Supplementary Information for

Modular Genetic Design of Multi-domain Functional Amyloids:

Insights into Self-assembly and Functional Properties

Mengkui Cui[a,b,c,#], Qi Qi[a#], Thomas Gurry[d,#], Tianxin Zhao[a,b,c], Bolin An[a], Jiahua Pu[a], Xinrui Gui[c,e], Allen A. Cheng[f], Siyu Zhang[a], Dongmin Xun[a], Michele Becce[f,g,h], Francesco Briatico-Vangosa[g], Cong Liu[e], Timothy K. Lu[f] and Chao Zhong[a,*]

[a]School of Physical Science and Technology, ShanghaiTech University, Shanghai 200120, China;
[b]University of Chinese Academy of Sciences, Beijing 100049, China;
[c]Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China;
[d]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA;
[e]Interdisciplinary Research Center on Biology and Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China
[f]Department of Electrical Engineering and Computer Science and Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA
[g]Dipartimento di Chimica Materiali e Ingegneria Chimica G. Natta, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italy
[h]Department of Materials, Imperial College London, London SW7 2AZ, United Kingdom (current address)

[#] M.K.C., Q.Q., and T.G. contributed equally to this work.

[*] To whom correspondence should be addressed. E-mail: zhongchao@shanghaitech.edu.cn

Supplementary Information:
Materials and Methods
Supplementary Figures 1-12
References

Materials and Methods

Table of Contents

1. Gene sequences and amino-acid sequences

Gene sequence and amino-acid sequences of CsgA

ATGGGTGTTGTTCCTCAGTACGGCGGCGGCGGTAACCACGGTGGTGGCGG
TAATAATAGCGGCCCAAATTCTGAGCTGAACATTTACCAGTACGGTGGCG
GTAACTCTGCACTTGCTCTGCAAACTGATGCCCGTAACTCTGACTTGACTA
TTACCCAGCATGGCGGCGGTAATGGTGCAGATGTTGGTCAGGGCTCAGAT
GACAGCTCAATCGATCTGACCCAACGTGGCTTCGGTAACAGCGCTACTCT
TGATCAGTGGAACGGCAAAAATTCTGAAATGACGGTTAAACAGTTCGGTG
GTGGCAACGGTGCTGCAGTTGACCAGACTGCATCTAACTCCTCCGTCAAC
GTGACTCAGGTTGGCTTTGGTAACAACGCGACCGCTCATCAGTAC

MGVVPQYGGGGNHGGGGNNSGPNSELNIYQYGGGNSALALQTDARNSDLTI
TQHGGGNGADVGQGSDDSSIDLTQRGFGNSATLDQWNGKNSEMTVKQFGG
GNGAAVDQTASNSSVNVTQVGFGNNATAHQY

Gene sequence and amino-acid sequences of CBD

ATGGCTTGGCAGGTCAACACAGCTTATACTGCGGGACAATTGGTCACATA
TAACGGCAAGACGTATAAATGTTTGCAGCCCCACACCTCCTTGGCAGGAT
GGGAACCATCCAACGTTCCTGCCTTGTGGCAGCTTCAA

MAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQ

Gene sequence and amino-acid sequences of SpyTag

GCGCACATCGTTATGGTCGATGCATATAAACCCACCAAA

AHIVMVDAYKPTK

Gene sequence and amino-acid sequences of Mfp3

ATGGCTGATTATTATGGTCCAAAGTATGGTCCTCCAAGACGCTACGGTGG
TGGCAACTACAATAGATATGGCAGACGTTATGGCGGGTATAAAGGCTGGA
ACAATGGTTGGAAAAGAGGTCGATGGGGACGAAAGTATTAT

MADYYGPKYGPPRRYGGGNYNRYGRRYGGYKGWNNGWKRGRWGRKYY

Gene sequence and amino-acid sequences of Mfp5

ATGAGTTCTGAAGAATACAAAGGTGGTTATTACCCAGGCAATACTTACCA
CTATCATTCAGGTGGTAGTTATCACGGATCCGGCTATCATGGAGGATATA
AGGGAAAGTATTACGGAAAGGCAAAGAAATACTATTATAAATATAAAAA

CAGCGGAAAATACAAGTATCTGAAGAAAGCTAGAAAATACCATAGAAAG
GGTTACAAGAAGTATTATGGAGGTGGTAGCAGT

MSSEEYKGGYYPGNTYHYHSGGSYHGSGYHGGYKGKYYGKAKKYYYKYK
NSGKYKYLKKARKYHRKGYKKYYGGGSS

Gene sequence and amino-acid sequences of Linker

CAGTACGGCGGGGGCTCCGGCGGGGGCTCC

GGGGSGGGGS

2. Protein sequences of two- and three-compositional CsgA-fusion proteins

CsgA-Mfp3 (CsgA-GGGGSGGGGS-Mfp3-HisTag)

MGVVPQYGGGGNHGGGGNNSGPN
SELNIYQYGGGNSALALQTDARN
SDLTITQHGGGNGADVGQGSDD
SSIDLTQRGFGNSATLDQWNGKN
SEMTVKQFGGGNGAAVDQTASN
SSVNVTQVGFGNNATAHQY
GGGGSGGGGS
ADYYGPKYGPPRRYGGGNYNRYG
RRYGGYKGWNNGWKRGRWGRKYYHHHHHH

Mfp5-CsgA (Mfp5-GGGGSGGGGS-CsgA-HisTag)

MSSEEYKGGYYPGNTYHYHSGGS
YHGSGYHGGYKGKYYGKAKKYY
YKYKNSGKYKYLKKARKYHRKG
YKKYYGGGSS
GGGGSGGGGS
GVVPQYGGGGNHGGGGNNSGPN
SELNIYQYGGGNSALALQTDARN
SDLTITQHGGGNGADVGQGSDD
SSIDLTQRGFGNSATLDQWNGKN
SEMTVKQFGGGNGAAVDQTASN
SSVNVTQVGFGNNATAHQYHHHHHH

CBD-CsgA-Mfp3 (CBD-GGGGSGGGGS-CsgA-GGGGSGGGGS-Mfp3-HisTag)

MAWQVNTAYTAGQLVTYNGKTYK
CLQPHTSLAGWEPSNVPALWQLQ
GGGGSGGGGS
GVVPQYGGGGNHGGGGNNSGPN
SELNIYQYGGGNSALALQTDARN
SDLTITQHGGGNGADVGQGSDD
SSIDLTQRGFGNSATLDQWNGKN
SEMTVKQFGGGNGAAVDQTASN
SSVNVTQVGFGNNATAHQY
GGGGSGGGGS
ADYYGPKYGPPRRYGGGNYNRYG
RRYGGYKGWNNGWKRGRWGRKYYHHHHHH

Mfp5-CsgA-CBD (Mfp5-GGGGSGGGGS-CsgA-GGGGSGGGGS-HisTag)

MSSEEYKGGYYPGNTYHYHSGGS
YHGSGYHGGYKGKYYGKAKKYY
YKYKNSGKYKYLKKARKYHRKG
YKKYYGGGSS
GGGGSGGGGS
GVVPQYGGGGNHGGGGNNSGPN
SELNIYQYGGGNSALALQTDARN
SDLTITQHGGGNGADVGQGSDD
SSIDLTQRGFGNSATLDQWNGKN
SEMTVKQFGGGNGAAVDQTASN
SSVNVTQVGFGNNATAHQY
GGGGSGGGGS
AWQVNTAYTAGQLVTYNGKTYK
CLQPHTSLAGWEPSNVPALWQLQHHHHHH

SpyTag-CsgA-CBD (Spytag-**GGGGSGGGGS-CsgA-GGGGSGGGGS-HisTag)**

AHIVMVDAYKPTK
GGGGSGGGGS
GVVPQYGGGGNHGGGGNNSGPN
SELNIYQYGGGNSALALQTDARN
SDLTITQHGGGNGADVGQGSDD
SSIDLTQRGFGNSATLDQWNGKN
SEMTVKQFGGGNGAAVDQTASN
SSVNVTQVGFGNNATAHQY
GGGGSGGGGS
AWQVNTAYTAGQLVTYNGKTYK
CLQPHTSLAGWEPSNVPALWQLQHHHHHH

3. Plasmid construction

CsgA, CBD, SpyTag, Mfp3, Mfp5 gene fragments were amplified with compatible overhangs for Gibson assembly by PCR from templates developed in the group. The pET-22b plasmid was cleaved by Nde I and Xho I (Fermentas FastDigest). The cleaved pET-22b plasmids and the corresponding gene fragments were mixed and incubated with 2×Gibson Assembly Master Mix (NEB) for 1 h at 50 °C, then transformed into DH5*a* competent *E. coli* cells. The transformed strains were smeared to the surface of agar plates containing specific antibiotics.

4. Gene Sequencing

Bacterial colonies growing against specific anbiotics were picked for amplified culture in LB media overnight. A typical mini-prep protocol was then applied to extract plasmids of bacterial colonies for gene sequencing. All sequencing services were provided by Life Technology. All of the sequencing results are illustrated in supplementary figures 1-6.

5. Protein expression, purification

Expression:

The verified recombinant plasmid, pET-22b/CsgA, pET-22b/CsgA-Mfp3, pET-22b/Mfp5-CsgA, pET-22b/CsgA-CBD, pET-22b/CBD-CsgA-Mfp3 and pET-22b/Mfp5-CsgA-CBD were transformed in DE3 (BL21) competent *E. coli*. The strains were cultured to OD_{600} ~1.5 in LB media containing 50 mg/mL ampicillin at 37°C for about 5 h. Protein expression were induced with 0.5 mM IPTG at 30 °C for 2 h. Strains were collected by centrifugation at 4000×g for 10 mins for three times, pellets were stored at -80°C for later use.

Purification:

Every 5 g cell pellets were lysed by 50 ml extraction solution (8 M guanidine hydrochloride (GdnHCl), 300 mM NaCl, 50 mM K2HPO4/KH2PO4, pH 7.2) for 12 h at room temperature. Supernatants of the lysates were collected by centrifugation (Eppendorf, Centrifuge 5810 R) at 12,000×g for 30 mins. 6 mL cobalt resins solution (Talon Metal Affinity Resin) were incubated with the supernatants for 1 hour. Then the cobalt resins that bound His-tagged proteins were collected by centrifugation at 2000×g for 2 min, and washed with 10mL of 50 mM potassium phosphate buffer (300 mM NaCl, 50 mM K2HPO4/KH2PO4, pH 7.2) twice. Resins were then mixed with 5 mL potassium phosphate buffer and were loaded on the gravity chromatography columns (Sangon). Resins were further washed with excess 20 mM imidazole washing buffer (20 mM imidazole and 50 mM potassium phosphate buffer, pH 7.2). GdnHCl and contaminated proteins could be removed by this step. 6 mL elution buffer (300 mM imidazole and 50 mM potassium phosphate buffer, pH 7.2) were added to elute the target proteins.

6. SDS-PAGE, Western Blotting

SDS-PAGE:
60 μL eluted protein solution was mixed with 20 μL of 4×LDS loading buffer (Life Technology). 20 μL of the mixed solution together with the standard protein ladder (Life Technology) were then loaded into the lanes of the mini gels (Life Technology), which were run at 165 volts for 35 min using Nupage MES SDS running buffer (NOVEX). The gels were stained with Coomassie Blue solution (Amresco) for 1 h and then destained with washing solution (methyl alcohol, acetic acid and distilled water with a volum ratio = 4:1:5) for 1 h twice, which were then imaged using a Bio-Rad ChemiDoc MP system.

Western Blotting:
Sample were loaded into the lanes of mini gels and then blotted onto polyvinylidene difluoride membranes using iBlot (Invitrogen). The membranes were blocked with 20% (m/v) milk solution for 1 h and then washed with 1× TBST solution (Sangon) for 2 min for 3 times. The membranes were incubated with anti-Higtag antibodies at a dilution of 1: 10000, followed by washing with 1×TBST solution for 10 min for 3 times. The membranes were then incubated with secondary anti-mouse antibodies at a dilution of 1:5,000 for 1 h. The blotting results were imaged using a Bio-Rad ChemiDoc MP system.

SDS-PAGE and Western Blotting data are shown in supplementary figures 7a.

7. Congo Red Staining
100 μL of 50 μg/mL proteins solutions sitting for a few days were spotted onto Protran BA83 nitrocellulose membranes (Whatman) with a dot blot manifold (Schleicher & Schuell Minifold-I Dot-Blot System). The membranes were then incubated with 20 mL of 0.002 (m/v%) Congo red solution at room temperature for 1 h. They were then washed with excess deionized water for 3 times and were then imaged with GE gel system (Amersham Imager 600). SDS-PAGE and Western Blotting data are shown in supplementary figures 7b.

8. Thioflavin T (ThT) assay
100 μL of freshly purified proteins (20 μM) were loaded into the 96-well plates. ThT solution was then added to a concentration of 20 μM. The fluorescence of the sample was tested every 3 min after shaking 5 s by a BioTek Synergy H1 Microplate Reader using BioTek GEN5 software set to 438 nm excitation and 495 nm emission with a 475 nm cutoff. ThT fluorescence for the polymerization of proteins was normalized by the equation $(F_i-F_0)/(F_{max}-F_0)$. F_i was the ThT intensity (arbitrary fluorescence units) of samples and F_0 was the ThT background intensity. F_{max} was the maximum ThT intensity of samples. Data are shown in supplementary figures 10b and 11b.

9. Atomic Force Microscopy (AFM) imaging
The purified protein solution was diluted to 1 μM. 10 μL of the solution was then incubated on mica surfaces for 18 h without complete drying out. The mica was

rinsed with excess deionized water and blow-dried using nitrogen flow. The sample was then imaged by AFM (Asylum MFP-3D) on AC air tapping mode using Veecoprobes Sb-doped Si cantilevers (ρ = 0.01-0.025 Ω-cm, k = 40 N/m, ν~ 300 kHz). Data are shown in supplementary figures 9 and 10a.

10. Circular Dichroism (CD) spectrum

Far-UV CD spectra were recorded on using an Applied Photophysics spectrometer. Spectra were recorded at room temperature as an average of 5 to 10 scans from 190 to 250 nm. The scan rate was 0.5 step with a response time of 1s. All samples were mixed before every reading. Each sample was tested three times, and the averaged spectrum was reported. The samples tested are aged protein solutions for 18 h and have varying protein concentrations (10 μM).

11. X-Ray fibre diffraction

Mature fibrils formed by different protein samples were pelleted by centrifugation, followed by washing with copious distilled water for several times to remove remaining salts. Fibril pellets were then suspended with 5 μL distilled water. 2 μL suspensions were pipetted between two fire-polished glass rods and drying for a couple of hours. The diffraction data were collected by an in-house X-ray machine equipped with a Rigaku Micromax-007 X-ray generator and an R-Axis IV++ area detector.

12. Quantitative assessment of Young's modulus by PK-QNM

Freshly purified protein solution was incubated on the surface of mica for 18 h. Samples were rinsed, dried and then determined with AFM (Bruker Fast Scan) in Peak force quantitative nanomechanics (QNM) mode by using RTESPA-100 (40N/m, 300kHz, Rotated Tip, Al Reflective Coating). Values of Young's modulus can be extracted from the DMT modulus channel. Quantitative analysis for each sample was collected for further analysis. Data are shown in supplementary figures 12.

13. Binding behaviors of amyloid fibrils towards chitin substrates

The functional activity of CBD was assessed using quantitative THT fluorescence assay. 50 μL washed chitin magnetic beads suspension were incubated in excess protein fibril solutions of different compositions. After incubating for 1 hour, chitin beads were deposited through magnet and were washed for three times with 500 μL chitin washing buffer to remove loosely adsorbent fibrils on the surfaces. Chitin magnetic beads were then mixed with 20 mM ThT solution and tested with THT assay using protocol described above. Functional activity of amyloid fibrils was compared by correlating quantitative fluorescence strength between chitin magnetic beads and CBD within amyloid fibrils.

14. Quartz Crystal Microbalance-Dissipation (QCM-D)

QCM-D measurements were performed with a Q-Sense Omaga Auto system. Quartz crystals with fundamental frequencies of ~5 MHz were spin coated with chitin

nanofibers using a spin coater (Polos). Fibrils suspensions were injected into the flow cell with a rate of typically 20 μL min^{-1}. Change in frequency (ΔF) and dissipation (ΔD) were measured at six overtones (n=3, 5 … 13) simultaneously. Protein fibrils were incubated on the surface for 2 h. The adsorbed protein on the crystal was then washed with distilled water for 1 h to remove any non-specifically bound material.

15. Construction of initial models

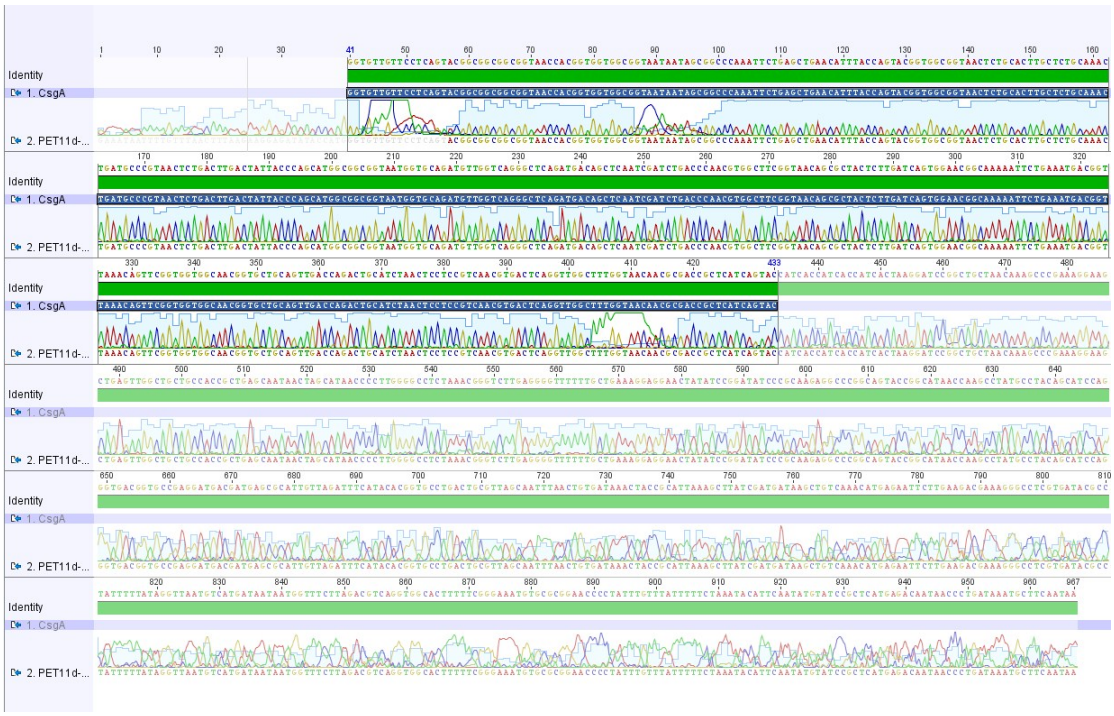
Models of the CBD-CsgA-Mfp3 and Mfp5-CsgA-CBD monomers and fibrils were constructed in a manner analogous to previous work.[1] Briefly, the CsgA moiety was constructed by threading the the CsgA amino-acid sequence onto a structure of an amyloid β42 fibril (PDB ID 2BEG)[2] using Modeller.[3] The linker and Mfp3/5 structures were constructed using CHARMM.[4] The CBD moiety was built from a solution NMR structure of the chitin binding domain of *B. circulans* chitinase (PDB ID 1ED7).[5] Individual segments were joined to create single monomers of the different fusion proteins, which were then minimized in CHARMM to prevent bad contacts. Fibril structures composed of four monomers each were constructed in CHARMM by fixing the relative motions of individual monomoers and minimizing the H-bond distances between adjacent monomers such that the stacked monomers made contacts in a non-staggered, in-register conformation.

16. Molecular dynamics simulations

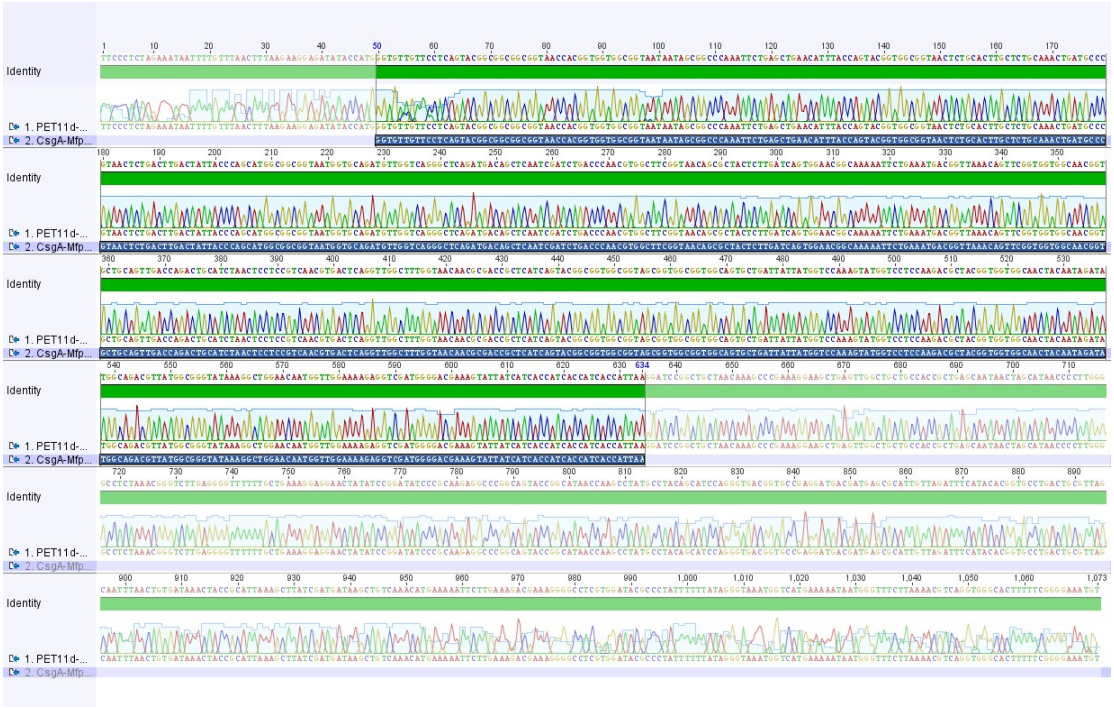
Simulations were performed using a polar hydrogen model in the CHARMM19 forcefield in EEF1 implicit solvent.[6] In all cases, the systems were linearly heated to a temperature of 293.15K over 100ps and coupled to a Nose thermostat at the same temperature. The fibril assembly steps were performed during a 1 ns equilibration run by introducing the appropriate H-bond restraints, and the fibrils were further equilibrated for 1 ns before fixing the CsgA moieties to increase computational speed, since the model of the fibril core has been previously shown to be stable over simulation timescales in both CsgA and fusion construct fibrils [1]. Monomer systems were simply equilibrated for 1 ns. Dynamics were run using a timestep of 2 fs at constant pressure and temperature using the Leapfrog Verlet integrator. Monomers were simulation for a total of 1us and fibrils for a total of 100ns.

17. Statistics

Fibril length and height was obtained by statistics from 40 randomly picked fibril spots. Young's modulus was obtained by statistics from 25 randomly picked fibril spots. Statistical analysis for ThT fluorescence was average of 5 independent tests. Data are presented as mean ± s. d. (standard deviation). A Student's t-test was used to compare data sets, and P-values less than 0.01 and 0.05 were considered statistically significant between two marked groups or samples. Data is shown in supplementary figures 8 and 11a.

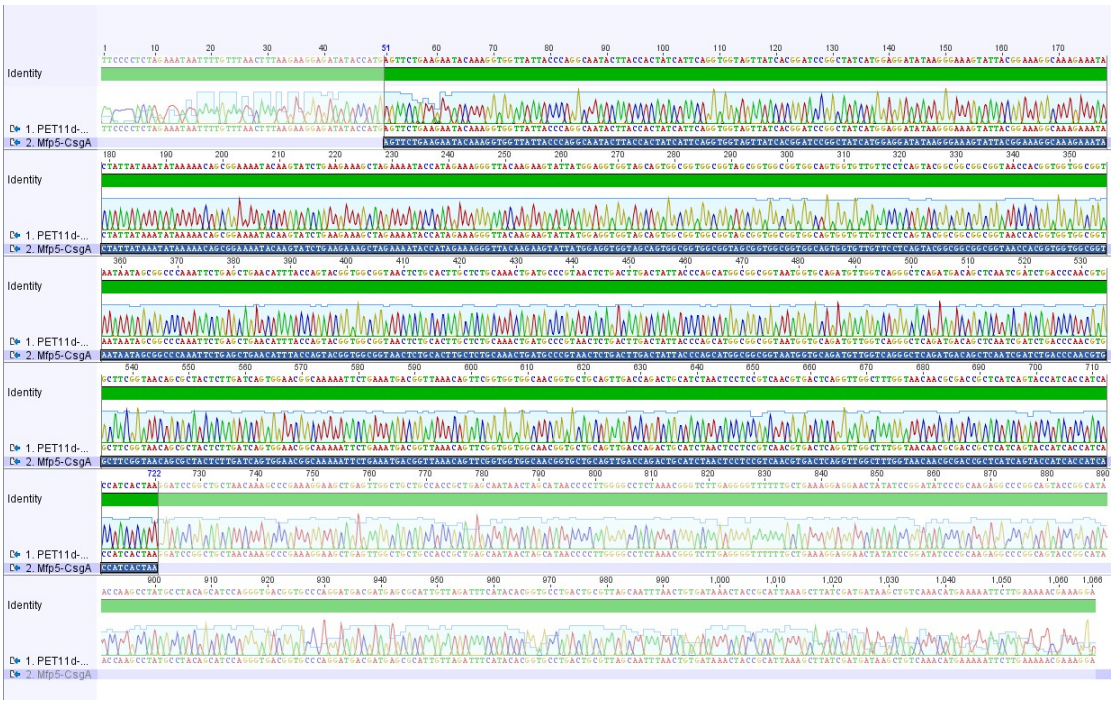


Supplementary Figure 1. Sequencing results of the vector expressing CsgA. The highlighted blue band is the reference sequence and the green band represents the consensus sequence between sequencing results and the reference sequence.

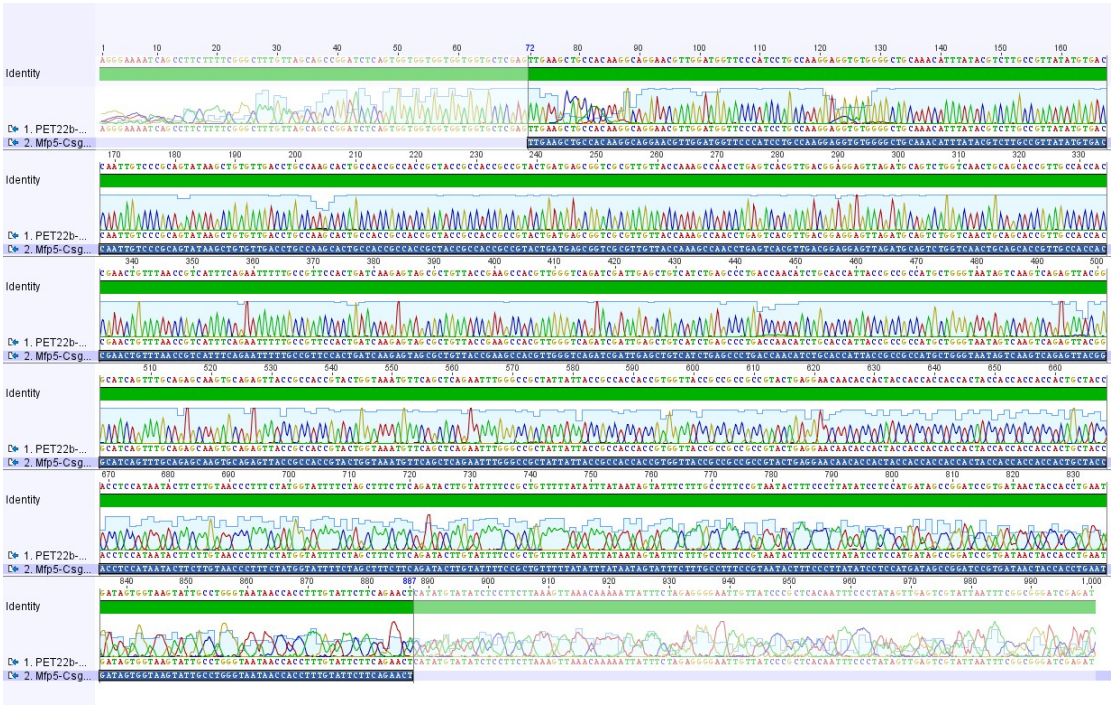


Supplementary Figure 2. Sequencing results of the vector expressing CsgA-Mfp3. The highlighted blue band is the reference sequence and the green band represents the consensus sequence between sequencing results and the reference sequence.

Supplementary Figure 3. Sequencing results of the vector expressing CBD-CsgA-Mfp3. The highlighted blue band is the reference sequence and the green band represents the consensus sequence between sequencing results and the reference sequence.

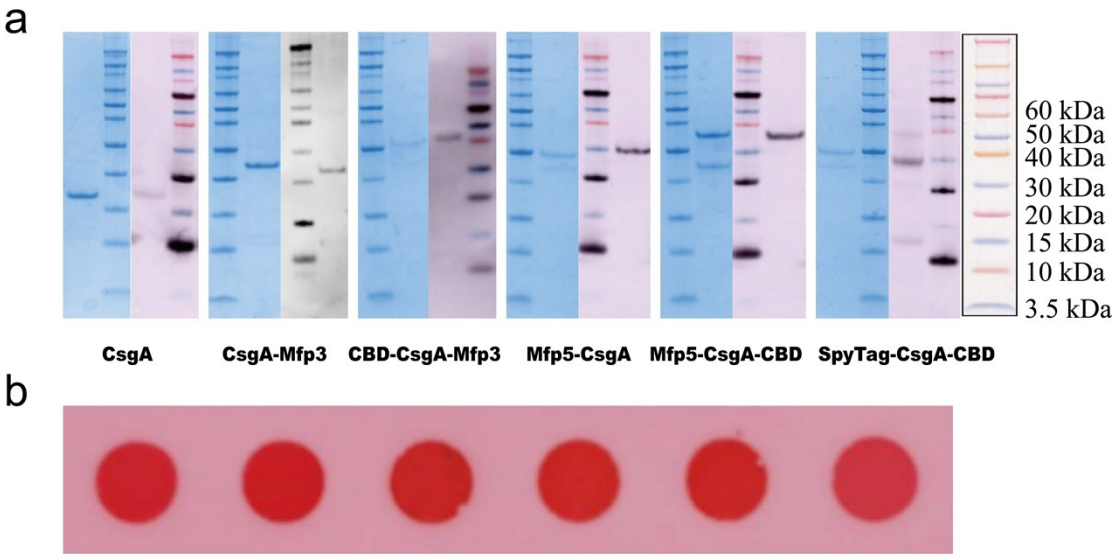


Supplementary Figure 4. Sequencing results of the vector expressing Mfp5-CsgA. The highlighted blue band is the reference sequence and the green band represents the consensus sequence between sequencing results and the reference sequence.

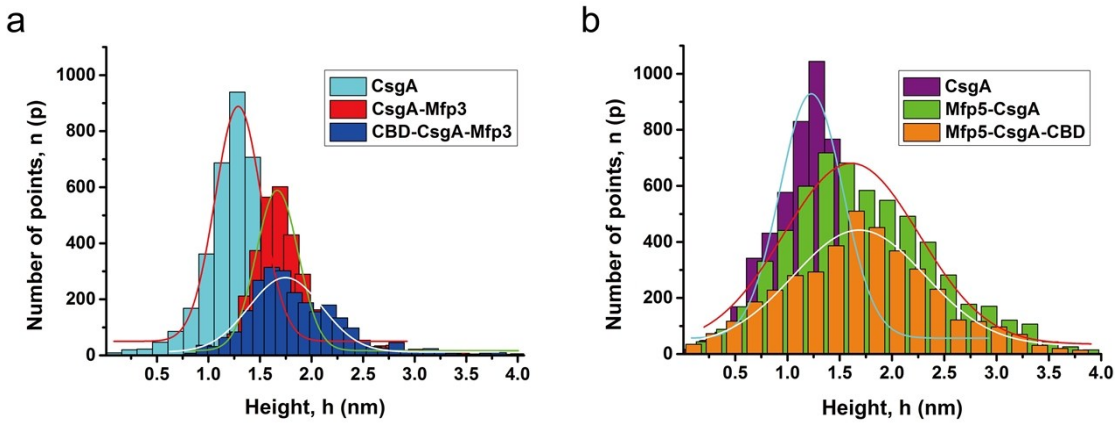


Supplementary Figure 5. Sequencing results of the vector expressing Mfp5-CsgA-CBD.The highlighted blue band is the reference sequence and the green band represents the consensus sequence between sequencing results and the reference sequence.

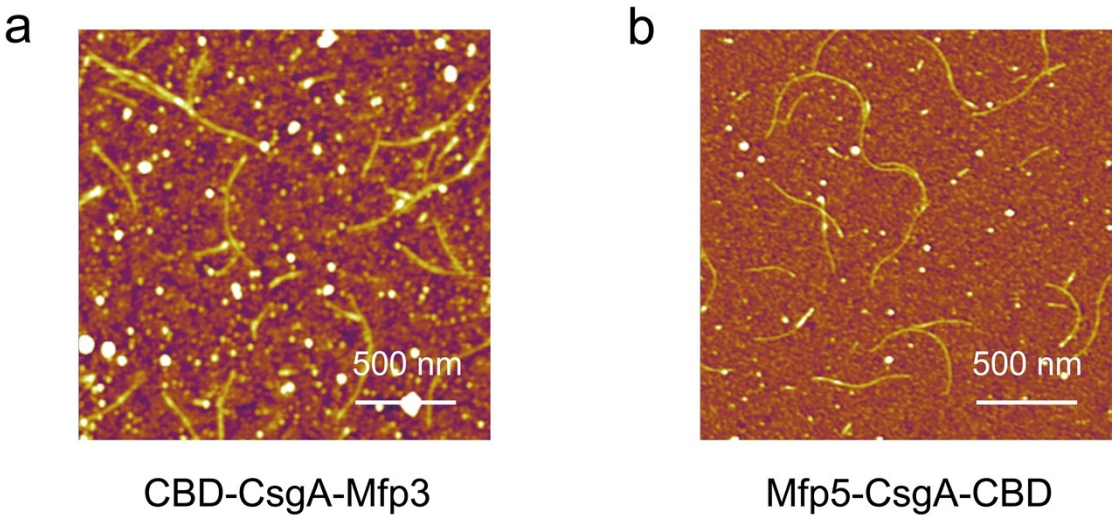
Supplementary Figure 6. Sequencing results of the vector expressing SpyTag-CsgA-CBD.The highlighted blue band is the reference sequence and the green band represents the consensus sequence between sequencing results and the reference sequence.



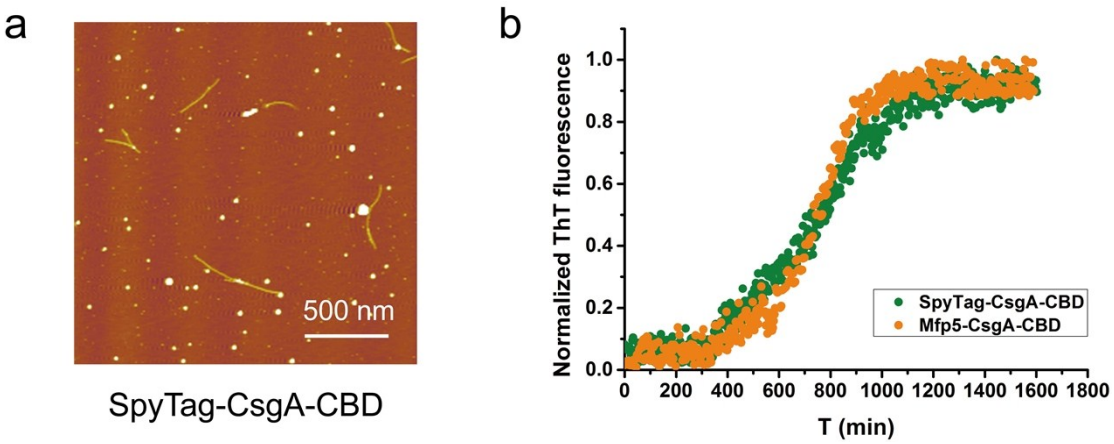
Supplementary Figure 7. (a) SDS-PAGE and western blots of proteins. (b) Congo red staining of fibril aggregations, which are corresponding to fibrils in figure (a).



Supplementary Figure 8. Statistical analysis of diameters of various protein fibrils based upon AFM height images. The number of measured points above certain height (Y axis) as a function of given height range (X axis) were plotted. After fitted with Gaussian function for each sample, diameter of fibrils of that type could be extracted. (a) The diameters of group 1 are: 1.39 ± 0.48 nm for CsgA, 1.66 ± 0.50 nm for CsgA-Mfp3 and 1.75 ± 0.46 nm for CBD-CsgA-Mfp3. (b) The diameters of group 2 are: 1.39 ± 0.48 nm for CsgA, 1.61 ± 0.45 nm for Mfp5-CsgA and 1.67 ± 0.43 nm for Mfp5-CsgA-CBD.



Supplementary Figure 9. AFM images of fibrils after incubation on mica surfaces for 48 h. (a) CBD-CsgA-Mfp3 fibrils. (b) Mfp5-CsgA-CBD fibrils. Compared with samples incubated for 18 h, the length of fibrils did not increase obviously, suggesting an obvious steric hindrance effect for fusion proteins containing three domains.

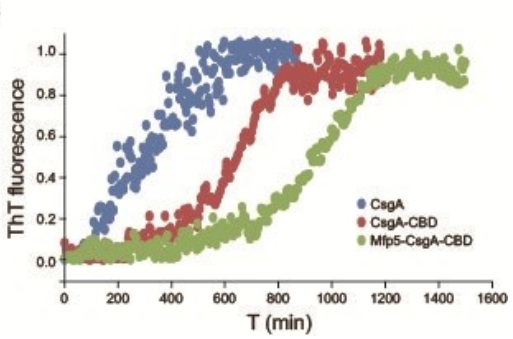


Supplementary Figure 10. Morphology and dynamic assembly process of Spytag-CsgA-CBD (compared with Mfp5-CsgA-CBD). (a) AFM images of SpyTag-CsgA-CBD fibrils after incubation on mica surfaces for 18 hours, resembling those collected for Mfp5-CsgA-CBD (Fig 2.) (b) ThT curves of SpyTag-CsgA-CBD and Mfp5-CsgA-CBD. Substituting Mfp5 domain with a smaller size SpyTag domain didn't significantly affect the morphology and assembly kinetics of fibrils, suggesting the size of a third domain has little influence on fibril assembly.

a

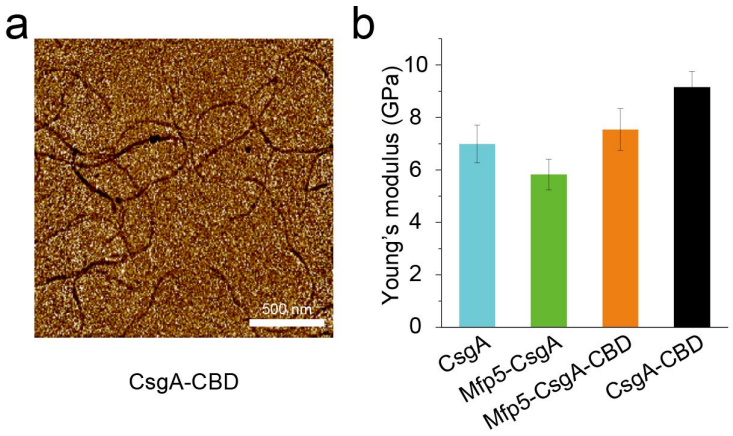
Fibril type	Fibril length (nm)	Fibril height (nm)
CsgA	3190.20 ± 1169.44	1.39 ± 0.48
CsgA-CBD	1120.28 ± 15.18	1.52± 0.05
Mfp5-CsgA-CBD	260.26 ± 125.88	1.67 ± 0.43

b



Supplementary Figure 11. Comparison of fibril morphology features and dynamic assembly process of CsgA, CsgA-CBD, and Mfp5-CsgA-CBD. (a) Fibril length and height comparison that summarized from AFM images of fibrils incubation on mica surfaces for 18 hours. (b) ThT curves of CsgA, CsgA-CBD and Mfp5-CsgA-CBD. This set of data shows similar influence trend of fusion domains on fibril structures and assembly kinetics as described in figure 3c.

Note: Although the number of amino acids of CBD domains is identical to that of Mfp3 and less than that of Mfp5, the height (1.52 ± 0.05) of the CsgA-CBD fibril is smaller than that of CsgA-Mfp3 and Mfp5-CsgA fibrils. This can be explained by the compact conformation of CBD domain, which likely self-folds into condensed β-sheet structure, in contrast with the loose conformation of random coil structure.



Supplementary Figure 12. (a) Modulus mapping images of fibril variants performed by AFM peak force quantitative nanomechanics (PF-QNM) following a Derjaguin, Muller, Toropov (DMT) model. Scale bars, 500 nm. (b) Stiffness comparison of CsgA and CsgA-fusion proteins (Mfp5-CsgA, Mfp5-CsgA-CBD, CsgA-CBD).

References

1. C. Zhong, T. Gurry, A. A. Cheng, J. Downey, Z. Deng, C. M. Stultz and T. K. Lu, *Nature nanotechnology*, 2014, **9**, 858-866.
2. T. Luhrs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Dobeli, D. Schubert and R. Riek, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 17342-17347.
3. B. Webb and A. Sali, *Current Protocols in Bioinformatics*, 2014, **47**, 1-32.
4. Bernard R. Brooks and M. K. e. al, *Journal of Computational Chemistry*, 1983, **4**, 187-217.
5. M. S. e. a. Takahisa Ikegami, *The Journal of biological chemistry*, 2000, **275**, 13654-13661.
6. Themis Lazaridis and M. Karplus, *PROTEINS: Structure, Function, and Genetics*, 1999, **35**, 133-152.