol* 10(7):1413–20.

67. Vlamakis H, Chai Y, Beauregard P, Losick R, Kolter R (2013) Sticking together: building a biofilm the *Bacillus subtilis* way. *Nat Rev Microbiol* 11(3):157–68.

68. Vlamakis H, Aguilar C, Losick R, Kolter R (2008) Control of cell fate by the formation of an architecturally complex bacterial community. *Genes Dev* 22(7):945–53.

69. Serra DO, Richter AM, Hengge R (2013) Cellulose as an architectural element in spatially structured *Escherichia coli* biofilms. *J Bacteriol* 195(24):5540–54.

70. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41(1):207–34.

71. Verhamme DT, Kiley TB, Stanley-Wall NR (2007) DegU co-ordinates multicellular behaviour exhibited by *Bacillus subtilis*. *Mol Microbiol* 65(2):554–68.

72. Patrick JE, Kearns DB (2008) MinJ (YvjD) is a topological determinant of cell division in *Bacillus subtilis*. *Mol Microbiol* 70(5):1166–79.

73. Arnaud M, Chastanet A, Débarbouillé M (2004) New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. *Appl Environ Microbiol* 70(11):6887–91.

74. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–20.

75. Bankevich A, et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455–77.

76. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M (2009) ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* 25(15):1968–9.

77. Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068–9.

78. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–9.

79. Koboldt DC, et al. (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 25(17):2283–5.

80. Schor M, Mey ASJS, Noé F, MacPhee CE (2015) Shedding Light on the Dock–Lock Mechanism

in Amyloid Fibril Growth Using Markov State Models. *J Phys Chem Lett* 6(6):1076–1081.

81. Branda SS, González-Pastor JE, Ben-Yehuda S, Losick R, Kolter R (2001) Fruiting body formation by *Bacillus subtilis*. *Proc Natl Acad Sci USA* 98(20):11621–6.

82. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.

83. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–402.

84. Dereeper A, et al. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36(Web Server):W465–W469.

85. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17(4):540–52.

86. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5):696–704.

87. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18(5):691–9.

88. Chevenet F, Brun C, Bañuls A-L, Jacq B, Christen R (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* 7(1):439.

89. Cianfanelli FR, et al. (2016) VgrG and PAAR Proteins Define Distinct Versions of a Functional Type VI Secretion System. *PLOS Pathog* 12(6):e1005735.

90. Makin OS, Serpell LC (2005) X-Ray Diffraction Studies of Amyloid Structure. *Amyloid Proteins* (Humana Press, New Jersey), pp 067–080.

91. Sumner Makin O, et al. (2007) *CLEARER* : a new tool for the analysis of X-ray fibre diffraction patterns and diffraction simulation from atomic structural models. *J Appl Crystallogr* 40(5):966–972.

92. Edelstein A, Amodaj N, Hoover K, Vale R, Stuurman N (2010) Computer control of microscopes using µManager. *Curr Protoc Mol Biol* Chapter 14, 14.20.

Figure 1: Recombinant TasA forms fibres. (**A-B**) Transmission electron microscopy images of recombinant fTasA stained with uranyl acetate shows the presence of fibres several microns in length and approximately 15 nm wide. At a higher magnification a repeat unit (app. 4nm) can be seen (B). (**C**) Solution state circular dichroism spectra of recombinant fTasA (Dotted black line) and mTasA (Solid grey line). (**D**) X-Ray Diffraction of recombinant fTasA protein fibres with exposure for 60 seconds where meridional and equatorial diffraction signals indicated in black and beige respectively.

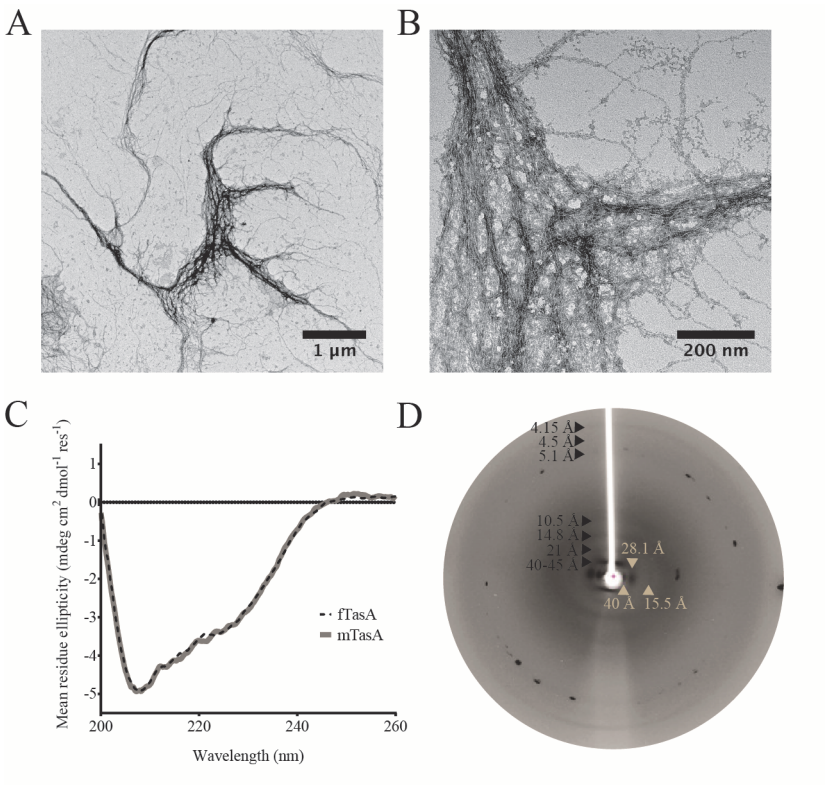


Figure 2: Recombinant fTasA is biologically active. (**A**) Biofilm phenotypes of wild type (NCIB3610), Δ*tasA* (NRS5267), Δ*sinIR* (NRS2012) and Δ*tasA sinR* (NRS5248) strains. (**B**) Biofilm phenotype of Δ*tasA* strain with the *ex vivo* addition of 10 µg recombinant fTasA or mTasA as indicated. **(C)** Immunoblot blot analysis of biofilm lysate collected from biofilms challenged with α-TasA antibody. **(D)** Integrity of 30 µg fTasA or mTasA pre-incubated for 24 hrs at 4°C or 60°C and subsequently incubated with cell-free NCIB3610 culture supernatant plus (+htSN) and minus (+SN) heat inactivation at 100°C. **(E)** Biofilm phenotype of Δ*tasA sinR* strain with the *ex vivo* addition of 10 µg recombinant fTasA or mTasA as indicated.

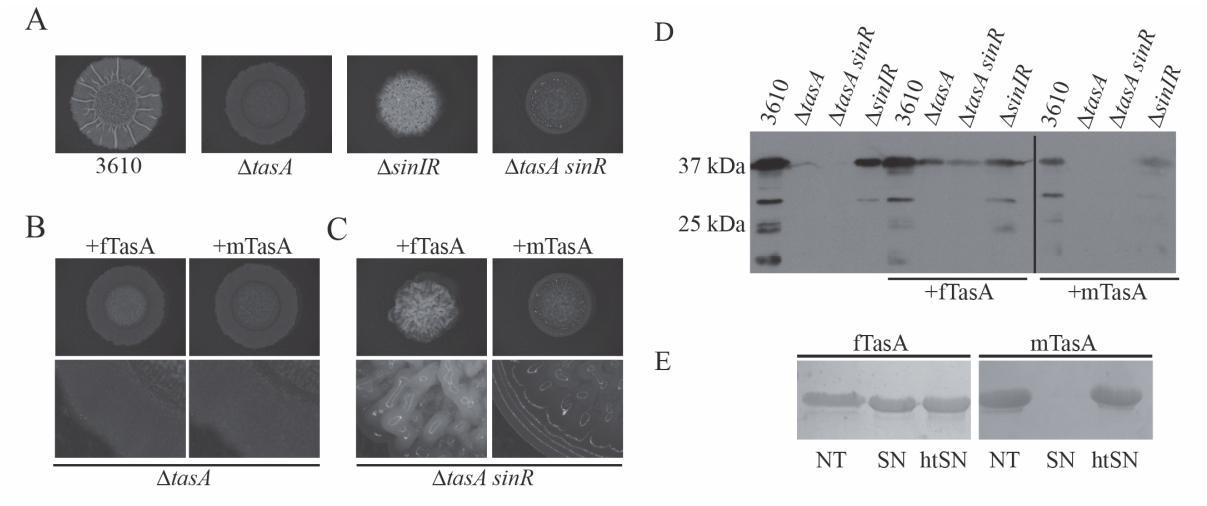


Figure 3: Recombinant fTasA does not need TapA but required the biofilm exopolysaccharide for activity (A) Biofilm phenotype of Δ*tapA* Δ*tasA sinR* (NRS5749) mutant upon addition of 10 µg fTasA *ex vivo.* **(B)** Biofilm phenotype of Δ*tasA sinR* (NRS5248) upon addition of transthyretin (TTR). **(C)** Biofilm phenotype of Δ*tasA sinR* P_{IPTG}-*epsA-O* (NRS5421) strain in the presence of 100 µM IPTG in absence and presence of *ex vivo* addition of 10 µg fTasA. An enlarged section of bottom left corner of the biofilm is shown.

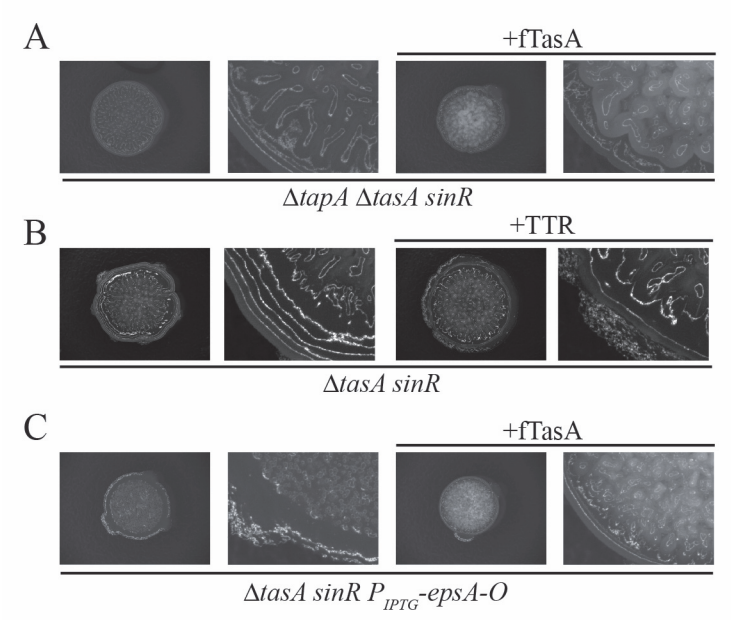


Figure 4: TasA undergoes a structural conversion at elevated temperatures. (A) Fluorescence intensity increase resulting from increased ThT binding upon incubation of recombinant fTasA (solid black line) and mTasA (dashed line) at 60°C, shown alongside buffer-only control (grey solid line). ThT fluorescence indicates mTasA converts faster than fTasA**. (B)** Circular dichroism spectra of recombinant mTasA (solid grey line) and fTasA (solid black line) upon incubation at 60°C for respectively 30 or 120 minutes show that both proteins undergo significant conformational changes. The CD spectrum for non-incubated fTasA (dotted line; also Fig 1B) is shown for reference. **(C)** TEM images of mTasA (left) and fTasA (right) incubated at 60°C for 8 hours show the presence of worm-like fibres.

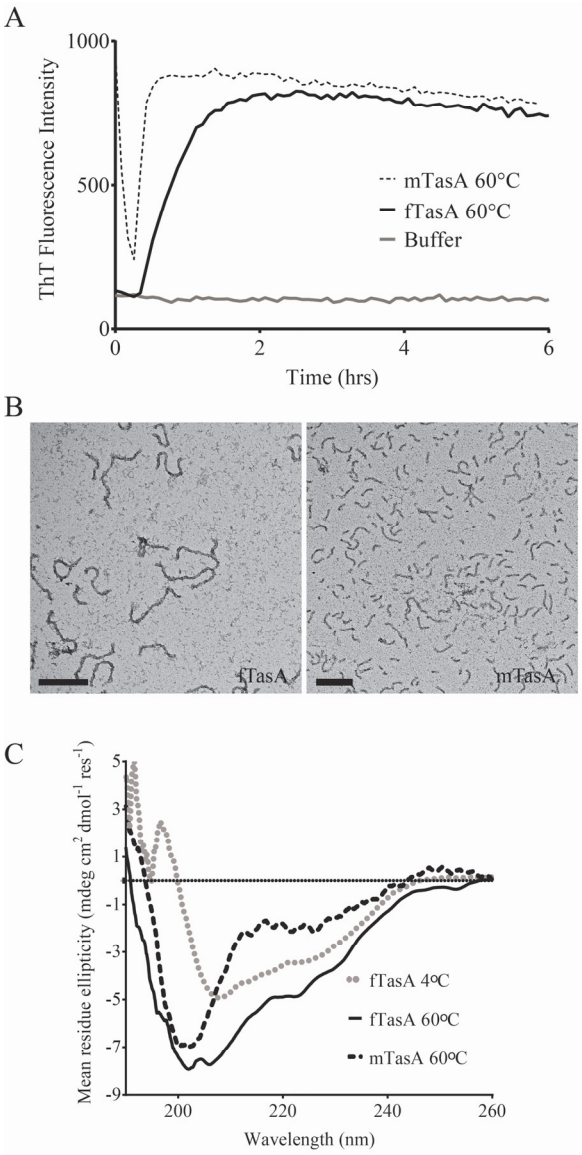


Figure 5: **Heat conversion impairs biological function.** (**A**) SDS-PAGE analysis of 10 µg recombinant protein incubated at 4°C or 60°C as indicated for 24 hours. (**B**) Biofilm phenotype of the Δ*tasA sinR* (NRS5248) strain with the exogenous addition of 10 µg protein from (A). The enlarged section shows detail of the bottom left of the biofilm. **(C)** Immunoblot blot analysis of biofilm lysate challenged with α-TasA antibody. **(D)** Integrity of 30 µg fTasA pre-incubated for 24 hrs at 4°C or 60°C subsequently incubated with cell-free NCIB3610 culture supernatant plus (+htSN) and minus (+SN) heat inactivation at 100°C with a fresh media (+media) as a control.

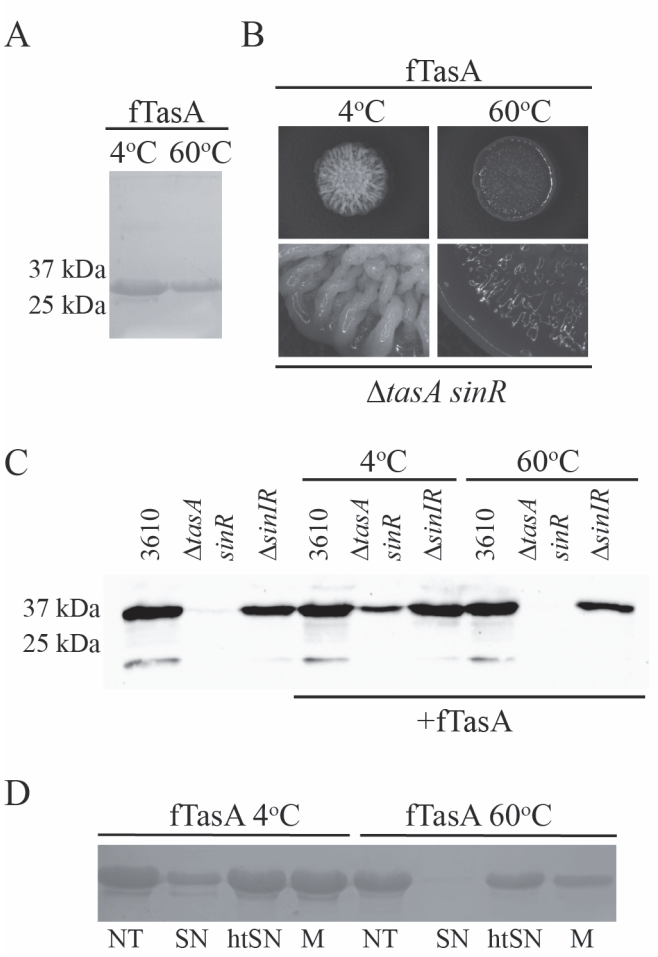


Figure 6: Complementation of *tasA sinR* null biofilm by recombinant orthologous TasA. (**A**) The phylogenetic tree was rooted using the midpoint method with the bootstrap value (red) given as a value between 0 and 1, where 1 is a high score. Highlighted are species chosen for subsequent analysis: *B. licheniformis* (orange), *B. amyloliquefaciens* (purple), *B. cereus* TasA (red) and CalY (green). For the protein sequence alignment and abbreviations of species names alongside accession numbers see Fig. S7. **(B)** Solution state circular dichroism spectra of recombinant *B. licheniformis*, *B. amyloliquefaciens*, *B. cereus* TasA and CalY. (**C**) Transmission electron microscopy images of recombinant orthologue TasA stained with uranyl acetate show the presence of fibres that are several micron in length and vary in width from 15 nm (*B. licheniformis*) to 25 nm (*B. amyloliquefaciens* and *B. cereus* CalY) and 60 nm (B. cereus TasA). A repeating unit at 4-5 nm is seen for all proteins. (**D**) Biofilm phenotypes shown for Δ*tasA sinR* (NRS5248) after co-culture with 10 µg recombinant *B. licheniformis* TasA, *B. amyloliquefaciens* TasA and *B. cereus* TasA and CalY fibres.

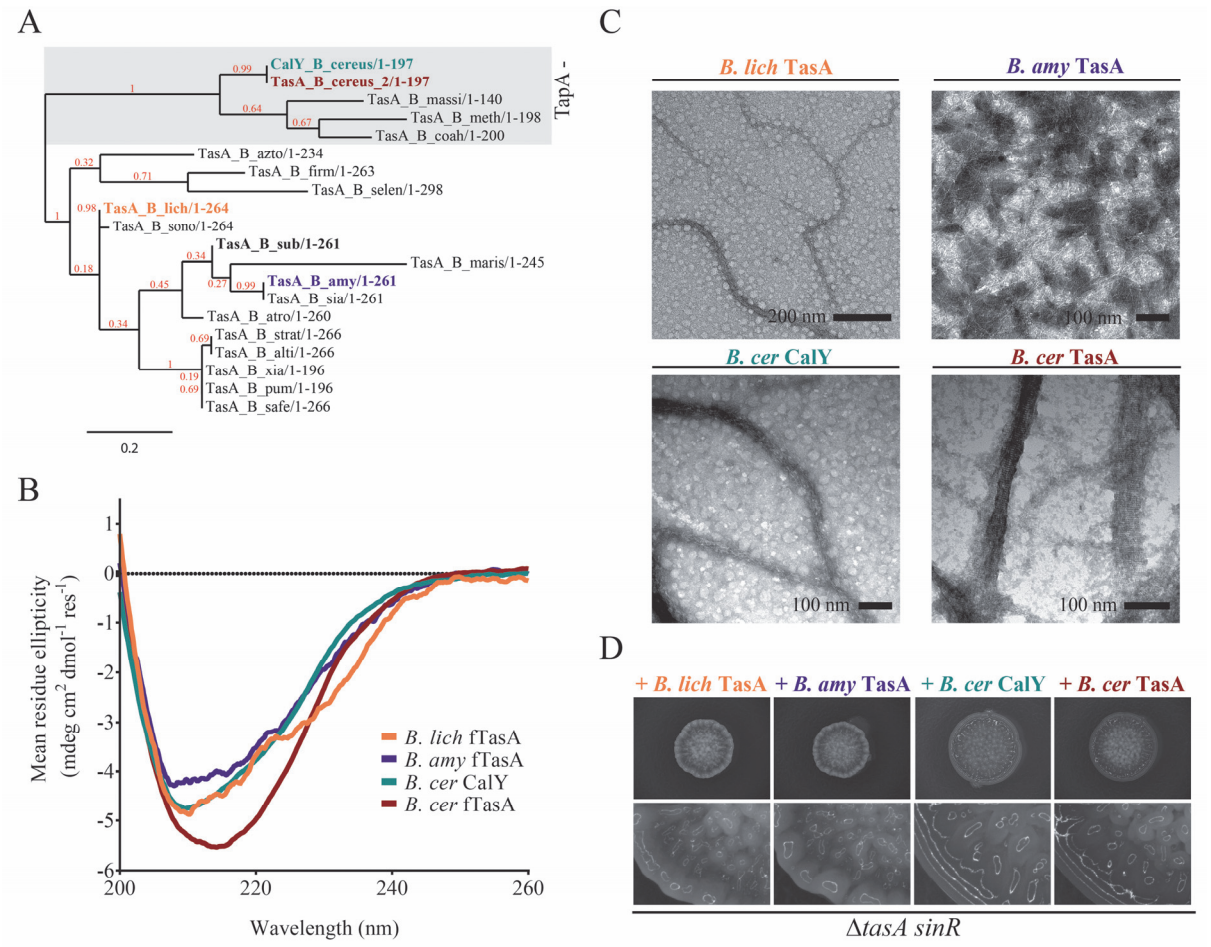


Table S1. *B. subtilis* and *E. coli* strains and plasmids used in this study

Strain	**Relevant genotype /Description**	**Source / Construction** [b,c]
MC1061	*E. coli F'lacIQ lacZM15 Tn10* (*tet*)	*E. coli* Genetic Stock Centre
BL21	F– *omp*T *hsd*SB(rB–, mB–) *gal dcm* (DE3)	(Studier and Moffat, 1986)
NCIB3610	Prototrophic wild-type strain	BGSC
168	*trpC2*	BGSC
NRS5048	168 pMiniMAD-Δ*tasA*	pNW1448 → 168
NRS5267	NCIB3610 Δ*tasA*	SPP1 NRS5048 → 3610
NRS5248	NCIB3610 Δ*tasA* $sinR_{(Phe65_Iso71dup)}$	SPP1 NRS5048 → 3610
NRS5255	NCIB3610 Δ*tasA* $sinR_{(Phe65_Iso71dup)}$ P_{IPTG}-*tasA* (spc)	pNW1432 → NRS5248
NRS1661	168 *epsA*::pBL584Φ(P*spachy*-*epsA*) (cml)	pBL584 → 168
NRS5421	NCIB3610 Δ*tasA* $sinR_{(Phe65_Iso71dup)}$ *epsA*::pBL584Φ(P*spachy*-*epsA*) (cml)	SPP1 NRS1661 → NRS5248
NRS3789	168 containing pNW685	pNW685 → 168
NRS3936	NCIB3610 Δ*tapA*	SPP1 NRS3789 → NRS 3610
NRS5748	NCIB3610 Δ*tapA ΔtasA*	SPP1 NRS3789 → NRS5267
NRS5749	NCIB3610 Δ*tapA ΔtasA sinR*	SPP1 NRS3789 → NRS5248
NRS5760	168 *amyE::*P_{IPTG}*-tapA-sipW-tasA-lacI (spc)*	pNW1804 → 168
NRS5762	NCIB3610 Δ*tapA ΔtasA*	SPP1 NRS5760 → NRS5748
NRS5763	NCIB3610 Δ*tapA ΔtasA* $sinR_{(Phe65_Iso71dup)}$ *amyE::*P_{IPTG}*-tapA-sipW-tasA-lacI (spc)*	SPP1 NRS5760 → NRS 5749
PY79 KO7	PY79 *nprE aprE epr mpr nprB vpr bpr*	BGSC
PY79 KO6	PY79 *nprE aprE epr mpr nprB vpr*	BGSC

1. Drug resistance cassettes are indicated as follows: *spc*, spectinomycin, *amp*: ampicillin, cml: chloramphenicol
2. BSGC represents the *Bacillus* genetic stock centre.
3. The direction of strain construction is indicated with plasmid DNA or phage (SPP1) (→) recipient strain.

Table S2 Plasmids used in this study

Plasmid	Description [a]	Source
pDR111	*B. subtilis* integration vector for IPTG-induced	Britton *et al*., 2002
pGEX-6P-1	Vector for overexpression of GST-fused proteins	GE Healthcare
pMiniMAD	Temp sensitive allelic replacement vector	Patrick & Kearns., 2008
pNW1448	pMinimad Δ*tasA*	This work
pNW1619	pDR110 P_{IPTG}-*sipW-tasA* (spc)[1]	This work
pNW1437	pDR110 P_{IPTG}-*tasA* (spc)[1]	This work
pNW685	pMAD Δ*tapA*	This work
pNW1432	pDR111 P_{IPTG}-*tasA* (spc)[1]	This work
pNW543	pGEX-6P-1- $tasA_{(28\text{-}261)BS}$ (amp)[1,2]	This work
pNW1437	pGEX-6P-1-TEV-$tasA_{(28\text{-}261)BS}$ (amp)[1,2]	This work
pNW1080	pGEX-6P-1-TEV-ser-$tasA_{(28\text{-}261)BS}$ (amp)[1,2]	This work
pNW1082	pGEX-6P-1-TEV-ala-$tasA_{(28\text{-}261)BS}$ (amp)[1,2]	This work
pNW1083	pGEX-6P-1-TEV-glu-$tasA_{(28\text{-}261)BS}$ (amp)[1,2]	This work
pNW1084	pGEX-6P-1-TEV-lys-$tasA_{(28\text{-}261)BS}$ (amp)[1,2]	This work
pNW1085	pGEX-6P-1-TEV-phe-$tasA_{(28\text{-}261)BS}$ (amp)[1,2]	This work
pNW1606	pGEX-6P-1-TEV-$tasA_{(28\text{-}264)BL}$ (amp)[1,2]	This work
pNW1608	pGEX-6P-1-TEV-$tasA_{(28\text{-}261)BA}$ (amp)[1,2]	This work
pNW1096	pGEX-6P-1-TEV-$tasA_{(28\text{-}197)BC1}$ (amp)[1,2]	This work
pNW1616	pGEX-6P-1-TEV-$tasA_{(28\text{-}197)BC2}$ (amp)[1,2]	This work
pBL584	*epsA*::pBL584Φ(P*spachy*-*epsA*)-*cat* (cml)[1]	(48)
pNW1804	pDR110- *tapA-sipW-tasA*	This work

1. Drug resistance cassettes are indicated as follows: *spc*, spectinomycin, *amp*: ampicillin, cml: chloramphenicol
2. Species abbreviations: BS (*B. subtilis*), BL (*B. licheniformis*), BA (*B. amyloliquefaciens*), BC1 (*B. cereus* TasA), BC2 (*B. cereus* CalY). The amino acids of the protein sequence encoded by the construct are detailed in brackets.

Table S3. Primers used in this study.

Primer	Sequence (5'-3')	Purpose	Plasmid
NSW2005	GCATGTCGACAGTTCTGTCAGGTTCGATGG	Upstream *tasA* region for deletion	pNW1448
NSW2006	GCATGGATCCACCCATGGTAAGCTCCCCTTTTA		
NSW2007	GCATGGATCCAATTAATAACAGCAAAAAAAAGAGACGGCCC	Downstream *tasA* region for deletion	
NSW2008	GCATGAATTCAAGTCTGCTCATCCTGGTCC		
NSW2100	*GAAAATTTATATTTTCAA***TCC**GCATTTAACGACATTAAA	Addition of serine to pNW1437	pNW1080
NSW2101	TTTAATGTCGTTAAATGC**GGA***TTGAAAATATAAATTTTC*		
NSW2110	*GAAAATTTATATTTTCAA***GCT**GCATTTAACGACATTAAA	Addition of Alanine to pNW1437	pNW1082
NSW2111	TTTAATGTCGTTAAATGC**AGC***TTGAAAATATAAATTTTC*		
NSW2112	*GAAAATTTATATTTTCAA***TTT**GCATTTAACGACATTAAA	Addition of Phenylalanine to pNW1437	pNW1085
NSW2113	TTTAATGTCGTTAAATGC**AAA***TTGAAAATATAAATTTTC*		
NSW2114	*GAAAATTTATATTTTCAA***AAA**GCATTTAACGACATTAAA	Addition of Lysine to pNW1437	pNW1084
NSW2115	TTTAATGTCGTTAAATGC**TTT***TTGAAAATATAAATTTTC*		
NSW2116	*GAAAATTTATATTTTCAA***GAA**GCATTTAACGACATTAAA	Addition of glutamic acid to pNW1437	pNW1082
NSW2117	TTTAATGTCGTTAAATGC**TTC***TTGAAAATATAAATTTTC*		
NSW2218	GCTAGTCGACCGGGGAAGAGGATGAAAAAAGCAATGAAG	*sipW-tasA* amplification (SalI/SphI)	pNW1619
NSW2219	GCTAGCATGCCTAGAGCTGTTATTAATTTTTATCCTCGCTATGC		
NSW1892	CAGGGGCCCCTGGGATCC*GAAAATTTATATTTTCAA*GCATTTAACGACATTAAA	Insertion of TEV protease site to pNW543	pNW1437
NSW1893	TTTAATGTCGTTAAATGC*TTGAAAATATAAATTTTC*GGATCCCAGGGGCCCCTG		
NSW660	GCATGGATCCGCATTTAACGACATTAAATCAA	*tasA* amplification (BamHI/XhoI)	pNW543
NSW661	GCATCTCGAGTTAATTTTTATCCTCGCTATGCGA		
NSW1308	GCATGGATCCCTCTTCCCATTTGGACATGTG	Upstream *tapA* region for deletion	pNW685
NSW1332	GGTAAGATATGTTTCGATTGGTCGACATGC		
NSW1333	GCATGTCGACCAGAAGGAAAGCGGGGAAGAG	Downstream *tapA* region for deletion	
NSW1334	GCATGAATTCATATCGAAACCTGTTGCCAGG		

Primer	Sequence (5'-3')	Purpose	Plasmid
NSW2126	CGAAAATTTATATTTTCAATTTTTCAGTGATAAAGAAGTG	Removal of serine from pNW1096	pNW1099
NSW2127	CACTTCTTTATCACTGAAAAA*TTGAAAATATAAATTTTC*G		
NSW1857	GCATGTCGACATAAAAGGGGAGCTTACCATGGGTATGAA	*tasA* gene amplification (SalI/SphI)	pNW1432
NSW1858	ATGCGCATGCTTATTAATTTTTATCCTCGCTATGC		
NSW1896	ATGCGTCGACTTTTACAGGAGGTAAGATATGTTTCG	Amplification of tapA operon: *tapA-sipW-tasA*	pNW1804
NSW2219	GCTAGCATGCCTAGAGCTGTTATTAATTTTTATCCTCGCTATGC		

* Underlined indicates insertion of a restriction site, the bold the insertion of codons for specified amino acids, and highlighted in italic underlined text represents the nucleotide sequence that encodes a TEV protease cleavage site.

Table S4 Single nucleotide polymorphisms identified by genomic sequencing

Genome co-ordinates	NCIB3610 & NRS5267	NRS5248	Mutation Type	Codon Substitution [a]	Locus Tag
490560	C	A			
490580	GT	GTT			
2552843	A	AGAAAC ATGAAAC CGAATAC	Frame shift + stop codon	gat/gaAACATGAAACCGAATACGAt	*sinR*
3874225	T	C			
4087227	A	T	None	ggA/ggT	*yxcE*
4087244	G	A	Missense	aGc/aAc	*yxcE*
4087248	A	T	None	ggA/ggT	*yxcE*
4087254	A	G	None	acA/acG	*yxcE*
4087260	T	A	None	ggT/ggA	*yxcE*
4087271	T	C	Missense	aTc/aCc	*yxcE*
4087272	C	A	None	atC/atA	*yxcE*
4087287	A	C	None	ccA/ccC	*yxcE*
4087317	C	A	None	ggC/ggA	*yxcE*
4087320	C	A	None	ccC/ccA	*yxcE*
4087332	T	C	None	tcT/tcC	*yxcE*
4087338	C	A	None	ggC/ggA	*yxcE*
4087347	C	T	None	ccC/ccT	*yxcE*
4087366	A	G	Missense	Atc/Gtc	*yxcE*
4087380	T	C	None	atT/atC	*yxcE*
4087491	C	T	None	gaC/gaT	*yxcD*

a. Uppercase font indicates nucleotide difference from NCIB3610

Figure SI 1: **Purification of recombinant fibre-forming TasA and monomeric TasA.** (**A**) Schematic of expression construct used for purification of TasA(28-261) (fTasA) using the plasmid backbone pGex-6-P-1 with TEV protease cleavage site. (**B**) SDS-PAGE analysis of 10 µg purified fTasA with molecular mass of 25731 kDa as calculated by lc-ms analysis **(C)** Mean square displacement (MSD) versus lag time for 2 mg/ml fTasA (solid line) and mTasA (dashed line). The slope of the MSD for mTasA scales as t^1 indicative of a viscous fluid medium. In contrast, the MSD slope for fTasA scales as $t^{1/2}$ which is indicative of a viscoelastic medium. **(D)** Sequence schematic of amino acid insertion, indicated by 'aa'. **(E)** Size exclusion chromatography (SEC) analysis (Superdex 200 10/300 GL) of N-terminal tagged TasA where the amino acids at N-terminus indicated (Ala, Glu, Phe, Lys). **(F)** SDS-PAGE analysis of 10µg purified serine tagged TasA (mTasA) with molecular mass of 25818 kDa as calculated by LC-MAS. **(G)** SEC of mTasA (Superdex 200 5/150 GL).

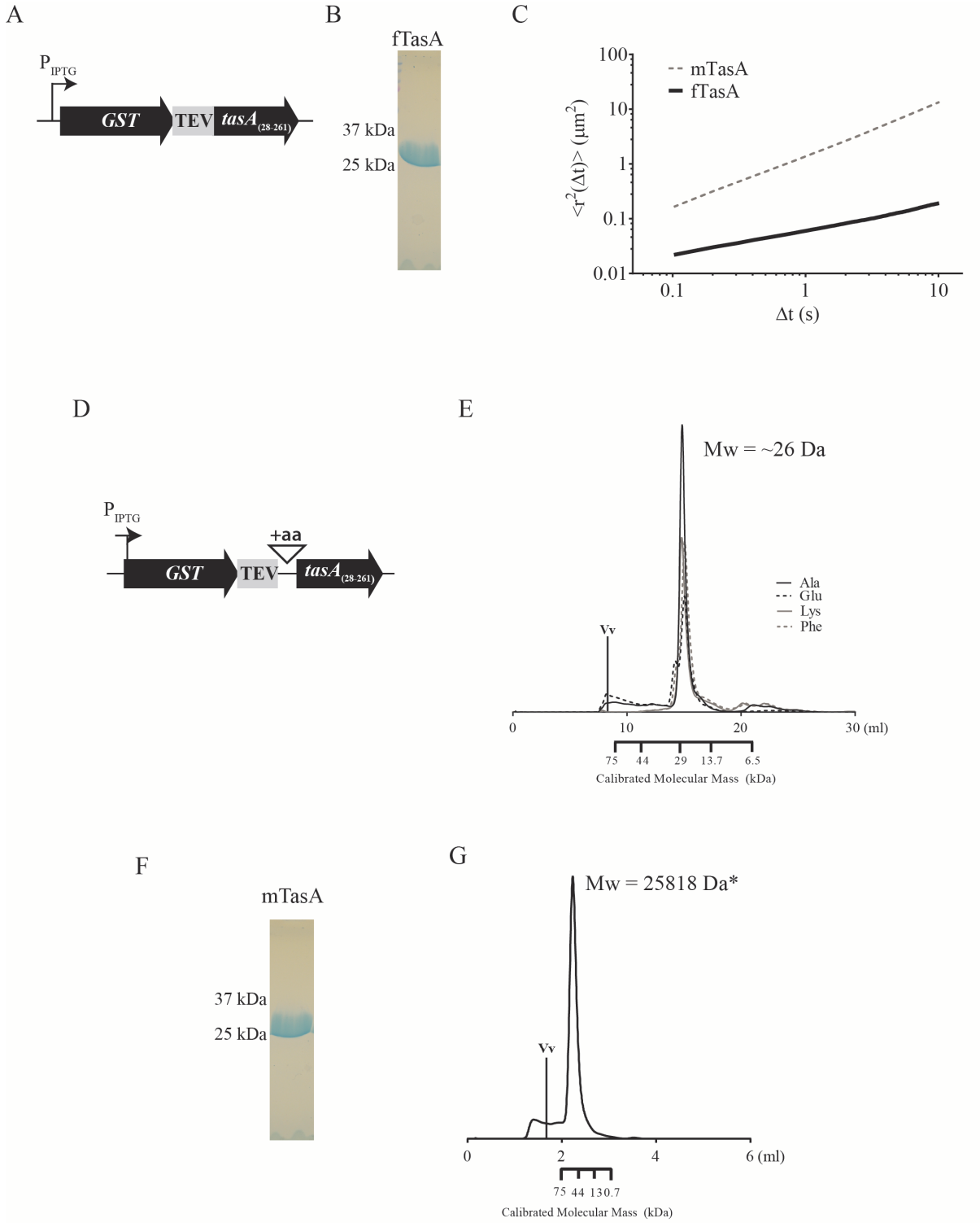

A
P IPTG
GST
TEV
tasA (28-261)
B
fTasA
37 kDa
25 kDa
C
mTasA
fTasA
<r²(Δt)> (μm²)
100
10
1
0.1
0.01
0.1
1
10
Δt (s)
D
P IPTG
+aa
GST
TEV
tasA (28-261)
E
Mw = ~26 Da
Ala
Glu
Lys
Phe
Vv
0
10
20
30 (ml)
75
44
29
13.7
6.5
Calibrated Molecular Mass (kDa)
F
mTasA
37 kDa
25 kDa
G
Mw = 25818 Da*
Vv
0
2
4
6 (ml)
75
44
13
0.7
Calibrated Molecular Mass (kDa)

Figure SI 2: Construction and genetic complementation of Δ*tasA* and Δ*tasA sinR*. (**A**) Sequence schematic of Δ*tasA* (NRS5267). **(B)** Sequence schematic of construct used to genetically complement Δ*tasA.* **(C)** Biofilm phenotype of complementation in presence and absence of 25 µM IPTG (NRS5313). **(D)** Sequence schematic of Δ*tasA sinR* with region of duplication in *sinR* as shown by whole-genomic sequencing (NRS5248). **(E)** Sequence schematic of construct used to genetically complement Δ*tasA sinR.* **(F)** Biofilm phenotype of complementation in presence and absence of 25 µM IPTG (NRS5255). **(G)** Sequence schematic of Δ*tapA* Δ*tasA sinR* (NRS5749). **(H)** Sequence schematic of whole operon complementation construct used to genetically complement Δ*tapA* Δ*tasA sinR*. **(I)** Biofilm phenotype of complementation in presence and absence of 25 µM IPTG (NRS5763).

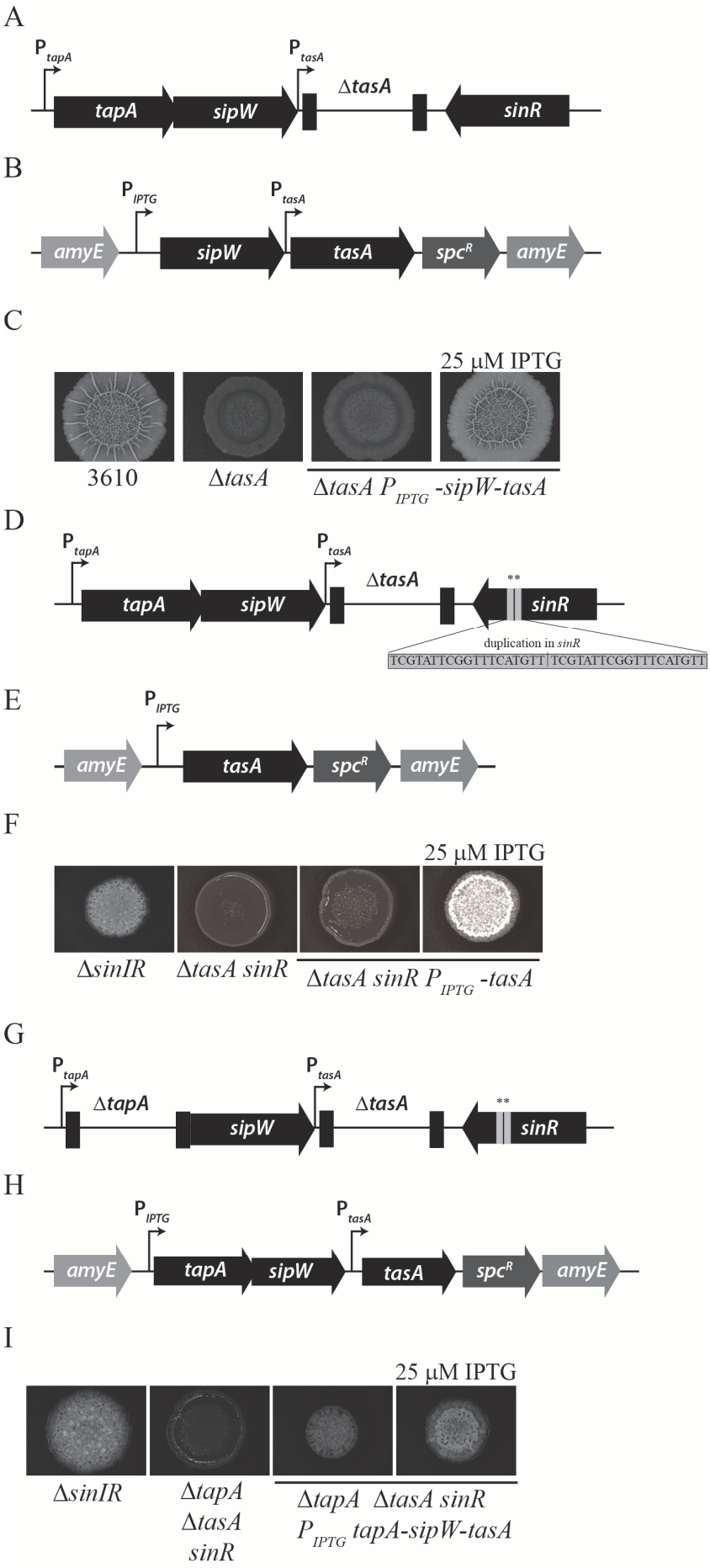

A
P tapA
P tasA
ΔtasA
tapA
sipW
sinR
B
P IPTG
P tasA
amyE
sipW
tasA
spcR
amyE
C
25 μM IPTG
3610
ΔtasA
ΔtasA P IPTG -sipW-tasA
D
P tapA
P tasA
ΔtasA
tapA
sipW
**
sinR
duplication in sinR
TCGTATTCGGTTTCATGTT|TCGTATTCGGTTTCATGTT
E
P IPTG
amyE
tasA
spcR
amyE
F
25 μM IPTG
ΔsinIR
ΔtasA sinR
ΔtasA sinR P IPTG -tasA
G
P tapA
P tasA
ΔtapA
sipW
ΔtasA
**
sinR
H
P IPTG
P tasA
amyE
tapA
sipW
tasA
spcR
amyE
I
25 μM IPTG
ΔsinIR
ΔtapA ΔtasA sinR
ΔtapA ΔtasA sinR
P IPTG tapA-sipW-tasA

Figure SI 3: Recombinant fTasA is biologically active. (A) Representative SDS-PAGE analysis of fTasA and mTasA, showing 10 µg protein used in (B). (**B**) Biofilm phenotypes of NCIB3610, *sinIR*, Δ*tasA sinR* and Δ*tasA* with the *ex vivo* addition of 10 µg purified protein as indicated. Images shown in Fig. 3B are repeated here. **(C-D)** Integrity of 30 µg fTasA **(C)** and mTasA **(D)** incubated for 24 hrs at 37°C analysed by SDS-PAGE. The protein was incubated with filtered spent supernatants collected from NCIB3610, plus and minus heat inactivation at 100°C; PY79 plus and minus heat inactivation at 100°C; PY79 KO6; PY79 KO7 and media only as a control.

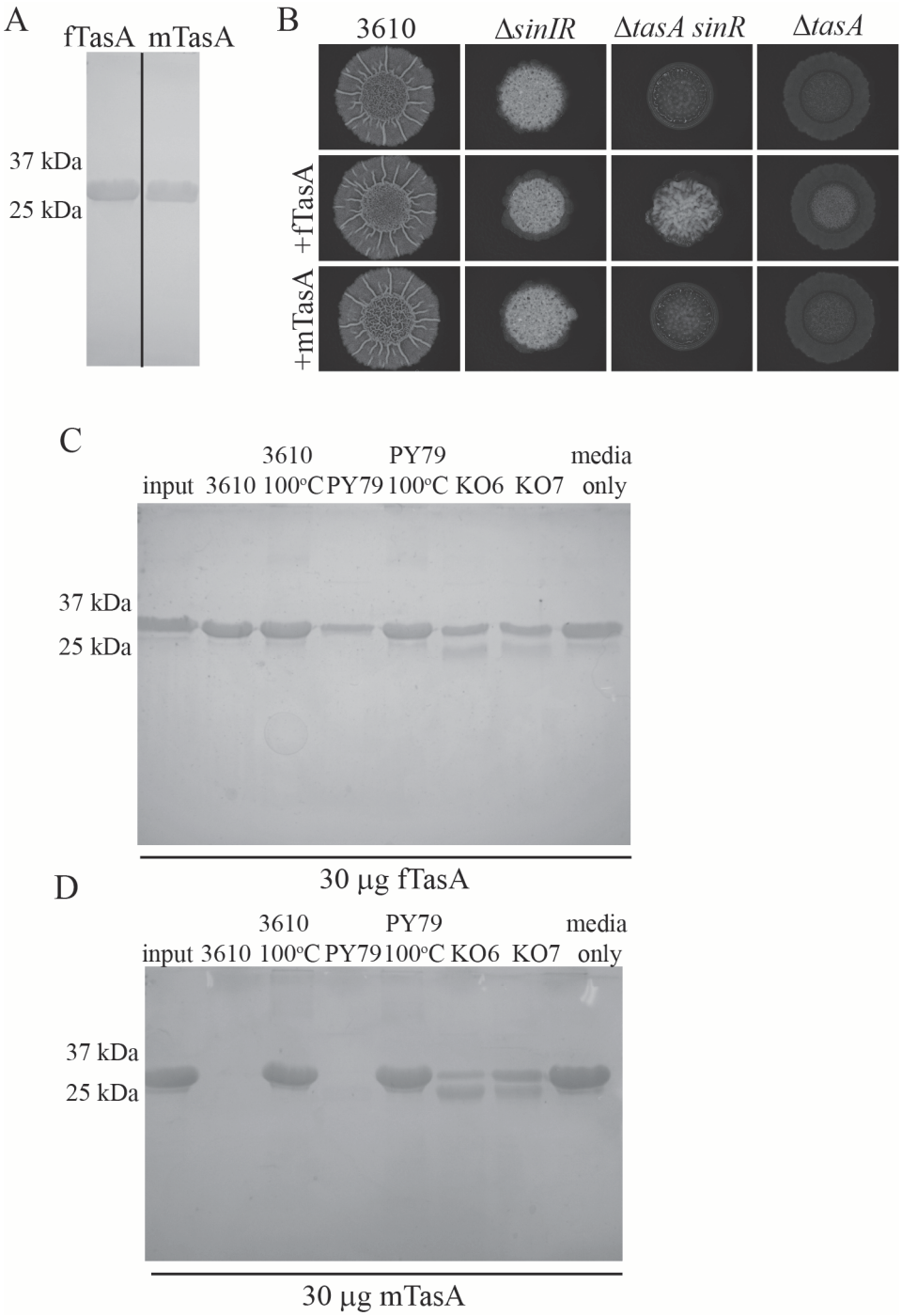


Figure SI 4: The biological activity of fTasA is independent of *tapA* but is dependent on specific matrix interaction. (**A**) Biofilm phenotypes of NCIB3610, Δ*tasA sinR* (NRS5248) and Δ*tapA* Δ*tasA sinR* (NRS5749) with the addition of 10 µg fTasA *ex vivo.* **(B)** Biofilm phenotype of Δ*tasA sinR* in presence and absence of the fibrous amyloid protein transthyretin (+TTR). **(C)** Biofilm phenotype of Δ*tasA sinR* P_{IPTG}-*eps* (NRS5421) in presence or absence of 100 µM IPTG and presence or absence of 10 µg fTasA where indicated by " –" and "+" respectively.

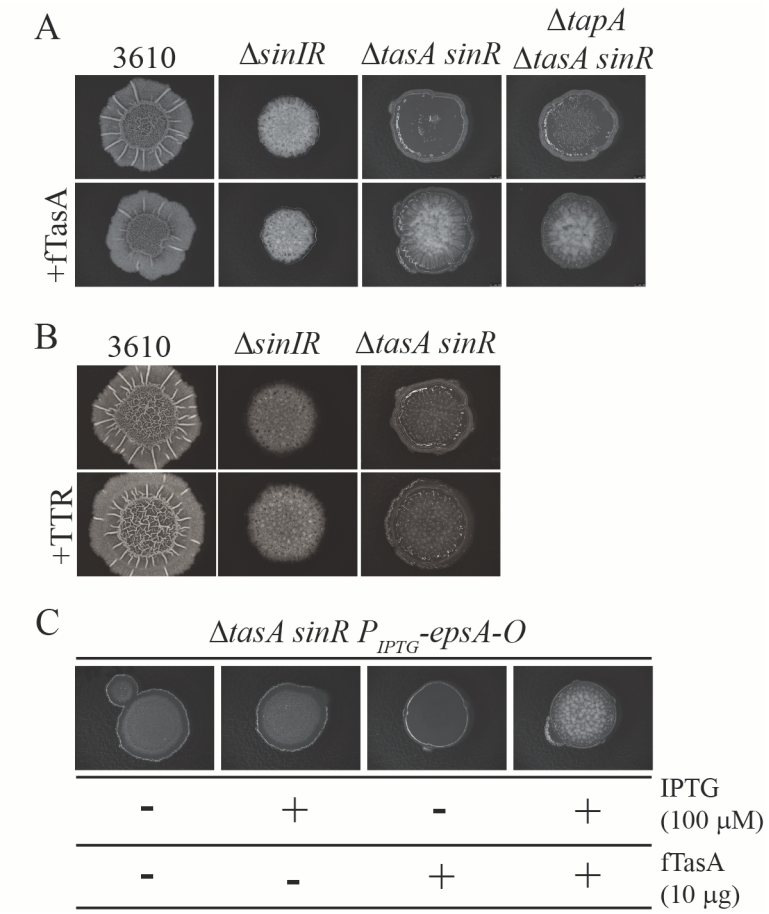


Fig SI 5: Heat treated fTasA has distinct properties compared to untreated fTasA. (A) Congo red absorbance spectra of untreated fTasA (red), 60°C incubated fTasA (green) with Congo red in buffer (dashed black) included as a negative control and bovine insuline incubated at pH 1.6 and 60°C as a positive control (blue). Both the insulin and heat treated fTasA spectra are red shifted compared to Congo red in buffer. **(B)** Difference between the protein spectra and the Congo red in buffer spectrum for untreated fTasA (red), heat-treated fTasA (green) and insulin (blue) indicate that both heat-treated fTasA and insulin bind Congo Red but untreated fTasA does not. **(C)** Normalised ThT fluorescence intensity for mTasA and fTasA incubated at pH7 and 60°C (light blue/light red), 45°C (green/orange) or 37°C (dark blue/dark red). **(D)** Normalised ThT fluorescence intensity of mTasA at 37°C and pH3 or pH7 (blue dashed/solid resp.), 45°C and pH3 or pH7 (green dashed/solid resp.), and fTasA at 37°C and pH3 or pH7 (red dashed/solid resp.), 45°C and pH3 or pH7 (orange dashed/solid resp.). **(E)** Mean square displacement (MSD) of 1 µm beads in a sample of 1 mg/ml fTasA that had been incubated at 60°C (dashed line) compared the MSD of the beads in a sample of fTasA that had been refrigerated at 4°C (solid line). Note, measurements were taken after samples had equilibrated to room temperature (22°C). The scaling of the MSD with time is comparable for both samples; however the fTasA that had not been heated is more viscous than the heated fTasA sample; (**F**) X-Ray diffraction of fTasA that had been pre-treated with 45°C heat and exposure for 60 seconds.

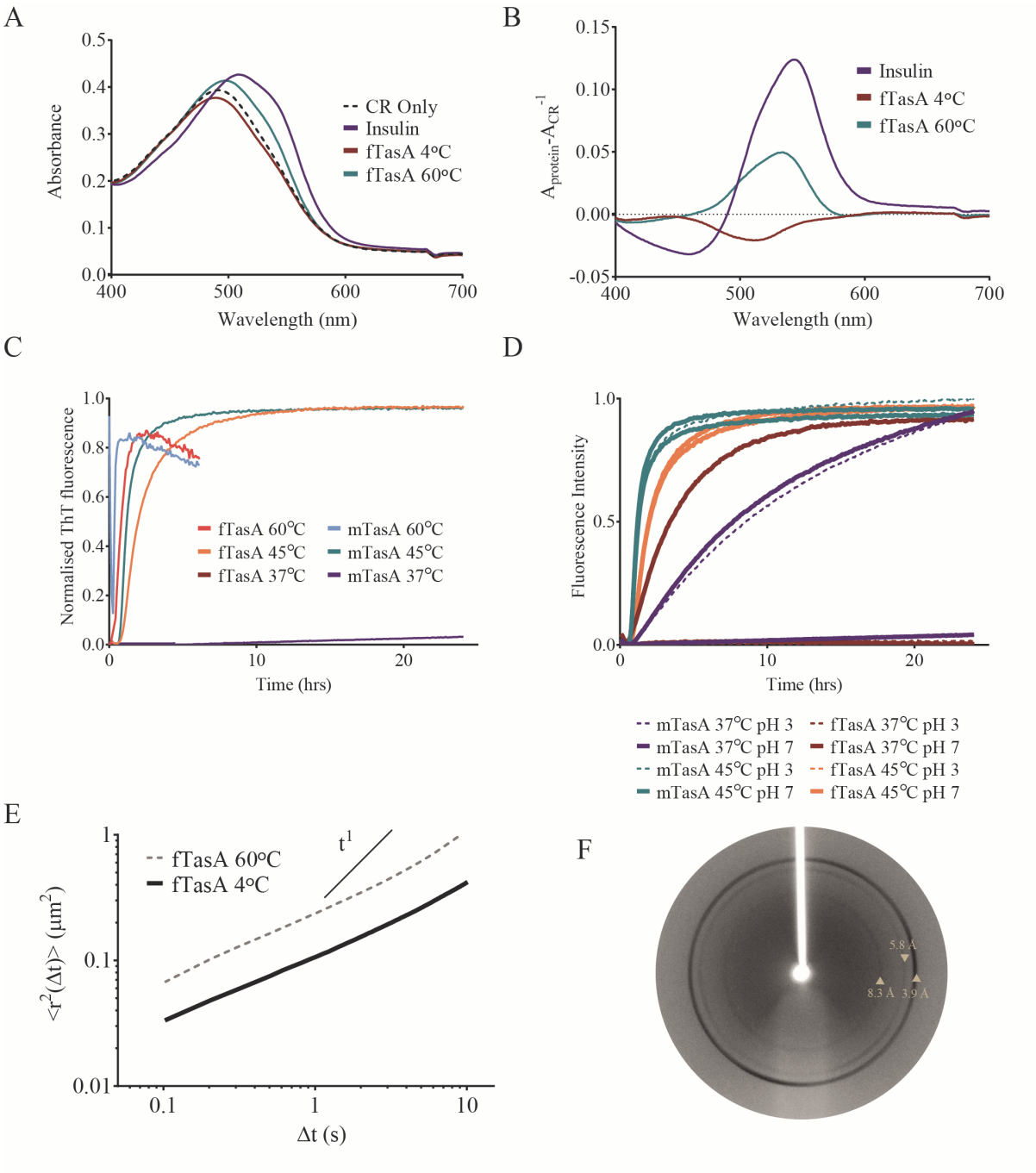
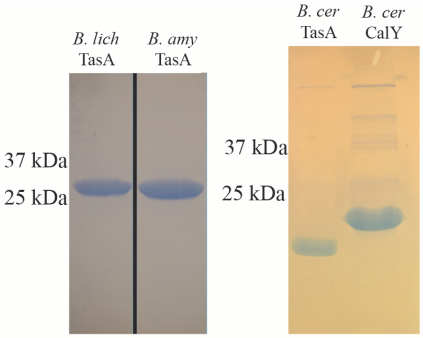

A
Absorbance
Wavelength (nm)
CR Only
Insulin
fTasA 4°C
fTasA 60°C
B
$A_{protein}$-A_{CR}^{-1}
Wavelength (nm)
Insulin
fTasA 4°C
fTasA 60°C
C
Normalised ThT fluorescence
Time (hrs)
fTasA 60°C
fTasA 45°C
fTasA 37°C
mTasA 60°C
mTasA 45°C
mTasA 37°C
D
Fluorescence Intensity
Time (hrs)
mTasA 37°C pH 3
mTasA 37°C pH 7
mTasA 45°C pH 3
mTasA 45°C pH 7
fTasA 37°C pH 3
fTasA 37°C pH 7
fTasA 45°C pH 3
fTasA 45°C pH 7
E
$<r^2(\Delta t)>$ (μm^2)
Δt (s)
fTasA 60°C
fTasA 4°C
t^1
F
5.8 Å
8.3 Å
3.9 Å

Fig SI 6: Alignment. Alignment of 20 TasA orthologues pulled from BlastP (82, 83) of *B. subtilis* TasA sequence (*Bacillus azotoformans* (B_azto) TapA (WP_003329523), TasA (WP_035196521.1). *Bacillus firmus* (B_firm) TapA (WP_082139007), TasA (WP_035326610.1). *Bacillus selenatarsenatis* (B_selen) TapA (WP_084135527), TasA (WP_041967097.1). *Bacillus licheniformis* (B_lich) TapA (WP_075747486), TasA (WP_043927876.1). *Bacillus sonorensis* (B_sono) TapA (WP_006637531), TasA (WP_006637529.1). *Bacillus marisflavi* (B_maris) TapA (WP_082139781), TasA (WP_048012492.1). *Bacillus amyloliquefaciens* (B_amy) TapA (WP_063094776), TasA (WP_044802563.1). *Bacillus siamensis* (B_sia) TapA (WP_029575370), TasA (WP_045926714.1). *Bacillus atrophaeus* (B_atro) TapA (WP_061670003), TasA (WP_010789194.1). *Bacillus stratosphericus* (B_strat) TapA (WP_039964022), TasA WP (007501219.1). *Bacillus altitudinis* (B_alti) TapA (WP_073413951), TasA (WP_039166017.1). *Bacillus xiamenensis* (B_xia) TapA (WP_034739525), TasA (WP_008360245.1). *Bacillus pumilis* (B_pum) TapA (WP_041106857), TasA (WP_041106853.1). *Bacillus safensis* (B_safe) TapA (WP_075623495), TasA (WP_034280222.1). *Bacillus methanolicus* (B_meth) TapA (None), TasA (WP_004434174.1). *Bacillus massiliosenegalensis* (B_massi) TapA (None), TasA (WP_019154446.1). *Bacillus coahuilensis* (B_coah) TapA (None), TasA (WP_010174983.1). *Bacillus cereus* (B_cer) TapA (None), TasA (WP_002201283.1)) by Clustal Omega (51). Signal sequence as predicted by SignalP v4.2 (42) underlined, 100% sequence identity highlighted in black.

```
TasA_B_massi  ------------------------------------------------------------ 000
CalY_B_cereus MSLKKKLGMGVASAALGLSLIGGGTFAFFSDKEVSNNTFAAGTLDLTLNP-----KTLVDIKDLK 060
TasA_B_cereus MSLKKKLGMGVASAALGLSLIGGGTFAFFSDKEVSNNTFAAGTLDLALNP-----KTIVDIKDLK 060
TasA_B_meth   MSIKKKLGLGIASAALGLSLVGGGTFAYFNDFAETSGTFASGTLDLNARP-----TTIINVDNIK 060
TasA_B_coah   MGIKQKLGLGVASAALGLSLVGGGTYAYFNDTETSTNTFAAGTLDLSVDP-----EVIINVDNIK 060
TasA_B_maris  MTIKKKLGLGVASAALGLSLIGGGTFAAFNDTATINNHFASGTLDLSVEKNGNNKPLNFDISNMK 065
TasA_B_selen  MSLKKKITMGALSATLGLSLVAGGTWAAFNDVETVSAGMEAGTLKLDLKK-YENKPFNFQISDLK 064
TasA_B_azto   MSIKKKLGLGVASAALGLSLIGGGTWAAFNDIETVGSSFAAGTLNLDVNN-GDGATALFDLSNLK 064
TasA_B_firm   MGIKKKIGLGVASAALGISLVGGGTWAAFNDVETLSASYAAGKLDLTAAD---TTSTGINLSNLK 062
TasA_B_pum    ------------------------------------------------------------ 000
TasA_B_safe   MAMKKSIRLGVLSGALGLALIGGGTWAAFNDIEKANAVYSTGELDLSAKE----NSGAINLANLK 061
TasA_B_xia    ------------------------------------------------------------ 000
TasA_B_strat  MAMKKSIRLGVLSGALGLALIGGGTWAAFNDVETANAVYSTGELNLSAKE----NSGAINLANIK 061
TasA_B_alti   MAMKKSIRLGVLSGALGLALIGGGTWAAFNDVETANAVYSTGELNLSAKE----NSGAINLANIK 061
TasA_B_lich   MGTKKKLGLGVASAALGLALVGGGTWAAFNDIETTQATYAAGTLDLNAKD----TSARVNLSNLK 061
TasA_B_sono   MGTKKKIGLGVASAALGLALVGGGTWAAFNDIETTQATYAAGTLDLNAKE----TSARVNLSNLK 061
TasA_B_sub    MGMKKKLSLGVASAALGLALVGGGTWAAFNDIKSKDATFASGTLDLSAKE----NSASVNLSNLK 061
TasA_B_amy    MGMKKKLSLGVASAALGLALVGGGTWAAFNDVKSTDATFASGTLDLSAKE----QSANVNLSNLK 061
TasA_B_sia    MGMKKKLSLGVASAALGLALVGGGTWAAFNDVKSTDATFASGTLDLSAKE----QSANVNLSNLK 061
TasA_B_atro   MGMKKKLSLGVASAALGLALVGGGTWAAFNDIKTTEATFASGTLDLSAKE----NSANVSLSNLK 061

TasA_B_massi  ----------LENNGTLDLSSIDLATSYSI---------------TDAAGDN-VDDFGKHIKVLF 039
CalY_B_cereus PGDSVKKEFLLKNNGTLAIKDVKLATKYSI---------------TDAKGDNAGEDFGKHIKVKF 110
TasA_B_cereus PGDSVKKEFLLQNNGTLAIKDVKLATKYSI---------------ADAKGDNAGEDFGKHIKVKF 110
TasA_B_meth   PGDKMLREFKLINDGSLDIAKVLLRTDYTV---------------KDAQNNNGNEDFGKHIKVNF 110
TasA_B_coah   PGDWMNRTFKLNNDGSLDISRVLLTTDYSV---------------TDAGGDN-INDMGEHIRVNF 109
TasA_B_maris  PGDSVQREFKLRNAGTLAIKDILLTVQAGGYS--------------NNESGATIDEFLSQFEVTL 116
TasA_B_selen  PGDKMTRNIKLVNSGTLAIRDVLMSIESVQFADYVPAEGEAGYEDDDTWGDNNVVQYLNQFRVTV 129
TasA_B_azto   PGDTMEREFLLSNAGSLAIKDVLMTTAITSFTQGE-----NEYVNIHGATDNSQEDFLNQFNVNI 124
TasA_B_firm   PGDVFQKEFELDNVGTLAIKDVLLKLTHSNFVDYV-----GGD-DTDIWGKNSAADFLRQFTITV 121
TasA_B_pum    ---------HFDNKGSLAINQVLMSLDYSQFT--------DGS-SAKNGGENTAEEFLSQFQVSV 047
TasA_B_safe   PGDRIKKEFHFDNKGSLAINQVLMSLDYSQFT--------DGS-SAKNGGKNTAEEFLSQFQVSV 117
TasA_B_xia    ---------NFENKGSLAINQVLMSLDYSQFK--------DGT-SAKNGGKNTAEEFLSQFQVSV 047
TasA_B_strat  PGDRIKKEFNFENKGSLAINQVLMSLDYSQFQ--------DGT-SAKNGGKNTAEEFLSQFQVSV 117
TasA_B_alti   PGDRIKKEFNFENKGSLAINQVLMSLDYSKFQ--------DGT-SAKNGGKNTAEEFLSQFQVSV 117
TasA_B_lich   PGDKFTKDFEFKNDGSLAIKEVLMQVGYSNFV--------DGN--AKNGGKSTAEDFLKQFKVSV 116
TasA_B_sono   PGDKFKKEFTFKNDGSLAIKEVLMQMGFSNFV--------DAN--AKNGGKNSAEDFLKQFKVSV 116
TasA_B_sub    PGDKLTKDFQFENNGSLAIKEVLMALNYGDFK--------A-----NGGSNTSPEDFLSQFEVTL 114
TasA_B_amy    PGDKLTKDFEFRNNGSLAIKEVLMALNFTDFK--------G-----AKKGNESAEDFLSQFEITV 114
TasA_B_sia    PGDKLTKDFEFRNNGSLAIKEVLMALNFTDFK--------G-----AKKGNESAEDFLSQFEITV 114
TasA_B_atro   PGDKLTKDFEFRNNGTLAIKEVLMALNYSGFN--------T-----AKGDNDSPEDFLSQFQITL 114

TasA_B_massi  LENADKTGDGWIIGDYNDIISE---TTLYDLQNMTPDAVE------------------------- 076
CalY_B_cereus LWNWDK---------QSEPVYE---TTLADLQNVDPDVLA------------------------- 138
TasA_B_cereus LWNWDK---------QSEPVYE---TTLADLQNVDPDVLA------------------------- 138
TasA_B_meth   LYNADK---------TDNVIYQ---TTLDQLKSMTPDAIE------------------------- 138
TasA_B_coah   LQNIDKSGI----VKPNNIIYS---TTLHDLKGTAPDAVQ------------------------- 142
TasA_B_maris  FKVNQSNDEY---DDFNSLVKTGQVLTLNDLVKGT----------------------L-----SEK 152
TasA_B_selen  MQVGAEGGSG---GFPRELIPADKDVRLGDFYLASGSLAGDDSKLNDGVTQADINTALSNVWGAV 191
TasA_B_azto   IDVDRQIT-----------VVQN--KTLKDLLTE------------------------------ 145
TasA_B_firm   ISSGTEGGN----GYSKDIVK-D--VNLFDFIEMTAGKGV---------------PP--NA-GSV 161
TasA_B_pum    LTVGAEGGN----GYPKNIILDH--ANLLDLHQLTSKQDQ---------------TAFEKL-RHA 090
TasA_B_safe   LTVGAEGGN----GYPKNIILDH--ANLLDLHQLTSKQDQ---------------TAFEKL-RHA 160
TasA_B_xia    LTVGAEGGN----GYPKNIILDH--ANLLDLHLLTSKQDQ---------------TAFE---RHA 088
TasA_B_strat  LTVGAEGGN----GYPKNIILDH--ANLLDLHLLTSKQDQ---------------TAFEKL-RHA 160
TasA_B_alti   LTVGAEGGN----GYPKNIILDH--ANLLDLHLLTSKQDQ---------------TAFEKL-RHA 160
TasA_B_lich   LTVGVEGGN----GYPKNIILDE--ANLYDLYNMSAKKDK---------------SAYEKV-KKA 159
TasA_B_sono   LTVGVEGGN----GYPKNIILDS--ANLYDLYNMSAKKDK---------------SSMEKI-KKL 159
TasA_B_sub    LTVGKEGGN----GYPKNIILDD--ANLKDLYLMSAKNDA---------------AAAEKI-KKQ 156
TasA_B_amy    LTVGKEGGN----GYPKNIILKA--ASLKDLYLMSTKQDK---------------AAAEAI-SKH 156
TasA_B_sia    LTVGKEGGN----GYPKNIILKA--ASLKDLYLMSTKQDK---------------AAAEAI-SKH 156
TasA_B_atro   LTVGKEGGN----GYPKNIILKA--ANLKDLYLMSAKQDK---------------TAAEKI-GQL 156

TasA_B_massi  ---------NLNSFLTWLLGLGGEDSGLPAGTSD--EMYVAFEEVDN-------GQDQNEFQGDS 123
CalY_B_cereus ---------KDI----FAPEWAE-NGGLAPNSED--YLWVQFVEEDD-------GKDQNKFQGDS 180
TasA_B_cereus ---------KDI----FAPEWGE-KGGLEAGTED--YLWVQFVEEDD-------GTDQNKFQGDS 180
TasA_B_meth   ---------NEV----FNKWLDEKGGKLAAGTSD--TLYVQFEEVDN-------GQDQNIFQGDS 181
TasA_B_coah   ---------EKI----WSFFGENS--GLKTGTSD--NFYVQFEEVDN-------QDDQNEFQGDS 183
TasA_B_maris  IKTDY-MSEGKINLAPIGIDNNP--KGLT--KTSANSVAMIVKEKEDDRKDSEGKYLQNKFMNNK 212
TasA_B_selen  DHTYIDAASRRINGATIN---PNEWTGLPVNPNDDDVVEISIEEVKDNTLNDKGTYVQNKYQGDT 252
TasA_B_azto   ----------TPNLAPI-NTNDPRYSGIPLNPADTENIRIQITEKEDPTVGINGEQDQNKYQGDS 199
TasA_B_firm   DSNYYDAESGRINVITKTGA-DQSYHGLPVNPDDKDKVLFTIKEEDIDDKDNKGRQIQNKYQGDS 225
TasA_B_pum    VDEKFLHESGKINVATVDGTVAPEYDGIPKNPFDYDKVEMIIEEINDQTKDKAGHYIQNKYQGDA 155
TasA_B_safe   VDEKFLHESGKINVATVDGTVAPEYDGIPKNPFDYDKVEMIIEEINDQTKDKAGHYIQNKYQGDA 225
TasA_B_xia    VDEKFLHESGKINVATVDGTAAPEYDGIPKNPFDYDKMEMIIEEVNDQTKDKAGHYIQNKYQGDA 153
TasA_B_strat  VDEKFLHESGKINVATVDGSTAPEYDGIPKNPFDYDKMEMIIEEVNDQTKDKAGHYIQNKYQGDA 225
TasA_B_alti   VDEKFLHESGKINVATVDGSTAPEYDGIPKNPFDYDKMEMIIEEVNDQTKDKAGNYIQNKYQGDA 225
TasA_B_lich   IEPAFLHDNGKINVATIDGKTAPEYDGIPKDPYDFDKVQLVIEEVNDKTTDASGRMVQNKYQGDS 224
TasA_B_sono   IEPGFLHDNGKINVATIKGKTAPQYDGIPKDPYDFDKVEMVIEEINDTTTDSNGNMIQNKYQGDS 224
TasA_B_sub    IDPKFLNASGKVNVATIDGKTAPEYDGVPKTPTDFDQVQMEIQEKDDKTKDEKGLMVQNKYQGNS 221
TasA_B_amy    IDPKFLSESGKVNVATINGKTAPEYDGVPKTPADYDQVRMEIKEKNDTAKTADGLSVQNKFQGNA 221
TasA_B_sia    IDPKFLSTSGKVNVATINGKTAPEYDGVPKTPVDYDQVRMEIQEKNDTAKGADGLSVQNKFQGNA 221
TasA_B_atro   IDSKFLSESGKINVATINGKTAPEYDGVPKTPVDYDQVQMVIEEKNDNAKDANGQMVQNKYQGNS 221

TasA_B_massi  LELEWTFTAHQTDGESK------------------------------ 140
CalY_B_cereus LNLEWTFNASQTDGEEK------------------------------ 197
TasA_B_cereus LNLEWTFNASQTDGEEK------------------------------ 197
TasA_B_meth   LELKWTFEGKQGAGQYK------------------------------ 198
TasA_B_coah   LSLEWTFEAKQTDGERR------------------------------ 200
TasA_B_maris  VDFKFNLEATQWNGVKIDTNDA---------NGEINNGNQVA----- 245
TasA_B_selen  ATVNFQFEARQWGGQEVTDSDIGAGLKGADKEGYIQTNERANNGN-- 298
TasA_B_azto   AQISFTLEATQWDGLTTD---------QTKDNGYLDINKEANSN--- 234
TasA_B_firm   ITLDFNFEARQWGGLTIE-------ESHVDEDGYISENEKSHSED-- 263
TasA_B_pum    IQIDLSFEATQWNGLTINP------KKHTDEKGYVKENEKAHSEDKK 196
TasA_B_safe   IQIDLSFEATQWNGLTINP------KKHTDEKGYVKENEKAHSEDKK 266
TasA_B_xia    IQIDLSFEATQWNGLTINP------KEHTDEKGYVKENEKAHSEDKK 194
TasA_B_strat  IQVDLSFEATQWNGLTINP------KKHTDEKGYVKENEKAHSEDKK 266
TasA_B_alti   IQVDLSFEATQWNGLTINP------KKHTDEKGYVKENEKAHSEDKK 266
TasA_B_lich   VQLDFSFEATQWNGLTIDG------KKHADEKGYVKENERAHSEDK- 264
TasA_B_sono   VQLDFSFEGTQWNGLTIDG------NKHADEKGYVKENERAHSEDK- 264
TasA_B_sub    IKLQFSFEATQWNGLTIK-------KDHTDKDGYVKENEKAHSEDKN 261
TasA_B_amy    ISLQFSFEATQWNGLTIT-------KDHTDKDGYVKENEKAHSEDKN 261
TasA_B_sia    ISLQFSFEATQWNGLTIT-------KDHTDKDGYVKENEKAHSEDKN 261
TasA_B_atro   IQLNFSFEATQWNGLTIK-------KDHTDKNGYIKENEKAHSEDK- 260
```

Figure SI 7: Characterisation of recombinant orthologous TasA. (A) SDS-PAGE of recombinant 10 µg *B. licheniformis*, *B. amyloliquefaciens* and *B. cereus* TasA and CalY generated in *E. coli*. **(B)** Coverage map of orthologue recombinant proteins sequences as determined by tandem mass spectrometry. Amino acid designated 1 signifies the first amino acid of the mature protein: *B. amyloliquefaciens* amino acid 27, *B. licheniformis* amino acid 28 and *B. cereus* for both CalY and TasA amino acid 30. **(C)** Transmission electron microscopy images of recombinant orthologue TasA stained with uranyl acetate show presence of fibres which are several micrometres long and approximately 22nm wide, with *B. cereus* TasA fibres being significantly wider at approximately 60nm. **(D-F)** Serial dilutions (from 4 mg/ml down to 0.0625 mg/ml) were prepared for *B. subtilis*, *B. licheniformis*, and *B. cereus* fTasA. Diffusion coefficients of 1µm beads were extracted from mean squared displacement curves. The diffusion coefficients were measured over successive days. The *B. licheniformis* and *B. cereus* fTasA samples gelled at lower concentrations compared to *B. subtilis* fTasA. Moreover, after 3 days, the *B. licheniformis* and *B. cereus* fTasA samples became gels for all concentrations studied indicating a slow, dynamical gelling process.

A



B

B. amyloliquefaciens Protein sequence coverage: 100%
Matched peptides shown in ***bold red***.

```
  1 AAFNDVKSTD ATFASGTLDL SAKEQSANVN LSNLKPGDKL TKDFEFRNNG
 51 SLAIKEVLMA LNFTDFKGAK KGNESAEDFL SQFEITVLTV GKEGGNGYPK
101 NIILKAASLK DLYLMSTKQD KAAAEAISKH IDPKFLSESG KVNVATINGK
151 TAPEYDGVPK TPADYDQVRM EIQFKNETAK TADGLSVQNK FQGNAISLQF
201 SFEATQWNGL TITKDHTDKD GYVKENEKAH SEDKN
```

B. licheniformis Protein sequence coverage: 96%
Matched peptides shown in ***bold red***.

```
  1 AFNDIETTQA TYAAGTLDLN AKDTSARVNL SNLKPGDKFT KDFEFKNDGS
 51 LAIKEVLMQV GYSNFVDGNA KNGGKSTAED FLKQFKVSVL TVGVEGGNGY
101 PKNIILDEAN LYDLYNMSAK KDKNAYEKVK KAIEPEFLHD NGKINVATIN
151 GKTAPEYDGI PKDPYDFDKV QLVIEFVNDK TTDASGRMVQ NKYQGDSVQL
201 DFSFEATQWN GLTIDGKKHA DEKGYVKENE RAHSEDK
```

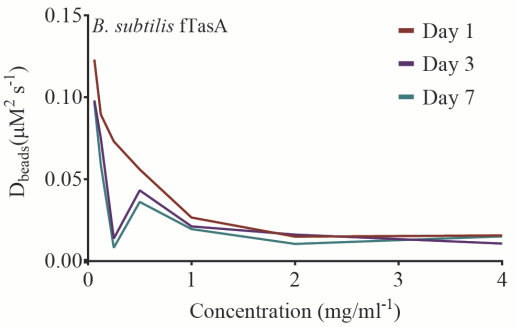
B. cereus TasA Protein sequence coverage: 100%
Matched peptides shown in ***bold red***.

```
  1 FFSDKEVSNN TFAAGTLDLT LNPKTLVDIK DLKPGDSVKK EFLLQNSGSL
 51 TIKDVKLATK YTVKDAKGDN AGEDFGKHVK VKFLWNWDKQ SEPVYETTLA
101 DLQKVDPDLL AKDIFAPEWG EKGGLAAGTE DYLWVQFEFV DDGKDQNIFQ
151 GDTLNLEWTF NANQEAGEEK
```

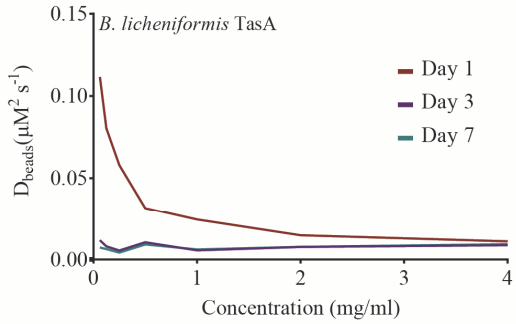
B. cereus CalY Protein sequence coverage: 97%
Matched peptides shown in ***bold red***.

```
  1 FFSDKEVSNN TFAAGTLDLE LNPSTVVNVS NLKPGDTIEK EFKLENKGSL
 51 DIKKVLLKTD YNVEDVKKDN KDDFGKHIKV TFLKNVDKHE TIVKQTTLDK
101 LKGDTLTAVD NDLSAWFWDE KGISAGKSDK FKVKFEFVDN GKDQNQFQGD
151 KLQLNWTFDA QQTAGEEK
```

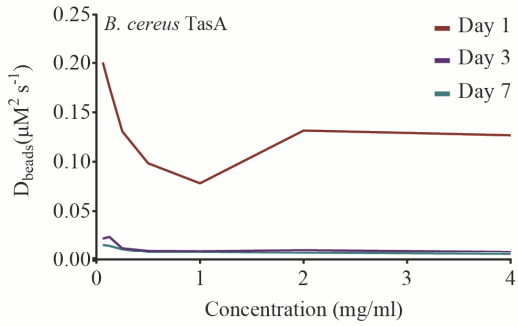
D



E



F



C

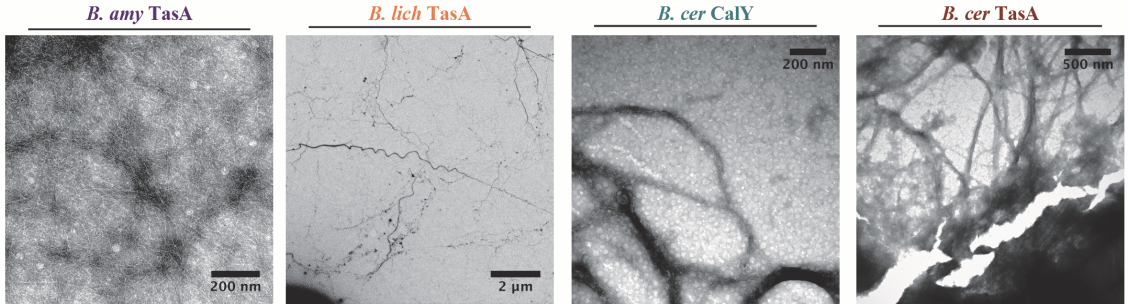
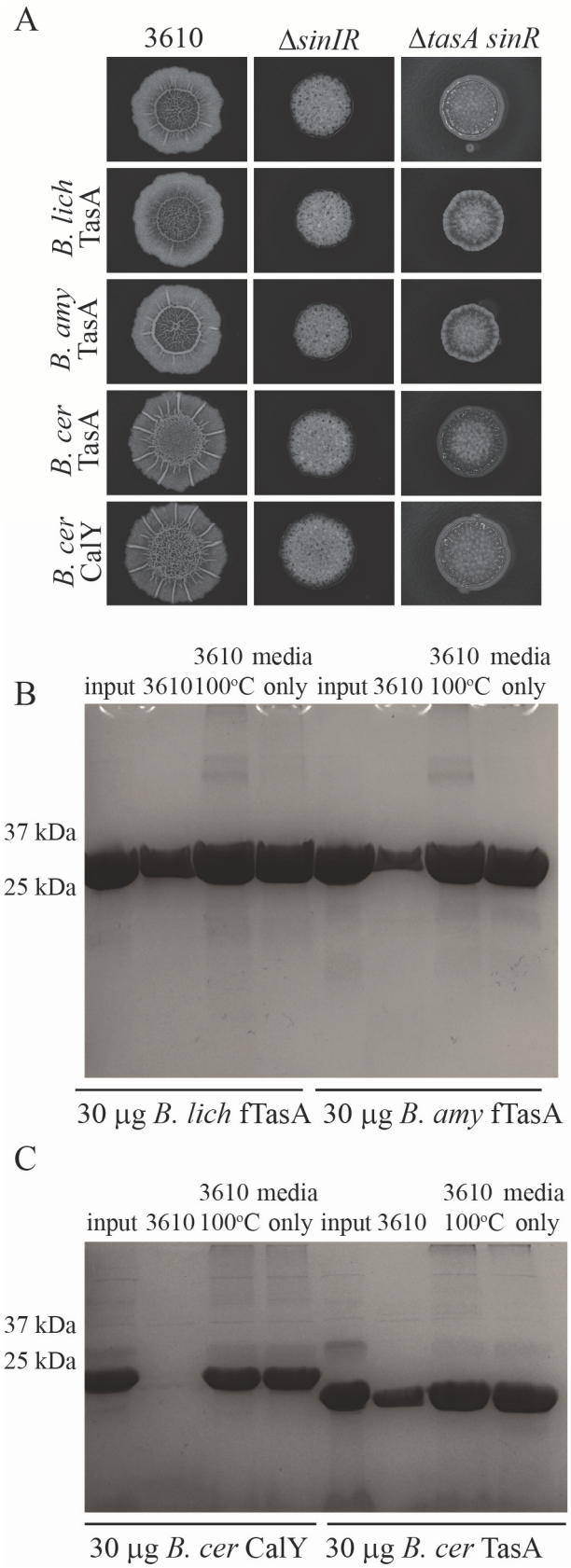


Figure SI 8: Complementation of Δ*tasA sinR* by recombinant orthologous fTasA. (A) Biofilm phenotypes of wild type (NCIB3610), Δ*tasA* (NRS5267), Δ*sinIR* (NRS2012) and Δ*tasA sinR* (NRS5248) strains with *ex vivo* addition of 10 µg recombinant orthologous TasA as indicated. **(B-C)** Integrity of 30 µg *B. licheniformis* TasA, *B. amyloliquefaciens* TasA and *B. cereus* CalY and TasA incubated for 24 hrs at 37°C then analysed by SDS-PAGE. The protein was incubated with filtered supernatant collected from NCIB3610, plus and minus heat inactivation at 100°C and a media only control.


A
3610
ΔsinIR
ΔtasA sinR
B. lich TasA
B. amy TasA
B. cer TasA
B. cer CalY
B
3610 media
input 3610 100°C only input 3610 100°C only
37 kDa
25 kDa
30 µg B. lich fTasA
30 µg B. amy fTasA
C
3610 media
input 3610 100°C only input 3610 100°C only
37 kDa
25 kDa
30 µg B. cer CalY
30 µg B. cer TasA