ACS Biomaterials SCIENCE & ENGINEERING

pubs.acs.org/journal/abseba

Review

Current Insights on the Diverse Structures and Functions in Bacterial Collagen-like Proteins

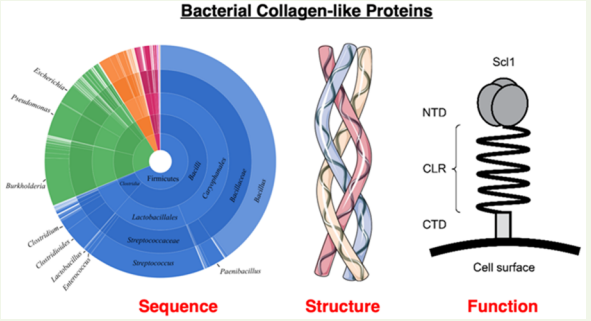
Yimin Qiu, Chenxi Zhai, Ling Chen, Xiaoyan Liu, and Jingjie Yeo*

Cite This: https://doi.org/10.1021/acsbiomaterials.1c00018

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The dearth of knowledge on the diverse structures and functions in bacterial collagen-like proteins is in stark contrast to the deep grasp of structures and functions in mammalian collagen, the ubiquitous triple-helical scleroprotein that plays a central role in tissue architecture, extracellular matrix organization, and signal transduction. To fill and highlight existing gaps due to the general paucity of data on bacterial CLPs, we comprehensively reviewed the latest insight into their functional and structural diversity from multiple perspectives of biology, computational simulations, and materials engineering. The origins and discovery of bacterial CLPs were explored. Their genetic distribution and molecular architecture were analyzed, and their structural and functional diversity in various bacterial genera was examined. The principal roles of computational techniques in understanding bacterial CLPs' structural stability, mechanical properties, and biological functions were also considered. This review serves to drive further interest and development of bacterial CLPs, not only for addressing fundamental biological problems in collagen but also for engineering novel biomaterials. Hence, both biology and materials communities will greatly benefit from intensified research into the diverse structures and functions in bacterial collagen-like proteins.



KEYWORDS: *collagen, genetic distribution, molecular diversity, structural diversity, functional diversity, modeling and simulation*

1. OVERVIEW OF ANIMAL-DERIVED COLLAGEN

Collagen is ubiquitously present in metazoans from the first multicellular animal (sponges) to humans[1] and plays a central role in tissue architecture, extracellular matrix organization, and signal transduction.[2] As the most abundant protein in vertebrates, collagen appears as supramolecular assemblies with diverse functions in different tissues, such as skin, cornea, tendon, cartilage, and bone.[3,4] A structural feature that is common to all collagen, the unique triple helical structure, is composed of three left-handed polyproline-II-like chains that are supercoiled into a right-handed triple helix.[5] The close packing of the polypeptide helices near the central axis is made possible by having repeated glycine amino acids as every third residue, leading to a characteristic repeating pattern $(Gly\text{-}Xaa\text{-}Yaa)_n$, where the Xaa and Yaa positions are often respectively occupied by proline and hydroxyproline in animal collagen.[6] Hydroxyproline (Hyp) is derived from the post-translational modification of proline by prolyl-4-hydroxylase (P4H).[7] The Hyp residue contributes greatly to stabilizing the triple helix through stereoelectronic effects,[8] as well as being essential for collagen self-association[9] and for interacting with cell receptors and molecular ligands.[10] For detailed discussions on collagen biosynthesis or collagen fibril assembly, we direct the interested reader to some excellent reviews.[6,11]

While the structure and biophysical properties of metazoan collagen were explored extensively in current scientific literature, discoveries in bacterial collagen are a burgeoning field for intensive research. Since the initial studies,[12,13] proteins containing large regions of Gly-Xaa-Yaa (GXY) repeats were uncovered in numerous bacterial genome databases. However, only a small fraction of the predicted bacterial collagen-like proteins (CLPs) was investigated so far, of which eight such CLPs were recombinantly expressed *in vitro* and confirmed to fold into the characteristic collagen triple helix even though bacterial CLPs lack the Hyp residue (see Yu et al. in 2014,[14] Brodsky and Ramshaw in 2017[15] for comprehensive reviews): Scl1 and Scl2 from *Streptococcus pyogenes* (*S. pyogenes*),[16] BclA from *Bacillus anthracis*,[17] Lcl from *Legionella pneumophila*,[18] and four unnamed proteins from *Clostridium perfringens*, *Solibacter usitatus*, *Rhodopseudomonas palustris* (*R. palustris*) and *Methylobacterium* sp. 4−46.[19] Hence, to fill and highlight existing gaps due to the general

Special Issue: Engineering Bioinspired and Biological Materials

Received: January 6, 2021
Accepted: April 7, 2021

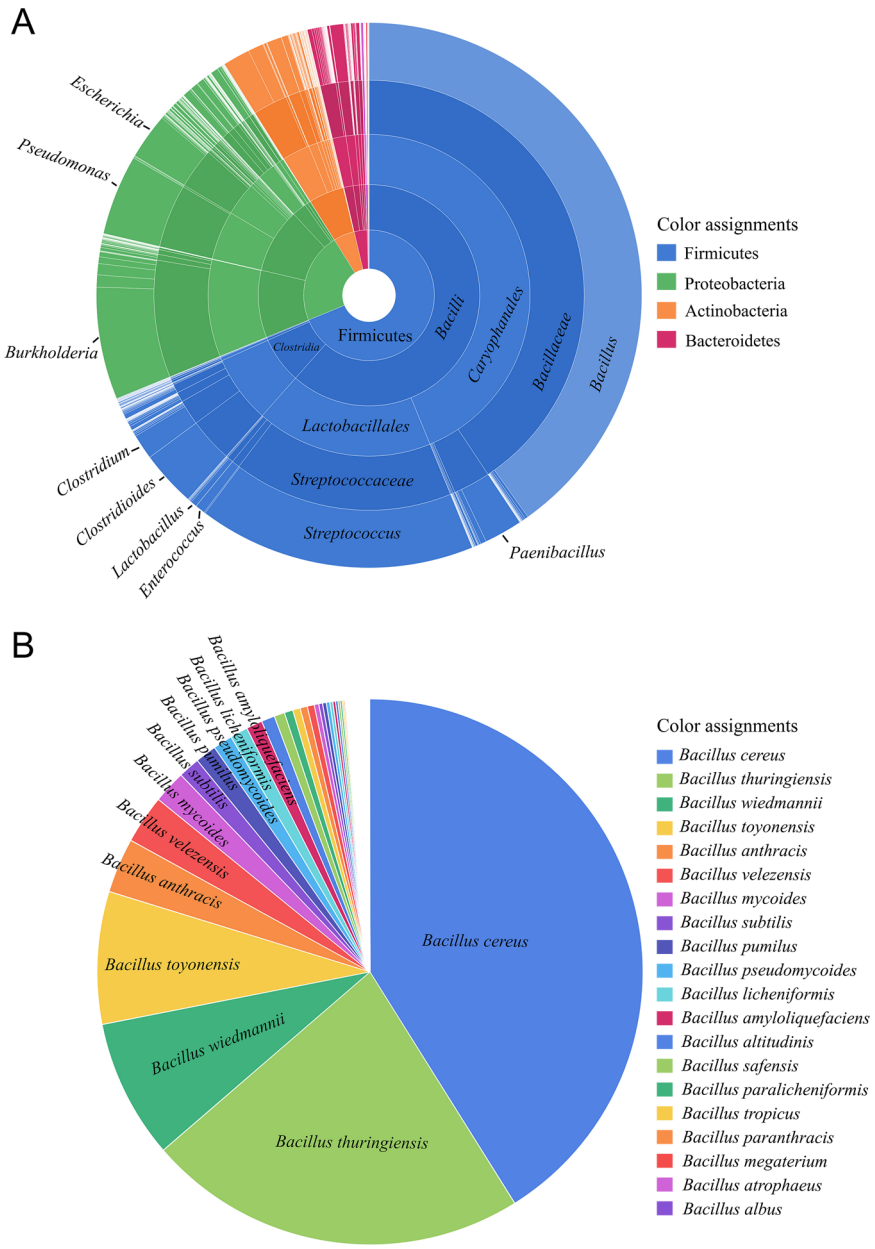


Figure 1. Distribution of the CLPs in the kingdom Bacteria (A) and the phylum Firmicutes (B). The survey involved 60,789 amino acid sequences of bacterial proteins annotated as collagen-like proteins. Data were retrieved from NCBI database (https://www.ncbi.nlm.nih.gov/protein) as of July 12, 2020.

paucity of data on bacterial CLPs, we aim to present the most recent data regarding bacterial collagen by providing a comprehensive overview of the latest insight into their functional and structural diversity from multiple perspectives of biology, computational simulations, and materials engineering. We strongly anticipate that this review will drive further interest and development of bacterial CLPs, not only for addressing fundamental biological problems in collagen but also for engineering novel biomaterials.

2. GENETIC DISTRIBUTION AND MOLECULAR DIVERSITY OF BACTERIAL COLLAGEN-LIKE PROTEINS

2.1. Discovery of Bacterial CLPs. The first clues to bacterial proteins that possessed collagen-like domains came about when repeating amino acid units of $(Gly\text{-}Xaa\text{-}Pro)_6$ were identified within the pullulanase from *Klebsiella pneumoniae* in 1988.[20] In 2000, the *S. pyogenes* proteins Scl1 and Scl2 were found to contain a substantial length of repeated GXY motifs and were subsequently demonstrated to form collagen-like triple helices.[16,21−23] Since then, a wide range of nonanimal, bacterial collagen sequences were identified, driven by the rapidly increasing availability of new genetic technologies and genomic information. Different collagen-like proteins were described in bacteria, such as streptococcal Scl proteins of *S. pyogenes*,[24] Bcl proteins of *B. anthracis*,[25,26] pneumococcal proteins Pcl,[27] Lcl of *L. pneumophila*,[28] Bucl proteins of *Burkholderia* spp.,[29] and Clps of *Bacillus amyloliquefaciens*.[30] The growing recognition of bacterial collagen-like proteins no longer arouses unexpected curiosities. Instead, this recognition presents an opportunity to define the biological properties of

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

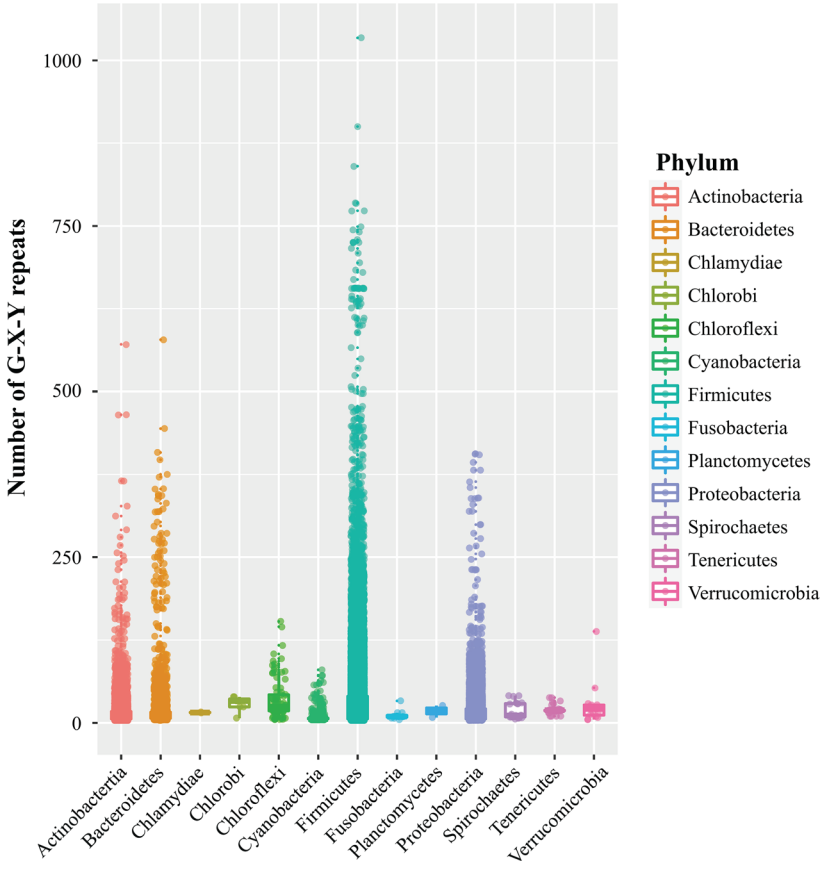


Figure 2. Numbers of the GXY triplets within the collagen-like region in the bacterial CLPs.

bacterial CLPs and provide deeper insight for engineering CLPs that cater to biomaterial applications.

2.2. Origin of Bacterial CLPs. Two hypotheses on the origins of collagen-like sequences (CLSs) in prokaryotes were proposed previously: (1) Collagenous sequences may have been horizontally transferred from eukaryotes to bacteria.[12,31] (2) Collagen-like repeats may have arisen *de novo* in bacteria as a result of spontaneous mutations, duplication of simple repeat sequences, and domain organization.[32,33] This latter hypothesis entails having eukaryotic and prokaryotic collagen sequences emerging through convergent evolution.[34] Lateral gene transfer mediated by viruses or bacteriophages may have further promoted the spread of collagenous sequences between bacteria, resulting in a patchy distribution of GXY repeats among bacterial genomes.[35,36]

2.3. Genetic Distribution of Bacterial CLPs. For the purposes of our review, we retrieved 60,789 amino acid sequences of bacterial proteins annotated as known and predicted CLPs from the NCBI database (https://www.ncbi.nlm.nih.gov/protein) to analyze their distribution and characteristics. These sequences included 14 bacterial phyla (Actinobacteria, Bacteroidetes, Chlamydiae, Chlorobi, Chloroflexi, Cyanobacteria, Firmicutes, Fusobacteria, Planctomycetes, Proteobacteria, Spirochaetes, Tenericutes, and Verrucomicrobia). Our bioinformatics analysis showed that the vast majority (about 98.9%) of bacterial CLPs currently reported could be primarily found in 4 phyla: Firmicutes, Proteobacteria, Actinobacteria, and Bacteroides (Table S1). The proportions were 68.1%, 22.2%, 5.3%, and 3.2%, respectively (Figure 1A). Among Firmicutes, 23,959 CLPs were found in the genus *Bacillus* with a proportion of 39.4%. Hence, *Bacillus* was the genus that possessed the most reported CLPs in bacteria. In the genus of *Bacillus*, more CLPs were distributed in *Bacillus cereus* and *Bacillus thuringiensis* (Figure 1B). A similar occurrence of proteins containing CLSs was found in the genomes of bacteria belonging to the Firmicutes phylum among 165 extremophiles.[37] Not surprisingly, members of Firmicutes (*Bacillus*) were isolated from several million-year-old samples such as in amber, thereby highlighting the resilience and longevity of these members. A broad range of Firmicutes are able to form dormant spores that are highly resistant to environmental stresses. Moreover, spores of some bacteria, such as *Bacillus anthracis* and *Clostridium difficile*, play a crucial role in interacting with the host in a manner that favors their survival.[25] Together with the earlier discovery and important biological functions of CLPs in Firmicutes, we anticipate that a greater focus on mining the Firmicutes genetic sequences will garner further discoveries of novel CLPs.

2.4. Molecular Architecture of Bacterial CLPs. As highlighted above, the most typical feature in the molecular architecture of bacterial CLPs is the presence of GXY amino acid repeats. These regions of GXY repeats divide the CLP chains into different functional domains. Here, we mainly focused on the characterization of the bacterial CLPs containing ≥4 GXY repeats. First, the lengths of GXY repeats within collagenous domains in bacterial CLPs were analyzed (Figure 2). The lengths of continuous GXY repeats in CLPs varied greatly among different bacterial phyla. Among the four main bacterial phyla harboring CLPs, the maximum number of GXY repeats of CLPs in Firmicutes reached 1,034 (WP_119996260.1, CLP from *Bacillus subtilis*), while those in Proteobacteria, Actinobacteria, and Bacteroides were 406,

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

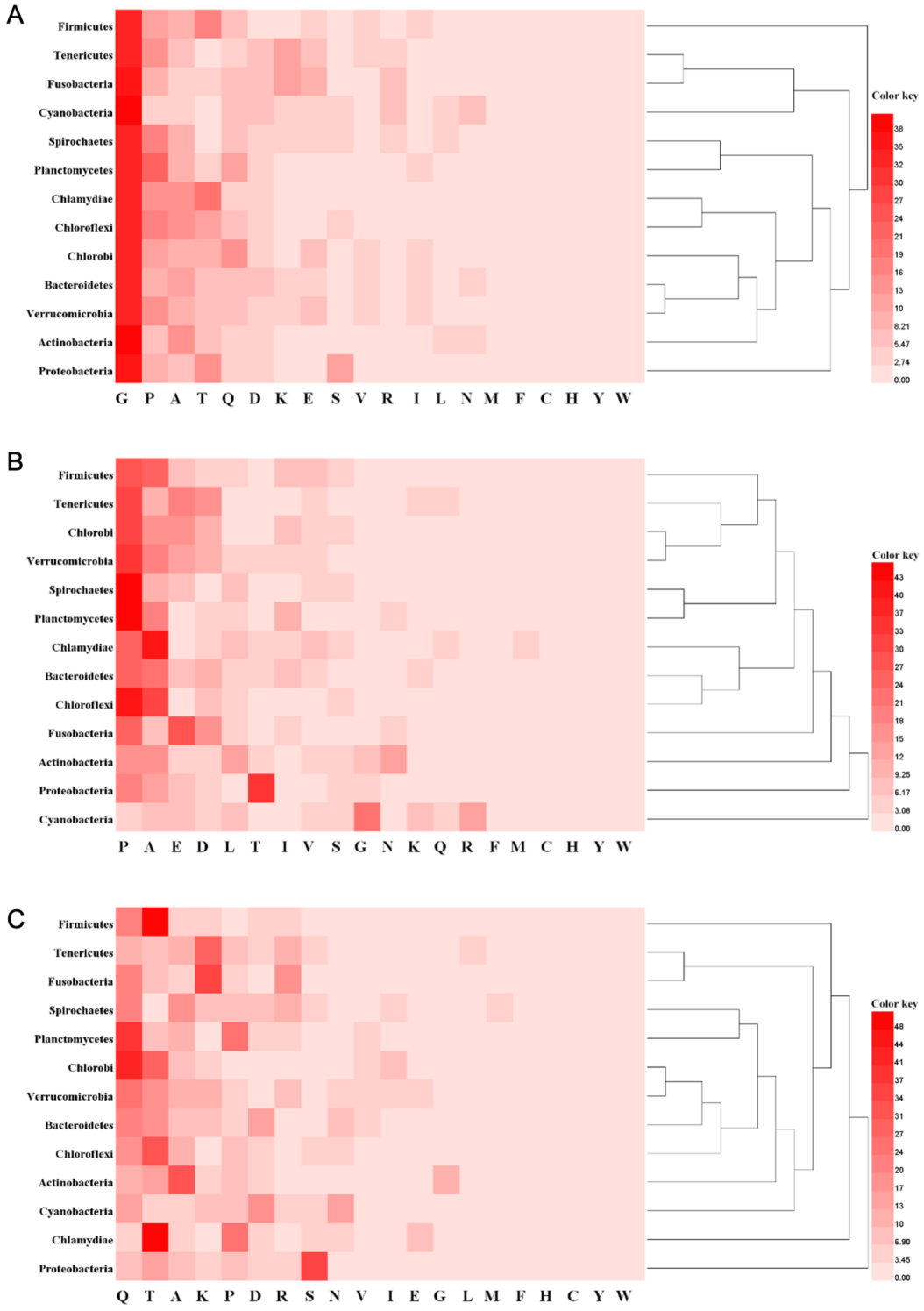


Figure 3. Heatmap reconstructed using the cluster analysis of GXY repeat regions of bacterial CLPs based on the percentage amino acid composition of the GXY triplets (A) and the amino acid residues at the Xaa position (B) and Yaa position (C) of GXY triplets. The color key shows the average percentage (%) of the individual amino acid residues in different bacterial phyla.

571, and 578, respectively. The tally of GXY repeats in Firmicutes CLPs appeared to vary more broadly compared with the numbers presented in other phyla. To date, 40 vertebrate collagen genes were identified, constituting 29 distinct homo- and/or heterotrimeric molecules.[7] Surveys of the minimal human collagenome revealed a total of 323 interruptions between G0G and G15G.[37,38] These data included 84 sequences with 12,234 triplets (36,702 residues in total) in contiguous $(GXY)_n$ stretches ($n \geq 3$) in the collagen domains. Fibrillar collagens I, II, III, V, and XI contained long, continuous collagenous domains of more than 1,000 amino acids, whereas type IV collagen chains contained 21−26 interruptions of different lengths in their triple helical domains of about 1,400 residues. In stark contrast, a total of 5,298 interruptions between G0G and G15G were identified in the collagen domains of the prokaryotic collagenome, which

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

Table 1. Biophysical Characteristics of the CLPs-Containing Bacterial Species and the Localization of CLPs Identified in Bacteria

genus	species	pathogenicity	spore-forming	proteins	CLPSs localization	ref
Streptococcus	*S. pyogenes*	pathogenic	−	SclA/Scl1, SclB/Scl2	cell surface protein	16
	S. pneumoniae	pathogenic	−	PclA	cell surface protein	27
	S. equi	pathogenic	−	SclC	cell surface protein	46
	S. equi	pathogenic	−	SclD to SclI	cell surface protein	154
Bacillus	*B. anthracis*	pathogenic	+	BclA	exosporium protein	17
	B. anthracis	pathogenic	+	BclB	exosporium protein	62
	B. cereus	pathogenic	+	ExsJ	exosporium protein	67
	B. cereus	pathogenic	+	BclA to BclE	exosporium protein	60
	B. thuringiensis	nonpathogenic	+	ExsJ	Exosporium protein	60
	B. amyloliquefaciens	nonpathogenic	+	ClpA, ClpD		30
	B. amyloliquefaciens	nonpathogenic	+	ClpB	localized on the flagella surface	72
	B. megaterium	nonpathogenic	+	BclA1, BclA2	putative exosporium proteins	59
Lysinibacillus	*L. sphaericus*	nonpathogenic	+	BclS	exosporium protein	56
Pasteuria	*P. ramosa*	pathogenic	+	Pcl1a	attached to the spore body	81
	P. ramosa	pathogenic	+	37 PCLs	in spore extracts	33
	P. ramosa	pathogenic	+	Pcl7	endospore components	83
	P. penetrans	pathogenic	+	Ppcl1 to Ppcl33		82
Clostridium	*C. perfringens*	pathogenic	+	unnamed		19
	C. difficile	pathogenic	+	BclA1, BclA2	in exosporium extracts	25
	C. difficile	pathogenic	+	BclA3	in exosporium extracts	85
	C. taeniosporum	nonpathogenic	+	GP85, CL2	spore appendage proteins	87
	C. sporogenes	nonpathogenic	+	BclA (partial), BclB	in exosporium extracts	90
Burkholderia	*B. pseudomallei, B. mallei*	pathogenic	−	13 Bucls		29
Escherichia	*E. coli* O157:*H7*	pathogenic	−	EPclA	included in *E. coli* prophages	36
Legionella	*L. pneumophila*	pathogenic	−	Lcl	putative outer membrane protein	18, 93
Glaesserella	*Haemophilus parasuis*	pathogenic	−	VtaA2	cell surface proteins	155
Methylobacterium	*Methylobacterium* sp. 4−46	nonpathogenic	−	unnamed		19
Rhodopseudomonas	*R. palustris*	nonpathogenic	−	unnamed		19
Candidatus Solibacter	*Can. S. usitatus*	nonpathogenic	−	unnamed		19

included 7,010 collagen-like sequences with 412,598 triplets or 1,237,794 residues in contiguous $(GXY)_n$ stretches ($n \geq 3$) in the collagen domains. Therefore, the prokaryotic collagenome is considerably more heterogeneous than the animal collagenome, in terms of both the length of the collagenous domains and their sequence composition.[38] It has been suggested that heterogeneity in bacterial collagen-like sequence may be linked to the bacterial adaption to their habitat and that bacteria have evolved mechanisms to maintain stability and activity in diverse environments.[37]

Investigation of animal collagens revealed that the thermal stability of collagens corresponded to the living temperature of the organism.[37] In cases of collagen-mimetic peptides and bacterial CLPs, the number of GXY repeats in collagen-like regions was associated with the thermal stability of collagenous triple helix.[39−41] A recent bioinformatics study of extremophilic and extremotolerant bacteria found that the proteins with CLSs (≥6 GXY repeats) were encoded by 64% of all studied extremophiles of different phyla. Unexpectedly, the length of CLSs from thermophilic Firmicutes tended to be shorter than those of mesophilic ones. The extreme thermophilic bacteria living at ≥70 °C encoded proteins containing relatively short lengths of CLSs composed of 6−7 GXY repeats. On the basis of reported data,[40] such short synthetic peptides without Hyp were not capable of forming stable triple helices. However, it cannot rule out the possibility that a short CLS could behave differently when integrated into a protein, even though it failed to adopt a stable triple helix *in vitro* or possessed low predicted thermal stability. The adjacent region within the CLS-embedded protein, interactions with other proteins, or amino acid modifications could increase the stability of the collagenous motifs.[42]

The bacterial CLPs encoded by different bacterial phyla are heterogeneous. To understand the evolutionary relationships among a large number of bacterial CLPs, the amino acid composition in the GXY repeat regions of bacterial CLPs was examined. Generally, the residue most frequently found in Firmicutes was Threonine (Thr, T), while the most common residue in Actinobacteria was Alanine (Ala, A) (Figure 3A). Contrary to observations that the proportion of Gly was about 34% in most bacterial CLPs, the Gly residue constituted approximately 41% of the amino-acid composition of CLPs in Actinobacteria and Cyanobacteria (Table S2). The amino acid composition at the Xaa and Yaa positions of GXY repeats is another crucial factor that influences the thermal stability of the triple helix.[37] The amino acid residues occupying the Xaa and Yaa positions within the GXY repeat varied significantly in different bacterial phyla. In the Xaa position, the most prevalent amino acid was Pro or Ala in most bacterial phyla (Figure 3B, Table S3). However, the individual amino acid residues occurring at the Yaa position were remarkably diverse (Figure 3C). Serine (Ser, S) residues occurred most frequently at the Yaa positions within CLPs in Proteobacteria, whereas Pro, Ala, and Thr were more commonly found in CLPs from other bacterial phyla (Table S4). Bacteria lacked prolyl-4-hydroxylase and hydroxyproline that contributed to the

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

stability of triple helices in animal collagens. Therefore, the distinctive amino acid compositions for bacterial CLPs might confer different mechanisms for imparting stability to the triple helical structures. For instance, the high content of charged amino acid in CLPs of *S. pyogenes*, *Streptococcus usitatus* (*S. usitatus*), and *R. palustris* provided stability to their collagenous regions through electrostatic interactions.[14,43] The *B. anthracis* BclA protein had a relatively high content of polar amino acids, especially Thr, which was commonly found in the Yaa position and often glycosylated.[17] Hydroxyl groups of Thr or the Thr-bound oligosaccharides might be involved in the formation of water bridges that are conducive to the thermostability of triple helices.[14] Moreover, glycosylated Thr residues in the BclA protein contributed to the adhesion of *B. anthracis* spores to target cells.[44,45] Pro-rich sequences, whose pyrrolidine ring has a constrained conformation, were more frequently found in putative CLPs from thermophilic bacteria than those in the psychrophilic bacteria.[37] Therefore, the amino acid composition diversity of different bacterial CLPs is responsible for various roles in biological activities or physical properties.

3. STRUCTURAL AND FUNCTIONAL DIVERSITY OF BACTERIAL COLLAGENS

Contrary to most mammalian collagens, knowledge concerning the structures and functions of bacterial collagens is far more limited. In particular, only a small number of reported microbial collagens was enzymatically or structurally characterized. Available data were mainly concentrated on just a few species, primarily in members of the Firmicutes phylum (Table 1). CLPs found in pathogenic bacterial species were typically cell surface associated[23] and seemed to participate in pathogenesis processes, including host cell adherence, invasion, and immune evasion.[29]

3.1. CLPs in the *Streptococcus* Genus. The most well-known bacterial collagens are from the *Streptococcus* strains, followed by *Bacillus anthracis*. Streptococcal collagen-like (Scl) proteins are present in many species of pathogenic streptococci, including *S. pyogenes*,[16] *S. pneumoniae*,[27] and *S. equi*.[46] Experiments and bioinformatic analyses provided more evidence that all Scl proteins were anchored to cell walls in a homotrimeric, lollipop-like structural organization.[32] The well-characterized Scl1 proteins of *S. pyogenes* consisted of an N-terminal noncollagenous variable domain (V), a central collagen-like domain (CL), and a C-terminal cell wall associated-region with an LPXTG motif (Figure 4).[47] The V domains forming the globular heads of the lollipop-like structures appeared to be important for efficient triple-helix assembly.[16] Additionally, relying on the serotypes, the Scl1 V domains showed binding specificities to a variety of host proteins, such as plasma lipoproteins LDL/HDL (LDL, low-density lipoprotein; HDL, high-density lipoprotein), complement factor H (CFH), factor H-related protein 1 (CFHR1) or the extracellular matrix (ECM) components fibronectin and laminin.[14] The CL domains of Scl1 forming the stalks conformation could directly bind to human collagen receptors, integrins $\alpha 2\beta 1$ and $\alpha 11\beta 1$, on the host cell surface, thereby facilitating *S. pyogenes* internalization to human cells and reemergence from host cells into the extracellular environment.[48] Interested readers can refer to an excellent review on the classification of Scl proteins and their roles in pathogenesis.[32]

3.2. CLPs in the *Bacillus* and *Lysinibacillus* Genera. *Bacillus* is a diverse genus composed of more than 200 species and widely distributed in different environments. *Bacillus* strains are Gram-positive, aerobic, or facultatively anaerobic, spore-forming rods. Large bodies of research on the *Bacillus* species were mainly focused on two *Bacillus* clades: (1) the *Bacillus subtilis* group, which consists of *B. subtilis* and its close relatives, including *Bacillus licheniformis*, *Bacillus amyloliquefaciens*, *Bacillus pumilus*, and *Bacillus mojavensis*; (2) the *Bacillus cereus* group, which consists of very closely related species, including *B. anthracis*, *B. cereus*, *Bacillus mycoides*, *Bacillus pseudomycoides*, *Bacillus thuringiensis*, and *Bacillus weihenstephanensis*.[49]

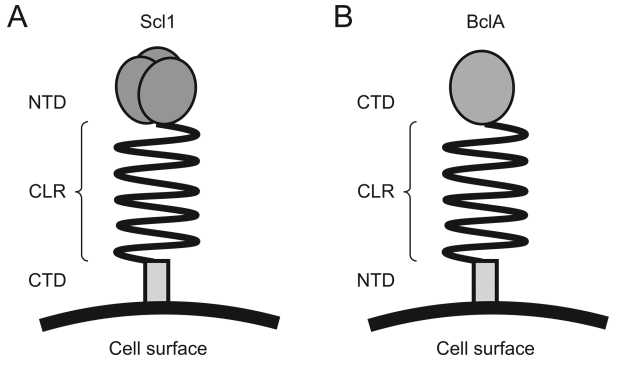


Figure 4. Domain organization of representative CLPs. (A) Mature Scl1.1 protein from *Streptococcus pyogenes* (UniProt: J7M2P0). The Scl protein consists of a variable N-terminal domain (NTD), a collagen-like region (CLR), and a C-terminal domain (CTD) containing the LPXTG cell wall anchor motif (Gram-positive anchor). (B) Mature BclA protein from *Bacillus anthracis* (UniProt: Q81JD7). The BclA protein also contains three domains: the N-terminal domain (NTD) of BclA involved in targeting and anchoring the protein in the exosporium; CLR, the collagen-like region; the C-terminal domain (CTD) of BclA that drives trimerization and forms the tips of the filaments in the exosporium hair-like nap and are the immunodominant target of vertebrate antibodies. Panel B is adapted with permission from ref 156. Copyright 2013 Springer Nature.

Spores of *Bacillus* spp. confer resistance to extreme temperature, radiation, desiccation, and other destructive agents. The coat is a multilayered structure surrounding the spore as a complex protein shell that protects the genome.[50] The structure of this spore coat differed considerably among species (Figures 5−6).[51] In *B. subtilis*, three morphologically distinct layers of the coat were visible in thin-section electron micrographs: a lamellar, lightly staining inner coat; a striated electron-dense outer coat; and an outermost layer called the crust (Figures 6A). A large subset of species, including the *B. cereus* group species and *Bacillus megaterium*, possessed a prominent loose-fitting structure surrounding the coat, termed the exosporium (Figure 6B). The exosporium was readily distinguished from the coat by a clear gap called the interspace,[52] and the structure of the exosporium also differed among species (Figure 6).[53] The exosporium layer was known in most detail for *B. anthracis* and *B. cereus*,[42] where the layer was composed of a paracrystalline basal layer with hexagonal periodicity and an external hairy nap layer. The filaments of the hair-like nap were formed by a single collagen-like glycoprotein, BclA.[17,26,54,55] As the outermost layer of the spore, the exosporium allowed Gram-positive pathogens to directly interact with the environment, particularly with the host immune system.[56] Moreover, the exosporium is a semipermeable barrier. It might serve as an environmental molecular sieve that permitted access to small germinant

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

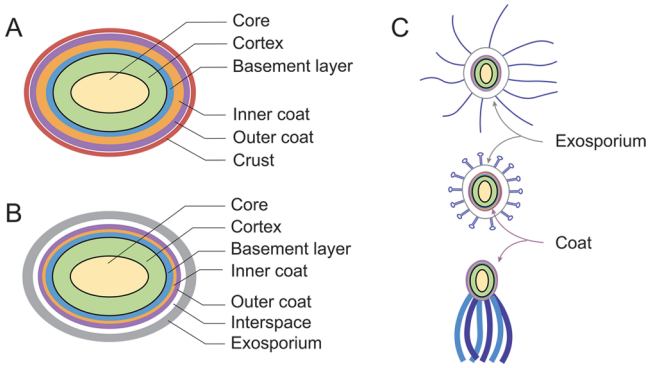


Figure 5. Schematic illustration of bacterial spore structures. Cartoon of a typical *B. subtilis* spore (A) and a *B. cereus* spore (B). The multiple layers of the spore are essential for protecting the bacterial chromosome, which is located within the spore core (yellow). The core is encased by two protective shells: an inner cortex shell (green), made of peptidoglycan, and an outer proteinaceous shell consisting of four layers [the basement layer (blue), inner coat (orange), outer coat (purple), and crust (red), which is only visible after ruthenium red staining]. The black lines between the core and cortex, and the cortex and basement layer, represent the inner and outer membranes, respectively. In a large number of species, the outermost protective layer of the spores is the exosporium (gray), which is separated from the coat by a clear gap termed the interspace. (C) Schematic illustration of the diversity in spore appendages. The spore is drawn following the same format in panel B. The exosporium is denoted by a gray line surrounding the coat. Appendages extend from the coat or exosporium. Top: pili-like structures on spores of some *B. cereus* strains. Middle: pin-like structures on *Clostridium bifermentans* spores. Bottom: ribbon-like structures on *C. taeniosporum* spores. Panel A is adapted with permission from ref 50. Copyright 2013 Springer Nature. Panel C is adapted with permission from ref 86. Copyright 2007 John Wiley and Sons.

molecules, while resisting environmental insults by excluding large molecules, such as antibodies and hydrolytic enzymes.[57−59]

Apart from BclA, several other *Bacillus* collagen-like (Bcl) proteins were identified in *B. cereus* group strains, including BclB, BclC, BclD, and BclE.[60] However, only BclA and BclB were demonstrated to integrate with the exosporium.[61] The length of the BclA collagen-like region (CLR) led to variations in exosporium filament length. Furthermore, the GXY repeats of the collagenous domain provided the primary anchor point for spore-specific sugars (i.e., rhamnose, 3-*O*-methyl rhamnose, and galactosamine) to BclA of *B. anthracis*.[62] Similarly, BclB is also a highly glycosylated protein. In contrast to *S. pyogenes* Scl1 protein, at least in *B. anthracis*, the N-terminal domain (NTD) of BclA was responsible for attachment to the basal layer, while the globular C-terminal domain (CTD) of BclA was oriented externally, forming the distal end of each exosporium filament (Figure 4).[17] The BclA fibers were once considered the major virulence factor of *B. anthracis*. However, subsequent studies confirmed that the fiber and exosporium were dispensable for full virulence of *B. anthracis*.[63] The exact role of the *B. anthracis* exosporium in pathogenesis or in the environment is still under intensive research. By investigating *bclA* delete mutants, the BclA fibers were shown to decrease the adherence of *B. anthracis* spores to nonprofessional phagocytes but not to macrophages. BclA, integrin $\alpha 2\beta 1$, and the complement component C1q were reported to be jointly involved in the entry of *B. anthracis* spores into epithelial cells.[64] Also, BclA mediated the activity of the integrin Mac-1 (CR3) directly, or indirectly via CD14, which ultimately led to enhanced Mac-1-dependent spore internalization. These *B. anthracis* spore−host interactions with BclA appeared to be essential for *B. anthracis* pathogenesis, at least during the initiation of infection.[63,65,66] Furthermore, among 10 more exosporium-associated proteins identified in *B. cereus* ATCC 10876, ExsJ was found containing the GXY repeats and was identical to that in *B. thuringiensis* subsp. Kurstaki.[67]

Interestingly, the bacterial CLP-encoding genes (*clpA* to *clpD*) were also found in *B. subtilis* group strain *B. velezensis* FZB42 (often referred to *B. amyloliquefaciens* FZB42),[30] which was the model strain for Gram-positive, plant growth-promoting, and biocontrol rhizobacteria.[68] Normally, plant growth-promoting rhizobacteria (PGPR) bolster plant growth by establishing themselves on the roots of the plant, while eliminating the presence of pathogens or suppressing pathogen colonization.[69] Recent studies on the root−microbe interaction indicated that the colonization ability and biocontrol activity of PGPR were tightly bound up with their biofilm-forming capabilities.[30,70,71] Root colonization and biofilm formation of PGPR were also considered to be a survival strategy that PGPR adopted to protect plants under various stress conditions.[71] The corresponding CLPs of *B. amyloliquefaciens* FZB42, from ClpA to ClpD, exhibited high homology with those in human pathogenic microorganisms, including *S. pyogenes* and *B. anthracis*. CLPs of the FZB42 strain were shown to be essential for biofilm formation, cellular autoaggregation, cell surface hydrophobicity, and colony wrinkling.[30] Combined with the presence of ClpB and ClpC on the flagella surface in immunogold-labeled images, the detection of ClpB and ClpC in the crude flagella extracts suggested that the two CLPs were associated with flagella and involved in construction of the architectural macrocolony biofilm in *B. amyloliquefaciens*.[72]

Bacillus megaterium forms a clade within the *Bacillus* genus, but it is quite distinct from both the *B. subtilis* and *B. cereus* clades. Similarly, research on the spores of *B. megaterium* was largely focused on the morphology and composition of exosporium.[73] In comparison to spores of the pathogenic *B. cereus* family which were entirely encased by a hairy nap layer, the exosporium of *B. megaterium* QM B1551 spores had filament-like structures localized to a single pole of the spore. The filaments of the pole-localized nap were formed by the BclA1 protein, which was encoded on pBM500, one of the indigenous plasmids required for exosporium assembly in *B. megaterium*.[59] In *B. anthracis*, assembly of BclA and BclB into the exosporium was dependent on their conserved N-terminal sequences, despite the differences in both the NTDs between the two proteins and how they interacted with the BxpB protein.[74] Although lacking the conserved motif within the orthologous BclA and BclB proteins, the NTD of BclA1 was demonstrated to be involved in localizing the putative exosporium basal layer and nap proteins to the developing forespore. Initially, the BclA1 proteins appeared to localize around the entire surface of the developing spore. Subsequently, these proteins detached from most of the spore surface, evidenced by the loss of fluorescence during spore maturation, and only pole-localized BclA1 fibers remained.[73] Furthermore, the BxpB protein, which was required for assembly of the BclA fibers on the surface of the *B. cereus* family exosporium,[61] appeared to be unnecessary for both exosporium assembly and nap attachment in *B. megaterium* spores.[73]

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

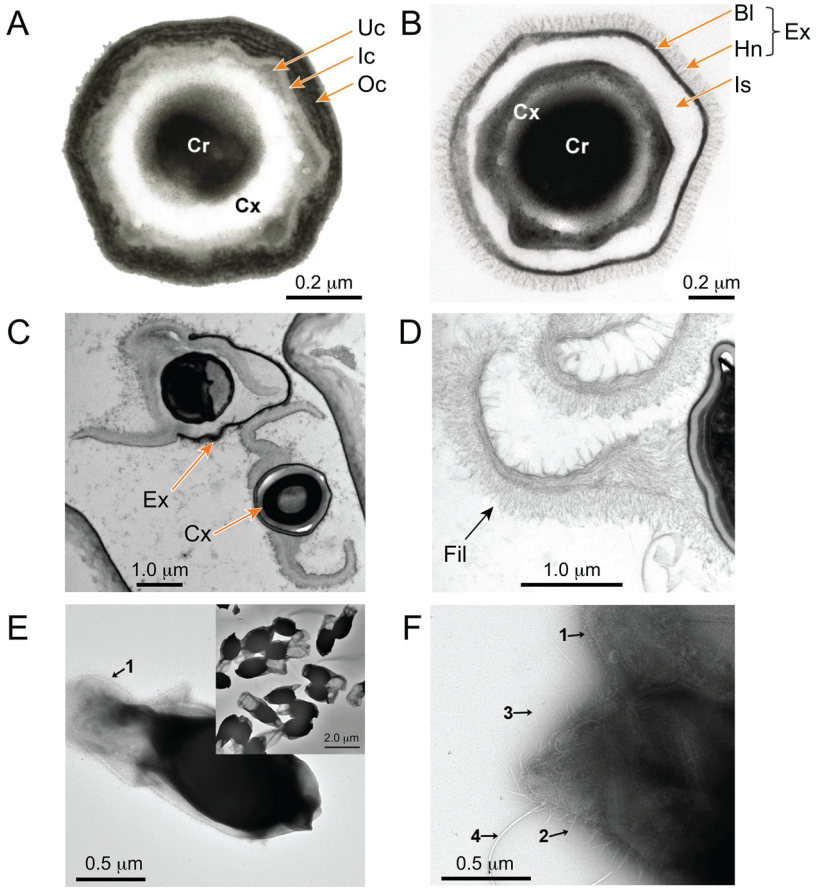


Figure 6. Spore ultrastructure. Electron micrographs of cross sections of spores of (A) *B. subtilis*, (B) *B. anthracis*, (C, D) *Pasteuria ramosa*, and (E, F) *Clostridium sporogenes*. The recognizable components or features of the spore are the spore core (Cr), the cortex peptidoglycan layer (Cx), the undercoat region (Uc), the inner (Ic) and the outer (Oc) coat layers, the interspace (Is), the exosporium (Ex), the basal layer (Bl), the hair-like glycoprotein nap (Hn), and the filaments (Fil). Panel C exhibits the transmission electron micrograph of *P. ramosa* "spore activation", showing the endospore being released from the exosporium. Panel D exhibits the transmission electron micrograph of *P. ramosa* exosporium, showing the densely packed filaments on the surface. Panel E exhibits the negative stain electron micrograph of whole spores of *C. sporogenes*, showing the electron-dense core with the folded sac-like exosporium extended at one pole of the spore. A hairy nap is present on the surface of the exosporium (arrow 1). The inset shows a lower magnification, wider view. Panel F exhibits the high magnification image of edge of exosporium fragments in *C. sporogenes*, showing a part of exosporium with various surface structures labeled: 1, hairy nap; 2, intermediate fibril; 3, beaded fibril; 4, large appendage. Panels A and B are adapted with permission from ref 50. Copyright 2013 Springer Nature. Panels C and D are adapted with permission from ref 83. Copyright 2020 Oxford University Press. Panels E and F are adapted from ref 90 under the terms of the Creative Commons CC-BY license.

Lysinibacillus sphaericus (formerly known as *Bacillus sphaericus*), is a Gram-positive, aerobic, and spore-forming bacterium that is commonly isolated from soils.[75] Among the mosquitocidal protein toxins, the binary toxins (Bin toxins, BinA and BinB) are the most potent mosquitocidal components of *L. sphaericus*.[76] The Bin toxins were produced during sporulation in the form of a parasporal crystal and were enveloped by a balloon-like exosporium.[56,77] Different from *B. cereus* group strains, the exosporium of *L. sphaericus* usually lacked the hair-like nap with BclA proteins as the major component. A collagen-like exosporium protein, BclS, was identified from *L. sphaericus* C3−41.[56] Similar to BclA of *B. anthracis*, the BclS is a glycoprotein that is important for the formation of filamentous structures in *L. sphaericus*. Notably, the BclS filaments might extend into the interspace from the exosporium basal layer. The roles of BclS in heat resistance, germination rate, and outgrowth capacities of *L. sphaericus* spores indicated that BclS and the exosporium were involved in stress tolerance and the persistence of spores.[56]

3.3. CLPs in the *Pasteuria* Genus. *Pasteuria* spp. belong to a group of genetically diverse endospore-forming bacteria (phylum: Firmicutes), the parasite of plant-parasitic nematodes and water fleas (*Daphnia* spp.).[78] As a close relative of genus *Bacillus*, the densely packed surface filaments are also a prominent anatomical feature of *Pasteuria* spores.[79,80] In the inert resting state, *P. ramosa* spores are encased within an intact exosporium. Ingestion by a *Daphnid* host triggers the release of the endospore from the exosporium (Figures 6C). The endospore has regions densely covered by filament structures, an appearance that is very similar to the hair-like nap of *Bacillus* spores (Figure 6D). It was speculated that these filaments were involved in the adhesion of *Pasteuria* endospores to their hosts via a Velcro-like mechanism.[80] CLPs were also discovered in *P. ramosa* and might be responsible for the infectivity of *P. ramosa*. Pcl1a (*Pasteuria* collagen-like protein 1a) was the first CLP identified on the surface of *P. ramosa* spores, based on differential 2D-gel electrophoresis patterns between two isolates with different infectivity.[81] Different from the highly conserved motifs (e.g., LPXTGE) identified in

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

Table 2. Characterization of Reported CLPs from Both Gram-Positive and Gram-Negative Bacteria

bacterium	protein	accession number	no. of amino acids[a]			collagen-like region[a]			ref
			total	N-terminus	C-terminus	no. of GXY repeats	T_m (°C)	triple-helix validation	
Gram-Positive									
B. anthracis[b]	BclA	WP_000069712.1	400	19	134	76	37.0	CD and trypsin	17
B. anthracis[b]	BclB	WP_011178707.1	348	33	165	50			62
B. cereus	ExsJ	AAN85822.1	430	157	164	with Gly interruption			67
B. amyloliquefaciens FZB42	ClpA	WP_012117055.1	228	67	134	9			30
B. amyloliquefaciens FZB42	ClpB	WP_012117056.1	665	158	135	124			30
B. amyloliquefaciens FZB42	ClpD	WP_012117058.1	459	72	174	71			30
B. megaterium QM B1551	BclA1	WP_013060180.1	954	159	135	220			59
B. megaterium QM B1551	BclA2	ADE72420.1	438	158	136	48			59
C. difficile 630	BclA1	WP_011860734.1	693	193	146	118	n.d.[f]	collagenase digestion	25
C. difficile 630	BclA2	WP_011861885.1	558	13	131	138	n.d.	collagenase digestion	25
C. difficile 630	BclA3	WP_011861926.1	661	50	147	with Gly interruption			85
C. perfringens SM101	unnamed	ABG86771.1	403	53	161	63	38.8	CD and trypsin	19
L. sphaericus C3−41	BclS	WP_012292195.1	1142	3	158	with Gly interruption			56
P. ramosa[c]	Pcl1a	ABY19404.1	625	75	7	181			81
S. equi (GCS)[e]	SclC	WP_015898629.1	302	59	90	39			46, 154
S. pneumoniae (GBS)[e]	PclA	AAL00207.1	2551	548	126	discontinuous GXY repeats			27
S. pyogenes (GAS)[e]	SclA/Scl1	WP_010922693.1	348	68	93	50	36.4	CD and trypsin	16
S. pyogenes (GAS)[e]	SclB/Scl2	AAL50184.1	444	73	100	79	37.6	CD and trypsin	16
Gram-Negative									
E. coli O157:H7 strains	EPclA	NP_310744.1	437	250	76	37	42.0[d]	CD	36
L. pneumophila	Lcl	WP_010948344.1	493	30	158	94	n.d.	CD and trypsin	18, 93
H. parasuis strain Nagasaki	VtaA2	EYE72921.1	1225	364	348	171			155
Can. S. usitatus	unnamed	WP_011683117.1	434	9	148	82	38.5	CD and trypsin	19
Methylobacterium sp. 4−46	unnamed	ACA18713.1	344	100	73	49	35.0	CD and trypsin	19
R. palustris	unnamed	WP_012497020.1	212	9	86	39	37.0	CD and trypsin	19

[a]The numbers of residues given are for a specific reported isolate. [b]*Bacillus anthracis* Sterne strain. [c]*Pasteuria ramosa* laboratory isolates P1 and P3. [d]Abbreviations in parentheses referred to group A, B, and C *Streptococcus*. [e]T_m of the full-length protein [f]n.d. − not determined.

Gram-positive bacterial surface proteins, there was a hydrophobic transmembrane region within the NTD of Pcl1a. Subsequently, more than 37 polymorphic collagen-like genes were discovered across genomes of 10 *P. ramosa* strains, which were extracted from *Daphnia* hosts with diverse species, geographic origins, and *P. ramosa* susceptibility.[33] They are likely to be associated with the interaction between *P. ramosa* and *Daphnia* host.[82] Additionally, an abundance of CLPs, including Pcl7, was identified from the endospore of *P. ramosa* by proteome sequencing.[83]

Likewise, a group of collagen-like putative genes was identified from the closely related species *Pasteuria penetrans* (strain: Res148). These 17 *pcl* genes formed five different phylogenetic clusters, suggesting that CLPs were a principal source of genetic diversity in animal pathogenic Firmicutes, including *Pasteuria*. It was demonstrated that the fibers on the surface of the *P. penetrans* endospores were collagen-like. However, further research is needed to confirm their structures.[80]

3.4. CLPs in the *Clostridium* Genus. *Clostridium difficile* is a spore-forming anaerobic pathogen, and the primary cause of antibiotic-associated diarrhea.[84] Spores of *C. difficile* play a vital role in infection, persistence, and transmission of *C. difficile* infections. *C. difficile* spores possess an exosporangial layer that surrounds the spore coat. The structures of this outer layer varied among isolates.[85] Three BclA orthologues, BclA1, BclA2, and BclA3, were identified in exosporium extracts of *C. difficile*. All three BclA were expressed during sporulation, and their localization to the spore depended on the N-terminal domains. BclA1 and BclA2 both contained triple helical structures due to their susceptibility to collagenase digestion. BclA1 was localized in the exosporium layer with its N-terminus buried inside of the exosporium while the C-terminus was oriented to the exosporium surface.[25] Moreover, the BclA3 homologue could be found on the spore surface of *C. difficile*. Characterization of *C. difficile* exosporangial surface extracts confirmed that BclA3 was a glycoprotein modified with chains of β-O-linked GlcNAc and with additional novel glycans.[85] It was firmly established that the surface-associated bacterial glycans played an indispensable role in the interactions between host and pathogenic bacteria. Consistent with previous reports, the glycans on the *C. difficile* spore surface conferred heat resistance to the spores and appeared to be involved in macrophage interactions.

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

In contrast to *B. cereus* spores that are enclosed in a sac-like exosporium,[26,54,55,67] many spores have external appendages organized as ribbons, pili, feathers, tubules, or swords (Figures 5).[86] Exosporia and appendages are produced separately: some species can produce either one while some species produce both. Also, formation of appendages is strain dependent. *Clostridium taeniosporum* was described as an appendage-forming *Clostridia* that attaches about 12 large, ribbon-like appendages through a common trunk at one spore pole to a layer outside the coat, called the "encasement". Unlike the loosely fitting exosporia, the encasement layer fits tightly around the spore.[87,88] In *C. taeniosporum*, fibrils were clearly visible as a hair-like nap along the appendage edge in atomic force microscope (AFM) images. Among the four appendage proteins identified in *C. taeniosporum* spores, a glycoprotein designated GP85 was shown to include a collagen-like region. Upstream of the collagenous region, the sequences of GP85 and the deduced CL2 protein were significantly homologous; the latter contained a collagen-like domain as well.[87] Although the GP85 and CL2 proteins shared some common features, there was no evidence that CL2 was a component of *C. taeniosporum* spore appendage.

Clostridium sporogenes is nonpathogenic, anaerobic, soil born bacterium, widely used as a surrogate for proteolytic strains of *Clostridium botulinum*.[89] As seen by electron microscopy (EM) and AFM, features of the *C. sporogenes* exosporium resembled those of the *B. cereus* group, such as the paracrystalline basal layer, the hair-like nap, and a deformable sac-like layer that was extensive at one pole and more intensive at the other end of the spore (Figures 6E).[59] Interestingly, beaded fibrils with bead-like structures at regular intervals were observed on the surface of *C. sporogenes* exosporium (Figures 6F). The BclA, a protein with a characteristic collagenous domain that was similar to the BclA of *B. anthracis*, was identified in *C. sporogenes* exosporium proteins. The *C. sporogenes* BclA might be a possible constituent of the hairy nap. Another collagen-like domain containing protein was identified, named BclB, since its CTD was homologous to that of *B. anthracis* BclB. However, the motifs within N-termini of *B. anthracis* BclA and BclB responsible for their localization to the exosporium were absent in the BclA and BclB proteins of *C. sporogenes*.[90]

3.5. Features and Functions of Polymorphic CLPs. Polymorphisms in the numbers of GXY triplets were detected in CLPs from both Gram-positive and Gram-negative bacteria (Table 2).[27,91] In *B. anthracis*, the length of the central collagenous domain of BclA was highly polymorphic, with 17−91 GXY tripeptides. The variation of exosporium filament hair length was dependent on the length of BclA collagenous domain.[54] Also, sequence polymorphisms could be found in the CLR region of Pcl1a in *P. ramosa*, implying that CLPs might be involved in the infectivity of *P. ramosa*.[80,81] Further investigations of Pcl1a and 37 novel putative PCLs (*P. ramosa* collagen-like protein, PCL) discovered from 10 *P. ramosa* strains revealed polymorphisms in PCL by bioinformatic characterization of structural features in these PCLs.[33] These polymorphisms appeared as differing numbers of GXY repeats, predicted protein length, and amino acid substitution. On the basis of these differences, five genotypes were discriminated among the 10 tested strains, which perfectly matched the pattern of *Daphnia*/*P. ramosa* interaction specificity. Recently, multiple polymorphisms were found within a single gene for the collagen-like protein Pcl7 in *P. ramosa* strains, which corresponded perfectly with one common and widespread infection phenotype.[83] These findings supported the hypothesis that polymorphic PCLs were responsible for high specificity in the interaction between *Pasteuria* and host. Sequence polymorphisms of the *Bacillus* collagen-like protein *bcl* genes were previously proposed as a tool for differentiating *B. anthracis* from other members of the *B. cereus* group.[54,92] Likewise, PCL polymorphisms revealed a strong potential for *Pasteuria* strain and species discrimination.[33]

Lcl, the collagen-like adhesin of *L. pneumophila*, is essential for bacterial attachment, autoaggregation, sedimentation, and biofilm formation.[18,28] Characterization of Lcl from different *L. pneumophila* isolates showed length polymorphisms in the collagenous repeats.[18] The intensity of the circular dichroism (CD) signature of the triple-helix increased with the number of Lcl collagenous repeats.[93] Similar length-dependent effects on the stability and aggregation propensity of the triple-helix domain were also observed in engineered collagenous proteins and collagen-like peptides.[93−95] Significantly, there was a positive correlation between the number of GXY tandem repeats in Lcl variants with the binding affinity of Lcl, as well as the biofilm-forming capacity of *L. pneumophila* strains.[93] The role of bacterial biofilms in pathogenesis was well-established for numerous infections, including *Legionella* infections.[96,97] Thus, the influence of number of Lcl tandem repeats on the clinical prevalence of *L. pneumophila* strains was investigated further. Interestingly, the results showed that Lcl variants containing 15 to 18 GXY repeats produced more robust biofilms than those with lower numbers of GXY repeats, while strains with more than 18 GXY repeats exhibited a reduced clinical incidence.[93] Therefore, the number of Lcl GXY repeats might serve as a virulence marker for identifying clinically relevant *L. pneumophila* strains.

3.6. Engineering Applications of Bacterial CLPs. As described in Tables 1 and 2, most of the well-characterized bacterial CLPs belonged to bacteria from the Firmicutes phylum. They were mainly cell surface-associated proteins, including endospore surface proteins, exosporium proteins, and components of the cell surface matrix. Correspondingly, conservative domains of known endospore surface proteins, such as BclA_C (PF18573) and Exospore_TM (superfamily cl14017), were identified in the CLPs from *B. anthracis*, *B. cereus*, *B. megaterium*, *B. amyloliquefaciens*, *C. perfringens*, and *C. difficile* (Table S5). Motifs characteristic of phage-related proteins and LPTX anchor motifs targeting proteins to the cell surface were also identified, which were consistent with the role in function or localization of CLPs. Additionally, recent research on CLPs in extremophiles predicted that some Firmicutes CLPs contained several conservative domains related to ligand-binding functions (e.g., cellulose-binding domains, chitin-binding domain ChiAl_BD) and enzymatic reactions (e.g., Lipase_GDSL_2, Peptidase_4 domains).[37] However, many conserved regions were annotated as domains of unknown function.

Owing to the noncytotoxicity and nonimmunogenicity, some bacterial CLPs found applications in the fields of cosmetics and biomedicine.[98] Crucially, the bacterial CLPs are thermally stable at body temperature without the need for secondary modification (prolyl hydroxylation).[99] The *S. pyogenes* Scl2 gene is studied most frequently. Similar to Scl1, the Scl2 protein consists of 3 main domains: a V domain at N-terminus, a CL domain of 78 triplets, and an attachment domain at C-terminus that is generally removed from recombinant proteins.[7,15] The triple-helix of Scl2 protein is

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

only about one-fifth the length of a human fibrillar collagen. Unlike the Scl1 protein, the CL domain of Scl2 protein was shown to be biologically inert. Thus, this protein could be used as a "blank slate" template for introducing various biological motifs and functions. Over the past few years, Scl2 protein was employed as a bioengineering model to define collagen function. A variety of specific sequences were introduced within the CL domain or between two CL domains, including integrin/heparin/fibronectin-binding domains, discoidin domain receptors, and MMP cleavage sites.[7] Sequence−structure−function relationships were also explored through substitutions of amino acids within the inserted biologically active sequences. Meanwhile, a number of studies demonstrated that recombinant Scl2 protein could be easily engineered with structural and functional modifications to meet biomedical requirements. For purposes of tissue engineering, binding domains or growth factors could also be incorporated in scaffolds of Scl2, in the form of porous sponges, films, or gels.[99,100] Modifying the N- and C-terminals and introducing cross-links could improve the strength and stability of collagen materials. Interested readers can refer to detailed reviews of bacterial collagens in biomedical applications by Yu et al.,[14] Ramshaw and Werkmeister,[7] and Brodsky and Ramshaw.[15]

Bacterial collagens are potential candidates for synthesizing biomaterials with tunable properties in biomedical applications. However, the following aspects should receive more attention when developing biomaterials based on the Scl proteins: (1) Scl2 shows no biological activity, but *S. pyogenes* is a pathogen, which raises concerns on the potential role of collagen in pathogenesis and hence possible side effects.[101] (2) Scl proteins can be recombinantly produced in high yields using *E. coli*, but the fermentation yields of the larger multimers are relatively low.[102] (3) The low thermal stability of the Scl proteins does not affect the production of three-dimensional scaffolds but impairs the spinnability of collagen fibers.[103] Although the majority of previous publications focused on Scl2, future research should be extended to collagen-like proteins from other bacteria to develop new features and variations for synthesizing novel biomaterials. The predicted CLP from *B. subtilis* may be a likely competitor for animal-derived collagen that can be useful for *in vivo* applications. This CLP contains continuous $(GXY)_{1034}$ repeats: a lengthy collagenous domain resembling that of human fibrillar collagen. The CLPs from extremophilic bacteria could be more robust and stable compared to the known mesophilic CLPs, thus being excellent candidates for new highly stable collagens.[37]

4. MODELING AND SIMULATION OF BACTERIAL CLPS

4.1. Structural Stability and Binding Affinity. Both mammalian and bacterial CLPs show remarkable stability due to the triple-helical structures that come about with or without Hyp residues. In mammalian collagens, molecular mechanics and *ab initio* simulations revealed that the gauche effect and electrostatic interactions mainly stabilized the pyrrolidine ring conformation to preserve the Hyp structural stability of collagen triple helix.[104] Some bacterial collagens shared the same triple-helical structures with mammalian collagens but lacked Hyp.[14] Pro and Hyp have a frequency of 27% and 38% in the Xaa and Yaa positions of mammalian collagens, respectively.[11] Bacterial collagens have 30% Pro residues in the Xaa position, but only 5% in the Yaa position.[32,105] Despite this difference, bacterial collagens retain a high thermal stability of 35−39 °C. Mohs et al. attributed this stability to electrostatic interactions and enthalpic contributions involving polar groups interactions, thereby stabilizing recombinant Scl2 proteins.[43] Interestingly, experimental validation and bioinformatics analysis showed that Scl2 could be modified by replacing the V trimerization domain with a *de novo* designed, stable coiled-coil to encourage or discourage folding.[106] Moreover, the crystal structure of the Scl2.3 globular domain was first reported using crystallographic data and molecular dynamics (MD) simulations.[107] This work expanded the design space of molecular tools in inhibiting streptococcal invasion. Furthermore, bacterial collagen stability could be quantitatively related to the amino acid sequence by an algorithm[40] that predicted the melting temperatures of multiple triple helical peptides relatively accurately, barring situations where there were oppositely charged residues or concentrations of similar charges.

Recombinant bacterial collagens have been employed as a tool to probe the biological consequences relative to the mutation site that may relate to the observed pattern of osteogenesis imperfecta mutations or Alport syndrome mutations. By combining MD simulations and experimentations, Gly → Ser mutations in the essential integrin binding motif (GFPGER) of the type I collagen triple helix played a role in preventing integrin from binding and inhibiting cell adhesion.[108,109] There was an asymmetry in biological consequences relative to the mutation site: Gly substitutions N-terminal to the GFPGER sequence decreased integrin binding, whereas the same mutation C-terminal to GFPGER did not affect integrin binding. Further investigation of the effects of Gly mutations on fibronectin−collagen binding revealed that Gly → Ser replacements hindered the interactions between the collagen β-strand and the β-sheet structure of the fibronectin modules.[110] These results provided insight into the effects of mutations and motivates future studies into the consequences of osteogenesis imperfecta mutations. Recently, on the basis of recombinant bacterial collagen system, Yeo et al. integrated MD simulations with experiments and discovered the adverse effects of glycine missense mutations in the collagenous domain of collagen type IV on structural stabilities, triple helical peptides degradation, and binding affinities between collagen and integrins, thereby paving a way toward developing new molecular therapies for Alport syndrome.[111]

Effects of peptide sequence on the lower critical solution temperature (LCST)-like transition temperatures of elastin-like peptide (ELP) and ELP-CLP conjugates were systematically studied with all-atomistic (AA) MD, coarse-grained (CG) MD simulations[112,113] and experiments.[114] The mechanism of guest residues affecting the LCST-like transition temperatures was clearly elucidated. Furthermore, human-like collagen (HLC), formed from recombinant collagen cDNA, could help tolerate probiotics by generating intermolecular hydrogen bonds. Hence, HLC might be beneficial hemostatic materials for scaffolds in tissues and organs.[115] Encapsulation of HLC and *Bifidobacterium longum* was further studied and evaluated in simulated gastrointestinal conditions,[116] proving that alginate-HLC could potentially be a powerful tool for delivering oral bioactive materials.

4.2. Biomechanics. Mechanical forces can alter the structure of collagen, thereby further impacting the synthesis

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

and degradation of collagen. However, the detailed mechanisms that relate mechanical loading and structural changes are challenging to unravel. Bacterial adhesion to collagen is the beginning, and the most critical part, of bacterial infections, but the underlying mechanism of adhesion is not well-elucidated. The mechanics in collagen binding were examined using steered MD simulations with umbrella sampling in *Staphylococcus aureus* to fundamentally understand the structural and conformational dynamics of collagen adhesin.[117] More importantly, their interactions with collagen leading to the final attachment were quantitatively revealed. A dynamic and multistep binding model named "Collagen Hub" was also developed to describe the binding of ligands to bacterial collagens.[118] In addition, *ab initio* MD simulations showed that bacterial adhesion facilitated hydration in collagen recognition, which could result in stronger interactions between the solute and the solvent, hence further enhancing protein−protein recognition.[119] Molecular modeling and experiments were also combined to gain insight into the structural elements of the bacterial collagen adhesin, which controls ligand binding.[120] This study showed how the interactions between type I collagen and microbial surface components impacted collagen binding affinity.[120] Classifications and ligand-binding specificities of CLPs in pathogenic streptococci were also significantly reviewed by a previous publication.[32] Moreover, An et al. reviewed a wide range of collagen interactions and their applications in drug design and delivery.[121]

Strain-dependent traits of collagen fibers degradation were studied with a multiscale model, which paves the way further toward accurate structural remodeling of collagenous tissues.[122] Furthermore, the effects of equi-biaxial strain rate of collagen and collagenase concentration on collagen catabolism were systematically studied,[123] showing that larger strain resulted in significantly lower median collagen concentrations, which indicated stronger catabolism. In addition, mechanical forces play an important part in enzyme reactions. Yi et al. found that mechanical forces enhanced the enzyme activity in general while a static strain of 20% prevented digestion compared to lower or higher strains, thereby providing greater insight into factors that influenced collagen degradation.[124] The mechanical performance of collagen molecules, especially in collagen degradation, collagen breakdown mediated by mechanical forces,[125] and collagen-based materials,[126] were extensively reviewed. These reviews concluded that the interplay between entropic elasticity, energetic elasticity, local structures, instabilities, and interactions with enzymes organically formed a network of interactions that determined the mechanical performance of collagen.

4.3. Bioinformatics. Bioinformatics are important tools in statistically interpreting biological phenomena by analyzing large and complex biological data sets to predict amino-acid sequences, protein structures, and locate genes in a given sequence. Via *in silico* analyses of genomic sequences and molecular modeling, 100 novel bacterial or viral proteins with collagen-related structural motif sequences were found in bacterial genome databases, which were largely different from mammalian collagens in amino acid content and distribution.[12]

Bioinformatics analyses were also applied to identify and screen genes or proteins for structural purposes with web-based tools ProtParam (https://www.expasy.org/resources/protparam), DAS (https://tmdas.bioinfo.se/DAS/index.html), and SignalP (http://www.cbs.dtu.dk/services/SignalP). A gene named *cne* from *S. equi* was identified successfully, which encoded a collagen-binding protein contributing to both bacterial adhesion and immune evasion.[127] To study the molecular interactions and binding energies of cross-linked type I and III collagens, bioinformatics analyses via quantitative structure−activity relationships successfully selected appropriate cross-linkers for stabilization.[128] In contrast to the mechanical properties and strengthening mechanisms of metallic and nonmetallic materials,[129−137] cross-linking of succinic acid with collagen yielded remarkable improvements in mechanical strength.[138] This advantage was mainly attributed to the noncovalent interactions. Gene ontology enrichment and functional annotation analyses were also used to screen specific proteins performing structural functions in eggshell membrane formation.[139] Collagen-processing- and cross-linking-related genes were identified in this study to obtain insight into the eggshell membrane structure and forming mechanisms.

Bioinformatics also contributed to monitoring and preventing various diseases with differential gene expression profile analysis. For example, several genes related to Dupuytren's disease pathogenesis were discovered from the transcriptome profiles.[140] To monitor clinical mastitis at the late infectious stage, isobaric tags for relative and absolute quantification (iTRAQ) analyses were used to quantitatively screen specific proteins associated with mastitis.[141] This approach showed the effects of mastitis on CLP and motivated future efforts in locating potential genes for mastitis susceptibility.[141] Using integrated bioinformatics, COL1A2 was found to be the key gene in a 37-gene network modulating cell mobility.[142] COL1A2 was highly expressed in human gastric cancer (GC), thereby stimulating further identification of diagnostic biomarkers for GC. A collagen type IV α1 (COL4A1) chain was found to possibly resist trastuzumab in GC, which was further validated by biological process annotation analyses and functionality analyses of relevant microRNAs.[143] At the same time, the molecular mechanisms of GC were investigated with the identification of six hub genes including COL4A1 and COL1A2 using an integrated bioinformatics approach that was further validated with experiments.[144] ECM−receptor interactions and focal adhesion pathways affected GC development and possible biomarkers were found for GC. A recent study pointed out that COL1A1 and COL1A2 could be potential GC biomarkers using integrated bioinformatics analysis.[145] Furthermore, a panel of overexpressed collagen genes in GC and their functions were identified and analyzed.[146] MicroRNA-29c-3p could potentially regulate and target COL1A1 and COL4A1.

In colorectal cancer (CRC), a recent study found that minor changes in collagen fiber arrangements after remodeling of collagen tended to be similar with normal collagen while major changes tended to promote cross-linking into bundles.[147] Collagen type I and III deposition increased while collagen type IV deposition were influenced by distant metastasis, revealing the correlation between collagen remodeling and the prognosis of CRC on the basis of protein functional annotation, differential mRNA expression, functional and pathway enrichment, cox proportional hazards regression, survival, and statistical analysis.[147] Gelatin speciation, mainly for collagen type I, II, and III, as well as adulteration were also accurately identified using peptide biomarkers with protein alignment and sequence analyses and mass spectrometry.[148] This study provided a validation method in food analysis for

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

effectively screening adulteration in food manufacturing. In pancreatic ductal adenocarcinoma (PDAC), molecular mechanisms were investigated statistically, showing that genes of collagen family contributed to the development of PDAC. Moreover, the interactions between the extracellular matrix and the receptor, as well as focal adhesion also played significant roles in the progression of PDAC.[149] Using hierarchical clustering, functional enrichment, and principal component analyses, mitofusin2 (Mfn2) was also identified to be a direct target gene of microRNA-214, which could mediate collagen synthesis in cardiac fibroblast.[150] Mfn2 could critically regulate cell proliferation and tissue fibrosis.

As noted above, a family of polymorphic bacterial CLPs was identified and characterized in *P. ramose*, describing more CLPs in the bacterial genome.[33] Using molecular modeling and phylogenetic analyses, a set of *Burkholderia* CLPs were systematically studied, facilitating better understanding of the pathogenicity factors, genome evolution, and infection detection.[29] Recently, the interactions between collagen and other proteins such as cellulose and pectin were systematically studied, revealing the magnitude of binding free energies, van der Waals bond, hydrogen bond, and desolvation energy between these complexes.[151] Furthermore, type I collagen from various sources could be interchangeably identified due to their unique traits, suggesting the use of collagen as a potential source of bioactive peptides.[152] To overcome the complicated computational tools used for biological research, an open bioimage informatics platform was introduced to facilitate reproducible computational study. This platform developed a community Web site and software with a visual interface to facilitate complex imaging workflows.[153]

5. CONCLUSIONS

Much is known about the structures and functions of mammalian collagen, the ubiquitous triple-helical scleroprotein that plays a central role in tissue architecture, extracellular matrix organization, and signal transduction. In stark contrast, there is a relative dearth of studies into bacterial collagen that is fueling a burgeoning field for intensive research. Currently, only a small fraction of the predicted bacterial collagen-like proteins (CLPs) was investigated in detail. However, there is tremendous potential for further discoveries as proteins containing large regions of Gly-Xaa-Yaa (GXY) repeats were uncovered in numerous bacterial genome databases. Therefore, our review aimed to fill and highlight existing gaps due to the general paucity of data on bacterial CLPs by providing a comprehensive overview of the latest insight into their functional and structural diversity from multiple perspectives of biology, computational simulations, and materials engineering. We explored the origins and discovery of bacterial CLPs, analyzed their genetic distribution and molecular architecture, and examined their structural and functional diversity in various bacterial genera. We also discussed the important roles of computational techniques in understanding bacterial CLPs' structural stability, mechanical properties, and biological functions. We strongly anticipate that this review will drive further interest and development of bacterial CLPs, not only for addressing fundamental biological problems in collagen but also for engineering novel biomaterials. For instance, the distinctive amino acid compositions for bacterial CLPs confer different mechanisms for imparting stability to collagenous triple-helical structures as bacteria lack the prolyl-4-hydroxylase and hydroxyproline responsible for the stability of triple helices in animal collagens. Processing methods that enhance trimerization by copolymerization may also lead to better yields from synthesis, thereby bolstering the production of environmentally sustainable biomaterials. Moreover, new production techniques are needed to overcome difficulties in aggregating bacterial collagens into large fibers. Some potential techniques include thermoplastic molding, spinning, or by forming composites. The relatively poor mechanical properties may also be surmounted and optimized by tuning the proportions of the binding and cross-linking peptides. By further connection with such fundamental revelations along the continuum of bacterial collagen sequence, structure, and function, novel means of harnessing bacterial collagen in tandem with synthetic biology may create an even broader array of functional engineering materials using a highly renewable resource. Furthermore, by embedding biologically active sequences from human collagen within the bacterial collagen domain, the resulting biomaterial can be adapted for specific environments and engineering applications. These biomaterials may also minimize hurdles posed by regulatory compliance for animal-derived products. Hence, both the biology and materials communities will greatly benefit from intensified research into the diverse structures and functions in bacterial collagen-like proteins.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsbiomaterials.1c00018.

Table S1 of annotated CLPs from different bacterial species (XLSX)

Table S2 of percentage amino acid composition, Table S3 of percentage of individual amino acid at the Xaa position, and Table S4 of percentage of individual amino acid at the Yaa position (PDF)

Table S5 of conservative domains identified in proteins containing collagen-like sequence (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Jingjie Yeo − *J^2 Lab for Engineering Living Materials, Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York 14850, United States;* orcid.org/0000-0002-6462-7422; Email: jingjieyeo@cornell.edu

Authors

Yimin Qiu − *National Biopesticide Engineering Technology Research Center, Hubei Biopesticide Engineering Research Center, Hubei Academy of Agricultural Sciences, Biopesticide Branch of Hubei Innovation Centre of Agricultural Science and Technology, Wuhan 430064, PR China; State Key Laboratory of Biocatalysis and Enzyme Engineering, School of Life Sciences, Hubei University, Wuhan 430062, PR China*

Chenxi Zhai − *J^2 Lab for Engineering Living Materials, Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York 14850, United States*

Ling Chen − *National Biopesticide Engineering Technology Research Center, Hubei Biopesticide Engineering Research Center, Hubei Academy of Agricultural Sciences, Biopesticide Branch of Hubei Innovation Centre of Agricultural Science and Technology, Wuhan 430064, PR China*

Xiaoyan Liu − *National Biopesticide Engineering Technology Research Center, Hubei Biopesticide Engineering Research Center, Hubei Academy of Agricultural Sciences, Biopesticide Branch of Hubei Innovation Centre of Agricultural Science and Technology, Wuhan 430064, PR China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsbiomaterials.1c00018

Author Contributions

Y.Q. and J.Y. designed the structure of the article and edited the manuscript. All authors contributed to the writing of the manuscript and have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.Y. acknowledge support from the US National Science Foundation (Grant No. 2038057) and Cornell University's Cornell-China Center and faculty startup grant. Y.Q., L.C., and X.L. acknowledge support from the National Natural Science Foundation of China (Grant No. 32000052), the Hubei Provincial Natural Science Foundation of China (No. 2020CFB249), the Youth Foundation of Hubei Academy of Agricultural Sciences (No. 2020NKYJJ16), and the Open Funding Project of the State Key Laboratory of Biocatalysis and Enzyme Engineering (No. SKLBEE2019004).

■ REFERENCES

(1) Duarte, A. S.; Correia, A.; Esteves, A. C. Bacterial collagenases − a review. *Crit. Rev. Microbiol.* **2016**, *42* (1), 106−126.
(2) Fidler, A. L.; Boudko, S. P.; Rokas, A.; Hudson, B. G. The triple helix of collagens - an ancient protein structure that enabled animal multicellularity and tissue evolution. *J. Cell Sci.* **2018**, *131* (7), jcs203950.
(3) Ricard-Blum, S. The collagen family. *Cold Spring Harbor Perspect. Biol.* **2011**, *3* (1), a004978−a004978.
(4) Shoulders, M. D.; Raines, R. T. Collagen structure and stability. *Annu. Rev. Biochem.* **2009**, *78*, 929−58.
(5) Brodsky, B.; Ramshaw, J. A. The collagen triple-helix structure. *Matrix Biol.* **1997**, *15* (8−9), 545−54.
(6) Brodsky, B.; Persikov, A. V. Molecular structure of the collagen triple helix. *Adv. Protein Chem.* **2005**, *70*, 301−39.
(7) Ramshaw, J. A. M.; Werkmeister, J. A. 2.23 Recombinant proteins as emerging biomaterials. *Comprehensive Biomaterials II* **2017**, 512−531.
(8) Bretscher, L. E.; Jenkins, C. L.; Taylor, K. M.; DeRider, M. L.; Raines, R. T. Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* **2001**, *123* (4), 777−778.
(9) Perret, S.; Merle, C.; Bernocco, S.; Berland, P.; Garrone, R.; Hulmes, D. J. S.; Theisen, M.; Ruggiero, F. Unhydroxylated triple helical collagen I produced in transgenic plants provides new clues on the role of hydroxyproline in collagen folding and fibril formation. *J. Biol. Chem.* **2001**, *276* (47), 43693−43698.
(10) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science* **1994**, *266* (5182), 75−81.
(11) Bella, J. Collagen structure: New tricks from a very old dog. *Biochem. J.* **2016**, *473* (8), 1001−25.
(12) Rasmussen, M.; Jacobsson, M.; Björck, L. Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J. Biol. Chem.* **2003**, *278* (34), 32313−32316.
(13) Engel, J.; Bächinger, H. P. Collagen-like sequences in phages and bacteria. *Proc. Indian Acad. Sci. (Chem. Sci.)* **1999**, *111* (1), 81−86.
(14) Yu, Z.; An, B.; Ramshaw, J. A. M.; Brodsky, B. Bacterial collagen-like proteins that form triple-helical structures. *J. Struct. Biol.* **2014**, *186* (3), 451−461.
(15) Brodsky, B.; Ramshaw, J. A. M. Bioengineered collagens. *Subcell. Biochem.* **2017**, *82*, 601−629.
(16) Xu, Y.; Keene, D. R.; Bujnicki, J. M.; Höök, M.; Lukomski, S. Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J. Biol. Chem.* **2002**, *277* (30), 27312−27318.
(17) Boydston, J. A.; Chen, P.; Steichen, C. T.; Turnbough, C. L. Orientation within the exosporium and structural stability of the collagen-like glycoprotein BclA of *Bacillus anthracis*. *J. Bacteriol.* **2005**, *187* (15), 5310−5317.
(18) Vandersmissen, L.; De Buck, E.; Saels, V.; Coil, D. A.; Anné, J. A *Legionella pneumophila* collagen-like protein encoded by a gene with a variable number of tandem repeats is involved in the adherence and invasion of host cells. *FEMS Microbiol. Lett.* **2010**, *306* (2), 168−176.
(19) Xu, C.; Yu, Z.; Inouye, M.; Brodsky, B.; Mirochnitchenko, O. Expanding the family of collagen proteins: recombinant bacterial collagens of varying composition form triple-helices of similar stability. *Biomacromolecules* **2010**, *11* (2), 348−356.
(20) Charalambous, B. M.; Keen, J. N.; McPherson, M. J. Collagen-like sequences stabilize homotrimers of a bacterial hydrolase. *EMBO J.* **1988**, *7* (9), 2903−2909.
(21) Whatmore, A. M. *Streptococcus pyogenes sclB* encodes a putative hypervariable surface protein with a collagen-like repetitive structure. *Microbiology* **2001**, *147* (2), 419−429.
(22) Rasmussen, M.; Björck, L. Unique regulation of SclB - a novel collagen-like surface protein of *Streptococcus pyogenes*. *Mol. Microbiol.* **2001**, *40* (6), 1427−38.
(23) Rasmussen, M.; Edén, A.; Björck, L. SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infect. Immun.* **2000**, *68* (11), 6370−6377.
(24) Lukomski, S.; Nakashima, K.; Abdi, I.; Cipriano, V. J.; Ireland, R. M.; Reid, S. D.; Adams, G. G.; Musser, J. M. Identification and characterization of the *scl* gene encoding a group A *Streptococcus* extracellular protein virulence factor with similarity to human collagen. *Infect. Immun.* **2000**, *68* (12), 6542−53.
(25) Pizarro-Guajardo, M.; Olguin-Araneda, V.; Barra-Carrasco, J.; Brito-Silva, C.; Sarker, M. R.; Paredes-Sabja, D. Characterization of the collagen-like exosporium protein, BclA1, of *Clostridium difficile* spores. *Anaerobe* **2014**, *25*, 18−30.
(26) Sylvestre, P.; Couture-Tosi, E.; Mock, M. A collagen-like surface glycoprotein is a structural component of the *Bacillus anthracis* exosporium. *Mol. Microbiol.* **2002**, *45* (1), 169−178.
(27) Paterson, G. K.; Nieminen, L.; Jefferies, J. M. C.; Mitchell, T. J. PclA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells. *FEMS Microbiol. Lett.* **2008**, *285* (2), 170−176.
(28) Duncan, C.; Prashar, A.; So, J.; Tang, P.; Low, D.; Terebiznik, M.; Guyard, C. Lcl of *Legionella pneumophila* is an immunogenic GAG binding adhesin that promotes interactions with lung epithelial cells and plays a crucial role in biofilm formation. *Infect. Immun.* **2011**, *79*, 2168−81.
(29) Bachert, B. A.; Choi, S. J.; Snyder, A. K.; Rio, R. V. M.; Durney, B. C.; Holland, L. A.; Amemiya, K.; Welkos, S. L.; Bozue, J. A.; Cote, C. K.; Berisio, R.; Lukomski, S. A unique set of the *Burkholderia* collagen-like proteins provides insight into pathogenesis, genome evolution and niche adaptation, and infection detection. *PLoS One* **2015**, *10* (9), e0137578.
(30) Zhao, X.; Wang, Y.; Shang, Q.; Li, Y.; Hao, H.; Zhang, Y.; Guo, Z.; Yang, G.; Xie, Z.; Wang, R. Collagen-Like proteins (ClpA, ClpB, ClpC, and ClpD) are required for biofilm formation and adhesion to plant roots by *Bacillus amyloliquefaciens* FZB42. *PLoS One* **2015**, *10* (2), e0117414−e0117414.

(31) Koonin, E. V.; Makarova, K. S.; Aravind, L. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.* **2001**, *55*, 709−42.
(32) Lukomski, S.; Bachert, B. A.; Squeglia, F.; Berisio, R. Collagen-like proteins of pathogenic streptococci. *Mol. Microbiol.* **2017**, *103* (6), 919−930.
(33) McElroy, K.; Mouton, L.; Du Pasquier, L.; Qi, W.; Ebert, D. Characterisation of a large family of polymorphic collagen-like proteins in the endospore-forming bacterium *Pasteuria ramosa*. *Res. Microbiol.* **2011**, *162* (7), 701−714.
(34) Doxey, A.; McConkey, B. Prediction of molecular mimicry candidates in human pathogenic bacteria. *Virulence* **2013**, *4*, 453.
(35) Singh, S. K.; Tripathi, T. Presence of collagen-like repeats in bacteriophage−encoded hyaluronate lyase. *J. Sci. Res. Banaras Hindu Univ.* **2020**, *64* (01), 193−197.
(36) Ghosh, N.; McKillop, T. J.; Jowitt, T. A.; Howard, M.; Davies, H.; Holmes, D. F.; Roberts, I. S.; Bella, J. Collagen-like proteins in pathogenic *E. coli* strains. *PLoS One* **2012**, *7*, e37872.
(37) Kananaviciute, R.; Kvederaviciute, K.; Dabkeviciene, D.; Mackevicius, G.; Kuisiene, N. Collagen-like sequences encoded by extremophilic and extremotolerant bacteria. *Genomics* **2020**, *112*, 2271.
(38) Bella, J. A first census of collagen interruptions: collagen's own stutters and stammers. *J. Struct. Biol.* **2014**, *186* (3), 438−50.
(39) Han, R.; Zwiefka, A.; Caswell, C.; Xu, Y.; Keene, D.; Lukomska, E.; Zhao, Z.; Höök, M.; Lukomski, S. Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers. *Appl. Microbiol. Biotechnol.* **2006**, *72*, 109−15.
(40) Persikov, A. V.; Ramshaw, J. A.; Brodsky, B. Prediction of collagen stability from amino acid sequence. *J. Biol. Chem.* **2005**, *280* (19), 19343−19349.
(41) Réty, S.; Salamitou, S.; Garcia-Verdugo, I.; Hulmes, D. J.; Le Hégarat, F.; Chaby, R.; Lewit-Bentley, A. The crystal structure of the *Bacillus anthracis* spore surface protein BclA shows remarkable similarity to mammalian proteins. *J. Biol. Chem.* **2005**, *280* (52), 43073−8.
(42) Henriques, A. O.; Moran, C. P., Jr. Structure, assembly, and function of the spore surface layers. *Annu. Rev. Microbiol.* **2007**, *61*, 555−88.
(43) Mohs, A.; Silva, T.; Yoshida, T.; Amin, R.; Lukomski, S.; Inouye, M.; Brodsky, B. Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J. Biol. Chem.* **2007**, *282* (41), 29757−29765.
(44) Lequette, Y.; Garénaux, E.; Combrouse, T.; Lima Dias, T. D.; Ronse, A.; Slomianny, C.; Trivelli, X.; Guerardel, Y.; Faille, C. Domains of BclA, the major surface glycoprotein of the *B. cereus* exosporium: glycosylation patterns and role in spore surface properties. *Biofouling* **2011**, *27* (7), 751−761.
(45) Daubenspeck, J. M.; Zeng, H.; Chen, P.; Dong, S.; Steichen, C. T.; Krishna, N. R.; Pritchard, D. G.; Turnbough, C. L. Novel oligosaccharide side chains of the collagen-like region of BclA, the major glycoprotein of the *Bacillus anthracis* exosporium. *J. Biol. Chem.* **2004**, *279* (30), 30945−30953.
(46) Karlström, Å.; Jacobsson, K.; Flock, M.; Flock, J.-I.; Guss, B. Identification of a novel collagen-like protein, SclC, in *Streptococcus equi* using signal sequence phage display. *Vet. Microbiol.* **2004**, *104* (3), 179−188.
(47) Kirchdoerfer, Robert N.; Herrin; Brantley, R.; Han; Byung, W.; Turnbough; Charles, L.; Cooper; Max, D.; Wilson; Ian, A. Variable lymphocyte receptor recognition of the immunodominant glycoprotein of *Bacillus anthracis* spores. *Structure* **2012**, *20* (3), 479−486.
(48) Caswell, C. C.; Lukomska, E.; Seo, N.-S.; Höök, M.; Lukomski, S. Scl1-dependent internalization of group A *Streptococcus* via direct interactions with the $\alpha_2\beta_1$ integrin enhances pathogen survival and re-emergence. *Mol. Microbiol.* **2007**, *64* (5), 1319−1331.
(49) Bhandari, V.; Ahmod, N. Z.; Shah, H. N.; Gupta, R. S. Molecular signatures for *Bacillus* species: Demarcation of the *Bacillus subtilis* and *Bacillus cereus* clades in molecular terms and proposal to limit the placement of new species into the genus *Bacillus*. *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 2712−2726.
(50) McKenney, P. T.; Driks, A.; Eichenberger, P. The *Bacillus subtilis* endospore: Assembly and functions of the multilayered coat. *Nat. Rev. Microbiol.* **2013**, *11* (1), 33−44.
(51) Qin, H.; Driks, A. Contrasting evolutionary patterns of spore coat proteins in two *Bacillus* species groups are linked to a difference in cellular structure. *BMC Evol. Biol.* **2013**, *13*, 261.
(52) Stewart, G. C. The exosporium layer of bacterial spores: a connection to the environment and the infected host. *Microbiol. Mol. Biol. Rev.* **2015**, *79* (4), 437−457.
(53) Kordzikowska, J. Investigation of a ubiquitous spore surface protein. *Master's Thesis*. 2016.
(54) Sylvestre, P.; Couture-Tosi, E.; Mock, M. l. Polymorphism in the collagen-like region of the *Bacillus anthracis* BclA protein leads to variation in exosporium filament length. *J. Bacteriol.* **2003**, *185* (5), 1555−1563.
(55) Steichen, C.; Chen, P.; Kearney, J. F.; Turnbough, C. L., Jr. Identification of the immunodominant protein and other proteins of the *Bacillus anthracis* exosporium. *J. Bacteriol.* **2003**, *185* (6), 1903−10.
(56) Zhao, N.; Ge, Y.; Shi, T. Y.; Hu, X. M.; Yuan, Z. M. Collagen-like glycoprotein BclS Is involved in the formation of filamentous structures of the *Lysinibacillus sphaericus* exosporium. *Appl. Environ. Microbiol.* **2014**, *80* (21), 6656−6663.
(57) Fazzini, M. M.; Schuch, R.; Fischetti, V. A. A novel spore protein, ExsM, regulates formation of the exosporium in *Bacillus cereus* and *Bacillus anthracis* and affects spore size and shape. *J. Bacteriol.* **2010**, *192* (15), 4012−4021.
(58) Ball, D. A.; Taylor, R.; Todd, S. J.; Redmond, C.; Couture-Tosi, E.; Sylvestre, P.; Moir, A.; Bullough, P. A. Structure of the exosporium and sublayers of spores of the *Bacillus cereus* family revealed by electron crystallography. *Mol. Microbiol.* **2008**, *68* (4), 947−958.
(59) Manetsberger, J.; Hall, E. A. H.; Christie, G. Plasmid-encoded genes influence exosporium assembly and morphology in *Bacillus megaterium* QM B1551 spores. *FEMS Microbiol. Lett.* **2015**, *362* (18), fnv147.
(60) Leski, T. A.; Caswell, C. C.; Pawlowski, M.; Klinke, D. J.; Bujnicki, J. M.; Hart, S. J.; Lukomski, S. Identification and classification of *bcl* genes and proteins of *Bacillus cereus* group organisms and their application in *Bacillus anthracis* detection and fingerprinting. *Appl. Environ. Microbiol.* **2009**, *75* (22), 7163−7172.
(61) Thompson, B. M.; Stewart, G. C. Targeting of the BclA and BclB proteins to the *Bacillus anthracis* spore surface. *Mol. Microbiol.* **2008**, *70* (2), 421−434.
(62) Thompson, B. M.; Waller, L. N.; Fox, K. F.; Fox, A.; Stewart, G. C. The BclB glycoprotein of *Bacillus anthracis* is involved in exosporium integrity. *J. Bacteriol.* **2007**, *189* (18), 6704−6713.
(63) Bozue, J.; Moody, K. L.; Cote, C. K.; Stiles, B. G.; Friedlander, A. M.; Welkos, S. L.; Hale, M. L. *Bacillus anthracis* spores of the *bclA* mutant exhibit increased adherence to epithelial cells, fibroblasts, and endothelial cells but not to macrophages. *Infect. Immun.* **2007**, *75* (9), 4498−505.
(64) Xue, Q.; Gu, C.; Rivera, J.; Hook, M.; Chen, X.; Pozzi, A.; Xu, Y. Entry of *Bacillus anthracis* spores into epithelial cells is mediated by the spore surface protein BclA, integrin $\alpha2\beta1$ and complement component C1q. *Cell. Microbiol.* **2011**, *13* (4), 620−34.
(65) Oliva, C.; Turnbough, C. L., Jr.; Kearney, J. F. CD14-Mac-1 interactions in *Bacillus anthracis* spore internalization by macrophages. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (33), 13957−62.
(66) Oliva, C. R.; Swiecki, M. K.; Griguer, C. E.; Lisanby, M. W.; Bullard, D. C.; Turnbough, C. L., Jr.; Kearney, J. F. The integrin Mac-1 (CR3) mediates internalization and directs *Bacillus anthracis* spores into professional phagocytes. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (4), 1261−6.
(67) Todd, S. J.; Moir, A. J. G.; Johnson, M. J.; Moir, A. Genes of *Bacillus cereus* and *Bacillus anthracis* encoding proteins of the exosporium. *J. Bacteriol.* **2003**, *185* (11), 3373−3378.

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

(68) Fan, B.; Wang, C.; Song, X.; Ding, X.; Wu, L.; Wu, H.; Gao, X.; Borriss, R. *Bacillus velezensis* FZB42 in 2018: The gram-positive model strain for plant growth promotion and biocontrol. *Front. Microbiol.* **2018**, DOI: 10.3389/fmicb.2018.02491.
(69) Lyu, D.; Backer, R.; Robinson, W. G.; Smith, D. L. Plant growth-promoting rhizobacteria for cannabis production: yield, cannabinoid profile and disease resistance. *Front. Microbiol.* **2019**, DOI: 10.3389/fmicb.2019.01761.
(70) Fan, B.; Chen, X. H.; Budiharjo, A.; Bleiss, W.; Vater, J.; Borriss, R. Efficient colonization of plant roots by the plant growth promoting bacterium *Bacillus amyloliquefaciens* FZB42, engineered to express green fluorescent protein. *J. Biotechnol.* **2011**, *151* (4), 303−311.
(71) Ahmad, I.; Khan, M. S.; Altaf, M. M.; Qais, F. A.; Ansari, F. A.; Rumbaugh, K. P. Biofilms: An overview of their significance in plant and soil health.*Biofilms in Plant and Soil Health*; Wiley, 2017; Chapter 1, pp 1−25, .
(72) Zhao, X.; Wang, R.; Shang, Q.; Hao, H.; Li, Y.; Zhang, Y.; Guo, Z.; Wang, Y.; Xie, Z. The new flagella-associated collagen-like proteins ClpB and ClpC of *Bacillus amyloliquefaciens* FZB42 are involved in bacterial motility. *Microbiol. Res.* **2016**, *184*, 25−31.
(73) Manetsberger, J.; Ghosh, A.; Hall, E. A. H.; Christie, G. Orthologues of *Bacillus subtilis* spore crust proteins have a structural role in the *Bacillus megaterium* QM B1551 spore exosporium. *Appl. Environ. Microbiol.* **2018**, *84* (20), 1−14.
(74) Thompson, B. M.; Hoelscher, B. C.; Driks, A.; Stewart, G. C. Assembly of the BclB glycoprotein into the exosporium and evidence for its role in the formation of the exosporium 'cap' structure in *Bacillus anthracis*. *Mol. Microbiol.* **2012**, *86* (5), 1073−1084.
(75) Xu, K.; Yuan, Z.; Rayner, S.; Hu, X. Genome comparison provides molecular insights into the phylogeny of the reassigned new genus *Lysinibacillus*. *BMC Genomics* **2015**, *16*, 140.
(76) Reinert, D. J.; Carpusca, I.; Aktories, K.; Schulz, G. E. Structure of the mosquitocidal toxin from *Bacillus sphaericus*. *J. Mol. Biol.* **2006**, *357* (4), 1226−36.
(77) Park, H.-W.; Bideshi, D. K.; Federici, B. A. Properties and applied use of the mosquitocidal bacterium, *Bacillus sphaericus*. *J. Asia-Pac. Entomol.* **2010**, *13* (3), 159−168.
(78) Bishop, A. H. *Pasteuria penetrans* and its parasitic interaction with plant parasitic nematodes. In *Endospore-forming soil bacteria*; Logan, N. A., Vos, P., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; pp 181−201. DOI: 10.1007/978-3-642-19577-8_9.
(79) Duneau, D.; Luijckx, P.; Ben-Ami, F.; Laforsch, C.; Ebert, D. Resolving the infection process reveals striking differences in the contribution of environment, genetics and phylogeny to host-parasite interactions. *BMC Biol.* **2011**, *9* (1), 11.
(80) Davies, K. G. Chapter 9 Understanding the interaction between an obligate hyperparasitic bacterium, *Pasteuria penetrans* and its obligate plant-parasitic nematode host, *Meloidogyne* spp. *Adv. Parasitol.* **2009**, *68*, 211−245.
(81) Mouton, L.; Traunecker, E.; McElroy, K.; Du Pasquier, L.; Ebert, D. Identification of a polymorphic collagen-like protein in the crustacean bacteria *Pasteuria ramosa*. *Res. Microbiol.* **2009**, *160* (10), 792−799.
(82) Srivastava, A.; Mohan, S.; Mauchline, T. H.; Davies, K. G. Evidence for diversifying selection of genetic regions of encoding putative collagen-like host-adhesive fibers in *Pasteuria penetrans*. *FEMS Microbiol. Ecol.* **2019**, *95*, fiy217.
(83) Andras, J. P.; Fields, P. D.; Du Pasquier, L.; Fredericksen, M.; Ebert, D. Genome-wide association analysis identifies a genetic basis of infectivity in a model bacterial pathogen. *Mol. Biol. Evol.* **2020**, *37*, 3439.
(84) Rupnik, M.; Wilcox, M. H.; Gerding, D. N. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat. Rev. Microbiol.* **2009**, *7* (7), 526−536.
(85) Strong, P. C. R.; Fulton, K. M.; Aubry, A.; Foote, S.; Twine, S. M.; Logan, S. M. Identification and characterization of glycoproteins on the spore surface of *Clostridium difficile*. *J. Bacteriol.* **2014**, *196* (14), 2627−2637.
(86) Driks, A. Surface appendages of bacterial spores. *Mol. Microbiol.* **2007**, *63* (3), 623−625.
(87) Walker, J. R.; Gnanam, A. J.; Blinkova, A. L.; Hermandson, M. J.; Karymov, M. A.; Lyubchenko, Y. L.; Graves, P. R.; Haystead, T. A.; Linse, K. D. *Clostridium taeniosporum* spore ribbon-like appendage structure, composition and genes. *Mol. Microbiol.* **2007**, *63* (3), 629−643.
(88) Cambridge, J. M.; Blinkova, A. L.; Salvador Rocha, E. I.; Bode Hernández, A.; Moreno, M.; Ginés-Candelaria, E.; Goetz, B. M.; Hunicke-Smith, S.; Satterwhite, E.; Tucker, H. O.; Walker, J. R. Genomics of *Clostridium taeniosporum*, an organism which forms endospores with ribbon-like appendages. *PLoS One* **2018**, *13* (1), e0189673.
(89) Parker, M. D.; Barrett, P.; Shepherd, J.; Price, L.; Bull, S. Characterisation of non-toxigenic *Clostridium* spp. strains, to use as surrogates for non-proteolytic *Clostridium botulinum* in chilled food challenge testing. *J. Microbiol. Methods* **2015**, *108*, 83−91.
(90) Janganan, T. K.; Mullin, N.; Tzokov, S. B.; Stringer, S.; Fagan, R. P.; Hobbs, J. K.; Moir, A.; Bullough, P. A. Characterization of the spore surface and exosporium proteins of *Clostridium sporogenes*; implications for *Clostridium botulinum* group I strains. *Food Microbiol.* **2016**, *59*, 205−12.
(91) Oliver-Kozup, H. A.; Elliott, M.; Bachert, B. A.; Martin, K. H.; Reid, S. D.; Schwegler-Berry, D. E.; Green, B. J.; Lukomski, S. The streptococcal collagen-like protein-1 (Scl1) is a significant determinant for biofilm formation by group A *Streptococcus*. *BMC Microbiol.* **2011**, *11* (1), 262−262.
(92) Castanha, E. R.; Swiger, R. R.; Senior, B.; Fox, A.; Waller, L. N.; Fox, K. F. Strain discrimination among *B. anthracis* and related organisms by characterization of *bclA* polymorphisms using PCR coupled with agarose gel or microchannel fluidics electrophoresis. *J. Microbiol. Methods* **2006**, *64* (1), 27−45.
(93) Abdel-Nour, M.; Su, H.; Duncan, C.; Li, S.; Raju, D.; Shamoun, F.; Valton, M.; Ginevra, C.; Jarraud, S.; Guyard, C.; Kerman, K.; Terebiznik, M. R. Polymorphisms of a collagen-like adhesin contributes to *Legionella pneumophila* adhesion, biofilm formation capacity and clinical prevalence. *Front. Microbiol.* **2019**, *10*, 604.
(94) Yu, Z.; Brodsky, B.; Inouye, M. Dissecting a bacterial collagen domain from *Streptococcus pyogenes*: sequence and length-dependent variations in triple helix stability and folding. *J. Biol. Chem.* **2011**, *286* (21), 18960−8.
(95) Kar, K.; Amin, P.; Bryan, M. A.; Persikov, A. V.; Mohs, A.; Wang, Y.-H.; Brodsky, B. Self-association of collagen triple helic peptides into higher order structures. *J. Biol. Chem.* **2006**, *281* (44), 33283−33290.
(96) Abdel-Nour, M.; Duncan, C.; Low, D. E.; Guyard, C. Biofilms: the stronghold of *Legionella pneumophila*. *Int. J. Mol. Sci.* **2013**, *14* (11), 21660−75.
(97) Bédard, E.; Lévesque, S.; Martin, P.; Pinsonneault, L.; Paranjape, K.; Lalancette, C.; Dolcé, C.-É.; Villion, M.; Valiquette, L.; Faucher, S.; Prevost, M. Energy conservation and the promotion of *Legionella pneumophila* growth: the probable role of heat exchangers in a nosocomial outbreak. *Infect Control Hosp Epidemiol* **2016**, *37*, 1475.
(98) Peng, Y. Y.; Yoshizumi, A.; Danon, S. J.; Glattauer, V.; Prokopenko, O.; Mirochnitchenko, O.; Yu, Z.; Inouye, M.; Werkmeister, J. A.; Brodsky, B.; Ramshaw, J. A. M. A *Streptococcus pyogenes* derived collagen-like protein as a non-cytotoxic and non-immunogenic cross-linkable biomaterial. *Biomaterials* **2010**, *31* (10), 2755−2761.
(99) An, B.; Kaplan, D. L.; Brodsky, B. Engineered recombinant bacterial collagen as an alternative collagen-based biomaterial for tissue engineering. *Front. Chem.* **2014**, *2*, 40.
(100) Peng, Y. Y.; Stoichevska, V.; Schacht, K.; Werkmeister, J. A.; Ramshaw, J. A. M. Engineering multiple biological functional motifs into a blank collagen-like protein template from *Streptococcus pyogenes*. *J. Biomed. Mater. Res., Part A* **2014**, *102* (7), 2189−2196.

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

(101) Ramshaw, J. A. M.; Werkmeister, J. A.; Glattauer, V. Recent progress with recombinant collagens produced in *Escherichia coli*. *Curr. Opin. Biomed. Eng.* **2019**, *10*, 149−155.
(102) Peng, Y. Y.; Stoichevska, V.; Howell, L.; Madsen, S.; Werkmeister, J. A.; Dumsday, G. J.; Ramshaw, J. A. M. Preparation and characterization of monomers to tetramers of a collagen-like domain from *Streptococcus pyogenes*. *Bioengineered* **2014**, *5* (6), 378−385.
(103) Golser, A. V.; Scheibel, T. Routes towards novel collagen-like biomaterials. *Fibers* **2018**, *6* (2), 21.
(104) Mooney, S. D.; Kollman, P. A.; Klein, T. E. Conformational preferences of substituted prolines in the collagen triple helix. *Biopolymers* **2002**, *64* (2), 63−71.
(105) Berisio, R.; Vitagliano, L. Polyproline and triple helix motifs in host-pathogen recognition. *Curr. Protein Pept. Sci.* **2012**, *13* (8), 855−865.
(106) Yoshizumi, A.; Fletcher, J. M.; Yu, Z.; Persikov, A. V.; Bartlett, G. J.; Boyle, A. L.; Vincent, T. L.; Woolfson, D. N.; Brodsky, B. Designed coiled coils promote folding of a recombinant bacterial collagen. *J. Biol. Chem.* **2011**, *286* (20), 17512−17520.
(107) Squeglia, F.; Bachert, B.; De Simone, A.; Lukomski, S.; Berisio, R. The crystal structure of the streptococcal collagen-like protein 2 globular domain from invasive M3-type group A *Streptococcus* shows significant similarity to immunomodulatory HIV protein gp41. *J. Biol. Chem.* **2014**, *289* (8), 5122−5133.
(108) Yigit, S.; Yu, H.; An, B.; Hamaia, S.; Farndale, R. W.; Kaplan, D. L.; Lin, Y.-S.; Brodsky, B. Mapping the effect of Gly mutations in collagen on $\alpha 2\beta 1$ integrin binding. *J. Biol. Chem.* **2016**, *291* (36), 19196−19207.
(109) Qiu, Y.; Mekkat, A.; Yu, H.; Yigit, S.; Hamaia, S.; Farndale, R. W.; Kaplan, D. L.; Lin, Y.-S.; Brodsky, B. Collagen Gly missense mutations: Effect of residue identity on collagen structure and integrin binding. *J. Struct. Biol.* **2018**, *203* (3), 255−262.
(110) Chhum, P.; Yu, H.; An, B.; Doyon, B. R.; Lin, Y.-S.; Brodsky, B. Consequences of glycine mutations in the fibronectin-binding sequence of collagen. *J. Biol. Chem.* **2016**, *291* (53), 27073−27086.
(111) Yeo, J.; Qiu, Y.; Jung, G. S.; Zhang, Y.-W.; Buehler, M. J.; Kaplan, D. L. Adverse effects of Alport syndrome-related Gly missense mutations on collagen type IV: Insights from molecular simulations and experiments. *Biomaterials* **2020**, *240*, 119857.
(112) Zhai, C.; Zhou, H.; Gao, T.; Zhao, L.; Lin, S. Electrostatically tuned microdomain morphology and phase-dependent ion transport anisotropy in single-ion conducting block copolyelectrolytes. *Macromolecules* **2018**, *51* (12), 4471−4483.
(113) Lin, S.; Cai, Z.; Wang, Y.; Zhao, L.; Zhai, C. Tailored morphology and highly enhanced phonon transport in polymer fibers: A multiscale computational framework. *NPJ. Comput. Mater.* **2019**, *5* (1), 126.
(114) Prhashanna, A.; Taylor, P. A.; Qin, J.; Kiick, K. L.; Jayaraman, A. Effect of peptide sequence on the LCST-like transition of elastin-like peptides and elastin-like peptide−collagen-like peptide conjugates: Simulations and experiments. *Biomacromolecules* **2019**, *20* (3), 1178−1189.
(115) Martín, M. J.; Lara-Villoslada, F.; Ruiz, M. A.; Morales, M. E. Microencapsulation of bacteria: A review of different technologies and their impact on the probiotic effects. *Innovative Food Sci. Emerging Technol.* **2015**, *27*, 15−25.
(116) Su, R.; Zhu, X.-L.; Fan, D.-D.; Mi, Y.; Yang, C.-Y.; Jia, X. Encapsulation of probiotic *Bifidobacterium longum* BIOMA 5920 with alginate−human-like collagen and evaluation of survival in simulated gastrointestinal conditions. *Int. J. Biol. Macromol.* **2011**, *49* (5), 979−984.
(117) Madani, A.; Garakani, K.; Mofrad, M. R. Molecular mechanics of *Staphylococcus aureus* adhesin, CNA, and the inhibition of bacterial adhesion by stretching collagen. *PLoS One* **2017**, *12* (6), e0179601.
(118) Zong, Y.; Xu, Y.; Liang, X.; Keene, D. R.; Höök, A.; Gurusiddappa, S.; Höök, M.; Narayana, S. V. A 'Collagen Hug'model for *Staphylococcus aureus* CNA binding to collagen. *EMBO J.* **2005**, *24* (24), 4224−4236.
(119) Vitagliano, L.; Berisio, R.; De Simone, A. Role of hydration in collagen recognition by bacterial adhesins. *Biophys. J.* **2011**, *100* (9), 2253−2261.
(120) Ross, C. L.; Liang, X.; Liu, Q.; Murray, B. E.; Höök, M.; Ganesh, V. K. Targeted protein engineering provides insights into binding mechanism and affinities of bacterial collagen adhesins. *J. Biol. Chem.* **2012**, *287* (41), 34856−34865.
(121) An, B.; Lin, Y.-S.; Brodsky, B. Collagen interactions: Drug design and delivery. *Adv. Drug Delivery Rev.* **2016**, *97*, 69−84.
(122) Hadi, M.; Sander, E.; Ruberti, J.; Barocas, V. H. Simulated remodeling of loaded collagen networks via strain-dependent enzymatic degradation and constant-rate fiber growth. *Mech. Mater.* **2012**, *44*, 72−82.
(123) Huang, S.; Huang, H.-Y. S. Biaxial stress relaxation of semilunar heart valve leaflets during simulated collagen catabolism: Effects of collagenase concentration and equibiaxial strain state. *Proc. Inst. Mech. Eng., Part H* **2015**, *229* (10), 721−731.
(124) Yi, E.; Sato, S.; Takahashi, A.; Parameswaran, H.; Blute, T. A.; Bartolák-Suki, E.; Suki, B. Mechanical forces accelerate collagen digestion by bacterial collagenase in lung tissue strips. *Front. Physiol.* **2016**, *7*, 287.
(125) Chang, S.-W.; Buehler, M. J. Molecular biomechanics of collagen molecules. *Mater. Today* **2014**, *17* (2), 70−76.
(126) Zhai, C.; Li, T.; Shi, H.; Yeo, J. Discovery and design of soft polymeric bio-inspired materials with multiscale simulations and artificial intelligence. *J. Mater. Chem. B* **2020**, *8*, 6562.
(127) Lannergård, J.; Frykberg, L.; Guss, B. CNE, a collagen-binding protein of *Streptococcus equi*. *FEMS Microbiol. Lett.* **2003**, *222* (1), 69−74.
(128) Swamy, R. N.; Gnanamani, A.; Shanmugasamy, S.; Gopal, R. K.; Mandal, A. B. Bioinformatics in crosslinking chemistry of collagen with selective cross linkers. *BMC Res. Notes* **2011**, *4* (1), 399.
(129) Wang, M.; Feng, Z.; Zhai, C.; Zhou, Q.; Wei, T.; Liu, J. Chromium carbide micro-whiskers: Preparation and strengthening effects in extreme conditions with experiments and molecular dynamics simulations. *J. Solid State Chem.* **2020**, *291*, 121598.
(130) Wang, M.; Feng, Z.; Zhai, C.; Zhang, J.; Li, Z.; Zhang, H.; Xu, X. Low-temperature in-situ grown mullite whiskers toughened heat-resistant inorganic adhesive. *J. Alloys Compd.* **2020**, *836*, 155349.
(131) Xu, X.; Wang, M.; Xie, J.; Zhai, C.; Li, H.; Su, X.; Wang, X. Cooperative effects of free volume and nanocrystallite on plastic deformation in bulk amorphous Al2O3−ZrO2−Y2O3. *Ceram. Int.* **2020**, *46*, 11049.
(132) Xu, X.; Zhai, C.; Ren, H.; Guo, L.; Hou, F.; Liu, J. Enhanced plasticity by deformation-induced nanocrysallization in bulk amorphous Al2O3-ZrO2-Y2O3. *Ceram. Int.* **2017**, *43* (3), 3453−3456.
(133) Zhai, C.; Zhong, Y.; Li, Z.; Wang, X.; Zhao, L.; Pan, L.; Zhang, J. Preparation and characterization of mechanical properties of foam glass for artificial floating island carrier. *Adv. Mech. Eng.* **2015**, *7* (6), 168781401558966.
(134) Zhai, C.; Li, Z.; Zhu, Y.; Zhang, J.; Wang, X.; Zhao, L.; Pan, L.; Wang, P. Effects of Sb2O3 on the mechanical properties of the borosilicate foam glasses sintered at low temperature. *Adv. Mater. Sci. Eng.* **2014**, *2014*, 703194.
(135) Xia, P.; Jiang, R.; Li, Z.; Zhu, Y.; Zhai, C.; Feng, D.; Sun, P. Effect of Y2O3 on the properties of vitrified bond and vitrified diamond composites. *Composites, Part B* **2014**, *67*, 515−520.
(136) Feng, D.; Wu, W.; Wang, P.; Zhu, Y.; Zhai, C.; Li, Z. Effects of Cu on properties of vitrified bond and vitrified CBN composites. *Int. J. Refract. Hard Met.* **2015**, *50*, 269−273.
(137) Zhai, C. Understanding slow dynamics in polymers using coarse-grained molecular dynamics simulations: Microphase separation in block copolyelectrolytes and photo-switching in azobenzene-polymers. *Thesis*. The Florida State University, 2020.
(138) Mitra, T.; Sailakshmi, G.; Gnanamani, A.; Mandal, A. Studies on cross-linking of succinic acid with chitosan/collagen. *Mater. Res.* **2013**, *16* (4), 755−765.
(139) Du, J.; Hincke, M. T.; Rose-Martel, M.; Hennequet-Antier, C.; Brionne, A.; Cogburn, L. A.; Nys, Y.; Gautron, J. Identifying specific

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX

proteins involved in eggshell membrane formation using gene expression analysis and bioinformatics. *BMC Genomics* **2015**, *16* (1), 792.
(140) Rehman, S.; Salway, F.; Stanley, J. K.; Ollier, W. E.; Day, P.; Bayat, A. Molecular phenotypic descriptors of Dupuytren's disease defined using informatics analysis of the transcriptome. *J. Hand Surg* **2008**, *33* (3), 359−372.
(141) Huang, J.; Luo, G.; Zhang, Z.; Wang, X.; Ju, Z.; Qi, C.; Zhang, Y.; Wang, C.; Li, R.; Li, J. iTRAQ-proteomics and bioinformatics analyses of mammary tissue from cows with clinical mastitis due to natural infection with *Staphylococci aureus*. *BMC Genomics* **2014**, *15* (1), 839.
(142) Rong, L.; Huang, W.; Tian, S.; Chi, X.; Zhao, P.; Liu, F. COL1A2 is a novel biomarker to improve clinical prediction in human gastric cancer: Integrating bioinformatics and meta-analysis. *Pathol. Oncol. Res.* **2018**, *24* (1), 129−134.
(143) Huang, R.; Gu, W.; Sun, B.; Gao, L. Identification of COL4A1 as a potential gene conferring trastuzumab resistance in gastric cancer based on bioinformatics analysis. *Mol. Med. Rep.* **2018**, *17* (5), 6387−6396.
(144) Cao, L.; Chen, Y.; Zhang, M.; Xu, D.-q.; Liu, Y.; Liu, T.; Liu, S.-x.; Wang, P. Identification of hub genes and potential molecular mechanisms in gastric cancer by integrated bioinformatics analysis. *PeerJ* **2018**, *6*, e5180.
(145) Li, Z.; Liu, Z.; Shao, Z.; Li, C.; Li, Y.; Liu, Q.; Zhang, Y.; Tan, B.; Liu, Y. Identifying multiple collagen gene family members as potential gastric cancer biomarkers using integrated bioinformatics analysis. *PeerJ* **2020**, *8*, e9123.
(146) Zhang, Q.-N.; Zhu, H.-L.; Xia, M.-T.; Liao, J.; Huang, X.-T.; Xiao, J.-W.; Yuan, C. A panel of collagen genes are associated with prognosis of patients with gastric cancer and regulated by microRNA-29c-3p: An integrated bioinformatics analysis and experimental validation. *Cancer Manage. Res.* **2019**, *11*, 4757.
(147) Liang, Y.; Lv, Z.; Huang, G.; Qin, J.; Li, H.; Nong, F.; Wen, B. Prognostic significance of abnormal matrix collagen remodeling in colorectal cancer based on histologic and bioinformatics analysis. *Oncol. Rep.* **2020**, *44* (4), 1671−1685.
(148) Yang, C. T.; Ghosh, D.; Beaudry, F. Detection of gelatin adulteration using bio-informatics, proteomics and high-resolution mass spectrometry. *Food Addit. Contam., Part A* **2018**, *35* (4), 599−608.
(149) Wang, Y.; Li, Y. Analysis of molecular pathways in pancreatic ductal adenocarcinomas with a bioinformatics approach. *Asian Pac J. Cancer Prev* **2015**, *16* (6), 2561−7.
(150) Sun, M.; Yu, H.; Zhang, Y.; Li, Z.; Gao, W. MicroRNA-214 mediates isoproterenol-induced proliferation and collagen synthesis in cardiac fibroblasts. *Sci. Rep.* **2015**, *5*, 18351.
(151) Ahmad, S. S.; Khalid, M.; Younis, K. Interaction study of dietary fibers (pectin and cellulose) with meat proteins using bioinformatics analysis: An In-Silico study. *LWT* **2020**, *119*, 108889.
(152) Nuñez, S. M.; Guzmán, F.; Valencia, P.; Almonacid, S.; Cárdenas, C. Collagen as a source of bioactive peptides: A bioinformatics approach. *Electron. J. Biotechnol.* **2020**, *48*, 101.
(153) De Chaumont, F.; Dallongeville, S.; Chenouard, N.; Hervé, N.; Pop, S.; Provoost, T.; Meas-Yedid, V.; Pankajakshan, P.; Lecomte, T.; Le Montagner, Y. ICY: an open bioimage informatics platform for extended reproducible research. *Nat. Methods* **2012**, *9* (7), 690−696.
(154) Karlström, A.; Jacobsson, K.; Guss, B. SclC is a member of a novel family of collagen-like proteins in *Streptococcus equi* subspecies *equi* that are recognised by antibodies against SclC. *Vet. Microbiol.* **2006**, *114* (1−2), 72−81.
(155) Costa-Hurtado, M.; Garcia-Rodriguez, L.; Lopez-Serrano, S.; Aragon, V. *Haemophilus parasuis* VtaA2 is involved in adhesion to extracellular proteins. *Vet. Res.* **2019**, *50* (1), 69.
(156) Park, T. J.; Heo, N. S.; Yim, S. S.; Park, J. H.; Jeong, K. J.; Lee, S. Y. Surface display of recombinant proteins on Escherichia coli by BclA exosporium of *Bacillus anthracis*. *Microb. Cell Fact.* **2013**, *12* (1), 81.

https://doi.org/10.1021/acsbiomaterials.1c00018
ACS Biomater. Sci. Eng. XXXX, XXX, XXX−XXX