Research Article

Conservation and Specificity in *Bacillus* Biofilm Dynamics: On Structure and Function of *B. cereus* Camelysins

Anne Diehl [1], Florian Lindemann [1], Nils Cremer [1], Yvette Roske [2], Matthias Hiller [1], Barth van Rossum [1], Martina Leidert [1], Kürşad Turgay [3,4], and Hartmut Oschkinat [1,5,*]

*1 - **Leibniz Research Institute for Molecular Pharmacology,** Robert-Rössle-Str. 10, 13125 Berlin, Germany*
*2 - **Max-Delbrück-Center for Molecular Medicine,** Robert-Rössle-Str. 10, 13125 Berlin, Germany*
*3 - **Max Planck Unit for the Science of Pathogens,** Charitéplatz 1, 10117 Berlin, Germany*
*4 - **Institute of Microbiology,** Leibniz University Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany*
*5 - **Institute for Chemistry and Biochemistry,** Free University Berlin 14195 Berlin, Germany*

Correspondence to Hartmut Oschkinat: Hartmut Oschkinat, Robert-Rössle-Str. 10, 13125 Berlin, Germany. *oschkinat@fmp-berlin.de (H. Oschkinat)*
https://doi.org/10.1016/j.jmb.2026.169661
Editor: Sheena Radford

Abstract

The *B. cereus* family comprises members highly pathogenic for mammals or insects, with *B. anthracis* and *B. thuringiensis* respectively as notable examples. The biofilm operon of these bacteria encodes two TasA-like proteins, the 60% identical Camelysins CalY1 and CalY2. In this study, we observed that at neutral pH CalY2 alone polymerizes readily into filaments, whereas CalY1 forms a polydispersed mixture of oligomers without filament formation. However, at basic or acidic pH CalY1 also modestly polymerizes. CalY2 polymerization into filaments involves ß-sheet remodeling via donor strand complementation, as demonstrated here by a combination of NMR and AlphaFold studies. In contrast to TasA of *B. subtilis*, this process is spontaneous and does not require initiation by a TapA homolog. NMR studies show that the functionally relevant region (β1–β2–β3) of the CalY2 monomer structure closely resembles that of *B. subtilis* TasA, and differs from AlphaFold models. A survey of AlphaFold 2 predictions on 12 homologous *B. cereus* group Camelysins yielded only four correctly predicted β1–β2–β3 segments, which decreased to one when using AlphaFold 3. Since crucial residues in the protomer contact region are conserved among TasA-like proteins, we investigated whether family members of different species could form mixed filaments. NMR revealed features in CalY2 filaments that are structurally conserved with TasA filaments but sequentially different, promoting specificity. These interactions and differences, respectively, involve the C-terminus and the beginning of β3, which most likely hinder joint TasA and CalY1 copolymerization. A protease activity could not be observed for the heterologously expressed *B. cereus* Camelysins.
Significance: The *B. cereus* group includes extremely harmful and surprisingly benign bacterial strains. The Anthrax-toxin-producing *B. anthracis* is one of the most toxic bacterial threats to man, whereas *B. thuringiensis* toxin is used as a biological insecticide. Other *B. cereus* strains pose problems in food production and medical implant usage. These bacteria can exist as biofilms allowing them to survive and proliferate, an essential feature of which are protein filaments. Here we characterize the *B. cereus* Camelysins CalY1 and CalY2 and compare their structure and filament formation with *B. subtilis* filaments

to understand principles determining patterns of conservation and specificity. This investigation provides the basis for developing novel means to suppress or enhance biofilms with potential benefits for plant protection.

Classification Biological sciences;
Biochemistry;
Structural biology;

Introduction

Bacteria are able to form biofilms, sessile multicellular communities protected by a self-produced matrix [1–3]. The matrix is formed by secreted fibril- or filament-forming proteins, exopolysaccharides and eDNA. In general, biofilm formation can promote resistance against environmental stress, antimicrobials and other detrimental agents [1]. An established model for studying the structural and functional aspects of such biofilms and their development is the Gram-positive soil-dwelling *B. subtilis* [3,4]. Its secreted TasA protein forms filamentous matrix components [5,6] that are different from the amyloid fibrils observed in biofilms of, for example, *E. coli* or *Pseudomonas* [7–9]. TasA sequence alignments [10] ranging from archaea to fungi revealed the potential of the *B. subtilis*-related *B. cereus* group bacterial strains [11] to form biofilms containing filaments of the TasA-like proteins CalY1 and CalY2, known as Camelysins [11–13]. The group contains subspecies such as *B. cereus*,*B. anthracis* and *B. thuringiensis* which are known for both their pathogenic and biotechnological potential.

Since biofilms can cause problems in food production chains and in the handling of medical devices, the overall aim of this project was to better understand the formation of B. cereus biofilms and thus support practical applications to combat them. Biofilms on contact surfaces, such as stainless steel and plastic protect bacilli during cleaning and sanitization processes. Furthermore, the biotechnological applications of *B. thuringiensis* as an insecticide may also benefit from a better knowledge of the principles of biofilm assembly.

Here, we characterize structural and functional properties of the Camelysins CalY1 and CalY2 of *B. cereus*, whose genes are located in the biofilm cluster *sipW-caly1-bc1280-caly2*. Within this gene cluster, *sipW* codes for a signal peptidase, and the product of *bc1280* was described recently as Camelysin-assisting polymerization protein (CapP) [14]. CalY1 and CalY2 have 62% sequence identity and have been identified as surface-bound biofilm components containing a signal sequence of 27 amino acids and forming a mature protein of about 19 kDa [15,16] (Figure 1). Similar expression levels for CalY1 and CalY2 and the necessity to transfer the complete operon from *B. cereus* to *B. subtilis* to achieve a 'normal' biofilm *B. subtilis* phenotype were found by the group of Romero [17,18]. Caro Astorga et al. distinguished the Camelysins *via* their gene context, renaming the genes to *sipW-tasA (calY1)-bc1280-calY(calY2)* [17]. In earlier, initial publications [15,16] the camelysins were characterized as metalloproteases (MEROPS family M73 [19]. TasA of *B. subtilis* was considered a member of the tripartite M73 due to its sequence homology of 37 % and 40 % to CalY1 and CalY2, respectively. Subsequent work has characterized Camelysins as unfolded proteins when isolated from inclusion bodies, forming amyloidic precipitants at pH 3.0 [20].

It was shown recently that *B. Subtilis* TapA and TasA fold as monomers and can employ a donor-strand exchange mechanism to build filaments composed of complemented Ig domains, induced by TapA that contains a complete Ig fold [5,6]. Interestingly, protomer-protomer contact sites harbor the most strongly conserved residues in the alignment ranging from archaea to fungi [10]. In TasA filaments, TasA-specific interactions involve largely C-terminal residues beyond the essential glutamine Q232 that are poorly conserved among *Bacillus* strains. These C-termini are very diverse in length and sequence, with the *B. cereus* group Camelysins having a very short C-terminus of 6 residues after the strictly conserved glutamine (Q191 in CalY1 and Q189 in CalY2) whereas *B. subtilis* TasA has a very long C-terminus of 28 residues (Figure 1, caption), enabling a number of species-specific contributions to the binding interface.

Our work contributes to the structural basis for understanding filament formation by the *B. cereus* Camelysins. To this end, we use negative stain electron microscopy (EM), and study possible interactions with *B. subtilis* proteins by analytical ultracentrifugation (AUC). We further investigate the structure of the CalY2 monomer by solution NMR and the CalY2 filament structure by solid-state NMR. Investigation of the CalY2 monomer structure focuses on the orientation of strand beta1 that is not well predicted by AlphaFold [21]. We compare structures of monomers and filaments with those of the *B. subtilis* system [22] to delineate

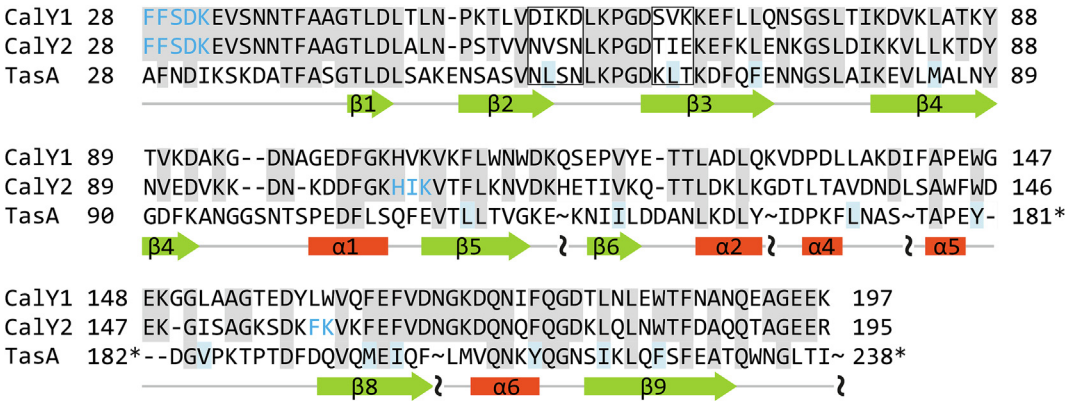


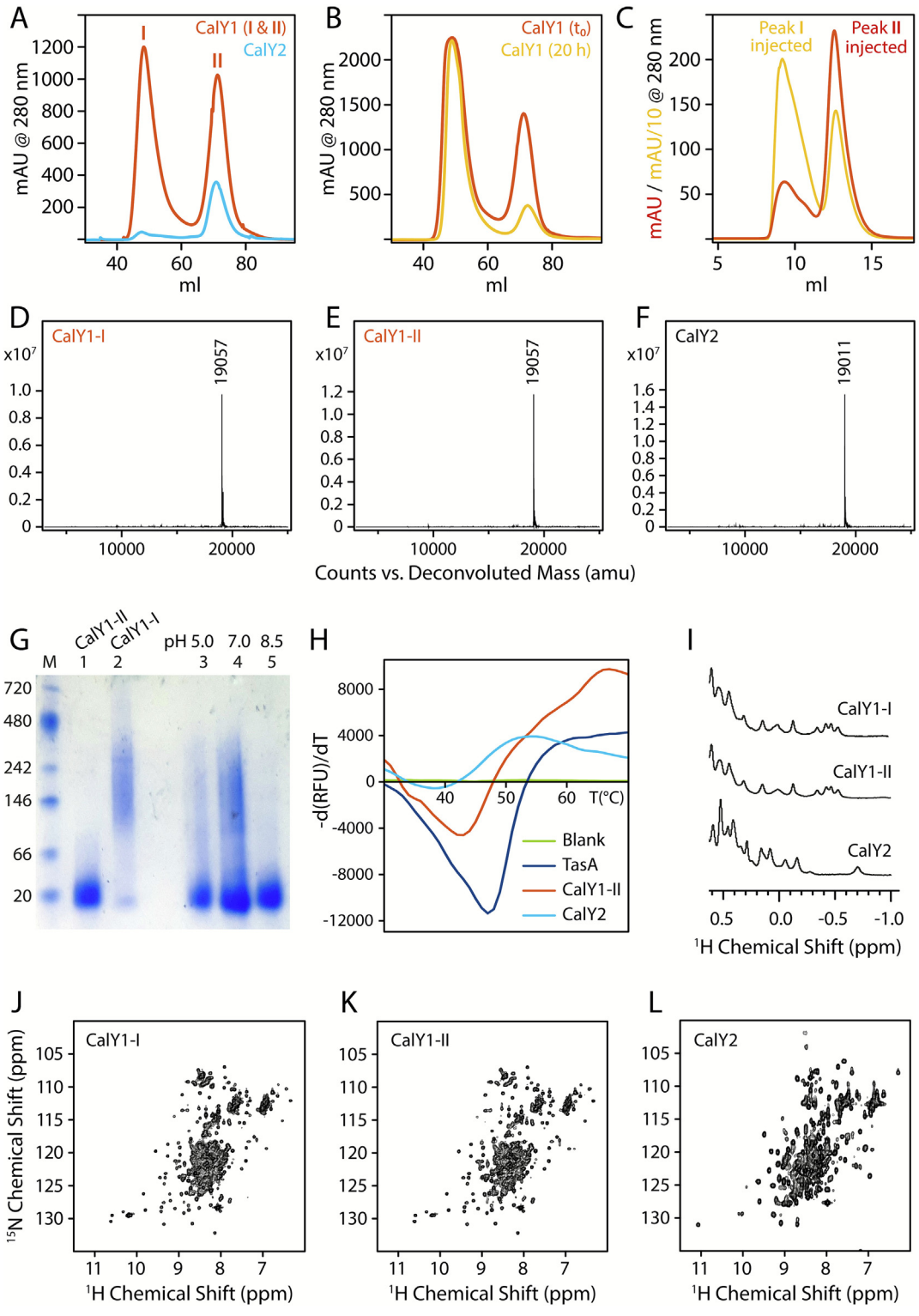
Figure 1. Amino acid sequence alignment of mature CalY1, CalY2 and TasA. The gray-shaded regions of the first two sequences are identical between CalY1 and CalY2. The TasA sequence is incomplete, only conserved areas are shown, interrupted by tildes. Numbers with a * refer to the first or last residue in the line. Segments of the TasA sequence conserved in the CalY sequence are also shaded gray, and conservative hydrophobic or aromatic replacements are indicated by blue background. The C-terminal residues of TasA ($_{239}$-KKDHTDKDGYVKENEKAHSEDKN$_{261}$) are not shown. The secondary structure below the alignment is derived from the TasA X-ray structure 5OF2. Blue letters indicate peptides not found in mass spectrometry analysis of the purified protein; black-lettered ranges of the sequence are covered. Within filaments, the non-conserved sequence areas that are indicated by black frames contribute specificity to filament formation (see text).

common principles and species- or even protein-specific aspects. Since bacteria form multi-species biofilms in nature, we further ask whether there is a possibility of developing mixed filaments by TasA-like proteins of different strains, or of the same strain, given the strict conservation of many interface residues. In a wider sense, this follows the hypothesis on “division of labor”, originally coined for processes inside a population of *B. subtilis* where cells may produce different biofilm components [23]. Hints on possible relevant interactions between *B. cereus* and *B. subtilis* are provided by Li et al [24], reporting that the joint application of both strains to farmed fish results in a curative effect on intestinal disorders, however, without discussing a possible molecular mechanism for this process.

Results

Recombinant expression and biochemical characterization of folded CalY1 and CalY2

In preparation for experiments to investigate the monomer structure, oligomerization and filament formation of the two Camelysins, we needed to express both proteins in mature form (sequences as in Figure 1) without any additional amino acids at the N- or C-terminus. To this end, recombinant soluble HisSumo_CalY1$_{28\text{-}197}$ and HisSumo_CalY2$_{28\text{-}195}$ were initially produced *via* a modular cloning system and vectors confirmed by sequencing. Initial purification by metal chelate chromatography (MC) was followed by tag cleavage, repetition of MC, and gel filtration. Under standard conditions with 150 mM NaCl, mature CalY1 eluted in two peaks (Figure 2A, red curve). The first peak (peak I) eluted shortly after the void volume of the column with a 10-fold apparent mass compared to the second peak (peak II), representing a monomer. In comparison, CalY2 appeared mainly in one peak that corresponds to a monomer (Figure 2A, light blue curve). The amount of CalY1 represented in peak I increased with time after elution, even at 8 °C (Figure 2B), as determined by two gel filtration runs, one performed immediately (red curve), one with a time lag of about 20 h (yellow curve). Re-chromatography of peak I and peak II on an analytical Superdex 75 column again yielded the same two species when run in both 150 mM NaCl (Figure 2C) and 300 mM NaCl. The identity of the proteins was verified by denatured LC/MS for CalY1 peak I (CalY1-I), peak II (CalY1-II) and CalY2, confirming the theoretical masses of 19,057 Da both for CalY-1 I and II and 19,010 Da for CalY2 (Figs. 2D-F). Checking protein identity by tryptic digest followed by MS analysis yielded more than 90% of sequence coverage for both CalY1 and CalY2. The non-detected peptides are indicated in Figure 1 by blue coloring of the residue codes. Given the structural integrity of CalY1 in peak I and II, we ran native PAGE (Figure 2G) of each peak fraction directly after elution from the gel filtration column. For peak I, we observed a polydispersed mixture of oligomers in the range of 100–400 kDa (Figure 2G, lane 2) together with a very small portion of monomers (~20 kDa). A defined, but broad band was observed for the protein eluting in peak II (Figure 2G, lane 1), corresponding to a monomer given the high


A
CalY1 (I & II)
CalY2
I
II
mAU @ 280 nm
ml
B
CalY1 (t0)
CalY1 (20 h)
C
Peak I injected
Peak II injected
mAU / mAU/10 @ 280 nm
D
CalY1-I
19057
E
CalY1-II
19057
F
CalY2
19011
Counts vs. Deconvoluted Mass (amu)
G
M
CalY1-II
CalY1-I
pH 5.0 7.0 8.5
H
-d(RFU)/dT
T(°C)
Blank
TasA
CalY1-II
CalY2
I
CalY1-I
CalY1-II
CalY2
1H Chemical Shift (ppm)
J
CalY1-I
15N Chemical Shift (ppm)
1H Chemical Shift (ppm)
K
CalY1-II
L
CalY2

amount loaded. Thermal shift experiments (Figure 2H) revealed melting points of 42 °C for CalY1-II and of 38 °C for CalY2, considerably lower than those observed for TasA (46 °C). We also analyzed the folding of both proteins by NMR (Figure 2I–L). Solution NMR is typically biased towards monomers and smaller oligomers since the transverse relaxation time increases with size. Signals of large oligomers are often broadened beyond recognition. In 1D ^{1}H NMR spectra, the region with negative chemical shifts shows only narrow signals if the protein is folded, and their line width gives a rough estimate of molecular size. Figure 2I shows two highly similar ^{1}H NMR spectra for CalY1-I and CalY1-II, where the apparent line width of the peaks is in agreement with monomeric protein, see for instance the separated signals between −0.1 and −0.2 ppm. As observed in Figure 2C and G, lanes 1 and 2, monomers form over time so that a monomer-like spectrum is observed for CalY1-I. The line widths of the CalY2 signals are similar, but a different signal pattern is observed. All three ^{1}H NMR spectra are characteristic of a folded protein as also apparent from their fingerprint in 2D ^{1}H–^{15}N correlations (Figure 2J–L). There, a considerable number of signals is observed in the ^{1}H spectral region between 8.8 and 11.0 ppm, indicating a large fraction of β-structure. The 2D spectra corroborate the presence of folded monomers in both CalY1-I and CalY1-II after storage of the samples.

In first reports on a possible functional role of Camelysins, protease activity was attributed to protein isolated from *B. cereus* cultures [15,16]. However, in recent investigations such activity was not found for *B. thuringiensis* Camelysin [27]. This led us to test for proteolytic activity of our recombinant CalY1-II and CalY2 by several independent assays, but we failed to find any (Supporting Text and Figures S2–S6).

Camelysin polymerization and interactions with TapA and TasA

The behavior of the two TasA-like proteins CalY1 and CalY2 is very different despite ∼60% sequence identity, which is already apparent during purification and sample handling. At concentrations >250 μM (ca. 5 mg/mL) CalY2 solutions containing 150 mM NaCl readily acquired gel-like properties. At higher salt concentration (300 mM NaCl) a 150 μM CalY2 solution could be further purified by size exclusion chromatography for following AUC experiments (Figure S1A). There, we used a salt concentration of 300 mM NaCl and a CalY2 concentration of 50 μM, observing that the protein remained in a monomeric state (Figure S1A). In negative-stain EM of CalY2 filaments, network-like structures were observed as shown in Figure 3A. In the contrary, CalY1 does not show a tendency to form polymers at neutral pH. A different picture was obtained at pH settings of 5.0 and 8.5, adjusted by dialysis. After ultracentrifugation of the pH 5.0 solution and resuspension of the pellet, extended polymer networks (Figure S1B) were observed by negative stain EM, whereas the control sample at neutral pH showed none (Figure S1C). At pH 8.5 similar but somewhat broader filaments, and precipitated protein were observed (Figure S1D). Samples of all three pH settings, taken before ultracentrifugation, yielded distinct profiles in a native gel (Figure 2G). Lanes 3–5 show all a large portion of monomeric CalY1-II, whereas higher oligomers of 100–300 kDa are evident only for the samples at pH 5.0 and 7.0 (lanes 3 and 4, respectively). The oligomer size distribution and the ratio to the monomer are comparable between pH 5.0 and pH 7.0, taking into account that twice as much protein at pH 7.0 was applied. At pH 8.5, with a comparable concentration as for pH 5.0, no higher oligomers can be observed. ThT assays of

Figure 2. Biophysical characterization of CalY1 and CalY2. **A** Gel filtration elution profiles of CalY1 and CalY2. For CalY1 two similar sized peaks are obtained (orange curve), CalY2 elutes mainly in one peak (light blue curve). **B** Gel filtration elution profiles of CalY1 directly after the metal chelate column (orange curve) and after a time of 20 h (yellow curve), indicating an increase in large oligomeric species. **C** Analytical gel filtration elution profiles of reinjected CalY1 peaks I and II, indicating the interconversion of forms. Both samples were stored in the freezer between the first and the second runs. **D-F** Molecular masses of the three peaks observed in the gel filtration profiles derived by denaturing mass spectrometry. CalY1 possesses a calculated mass of 19,057 Da, here observed for both samples (peak I and II in **A**). CalY2 has a mass of 19,011 Da, here 19,010 is observed. **G** Native PAGE of CalY1-II (lane 1), CalY1-I (lane 2) directly after gel filtration. Lanes 3, 5, and 4 show the profiles of CalY1-II subjected to dialysis against buffer with pH 5.0 and 8.5, and the non-dialyzed, neutral sample, respectively. M = Marker with indicated masses. In this representation of the gel, the aspect ratio was changed and its contrast increased. **H** Thermal shift assays of TasA (blue curve) and CalY1-II (orange curve), each applied as 50 μM solution, and of CalY2 (light blue curve) in a 100 μM concentration. For melting points see main text. **I** Methyl region (−1.0 to 0.5 ppm) of ^{1}H NMR spectra. Both CalY1 samples were stored in the freezer before assaying. **J–L** Two-dimensional ^{1}H–^{15}N correlation spectra exhibiting the fingerprints of CalY1 and CalY2. Note the high similarity of CalY1-I and CalY1-II spectra after freezing prior to experiments.

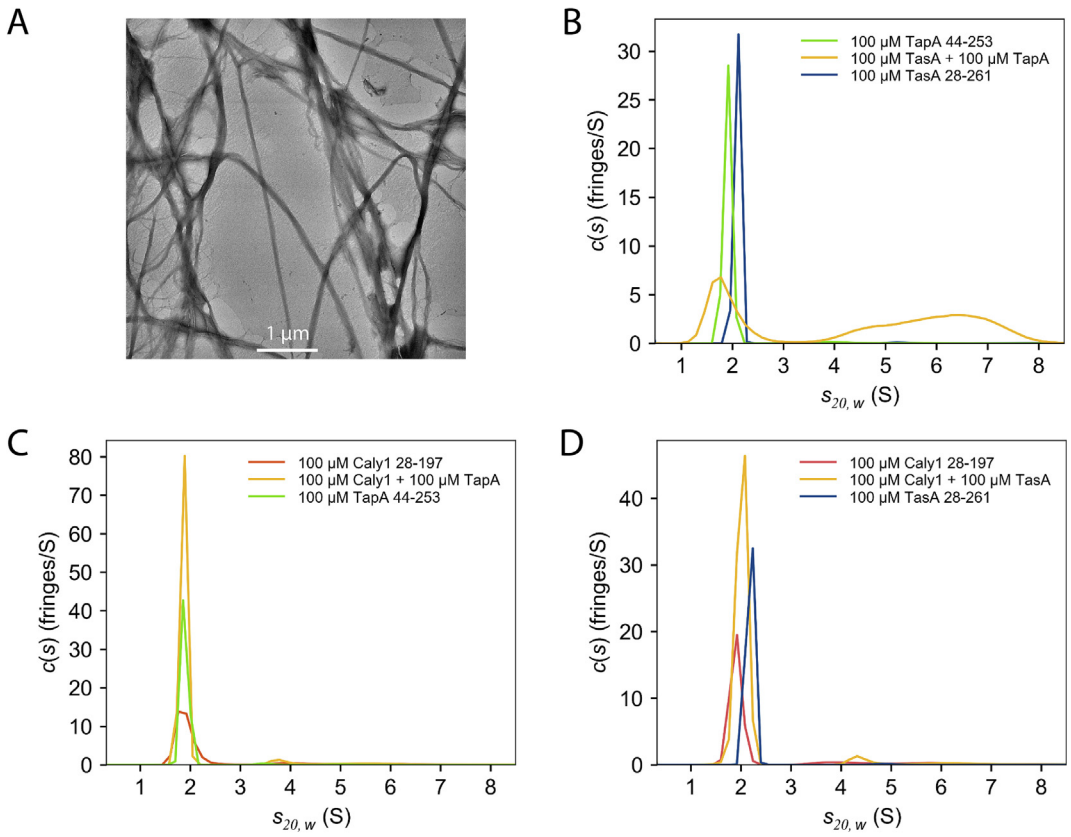


Figure 3. Analytical ultracentrifugation (AUC). **A** Negative stain EM micrograph of filaments formed spontaneously by CalY2. **B** AUC of TasA and TapA separately and in a 1:1 mixture. As individual proteins, they remain monomeric, and together they form filaments with a size distribution from 4 to 8 S at 20 °C. The experiment was repeated for comparison (see Roske et al. 2023 [11] **C** CalY1 and TapA alone remain monomeric, with CalY1 showing a fraction of oligomers (see Figure 2A and G). In combination the monomers remain stable and no significant filament formation occurs. (conditions as in **B**). **D** CalY1 and TasA remain monomeric, also in combination, and no significant filament formation occurs (conditions as in **B**).

the CalY1 filaments occurring at pH 5.0 and 8.5 did not yield an amyloid-like response suggesting that non-amyloidic filaments are formed (Figure S1E). Altogether, these results indicate very different association behaviors of CalY1 and CalY2, with CalY2 spontaneously polymerizing at neutral pH whereas CalY1 does not. However, for CalY1 at neutral pH, we observe the presence of a polydispersed equilibrium of monomeric and oligomeric forms that do not polymerize and are thus expected to be structurally different from filaments.

To develop a strategy for further testing the propensity of CalY1 to form polymers, we considered the high conservation of residues in the protomer-protomer interface of TasA-like proteins [5,6] and investigated the possibility of cross-strain interactions. This enabled us to make use of the comprehensive characterization of *B. subtilis* TapA and TasA, which have two definite tasks in biofilm formation as initiator and polymer-forming proteins, respectively. By applying our previously successful AUC experiment, we investigated the possibility of whether TapA enhances CalY1 oligomerization and whether CalY1 induces polymerization of TasA. In Figure 3B, a reproduction of the AUC experiment [10] revealing the interaction of mature *B. subtilis* TasA and TapA and the concomitant formation of a large species is shown to enable comparisons. Since TapA induced TasA oligomerization, and the most N-terminal residues of TasA are conserved in CalY1, we tried to achieve a similar effect by subjecting CalY1 mixed with TapA to AUC (Figure 3C). However, no broad distribution of oligomers was obtained as in Figure 3B. Similarly, CalY1 was mixed with TasA, but again no large concentrations of multimers were observed (Figure 3D). Figures 3C and D show small populations of oligomers at an *S* value of around 3.8 that could well be associated with the previously observed oligomeric form of CalY1 (CalY1 I, Figure 2A and G) that appears here in small quantities.

Since the processes underlying biofilm formation may be more complex, requiring additional factors, we extended our search for conditions of CalY1 polymerization to *in-vivo* detection of

possible cross-interactions with *B. subtilis* proteins. These experiments follow earlier genetic experiments introducing *B. cereus* gene clusters into *B. subtilis* [17]. At the same time, this complementation assay represents another test for cross-species interactions in biofilms. For this purpose, we employed live *B. subtilis* cultures and probed for changes in pellicle phenotype. MOLP [25] and mineral salt medium with glycerol and glutamate (MSGG) [26] promote wrinkles on a *B. subtilis* biofilm. Utilizing such MOLP medium, the degree of wrinkle formation may be taken as an indicator for the extent of the produced biofilm. However, *B. cereus* shows more or less a skin-like appearance on the surface of a static culture and not the same kind of wrinkles. In our experiment, we investigated complementation effects of CalY1 and CalY2 on biofilms of *ΔtasA B. subtilis*. After 46 h (Figure S7), the biofilm of wildtype *B. subtilis* is clearly wrinkled in contrast to a culture of the *ΔtasA B. subtilis* strain [27], highlighting the importance of TasA as a major extracellular matrix protein. Supplementation of the *ΔtasA* setup with 5 μM *B. subtilis* TasA rescues the biofilm (Figure S2, well S2 and S3), whereas the supplementation of CalY1 and CalY2 alone or in combination does not recover the intense wrinkling observed in the case of the wildtype (wells 4, 5 and 6). This suggests that CalY1 and CalY2 are not significantly promoting *B. subtilis* biofilm formation.

Comparison of CalY1 and CalY2 monomer models with the TasA X-ray structure

Structural features of both *B. cereus* Camelysins in their monomeric form were analyzed in the following section by inspecting AlphaFold 2 predictions [29,30] and homology models [10], the latter based on the X-ray structure of the *B. subtilis* TasA monomer (PDB: 5OF1), as well as *via* an NMR investigation. NMR was applied to determine the position and orientation of β1 with respect to β3 in the CalY2 monomer, since there are structural differences between the experimentally determined TasA monomer structure and the AlphaFold predictions in the conserved and functionally relevant β1–β2–β3 region.

The *B. cereus* Camelysin sequences are much shorter than those of *B. subtilis* TasA, hence the secondary structure in the CalY1 and CalY2 AlphaFold 2 and homology models are obviously different from the TasA X-ray structure as a result of the sequence deletions (Figure S7). As a consequence, a much shorter loop region between Camelysin strands β7 and β8 (Figure S7B and C) is observed, and helix α6 of TasA is missing entirely (Figure S7A, small ellipsoid) in comparison to the TasA X-ray structure. The largest structural difference occurs between α2 and strand β8 (of TasA) where the positions of helices is different, and an equivalent of α3 and β7 of TasA is missing (Figures S7 and S8A, B, large ellipsoids in each). The superimposition of the modeled structures of *B. cereus* CalY1 and Caly2 and the Alphafold 2 predictions show considerable differences in this region (Figure S8C and D). Overall, a superposition of AlphaFold 2 CalY1 and CalY2 structures with the modeled ones revealed a root mean square deviation (rmsd) of 3.8 and 3.5 Å (aligning only 135 residues, mostly β-sheet), respectively. A similar rmsd of 3.8 Å was observed for the overlay of TasA onto AlphaFold 2 CalY2. Both of the AlphaFold CalY structures are very similar, yet there are differences in the region of α3 (Figure S9A), whereas the homology models

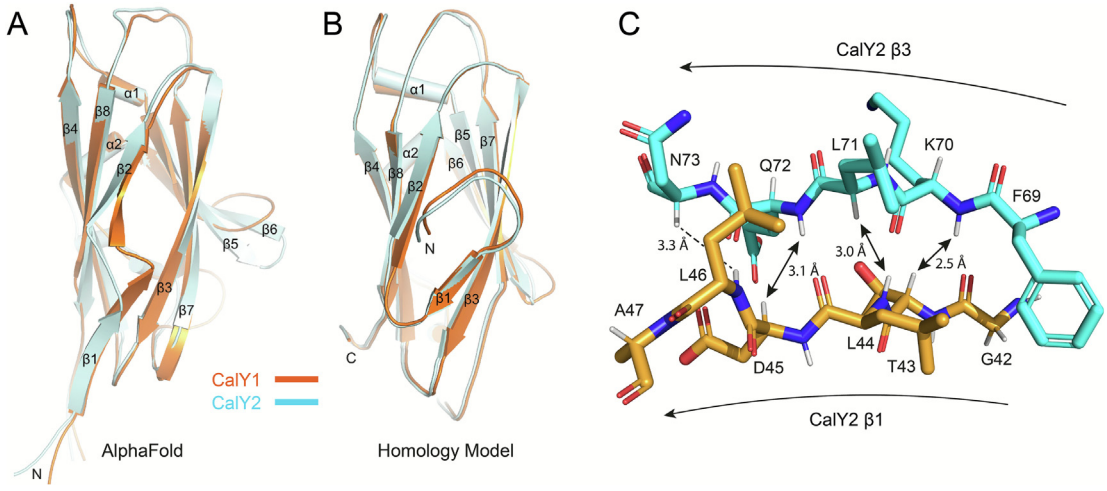


Figure 4. Structural comparison of CalY1 (light blue) and CalY2 (orange) models. **A** Superimposition of AlphaFold 2 models. **B** Overlay of homology models based on the TasA X-ray structure (pdb code: 5OF1). **C** Interactions between CalY2 strands β1 and β3 detected as cross-strand NOE between H_N and $H_{C\alpha}$ in a 3D ^{15}N–^{1}H–^{1}H NOESY NMR spectrum (solid arrows). The dotted line indicates a close contact that was not observed as a cross peak in the 3D spectrum.

based on TasA appear more homogeneous (Figure S9B).

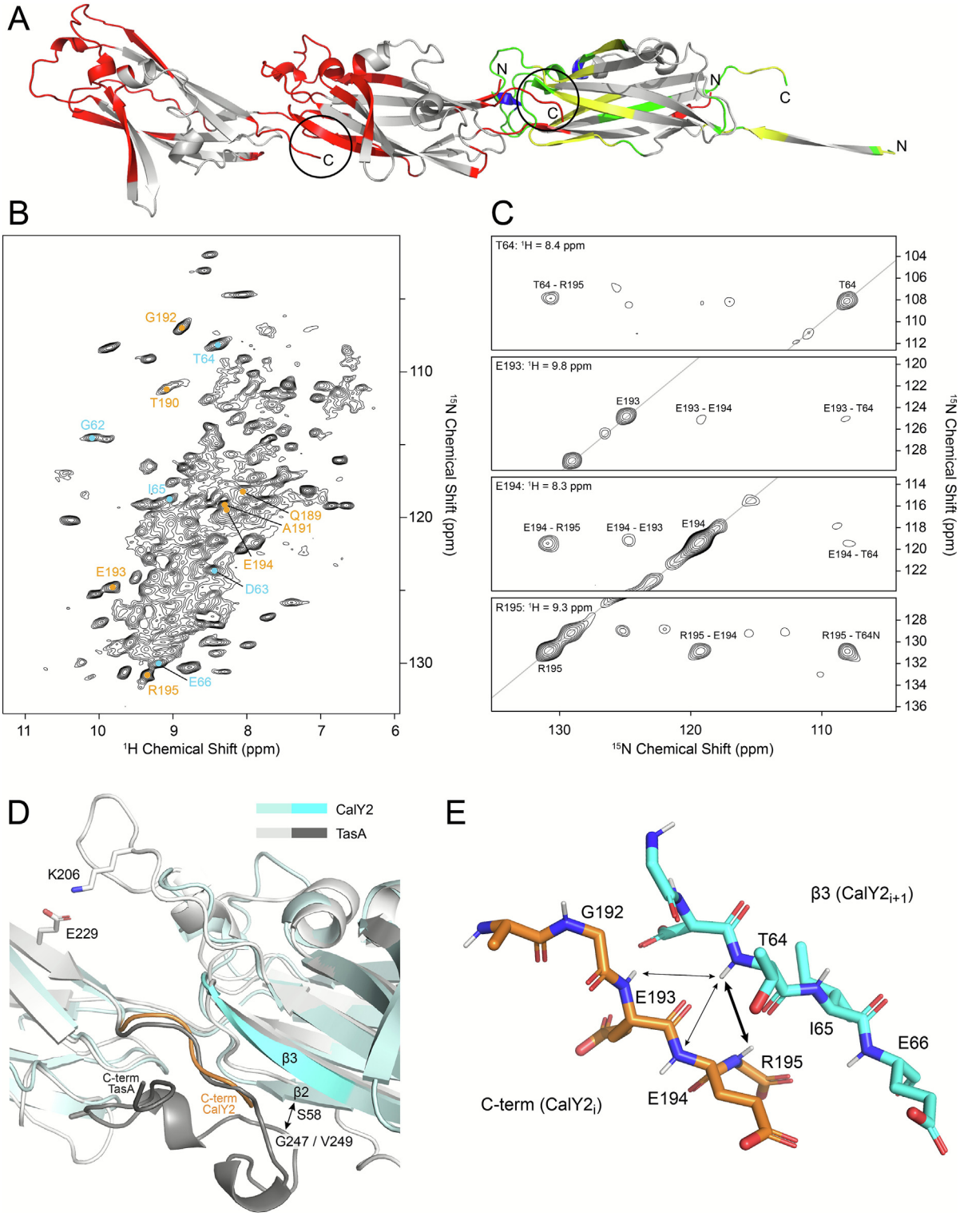
The most remarkable difference between the AlphaFold 2 predictions shown in Figure 4A and those modeled on the basis of TasA (Figure 4B) concerns the functional β1 (comprising the strictly conserved residues $_{42}$GTLD$_{45}$), which is parallel to β3 in the X-ray structure of TasA. However, in most Camelysin AlphaFold 2 and 3 monomer models there is an antiparallel strand β1, made up by the residues 46–49, LTLN in CalY1 and LALN in CalY2. These are exactly the same residues that form β1 antiparallel to β3 in the AlphaFold 2 multimer models. Not surprisingly, the same situation is found in the experimentally determined TasA monomers (β1 = $_{42}$GTLD$_{45}$) and filaments (β1 = $_{46}$LSAK$_{49}$). β2 is also somewhat shifted, comprising residues 54–56 in the AlphaFold monomer and the filament model. Furthermore, there is a traverse to a β2 antiparallel to β8 as observed in the filament structure of TasA. In summary, the AlphaFold monomer prediction shows a unit β1–β2 that resembles the corresponding structure in a v-set Ig fold as present in the filament. We have further predicted the structures of a set of 12 *B. cereus* group Camelysins with AlphaFold 3, including those of *B. cereus*, *B. mycoides*, *B. anthracis*, and *B. thuringiensis* (Figures S10 and S11), which showed β1 as antiparallel to β3 in 11 of the 12 Camelysin AlphaFold 3 predictions and parallel only in case of *B. thuringiensis* CalY1. In several predictions, this β2 is not fully developed, presumably since the inserting strand β0 of the previous protomer in the filament is missing. In summary, AlphaFold 3 imposes a TasA-like filament structure of the β1–β3 unit onto the monomer in 11 of 12 cases, with mostly confidence-signaling plDDT (predicted local distance difference test) values in the range 70–90 for the atoms of the β1 region. The overall score pTM (predicted template modeling) is around 0.7 for each of the 12 predictions. For the same sequences, AlphaFold 2 showed 4 results with parallel arrangements of β1–β3.

Tackling these discrepancies by an experimental approach, we applied solution NMR to determine the position of strand β1 in a non-polymerizing construct of CalY2, comprising residues 42–189 and thus lacking the N- and C-termini. An extended NMR dataset was recorded and a sufficient number of backbone NMR signals was assigned in an automated manner using the software ARTINA (Table S1) [31,32]. In particular, the stretches $_{42}$GTLDLA$_{47}$ and $_{69}$FKLENKGS$_{76}$ could be unambiguously assigned (see assignment plots in Figures S12 and S13). Since the data set included 3D ^{15}N-edited ^{1}H–^{1}H NOESY spectra of protonated and deuterated samples, we were able to detect interactions of backbone H_N and Hα, see Figure 4C. This enabled us to assess whether the observed patterns are in line with a parallel orientation of β1 and β3 as in the TasA-derived homology model or with an antiparallel arrangement predicted by AlphaFold. The parallel orientation is confirmed by cross peaks occurring between K70$_{HN}$ and T43$_{Hα}$; L71$_{Hα}$ and L44$_{HN}$; E72$_{HN}$ and D45$_{Hα}$ in the 3D ^{15}N-resolved ^{1}H–^{1}H NOESY spectrum (Figures 4C and S12, S13). A cross-peak correlating L46$_{HN}$ and N73$_{Hα}$ was not observed. The expected correlations between the H_N of L44, E72 and L46 appear all too close to the diagonal (9.034, 8.921 and 9.027 ppm, respectively) to add evidence. In the antiparallel model predicted by AlphaFold 2, hydrogen bonds are predicted between T43 and S76. However, no cross-peak was observed between their H_N resonances or between relevant cross-strand Hα atoms that could support this model. Likewise, no cross-peak is found between D45$_{HN}$ and G75$_{Hα}$. Expected cross-strand correlations between D45$_{HN}$ and G75$_{HN}$, and between A47$_{HN}$ and K74$_{HN}$ are both too close to the diagonal to be resolved. The interaction between A47$_{HN}$ and E72$_{HN}$ would not be hampered by signal overlap of the two H_N. However, it is not represented by a symmetric cross-peak set, i.e. the cross-peak on one side of the diagonal is missing whereas on the other a cross-peak due to another interaction is present. For the predicted antiparallel interaction between K74$_{HN}$ and L46$_{Hα}$ small cross-peaks were found, however, the respective nuclei are also relatively close in the parallel arrangement (5.3 Å). Due to the unique fit of our data to the parallel arrangement, we consider the TasA-derived homology model as a valid representation of the β1–β3 unit. In conclusion, the structural investigation on the CalY2 42–189 monomer in solution provides strong evidence for the conservation of the parallel β1–β3 arrangement as first observed in TasA. It is thus different to the AlphaFold prediction as in the TasA case.

Finally, we used our CalY2 model to check whether the amino acids proposed by MEROPS to form the active center of the putative protease (Asp31, Asp63, Asn73, Asp101, Gln173 and Gln177) are in a suitable arrangement for metal binding and protease activity. As Figure S14 shows, they are all too distant from each other to contribute to a metal binding site.

Structure of CalY2 filaments

An AlphaFold 2 model of CalY2 filaments (Figure 5A) was generated that follows the principles observed for TasA filaments [5]. Since the residues maintaining the head-to-tail interactions are the most conserved among TasA-like proteins, the main contributions to specificity must result from N- and C-terminal sites, and residues in β2 and β3 if the CalY2 filament structure is similar to that of TasA. In general, the N-terminal strand shows considerable conservation at its beginning, with the mechanism of insertion likely conserved. However, the


A
N
C
C
N
N
C
B
G192
T64
T190
G62
I65
Q189
A191
E194
E193
D63
E66
R195
110
120
130
^{15}N Chemical Shift (ppm)
11
10
9
8
7
6
^{1}H Chemical Shift (ppm)
C
T64: ^{1}H = 8.4 ppm
T64 - R195
T64
E193: ^{1}H = 9.8 ppm
E193
E193 - E194
E193 - T64
E194: ^{1}H = 8.3 ppm
E194 - R195
E194 - E193
E194
E194 - T64
R195: ^{1}H = 9.3 ppm
R195 - E194
R195 - T64N
R195
104 106 108 110 112
120 122 124 126 128
114 116 118 120 122
128 130 132 134 136
^{15}N Chemical Shift (ppm)
130
120
110
^{15}N Chemical Shift (ppm)
D
CalY2
TasA
K206
E229
β3
C-term TasA
C-term CalY2
β2
S58
G247 / V249
E
β3 (CalY2$_{i+1}$)
G192
T64
E193
I65
E66
R195
C-term (CalY2$_i$)
E194

very different C-termini of TasA-like proteins may adopt different structures and contribute differently to the binding interface. In fact, the C-terminal residues are the most diverse of the entire protein in an alignment ranging from archea to fungi [10], and even within the genus *Bacillus.* The very similar CalY1 and CalY2 have only a few residues following the essential C-terminal glutamine which is equivalent to Q232 in TasA (Figure 1). The subsequent C-terminal residues are $_{190}$TAGEER$_{195}$ (CalY2) and $_{192}$EAGE EK$_{197}$ (CalY1), whereas TasA contains 29 residues ($_{233}$WNGLTIKKDHTDKDGYVKE NEKAH SEDKN$_{261}$) that maintain hydrogen bonds with β2 and β3. Due to these sequence differences we decided to determine the position of the C-terminus in CalY2 filaments experimentally, performing solid-state MAS NMR studies on ^{13}C, ^{15}N- and ^{2}H, ^{15}N, ^{13}C-labeled samples of CalY2 polymers (Figure 3A, for molecular AlphaFold 2 prediction see Figure 5A). This would also test the AlphaFold model and, implicitly, confirm the preservation of the overall structure. The CalY2 filament preparation yield ^{1}H-^{15}N MAS NMR spectra with a considerable number of sharp cross peaks (Figure 5B) and large ^{1}H and ^{15}N chemical shift ranges. The ^{1}H chemical shift area of 8.8–11 ppm exposes many cross peaks as typical for proteins with a high β-sheet content. An extended dataset of the ^{2}H, ^{15}N, ^{13}C-labeled and 100% back-exchanged sample was recorded at 60 kHz MAS (Table S2). The spectra enabled the sequence-specific assignments of H_N, N, C_α, C', and C_β resonances.

The assignment covers approximately 50% of the protein backbone resonances (Table S3), as indicated by red color in the AlphaFold 2 model in Figure 5A. The rightmost protomer was colored according to secondary structure propensities. There, residues predicted to form strands by Talos + are shown in yellow, loop-like structures or random coil arrangements are colored green, and helices are shown in blue. The chemical shift index fits well to the filament model, except for the middle section of β9, which appears to have a less regular backbone. In spectra of the sample with perdeuterated non-exchanging sites, contacts between backbone H_N are readily detected, their proximity often coinciding with hydrogen bonding patterns in secondary structure elements, especially within β-sheets [33]. Regarding the position of the C-terminus, we found distinct H_N-H_N cross peaks involving T64 and the C-terminal residues E193, E194, and R195 in both hNHH and hNhhNH (Figure 5C) spectra. There is no cross peak between E193 H_N and R195 H_N, indicating that those are spatially apart. All four residues involved in Figure 5C are located in unique stretches ($_{61}$PGD**T**I$_{65}$ and $_{189}$QTAG**EER**$_{195}$) and can unambiguously be assigned due to characteristic chemical shifts.

The detected restraints confirm the AlphaFold 2 multimer model of CalY2 filaments shown in Figure 5A and thus the β-sheet complementation by β0. We simultaneously determined the position of the CalY2 C-terminus relative to β3 experimentally. This part of the model is depicted in Figure 5D with the distance restraints indicated in Figure 5E, together with the hydrogen bonds that are characteristic for the modeled secondary structure. As observed from the overlay of CalY2 and TasA filament structures in Figure 5D, the CalY2 C-terminus runs coherently with that of TasA, despite very different types of amino acids involved on both the C-terminus and β3. In this way, the C-terminal residues directly following the conserved glutamine and those at the beginning of β3 (see framed residues in Figure 1) encode specificity in filament formation of TasA-like proteins, together with residues at the end of β2 that interact with the conserved end of β0.

Discussion

CalY1 and 2 – oligomerization vs. polymerization

CalY1 and CalY2 of *B. cereus* were previously characterized as TasA-like proteins [12] and models of their structure presented [10]. With a sequence identity of 60% (see alignment in Figure 1), and conservative replacement of further residues, we expected a tendency to form filaments similar to TasA, and potentially even mixed fila-

Figure 5. A AF2-multimer model of a CalY2 filament showing three molecules. The circles indicate the site of intermolecular contacts of the C-terminus. Red color indicates residues with NMR signal assignments. The chemical shift index-based secondary structure propensity is indicated in the rightmost protomer by blue (helix), green (random) and yellow (strand) coloring. **B** ^{1}H–^{15}N correlation spectrum of filamentous CalY2 obtained *via* a CP-based pulse sequence in a 1.3 mm ZrO_2 rotor at 60 kHz magic-angle spinning. The assignments of the cross peaks to residues involved in C-terminal intermolecular interactions are indicated in cyan (β3) and orange (C-terminus). **C** Through-space contacts detected in the hNhhNH spectrum. Strong signals, corresponding to a short spatial distance, are present between T64-R195 and E194-R195. Weaker signals, indicative of inter-mediate distance, are detected between T64–E193 and E193–E194. **D** Overlay of TasA (gray) and CalY2 (light blue) filament models, focusing on the protomer interaction site. TasA-specific intermolecular contacts are indicated. The specific contact site in the filament of CalY2 is indicated in orange (C-terminus) and cyan (β3). **E** Interactions detected in **C** between the two strands (C-term and β3) highlighted in **D**.

ments. As a prerequisite for following up on our hypothesis, recombinant mature CalY1 and CalY2 were produced as soluble and folded proteins. Surprisingly, CalY1 appears as two species I and II in gel filtration experiments which dynamically convert into each other. According to native PAGE, CalY1-II corresponds to monomeric protein, and CalY1-I to a polydisperse mixture of oligomers. By contrast, CalY2 appears initially as a monomer in solution and is found to spontaneously form gel-like dense networks of filaments immediately after cleavage of the Sumo fusion. CalY1 does not form filaments at neutral pH. Only at acidic and basic pH settings are small amounts of filaments detected, correlating at pH 8.5 with reduced amounts of polydispersed oligomers. This different behavior between CalY1 and CalY2 is surprising, given 62% sequence identity, and raises the question of which features in the sequences are responsible. Most likely, the polydispersed nature of CalY1 solutions interferes with filament formation. Among the sequence differences between the two Camelysins is a region on the β5/β6 side of the protein (Figure S15, top end of the model), where CalY2 forms a long, two-stranded β-sheet made up by β5 and β6, and in CalY1 this sheet is in part collapsed, forming a disordered loop involving the stretch $_{112}$-WNWDKQSEP$_{120}$, exposing two tryptophan residues. A second poorly-conserved neighboring region is $_{130}$QKVDPDLLAK$_{139}$. It is tempting to speculate that oligomerization into a polydispersed mixture is supported by these residues, especially as they are considered in part flexible.

Contrary to the first descriptions of Camelysins [15,16] we could not find any protease activity for recombinant CalY1 and CalY2, even after supplementing metal ions. As all *Bacillus* strains produce large amounts of proteases, the protease activity measured formerly on material from natural sources may originate from other co-purified impurities. It was previously demonstrated that CalY in a strain of *B. thuringiensis* is not a protease, but plays a role in Biofilm formation and cell adhesion [28]. As shown in Figure S14, the CalY1 and CalY2 residues proposed by MEROPS to form a potential metal binding site are all too distant to support the metalloprotease hypotheses.

Conservation of monomer and filament structure and the usefulness of validating AlphaFold predictions

When comparing the CalY2 AlphaFold 2 and homology models (Figure 4A and B) with the X-ray structure of monomeric TasA (Figure S8A), our NMR studies confirmed the homology model regarding the parallel arrangement of the functionally important β1–β3 strand interaction. This is in stark contrast to the AlphaFold prediction that produces an antiparallel β1–β3 orientation, resembling the polymer structure. The structural arrangements in the β5/β6 and α2–α4 regions of the models as shown in Figures S8B–D and S9 are strain-specific due to the strong sequence differences, even between CalY1 and CalY2. Here, we cannot determine which models are more accurate, the AlphaFold predictions (Figure S9A) or the homology models shown in Figure S9B. For the CalY2 filament, we were interested in the structure adopted by the C-terminal residues since they contribute considerably to the protomer interface in TasA filaments. Here, we also found conservation of the structure by NMR, despite very different amino acid sequences and especially lengths.

The inability of AlphaFold to predict the 'active center' of the monomer structure, i.e. the β1–β2–β3 part which undergoes a structural rearrangement upon filament formation, signals the necessity of a certain caution in the handling of AlphaFold predictions. Nevertheless, this result may provide important clues regarding the proteins under study. Generally speaking, AlphaFold provides several probability-ranked solutions to one interpretation of a single, deep multiple sequence alignment, and it does not consider predictions from subsets of alignments. Thus AlphaFold may not predict two different forms of a protein such as the monomer and the complex (filament) at the same time correctly if considerable structural rearrangements are involved. However, larger structural rearrangements may occur more often than expected since, for example, more than 30% of genomes represent whole proteins or protein segments that are disordered in the absence of an interaction partner. In our study, AlphaFold reproducibly predicts the correct filament structure, but reproducibly incorrectly predicts the active center of the monomer. In fact, AlphaFold 2 predicts 4 structures out of 12 sequences supplied with a parallel β1–β3 arrangement, and AlphaFold 3 only one, with 11 structures showing an antiparallel arrangement. Intriguingly, these results allow for interesting conclusions. AlphaFold interprets correlated mutations that are expected to occur primarily when the protein structure or – more importantly – function is impaired by an initial change, and these deleterious effects are subsequently compensated for by a second change. It is fair to assume that the delicate protomer–protomer interactions in the filament are not very tolerant of amino acid changes, requiring matching mutations to maintain structure and function. If the position of β1 in the monomer is functionally important, then it should also be represented by associated mutations. The fact that we observed largely the β1–β2 arrangement of the filament structure in the prediction of the monomer suggests that the filament structure is more important for the organism than the monomer structure, in which the other orientation of β1 might be a mechanism for auto-inhibition of

polymerization. A second consideration arising from the predictions is connected to the very similar filament model obtained for CalY1 sequences. In obtaining the filament model, we recognize the need for corrections of initial, random mutations that the organism needs for survival and successful reproduction. We therefore expect that CalY1 filaments still play a role for *B. cereus*, despite the fact that we did not obtain them in vitro at neutral pH.

On the mechanism of biofilm genesis in B. cereus

A key result of this study is in the finding that CalY2 readily forms filaments; the energy barrier towards polymerization is very low. An initiation by CalY1 as in the case of *B. subtilis* where TapA helps TasA to polymerize was not found, in part because CalY2 polymerization is extremely fast and initiation is thus not required. In the *B. cereus* biofilm operon there is no direct TapA homologue, although a report assigns a homologous function to the product of *bc1280* termed CapP [14]. However, the potential mechanism is unclear and an N-terminal motif that is homologous to the CalY2 N-terminus is not present in its sequence (Uniprot: Q81GC7_BACCR), as would be required for it to follow the same mechanistic principles as for TapA/TasA. In this case the TapA N-terminus is homologous to the TasA one, and provides the template of strand β0 around which TasA folds, initiating the chain reaction. Worthy of note, the CapP sequence has 14 EEQKK repeats separated by V or A at its C-terminus, and the AlphaFold prediction shows a largely helical protein. Altogether, this suggests that the *B. cereus* biofilm operon does not contain a TapA-equivalent gene, although CapP might still help in polymerization. Nevertheless, the question remains regarding which function(s) CalY1 could fulfill apart from forming filaments outside the neutral pH range. Its sequence certainly bears the properties required for filament formation, summarized in the fact that AlphaFold builds a proper filament.

Indeed, following earlier work by Caro-Astorga et al. [17], Candela et al. [28] observe filaments of *B. thuringiensis* CalY1, yet solely under certain conditions. CalY1 expression rises sharply from early-stationary to mid-stationary phase. At the beginning of the stationary phase, filaments are largely found at the bacterial surface, but later are absent from the cell surface and free filaments are observed within the biofilm. Filaments were not successfully obtained *in vitro* in this study, but the authors demonstrated that components of a planktonic culture of a ΔcalY1 strain were able to induce filament formation only during and after stationary phase. Furthermore, *B. thuringiensis* CalY1 was found to be essential for binding of the bacteria to HeLa cell surfaces. This further supports the conclusion that a certain factor is necessary for CalY1 to polymerize. These findings are in line with the observation that *B. cereus* CalY1 is required for the colonization of plant roots [15]. It is tempting to speculate about the mechanism, which might include binding of the additional factor, potentially an oligosaccharide, to the disordered region between β5 and β6, thus reducing polydisperse oligomer formation and enabling the formation of stable filaments. In this long flexible loop, a motif comprising two tryptophan residues might be suited to oligosaccharide binding.

Specificity in Bacillus biofilm formation – and why CalY1 does not form filaments

The high similarity of the CalY1 and CalY2 sequences and their very different biochemical behavior provokes discussion of specificity-determining sequence segments implicated in filament formation, since the inability of CalY1 to form polymers at neutral pH must be also associated with lower CalY1 filament stability as compared to CalY2 filaments. For the two Camelysins, strand β0 is identical between CalY1 and CalY2, and the C-termini are strongly homologous. Furthermore, our experimental study confirmed the CalY2 filament model and also a high structural similarity of the CalY2 filaments with those of TasA around the C-terminal protomer interaction site (Figure 5D). Therefore, differences in β2, β3 and β8 of the Camelysins (or β9 of other TasA-like proteins) are prime candidates for influencing filament stability.

Starting with the interaction of the C-terminal residues and β3, it appears that even the beginning of β3 differs between CalY1 and CalY2 (Figure 1, blue shaded area after LKPGD). Furthermore, β2 plus the two following residues reaching into the head-to-tail binding interface have a different sequence in CalY1 (-LVDIKD-) compared to CalY2 (-VVNVSN-), which both align with the conserved -NTFAA- of β0 (Figure 1, blue shaded area before LKPGD). In this relevant section of β2 two charges are modified, whereby one negative charge is introduced into the binding interface opposite the conserved threonine of β0. The differing residues relevant for filament stability are displayed in green in a protomer of an AlphaFold 2 prediction of CalY1 filaments (Figure S14B). Similar to the C-termini, the discussed sections of β2 and β3 are also very divergent in an extended sequence alignment of TasA-like proteins, hence those sections may be considered as specificity-determining. Between CalY1 and CalY2 are more mutations, yet the hydrophobic amino acids oriented towards the interior are all conserved or conservatively substituted, ensuring that the backbone positions are not changed. The side chains of substantially differing amino acids are all situated on the surface, as often in β-sheet proteins. Whether these substitutions destabilize the filament cannot be deduced here, but it seems unlikely.

In conclusion, we followed the hypothesis that TasA-like proteins from related *Bacillus* strains could interact, but found that CalY2 readily polymerizes whereas CalY1 does not, and CalY1 does not form mixed filaments with *B. subtilis* TasA and TapA. We experimentally confirm structural conservation of the parallel arrangement of β1–β3 in CalY2 monomers, contrasted by a tendency of AlphaFold to predict mainly filament structural features in this region. It may be speculated that the parallel orientation is a modest inhibitor of polymerization that regulates initiation, thereby helping to optimally modulate biofilm formation. The local structure around the C-terminal residues in CalY2 filaments was determined by solid-state NMR approaches and also shows structural homology with TasA filaments despite very different local sequences. Altogether, these investigations enable the definition of specificity-determining regions in the sequences of TasA-like proteins at the end of β2, the beginning of β1, and the C-terminus. Throughout this investigation, NMR has proved to be a suitable method to test AlphaFold predictions in specific regions of the models.

Material and Methods

Constructs

For CalY1 (197 amino acids, Q8GJ76; Q8GJ76_BACCE; AF-Q8GJ76-F1) and CalY2 (195 amino acids, A0A0E3SV09_BACCE; AF-A0A0E3SV09-F1) codon-optimized strings were synthesized by Geneart (Regensburg, Germany). The 5‘and 3‘ends flanking the open reading frame contain BsaI sites to generate pET-based expression vectors under T7 promotor control via an in-house Modular Cloning System. 2 vectors coding for HisSumo_CalY1_28-197 and HisSumo_CalY2_28-195 were confirmed by sequencing. Constructs for TasA and TapA have been described elsewhere [7,10].

Recombinant production

Protocols were described before [10]. In brief, both CalY1 and CalY2 constructs did yield well-soluble protein after expression on rich and minimal medium. Approximately 60–80 mg of fusion protein were obtained from a 1 L culture after the first metal chelate (MC) column. Following cleavage of the HisSumo tag by Sumoprotease under dialysis and a second MC column, the CalYs were concentrated and applied to a 120 or 320 ml Superdex 75 gel filtration column.

Top-down mass spectrometry

Protein intact mass analyses were conducted on an Agilent 1290 Infinity II UHPLC system coupled to an Agilent 6230B time-of-flight (TOF) LC/MS instrument equipped with an AJS (Agilent Jet Stream Technology) ion source operated in positive ion mode. Protein samples were desalted using a Zorbax 300SB-C3 guard column (2.1 × 12.5 mm, 5 μm). Approximately 0.3 μg of denatured sample was injected for each analysis. LC/MS parameters were adapted from Chalk et al. (2017) [34]. The ion source was operated with the capillary voltage at 4000 V, nebulizer pressure at 50 psi, drying and sheath gas at 350 °C, and drying and sheath gas flow rate at 12 and 11 l/min, respectively. The instrument ion optic voltages were as follows: fragmentor 250 V, skimmer 65 V, and octopole RF 750 V. MS data were analyzed using the Protein Deconvolution feature of the MassHunter BioConfirm Version 10.0 software (Agilent) that uses the Maximum Entropy algorithm for accurate molecular mass calculation. Deconvolution was performed between mass a range of 800–2500 *m/z*, using peaks with a ratio of signal to noise greater than 30:1. The deconvoluted mass range was set at 5–25 kDa and the step mass was 1 Da.

Native gel (Figure 2G)

A commercial pre-prepared gel (TGX 4–20%) was purchased from Bio-Rad, including running buffer 161-0734 that contained 25 mM Tris and 192 mM glycine at pH 8.3. In all cases, 5 μl of protein solution were loaded onto the gel. To investigate the nature of CalY1-I and CalY2-II, protein was prepared freshly up to the final step of the purification protocol, the SEC. The eluted peaks were directly loaded onto the gel in lanes 1 (CalY1-II) and 2 (CalY1-I). Lanes 3–5 contain the CalY1 samples subjected to polymerization trials, with 2.6 mg/ml at pH 5.0 loaded on lane 3, 5.3 mg/ml at pH 7.0 loaded on lane 4, and 2.8 mg/ml at pH 8.5 loaded on lane 5.

Thermal shift assay (Figure 2H)

Protein melting temperatures were determined as described by Roske et al. [7].

Polymerisation trials with CalY1 at different pH

In order to investigate the polymerization behavior of CalY1, 800 μl of sample (5.3 mg/ml) was taken out of the freezer, split into three batches, and diluted into 1 × 20 mM Tris buffer at pH 7.0, containing 150 mM NaCl. Then one batch was left at pH 7.0 and the other two were dialyzed within 48 h at 8 °C against pH 5.0 and pH 8.5 buffer that was otherwise identical. Equal volumes of all three pH settings, yet containing different amounts of protein, were taken and loaded onto the native gel lanes 3–5. The concentrations were 2.0 μM for pH 5.0, 5.3 μM for pH 7.0, and 2.6 μM for pH 8.5. For investigating the filament content, all three samples were subjected to ultracentrifugation for 2 h at 8 °C and the pellet

resuspended in 20 μL buffer for transfer to negative stain EM.

Thioflavin T (ThT) assay

5 μl samples of CalY1 solutions in 20 mM Tris, 150 mM NaCl at pH 5, 7 and 8.5 with a concentration of 280 μM were diluted with 95 μl 20 mM Tris pH 7, 150 mM NaCl while transferred into a 96-well plate (Costar 3615) and mixed 1:1 with 40 μM ThT solution in the same buffer. Fluorescence (Ex 438 nm/ Em 495 nm) was acquired using a Tecan reader.

Analytical Ultracentrifugation (AUC)

Sedimentation velocity experiments [35] were performed with a Beckman Optima XL-I analytical ultracentrifuge at 20 °C and at a rotor speed of 40,000 rpm (129,000*g* at cell bottom). The samples were measured in dialysis buffer (20 mM Na phosphate pH 7.0, 150 mM NaCl or with 300 mM NaCl for experiments with CalY2) separately or in combination at a final concentration of 100 μM in experiments with CalY1 or 50 μM in experiments with CalY2 as indicated. Interference data with a total of 140 scans were recorded every 5 min. Sedimentation coefficient distributions c(s) were analyzed with the program SEDFIT [36]. The protein partial-specific volume and the buffer physical constants were calculated from amino acid and buffer composition using SEDNTERP [37]. Results were plotted with GUSSI.

Dot blot

TasA and Camelysin solutions as well as 2 negative controls were diluted to a protein concentration of 1 mg/ml. Defined spots were applied to a nitrocellulose membrane by pipetting 2 μl each. Following a blocking step, the membrane was incubated with the anti-TasA primary antibody (generous gift of K. Turgay) at a 1:5000 dilution for 1 h. After a subsequent washing step, the membrane was incubated with the peroxidase-conjugated secondary antibody (Rockland #611-1302) at a 1:5000 dilution for 1 h. After a final washing step, the membrane was treated with SuperSignal™ West Pico Chemiluminescent Substrate (Thermo #34077). Chemiluminescence signal was detected using the Vilber Fusion FX imaging system.

Biofilm experiments

The appearance of live biofilms as pellicles on culture medium depends on the employed growth conditions, and the investigated bacterial strains. MOLP [25] and mineral salt medium with glycerol and glutamate [26] promote wrinkled biofilm that indicates an extended layer of carbohydrates. The appearances of the pellicle-type *B. subtilis* and *B. cereus* biofilms differ. *B. subtilis* can form a strongly wrinkled biofilm and *B. cereus* shows more or less a skin-like appearance on the surface of a static culture. For this reason, a possible interaction of *C. cereus* and *B. subtilis* biofilm proteins was investigated by using *B. subtilis* cells, monitoring the degree of wrinkle formation. The addition of 0.5 % ethanol or acetoin enhances a wrinkled phenotype. Importantly, natural isolates or laboratory strains can vary strongly in their ability to form pellicles [38].

For these biofilm experiments with pellicule cultures, wild type DK1042 (NCIB 3610 *comIQ12L*) and *ΔtasA* (DK1042 *tasA::kan*) strains were provided by the group of K. Turgay (MPI, Berlin). WT and *ΔtasA* from glycerol stocks were grown under shaking on LB (*ΔtasA* with 10 μg/ml kanamycin) over night at 30 °C. At the next day, the culture was refreshed and grown to mid log phase (about 4–6 h). The biofilm setups with 1,4 ml MOLP medium per cavity in a 24well plate were inoculated 1:100 with the fresh culture. For rescue experiments 5 μM protein (TasA or Camelysins) was added after 1 h at 30 °C. The incubation without shaking took 46 h.

Structural investigations by solution NMR

All solution NMR experiments were done on recombinant CalY2 42–189 dissolved in 20 mM Tris–HCl pH7.0 buffer containing 300 mM NaCl and 10% (v/v) D_2O at protein concentrations of 500–600 μM. For spectral assignment a series of triple resonance experiments, CBCANH, CBCA(CO)NH, HNCO, HBHA(CO)NH, HCC(CO)NH, ^{15}N-edited NOESY, and a ^{15}N-edited HSQC experiment were recorded on uniformly ^{13}C, ^{15}N-labeled samples. All experiments were acquired at 300 K using an AVANCE III 600 MHz spectrometer equipped with a QCI CryoProbe (Bruker BioSpin) and pulse programs closely related to the Bruker pulse program library. Spectra were processed with TOPSPIN 3.5pl7 (Bruker BioSpin), referenced to 2,2-dimethyl-2-sila pentane-5-sulfonic acid (DSS, 0 ppm), and further analyzed with CcpNmr AnalysisAssign [39]. The chemical shift assignment was achieved by using the ARTINA [31] chemical shift assignment procedure available on the NMRtist web-page. Chemical shifts of ^{1}H resonances were referenced to DSS (0 ppm), while ^{15}N and ^{13}C chemical shifts were referenced indirectly using the gyromagnetic ratios of ^{15}N, ^{13}C, and ^{1}H ($\gamma^{15}N/\gamma^{1}H = 0.101329118$, $\gamma^{13}C/\gamma^{1}H = 0.251449530$).

Solid-state NMR

CalY2 was expressed uniformly in [^{2}H, ^{13}C, ^{15}N]-labeled form and purified in aqueous buffers (100% ^{1}H back-exchange). After SEC, the protein spontaneously formed filaments in 20 mM Tris, pH 7.5 and 330 mM NaCl, observable as a drastic increase in solution viscosity and by the appearance of assemblies in negative stain EM.

The filaments were sedimented by ultracentrifugation in a TLA110 rotor from Beckman at 75,000 rpm (~230,000*g*) for 30 min at 8 °C. Afterwards, the pellet was washed once with a low-salt buffer (20 mM Tris, pH 7.5, and 50 mM NaCl) and the centrifugation step was repeated for 1 h. Subsequently, the protein was transferred into a 1.3 mm ZrO_2 NMR rotor. After initial measurements, the rotor was opened and an additional portion of the CalY2 filament pellet was added. The rotor caps were sealed with a thin layer of glue.

All spectra were recorded on a Bruker Avance III spectrometer operating at a 600 MHz ^{1}H Larmor frequency, using a triple-channel (^{1}H, ^{13}C, ^{15}N) 1.3 mm MAS NMR probehead (Bruker). A MAS II unit was used to adjust the spinning rate to 60 kHz; temperature control during measurements was achieved with a BCU II unit, applying a gas flow of 1200 L/h, and a cooling gas temperature of 252 K, resulting in a sample temperature of approximately 297 K.

90° hard pulses for NMR measurements were set to 2.5 μs (^{1}H), 5 μs (^{13}C), and 7 μs (^{15}N). For basic assignment spectra, manufacturer-provided pulse programs hNH2D.dcp, hCONH3D.tcp, hcoCacoNH3D.tcp, and hcaCbcaNH3D.tcp with cross-polarization (CP) [40–42] for hetero-nuclear magnetization transfer were used. In those pulse sequences, homo-nuclear (^{13}C–^{13}C) transfer was mediated by *J*-coupling as described in [43]. Provided parameter recommendations were closely followed and subsequently optimized for existing conditions.

3D (hCaNH, hCOcaNH) and 4D (hCOCaNH, hCaCONH, hCxCaNH, hCxCacoNH, each with 5% non-uniform sampling) spectra were recorded sensitivity-enhanced as previously published [44,45] In short, CP with a linearly ramped amplitude on ^{1}H was used for the initial ^{1}H–^{13}C magnetization transfer and sensitivity-enhanced transverse-mixing pulses (TROP) for all subsequent transfers in the protein backbone (H_N, N, Ca, CO). Original pulse sequences and shapes for 600 MHz ^{1}H Larmor frequency are available at https://www.optimal-nmr.net and were adapted to a spinning rate of 60 kHz by adjusting the length of the TROP pulse components each to the duration of a rotation period. For hCxCaNH and hCxCacoNH experiments, magnetization transfer into the side chain was mediated by DIPSI-3 [46].

Information about spatial proximity were extracted from hNHH and hNhhNH pulse sequences. The hNHH sequence consists of an hNH CP step yielding two indirect dimensions followed by a 1.5 ms ^{1}H–^{1}H radio frequency-driven dipolar recoupling (RFDR) [47] step prior to acquisition. The hNhhNH pulse sequence includes an additional, subsequent hNH CP step. Acquisition and processing parameters for all spectra can be found in Tabless S3 and S4, respectively.

Uniformly sampled spectra were processed using TopSpin v4. NUS spectra were reconstructed and processed using nmrPipe [48]; NUS lists were prepared using the tool available at https://gwagner.med.harvard.edu/intranet/hmsIST [49,50]. The analysis of all spectra and the signal assignment were performed with CcpNmr AnalysisAssign, version 3.3.2.1 [35].

CRediT authorship contribution statement

Anne Diehl: Writing – original draft, Investigation, Conceptualization. **Florian Lindemann:** Methodology, Investigation. **Nils Cremer:** Methodology, Investigation. **Yvette Roske:** Methodology, Investigation. **Matthias Hiller:** Methodology, Investigation. **Barth van Rossum:** Conceptualization, Investigation, Methodology, Visualization, Writing – original draft. **Martina Leidert:** Methodology, Investigation. **Kürşad Turgay:** Conceptualization, Formal analysis, Writing – review & editing. **Hartmut Oschkinat:** Writing – review & editing, Writing – original draft, Data curation, Conceptualization.

DATA AVAILABILITY

NMR raw data, chemical shift and peak lists of CalY2 filaments are deposited in the BMRB as entry XXXXX. In addition, raw and processed CalY2 filament spectra, nmrPipe scripts used for reconstruction, the CcpNmr AnalysisAssign assignment project, TALOS+ predictions, and the AlphaFold model are uploaded to zenodo (10.5281/zenodo.15528727).

DECLARATION OF COMPETING INTEREST

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Hartmut Oschkinat reports financial support was provided by German Research Foundation. Kuersad Turgay reports financial support was provided by German Research Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We indebted to Anja Schütz and the team of the Protein Production & Characterization Platform at the Max Delbrück Center (MDC) for Molecular Medicine, Berlin, Germany, for access to protein mass analyses. The modular cloning Kit was provided by Martin Bommer during his time at the

MDC. For NMR experiments we thank Peter Schmieder (FMP). The authors are grateful to Quentin Sattentau, Oxford, for style-editing the manuscript.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.jmb.2026.169661.

Received 15 October 2025;
Accepted 19 January 2026;
Available online 27 January 2026

Keywords:
Bacillus subtilis;
Bacillus cereus;
biofilm;
CalY;
TasA

Abbreviations:
AA, amino acid(s); NMR, nuclear magnetic resonance; AUC, analytical ultracentrifugation; EM, electron microscopy; MAS, magic-angle spinning

References

[1]. Flemming, H.-C. et al, (2016). Biofilms: an emergent form of bacterial life. *Nature Rev. Microbiol.* **14**, 563–575.
[2]. Hobley, L., Harking, C., MacPhee, C.E., Stanley-Wall, N. R., (2015). Giving structure to the biofilm matrix: an overview of individual strategies and emerging common themes. *FEMS Microbiol. Rev.* **39**, 649–669.
[3]. Vlamakis, H., Chai, Y., Beauregard, P., Losick, R., Kolter, R., (2013). Sticking together: building a biofilm the *Bacillus subtilis* way. *Nature Rev. Microbiol.* **11**, 157–168.
[4]. Arnaouteli, S., Bamfort, N.C., Stanley-Wall, N.R., Kovacs, A.T., (2021). *Bacillus subtilis* biofilm formation and social interactions. *Nature Rev. Microbiol.* **19** (9), 600–614.
[5]. Böhning, J., Ghrayeb, M., Pedebos, C., Abbas, D.K., Khalid, S., Chai, L., Bharat, T.A.M., (2022). Donor-strand exchange drives assembly of the TasA scaffold in *Bacillus subtilis* biofilms. *Nature Commun.* **13**, 7082.
[6]. Roske, Y., Lindemann, F., Diehl, A., Cremer, N., Higman, V.A., Schlegel, B., et al., (2023). TapA acts as specific chaperone in TasA filament formation by strand complementation. *PNAS* **120**, (17)e2217070120
[7]. van Gerven, N., Klein, R.D., Hultgren, S.J., Remaut, H., (2015). Bacterial amyloid formation: structural insights into curli biogenesis. *Trends Biomol.* **23**, 693–706.
[8]. Hansen, K.H., Byeon, C.H., Liu, Q., Drace, T., Boesen, T., Conway, J.F., Akbey, Ü., (2024). Structure of biofilm-forming functional amyloid PSMα1 from *Staphylococcus aureus*. *PNAS* **121**, (33)e2406775121
[9]. Böhning, J., Tarafder, A.K., Bharat, T.A.M., (2024). The role of filamentous matrix molecules in shaping the architecture and emergent properties of bacterial biofilms. *Biochem. J.* **481** (4), 245–263.
[10]. Diehl, A., Roske, Y., Ball, L., Chowdhury, A., Hiller, M., Molière, N., Kramer, R., Stöppler, D., Worth, C.L., Schlegel, B., Leidert, M., Cremer, N., Erdmann, N., Lopez, D., Stephanowitz, H., Krause, E., van Rossum, B.-J., Schmieder, P., Heinemann, U., Turgay, K., Akbey, Ü., Oschkinat, H., (2018). Structural changes of TasA in biofilm formation of *Bacillus subtilis*. *PNAS* **115** (13), 3237–3242.
[11]. Ehling-Schulz, M., Lereclus, D., Koehler, T.M., (2018). The *Bacillus cereus* group: *Bacillus* species with pathogenic potential. *Microbiol. Spectr.* **7** (3) https://doi.org/10.1128/microbiolspec.gpp3-0032-2018.
[12]. Gao, T., Foulston, L., Chai, Y., Wang, Q., Losick, R., (2015). Alternative modes of biofilm formation by plant-associated *Bacillus cereus*. *Microbiologyopen* **4** (3), 452–464.
[13]. Lin, Y., Briandet, R., Kovács, Á.T., (2022). *Bacillus cereus sensu lato* biofilm formation and its ecological importance. *Biofilm* **4**, 100070.
[14]. Álvarez-Mena, A., Shukkoor, M.B.A., Caro-Astorga, J., Berbon, M., Antequera-Gómez, M.L., Grélard, A., et al., (2015). CapP mediates the structural formation of biofilm-specific pili in the opportunistic human pathogen *Bacillus cereus* 2024.02.27.582368 *bioRxiv* http://biorxiv.org/content/early/2025/01/08/2024.02.27.582368.
[15]. Fricke, B., Drössler, K., Willhardt, I., Schierhorn, A., Menge, S., Rücknagel, P., (2001). The cell envelope-bound metalloprotease (camelysin) from *Bacillus cereus* is a possible pathogenic factor. *Biochim. Biophys. Acta* **1537** (2), 132–146.
[16]. Grass, G., Schierhorn, A., Sorkau, E., Müller, H., Rücknagel, P., Nies, D.H., et al., (2004). Camelysin is a novel surface metalloproteinase from *Bacillus cereus*. *Infect. Immun.* **72** (1), 219–228.
[17]. Caro-Astorga, J., Pérez-García, A., de Vicente, A., Romero, D., (2014). A genomic region involved in the formation of adhesin fibers in *Bacillus cereus* biofilms. *Front. Microbiol.* **5**, 1–11.
[18]. Caro-Astorga, J., Frenzel, E., Perkins, J.R., Álvarez-Mena, A., de Vicente, A., Ranea, J.A.G., Kuipers, O.P., Romero, D., (2020). Biofilm formation displays intrinsic offensive and defensive features of *Bacillus cereus*. *Npj Biofilm Microbiomes* **6**, 3.
[19]. Rawlings, N.D., Tolle, D.P., Barret, A.J., (2004). MEROPS: the peptidase database. *Nucleic Acids Res.* **32**, D160–D164.
[20]. El Mammeri, N., Hierrezuelo, J., Tolchard, J., Cámara-Almirón, J., Caro-Astorga, J., Álvarez-Mena, A., et al., (2019). Molecular architecture of bacterial amyloids in *Bacillus* biofilms. *FASEB J.* **33** (11), 12146–12163.
[21]. Jumper, J. et al, (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.
[22]. Arnaouteli, S., Bamford, N.C., Stanley-Wall, N.R., Kovacs, A.T., (2021). *Bacillus subtilis* biofilm formation and social interactions. *Nature Rev. Microbio.* **19**, 600–614.
[23]. Dragoš, A., Kiesewalter, H., Martin, M., Hsu, C.Y., Hartmann, R., Wechsler, T., et al., (2018). Division of labor during biofilm matrix production. *Curr. Biol.* **28** (12), 1903–1913.e5.
[24]. Li, J., Fang, P., Yi, X., Kumar, V., Peng, M., (2022). Probiotics *Bacillus cereus* and *B. subtilis* reshape the intestinal microbiota of Pengze crucian carp (*Carassius auratus* var. Pengze) fed with high plant protein diets. *Front. Nutr.* **9**

[25]. Ahimou, F., Jacques, P., Deleu, M., (2000). Surfactin and iturin A effects on *Bacillus subtilis* surface hydrophobicity. *Enzyme Microb. Technol.* **27** (10), 749–754.
[26]. Branda, S.S., González-Pastor, J.E., Ben-Yehuda, S., Losick, R., Kolter, R., (2001). Fruiting body formation by *Bacillus subtilis*. *PNAS* **98** (20), 11621–11626.
[27]. McLoon, A.L., Guttenplan, S.B., Kearns, D.B., Kolter, R., Losick, R., (2011). Tracing the domestication of a biofilm-forming bacterium. *J. Bacteriol.* **193** (8), 2027–2034.
[28]. Candela, T., Fagerlund, A., Buisson, C., Gilois, N., Kolstø, A.B., Økstad, O.A., et al., (2019). CalY is a major virulence factor and a biofilm matrix protein. *Mol. Microbiol.* **111** (6), 1416–1429.
[29]. Varadi, M. et al, (2024). AlphaFold protein structure database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **52** (D1), D368–D375.
[30]. The UniProt Consortium, (2025). UniProt: the Universal protein knowledgebase in 2025. *Nucleic Acids Res.* **53** (D1), D609–D617.
[31]. Klukowski, P., Riek, R., Günter, P., (2022). Rapid protein assignments and structures from raw NMR spectra with the deep learning technique ARTINA. *Nature Commun.* **13**, 6151.
[32]. Klukowski, P., Riek, R., Günter, P., (2023). NMRtist: an online platform for automated biomolecular NMR spectra analysis. *Bioinformatics* **39**, (2)btad066
[33]. Friedrich, D., Perodeau, J., Nieuwkoop, A.J., Oschkinat, H., (2020). MAS NMR detection of hydrogen bonds for protein secondary structure characterization. *J. Biomol. NMR* **74**, 247–256.
[34]. Chalk, R., (2017). Mass sSpectrometric analysis of proteins. *Methods Mol. Biol.* **1586**, 373–395.
[35]. Schuck, P., (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78** (3), 1606–1619.
[36]. Laue, T.M., Shah, B., Ridgeway, T., Pelletier, S., (1992). Computer-aided interpretation of sedimentation data for proteins. In: Harding, S.E., Rowe, A.J., Horton, J.C. (Eds.), *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Royal Society of Chemistry, pp. 90–125.
[37]. Brautigam, C.A., (2015). Calculations and publication-quality illustrations for analytical ultracentrifugation data. *In: Methods in Enzymology*. first ed. Elsevier Inc., pp. 109–133.
[38]. Yan, F., Yu, Y., Gozzi, K., Chen, Y., Guo, J.H., Chai, Y., (2017). Genome-wide investigation of biofilm formation in *Bacillus cereus*. *Appl. Environ. Microbiol.* **83**, (13)e00561-17
[39]. Skinner, S.P., Fogh, R.H., Boucher, W., Ragan, T.J., Mureddu, L.G., Vuister, G.W., (2016). CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR* **66** (2), 111–124.
[40]. Öster, C., Lange, S., Hendriks, K., Lange, A., (2024). Detecting bound ions in ion channels by solid-state NMR experiments on 15N-labelled ammonium ions. *Methods Mol. Biol.* **2796**, 23–24.
[41]. Pines, A., Gibby, M.G., Waugh, J.S., (1972). Proton-enhanced nuclear induction spectroscopy. A method for high resolution NMR of dilute spins in solids. *J. Chem. Phys.* **56**, 1776–1777.
[42]. Barbet-Massin, E., Pell, A.J., Retel, J.S., Andreas, L.B., Jaudzems, K., Franks, W.T., Nieuwkoop, A.J., Hiller, M., Higman, V., Guerry, P., Bertarello, A., Knight, M.J., Felletti, M., Le Marchand, T., Kotelovica, S., Akopjana, I., Tars, K., Stoppini, M., Bellotti, V., Bolognesi, M., Ricagno, S., Chou, J.J., Griffin, R.G., Oschkinat, H., Lesage, A., Emsley, L., Herrmann, T., Pintacuda, G., (2014). Rapid proton-detected NMR assignment for proteins with fast magic angle spinning. *J. Am. Chem. Soc.* **136** (35), 12489–12497.
[43]. Öster, C., Chevelkov, V., Lange, A., (2025). Evaluation of TOCSY mixing for sensitivity-enhancement in solid-state NMR and application of 4D experiments for side-chain assignments of the full-length 30 kDa membrane protein GlpG. *J. Biomol. NMR* **79**, 25–34.
[44]. Blahut, J., Brand, M.J., Pradhan, T., Reif, B., Tošner, Z., (2022). Sensitivity-enhanced multidimensional solid-state NMR spectroscopy by optimal-control-based transverse mixing sequences. *J. Am. Chem. Soc.* **144** (38), 17336–17340.
[45]. Blahut, J., Brandl, M.J., Sarkar, R., Reif, B., Tošner, Z., (2023). Optimal control derived sensitivity-enhanced CA-CO mixing sequences for MAS solid-state NMR – applications in sequential protein backbone assignments. *J. Magn. Reson.* **16–17**, 100122.
[46]. Shaka, A.J., Lee, C.J., Pines, A., (1988). Iterative schemes for bilinear operators; application to spin decoupling. *J. Magn. Reson.* **77** (2), 274–293.
[47]. Bennett, A.E., Griffin, R.G., Ok, J.H., Vega, S., (1992). Chemical shift correlation spectroscopy in rotating solids: radio frequency-driven dipolar recoupling and longitudinal exchange available. *J. Chem. Phys.* **96**, 8624–8627.
[48]. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., Bax, A., (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293.
[49]. Hyberts, S.G., Takeuchi, K., Wagner, G., (2010). Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J. Am. Chem. Soc.* **132** (7)
[50]. Hyberts, S.G., Milbradt, A.G., Wagner, A.B., Arthanari, H., Wagner, G., (2012). Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J. Biomol. NMR* **52**, 315–327.