Extracellular expression of scaffoldin mini-CipA from *Clostridium thermocellum* by *Bacillus subtilis*

Nguyen Hoang Ngoc Phuong[1*], Phan Thi Phuong Trang[2], Tran Linh Thuoc[1] and Nguyen Duc Hoang[2]
1. Faculty of Biology and Biotechnology, Vietnam National University Ho Chi Minh City-University of Science, VIETNAM
2. Center for Bioscience and Biotechnology, Vietnam National University Ho Chi Minh City-University of Science, VIETNAM
*nhnphuong@hcmus.edu.vn

Abstract

One of the strategies for producing bioproducts from the abundant plant biomass is designing cell factories for consolidated bioprocessing. Clostridium thermocellum (Acetivibrio thermocellus), an anaerobic and thermophilic bacterium, secrets protein complexes called cellulosomes which can degrade low-accessibility lignocellulose due to synergistic activities from multi-functional cellulosomal enzymes. Scaffoldin CipA is a primary structural component of the cellulosome from C. thermocellum which is responsible for the assembly of cellulosomal enzymes on plant cell walls and the attachment of the cellulosome to carbohydrate fibers.

This research focused on designing a truncated extracellular scaffoldin CipA, called mini-CipA, by cloning it in Escherichia coli and expressing it extracellularly in Bacillus subtilis WB800N. This safe facultative aerobic probiotic can multiply to high cell density on inexpensive substrates. SDS-PAGE and Western blot analyses of the culture media of Bacillus subtilis expressing mini-CipA showed bands lower than 100 kDa which were consistent with the theoretical molecular weight of mini-CipA. Additionally, ten peptides from mini-CipA were detected using High-performance liquid chromatography - tandem mass spectrometry. The number of detected proteins in the secretome of Bacillus subtilis expressing mini-CipA was approximately 30% of those in the control sample.

This research shows promise in utilizing mini-CipA from culture media without the need for a purification step.

Keywords: *Clostridia*, CipA, mini-cellulosome, protein engineering, secretion.

Introduction

Plant biomass, an abundant renewable resource, has many applications in food packaging, water treatment, bioproducts and bioenergy[24]. In the pre-processing of plant biomass, several conventional chemical or physical methods such as acid hydrolysis or steam explosion have been used to overcome lignocellulose recalcitrance. However, these methods are energy-consuming, environmentally unfriendly and high maintenance demanding. On the other hand, a more affordable method of biocatalysis has played an increasingly important role in plant cell wall treatment[32]. High concentrations and diversity of enzymes on crystalline regions are necessary to overcome the barriers of lignocellulose degradation[23].

Cellulosomes are known as highly efficient macromolecular protein complexes for deconstructing plant cell wall complex carbohydrates which are secreted by cellulolytic microorganisms. They can bind to cell surfaces or can be free in culture media[21]. The primary structural components responsible for assembling of multifunctional enzymes on lignocellulose are scaffoldins which enhance hydrolysis in low-accessibility cellulose (Avicel) more than in high-accessibility amorphous cellulose[38]. In *C. thermocellum*, the scaffoldin CipA comprises of nine type-1 cohesins, a carbohydrate binding module (CBM) and a type-2 dockerin[8]. The type-1 cohesins associate specifically with type-1 dockerin in the catalytic enzymes. The CBM attaches the cellulosome to lignocellulose fibers. The type-2 dockerin interacts specifically with the type-2 cohesin of the cell surface proteins. A CipA deletion mutant in *C. thermocellum* reduced the cellulose solubilization rate by 100-fold[26].

Because *C. thermocellum* and other microbes secreting cellulosomes are anaerobic microorganisms that grow slowly and require high maintenance conditions in large scale cultivation, modified industrial bacteria and yeasts have been used to express cellulosomal genes. However, the large molecular weight of approximately 197 kDa is an obstacle for full-length transgenesis of scaffoldin CipA. Therefore, many experiments on designer scaffoldins containing from one to nine type-1 cohesins have been performed in *Escherichia coli* DH5α[9], *Bacillus subtilis* RM125[5], *Lactococcus lactis* HtrA NZ9000[34], *Kluyveromyces marxianus*[1] and *Saccharomyces cerevisiae* EBY100[6].

Cellulosomes and their cellulosomal catalytic units have been intracellularly expressed to study their biochemical and physicochemical characteristics *in vitro*. They have been also genetically modified to increase the durability of the cellulosomal enzymes[17,37] and expand their ability to use different substrates[7]. However, the expression of cellulosomal proteins intracellularly in *E. coli* has resulted in the formation inclusion bodies[22,33].

Additionally, the approach of cell-surface expression using anchoring proteins or linkers to mimic natural cellulosomes has also been experimented[1,31,34]. The hydrolytic activity of the modified strains was higher than that of mixed free enzymes but the expression density on the cell surface

decreased when increasing the size of cellulosomes on surface expression.

Another approach of secreting expression has some drawbacks such as low natural secretion of proteins[18] and protein degradation by host-cell proteases[16]. A derivative of the *B. subtilis* 168 strain deficient in 8 proteases, *B. subtlilis* WB800N[15], could reduce protein degradation in extracellular expression. In this research, we designed extracellular mini-CipA from *C. thermocellum* and analyzed the secreted proteins in *B. subtilis* WB800N culture media.

Material and Methods

Plasmids, strains and culture media: The *E. coli* - *B. subtilis* shuttle plasmids pHT were used for gene manipulation in *E. coli* OmniMAX™ (Invitrogen, USA) and gene expression in *B. subtlis* WB800N (MoBiTec, Germany). They contained an expression cassette controlled by an isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible P_{grac} promoter, as well as two antibiotic resistant genes (*bla* and *cat*). Selection markers included 100 µg/ml ampicillin (Amp) for *E. coli* and 10 µg/ml chloramphenicol (Cm) for *B. subtilis* at final concentration with cultures grown in Luria Broth (LB). *C. thermocellum* DSM1237/ATCC27405 (DSMZ, Germany) were cultured in DSMZ #122 medium for genome extraction to amplify cellulosomal genes[4].

Mini-CipA cloning: Mini-CipA was created by removing six type-1 cohesins from the original CipA in two steps (Fig. 1A). In the first step, the 5' terminal DNA fragment of *cipA* (EMBL/EBI: L08665.1) coding for two type-1 cohesins and a CBM (*coh1-coh2-cbm*) was amplified by PCR with the *C. thermocellum* chromosome (Genbank:CP000568.1) as the template. This fragment was inserted into pHT backbone plasmid to create pHT1725. In the second step, the 3' terminal DNA fragment of *cipA* coding for the last type-1 cohesin and the type-2 dockerin of *cipA* (*coh9-dock2*) was amplified and inserted into to create pHT1701. The plasmids were screened by colony PCR before DNA sequencing.

E. coli OmniMAX™ carrying plasmids were cultured in Luria Broth containing ampicillin (LB-Am) media at 37°C with shaking at 250 rpm overnight. Plasmid characteristics and primer sequences were provided in table 1 and table 2. A DNA ladder (200 to 2000 bp ZipRuler Express DNA Ladder, Fermentas) was used as a weight marker in DNA gel electrophoresis. Primer synthesis and DNA sequencing were conducted by Macrogen, Korea.

Integration of mini-CipA into the *B. subtilis* WB800N genome: *B. subtilis* WB800N colonies transformed by pHT1701 were cultured in Luria Broth medium agar containing chloramphenicol (LB-Cm) at 37°C overnight and used as templates for amplifying the fragments *yvfM-5'lacA*, *3'lacA-yvfO* and *mini-CipA*. Three pairs of PCR primers were used to identify the *B. subtilis* WB800N colonies in which *mini-cipA* integrated into their chromosome at the *lacA* locus (Fig. 2A).

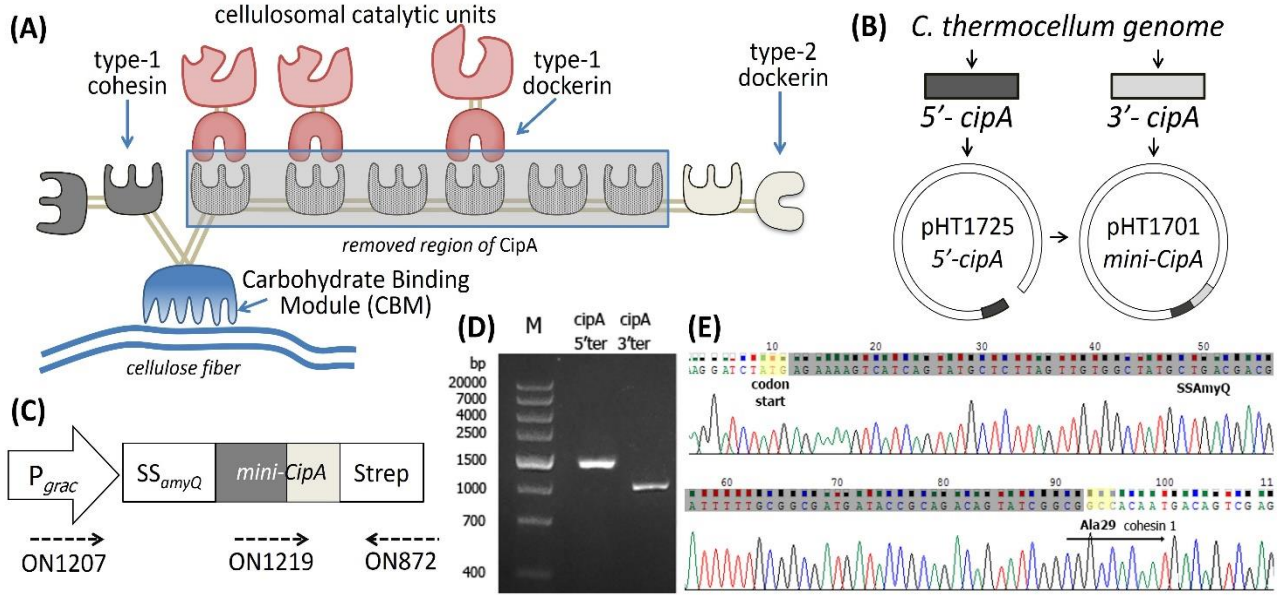


Fig. 1: Plasmids coding for mini-CipA were designed in this study.

(A) Mini-CipA which was created by removing the middle region of CipA contained three type-1 cohesins, a Carbohydrate Binding Module (CBM) and a type-2 dockerin. (B) pHT1701 coding for mini-CipA was created via 2 steps. Step 1: 5'*cipA* was amplified from the *C. thermocellum* genome and inserted into the pHT backbone plasmid to create pHT1725. Step 2: 3'*cipA* was amplified from *C. thermocellum* genome and inserted into the pHT1725 to create pHT1701. (C) Expression cassette of pHT1701 contains the promoter P_{grac}, signal sequence SS_{AmyQ}, *mini-cipA*, Strep II tag and primers for sequencing (dashed arrows). (D) PCR products of 5' terminus (1608 bp) and 3' terminus (1020 bp) of *mini-cipA* amplified from *C. thermocellum*. (E) DNA sequencing of pHT1701 using the primer ON1207 for the regions of SS_{AmyQ} and type-1 cohesin.

Sodium Dodecyl Sulfate-Polyacrylamide gel electrophoresis (SDS-PAGE): *B. subtilis* WB800N carrying the plasmids were cultured in 5 mL of LB-Cm media with 1 mM IPTG at 37°C for 18 hours. Supernatants were collected by centrifuging the culture media at 13,000 g for 10 minutes. 1600 µL of supernatants were mixed with 400 µL of 50% trichloroacetic acid (TCA) in Eppendorf tubes and then placed on ice for 30 minutes for protein precipitation.

The mixtures were then centrifuged at 13,000 g for 10 minutes and the pellets were collected. The pellets were washed twice with chilled acetone of 400 µL each time and centrifuged at 13,000 g for 5 minutes.

Table 1
***E. coli* - *B. subtilis* shuttle plasmids were designed in this research.**

S.N.	Plasmids	Coding genes	Characteristics
1	pHT1725	*coh1-coh2-cbm*	Amp^R, Cm^R, 5'-*lacA*, 3'-*lacA*, P_{grac}, SS_{AmyQ}, *coh1-coh2-cbm*, Strep-tag
2	pHT1701	*coh1-coh2-cbm-coh9-dock2*	Amp^R, Cm^R, 5'-*lacA*, 3'-*lacA*, P_{grac}, SS_{AmyQ}, *coh1-coh2-cbm-coh9*-dock2, StrepII-tag

pHT: *E. coli* - *B. subtilis* shuttle vectors. Amp^R: Ampicilin resistant gene, also known as *bla* gene, codes for enzyme β-lactamase. Cm^R: Chloramphenicol resistant gene, also known as *cat* gene, codes for the enzyme Chloramphenicol acetyltransferase. 5'-*lacA* and 3'-*lacA*: 5-terminal and 3'-terminal regions of *lacA* are used for integration into *B. subtilis* WB800N chromosome. P_{grac}: Promoter P_{grac}, SS_{AmyQ}: signal sequence o f enzyme amylase AmyQ (SS_{AmyQ}) from *Bacillus amyloliquefasciens*, *coh*: domain coding for type-1 cohesin, *dock2*: domain coding for type-2 dockerin, *cbm*: domain coding for Carbohydrate Binding Module, Strep-tag: Streptavidin II tag.

Table 2
Primers were used in this research.

Purpose	Primers	Templates	Products (bp)	Nucleic acid sequence (5'-3')
Obtaining *coh1-coh2-cbm*	ON1211	*C. thermocellum* genome	1608	CCAGGTCTCAGATCTATGAGAAAAGTCATCAGTATGCTCTTAGTTG
	ON1214			CCAGGTCTCTCTAGATACACTGCCACCGGGTTCTTTAC
Obtaining *coh9- dock2*	ON1215		1020	AGGATCCATGTCTAGAGATACAACAGTACCTACAACATCGCC
	ON1216			GGATGGCTCCAAGCGACGTCCTGTGCGTCGTAATCACTTGATGTAG
PCR colony for pHT1725	ON1211	*E. coli* /pHT1725	1518	CCAGGTCTCAGATCTATGAGAAAAGTCATCAGTATGCTCTTAGTTG
	ON904			TTTACATCAATCACGGTTGCATAATCC
PCR colony for pHT1701	ON872	*E. coli* /pHT1701	1200	AGCATCAGCAGGATCCTATGATGCTTCACTTATTCCGAATC
	ON1219			TAGGCGGGCTGCCCCGGGTTAG
Sequencing pHT1701	ON1207	pHT1701		GGGAGATTCTTTATTATAAGAATTG
	ON872			TCCTTTTTTTGAATTCAAGATCTCCATGGACGCGTGACG
	ON1219			GTCTGCTTCACAGTTTGTTGAATG
Testing *yvfM-lacA5* region	ON1441	*B. subtilis* WB800N /pHT1701	879	CTATACGACATTTGCGGCCGGAG
	ON1691			GGTTATCATGCAGGATTGTTTATGAACTC
Testing *lacA3-yvfO* region	ON945		747	GCGTCCATGGAGATCTATCCGGTTGTTACTCGCTCACATTTATCG
	ON1442			GATCCTCTGCCCGAAGCTCTGAC
Testing *mini-CipA* region	ON1345		563	CTAAATTATACGGCGACGTTGTACCGACAAACACACCGACA
	ON1214			CCAGGTCTCTCTAGATACACTGCCACCGGGTTCTTTAC

Laemmli sample buffer of 40 µL was added to the pellets for SDS-PAGE. The mixtures were vortexed and heated at 80°C for 10 minutes for protein denaturation. Subsequently, 10 µL of the protein sample was added per well of a 1% agarose gel for SDS-PAGE electrophoresis at 25 mA and 400 mV for 70 minutes. A protein ladder (10 to 200 kDa PageRuler™ Unstained Protein Ladder, Thermo Fisher Scientific) was used as weight markers.

Western Blot: Samples for SDS-PAGE electrophoresis were prepared following the SDS-PAGE analysis procedure except a pre-stained protein ladder (10 to 250 kDa PageRuler™ Pre-stained Protein Ladder, Thermo Fisher Scientific). After completing SDS-PAGE electrophoresis, the proteins in the agarose gel were transferred to nitrocellulose membrane. The membrane was then blocked in phosphate-buffered saline with tween (PBS-T) containing 1% bovine serum albumin (BSA) and 0.1% tween-20 for 1 hour at room temperature.

The blocked membrane was gently shaken at room temperature for 1 hour in 20 mL of PBS-T containing an anti-Strep mouse antibody IgG linked with horseradish peroxidase (GenScript, anti-StrepII mAb-HRP, 5A9F9) to detect *mini-CipA* fused with a StrepII tag at the C-terminal region. A chromogenic substrate of HRP, 3,3′,5,5′-Tetramethylbenzidine (TMB, Thermo Fisher Scientific), was added to the membrane to detect the mini-CipA-antibody interaction.

High Performance Liquid Chromatography-Tandem Mass Spectrometry (HPLC-MS/MS): *B. subtilis* WB800N carrying pHT1701 and secreting mini-CipA were cultured in 5 mL of LB-Cm with 1 mM IPTG at 37°C for 18 hours. The control sample was prepared with *B. subtilis* WB800N carrying the plasmid backbone of non-coding genes. Supernatants containing secreted proteins were collected by centrifuging the culture media at 13,000 g for 10 minutes. After precipitation with chilled acetone at a 1:4 ratio for 4 hours, the secreted proteins were digested using trypsin/P (Promega) in 10 mM $NaHCO_3$ pH 8 with a ratio of 1/100 (w/w) and incubated at 37°C for 24 hours. The peptides were then cleaned using a C18 desalting tip and injected into a HPLC system (Easy-nLC 1000 L, Thermo Scientific) paired with a Quadruple Time-of-flight and Orbitrap Spectrometer (Q-Exactive, Thermo Scientific).

The data was acquired using Thermo Xcalibur and analyzed using Matrix Science's Mascot algorithm in Thermo Proteome Discoverer v1.4. Protein sequences were obtained from the *B. subtilis* 168 protein database (Uniprot txid 224308) combined with the protein sequences for mini-CipA from pHT1701 and common contaminants.

Results

Cloning mini-CipA in *E. coli*: Gene amplification from the *C. thermocellum* chromosome produced the *5'-cipA* and *3'-cipA* fragments, measuring 1608 bp and 1020 bp in length respectively (Fig. 1D). These fragments were cloned into *E. coli* to generate pHT1725 and pHT1701. DNA sequencing confirmed that pHT1701 encoded the designer mini-CipA (Fig. 1E).

Integration of *mini-CipA* into *B. subtilis* WB800N genome: PCR products from *B. subtilis* WB800N colonies successfully integrated with *mini-cipA* at the *lacA* locus contained three fragments: *yvfM-lacA*5' (879 bp), *lacA*3'-*yvfO* (747 bp) and *mini-cipA* (563 bp). *Mini-cipA* integrated successfully into the genomes of colony 1 and colony 2 while the colony 3, which only showed the PCR product of *mini-cipA*, failed to integrate into the genome (Fig. 2B).

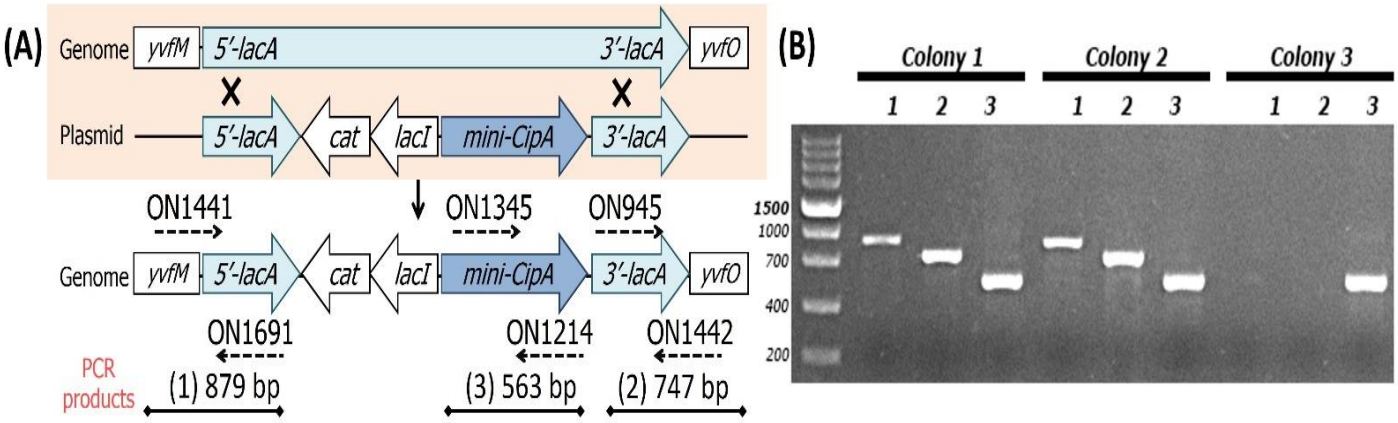


Fig. 2: Integration of mini-CipA into the *Bacillus subtilis* WB800N genome.
(A) *lacA* is between *yvfM* and *yvfO* in *Bacillus subtilis* WB800N genome. Plasmid pHT1701 contains the 5'-terminal and 3'-terminal regions of *lacA* labeled 5'-*lacA* and 3'-*lacA*. The crossover between the *lacA* regions in the genome and the plasmid results in the integration of *mini-cipA* into the genome. Colony PCR products using three pairs of primers (shown as dashed arrows) for three regions, *yvfM-5'lacA*, *3'lacA-yvfO* and *mini-CipA*. (B) Successful integration of mini-CipA into the genomes was confirmed by colony PCR products with three pairs of primers in colony 1 and colony 2. In contrast, the unsuccessful integration in colony 3 only showed the PCR product of *mini-cipA* (563 bp).

SDS-PAGE and Western blot: *B. subtilis*/pHT1701 samples displayed a protein band below 100 kDa, consistent with the theoretical molecular weight of mini-CipA (89 kDa) on the SDS-PAGE gel (Fig. 3A, red dot). The Western blot results using anti-strep antibody also showed a band consistent with mini-CipA fused with a strep tag at the C-terminus (Fig. 3B, red dot). However, both control and mini-CipA extracellular samples displayed a band below 130 kDa of an unknown protein from the host strain.

HPLC-MS/MS: The extracellular secretome of *B. subtilis* carrying pHT1701 was analyzed using HPLC-MS/MS. Ten peptides with high Mascot score (≥ 25) covered functional regions of mini-CipA (Fig. 3C and table 3). Fragment ions of the 6th peptide (VTNTGSSAIDLSK) in the CBM region were displayed in fig. 3E. Most theoretical b and y ions were detected in the observed data. The results of SDS-PAGE, Western blot and HPLC-MS/MS indicated that mini-CipA was expressed extracellularly by *B. subtilis*.

Secretome analysis: In the mini-CipA sample, 120 proteins were detected which was approximately 30% of the 390 proteins in the control sample as shown in fig. 4. Proteins involved in metabolic pathways related to carbohydrate degradation and cell wall biogenesis were more prominent than other pathways.

Fig 5 and table 4 display the proteins with molecular weights ranging from 60 to 150 kDa detected in the control and mini-CipA samples. Their emPAI values indicated the expression levels of detected proteins[14]. While the unknown protein from the host strain at 130 kDa band in the Western blot result appeared in both the control and mini-CipA samples, many proteins existed only in each sample. We suspect the protein at 130 kDa in the Western blot result could be isoleucine-tRNA ligase (Q45477, Fig. 5) because its mass was consistent with the band. Neighboring proteins that appeared in both samples, such as ribonucleoside-diphosphate reductase NrdEB subunit alpha (O31875) and cell wall-associated protease (P54423), could potentially be cleaved into smaller proteins according to the protein database Uniprot.

Discussion

Previous reports on cell surface expression of mini-cellulosome often involved removing the type-2 dockerin of mini-CipA and fusing with a cell wall sorting signal (CWSS) to covalently attach the mini-cellulosome to the *B. subtilis* host cell wall[2,13]

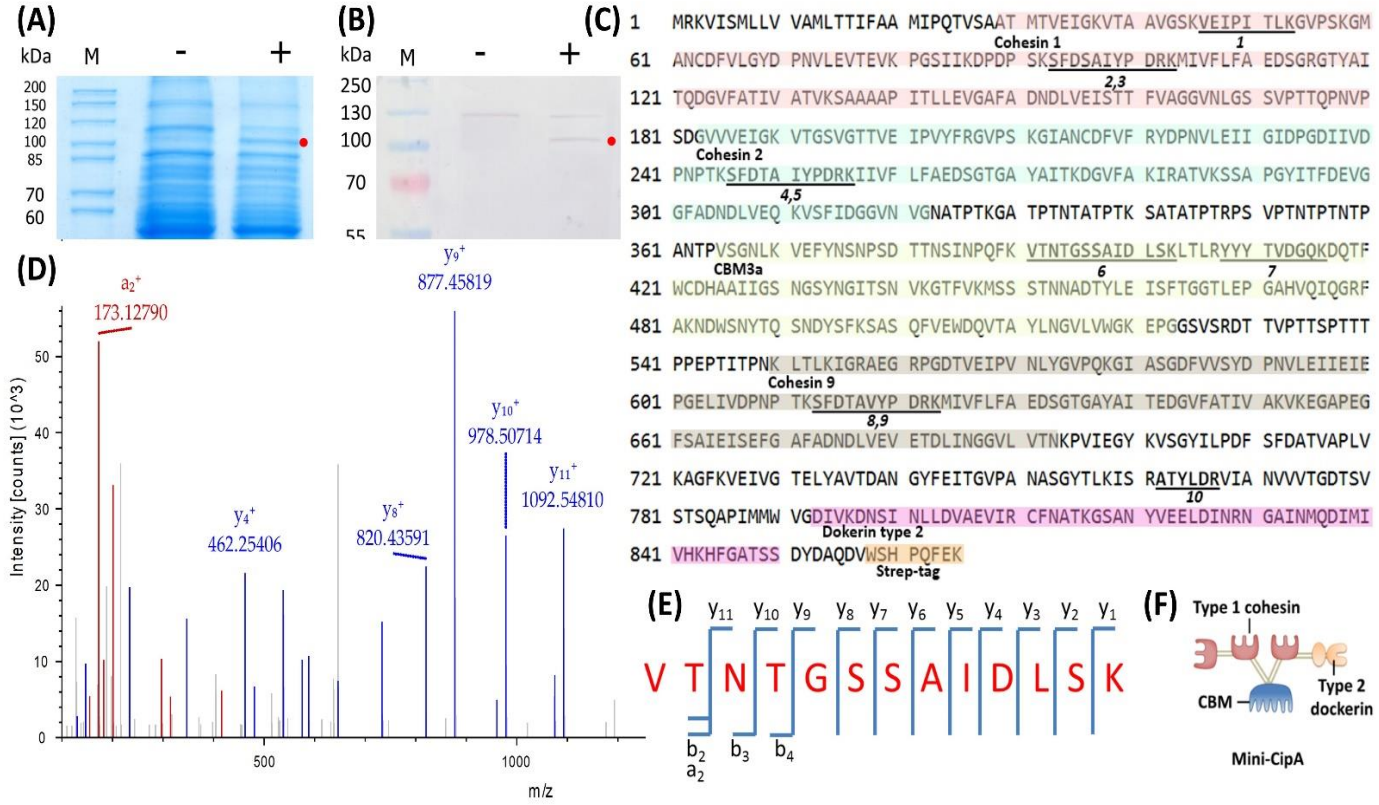


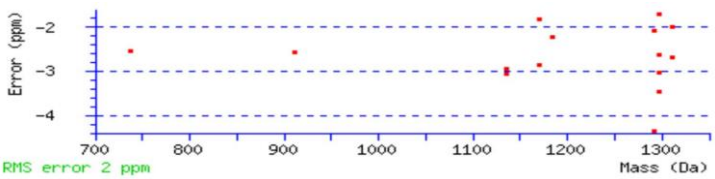
Fig. 3: Analysis of the extracellular mini-CipA from *B. subtilis* WB800N culture media.
(A) SDS-PAGE of culture media from B. subtilis WB800N expressing mini-CipA. (B) Western blot with anti-Strep tag antibody revealed the bands of mini-CipA (red dots). (+/-) culture media with or without the addition of IPTG. (C) Ten peptides (underline) with Mascot score above 25 were detected by HPLC-MS/MS in the mini-CipA modules (colored sequences). (D) The mass spectra of the observed fragment ions (colored) were assigned to peptide 6 (VTNTGSSAIDLSK). (E) The sequence of the peptide 6 and its detected fragment ions by HPLC-MS/MS. (F) Mini-CipA contains modules comprising three type-1 cohesins, a CBM and a type-2 dockerin.

Table 3
Mascot result

Protein View: pHT1701/Mini-CipA
protein mini-CipA
OS=Clostridium thermocellum (strain ATCC 27405 / DSM 1237 / NBRC 103400 / NCIMB 10682 / NRRL B-4536 / VPI 7372) GN=cipA PE=1 SV=2
Database: Bacillus subtilis 168 + pHT1701/Mini-CipA
Score: 418
Nominal mass (Mr): 92120
Calculated pI: 4.57
Sequence similarity is available as an NCBI BLAST search of pHT1701 against nr.
Search parameters
Enzyme: Trypsin/P: cuts C-term side of KR.
Variable modifications: Oxidation (M), Carbamidomethyl (C)
Protein sequence coverage: 6%
Matched peptides shown in underline.

1 MRKVISMLLV VAMLTTIFAA MIPQTVSAAT MTVEIGKVTA AVGSK**VEIPI TLK**GVPSKGM
61 ANCDFVLGYD PNVLEVTEVK PGSIIKDPDP SK**SFDSAIYP DRK**MIVFLFA EDSGRGTYAI
121 TQDGVFATIV ATVKSAAAAP ITLLEVGAFA DNDLVEISTT FVAGGVNLGS SVPTTQPNVP
181 SDGVVVEIGK VTGSVGTTVE IPVYFRGVPS KGIANCDFVF RYDPNVLEII GIDPGDIIVD
241 PNPTK**SFDTA IYPDRK**IIVF LFAEDSGTGA YAITKDGVFA KIRATVKSSA PGYITFDEVG
301 GFADNDLVEQ KVSFIDGGVN VGNATPTKGA TPTNTATPTK SATATPTRPS VPTNTPTNTP
361 ANTPVSGNLK VEFYNSNPSD TTNSINPQFK **VTNTGSSAID LSK**LTLR**YYY TVDGQK**DQTF
421 WCDHAAIIGS NGSYNGITSN VKGTFVKMSS STNNADTYLE ISFTGGTLEP GAHVQIQGRF
481 AKNDWSNYTQ SNDYSFKSAS QFVEWDQVTA YLNGVLVWGK EPGGSVSRDT TVPTTSPTTT
541 PPEPTITPNK LTLKIGRAEG RPGDTVEIPV NLYGVPQKGI ASGDFVVSYD PNVLEIIEIE
601 PGELIVDPNP TK**SFDTAVYP DRK**MIVFLFA EDSGTGAYAI TEDGVFATIV AKVKEGAPEG
661 FSAIEISEFG AFADNDLVEV ETDLINGGVL VTNKPVIEGY KVSGYILPDF SFDATVAPLV
721 KAGFKVEIVG TELYAVTDAN GYFEITGVPA NASGYTLKIS R**ATYLDR**VIA NVVVTGDTSV
781 STSQAPIMMW VGDIVKDNSI NLLDVAEVIR CFNATKGSAN YVEELDINRN GAINMQDIMI
841 VHKHFGATSS DYDAQDVWSH PQFEK
Unformatted sequence string: 865 residues.

Query	Start	– End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Rank
7509	46	– 53	456.7907	911.5668	911.5692	-2.57	0	20	0.0093	1
8839	93	– 102	585.7732	1169.5319	1169.5353	-2.86	0	69	1.1e-007	1
8840	93	– 102	585.7739	1169.5332	1169.5353	-1.82	0	32	0.00092	2
9527	93	– 103	433.5492	1297.6257	1297.6302	-3.47	1	1	1.2	2
9528	93	– 103	649.8204	1297.6263	1297.6302	-3.04	1	1	1	2
9529	93	– 103	433.5496	1297.6269	1297.6302	-2.62	1	20	0.015	1
9530	93	– 103	649.8213	1297.6280	1297.6302	-1.71	1	51	1.6e-005	1
8916	246	– 255	592.7814	1183.5483	1183.5509	-2.23	0	46	2.5e-005	1
9604	246	– 256	438.2214	1311.6424	1311.6459	-2.69	1	29	0.0011	1
9605	246	– 256	656.8289	1311.6433	1311.6459	-2.00	1	57	2e-006	1
9488	391	– 403	646.8355	1291.6563	1291.6620	-4.34	0	87	1.9e-009	1
9489	391	– 403	646.8369	1291.6593	1291.6620	-2.08	0	56	2.6e-006	1
8649	408	– 416	568.7648	1135.5151	1135.5186	-3.06	0	49	1.3e-005	1
8650	408	– 416	568.7649	1135.5152	1135.5186	-2.96	0	49	1.2e-005	1
8840	624	– 633	585.7739	1169.5332	1169.5353	-1.83	0	71	8e-008	1
9527	624	– 634	433.5492	1297.6257	1297.6303	-3.48	1	30	0.0013	1
9528	624	– 634	649.8204	1297.6263	1297.6303	-3.05	1	33	0.00066	1
6441	763	– 768	369.6917	737.3689	737.3708	-2.52	0	10	0.11	1



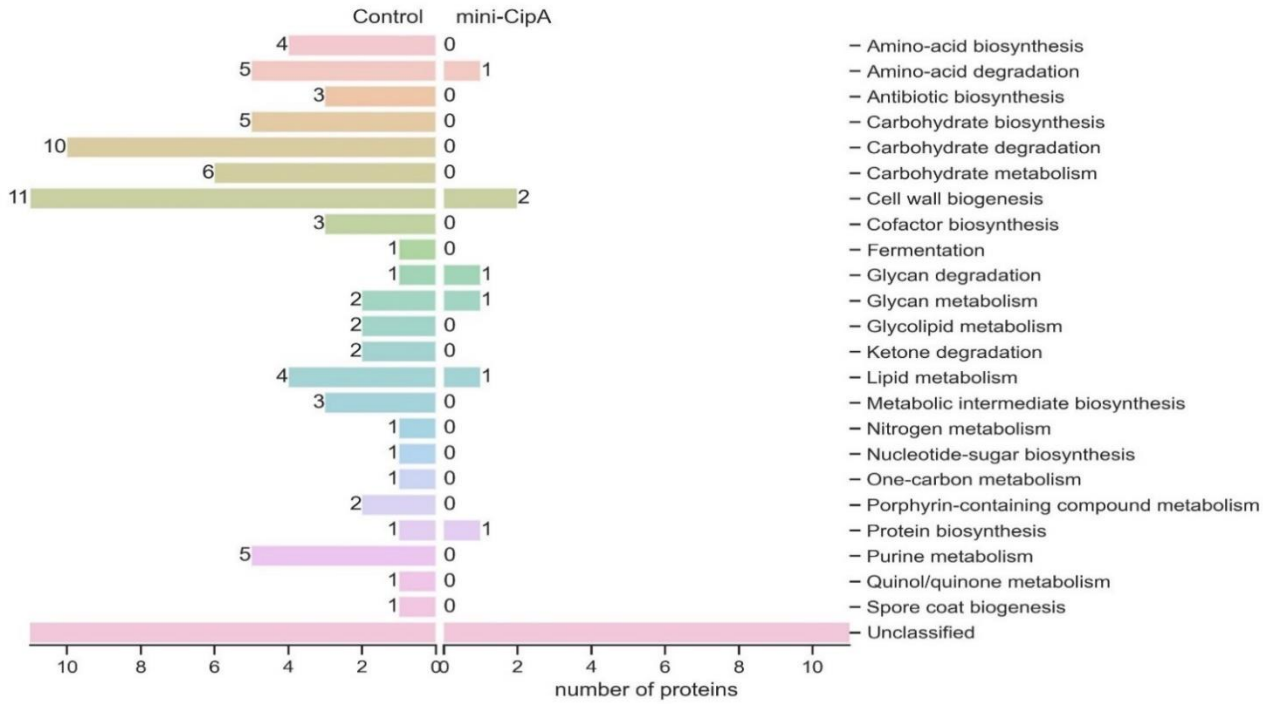


Fig. 4: Metabolic pathways of detected proteins in *B. subtilis* WB800N extracellularly expressing mini-CipA.
In the control sample of *B. subtilis* WB800N, 315 proteins of unclassified pathways were among the 390 detected proteins. In the secretome of *B. subtilis* WB800N expressing mini-CipA, 113 proteins of unclassified pathways were among the 120 detected proteins.

Table 4
The emPAI of detected proteins from *B. subtilis* WB800N in the control and mini-CipA samples in the range 60 – 150 kDa.

S.N.	Protein Accession	Protein names	Mass (Da)	Pathway	mini-CipA	Control
1	O31875	Ribonucleoside-diphosphate reductase NrdEB subunit alpha (EC 1.17.4.1) (Ribonucleotide reductase large subunit) [Cleaved into: Bsu nrdEB intein]	124621	-	0.02	0.02
2	C0SP86	DNA translocase SftA (Septum-associated FtsK-like translocase of DNA)	106823	-	0.03	-
3	Q45477	Isoleucine--tRNA ligase (EC 6.1.1.5) (Isoleucyl-tRNA synthetase) (IleRS)	104845	-	0.03	0.09
4	P54423	Cell wall-associated protease (EC 3.4.21.-) [Cleaved into: Cell wall-associated polypeptide CWBP23 (CWBP23); Cell wall-associated polypeptide CWBP52 (CWBP52)]	96488	-	0.2	0.44
5	P29141	Minor extracellular protease Vpr (EC 3.4.21.-)	85608	-	0.07	0.07
6	P68579	SPbeta prophage-derived sublancin-168-processing and transport ATP-binding protein SunT (EC 3.4.22.-) (EC 7.-.-.-)	81564	-	0.04	0.04
7	Q796K8	Penicillin-binding protein H (EC 3.4.16.4)	79348	Cell wall biogenesis	0.04	0.16
8	O34952	Lipoteichoic acid synthase 2 [Cleaved into: Glycerol phosphate lipoteichoic acid synthase 2 (LTA synthase 1) (EC 2.7.8.-) (Polyglycerol phosphate synthase 2); Processed glycerol phosphate lipoteichoic acid synthase 2]	74234	Cell wall biogenesis	0.04	0.32
9	P00691	Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan glucanohydrolase)	72378	-	0.04	0.18
10	O31980	SPbeta prophage-derived uncharacterized protein YomE	71098	-	0.04	-
11	P42061	Oligopeptide-binding protein AppA	61917	-	0.1	-
12	P24141	Oligopeptide-binding protein OppA (Stage 0 sporulation protein KA)	61525	-	0.15	0.1

-: Not detected or Unclassified

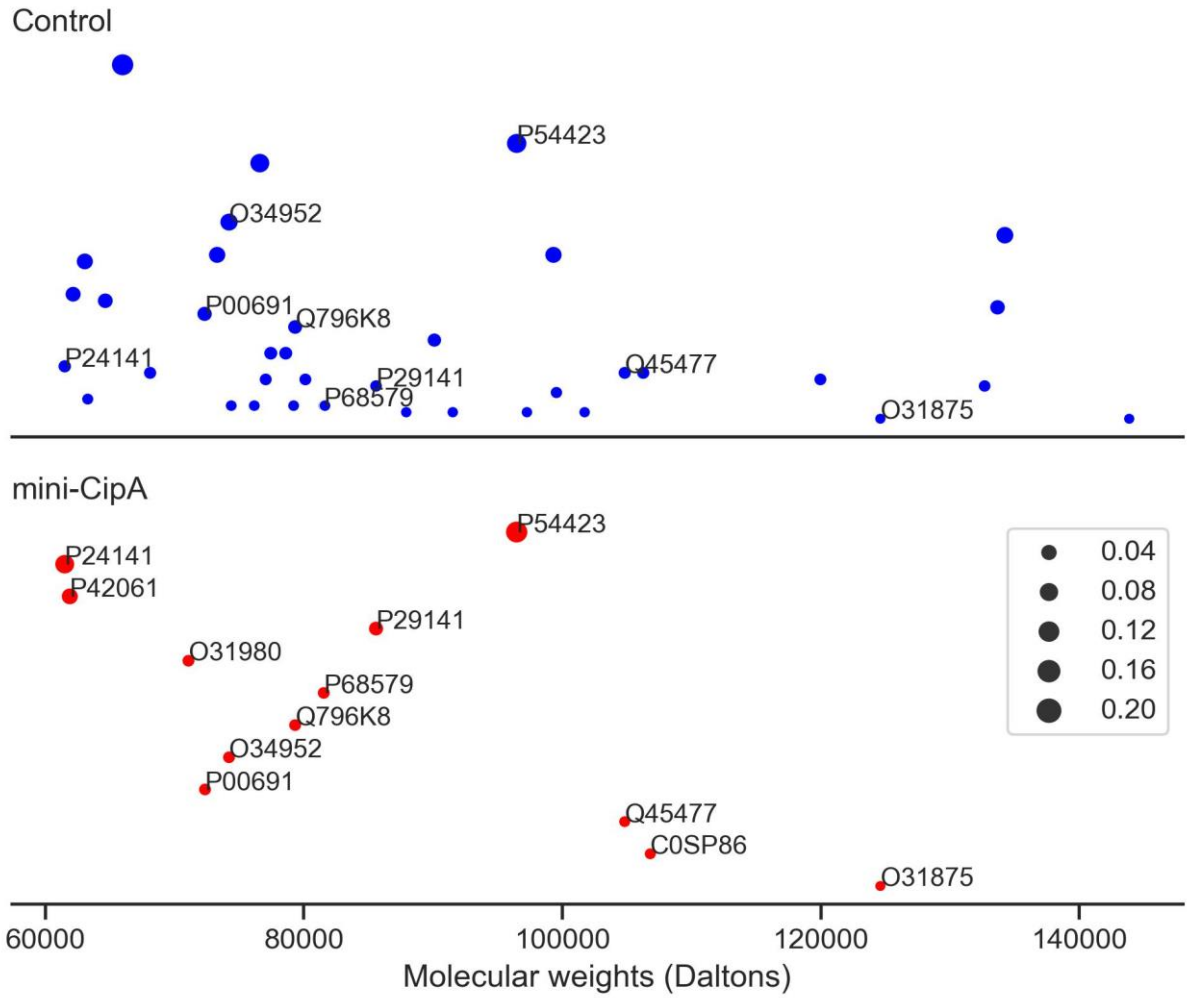


Fig. 5: EmPAI values of 60 to 150 kDa proteins in the control and mini-CipA samples from *B. subtilis* WB800N
P24141: Oligopeptide-binding protein OppA (61525 Da), P42061: Oligopeptide-binding protein AppA (61917 Da), O31980: SPbeta prophage-derived uncharacterized protein YomE (71098 Da), P00691: Alpha-amylase (72378 Da), O34952: Lipoteichoic acid synthase 2 (74234 Da), Q796K8: Penicillin-binding protein H (79348 Da), P68579: SPbeta prophage-derived sublancin-168-processing and transport ATP-binding protein SunT (81564 Da), P29141: Minor extracellular protease Vpr (85608 Da), P54423: Cell wall-associated protease (96488 Da), Q45477: Isoleucine--tRNA ligase (104845 Da), C0SP86: DNA translocase SftA (106823 Da), O31875: Ribonucleoside-diphosphate reductase NrdEB subunit alpha (124621 Da). Unlabeled dots: proteins detected only in control sample.

However, the designer mini-CipA in this study retained the type-2 dockerin for binding type-2 cohesins of cell surface proteins for future research. While qualitative results have been provided so far, the quantitative analysis of mini-CipA in culture media remains unclear. Quantitative methods such as Western blot[27] and HPLC-MS/MS[36] can be utilized, but determining which secreted house-keeping proteins in *B. subtilis* can serve as internal standards for quantitative Western blot, requires further investigation. Although the emPAI value can provide an absolute quantification of proteins[14], it is better used for horizontal comparison of target proteins across multiple samples because variations in the number of observable peptides for different proteins can impact emPAI value calculation. Futhermore, the bands around 130 kDa in SDS-PAGE and Western blot gels have not been identified, suggesting that tryptic digestion of in-gel proteins could be performed for more precise detection through mass spectrometry[12].

Compared to intracellular expression, extracellular expression provides many advantages such as reducing protein aggregation and simplifying downstream purification. The expression of foreign proteins requires host cell resources and leads to the decreased growth rate and viability of the host cell due to the metabolic burden[3,29]. This might explain why the number of detected proteins from the host cell in the mini-CipA sample was approximately 30% of the control sample. An estimated five billion metric tons of agricultural residues are produced annually worldwide[30]. A significant portion of crop residues is burnt particularly in Asian countries where the economic importance of crop residue management is often underestimated[19,28]. These unsustainable practices contribute to air pollution and hinder the recycling of soil nutrients. Designing cellulosomes presents an innovative approach to adding more value to agricultural by-products. Consolidated bioprocessing utilizing cellulosomes facilitates the degradation of

lignocellulose into short chains of saccharides and fatty acids, which are used as nutrients for ruminants and livestock or fermented into biofuels. In the gut microbiota of ancient humans, hunter-gatherers and rural populations, *Ruminococcus* producing cellulosomes were more prevalent compared to industrialized societies[25].

The significant connections between gut microbiota and human health suggest that intervening in microbial function could be a potent therapeutic strategy[11]. *B. subtilis*, a species generally recognized as safe by the United States Food and Drug Administration, has been applied for production of bio-pharmaceutical proteins such as antibodies, hormones and growth factors[20]. Some *B. subtilis* strains isolated from the human gastrointestinal tract could be health supporting probiotics[10,35]. Other safe host-cell probiotics such as *Lactobacillus plantarum* and *Kluyveromyces marxianus* have been employed to express designer cellulosomes[1,31]. Further exploration of designer cellulosomes for advanced probiotics in disease prevention and treatment by modulating gut microbiomes requires more extensive research.

Conclusion

Scaffoldin CipA is the primary structural component of cellulosomes in *Clostridium thermocellum* which assembles cellulosomal multienzymes and attaches the protein complex to cellulose. In this study, a designer extracellular mini-CipA which was created in *B. subtilis* WB800N consists of three type-1 cohesins, a Carbohydrate Binding Module, and a type-2 dockerin.

Qualitative analyses using SDS-PAGE, Western blot and HPLC-MS/MS confirmed that *B. subtilis* WB800N secreted the designer mini-CipA. The secretome of the *B. subtilis* WB800N expressing the mini-CipA contained approximately 30% of the number of proteins found in the control sample.

Acknowledgement

The research was sponsored by the International Foundation for Science, Sweden (Grant No. F/5540-1) for Nguyen H.N.P. Nguyen acknowledges the Toshiba Fellowships and Merck Vietnam award. The authors thank Prof. W. Schumann at the University of Bayreuth, Germany for sharing biomaterials. Additionally, the authors express gratitude to Prof. Joseph Loo, Dr. Rachel Loo and Dr. Hong Hanh Nguyen at the Center for Global Mentoring, Molecular Biology Institute, University of California, Los Angeles, USA for their assistance in mass spectrometry analysis. The resulting biomaterials were provided to the Center for Bioscience and Biotechnology, VNUHCM-US.

References

1. Anandharaj M., Lin Y.J., Rani R.P., Nadendla E.K., Ho M.C., Huang C.C., Cheng J.F., Chang J.J. and Li W.H., Constructing a yeast to express the largest cellulosome complex on the cell surface, Proceedings of the National Academy of Sciences, **117(5)**, 2385–2394 **(2020)**

2. Arfi Y., Shamshoum M., Rogachev I., Peleg Y. and Bayer E.A., Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation, Proceedings of the National Academy of Sciences, **111(25)**, 9109–9114 **(2014)**

3. Birnbaum S. and Bailey J.E., Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*, *Biotechnology and Bioengineering*, **37(8)**, 736–745 **(1991)**

4. Chachaty E. and Saulnier P., Isolating chromosomal DNA from bacteria, Totowa, NJ, Humana Press, 29–32 **(2000)**

5. Chang J.J., Anandharaj M., Ho C.Y., Tsuge K., Tsai T.Y., Ke H.M., Lin Y.J., Ha Tran M.D., Li W.H. and Huang C.C., Biomimetic strategy for constructing Clostridium thermocellum cellulosomal operons in Bacillus subtilis, *Biotechnology for Biofuels*, **11(1)**, 157 **(2018)**

6. Fan L.H., Zhang Z.J., Yu X.Y., Xue Y.X. and Tan T.W., Self-surface assembly of cellulosomes with two miniscaffoldins on Saccharomyces cerevisiae for cellulosic ethanol production, Proceedings of the National Academy of Sciences of the United States of America, **109(33)**, 13260–13265 **(2012)**

7. Fierobe H.P., Mechaly A., Tardif C., Belaich A., Lamed R., Shoham Y., Belaich J.P. and Bayer E.A., Design and production of active cellulosome chimeras: selective incorporation of dockerin-containing enzymes into defined functional complexes, *Journal of Biological Chemistry*, **276(24)**, 21257–21261 **(2001)**

8. Fontes C.M.G.A. and Gilbert H.J., Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates, *Annual Review of Biochemistry*, **79**, 655–681 **(2010)**

9. Hamann P.R.V., de M.B. Silva L., Gomes T.C. and Noronha E.F., Assembling mini-xylanosomes with Clostridium thermocellum XynA and their properties in lignocellulose deconstruction, *Enzyme and Microbial Technology*, **150**, 109887 **(2021)**

10. Hong H.A., Khaneja R., Tam N.M.K., Cazzato A., Tan S., Urdaci M., Brisson A., Gasbarrini A., Barnes I. and Cutting S.M., *Bacillus subtilis* isolated from the human gastrointestinal tract, *Research in Microbiology*, **160(2)**, 134–143 **(2009)**

11. Hsu B.B., Plant I.N., Lyon L., Anastassacos F.M., Way J.C. and Silver P.A., *In situ* reprogramming of gut bacteria by oral delivery, *Nature Communications*, **11(1)**, 5030 **(2020)**

12. Huynh M.L., Russell P. and Walsh B., Tryptic digestion of in-gel proteins for mass spectrometry analysis, *Methods in Molecular Biology*, **519**, 507–513 **(2009)**

13. Hyeon J.E., Kang D.H., Kim Y.I., Jeon S.D., You S.K., Kim K.Y., Kim S.W. and Han S.O., Production of functional agarolytic nano-complex for the synergistic hydrolysis of marine biomass and its potential application in carbohydrate-binding module-utilizing

one-step purification, *Process Biochemistry*, **47(5)**, 877–881 **(2012)**

14. Ishihama Y., Oda Y., Tabata T., Sato T., Nagasu T., Rappsilber J. and Mann M., Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein, *Molecular & Cellular Proteomics*, **4(9)**, 1265–1272 **(2005)**

15. Jeong H., Jeong D.E., Park S.H., Kim S.J. and Choi S.K., Complete genome sequence of Bacillus subtilis strain WB800N, an extracellular protease-deficient derivative of strain 168, *Microbiology Resource Announcements*, **7(18)**, e01380-18 **(2018)**

16. Joliff G., Edelman A., Klier A. and Rapoport G., Inducible secretion of a cellulase from Clostridium thermocellum in *Bacillus subtilis*, *Applied and Environmental Microbiology*, **55(11)**, 2739–44 **(1989)**

17. Kataeva I.A., Seidel R.D., Li X.L. and Ljungdahl L.G., Properties and mutation analysis of the CelK cellulose-binding domain from the Clostridium thermocellum cellulosome, *Journal of Bacteriology*, **183(5)**, 1552–1559 **(2001)**

18. Kleiner-Grote G.R.M., Risse J.M. and Friehs K., Secretion of recombinant proteins from *E. coli*, *Engineering in Life Sciences*, **18(8)**, 532–550 **(2018)**

19. Krishna V.V. and Mkondiwa M., Economics of crop residue management, *Annual Review of Resource Economics*, **15**, 19–39 **(2023)**

20. Lakowitz A., Godard T., Biedendieck R. and Krull R., Mini review: Recombinant production of tailored bio-pharmaceuticals in different *Bacillus* strains and future perspectives, *European Journal of Pharmaceutics and Biopharmaceutics*, **126**, 27–39 **(2018)**

21. Lamed R. and Bayer E.A., The cellulosome of Clostridium thermocellum, Academic Press, **33**, 1–46 **(1988)**

22. Lee H.J. et al, Functional expression and extracellular secretion of Clostridium thermocellum Cel48S cellulase in Escherichia coli via the signal recognition particle-dependent translocation pathway, *Enzyme and Microbial Technology*, **151**, 109–918 **(2021)**

23. Li X., Xiao Y., Feng Y., Li B., Li W. and Cui Q., The spatial proximity effect of beta-glucosidase and cellulosomes on cellulose degradation, *Enzyme and Microbial Technology*, **115**, 52–61 **(2018)**

24. Meraj A. and Jawaid M., 1 - Plant biomass-based materials: an overview, Academic Press, 1–24 **(2024)**

25. Moraïs S., Winkler S., Zorea A., Levin L., Nagies F.S.P., Kapust N., Lamed E., Artan-Furman A., Bolam D.N., Yadav M.P., Bayer E.A., Martin W.F. and Mizrahi I., Cryptic diversity of cellulose-degrading gut bacteria in industrialized humans, *Science*, **383(6688)**, eadj9223 **(2024)**

26. Olson D.G., Giannone R.J., Hettich R.L. and Lynd L.R., Role of the CipA scaffoldin protein in cellulose solubilization, as determined by targeted gene deletion and complementation in Clostridium thermocellum, *Journal of Bacteriology*, **195(4)**, 733–739 **(2013)**

27. Pillai-Kastoori L., Schutz-Geschwender A.R. and Harford J.A., A systematic approach to quantitative Western blot analysis, *Analytical Biochemistry*, **593**, 113608 **(2020)**

28. Ravindra K., Singh T. and Mor S., Emissions of air pollutants from primary crop residue burning in India and their mitigation strategies for cleaner emissions, *Journal of Cleaner Production*, **208**, 261–273 **(2019)**

29. Rouches M.V., Xu Y., Cortes L.B.G. and Lambert G., A plasmid system with tunable copy number, *Nature Communications*, **13(1)**, 3908 **(2022)**

30. Shinde R., Shahi D.K., Mahapatra P., Singh C.S., Naik S.K., Thombare N. and Singh A.K., Management of crop residues with special reference to the on-farm utilization methods: A review, *Industrial Crops and Products*, **181**, 114772 **(2022)**

31. Stern J., Moraïs S., Ben-David Y., Salama R., Shamshoum M., Lamed R., Shoham Y., Bayer E.A. and Mizrahi I., Assembly of synthetic functional cellulosomal structures onto the cell surface of Lactobacillus plantarum, a potent member of the gut microbiome, *Applied and Environmental Microbiology*, **84(8)**, e00282-18 **(2018)**

32. Tan Y.H., Chin B.L.F., Khor H.Q., Mubarak N.M., Khalid M. and Yek P.N.Y., 2 - Origin, preparation and processing of plant biomass materials, Academic Press, 25–47 **(2024)**

33. Tokatlidis K., Dhurjati P., Millet J., Béguin P. and Aubert J.P., High activity of inclusion bodies formed in Escherichia coli overproducing Clostridium thermocellum endoglucanase D, *FEBS Letters*, **282(1)**, 205–208 **(1991)**

34. Wieczorek A.S. and Martin V.J., Engineering the cell surface display of cohesins for assembly of cellulosome-inspired enzyme complexes on Lactococcus lactis, *Microbial Cell Factories*, **9(1)**, 69 **(2010)**

35. Williams N. and Weir T.L., Spore-Based Probiotic Bacillus subtilis: Current Applications in Humans and Future Perspectives, *Fermentation*, **10(2)**, 78 **(2024)**

36. Wong J.W.H., Sullivan M.J. and Cagney G., Computational methods for the comparative quantification of proteins in label-free LCn-MS experiments, *Briefings in Bioinformatics*, **9(2)**, 156–165 **(2008)**

37. Yi Z.L. and Wu Z.L., Mutations from a family-shuffling-library reveal amino acid residues responsible for the thermostability of endoglucanase CelA from Clostridium thermocellum, *Biotechnology Letters*, **32(12)**, 1869–1875 **(2010)**

38. You C., Zhang X.Z. and Zhang Y.H.P., Mini-scaffoldin enhanced mini-cellulosome hydrolysis performance on low-accessibility cellulose (Avicel) more than on high-accessibility amorphous cellulose, *Biochemical Engineering Journal*, **63**, 57–65 **(2012)**.

(Received 09th December 2024, accepted 11th January 2025)