

ARTICLE OPEN



Molecular characterization of the N-terminal half of TasA during amyloid-like assembly and its contribution to *Bacillus subtilis* biofilm formation

Jesús Cámara-Almirón^{1,3}, Laura Domínguez-García¹, Nadia El Mammeri^{2,4}, Alons Lends^{2,5}, Birgit Habenstein^{1,2}, Antonio de Vicente¹, Antoine Loquet^{1,2} and Diego Romero¹✉

Biofilms are bacterial communities that result from a cell differentiation process leading to the secretion of an extracellular matrix (ECM) by part of the population. In *Bacillus subtilis*, the main protein component of the ECM is TasA, which forms a fiber-based scaffold that confers structure to the ECM. The N-terminal half of TasA is strongly conserved among *Bacillus* species and contains a protein domain, the rigid core (RcTasA), which is critical for the structural and functional properties of the recombinant protein. In this study, we demonstrate that recombinantly purified RcTasA in vitro retains biochemical properties previously observed for the entire protein. Further analysis of the RcTasA amino acid sequence revealed two aggregation-prone stretches and a region of imperfect amino acid repeats, which are known to contribute to functional amyloid assembly. Biochemical characterization of these stretches found in RcTasA revealed their amyloid-like capacity in vitro, contributing to the amyloid nature of RcTasA. Moreover, the study of the imperfect amino acid repeats revealed the critical role of residues D64, K68 and D69 in the structural function of TasA. Experiments with versions of TasA carrying the substitutions D64A and K68AD69A demonstrated a partial loss of function of the protein either in the assembly of the ECM or in the stability of the core and amyloid-like properties. Taken together, our findings allow us to better understand the polymerization process of TasA during biofilm formation and provide knowledge into the sequence determinants that promote the molecular behavior of protein filaments in bacteria.

npj Biofilms and Microbiomes (2023)9:68; <https://doi.org/10.1038/s41522-023-00437-w>

INTRODUCTION

Biofilm formation is an intrinsic ability of bacterial species that allows their association in communities in which bacterial cells are closely linked to each other by the secretion of a self-produced extracellular matrix (ECM)¹. The ECM is mostly comprised of extracellular polysaccharides, proteins and other molecules that allow for a more efficient interaction of the microbial community with the environment and act as a shield that protects cells against external stressors^{1,2}. Functional amyloids and other proteinaceous filaments play a key role in the assembly of these complex multicellular communities, providing mainly structural support.

Amyloids are proteins that have the ability to transition from monomers into insoluble fibers that share a common quaternary structure enriched in beta-sheets stacked in the so-called cross-beta architecture³. These proteins have traditionally been studied in the context of human diseases, as several amyloid proteins and peptides are considered the causative agents of many protein misfolding disorders that lead to known neurodegenerative pathologies^{4,5}. However, the beneficial role of amyloid proteins in many biological functions required for homeostasis is currently well established, which is why part of this broad family of proteins has been subclassified as functional amyloids⁵. Alternative functions reported for these proteins in bacteria are antimicrobial, promotion of bacterial virulence in plants or providing a structural scaffold during biofilm formation (see^{6–8} for extended reviews on the subject).

The formation of the amyloid fold is not a trait exclusively restricted to amyloid proteins but is more a universal property of polypeptide chains when certain physicochemical conditions are achieved³. In amyloids, this specific folding pattern is favored, which is partly due to the amino acid context where certain regions, the presence of different amino acid patterns (such as imperfect repeats^{9,10}) or the appearance of specific amino acids in certain conformations determine the amyloid propensity of a given polypeptide chain under physiological conditions¹¹. These regions have been studied extensively in amyloid proteins, as they form what is known as the rigid core of the protein, the part of the protein that contributes more to the cross-beta structure and is more tightly folded¹², conferring the fibers their unique physicochemical properties.

In species from the *Bacillus* genus, the main protein component of biofilms is TasA¹³, which differs between the two phylogenetically distinct groups of *Bacillus*, the *subtilis* group and the *cereus* group¹⁴. TasA, in its monomeric form, is a stable soluble globular protein that is able to transition during biofilm formation into filaments that can be amyloid or non-amyloid in nature and that have been characterized from a functional, morphological, biochemical and structural perspective. TasA native fibers have been shown to form in vivo, by direct observation in biofilm samples^{15,16}, and in vitro, in protein extracted from biofilm samples in a pre-aggregated conformation^{16–19}. In this context, TasA filaments exhibit a structural arrangement that is enriched in

¹Departamento de Microbiología, Instituto de Hortofruticultura Subtropical y Mediterránea “La Mayora”, Universidad de Málaga, (Campus Universitario de Teatinos), Málaga, Spain. ²CNRS, Chemistry and Biology of Membranes and Nanoobjects (CBMN), Institut Européen de Chimie et Biologie (IECB), University of Bordeaux, Pessac, France. ³Present address: Department of Fundamental Microbiology, Faculty of Biology and Medicine, University of Lausanne, Biophore Building, Lausanne, Switzerland. ⁴Present address: Department of Chemistry, Massachusetts Institute of Technology, 170 Albany Street, Cambridge, MA 02139, USA. ⁵Present address: Latvian Institute of Organic Synthesis, Aizkraukles 21, Riga LV 1006, Latvia. ✉email: diego_romero@uma.es

beta-sheets but maintaining the structure of the monomer, in which the polymerization is driven by donor strand complementation and thus, its structure differs from the canonical cross-beta architecture that is associated to amyloid proteins¹⁶. Additionally, recombinant TasA heterologously expressed in *E. coli* has been shown in vitro to form filaments exhibiting typical amyloid features and a cross-beta architecture^{14,15,20}, but also non-amyloid fibers under different experimental conditions²¹. The assembly of TasA is assisted by TapA, a two-domain partially disordered accessory protein, with flexible N- and C-termini^{22,23} that is also essential for biofilm formation, which promotes TasA polymerization in a chaperone-like manner, supporting the donor strand complementation model, and anchors the fibers to the cell surface^{14,23,24}. The main biological function of these fibers in biofilms is structural, creating a scaffold that provides integrity to the ECM¹⁵. However, the protein plays other biologically relevant roles, contributing to membrane stability and dynamics or acting as a developmental signal that maintains a motile subpopulation within biofilms^{25,26}.

In previous works¹⁴, we studied the structural properties of TasA amyloid-like thioflavin-t (ThT) stainable filaments, where a rigid region, designated as the rigid core and comprising mostly the N-terminal half of the protein, was delimited. In the present work, we have revealed some of the sequence determinants of this core that promote the amyloid-like biochemical behavior of TasA under these specific conditions. By combining bioinformatics, mutational analysis and structural, biochemical, and biophysical studies, we defined specific residues critical for the structural function of TasA

in the ECM. All these results allow for a better comprehension of the structural function of TasA during ECM production and biofilm formation and thus improve our understanding of the role of functional amyloids and other filamentous proteins in the persistence and interaction of bacteria in and with the environment.

RESULTS

TasA contains an N-terminal rigid core that retains amyloid-like properties in vitro

Solid-state nuclear magnetic resonance spectroscopy analysis (SSNMR) on assembled recombinant TasA fibers allowed us to delimitate a rigid core in the recombinant ThT-stainable filaments, comprising 110 amino acids that extends from residues K35 to K144¹⁴ in TasA sequence (Fig. 1 residues labeled in red) and that we have designated as rigid core TasA (RcTasA). RcTasA is located in the N-terminal half of the protein, which is the region of TasA that is highly conserved across different *Bacillus* species (Fig. 1). Based on these findings, we hypothesized a relevant contribution of this region to the biochemical and biophysical properties of the protein and to the final architecture of the TasA filament and functionality. We analyzed the properties of the RcTasA sequence using two complementary bioinformatic approaches. First, we searched for amino acid repeats within the sequence, which have been demonstrated to be involved in the polymerization process of other amyloid proteins, including bacterial functional amyloids^{9,10,27,28}; second, we looked for aggregation-prone regions in

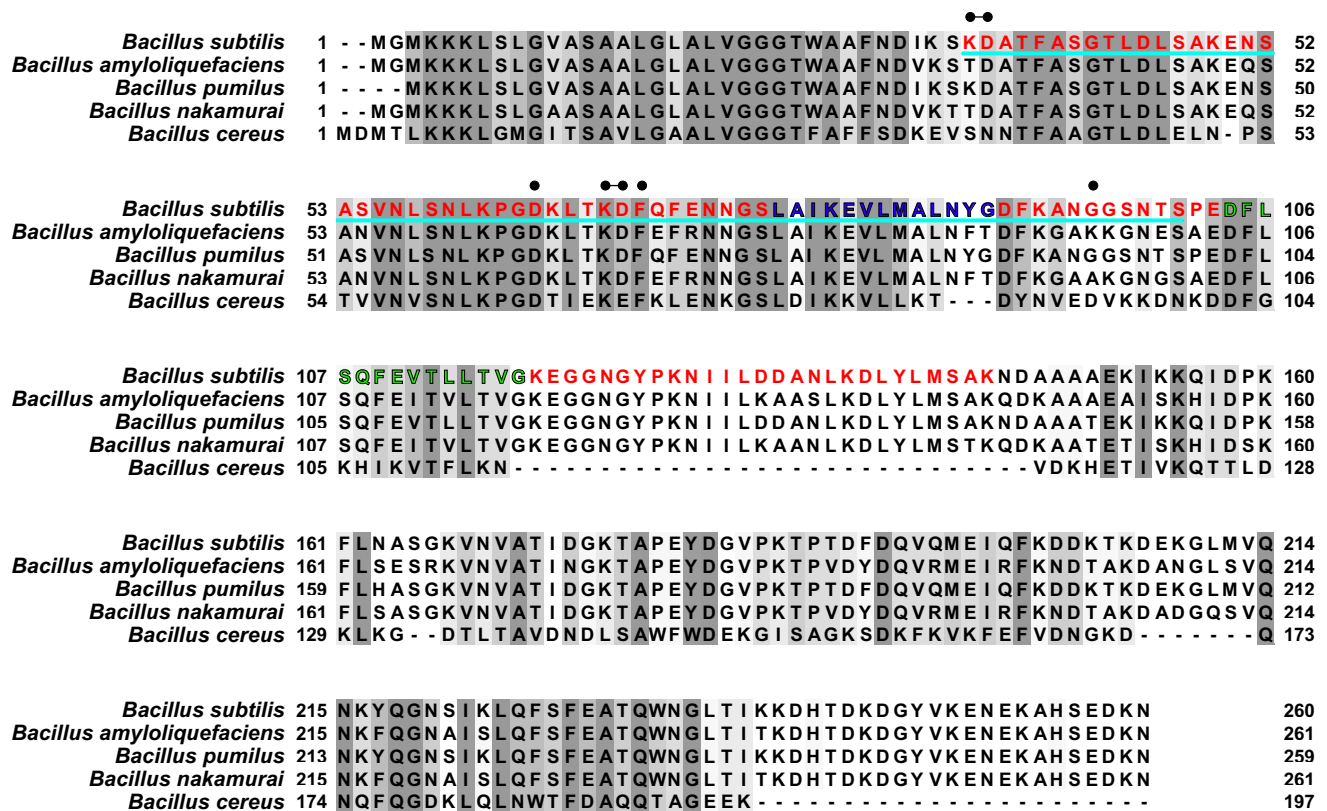


Fig. 1 Sequence analysis of the N-terminal domain of TasA reveal amyloidogenic stretches and imperfect amino acid repeats within the amyloid core region of TasA. Sequence comparison of the N-terminal domain of TasA between different species of *Bacillus*. The different features detected by bioinformatic tools are labeled over the sequence. The gray colors indicate different degrees of sequence conservation. The region corresponding to the amyloid core of TasA is labeled in red (from K35 to K144). The two amyloidogenic stretches are labeled in blue (L78 to G90, LG-13) and green (D104, G117, DG-14). The sequence underlined in cyan correspond to the imperfect amino acid repeat in which the sequence $KD_{xx}F_{xxxxxx}L_{xx}K_{xxxx}N_{xxxx}G_{xxxx}$ is repeated twice in the core. Black dots indicate residues that have been analyzed by site-directed mutagenesis.

RcTasA. These tools revealed the presence of (i) two aggregation-prone regions that extend from L78 to G90 (called here the segment LG-13) and from D104 to G117 (called DG-14) (Fig. 1, labeled, respectively, in blue and green) and (ii) a sequence of imperfect amino acid repeats in which the sequence KDxxFxxxxxxLxxKExxxxxNxxxxKxxxGxxxx is repeated twice and extends from K35 to S101 (Fig. 1 underlined in cyan).

To experimentally corroborate the predicted amyloid-like properties of this region, RcTasA, heterologously expressed in *E. coli*, was purified to homogeneity and maintained soluble in monomeric form in a 1% acetic acid solution¹⁴ prior to studying the polymerization on buffer at physiological pH (20 mM Tris, 50 mM NaCl at pH 7). Transmission electron microscopy (TEM) analysis of RcTasA in buffer after 1 week of incubation revealed filaments with a bundle-like supramolecular organization that resembled those formed by the entire TasA protein (Fig. 2a, left and more detail in right micrographs). The bundle widths varied from ~5 to ~25 nm with a tendency to stack themselves to form larger objects. The smallest molecular entity observable, presumably protofibrils, averaged ~3.5 nm wide. Given the ability of RcTasA to aggregate and form filament-like structures, we further characterized the kinetics of the aggregation process over time using dynamic light scattering (DLS), a noninvasive technique that allows us to monitor the size distribution of the aggregates found in the sample based on the signal intensity associated with each hydrodynamic radius (R_H). The absence of major changes in the size distribution of the population during the time studied and the mean R_H of ~10 nm of the population immediately after buffer exchange (time 0 h), which is far above what is attributed to WT TasA^{14,17}, were two complementary pieces of evidence suggesting that polymerization of RcTasA occurs rapidly upon buffer exchange to neutral pH (Fig. 2b). TEM analysis of RcTasA samples at $t=0$ h, immediately after buffer exchange and thus prior to their assembly, revealed the presence of large aggregates of several microns of diameter that were previously undetected by DLS (probably due to their large size), supporting the hypothesis of a rapid aggregation in these experimental conditions (Fig. 2c). After this initial stage, two populations of similar abundance values and mean R_H values of ~12 nm and ~22 nm were observed at 72 h, indicative of a transition into larger aggregates. According to this dynamic, a more homogeneous size distribution of aggregates with a mean R_H of ~25 nm dominated at 144 h (Fig. 2b, middle and bottom graphs). Overall, this set of experiments confirms that RcTasA rapidly transitions at pH 7 from monomers to aggregates that exhibit an unequivocal fibrillar form. Next, we wondered whether these RcTasA assemblies exhibited biochemical features similar to those of full-length WT TasA.

Amyloid proteins exhibit unique tinctorial properties as they are able to bind to specific dyes. One such dye is thioflavin-T (ThT). ThT has been extensively used to characterize amyloid proteins in the past and is still one of the most widely used methods to infer the presence of amyloid and amyloid-like proteins in samples nowadays in combination with other techniques²⁹. ThT binds to the beta-sheets that are present in amyloid samples, which leads to a shift in the fluorescence emission maximum of the dye that can be used to estimate the gain of beta-sheet secondary structure; this occurs concomitantly with the transition from the monomeric to the fibrillar form^{30,31}. ThT binding assays of purified RcTasA showed a progressive increase in the fluorescence intensity of ThT at the amyloid-specific wavelength over time in a concentration-dependent manner (Fig. 2d top). Moreover, this ThT binding kinetics was shown in a polymerization curve similar to that observed for WT TasA in vitro and other amyloid proteins^{14,32–34}. To exclude any effect of scattering in the observed signal, we recorded the complete emission spectra of RcTasA between 455 and 650 nm at an excitation wavelength of 450 nm in the presence or absence of ThT (Fig. 2d bottom). The spectra

showed that the effect of scattering was negligible, demonstrating that the observed increase in fluorescence over time is due to the binding of ThT to the RcTasA samples.

As mentioned above, the rigid core is the part of the protein more tightly folded in the final tridimensional structure of the fiber, and therefore, it exhibits remarkable physicochemical robustness, including resistance to proteases³⁵. In our previous work, RcTasA was detected after TasA fibrils were subjected to partial proteinase K digestion¹⁴. Thus, we wondered if the aggregation properties of isolated RcTasA preserved the structural fold that confers TasA fibers their resistance against proteolytic activity. In a similar protease digestion experiment, polymerized RcTasA resisted treatment with proteinase K even 1 h after incubation, as shown by the ~13 kDa band resolved in the Coomassie-stained SDS-PAGE gel of the digestion samples compared to the untreated control (Fig. 2e, last lane on the right and lane labeled with a “C”, respectively). A similar molecular behavior of RcTasA samples was observed at $t=0$ h of incubation, prior to their assembly. Digestion with proteinase K showed more level of degradation (less material) of this sample (0 h) compared to the 1 week assembled samples, however, the level of resistance observed at short times of exposure to proteinase K (Fig. 2f) reflected the remarkable robustness of the RcTasA from the very early stages of polymerization.

Amyloidogenic features of the N-terminal half of TasA contribute to the molecular behavior of RcTasA

Our previous findings demonstrated that RcTasA exhibits biochemical features similar to those observed for the whole TasA protein and proved that these properties are retained in the N-terminal half. Thus, we asked if the features previously predicted in our in silico analysis (Fig. 1) were relevant to defining the structure or functionality of TasA. Our bioinformatic predictions of aggregation-prone stretches consistently showed two regions of RcTasA, designated LG-13 and DG-14, with a high tendency for aggregation (Fig. 3a) (see Fig. 1, labeled in blue and green, respectively). To study whether these two regions contribute to biofilm formation and have any impact on the structural function of TasA, we performed a mutagenesis analysis in which either of these two stretches was deleted from the protein sequence. As described in a previous work²⁵, to avoid undesirable effects when manipulating the *tasA* gene in the endogenous operon, these experiments were performed using a strain lacking the whole *tapA-sipW-tasA* operon and complemented with a version of the operon containing the mutated versions of *tasA* and integrated in the neutral *lacA* locus. Complementation of the $\Delta(tapA-sipW-tasA)$ strain with an operon containing a native version of TasA rescued the wrinkled phenotype associated with the WT strain in solid and liquid cultures (Fig. 3b, left images). However, complementation with the operon containing a TasA version lacking amino acids from E82 to N88 ($\Delta 82-88$, LG-13 region) or from Q108 to V116 ($\Delta 108-116$, DG-14 region) failed to restore the biofilm formation phenotype (Fig. 3b, middle and right images). To check if these mutated versions of TasA were stable, we performed a biofilm fractionation assay followed by protein precipitation and Western blot analysis using anti-TasA antibodies. A strong anti-TasA reacting band was observed in samples from WT or the control strains, but the signal was absent in samples of the strains carrying the mutated alleles (Fig. 3c). This result, along with the fact that the two strains bearing the mutated versions of TasA showed the same phenotype, suggested that the deletion of these regions renders the protein unstable, which most likely leads to the degradation of the protein by quality control proteases.

Complementarily, we studied the amyloidogenic behavior of these two amino acids stretches in vitro using synthetic peptides. None of the peptides showed ThT binding activity (β -sheet formation) during the first 20 h of incubation. However, after this

latent phase, the LG-13 peptide exhibited an exponential increase in the fluorescence signal, reaching saturation a few hours later (Fig. 3d, blue). Transmission electron microscopy analysis at the end of the ThT experiment revealed molecular entities of small size without any significant organization for the DG-14 peptide

(Fig. 3d, left image). Abundant fibrillar assemblies made of bundles comprised of several protofilaments (~5 nm diameter each) ranging between 6 and 14 nm in diameter, which is a molecular organization very similar to the full-length mature TasA assemblies observed *in vitro*, characterized samples of the LG-13

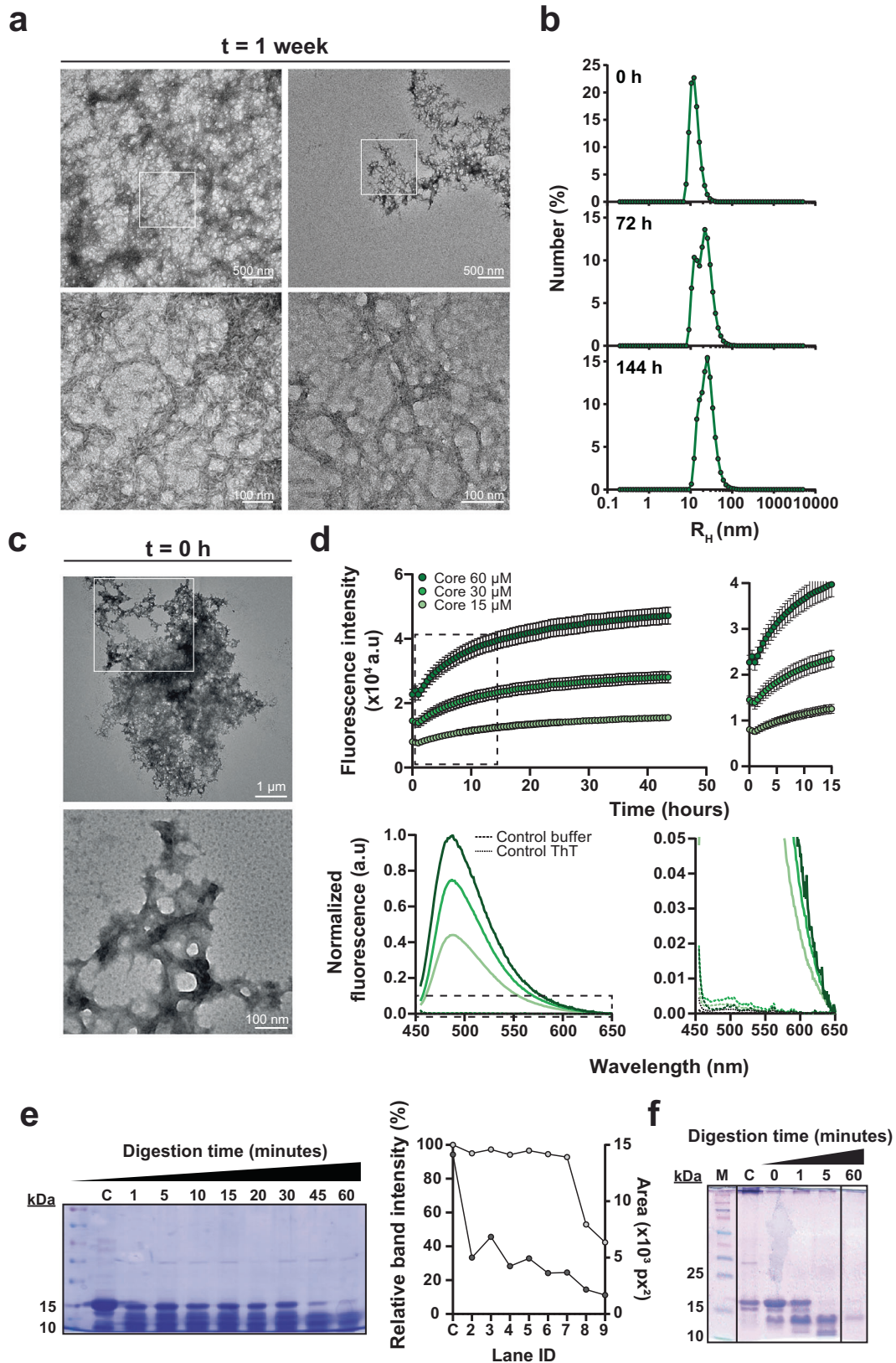


Fig. 2 The TasA rigid core region exhibits amyloid properties in vitro. **a** Transmission electron micrographs of negatively stained RcTasA samples show dense bundles of fibrillar assemblies. The scale bars are 500 nm (top) and 100 nm (bottom). White squares over the images indicate the areas that were zoomed in during the imaging. The right images show the structures formed by the RcTasA in detail. **b** Aggregation kinetics of RcTasA as measured by DLS at different time points. **c** Transmission electron micrographs of negatively stained RcTasA samples of 0 h of incubation after buffer exchange. White squares over the image indicate the area that was zoomed in during the imaging. The scale bars are, from top to bottom, 1 μ m and 100 nm. **d** Top: Kinetics of fluorescence emission at 480 nm of RcTasA samples indicates ThT binding properties. Experiments were performed at different protein concentrations. Average values of three independent experiments are represented. Error bars indicate the SEM. Dashed rectangle indicates the area of the graph that has been enlarged in the right graph. Bottom: Fluorescence emission spectra of the RcTasA samples in the presence or absence of ThT after 40 h of incubation under the same experimental conditions. The spectra were recorded between 455 and 650 nm at an excitation wavelength of 450 nm. Dashed rectangle indicates the area of the graph that has been enlarged in the right graph. **e** Left: Coomassie stained SDS–PAGE gel of the RcTasA samples assembled for 1 week and digested with proteinase K at different time points. The first lane indicates the molecular marker. C refers to the untreated control. Each lane contains a sample corresponding to a specific digestion time. Right: Quantification of the relative intensity (light gray) and area (dark gray) of the ~12 kDa proteinase K resistant band at different time points. **f** Coomassie stained SDS–PAGE gel of the RcTasA samples assembled for 0 h and digested with proteinase K at different time points. M indicates the molecular marker. C refers to the untreated control. Each lane contains a sample corresponding to a specific digestion time. SDS–PAGE gel images have been cropped and spliced for illustrative purposes. The lines over the gel images indicate the boundaries of the image splicing. The three slices in the gel image are derived from a single gel, respectively.

peptide. We further studied the rigid core of the self-assembled fibrils made of LG-13 or the molecular entities of DG-14 TasA peptides by solid-state NMR spectroscopy. Using magic-angle spinning NMR, we measured 1D dipolar-based ^{13}C cross-polarization (CP) spectra to assess the secondary structure conformation of TasA subunits along the fibril axis (Fig. 3e). Both peptides showed strong and sharp ^{13}C NMR signals (Fig. 3e), indicating that the structural core of these structures is made of well-ordered and rigid subunits. We compared those 1D ^{13}C spectra to that of full-length TasA fibrillated in vitro¹⁴. Overall, the spectral envelopes of the three samples are quite similar, indicating a very similar secondary conformation. In particular, the backbone carbonyl ^{13}C atoms accurately probe the conformation of amino acids using NMR chemical shifts. Indeed, β -sheet carbonyl chemical shifts are ~173 ppm, while α -helical residues are found at ~178 ppm (using a DSS-based NMR scale) (Fig. 3f). Similar to full-length TasA, our structural analysis by NMR suggests that LG-13 and DG-14 peptides form β -rich entities with different supramolecular organization. Overall, these results suggest that LG-13 and DG-14 contribute to the amyloid properties exhibited by TasA and maintain the same secondary structure conformation. Although the LG-13 region is the only one able to self-assemble into ThT-stainable filaments, the SSNMR analysis indicates that the molecular entities formed by the two peptides contain well-ordered subunits associated with low structural polymorphism.

RcTasA contains an imperfect amino acid repeat with specific amino acids critical for protein functionality

Repeated regions within amyloid proteins have been demonstrated to contribute to the assembly of amyloid filaments^{9,10,36}, and an imperfect amino acid repeat, KDxxFxxxxLxxKExxxxNxxxxKxxxGxxxx, which is repeated twice, was predicted within the RcTasA region (Fig. 1, cyan line). To explore the contribution of this region to the structure and functionality of TasA, we performed a mutational analysis by alanine scanning on selected amino acids K35, D36, D64, K68, D69, F72 and G96, which showed different levels of conservation (Fig. 1, black dots). The strain expressing the native *tasA* allele was unaffected in its ability to form a WT-like biofilm in solid or liquid medium (Fig. 4a). Strains expressing some of the mutated alleles (K35A, D36A; F72A or G96A) also displayed the characteristic wrinkled morphology of *B. subtilis* biofilms (Fig. 4a). However, the strains that expressed the D64A or the K68AD69A alleles failed to fully restore the WT biofilm formation phenotype (Fig. 4a).

We speculated that this partial loss of phenotype could be related to the instability of the mutated TasA alleles due to the alanine substitutions. Biofilm fractionation assay from pellicles followed by immunoblot analysis using the anti-TasA antibody

(Fig. 4b) showed the presence of an anti-TasA reacting band in all the fractions from the WT strain, the strain carrying the native TasA allele or the strain carrying the TasA D64A allele. However, a weaker signal was observed in fractions of cells expressing the TasA K68AD69A version. In addition, two well-defined reacting bands appeared in the medium and ECM fractions of samples from this strain, which suggests interference of the amino acid substitution with protein processing. Nonetheless, we concluded from this experiment that the TasA version of the mutants with an altered biofilm formation phenotype is sufficiently stable to be detected in protein extracts; thus, the mutant phenotype cannot be explained by the instability and degradation of TasA. TEM analysis of 48 h pellicles of these two mutated strains showed a network of branched fibers decorating the surfaces of cells expressing the native TasA or the TasA D64A alleles (Supplementary Fig. 1, top and second from the top micrograph panels). In contrast, straight fibrillar structures were scarcely observed in cell preparations of the strain expressing the TasA K68AD69A allele (Supplementary Fig. 1).

Thus, it appeared that the structural function of TasA in the formation of the filaments and required for biofilm formation was compromised in this strain. We then performed an in silico analysis in which the mutated proteins were modeled and compared to the WT protein (Fig. 5). The TasA WT structure had to be predicted and used as a reference when comparing the variant proteins, although the structure of the TasA monomer has already been experimentally determined and deposited in the Protein Data Bank²⁰. Differences between the modeled WT TasA predicted by AlphaFold and the crystallographic model were found in the loop between K118 and Y124, which was not modeled in the crystal structure, or in some regions that are modeled differently in the predicted WT structure, such as the first amino acids or the absence of beta-sheet $\beta 7$ (Supplementary Fig. 2A). We represented the different protein models using the crystal structure as a reference, using the same schematic representation of the different structural features of the protein that was used in another study²⁰ (Supplementary Fig. 2A). When the models of the mutated versions of TasA and WT TasA were compared (Fig. 5a, b), some noticeable structural differences were found: (i) TasA D64A showed the disruption of the $\beta 6$ beta-sheet by a random coil conformation adopted by residues from K126 to D130 and (ii) the appearance of a new alpha-helix, labeled $\alpha 5'$, from residues G175 to T177 (Fig. 5a). The disruption of the alpha-helix $\alpha 5$ was the only difference between the model of TasA K68AD69A and the TasA WT model (Fig. 5b). To further analyze these structural differences, we additionally modeled the possible hydrogen bonds within each protein structure. The D64 residue present in the loop between $\beta 2$ and $\beta 3$ in the WT protein forms five hydrogen bonds with neighboring amino acids: two with K61 and one with F200

(Fig. 5c top, left image). The substitution of this residue seems to interfere with one of the hydrogen bonds formed with K61 (Fig. 5c top). This single change seems to perturb the structure of the protein and might explain the structural differences when compared to the TasA WT model. In the case of TasA K68A and D69A, the K68 residue is bound to M196 from β 8 through two hydrogen bonds (Fig. 5c bottom). The substitution of these amino acids preserves the hydrogen bonds connecting the two beta-sheets by A68, but an additional bond is formed between A69 and

G42. This change in the hydrogen bond configuration might result in the loss of α 5, whose residues are modeled as random coils in the predicted structure of this TasA variant.

Similarly, we analyzed the effects of these amino acid substitutions over the structure of the non-amyloid TasA filament using the structural data that is currently available¹⁶. The structure of the TasA filament directly isolated from *B. subtilis* biofilms was recently solved using cryo-electron microscopy and deposited in the Protein Data Bank¹⁶. The structural analysis revealed that the

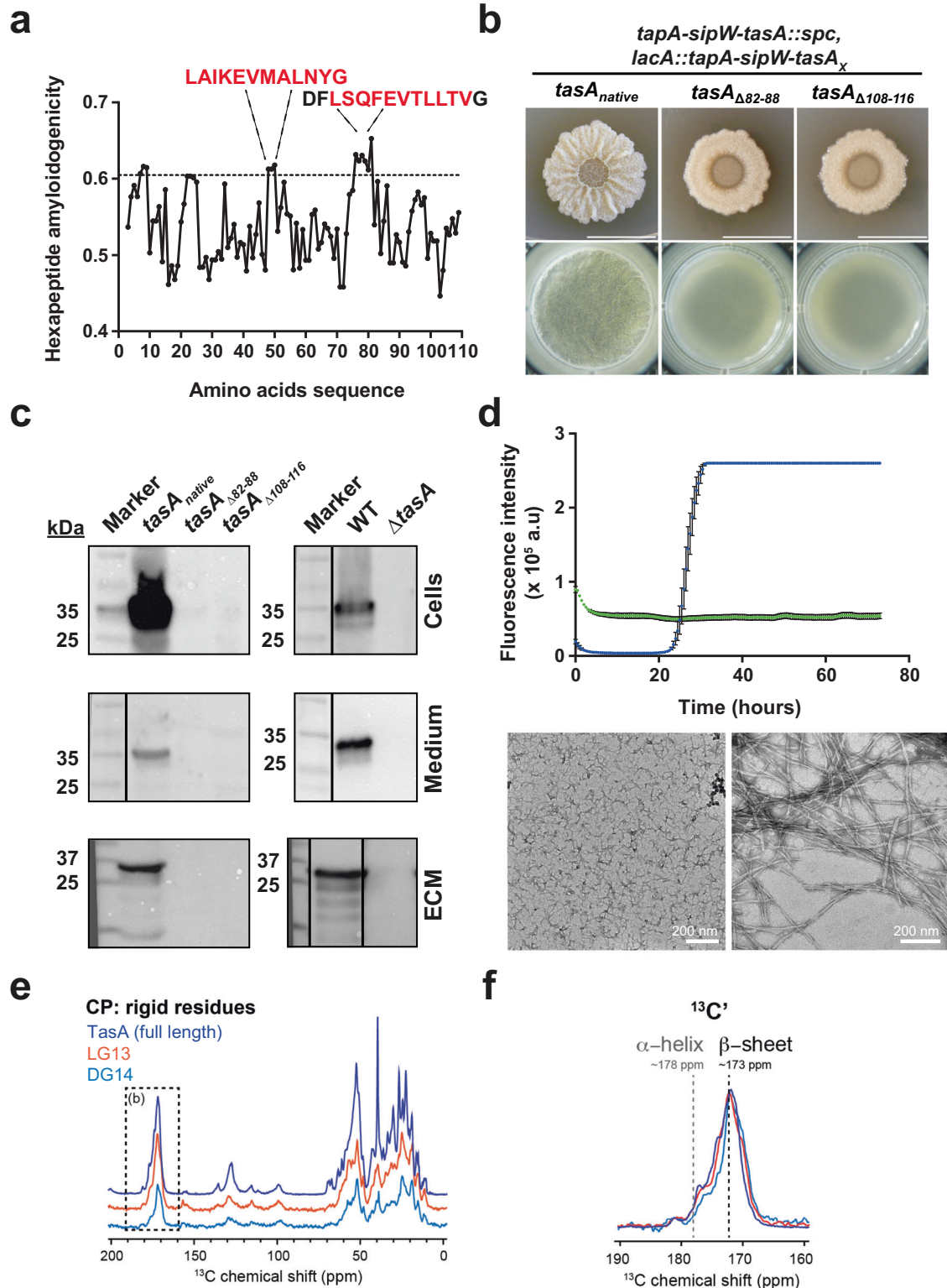


Fig. 3 Amyloidogenic regions are important for protein folding and contribute to the amyloid properties of RcTasA. **a** The hexapeptide amyloidogenicity profile of RcTasA as predicted by MetAmyl reveals two amyloidogenic stretches found in this sequence. **b** Mutagenesis analysis of RcTasA in which the predicted amyloid stretches were deleted from the TasA sequence and the mutated alleles were introduced in a $\Delta(\text{tapA-sipW-tasA})$ background. The images show the phenotypes of *B. subtilis* biofilms grown in solid and liquid biofilm-inducing media of the strains carrying either the native allele (unmodified) or the alleles carrying the deletions. Scale bars = 1 cm. **c** Western blot using an α -TasA antibody of protein extracts corresponding to the different biofilm fractions (cells, medium and matrix) of the strains carrying the abovementioned alleles. Immunoblot images have been cropped and spliced for illustrative purposes. Black lines over the blot images delineate boundaries of immunoblot splicing. In all cases, all the slices shown were derived from a single blot. **d** Top. Kinetics of ThT fluorescence emission at 480 nm of the synthetic peptides corresponding to the amyloid stretch predicted. The average of three independent experiments is shown. Error bars indicate the SEM. Bottom. Transmission electron microscopy micrographs of negatively stained LG-13 and DG-14 synthetic peptides at the final experimental time-point. Scale bar = 200 nm. **e** $^1\text{D } ^{13}\text{C}$ cross-polarization spectra of TasA fibrils (dark blue) and the assemblies formed by peptides LG-13 (red) and DG14 (light blue). **f** Zoomed region of the 1D ^{13}C cross-polarization spectra shown in E between 160 and 190 ppm.

filament, under these experimental conditions, is formed by a process of donor strand complementation of the different subunits in which the first residues of the N-terminal half of the first monomer (n-1) are extended away in a new beta strand that complements the core of the next subunit (n0)¹⁶ leading to the formation of a filament that lacks the characteristic cross- β architecture. To investigate how these amino acid substitutions affect this filament structure, we followed a similar approach as the one described in the previous paragraph. The structure of the WT filament was first predicted using the deposited structure of the TasA filament as a template to find any structural differences between the experimentally resolved structure and the structure predicted by AlphaFold that could interfere with our subsequent analysis (Supplementary Fig. 2B). The structure of TasA obtained with AlphaFold was very accurate, with only minor differences found when comparing the filament subunits of the real structure with the predicted TasA WT model (Supplementary Fig. 2B center and right). We then predicted the structure of filaments containing TasA subunits carrying the amino acid substitutions D64A or K68AD69A to study putative effects on the donor strand complemented filament model (Supplementary Fig. 3A, B). Filaments containing the mutated subunits were structurally similar to the WT version, suggesting that these mutations do not have a major impact in the secondary structure of the filament. However, when these amino acid substitutions are analyzed more closely in the mutated filaments we can find differences in terms of hydrogen bond formation between the different residues that interact with the mutated amino acids. Residue D64 is located in the WT filament in the interface between subunits and forms hydrogen bonds with amino acids from the same subunit (n0, K61, F200), but also mediates the contact with the previous subunit (n-1) by interacting with L236 and N252. These bonds are lost when this amino acid is substituted by alanine (Supplementary Fig. 3A). This result suggests that D64 might play an important role in stabilizing the structure of the filament and the contact between different subunits. On the contrary, residues K68 and D69 are located in the most internal part of the filament's core, as part of β 3 in the filament which does not undergo any major structural change compared to the structure of the TasA monomer (Fig. 5a and Supplementary Fig. 3B). These amino acids form hydrogen bonds in the filament with N51, S52, V55, K118 and M196 and their substitution by alanine disrupts the interaction with V55 and K118, potentially contributing to a lower stability of the filament's core.

Molecular characterization of the TasA variants revealed two different alterations in the core of the mutated proteins

Three residues of the RcTasA core, D64, K68 and D69, are relevant for the functionality of TasA, and the predicted structural changes of TasA associated with their substitution explain their high degree of conservation within the RcTasA sequence (see Fig. 1). Thus, we reasoned that these amino acids could participate in the polymerization of ThT-stainable amyloid-like TasA fibrils. The two

versions of TasA were heterologously expressed and purified, and their molecular properties were studied *in vitro* using the protocols previously optimized for TasA, TapA or RcTasA¹⁴. The structural changes at the monomer and filament level observed for these two versions of TasA were suggestive of alterations at the level of structural stability, amyloid-like behavior and fiber formation compared to the WT TasA.

We further analyzed the aggregation kinetics of the two alleles using DLS. TasA D64A showed a mean R_H of ~ 1.6 nm at the beginning of the experiment, which was slightly smaller than that reported for TasA WT¹⁴ (Fig. 6a left). In alignment to this observation, small and disperse aggregates were observed via TEM analysis at $t = 0$ h of assembly (Supplementary Fig. 4A, left). However, the most representative population in the TasA D64A sample reached a mean R_H between ~ 7.8 and ~ 9 nm after 72 h of incubation and finally reached a mean size between ~ 9 and ~ 10 nm after 144 h (Fig. 6a left). The kinetics of aggregation of TasA K68AD69A were, however, substantially different, given that the mean size distribution of the particles present in the sample was ~ 2 nm at $t = 0$ h which is accordance to the disperse and small protein entities observed via TEM analysis (Supplementary Fig. 4A, right). Conversely, the mean particle size detected by DLS for this TasA variant remained nearly invariable over time (Fig. 6a right). This finding suggested that the aggregation kinetics of this protein are slower than those observed for the WT protein or the TasA D64A allele, indicating a defect of TasA K68AD69A in aggregation. We next measured the characteristic ThT affinity of these molecular assemblies by ThT binding kinetics experiments. The dynamics of fluorescence emission of TasA D64A (Fig. 6b blue) at the amyloid-specific wavelength were similar to those of TasA WT (Fig. 6b gray), indicating a similar aggregation kinetics. However, the fluorescent signal of TasA K68AD69A saturated earlier than that of the TasA WT or TasA K68AD69A proteins, even at the higher protein concentration used in the experiment, and the maximum intensity of the signal was, overall, nearly three times lower than that observed for the TasA WT or TasA D64A proteins (Fig. 6b red), suggesting limited assembly of amyloid-like structures. Morphological characterization by TEM of the molecular assemblies formed by the two mutated variants showed a tendency to form larger and mostly amorphous aggregates compared to TasA WT (Fig. 6c). The aggregates of TasA D64A were composed of associated bundles of what presumably look like aggregated filaments of different sizes, from ~ 5 nm to hundreds of nanometers. Some individual fibrillar entities could also be observed within these bundles, with an average size between ~ 1.5 and ~ 2.5 nm wide and with variable length (Fig. 6c center panel). TasA K68AD69A, however, formed large and densely packed aggregates with individual fibrillar entities that were hardly visible (Fig. 6c right panel). The partial proteinase-K digestion experiment at $t = 0$ h of aggregation shows evident signs of degradation in both TasA variants, as indicated by the very faint and diffuse protein bands observed in the SDS-PAGE gel image of the digestion samples, even at short treatment times

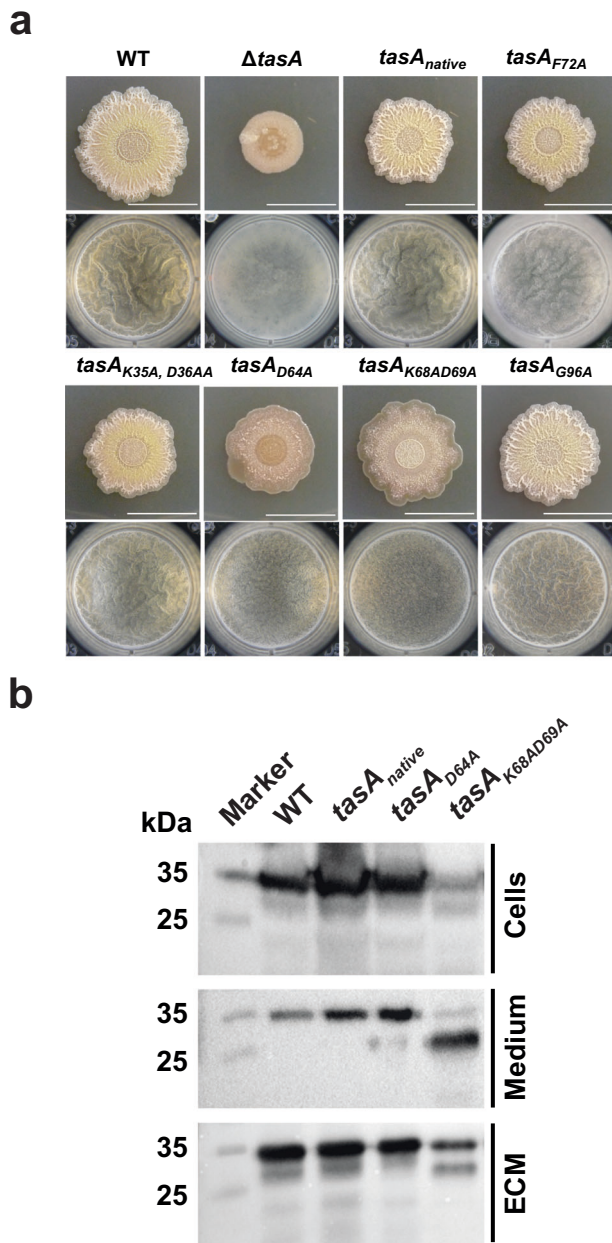


Fig. 4 Mutagenesis analysis of the imperfect amino acid repeat reveals amino acids essential for biofilm formation. **a** Alanine scanning site-directed mutagenesis analysis of RcTasA in which specific amino acids from the imperfect repeat region of RcTasA were substituted from the TasA sequence and the mutated alleles were introduced in a $\Delta(tapA-sipW-tasA)$ background. The images show the phenotypes of *B. subtilis* biofilms grown in solid and liquid biofilm-inducing media of the strains carrying either the native allele (unmodified) or the alleles carrying the substitutions. Scale bars = 1 cm. **b** Western blot using an α -TasA antibody of protein extracts corresponding to the different biofilm fractions (cells, medium and matrix) of the strains carrying the abovementioned alleles that were affected in colony morphology or pellicle formation. Immunoblot images have been cropped for illustrative purposes.

(Supplementary Fig. 4B, left images, top and bottom). This susceptibility to degradation was especially noticeable for TasA D64A at $t = 0$ h of assembly, given that most of the protein was degraded after 5 min of digestion with proteinase K. In contrast, at $t = 1$ week the assemblies formed by both mutated proteins were

resistant to protease activity, as indicated by the bands that were present after digestion in the Coomassie-stained polyacrylamide gels (Supplementary Fig. 4B, middle images, top and bottom). Differences were, however, noticeable after 20 min of digestion. Treatment with TasA D64A rendered resistance bands between 15 and 25 kDa and between 10 and 15 kDa, which progressively disappeared over time. A similar pattern was obtained for TasA K68AD69A; however, the band between 10 and 15 kDa remained after 60 min of digestion, indicating that this variant is likely more resistant to digestion than TasA D64A which is in alignment to the results obtained with the samples at 0 h of assembly (Supplementary Fig. 4B middle images top and bottom and right graphs top and bottom).

All these results indicate that both proteins have the ability to transition from monomers to aggregates. TasA D64A is a stable protein in cells with similar biochemical stability and behavior compared to TasA WT and is, in principle, not affected in its ability to form ThT-stainable aggregates that although resistant to proteinase degradation, shows a lower tolerance to this aggression, and therefore, lower robustness than the WT or the K68AD69A counterparts in the aggregated form. However, TasA K68AD69A appear to be less stable than TasA WT in the cells and exhibit a stronger tendency to form aggregates with limited amyloid-like features. These two distinct molecular behaviors reflect the importance of these residues in the core's contribution to TasA functionality and explain the two different phenotypes derived from these amino acid substitutions.

To further characterize the molecular conformation of TasA D64A, TasA K68AD69A amyloid-like fibrils, we employed solid-state NMR spectroscopy. Recombinant D64A or K68AD69A mutant proteins were overexpressed in *E. coli* and ^{13}C -labeled following protocols previously optimized¹⁴. In vitro self-assembled filaments of the mutant proteins were investigated using magic-angle spinning SSNMR. We recorded a two-dimensional ^{13}C - ^{13}C correlation experiment using a first cross-polarization ^1H - ^{13}C step to reveal the conformation of rigid residues involved in the rigid core. Comparison of ^{13}C - ^{13}C spectral fingerprints (Fig. 7a in red and green) showed that D64A, K68AD69A assemblies exhibit broader NMR signals. ^{13}C line-widths for the two mutants are ~200-400 Hz (measured with full-width at half-height), suggesting a high propensity of the 2 mutants to adopt polymorphic amyloid structures. For WT assemblies (Fig. 7a, in black), narrow peaks are observed with ^{13}C line-widths of ~100-200 Hz, indicating a low amount of structural polymorphism at the local level. These line-width values are not surprising for WT fibrils, since functional amyloids often lead to the aggregation of protein subunits into a unique polymorphic structure characterized by well-resolved SSNMR signals³⁷. Next, we scrutinized the chemical shift values for the alanine $\text{C}\alpha$ - $\text{C}\beta$ spectral region. This spectral region is helpful for probing the secondary structure. Following the procedure described for WT TasA fibrils¹⁴, we compared the two mutants to WT fibrils (Fig. 7b). We observed two trends for the chemical shift values of the mutant fibril NMR signals: (i) several signals are conserved in their chemical shift values between the two mutants and the WT sample, suggesting that these residues have a conserved local conformation between the three samples. Nevertheless, (ii) numerous signals observed in the WT sample disappeared for the two mutants, indicating a partial loss of the structural integrity for the mutants. This observation was also made for other spectral regions of the two-dimensional experiments. Altogether, this chemical shift analysis suggests that the well-ordered amyloid rigid core observed for WT TasA fibrils is only partially conserved for the two mutants, and additionally, a significant number of residues in the rigid core of D64A and K68AD69A mutants have lost their structural stability. Although the chemical shift analysis did not reveal a significant difference between D64A and K68AD69A mutants, the comparison of cross-polarization (CP) and J-based (INEPT) polarization transfer showed

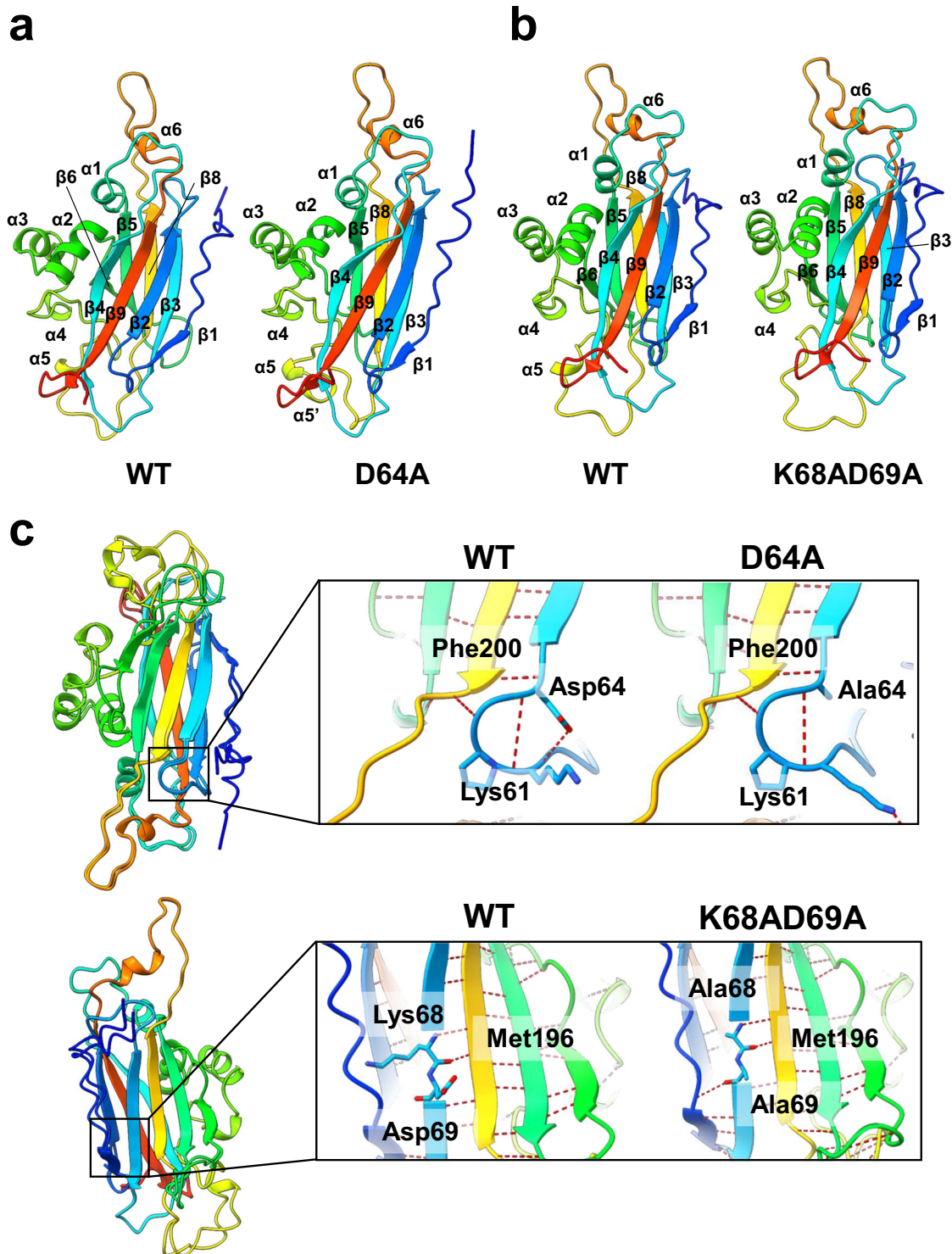


Fig. 5 Structure predictions and comparison of the WT TasA and variant proteins. **a** Structure comparison between the TasA model and the D64A allele. The numbering in the scheme follows the representation described for the crystal structure of TasA in its monomeric form published by Diehl et al.²⁰. The coloring scheme indicates the position of the residue in the sequence of the protein, where blue indicates the N-terminus and red indicates the C-terminus. Warmer colors indicate proximity to the C-terminal end. **b** Structure comparison between the TasA model and the K68A and D69A alleles. **c** Superimposition of the WT—D64A models (top) and WT—K68AD69A models (bottom) and a detailed view of the hydrogen bond analysis of the mutated proteins.

a different behavior. While CP transfer reveals rigid residues of the structural core, INEPT signals probe the highly mobile residues of the assembly. We derived the CP/INEPT signal ratio from ¹D ¹³C experiments (Fig. 7c), and we observed that the ratio was

comparable for WT and D64A samples (ratio ~2) but decreased for K68A and D69A samples (ratio ~1). The relative loss of CP signals for the K68AD69A samples suggests overall more dynamic assemblies, in line with the TEM observations.

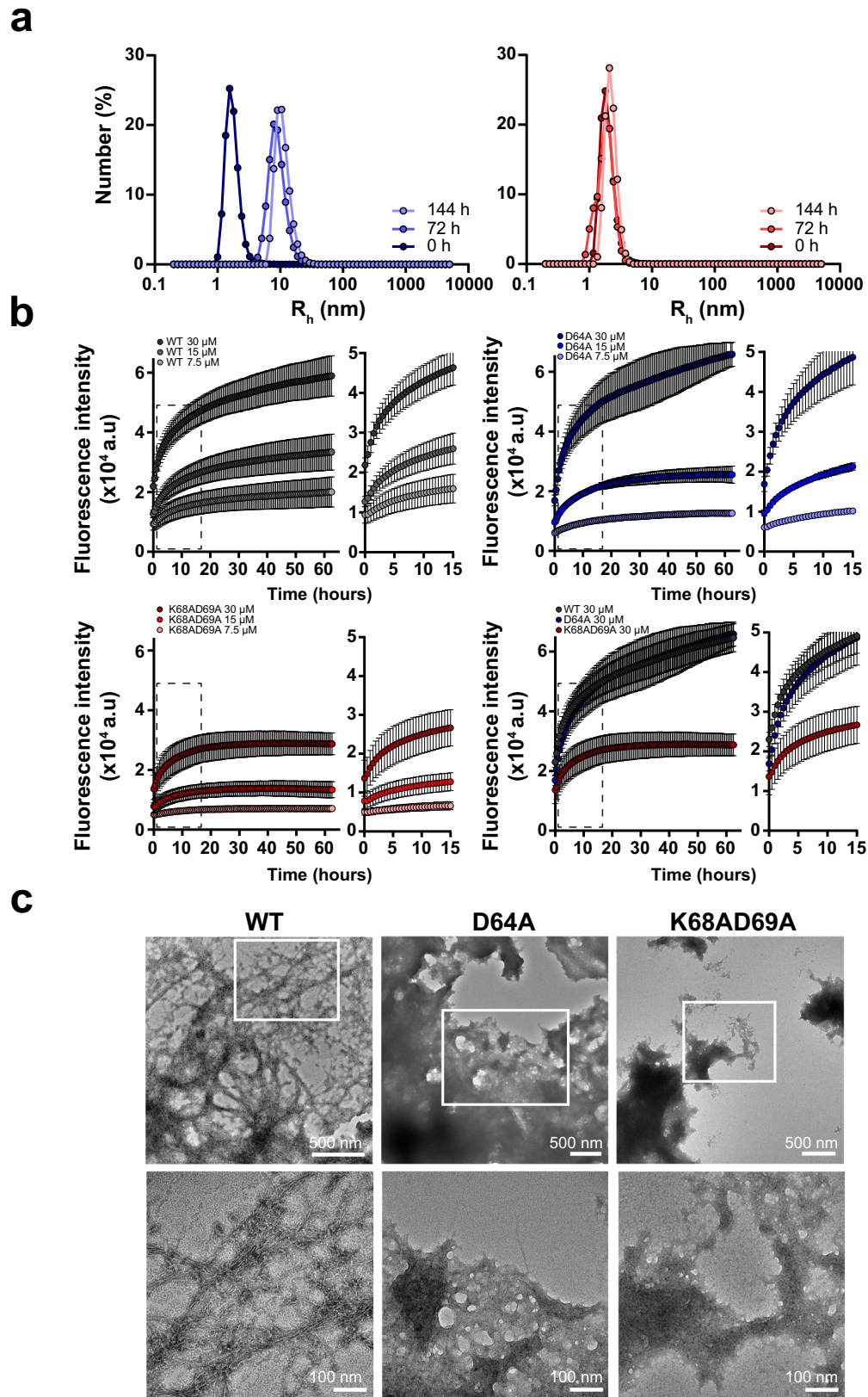


Fig. 6 **In vitro** analysis of the *TasA* variants reveals two types of molecular behaviors. **a** Aggregation kinetics of the D69A allele (left) and K68AD69A allele (right) as measured by DLS at different time points. **b** Kinetics of ThT fluorescence emission at 480 nm of the WT, D64A or K68AD69A alleles and superimposition of the plots corresponding to the 30 μ M protein concentration. Experiments were performed at different protein concentrations. Average values of three independent experiments are represented. Error bars indicate the SEM. **c** Transmission electron microscopy micrographs of negatively stained samples of the WT and *TasA* variants. White squares indicate areas of the pictures that have been zoomed in. Scale bars = 500 and 100 nm from top to bottom.

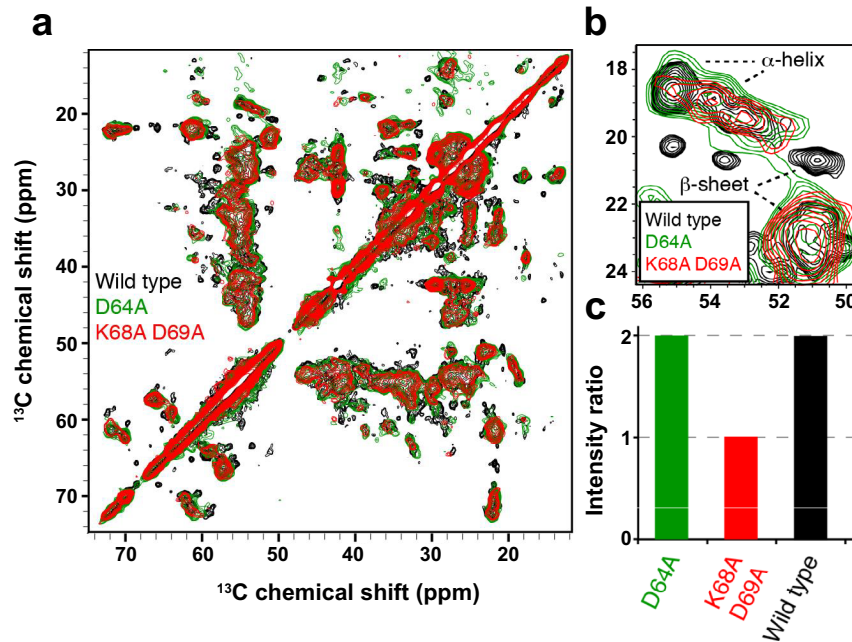


Fig. 7 Recombinant WT TasA, TasA D64A or TasA K68AD69A show similar structural fingerprints. **a** 2D ^{13}C - ^{13}C SSNMR experiments of recombinant TasA WT (black), TasA D64A (green) and TasA K68AD69A (red) filaments revealing the rigid residues. **b** Overlay of 2D SSNMR ^{13}C - ^{13}C spectra of TasA WT (black), TasA D64A (green) and TasA K68A D69A (red) fibrils in the alanine C α -C β spectral region. **c** CP/INEPT signal ratio from ^1D ^{13}C experiments in samples of the WT TasA, TasA D64A and TasA K68A and D69A fibrils.

In vivo characterization of the TasA variants demonstrates a complete loss of structural function in the K68AD69A allele

Previous experiments have demonstrated that the two amino acid substitutions introduced in the imperfect repeat region of RcTasA have an effect on morphology and biofilm formation and have revealed that (i) the D64A allele shows a molecular behavior similar to that of the WT protein, which is able to self-assemble with a tendency to form dense aggregates consisting of large bundles of ThT-stainable fibers that exhibit amyloid-like properties and (ii) the K68AD69A protein is unstable, exhibits limited amyloid-like properties and a tendency toward the formation of aggregates instead of fibers.

Despite the loss of functionality of both TasA variants compared to the WT protein, the strains expressing these two versions of the protein still formed a pellicle, although it was different from that of the WT strain or the strain expressing the native protein. One possible explanation for this contradictory finding is the compensatory structural activity mediated by EPS, the other major component of the ECM in *B. subtilis* biofilms. To test this hypothesis, the EPS operon was deleted in all the strains expressing the different versions of TasA. The biofilm formation phenotype of the strains carrying the native or D64A variant proteins resembled that of the $\Delta\text{epsA-O}$ strain alone, in which a weak and fragmented pellicle floated in the liquid medium (Fig. 8a). However, the phenotype of the strain expressing $\text{tasA}_{\text{K68AD69A}}$ mirrored that of the double ΔtasA , $\Delta\text{epsA-O}$ deletion strain, showing a complete absence of the pellicle (Fig. 8a right image). According to these biofilm phenotypes, cells of the $\Delta\text{epsA-O}$, $\Delta\text{epsA-O}$ $\text{tasA}_{\text{native}}$, and $\Delta\text{epsA-O}$ $\text{tasA}_{\text{D64A}}$ strains were uniformly decorated with a network of fibers that was morphologically indistinguishable between these strains (Fig. 8b). The identical growth dynamics of all the strains (Fig. 8c) and the estimated cell density (Fig. 8d) in MSgg led us to discard any growth defect of the strain expressing $\text{tasA}_{\text{K68AD69A}}$.

As previously reported²⁴, the mixture of ΔtasA and Δeps strains complement themselves when coinoculated in liquid medium, providing each other externally with their missing ECM components and restoring the wrinkled phenotype characteristic of *B. subtilis* biofilms. The same phenomenon takes place in solid medium, in

which the coinoculation of both strains in a 1:1 proportion results in a phenotype very different from the other two mutant strains, leading to the formation of wrinkles (Fig. 9, top images). The three $\Delta\text{epsA-O}$ strains expressing the different versions of TasA showed a phenotype that resembled that of the Δeps strain in solid medium (Fig. 9 middle row images). Restoration of the wrinkly phenotype was only achieved when ΔtasA was cocultured with $\Delta\text{epsA-O}$ carrying the $\text{tasA}_{\text{native}}$ or $\text{tasA}_{\text{D64A}}$ alleles (Fig. 9, bottom images), demonstrating that the biofilm-defective phenotype of the $\Delta\text{epsA-O}$ $\text{tasA}_{\text{K68AD69A}}$ strain is mostly due to a lack of structural functionality of RcTasA and the inability of this strain to assemble a functional ECM.

DISCUSSION

TasA has been shown to form non-amyloid fibers and ThT-stainable amyloid fibers under different experimental conditions^{15,16,20,21,23}. In this work we have characterized the N-terminal domain of TasA under experimental conditions that lead to the formation of ThT-stainable amyloid-like filaments. The N-terminal domain of TasA contains two aggregation-prone stretches and one sequence of imperfect amino acid repeats. Our study has proven that at least one of the predicted stretches found in RcTasA, LG-13, which exhibits amyloid-like properties, has the potential to contribute to the process of polymerization of TasA into ThT-stainable fibers and that some highly conserved amino acids within this imperfect repeat are potential determinants of the amyloid-like nature observed in the recombinant ThT-stainable fibers and the related structural functionality of TasA. Counter intuitive results were obtained when studying the amyloid stretch DG-14 from the RcTasA, as we observed no ThT binding and small molecular assemblies by TEM despite showing a ssNMR signature compatible with beta-rich protein entities (Fig. 3d). We speculate that the high and constant signal in the ThT binding experiment is related to the high scattering due to aggregation and precipitation of the peptide under these experimental conditions, which might indicate that the progression of the aggregation kinetics of this peptide could be arrested in a state that does not allow for the formation of large amyloid-like fibers. However, the solid material recovered from the

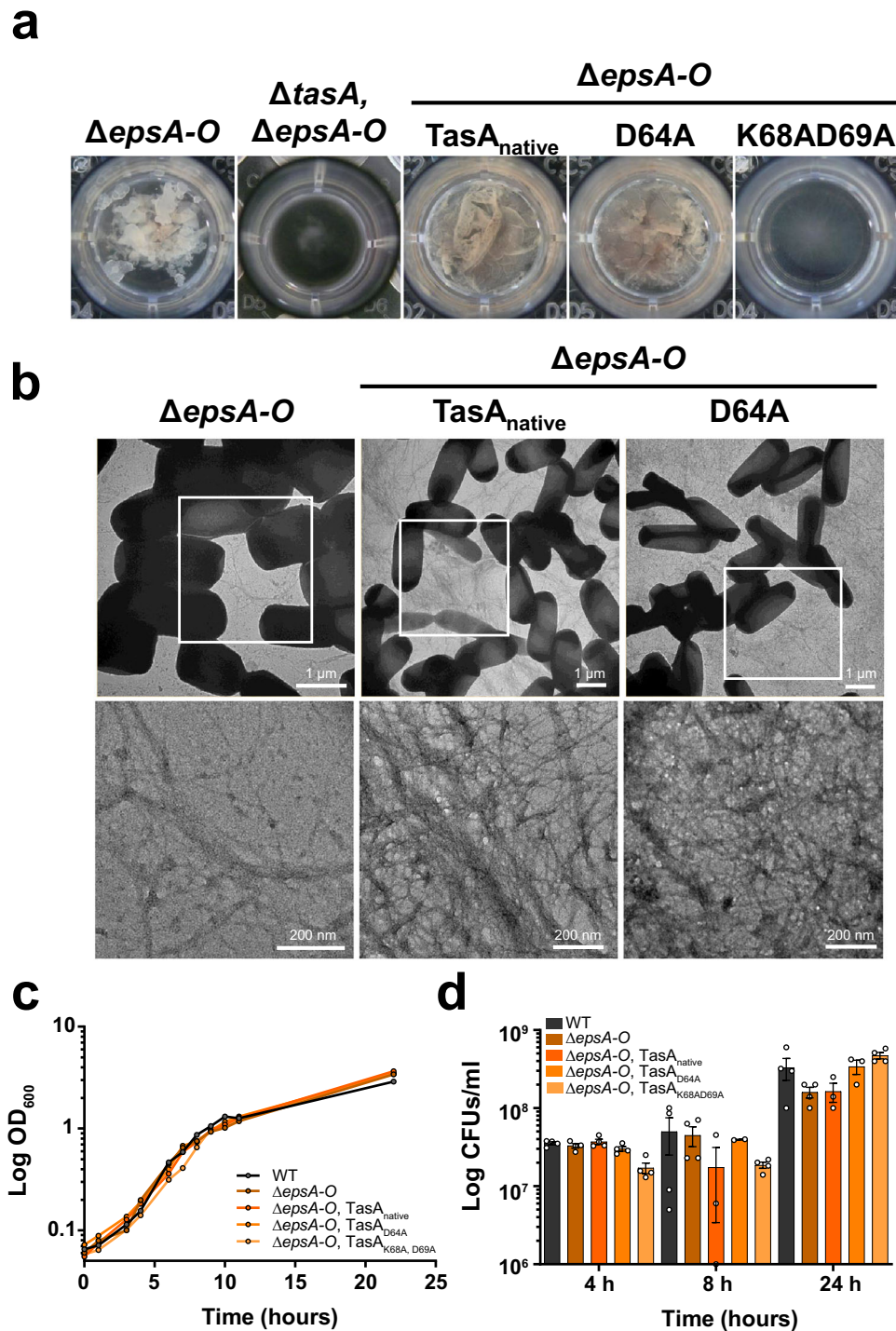


Fig. 8 **K68AD69A completely abrogated biofilm formation.** **a** The images show the biofilm formation phenotypes of the Δ epsA-O and Δ tasA Δ epsA-O strains, and Δ epsA-O strains carrying the different TasA variant alleles. **b** Transmission electron microscopy micrographs of negatively stained samples from the Δ epsA-O or Δ epsA-O strain carrying the native TasA allele or the D64A allele grown under biofilm-inducing conditions. White squares indicate areas of the images that have been zoomed in. Scale bars = 1 μ m (top) and 200 nm (bottom). **c** Growth curves of the WT, Δ epsA-O, and Δ epsA-O strains carrying the different TasA variant alleles in liquid MSgg medium. **d** Colony counts at different time points of the growth curve of the strains WT, Δ epsA-O, and Δ epsA-O carrying the different TasA variant alleles in liquid MSgg medium. Error bars indicate the SEM.

sample and used for ssNMR indicates that the sample is enriched in beta-sheets, showing a similar molecular signature as the other assemblies. Unfortunately, we were unable to characterize the TasA variants lacking these amyloidogenic regions, as their deletion resulted in the degradation of the protein. These regions are known

to be of utmost importance to explain the molecular behavior of amyloid proteins^{38,39}, and given the ability of amyloids to self-assemble through protein–protein interactions, these stretches serve novel approaches in synthetic biology and biotechnology, for example, to capture free amyloid peptides or proteins or even

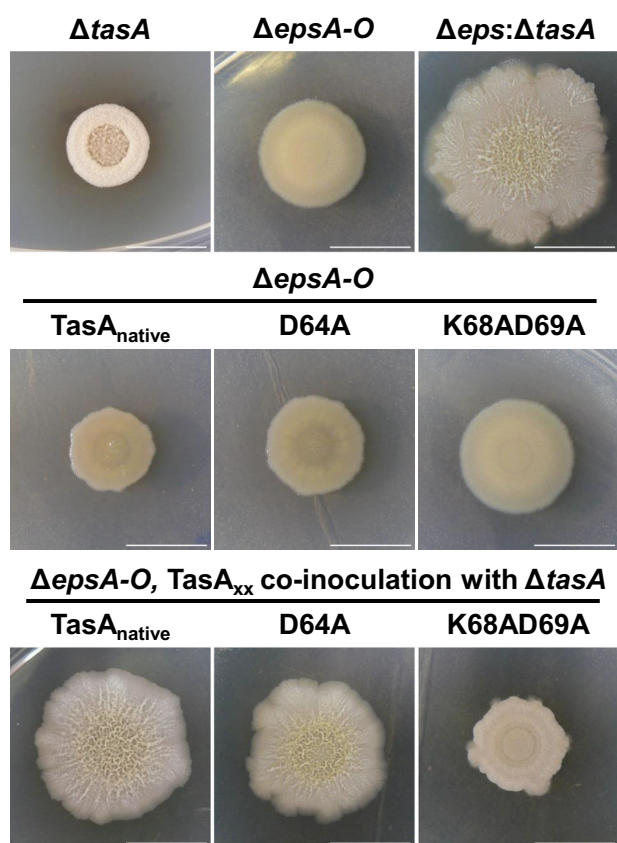


Fig. 9 The TasA D64A allele, but not the K68AD69A allele, is able to perform extracellular complementation when mixed with a *ΔtasA* strain. Colony morphology phenotypes of *ΔtasA* and *ΔepsA-O* strains, of the coinoculation between *ΔtasA* and *ΔepsA-O*, of the *ΔepsA-O* strains carrying the different TasA alleles and of the coinoculation of these strains with the *ΔtasA* strain. Scale bars = 1 cm.

bacteria themselves when exhibiting these sequences on their surface⁴⁰. The fact that these mutated proteins are eliminated to undetectable levels and the different phenotypes of strains expressing their alleles compared to the *ΔtasA* strain or any other ECM-defective strain suggest a relevant contribution to the native folding features of TasA.

The unique amino acid substitution D64A in RcTasA was sufficient to render a mutated protein that retained amyloid features but was incapable of contributing to biofilm formation. The fact that this variant is able to complement biofilm formation, at least partially, when it is produced in a Δ EPS background and coinoculated with a *ΔtasA* strain, suggests that when both components are produced in the same cell, they are not completely functional. Indeed, this is reversed when they are expressed in different cell populations. We speculate that the fibers formed by TasA D64A, although functional from a structural and biochemical point of view, modify the surface of the cells, altering the deposition of the EPS in the ECM or disturbing any other putative interaction between the filaments and the EPS. Such a phenomenon would not be unprecedented, given that similar interactions between functional amyloids and polysaccharides have been reported in other bacterial species. In biofilms of *Streptomyces coelicolor*, the fimbriae that mediate adhesion to surfaces are composed of bundles of amyloid fibers formed by chaplins, functional amyloids involved in reducing the water surface tension and promoting the development of aerial structures⁴¹. These fimbriae are maintained attached to the cell surface via their interaction with cellulose, which seems to serve as a template for the polymerization of the fimbriae⁴². Moreover, a more recent work has proposed that

coiled-coil domains in intermediate filament-like proteins have an intrinsic affinity for cellulose in different bacterial species⁴³, demonstrating a connection between specific domains in filamentous proteins and interactions with polysaccharides.

The substitutions K68AD69A, which have been characterized in this work, were previously studied in a work in which an additional function for TasA in cell physiology via its localization to the cell membrane was demonstrated²⁵. A strain expressing this version of TasA has a defective biofilm formation phenotype, which is disconnected from the cell membrane stability required for the proper physiological function of TasA²⁵. These amino acids are well conserved in the RcTasA region of imperfect repeats, and contrary to the other TasA variant investigated, their substitution affects the amyloid-like nature and stability of the entire protein, compromising its structural function in biofilm formation. Imperfect amino acid repeats are a common theme in many functional and nonfunctional amyloids^{9,10,27,36,44,45}, a reason for their use for the localization of important residues for protein fibrillation. Indeed, the alteration of these regions can render altered amyloid-like behavior, either inhibiting or improving the process of amyloidogenesis.

A recent work has provided insight into the molecular structure of native TasA filaments purified from *B. subtilis* biofilms with atomic resolution¹⁶. It was proposed that, as a monomer, the unstructured N-terminal end of the protein (amino acids from A28 to S41) along with the β 1 sheet undergoes a conformational change that results in these residues extending away from the monomer folding into another beta-sheet that complements the next subunit in a process known as donor strand complementation and the refolding of β 3. Based on this finding, and although native TasA filaments have previously been reported to have amyloid-like properties both *in vivo* and *in vitro*^{15,20}, the authors propose that the structural features of the native TasA filaments fit in the category of non canonical amyloid, as these fibers are neither a linear arrangement of globular subunits nor is their structure representative of the cross- β architecture typical of amyloid proteins. These results are supported by recent structural insight of the interaction of TapA with TasA, which demonstrated that non-ThT-stainable recombinant TasA filaments *in vitro* are formed by the intercalation of N-terminal peptide of TasA with the next subunit of the fiber maintaining the β -sandwich fold, similar to what occurs in type I pili, in which TapA should act as a chaperone and could potentially promote the growth of TasA filaments, presumably at the beginning of the filament, where results indicate that TapA could contribute to the donor strand complementation²³. Nonetheless, it has been previously demonstrated that recombinant TasA filaments *in vitro* can be amyloid, as indicated by the presence of a cross- β architecture in X-ray diffraction experiments^{14,20}, clearing any doubts regarding this question, at least concerning recombinant TasA, and despite the ability of the recombinant protein to also form non-amyloid fibers *in vitro*^{21,23}. These two different structural conformations of TasA filaments under different experimental conditions are not necessarily mutually exclusive, given that proteins can adopt different conformations depending on the environmental conditions, to fulfill specific functions and therefore, the amyloid filaments might contribute to any stage of the process of ECM assembly and biofilm formation. This is especially true for functional amyloids, which can exist in an amyloid and non-amyloid conformation depending on the surrounding conditions and the presence of other proteins or molecules. For instance, it is known that some intrinsic characteristics of the protein sequence such as the presence of conserved proline or glycine residues (which act as beta-breakers) or of the so-called “gatekeeper residues”, which are residues with low aggregation tendency (proline, lysine, arginine, glutamate and aspartate) in the border regions of aggregation-prone stretches can prevent the transition of some proteins toward the formation of amyloid aggregates and

fibers under specific physicochemical-conditions⁴⁶. Thus, it can be deduced that changes in the environment can trigger the amyloid conformation, especially in proteins with high aggregation propensity. This phenomenon has been characterized in the Curli fiber of *Escherichia coli*, a functional amyloid with the ability to form fibers important for biofilm architecture⁴⁷. The presence of aspartate or glycine residues in some of the imperfect repeats of the amyloid core of CsgA (the major protein component of the Curli fiber), important for the amyloid nucleation of the protein, partially inhibit its tendency toward aggregation as an attempt of modulating polymerization and thus, the potential toxicity of these protein aggregates⁴⁸. Another example of functional amyloid that exists in amyloid and non-amyloid conformation is the microcin E492 from *Klebsiella pneumoniae*, a toxic bacteriocin that form aggregates with antibacterial activity due to its pore-forming ability⁴⁹. However, these toxic aggregates are inactivated in stationary phase by polymerization into fibers with characteristic cross-beta architecture (in a process that is also controlled by posttranslational modifications⁵⁰), that are non-toxic and can act as a reservoir of toxic aggregates when the environmental conditions change⁵¹. Similarly, the listeriolysin O toxin from *Listeria monocytogenes* has the ability to transition from an active toxic form to an amyloid-like form depending on the pH conditions^{6,52}. In fact, TasA itself has been previously shown to adopt different conformations, amyloid and non-amyloid, upon interaction of the protein with different model membranes⁵³. Thus, it is tempting to speculate that TasA might be a functional "amyloidogenic" protein, in which the non-amyloid conformation of TasA may coexist with amyloid-like filaments, due to the intrinsic amyloidogenic nature of TasA associated at least to the sequence determinants analyzed in the present work, which might permit the transition of monomers to amyloid-like filaments under specific environmental conditions either within the biofilm or externally. This structural versatility is precisely one of the features frequently associated to functional amyloid proteins.

With the data that are currently available, the rigid core region of TasA analyzed in this work fits rather well in the proposed donor-strand complementation model in which the N-terminal domain of TasA is involved in donating the strand required for the assembly of the subunits and for the interactions that stabilize this strand. Our results analyzing the effect of these mutations in the structure of the native fibers suggest that amino acid D64 could be involved in stabilizing the interaction between different subunits. Thus, the mutation of this residue could lead to the formation of filaments that are stable enough so they are present in biofilms in vivo and exhibit a biochemical behavior that is similar to that of the WT protein, but structurally slightly different, failing to support the formation of the ECM. Amino acids K68 and D69 are located on the inner part of the rigid core as an integral part of the filament's β_3 and their substitution abrogates the formation of the filament, rather the mutated protein shows a tendency towards the formation of amorphous assemblies with less affinity for ThT (Fig. 6c right images and Fig. 8), demonstrating the essentiality of these two amino acids for the structural function of TasA.

Altogether, our findings, along with those from previously published works, lead us to a better understanding of the complex polymerization process of TasA during biofilm formation and provide knowledge into the sequence determinants that promote the molecular behavior of protein filaments in the ECM.

METHODS

Bacterial strains and culture conditions

The bacterial strains used in this study are listed in Supplementary Table 1. Bacterial cultures were grown at 37 °C from frozen stocks on Luria-Bertani (LB: 1% tryptone (Oxoid), 0.5% yeast extract

(Oxoid) and 0.5% NaCl) plates. Isolated bacteria were inoculated in appropriate medium. Biofilm assays were performed on MSgg medium: 100 mM morpholinepropane sulfonic acid (MOPS) (pH 7), 0.5% glycerol, 0.5% glutamate, 5 mM potassium phosphate (pH 7), 50 μ g/ml tryptophan, 50 μ g/ml phenylalanine, 50 μ g/ml threonine, 2 mM MgCl₂, 700 μ M CaCl₂, 50 μ M FeCl₃, 50 μ M MnCl₂, 2 μ M thiamine, and 1 μ M ZnCl₂. For cloning and plasmid replication, *Escherichia coli* DH5a was used. For protein purification, *Escherichia coli* BL21(DE3) was used. *Bacillus subtilis* 168 is a naturally competent domesticated strain used as a first step to transform the different constructs into *Bacillus subtilis* NCIB3610 by SPP1 phage-mediated generalized transduction. The final antibiotic concentrations for *B. subtilis* were MLS (1 μ g/ml erythromycin, 25 μ g/ml lincomycin), spectinomycin (100 μ g/ml), tetracycline (10 μ g/ml), chloramphenicol (5 μ g/ml), and kanamycin (10 μ g/ml). For the selection of plasmids in *E. coli*, ampicillin at 100 μ g/ml was used.

Plasmid and strain construction

Primers used in this study are listed in Supplementary Table 2. TasA was purified using the pDFR6 (pET22b-tasA) plasmid containing the ORF of TasA without the signal peptide or the stop codon, which was constructed with oligos TasAH-F (5'-aaaaaaaa-CATATGgcatttaacgacattaa-3') and TasAH-R (5'-aaaaaaaa-CTCGAGattttatcctcgctatgctgc-3') followed by digestion of the DNA fragment and the pET22b vector with enzymes NdeI and XhoI and ligation²⁴. To purify TasA_{D64A} and TasA_{K68AD69A}, the ORF of the mutated alleles excluding the signal peptide were amplified from strains JC78 and JC81 using the primers TasA_Exp_C_NdeI_F and TasA_Exp_C_XhoI_R. The resulting PCR products were then cloned into pET22b by digestion with NdeI and XhoI, and the final plasmids were maintained in *E. coli* (strains JC104 and JC106, respectively).

The region corresponding to the rigid core of TasA was amplified from the *tasA* ORF using primers TasA_Exp_C_NdeI_F and TasA_Exp_C_XhoI_R. The resulting PCR product was digested using NdeI and XhoI and cloned into pET22b. The final plasmid was maintained in *E. coli* (strain JC118).

Strains JC72, JC75, JC76, JC80 and JC82 were generated by site-directed mutagenesis using the primer pairs del82-88, del82-88-antisense, del108-116, del108-116-antisense KD_AA_35-36, KD_AA_35-36_antisense, F_A_72, F_A_72_antisense, E_A_82, E_A_82_antisense, G_A_96, G_A_96_antisense and D_A_64, D_A_64_antisense with a QuickChange Lightning Site Directed Mutagenesis Kit (Agilent Technologies), following the manufacturer's instructions.

All of the *B. subtilis* strains generated were constructed by transforming *B. subtilis* 168 via its natural competence and then using the positive clones as donors for transferring the constructs into *B. subtilis* NCIB3610 via generalized SPP1 phage transduction⁵⁴.

Bioinformatic analysis of the TasA sequence

For the prediction of amyloidogenic regions within the N-terminal region of the TasA protein sequence, we used different sequence-based tools that are freely available online and applied different computational methods⁵⁵. MetAmyl⁵⁶ and AmylPred 2.0⁵⁷ were used as consensus predictors, as both applications integrate the results provided by different online tools that predict protein aggregation or amyloidogenicity based on sequence analysis. APPNN⁵⁸ and FISH amyloid⁵⁹ were used as predictors based on machine-learning approaches. The consensus regions obtained from the different methods were selected for further experimental analyses.

The analysis of imperfect amino acid repeats present within the TasA sequence was performed using the webtool RADAR⁶⁰, available at the EMBL-EBI website.

The structure prediction of WT TasA, TasA_{D64A} and TasA_{K68AD69A} was performed using the crystal structure of the TasA monomer

(PDB ID: 5OF1) or the TasA fiber (PDB ID: 8AUR) from the Protein Data Bank as a template. Structures were predicted using AlphaFold⁶¹ in a Google Colab notebook (ColabFold)⁶². Protein models were visualized and analyzed using UCSF ChimeraX (v1.3)⁶³.

Biofilm formation assays

To analyze biofilm formation and the colony architecture on plates and pellicle formation in liquid medium, the different strains were grown in LB plates at 37 °C overnight. Then, suspensions of the corresponding strains were made in PBS, and these were adjusted to an OD₆₀₀ of 1. To analyze biofilm formation on the plates, 2 µl drops of the bacterial suspensions were spotted onto the surface of MSgg plates that were incubated at 30 °C. Pictures were taken at 24, 48 or 72 h to observe the colony architecture at different time points. To analyze the formation of pellicles in the air liquid interface of liquid media, 10 µl of each bacterial suspension was added to 1 ml of liquid MSgg medium in 24-well plates. Then, the plates were incubated at 30 °C without agitation or aeration for 48 h.

Biofilm fractionation

The presence of TasA was analyzed in the studied strains by fractionating the biofilms in medium, cells or ECM¹³. The fractionation was performed as follows: to have sufficient biological material, one 24-well plate was used for each bacterial strain, of which only the 8 central wells were used. Briefly, pellicles were carefully lifted from the wells and placed in a separate tube containing 8 ml of MS medium (MSgg without glycerol and glutamate). The remaining spent medium of the well was filtered sterilized and kept at 4 °C for further analysis. Then, to separate the cells from the ECM, the pellicles were subjected to mild sonication (approximately 10 pulses of 30 s at 20% amplitude in a Branson 450 digital sonifier, with 30 s pauses in ice), and the resulting suspensions were centrifuged at 9000 × g to separate the ECM from the cells. The supernatant (ECM fraction) was filtered, sterilized and kept at 4 °C. The cell fraction was resuspended in 8 ml of MS medium and kept at 4 °C until further processing.

Synthetic peptides

Peptides corresponding to the amyloidogenic regions detected within the TasA rigid core were synthetically produced and purified by Proteogenix (Schiltigheim, France). The sequence of the peptides was: for peptide LG-13 NH₂-LAIKEVMALNYG-COOH and for peptide DG-14 NH₂-DFLSQFEVTLTVG-COOH. A stock solution of each peptide was prepared in DMSO. Then, the peptides were diluted in Tris 20 mM and NaCl 50 mM to prepare the working solution for the different experiments at the appropriate concentration.

Protein precipitation

To precipitate the proteins of each biofilm fraction, 2 ml of the medium and ECM fractions were directly treated with 10% trichloroacetic acid (TCA) and incubated on ice for 1 h. For the cell fractions, 2 ml of the suspension was treated with 0.1 mg/ml lysozyme and incubated at 37 °C for 30 min. Then, the lysozyme-treated suspension was subjected to TCA treatment as described above. After incubation, proteins were collected by centrifugation at 13000 × g for 20 min at 4 °C. The pellets were washed twice with cold acetone and dried by air. The proteins were resolubilized in 8 M urea, 100 mM Tris, and 150 mM NaCl by incubating at 100 °C for 5 mins.

Protein expression and purification

E. coli BL21(DE3) cells were freshly transformed with the pET22b plasmid containing either WT TasA (pDFR6 plasmid), the different mutated versions of the protein (D64A, and K68AD69A) or the rigid core region. Colonies were selected from the plates, resuspended in 10 mL of LB with 100 µg/mL ampicillin and incubated overnight at

37 °C with shaking. This preinoculum was then used to inoculate 500 mL of LB + ampicillin, and the culture was incubated at 37 °C until an OD₆₀₀ of 0.7–0.8 was reached. Next, the culture was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated O/N at 30 °C with shaking to induce the formation of inclusion bodies. After that, cells were harvested by centrifugation (5000 × g, 15 min, 4 °C), resuspended in buffer A (Tris 50 mM, 150 mM NaCl, pH 8), and then centrifuged again. These pellets were stored frozen at –80 °C until use. After thawing, cells were resuspended in buffer A and broken down by sonication on ice using a Branson 450 digital sonifier (3 × 45 s, 60% amplitude). After sonication, the lysates were centrifuged (15,000 × g, 60 min, 4 °C), and the supernatant was discarded, as proteins were mainly expressed in inclusion bodies. The proteinaceous pellet was resuspended in buffer A supplemented with 2% Triton X-100, incubated at 37 °C with shaking for 20 min to further eliminate any remaining cell debris, and centrifuged (15,000 × g, 10 min, 4 °C). The pellet was then extensively washed with buffer A (37 °C, 2 h), centrifuged (15,000 × g for 10 min, 4 °C), resuspended in denaturing buffer (Tris 50 mM NaCl 500 mM, 6 M GuHCl), and incubated at 60 °C overnight to completely solubilize the inclusion bodies. Lysates were clarified via sonication on ice (3 × 45 s, 60% amplitude) and centrifugation (15,000 × g, 1 h, 16 °C) and were then passed through a 0.45-µm filter prior to affinity chromatography. Proteins were purified using an AKTA Start FPLC system (GE Healthcare). The lysates were loaded into a HisTrap HP 5 mL column (GE Healthcare) previously equilibrated with binding buffer (50 mM Tris, 0.5 M NaCl, 20 mM imidazole, 8 M urea, pH 8). Protein was eluted from the column with elution buffer (50 mM Tris, 0.5 M NaCl, 500 mM imidazole, 8 M urea, pH 8). After the affinity chromatography step, the buffer was exchanged with 1% acetic acid at pH 3 and 0.02% sodium azide by using a HiPrep 26/10 desalting column (GE Healthcare). This ensured that the proteins were maintained in their monomeric form. The purified proteins were stored under these conditions at 4 °C (maximum 1 month) until further use.

Assembly of TasA filaments

To assemble the TasA filaments in vitro from the corresponding monomeric purified proteins, first, the acidic buffer was removed and exchanged with Tris 20 mM NaCl 50 mM, pH 7, using FPLC and a HiPrep 26/10 desalting column as described above. Filaments were assembled with a protein concentration above 1 mg/ml and incubated at 30 °C in standing conditions for one week.

SDS–PAGE and western blot

Protein samples from the different assays were diluted in 2x Laemmli sample buffer (Bio-Rad) and heated at 100 °C for 5 min. Proteins were separated by SDS–PAGE in 12% acrylamide gels and then transferred onto PVDF membranes using a Trans-Blot Turbo Transfer System (Bio-Rad) and PVDF transfer packs (Bio-Rad). For immunodetection of TasA, the membranes were incubated with an anti-TasA antibody (rabbit) at a 1:20,000 dilution in Pierce Protein-Free (TBS) blocking buffer (Thermo Fisher). A secondary anti-rabbit IgG antibody conjugated to horseradish peroxidase (Bio-Rad #1706515) was used at a 1:3000 dilution in the same buffer. The membranes were developed using Pierce ECL Western blotting Substrate (Thermo Fisher).

Transmission electron microscopy

To visualize the fibers formed by the different TasA variants, the assembled protein samples were deposited over carbon-coated copper grids and incubated for 1 h. Next, the excess sample was blotted off the grid, and the samples were negatively stained by floating the grids in a 2% uranyl acetate solution for 30 s and then washed by floating the grid in distilled water for 30 s.

To visualize the filaments formed by the synthetic peptides corresponding to the amyloidogenic regions of the rigid core of TasA, the samples were taken from the microplates used in the thioflavin-T binding experiments after saturation of the signal (36 h of incubation) and were deposited in grids and negatively stained as described above.

To study the morphology of fibers on the surface of cells, pellicles of the different strains were grown as described above. After 48 h of incubation, carbon-coated copper grids were deposited into the wells over the fully formed pellicles and incubated for 2 h at 30 °C in standing conditions. After incubation, the grids were washed twice with distilled water for 30 s, negatively stained with 1% uranyl acetate for 20 s and washed again once with distilled water for 30 s.

All the samples were visualized in an FEI Tecnai G2 20 TWIN transmission electron microscope at an accelerating voltage of 80 kV. The images were taken using a side-mounted CCD Olympus Veleta with 2k x 2k Mp.

Thioflavin-T binding assays

To ensure that the assay is performed when TasA is in monomeric form, the purified proteins or the rigid core were buffer exchanged and immediately used afterward in 20 mM Tris and 50 mM NaCl pH 7. The assay was performed in 96-well microplates. TasA or the different variants were mixed at different final concentrations of 7.5, 15 or 30 µM with 30 µM thioflavin-T in a final volume of 200 µl. In the experiments with the rigid core region, the purified core was mixed at a final concentration of 15, 30 or 60 µM with thioflavin-T under the same conditions. Measurements were performed at 30 °C every 30 min in a Fluostar Omega plate reader (BMG Labtech) equipped with 450 nm and 480 nm filters for excitation and emission, respectively. Complete emission spectra of the RcTasA and TasA WT in the presence or absence of ThT were recorded after 40 h of ThT binding assay carried out in the same experimental conditions as described above. The excitation wavelength was set at 450 nm and fluorescence emission was recorded from 455 to 650 nm with 1 nm steps in a Edinburgh Instruments FLS920 fluorimeter.

For the experiments performed with peptide samples, working solutions of both peptides were prepared in 20 mM Tris and 50 mM NaCl, and the peptides were used at a final concentration of 200 µM.

Dynamic light scattering experiments

Freshly purified TasA, TasA rigid core or the different TasA variants were buffer exchanged to 50 mM Tris and 20 mM NaCl containing 0.02% sodium azide immediately prior to the experiment. The final concentration was adjusted to 30 µM for all the samples, and these were incubated at 30 °C in standing conditions. Samples were taken at different time points to measure particle size in a Zetasizer Nano ZS (Malvern) equipped with a 632.8 nm laser as the excitation source and using 1 cm path length polystyrene cuvettes. Zetasizer software version 8.01 (Malvern) was used for data analysis.

SSNMR spectroscopy

All experiments were recorded at 600 MHz on a Bruker spectrometer equipped with a triple resonance 3.2 mm probe using a magic-angle spinning frequency of 11 kHz. The sample temperature was set to 278 K, and chemical shifts were calibrated using DSS as an internal reference. A decoupling strength of 90 kHz was used with small phase incremental alternation with 64 steps (SPINAL-64). 1D ¹³C experiments on LG-13 and DG-14 were recorded using a cross-polarization contact time of 1 ms and 20k scans. 2D ¹³C-¹³C experiments were recorded using a proton-driven spin diffusion mixing time of 50 ms.

Partial proteinase-K digestion experiments

Experiments were performed as published elsewhere^{14,64}. Briefly, 50 µg of assembled WT TasA, the different variants or the rigid core were incubated for 1 h at 37 °C with 0.02 mg/ml proteinase K in 20 mM Tris pH 8 and 50 mM NaCl, and samples were taken at different time points for further analysis (1, 5, 15, 20, 30, 45 or 60 min). Reactions were stopped in each sample by the addition of 1 vol. Laemmli buffer and heated at 100 °C for 5 min. Then, samples were separated by SDS-PAGE and stained with Coomassie Brilliant Blue.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The data that supports the findings of this study are available within the manuscript and its supplementary materials. Additional data related to this work may be obtained from the corresponding author upon reasonable request. Protein models used for the structure prediction of the TasA variants in the manuscript were obtained from the Protein Data Bank (PDB ID: 5OF1 and 8AUR).

Received: 9 June 2023; Accepted: 12 September 2023;

Published online: 22 September 2023

REFERENCES

1. Yin, W., Wang, Y., Liu, L. & He, J. Biofilms: the microbial “protective clothing” in extreme environments. *Int J. Mol. Sci.* **20**, 3423 (2019).
2. Yan, J. & Bassler, B. L. Surviving as a community: Antibiotic tolerance and persistence in bacterial biofilms. *Cell Host Microbe* **26**, 15–21 (2019).
3. Gallardo, R., Ranson, N. A. & Radford, S. E. Amyloid structures: much more than just a cross-beta fold. *Curr. Opin. Struct. Biol.* **60**, 7–16 (2020).
4. Ke, P. C. et al. Half a century of amyloids: past, present and future. *Chem. Soc. Rev.* **49**, 5473–5509 (2020).
5. Hassan, M. N. et al. The amyloid state of proteins: A boon or bane. *Int J. Biol. Macromol.* **200**, 593–617 (2022).
6. Alvarez-Mena, A., Camara-Almirón, J., de Vicente, A. & Romero, D. Multifunctional amyloids in the biology of gram-positive bacteria. *Microorganisms* **8**, 2020 (2020).
7. Levkovich, S. A., Gazit, E. & Bar-Yosef, D. L. Two decades of studying functional amyloids in microorganisms. *Trends Microbiol.* **29**, 251–265 (2021).
8. Salinas, N., Povolotsky, T. L., Landau, M. & Kolodkin-Gal, I. Emerging roles of functional bacterial amyloids in gene regulation, toxicity, and immunomodulation. *Microbiol. Mol. Biol. Rev.* **85**, e00062–00020 (2020).
9. Szulc, N. et al. Variability of amyloid propensity in imperfect repeats of CsgA protein of *Salmonella enterica* and *Escherichia coli*. *Int. J. Mol. Sci.* **22**, 5127 (2021).
10. Rasmussen, C. B. et al. Imperfect repeats in the functional amyloid protein FapC reduce the tendency to fragment during fibrillation. *Protein Sci.* **28**, 633–642 (2019).
11. Knowles, T. P., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384–396 (2014).
12. Ulamec, S. M., Brockwell, D. J. & Radford, S. E. Looking beyond the core: The role of flanking regions in the aggregation of amyloidogenic peptides and proteins. *Front Neurosci.* **14**, 611285 (2020).
13. Branda, S. S., Chu, F., Kearns, D. B., Losick, R. & Kolter, R. A major protein component of the *Bacillus subtilis* biofilm matrix. *Mol. Microbiol.* **59**, 1229–1238 (2006).
14. El Mammeri, N. et al. Molecular architecture of bacterial amyloids in *Bacillus* biofilms. *FASEB J.* **33**, 12146–12163 (2019).
15. Romero, D., Aguilar, C., Losick, R. & Kolter, R. Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms. *Proc. Natl Acad. Sci. USA* **107**, 2230–2234 (2010).
16. Bohning, J. et al. Donor-strand exchange drives assembly of the TasA scaffold in *Bacillus subtilis* biofilms. *Nat. Commun.* **13**, 7082 (2022).
17. Chai, L. et al. Isolation, characterization, and aggregation of a structured bacterial matrix precursor. *J. Biol. Chem.* **288**, 17559–17568 (2013).
18. Azulay, D. N. et al. Colloidal-like aggregation of a functional amyloid protein. *Phys. Chem. Chem. Phys.* **22**, 23286–23294 (2020).
19. Ghayeb, M., Hayet, S., Lester-Zer, N., Levi-Kalisman, Y. & Chai, L. Fibrillar polymorphism of the bacterial extracellular matrix protein TasA. *Microorganisms* **9**, 529 (2021).
20. Diehl, A. et al. Structural changes of TasA in biofilm formation of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **115**, 3237–3242 (2018).

21. Erskine, E. et al. Formation of functional, non-amyloidogenic fibres by recombinant *Bacillus subtilis* TasA. *Mol. Microbiol.* **110**, 897–913 (2018).
22. Abbasi, R. et al. The bacterial extracellular matrix protein TapA is a two-domain partially disordered protein. *ChemBiochem* **20**, 355–359 (2019).
23. Roske, Y. et al. TapA acts as specific chaperone in TasA filament formation by strand complementation. *Proc. Natl. Acad. Sci. USA* **120**, e2217070120 (2023).
24. Romero, D., Vlamakis, H., Losick, R. & Kolter, R. An accessory protein required for anchoring and assembly of amyloid fibres in *B. subtilis* biofilms. *Mol. Microbiol.* **80**, 1155–1168 (2011).
25. Cámara-Almirón, J. et al. Dual functionality of the amyloid protein TasA in *Bacillus* physiology and fitness on the phylloplane. *Nat. Commun.* **11**, 1859 (2020).
26. Steinberg, N. et al. The extracellular matrix protein TasA is a developmental cue that maintains a motile subpopulation within *Bacillus subtilis* biofilms. *Sci. Signal* **13**, eaaw8905 (2020).
27. Harada, R. et al. The effect of amino acid substitution in the imperfect repeat sequences of alpha-synuclein on fibrillation. *Biochim Biophys. Acta* **1792**, 998–1003 (2009).
28. Wang, X. & Chapman, M. R. Sequence determinants of bacterial amyloid formation. *J. Mol. Biol.* **380**, 570–580 (2008).
29. Gade Malmos, K. et al. ThT 101: a primer on the use of thioflavin T to investigate amyloid formation. *Amyloid* **24**, 1–16 (2017).
30. Xue, C., Lin, T. Y., Chang, D. & Guo, Z. Thioflavin T as an amyloid dye: fibril quantification, optimal concentration and effect on aggregation. *R. Soc. Open Sci.* **4**, 160696 (2017).
31. Biancalana, M. & Koide, S. Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta* **1804**, 1405–1412 (2010).
32. Horvath, I. et al. Mechanisms of protein oligomerization: inhibitor of functional amyloids templates alpha-synuclein fibrillation. *J. Am. Chem. Soc.* **134**, 3439–3444 (2012).
33. Perov, S. et al. Structural insights into curli CsgA cross-beta fibril architecture inspire repurposing of anti-amyloid compounds as anti-biofilm agents. *PLoS Pathog.* **15**, e1007978 (2019).
34. Yang, W. et al. The propensity of the bacterial rodlin protein RdlB to form amyloid fibrils determines its function in *Streptomyces coelicolor*. *Sci. Rep.* **7**, 42867 (2017).
35. Kushnirov, V. V., Dergalev, A. A. & Alexandrov, A. I. Proteinase K resistant cores of prions and amyloids. *Prion* **14**, 11–19 (2020).
36. Cherny, I. et al. The formation of *Escherichia coli* curli amyloid fibrils is mediated by prion-like peptide repeats. *J. Mol. Biol.* **352**, 245–252 (2005).
37. Daskalov, A. et al. Structures of Pathological and Functional Amyloids and Prions, a Solid-State NMR Perspective. *Front. Mol. Neurosci.* **14**, 670513 (2021).
38. De La Paz, M. L. & Serrano, L. Sequence determinants of amyloid fibril formation. *Proc. Natl. Acad. Sci. USA* **101**, 87–92 (2004).
39. Ventura, S. et al. Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proc. Natl. Acad. Sci. USA* **101**, 7258–7263 (2004).
40. Vendrell-Fernandez, S., Lozano-Picazo, P., Cuadros-Sanchez, P., Tejero-Ojeda, M. M. & Giraldo, R. Conversion of the Ompf porin into a device to gather amyloids on the *E. coli* outer membrane. *ACS Synth. Biol.* **11**, 655–667 (2022).
41. Claessen, D. et al. A novel class of secreted hydrophobic proteins is involved in aerial hyphae formation in *Streptomyces coelicolor* by forming amyloid-like fibrils. *Genes Dev.* **17**, 1714–1726 (2003).
42. de Jong, W., Wosten, H. A., Dijkhuizen, L. & Claessen, D. Attachment of *Streptomyces coelicolor* is mediated by amyloid-like fimbriae that are anchored to the cell surface via cellulose. *Mol. Microbiol.* **73**, 1128–1140 (2009).
43. Soderholm, N., Javadi, A., Flores, I. S., Flardh, K. & Sandblad, L. Affinity to cellulose is a shared property among coiled-coil domains of intermediate filaments and prokaryotic intermediate filament-like proteins. *Sci. Rep.* **8**, 16524 (2018).
44. McGlinchey, R. P. et al. The repeat domain of the melanosome fibril protein Pmel17 forms the amyloid core promoting melanin synthesis. *Proc. Natl. Acad. Sci. USA* **106**, 13731–13736 (2009).
45. Daskalov, A., Dyrka, W. & Saupé, S. J. Theme and variations: evolutionary diversification of the HET-s functional amyloid motif. *Sci. Rep.* **5**, 12494 (2015).
46. Monsellier, E. & Chiti, F. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO Rep.* **8**, 737–742 (2007).
47. Van Gerven, N., Klein, R. D., Hultgren, S. J. & Remaut, H. Bacterial amyloid formation: structural insights into curli biogenesis. *Trends Microbiol.* **23**, 693–706 (2015).
48. Wang, X., Zhou, Y., Ren, J. J., Hammer, N. D. & Chapman, M. R. Gatekeeper residues in the major curlin subunit modulate bacterial amyloid fiber biogenesis. *Proc. Natl. Acad. Sci. USA* **107**, 163–168 (2010).
49. Bieler, S. et al. Amyloid formation modulates the biological activity of a bacterial protein. *J. Biol. Chem.* **280**, 26880–26885 (2005).
50. Marcoleta, A. et al. Microcin e492 amyloid formation is retarded by posttranslational modification. *J. Bacteriol.* **195**, 3995–4004 (2013).
51. Shahnavaz, M. & Soto, C. Microcin amyloid fibrils A are reservoir of toxic oligomeric species. *J. Biol. Chem.* **287**, 11665–11676 (2012).
52. Bavdek, A. et al. pH dependence of listeriolysin O aggregation and pore-forming ability. *FEBS J.* **279**, 126–141 (2012).
53. Malishev, R., Abbasi, R., Jelinek, R. & Chai, L. Bacterial model membranes reshape fibrillation of a functional amyloid protein. *Biochemistry* **57**, 5230–5238 (2018).
54. Yasbin, R. E. & Young, F. E. Transduction in *Bacillus subtilis* by bacteriophage SPP1. *J. Virol.* **14**, 1343–1348 (1974).
55. Santos, J., Pujols, J., Pallares, I., Iglesias, V. & Ventura, S. Computational prediction of protein aggregation: Advances in proteomics, conformation-specific algorithms and biotechnological applications. *Comput. Struct. Biotechnol. J.* **18**, 1403–1413 (2020).
56. Emily, M., Talvas, A. & Delamarque, C. MetAmyl: a META-predictor for AMYLOID proteins. *PLoS One* **8**, e79722 (2013).
57. Tsolis, A. C., Papandreou, N. C., Iconomidou, V. A. & Hamodrakas, S. J. A consensus method for the prediction of ‘aggregation-prone’ peptides in globular proteins. *PLoS One* **8**, e54175 (2013).
58. Familia, C., Dennison, S. R., Quintas, A. & Phoenix, D. A. Prediction of peptide and protein propensity for amyloid formation. *PLoS One* **10**, e0134679 (2015).
59. Gasior, P. & Kotulska, M. FISH Amyloid—a new method for finding amyloidogenic segments in proteins based on site specific co-occurrence of aminoacids. *BMC Bioinforma.* **15**, 54 (2014).
60. Heger, A. & Holm, L. Rapid automatic detection and alignment of repeats in protein sequences. *Proteins* **41**, 224–237 (2000).
61. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
62. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
63. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
64. Dos Reis, S. et al. The HET-s prion protein of the filamentous fungus *Podospora anserina* aggregates in vitro into amyloid-like fibrils. *J. Biol. Chem.* **277**, 5703–5706 (2002).

ACKNOWLEDGEMENTS

We thank Juan Félix López Téllez from the Nanoimaging Unit of the Plataforma Bionand, Instituto de Investigación Biomédica de Málaga (IBIMA), for his technical assistance in the transmission electron microscopy imaging and analysis. We thank José Luis Zafra Paredes and José María Montenegro Martos from the Vibrational Spectroscopy unit and the Electronic Spectroscopy unit, respectively, of the Servicios Centrales de Apoyo a la Investigación (SCAI) from the University of Málaga for their technical assistance with the circular dichroism and dynamic light scattering experiments. We thank David Vela Corcía for his help with the protein modeling and hydrogen bond analysis. We also thank Estelle Morvan, Axelle Grélard and the Biophysical and Structural Chemistry Platform at IECB, CNRS UAR 3033, INSERM US001, University of Bordeaux. This work was supported by grants from ERC Starting Grant (BacBio 637971), Plan Nacional de I+D+i of Ministerio de Ciencia e Innovación (PID2019-107724GB-I00), Proyectos I+D+I en el marco del Programa Operativo FEDER Andalucía (UMA18-FEDERJA-055) and Ayudas a Proyectos I+D+I Programa PAIDI 2020 of Junta de Andalucía (P20_00479).

AUTHOR CONTRIBUTIONS

D.R. conceived the study. D.R. and J.C.-A. designed the experiments. J.C.-A. performed the main experimental work. L.D.-G. performed the experiments with the RcTasA and provided assistance with the experimental work. A.L. and B.H. designed the SSNMR spectroscopy experiments. N.E.I.M. and A.L. carried out the SSNMR spectroscopy experiments. J.C.-A., D.R., A.L. and N.E.I.M. wrote the manuscript. L.D.-G., N.E.I.M., A.L., B.H., A.L. and A.d.V. contributed critically to writing the final version of the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41522-023-00437-w>.

Correspondence and requests for materials should be addressed to Diego Romero.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023