Journal of Structural Biology 186 (2014) 451–461

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Bacterial collagen-like proteins that form triple-helical structures

Zhuoxin Yu [a,1], Bo An [b], John A.M. Ramshaw [c,*], Barbara Brodsky [b]

[a] Department of Biochemistry, Robert Wood Johnson Medical School, Rutgers University, Piscataway, NJ 08854, USA
[b] Department of Biomedical Engineering, Tufts University, Medford, MA 02155, USA
[c] CSIRO Materials Science and Engineering, Bayview Avenue, Clayton, VIC 3169, Australia

ARTICLE INFO

Article history:
Available online 14 January 2014

Keywords:
Collagen
Triple-helix
Recombinant expression
Thermal stability
Prokaryote
Biomedical material

ABSTRACT

A large number of collagen-like proteins have been identified in bacteria during the past 10 years, principally from analysis of genome databases. These bacterial collagens share the distinctive Gly-Xaa-Yaa repeating amino acid sequence of animal collagens which underlies their unique triple-helical structure. A number of the bacterial collagens have been expressed in *Escherichia coli*, and they all adopt a triple-helix conformation. Unlike animal collagens, these bacterial proteins do not contain the post-translationally modified amino acid, hydroxyproline, which is known to stabilize the triple-helix structure and may promote self-assembly. Despite the absence of collagen hydroxylation, the triple-helix structures of the bacterial collagens studied exhibit a high thermal stability of 35–39 °C, close to that seen for mammalian collagens. These bacterial collagens are readily produced in large quantities by recombinant methods, either in the original amino acid sequence or in genetically manipulated sequences. This new family of recombinant, easy to modify collagens could provide a novel system for investigating structural and functional motifs in animal collagens and could also form the basis of new biomedical materials with designed structural properties and functions.

1. Discovery of bacterial collagens

Collagen is the most abundant protein in mammals, and plays a critical role in extracellular matrix structural properties and cell signaling. The defining feature of a collagen is its molecular structure, which is the unique supercoiled triple-helix. This conformation is made up of three left-handed polyproline-like chains twisted together into a right-handed triple-helix (Brodsky and Ramshaw, 1997). The tight packing of the triple-helix requires that every third residue in the primary sequence be Gly, because there is no space for any larger amino acid in the interior axis of the triple-helix. This leads to the repetitive sequence pattern $(Gly\text{-}Xaa\text{-}Yaa)_n$, which is a distinguishing feature of collagens. Another characteristic of animal collagens is the presence of a high content of Pro and, notably, a high content (>10% of residues) of the post-translationally formed hydroxyproline (Hyp) (Myllyharju, 2003). The enzyme prolyl hydroxylase hydroxylates all Pro residues in the Yaa position of the Gly-Xaa-Yaa repeat in collagens. Hyp residues make a critical contribution to the stability of the triple-helix through stereoelectronic effects (Bretscher et al., 2001) and/or hydration (Bella et al., 1994), and also appear essential for collagen self-association (Perret et al., 2001) and for some receptor interactions. Although collagens were originally thought to be found only in multicellular animals and to require the Hyp residue, it has recently been demonstrated that there are collagen-like proteins in bacteria and that they can form triple-helix structures even though they lack Hyp (Table 1).

In 2000, two proteins in the gram negative bacterium *Streptococcus pyogenes* were found to contain a substantial length of $(Gly\text{-}Xaa\text{-}Yaa)_n$ amino acid sequence and it was postulated that these form collagenous structures (Lukomski et al., 2000; Rasmussen et al., 2000). As increasing numbers of genomic sequences were being reported, an analysis was carried out on 136 eubacterial genomes (Rasmussen et al., 2003), searching for sequences with homology to $(Gly\text{-}Pro\text{-}Pro)_n$. Hits were found for 56 proteins in 25 bacterial genomes, with none seen in any of the 15 archaebacterial genomes. The number of Gly-Xaa-Yaa tripeptides varied from 7 to 745, with an average length of 76 triplets, and these collagen-like sequences are always flanked by non-collagenous domains. The collagen-like sequences from different bacteria all had a relatively high Pro content, and Rasmussen et al. (2003) found distinctive amino acid compositions for different potential proteins which could be categorized as Thr-rich, Pro-rich, or rich in charged residues. Pro was preferentially found in the X position in bacterial proteins in contrast to mammalian collagens where there are typically half or more of the Pro residues in the Yaa-position, which are subsequently hydroxylated. Conversely, in the bacterial

* Corresponding author.
E-mail addresses: zhuoxin_yu@yahoo.com (Z. Yu), Bo.An@tufts.edu (B. An), John.Ramshaw@csiro.au (J.A.M. Ramshaw), Barbara.Brodsky@tufts.edu (B. Brodsky).
[1] Present address: Brightech International, Somerset, NJ 08873, USA.

1047-8477/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2014.01.003

Table 1
Comparison of bacterial collagens which have been characterized with mammalian collagens.

		Mammalian collagens	Bacterial collagens
Similarities	Gly-Xaa-Yaa repeats	Yes	Yes
	Triple-helix	Yes	Yes
	Trypsin resistance	Yes	Yes
	Thermal stability	T_m = ~37 °C	T_m = ~35–39 °C
	Calorimetric enthalpy	Very high	Relatively high, but lower than mammalian collagen
	Trimerization domain	At either end	At either end
Differences	Length	~350 triplets for fibrillar collagens	~35–82 triplets (characterized so far)
	Hydroxyproline	Yes	No
	Heterotrimer	Some of them	Not found yet
	Sequence and stabilization	Pro/Hyp rich for stability	Strategies include electrostatic interactions; glycosylation of Thr residues; very high Pro
	Types	28 different types	Species dependent – can be many species variants
	Interruptions	In non-fibrillar types	None characterized
	Fibril formation	In the 5 fibrillar types	Non reported

collagens Thr and Gln are much more frequent in the Yaa-position than observed for mammalian collagens (Rasmussen et al., 2003). Several bacterial genomes contained multiple collagen-like sequences, up to 9 in some cases, so it is possible that heterotrimers, with two or three different chains, are formed in these cases. It has been suggested that bacterial collagen sequences arise from horizontal gene transfer from eukaryotes to bacteria (Rasmussen et al., 2003).

Clearly, in the 10 years since this initial study (Rasmussen et al., 2003), the extent of the genomic information has increased many-fold and a large number of additional genomes are available for interrogation. Further studies on several of these bacterial proteins have confirmed that they have the characteristic triple-helix structure of collagen and suggest they may play important roles in pathogenesis. These proteins, which are being recognized in increasing numbers, are no longer unexpected curiosities, but represent an opportunity for approaching basic science problems in collagen and for biomaterial applications.

2. Biological role of bacterial collagen-like proteins

While there are many bacterial species that contain collagen-like sequences in the genome (Rasmussen et al., 2003), there is evidence for their natural expression for only a few cases (Karlstrom et al., 2004, 2006). A few pathogenic bacterial systems have been well characterized and these cases suggest the collagen protein may interact with the host to assist invasion or help a pathogen evade the host immune system. The two *S. pyogenes* bacterial collagens, Scl1 and Scl2, have sequences indicating they are anchored on the cell surface and have been shown to bind to a variety of host proteins. Depending on the specific serotype, the non-collagenous V-domain of Scl1 may bind to high-density lipoprotein (HDL) (Gao et al., 2010), low-density lipoprotein (LDL) (Han et al., 2006a), factor H (Caswell et al., 2008a), complement factor H-related protein 1 (CFHR1) (Reuter et al., 2010), or the extra cellular matrix (ECM) proteins fibronectin and laminin (Caswell et al., 2009). Binding to these components may help *S. pyogenes* escape from complement-regulated phagocytosis and enhance its adherence to the macrophages and ECM. Both Scl1 and Scl2 bind to thrombin-activatable fibrinolysis inhibitor (TAFI, procarboxypeptidase) and recruit it to *S. pyogenes* cell surface, counteracting the host response through regulating the proteolysis by activated TAFI (Pahlman et al., 2007) and redirecting inflammation from a transient state to a chronic state (Seron et al., 2011). The collagenous domain of Scl1 (denoted CL) mimics mammalian collagens by interacting with collagen receptor integrins $\alpha2\beta1$ and $\alpha11\beta1$ through a GLPGER binding site (Caswell et al., 2008b). This interaction facilitates *S. pyogenes* adherence to host cells and activates intracellular signaling (Humtsoe et al., 2005). It also enhances the internalization of *S. pyogenes* by host cells and re-emergence from host cells into extracellular environment (Caswell et al., 2007). More recently, it was found that Scl1 protein plays an important role in biofilm formation by targeting EDA-containing cellular fibronectin (Oliver-Kozup et al., 2011, 2013).

A very different role appears for the two collagen-like proteins, BclA and BclB, found in the pathogenic bacteria *Bacillus anthracis* (Sylvestre et al., 2002; Waller et al., 2005). These glycosylated proteins are structural components of the *Bacillus* exosporium and have been shown to be present in thin hair-like surface filaments. Similar to Scl1 and Scl2, the central part of BclA and BclB is the collagenous region with a $(Gly\text{-}Xaa\text{-}Yaa)_n$ sequence (Boydston et al., 2005). The length of the central collagenous domain is highly polymorphic, with 17–91 Gly-Xaa-Yaa tri-peptides, and the variation of exosporium filament hair length is dependent on the length of BclA collagenous domain (Sylvestre et al., 2003). A globular C-terminal domain is located at the distal end of the filaments and forms a rugged permeability barrier or shield around the spore (Boydston et al., 2005).

Even systems which have only been partly characterized hint at the complexities of quaternary structure, interactions and function that may be involved with bacterial collagen-like proteins. For example, collagen-like sequences have been found as part of the spore appendages of *Clostridium taeniosporum* (Walker et al., 2007). Two of the 4 appendage proteins have collagen-like sequences: GP85 has 53 Gly-Xaa-Yaa repeats, while CL2 has 43 Gly-Xaa-Yaa repeats (Walker et al., 2007). In other species, such as *B. anthracis* (Steichen et al., 2003), an external nap has been associated with triple helical collagen, so this may also prove to be the case for *C. taeniosporum*, but the formation of triple helical structure has not yet been shown. Another partly characterized system is the collagen-like domains reported in *Pasteuria ramosa* (Mouton et al., 2009; McElroy et al., 2011), where a triple-helical structure has been inferred by comparison to the *Bacillus* structure (Mouton et al., 2009; McElroy et al., 2011). Recent studies (McElroy et al., 2011), using analysis of an incomplete genome analysis for *P. ramosa*, have suggested huge complexity for the collagens in this species.

The bacterial collagens are frequently associated with the outer membrane of the organisms. In mammalian systems there are also certain collagens, for example types XIII, XVII, XXIII and XXV that are transmembrane collagens (Franzke et al., 2005; Ricard-Blum, 2011). The ectodomains of mammalian transmembrane collagens and certain bacterial collagens both show cell adhesive properties.