Isopeptide bond in collagen- and fibrinogen-binding MSCRAMMs

Upasana Sridharan¹ · Karthe Ponnuraj¹

Received: 21 October 2015 / Accepted: 14 December 2015 / Published online: 11 January 2016
© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag Berlin Heidelberg 2016

Abstract The internal isopeptide bonds are amide bonds formed autocatalytically between the side chains of Lys and Asn/Asp residues and have been discovered recently. These bonds are well conserved in Gram-positive bacterial pilin proteins and are also observed over a wide range of Gram-positive bacterial surface proteins. The presence of these bonds confers the pilus subunits with remarkable properties in terms of thermal stability and resistance to proteases. Like pili, microbial surface components recognizing adhesive matrix molecules (MSCRAMMs) are also surface proteins found only in Gram-positive bacteria. They specifically interact with the extracellular matrix (ECM) molecules like collagen, fibrinogen, fibronectin, laminin, etc. Many biophysical and biochemical studies have been carried out to characterize the isopeptide bonds in pili proteins from Gram-positive bacteria, but no attempts have been made to study the isopeptide bonds in MSCRAMMs. This short review aims to study the significance of the isopeptide bonds in relation to their function, by analyzing the crystal structures of collagen- and fibrinogen-binding MSCRAMMs. In this analysis, interestingly, we observed that the putative isopeptide bonds are restricted to the collagen-binding MSCRAMMs. Based on analogy with bacterial pilus subunits, we hypothesize that the collagen-binding MSCRAMMs possessing putative isopeptide bonds exhibit similar structural properties, which could help the bacteria in colonizing the host and provide resistance against host–defense mechanisms.

Keywords Isopeptide bonds · MSCRAMMs · Adhesin · Stability · Gram-positive bacteria

Introduction

Host tissue adhesion is the first critical step for pathogenesis. The Gram-positive and -negative bacteria use a multitude of surface proteins to adhere to the host tissues. This attachment process is important for colonization and infection of the host. Microbial surface components recognizing adhesive matrix molecules (MSCRAMMs) belong to a specific type of bacterial adhesion proteins that are covalently linked to the peptidoglycans of the cell wall of Gram-positive bacteria. Many of the MSCRAMMs are mainly targeted towards the extracellular matrix (ECM) molecules, such as collagen, fibrinogen, fibronectin, laminin, etc. (Patti et al. 1994; Foster et al. 2014). The structural studies on MSCRAMMs conducted in the past 15 years have revealed a highly similar structural folding involving an LPXTG motif conserved among them (Vengadesan and Narayana 2011). The domain organization for MSCRAMMs shows a typical N-terminal signal peptide sequence, followed by a non-repetitive ligand-binding ‘A’ region and a varying number of repetitive domains, followed by a cell wall anchoring region, including an LPXTG motif in the cell wall sorting region, a hydrophobic domain spanning the membrane region, and a cytoplasmic positively charged C-terminal sequence (Fig. 1). The ligand-binding ‘A’ region is generally composed of two or three domains, and the domains involved in the ligand binding adopt a novel DEv-IgG fold (Deivanayagam et al. 2002) that belongs to the immunoglobulin superfamily having a V-type fold. This fold provides a solitary model for MSCRAMMs that are highly efficient in adhering to the ECM molecules, followed by a biofilm formation and immune evasion (Vengadesan and Narayana 2011).

✉ Karthe Ponnuraj
pkarthe@hotmail.com; karthe@unom.ac.in

¹ Centre of Advanced Study in Crystallography and Biophysics,
University of Madras, Guindy Campus, Chennai 600 025, India

Fig. 1 The domain organization of some collagen- and fibrinogen-binding microbial surface components recognizing adhesive matrix molecules (MSCRAMMs). Common structural motifs present in MSCRAMMs are an N-terminal signal sequence (S) followed by a non-repetitive ligand-binding region [A-region; green (N-terminal domain) and yellow (C-terminal domain)] and a varying number of repetitive regions suitable for stalking/projecting the ligand-binding region and a

C-terminal region for cell wall anchoring that includes a cell wall sorting region (W) containing the LPXTG motif, the membrane-spanning hydrophobic domain (M), and the cytoplasmic positively charged C-terminal tail (C). The N- and C-terminal domains of the ligand-binding A-region exhibit DEv-IgG fold, whereas the domains of the B-repeat region exhibit inv-IgG fold

The disulfide bond is the only well-known covalent crosslink in proteins where two cysteine residues are bound together in redox conditions spontaneously. However, studies have presently highlighted that other crosslinks like ester (Baker et al. 2015) and isopeptide bonds (Kang et al. 2007) are available in proteins. Isopeptide bonds are intramolecular covalent bonds that form autocatalytically between the side chains of lysine and asparagine/aspartic acid. These are found to confer high stability to the protein, conveying structural solidity (Wang et al. 2013b). Though isopeptide bonds were known previously for their crosslink formation in ubiquitylation (Pickart 2001), sumoylation, and transglutamination (Greenberg et al. 1991), it was discovered recently that they can form autocatalytically in proteins when a suitable hydrophobic environment is created (Kang et al. 2007).

The autocatalytically formed isopeptide bonds are amide bonds that are quite prevalent. Recent studies have unfolded this type of bonds, mainly in bacterial pili proteins, which are individual proteinaceous and non-flagellar surface appendages involved in the host adhesion mechanism (Kline et al. 2010). The main functional implication of pili involves, similar to MSCRAMMs and other surface proteins, actions such as invasion, colonization, signaling, and biofilm formation of the anchored cell (Mandlik et al. 2008). Pili are more common in

Gram-negative bacteria; however, these have been observed in some species of Gram-positive bacteria as well (Telford et al. 2006). The isopeptide bonds in pili are found in a few Gram-positive pathogenic bacteria, including *Streptococcus pyogenes* (proteins such as Spy0128, FctB) (Kang et al. 2007; Linke et al. 2010), *Corynebacterium diphtheriae* (SpaA) (Kang et al. 2009), *Bacillus cereus* (BcpA) (Budzik et al. 2009; Hendrickx et al. 2012), *S. pneumoniae* (RrgA and RrgB) (Spraggon et al. 2010), *S. gordonii* (SspB) (Forsgren et al. 2010), *S. mutans* (SpaP) (Larson et al. 2011; Nylander et al. 2011), and *S. agalactiae* (GBS52) (Krishnan et al. 2007). The functional implications of isopeptide bonds in pili have been studied significantly (Kang and Baker 2009; Alegre-Cebollada et al. 2010; Paterson and Baker 2011; Wang et al. 2013b), but not with reference to MSCRAMMs. For example, the structure of the collagen-binding domain of *S. aureus* Cna, a prototype MSCRAMM, was solved nearly two decades ago (Symersky et al. 1997). This structure contains a putative isopeptide bond (analyzed in the present study), but the significance of the bond has not yet been well studied. MSCRAMMs and pili differ in their structural organization (Vengadesan and Narayana 2011; Krishnan 2015); however, they share some similarities, including these crosslinking isopeptide bonds and sortase-mediated anchoring to the cell wall. This review focuses on the isopeptide bonds in the MSCRAMM structures, their structural

conservation, and possible contribution to the function and stability of the adhesin.

Chemical structure, occurrence, and formation of isopeptide bonds

An intramolecular isopeptide bond is a crosslink, formed between the side chain amine of lysine residue and the side chain of asparagine/aspartic acid, along with an essential base glutamic acid/aspartic acid under a favorable hydrophobic environment (Fig. 2). The isopeptide bond takes up two isomeric forms, *cis* and *trans* configurations, and can be classified into two types, the D-type and the E-type, based on their topology (Kang et al. 2007; Krishnan 2015). The isopeptide bond formation is described as a proximity-induced reaction (Kang and Baker 2011). It depends primarily on the geometry of the participating amino acids and the hydrophobic moiety in which the isopeptide unit is formed. Such an environment would lower the pK_a of lysine residue, inducing a nucleophilic attack by the unprotonated ϵ -amino group of the Lys residue on the γ -carbon group of Asn/Asp, leading to a tetrahedral intermediate, followed by a proton shuttle from the Lys residue to the leaving group, i.e., either $-NH_2$ or $-OH$, via a trivial neighboring base, Glu/Asp (Kang and Baker 2011). These two residues might also help to polarize the carbon atom from Asn/Asp, thereby making it more prone to the nucleophilic attack. The Lys–Asn pair releases ammonia and the Lys–Asp pair releases water as products after the bond formation; otherwise, the reaction is the same and isopeptide bonds are formed at the same rate (Kang and Baker 2011).

Isopeptide bonds in MSCRAMMs

MSCRAMMs are modular proteins, and different domains of MSCRAMMs generally carry an immunoglobulin-G (IgG) topology with a novel variation. The ligand-binding A-region is usually made up of two or three domains (Fig. 1), and each

Fig. 2 Schematic representation of the formation of an isopeptide bond. Amide bond formation between the lysine and asparagine/aspartate side chains. The neighboring glutamine acts as a catalyst in this spontaneous isopeptide bond formation

domain exhibits DEv-IgG fold. The DEv-IgG fold is very similar to the IgG-C fold (Fig. 3a), but has additional secondary structural elements between the strands D and E (Fig. 3b). Many collagen- and fibrinogen-binding MSCRAMMs contain a repetitive B-region, which follows the A-region (Fig. 1), and the individual domains of the B-region adopt a fold called the inverse-IgG (inv-IgG) fold (Fig. 3c). The inv-IgG fold is similar to the IgG-C fold, but arranged in a reverse order. The DEv-IgG and inv-IgG folds are also referred as CnaA and CnaB folds, respectively (Vengadesan and Narayana 2011).

Though the crystal structures of many MSCRAMMs have been solved to date, none of them have reported the presence of isopeptide bonds in those structures. While solving the structures of MSCRAMMs like collagen-binding Cna (Symersky et al. 1997; Zong et al. 2005), Ace (Liu et al. 2007; Ponnuraj and Narayana 2007), and B-repeats of Cna (Deivanayagam et al. 2000), any continuous electron density joining the side chain of Lys and Asn would have been considered as an unexplained density and not designated as an isopeptide bond. Autocatalytically formed isopeptide bonds were first discovered in the structure of the major pilin subunit Spy0128 of *S. pyogenes* (Kang et al. 2007).

So far, the intramolecular isopeptide bonds have been found only in DEv-IgG (CnaA) and inv-IgG (CnaB) type

Fig. 3 Topology of the IgG constant domain and its variant DEv-IgG and Inv-IgG folds. **a** IgG constant domain. It consists of a four-stranded β -sheet (ABED) on one side and a three-stranded β -sheet (CFG) on the other side, thus forming a barrel. **b** The DEv-IgG fold. It is similar to the IgG-C fold, but consists of a few extra secondary structural elements (light brown) that are intervened between the strands D and E. The isopeptide bond found in the C-terminal domain of the ligand-binding region of collagen-binding Cna, Ace, and RspB structures is indicated by the black dotted line between the A and F strands. **c** The Inv-IgG fold. It consists of a four- (DAGF) and three- (EBC) or a three- (DAG) and four- (FEBC) stranded β -sheets similar to the IgG-C fold, but stacked in a reverse manner. The putative isopeptide bond observed in the B-repeats of Cna between the A and G strands is shown by the black dotted line

folded domains (Baker et al. 2015). There are two types of isopeptide bonds, the D-type and the E-type. The former is observed in DEv-IgG folds and the latter in Inv-IgG folds (Vengadesan and Narayana 2011). The E-type isopeptide bond is more commonly found in several Gram-positive pilins (Krishnan 2015). The D-type isopeptide bond bridges two β -sheets, where the target Lys residue is found in the 'A' strand of sheet I and the host Asn/Asp is found in the 'F' strand of sheet II, along with a conserved Asp residue in the 'C' strand of sheet II (Fig. 3b). The strands 'A' and 'F' are in an antiparallel fashion. A similar bonding is also observed in the E-type, where the target and host residues are placed within the same β -sheet. The strands carrying the Lys and Asn residues are arranged in a parallel fashion, along with a conserved Glu residue present at a hydrogen bond forming distance to the Lys residue (Fig. 3c).

Collagen-binding MSCRAMMs

Collagens are the main structural proteins constituting the extracellular space in most life forms. Collagen-binding adhesins are well expressed in all vertebrates, eukaryotes, and prokaryotes. With reference to MSCRAMMs, the collagen adhesin (Cna) of *S. aureus* is considered a prototype. In Cna, the collagen-binding A-region (residues 31–344) is made up of two domains (Fig. 1), named N1 and N2 (Fig. 4a). Symersky et al. (1997) reported the first crystal structure of an MSCRAMM, which is the N2 domain of Cna. Later, Zong et al. (2005) reported the crystal structure of the collagen-binding A-region (consisting of both N1 and N2 domains) as an apoprotein and also in complex with collagen triple helical peptide. These structures revealed that both N1 (N-terminal domain) and N2 (C-terminal domain) adopt the DEv-IgG fold, and the interdomain region serves as a collagen-binding site (Zong et al. 2005). Following Cna, two more collagen-binding MSCRAMMs structures were solved, namely Ace from *Enterococcus faecalis* (Liu et al. 2007; Ponnuraj and Narayana 2007) and RspB from *Erysipelothrix rhusiopathiae* (Devi et al. 2012). In the structural report of these proteins (Cna, Ace, and RspB), there is no mention of the presence of any isopeptide bonds. However, the presence of a putative isopeptide bond in the N2 domain of Cna was reported during the discovery of an isopeptide bond in Spy0128 (Kang et al. 2007).

The multiple sequence alignment of the ligand-binding region of Cna (31–330), Ace (32–327), and RspB (30–348) revealed a low sequence identity between them, but they exhibit similar structural fold (Devi et al. 2012). This observation prompted the question as to whether Ace and RspB structures could also contain an isopeptide bond, as observed in Cna. Therefore, we analyzed the structures of full-length Cna (PDB ID 2F6A), Ace (PDB ID 2Z1P), RspB (PDB ID 3V10),

and the minimal collagen-binding domain of Cna (termed Cna19; PDB ID 1AMX) and Ace (Ace19; PDB ID 2OKM). The electron density maps of Cna19, Ace19, and RspB clearly revealed that all these structures contain a D-type isopeptide bond. Further, it is surprising to note that the isopeptide bond is found only in the N2 domain (Fig. 4a, b), although the N1 and N2 domains are structurally homologous. In Cna19, the isopeptide bond is formed between Lys176 and Asn293, along with Asp209 (PDB ID 1AMX) (Fig. 4c). In Ace19, it is formed between Lys42 and Asn155, along with Asp74 (PDB ID 2OKM) (Fig. 4d), whereas in RspB, Lys198 and Asn308 along with Asp231 are (PDB ID 3V10) involved in the similar bond formation (Fig. 4e). It is also observed that the isopeptide bond is structurally conserved in all three proteins (Fig. 4b).

We analyzed the full-length Cna structure (PDB ID 2F68), especially near the ligand-binding region, to find any putative isopeptide bond(s) which may be functionally important. This analysis revealed that the residues Lys289, Asn320, and Glu132 are located near the latching region, with Asn320 at the beginning of the latching strand (Fig. 4f). However, the distances between the participating amino acids are not in the hydrogen bonding range. But minor structural rearrangement/side-chain re-orientation could bring these residues close and, under suitable conditions, there is a possibility of isopeptide bond formation between Lys289 and Asn320. The Lys–Asn–Glu triad found in Cna is not structurally conserved in Ace and RspB; however, the latter two structures also contain the Lys–Asn–Glu triad at different positions near the beginning of the latch. But, similar to Cna, the side chains of the participating amino acids are out of the hydrogen bonding range. In the ligand-binding mechanism involving collagen/fibrinogen, the final “latching event” is critical for tightly securing the ligand in its place, i.e., in between the two domains (Ponnuraj et al. 2003; Zong et al. 2005). If the Lys–Asn–Glu triad found near the latching region of Cna, Ace, and RspB forms an isopeptide bond during some biological process (for example, biofilm formation), this could result in a “rigid latch”. This event could lead to tighter capture of collagen, which is functionally very important.

Concerning the repetitive B-region in collagen-binding MSCRAMMs, so far, the crystal structure consisting of two (PDB ID 1D2O) and four B-repeats (PDB ID 1D2P) (we considered one CnaB fold domain as one repeat) are available for collagen-binding Cna of *S. aureus* (Deivanayagam et al. 2000). Despite the availability of these structures, the function of this Cna B-region is still unclear and is hypothesized to be functioning independently from the Cna A-region (Snodgrass et al. 1999). Kang et al. (2007) reported the presence of a putative isopeptide bond in the structure of Cna B-repeat. To get more insight into this, we analyzed the structures of two

Fig. 4 The isopeptide bond and Lys–Asn–Glu triad in collagen-binding MSCRAMMs. **a** Ribbon representation showing the ligand-binding domains [green (N-terminal domain) and yellow (C-terminal domain)] of Cna in complex with a collagen peptide (PDB ID 2F6A). The collagen peptide is bound in between the N1 and N2 domains. The red circle in the N2 domain indicates the location of the isopeptide bond. The red dotted line at the latching strand indicates the hydrogen bond. **b** Structural superposition of the minimal collagen-binding domain of Cna (gray, PDB ID 1AMX), Ace (magenta, PDB ID 2OKM), and the N2 domain of RspB (green, PDB ID 3V10). The 2Fo–Fc electron density

map contoured at 1.0σ shows the continuous density between the side chain of Lys and Asn in the structure of the: **c** minimal collagen-binding domain of Cna (PDB ID 1AMX, resolution 2.0 Å); **d** minimal collagen-binding domain of Ace (PDB ID 2OKM, resolution 1.65 Å); and **e** RspB (PDB ID 3V10, resolution 1.75 Å). The continuous electron density between Lys and Asn residues indicates the presence of an isopeptide bond. **f** The Lys–Asn–Glu triad at the latch region of Cna (PDB ID 2F68). The distance between the residues is given in Angstrom units and is shown by the red dotted line

and four B-repeats of Cna. This analysis revealed that one Lys–Asn–Glu triad is present per repeat, with the participating residues at the hydrogen bonding distance. Since the structure factors are not available in the Protein Data Bank (PDB) for the structures of Cna B-repeats, we were unable to generate its electron density maps to check if there is an amide bond between the side chain of Lys and Asn or not. But based on the geometric arrangement of the residues in the Lys–Asn–Glu triad, it is highly likely that the isopeptide bond is formed between the Lys–Asn residues of the triad, which is present in all four B-repeats (B1–B4). It is noted that the triads Lys543–Asn620–Glu603, Lys633–Asn717–Glu694, Lys730–Asn807–Glu790, and Lys820–Asn904–Glu881 are capable of forming an isopeptide bond in each of the four B-repeats (Fig. 5a, b, c, e). Each B-repeat region adopts the inverse of the IgG fold and it holds the E-type putative isopeptide bond that is specific for this kind of reverse orientation of the IgG fold (Krishnan 2015). The superposition of the four individual repeats shows that the Lys–Asn–Glu triad is structurally well conserved (Fig. 5f). This is similar to the conservation of isopeptide bond in the N2 domain of the collagen-binding A-region of Cna, Ace, and RspB (Fig. 4b).

To date, many pilin subunit structures have been determined and the CnaB-type fold is found to be dominant in those structures (Krishnan 2015). It is our interest to compare the CnaB-type pilin subunit structure with the Cna B-repeat to identify any structural conservation on the isopeptide bond. The N2 domain of GBS52, the minor pilin from *S. agalactiae* (PDB ID 3PHS), was compared with the Cna B-repeat region comprising residues 722–811. The structure superposition revealed that the isopeptide bond (Lys160–Asn256) in pilin is structurally conserved with the putative isopeptide bond of the Cna B-repeat (Fig. 5g). As many pilin subunit structures are available, further structural and sequence analysis is required to obtain some insights on the structural conservation of the isopeptide bond seen here.

Fibrinogen-binding MSCRAMMs

Fibrinogen is a glycoprotein that is found in blood plasma and plays integral roles in homeostasis and coagulation. It is a

Fig. 5 Putative isopeptide bond in the B-repeats of Cna. **a** Ribbon representation of the structure of the four repeats (B1, B2, B3, and B4) of Cna B-region (PDB ID 1D2P). One CnaB (Inv-IgG) type fold is considered as one repeat. The putative isopeptide bond forming residues are shown as sticks in four B-repeats. **b–e** Close-up views show the geometric arrangement of the Lys–Asn–Glu triad in the B1, B2, B3, and B4 repeats, respectively. The distance between the residues

is given in Angstrom units and is shown by the red dotted lines. **f** Structural superposition of the B1, B2, B3, and B4 repeats. The Lys–Asn–Glu triad (shown as a stick) is structurally conserved. **g** Structural superposition of the B3 repeat of Cna (residues 722–811, gray) and the N2 domain of GBS52 (yellow). In GBS52, the isopeptide bond is formed between Lys160–Asn256 (shown as a stick), whereas in the B3 repeat of Cna, the corresponding residues are at the hydrogen bonding distance

symmetrical dimeric complex structure made up of six polypeptide chains consisting of a pair of each α -, β - and γ -chains. All three chains have been shown to be critical for pathogenic adherence (Cheung et al. 1991; Otto 2013). The crystal structure of several fibrinogen-binding MSCRAMMs have been solved so far, including ClfA (PDB ID 1N67) (Deivanayagam et al. 2002), ClfB–Fg peptide complex (PDB ID 4F20) (Xiang et al. 2012), SdrD (PDB ID 4JDZ) (Wang et al. 2013a) from *S. aureus*, Srr-1 (PDB ID 4MBO) and Srr-2 (PDB ID 4MBR) (Seo et al. 2013) from *S. agalactiae*, and SdrG (PDB ID 1R19 and 1R17, native and complex with fibrinopeptide, respectively) from *Staphylococcus epidermidis* (Ponnuraj et al. 2003). When superposed, they share a highly conserved structural homology, but exhibit much less sequence identity (<20%) than that observed in collagen-binding MSCRAMMs. The fibrinogen-binding A-region is made up of two domains, named N2 (N-terminal domain) and N3 (C-terminal domain) (Fig. 6a). Both domains exhibit DEv-IgG fold (Ponnuraj et al. 2003).

A detailed analysis of these structures revealed that most of the proteins under study here contain the Lys–

Asn–Glu triad, but the participating residues are not in the hydrogen bonding distance. If the Lys and Asn residues in the triad have to form an isopeptide bond, a minor conformational change or side-chain reorientation is necessary. The superposition of these structures shows that these triads are not structurally conserved and are located in different regions of the molecule (Fig. 6a–d). One interesting observation is that, in the structure of SdrG–fibrinopeptide complex (PDB ID 1R17), the triad is located at the tail of the latching region (Fig. 6d), and this region is functionally critical (Ponnuraj et al. 2003). If there is an isopeptide bond at the latching region, that would lead to a very sturdy latch. This will result in a highly secured adhesin–host complex.

Concerning the B-repetitive region in fibrinogen-binding MSCRAMMs, the crystal structure is available only for one B-repeat region of SdrD of *S. aureus* (Wang et al. 2013a). This structure also contains the ligand-binding A-region (N2 and N3 domains), which precedes the B-repeat. This is the only MSCRAMM structure which contains both the ligand-binding A-region and

Fig. 6 The Lys–Asn–Glu triad in fibrinogen-binding MSCRAMMs. **a** Ribbon representation showing the ligand-binding domains (green and yellow) of SdrG in complex with fibrinogen peptide (pink) (PDB ID 1R17). This structure is considered a representative structure for fibrinogen-binding MSCRAMMs, since all the fibrinogen-binding MSCRAMMs are structurally highly homologous. The Lys–Asn–Glu triad is shown in various fibrinogen-binding MSCRAMMs, such as **b** SdrD (PDB ID 4JDZ), **c** Srr-2 (PDB ID 4MBR), and **d** SdrG–fibrinogen peptide complex (PDB ID 1R17). **e** Structural comparison of

the B-repeat of the collagen- and fibrinogen-binding MSCRAMMs. Ribbon representation showing the superposition of the collagen-binding Cna B-repeat (yellow, residues 625–721, PDB ID 1D2P) and the fibrinogen-binding SdrD B-repeat (brown, residues 556–680, PDB ID 4JDZ). Both the structures are highly identical and adopt CnaB fold. Differences are observed only in the loop regions. Despite the high structural homology, the putative isopeptide bond is found only in the Cna B-repeat, whereas it is absent in SdrD

the B-repeat segment. The B-repeat region of SdrD possesses a CnaB fold and there is a good structural identity between the B-repeat region of SdrD and the Cna B-repeat (backbone RMSD of 2.01 Å, Fig. 6e). Despite the conservation of the structural fold, it is surprising to note that the B-repeat region of SdrD lacks the isopeptide bond. The SdrD structure contains three Ca^{2+} cations in the B-repetitive region that are coordinated to the neighboring loops. Perhaps, these ions contribute to stabilizing the protein structure and compensate for the occurrence of isopeptide bonds functionally.

Conclusion

The intramolecular isopeptide bonds in bacterial pili have been helpful in revealing the prevalence of their occurrence in MSCRAMMs of different bacterial species. Though MSCRAMM structures contain intramolecular

Lys–Asn isopeptide bond and the Lys–Asn–Glu triad, the detailed functional significance of these bonds has not been fully studied. On the other hand, in the pili which are also like MSCRAMMs found on the bacterial surface, mediating adherence and colonization, containing CnaA-type and CnaB-type domains, the significance of the isopeptide bonds have been studied in great detail. The isopeptide bonds studied in the major and minor pilin proteins from Gram-positive bacteria so far have been suggested to provide structural stability to the overall molecule, as the protein withstood against chemical, mechanical, and thermal stress. For example, an atomic force microscopy (AFM) analysis of Spy0128, the major pilin protein of *S. pyogenes*, revealed that the subunits of Spy0128 are inductile despite high pull forces up to 800 pN being applied (Alegre-Cebollada et al. 2010). Destruction of intramolecular isopeptide bonds by site-directed mutagenesis in Spy0128 reduced the mechanical flexibility of the

protein incredibly (Alegre-Cebollada et al. 2010). The thermal stability of Spy0128 and its ability to resist proteases were severely affected by the loss of the isopeptide bonds (Kang and Baker 2009). Even circular dichroism (CD) studies conducted in Spy0128 showed a similar scenario, where the thermal unfolding showed a striking destabilization in the absence of the intramolecular isopeptide bonds. A decrease in melting temperature (T_m) was also noted during this study (Kang and Baker 2009). Mutational analysis combined with electron microscopic and atomic force microscopic studies on the major pilin protein BcpA revealed that the isopeptide bonds are required for the bundle formation of BcpA pili (Hendrickx et al. 2012).

The intramolecular isopeptide bonds have been found only in CnaA-type and CnaB-type domains, and these domains are predicted to occur in a wide range of bacterial surface proteins. This conservation conveys a message about how dominant these isopeptide bonds are throughout the evolution. Besides this conversation, it is interesting to note that these bonds are not universally found in all CnaA-type and CnaB-type domains, and the reason behind the selectivity is not known.

The bacteria probably would adapt all necessary steps for its adhesive molecule to interact with the host molecule in a lock-and-key fashion. In this context, the Lys–Asn–Glu triad found in the latching region of the collagen- and fibrinogen-binding MSCRAMMs likely holds some functional significance, as a formation of this amide bond would lead to “rigid latch”. As a result, the ligand would be tightly secured and would not easily dissociate from the ligand binding cleft.

Not many experimental studies, such as mutational analysis, CD, AFM, electron microscopy, nuclear magnetic resonance (NMR) spectroscopy, mass spectrometry, and proteolysis, have been conducted in MSCRAMMs on isopeptide bonds. However, we hypothesize that, based on similar structural context with pili, MSCRAMMs possessing isopeptide bonds would also exhibit similar structural properties. Our structural analysis revealed that, although the ligand-binding region of collagen- and fibrinogen-binding MSCRAMMs share similar structural features, such as it is being made up of two domains and these domains exhibiting DEv-IgG fold, the putative isopeptide bond is found only in collagen-binding MSCRAMMs. This also holds true for the B-repeat region, as Cna B-repeats (inv-IgG fold) only contain the putative isopeptide bonds, whereas it is not observed in fibrinogen-binding SdrD. This observation is definitely one of the unique features of collagen-binding MSCRAMMs, and its functional implication needs to be further explored using biophysical and biochemical studies.

Acknowledgments KP thanks the Department of Science and Technology (DST), Government of India for the financial support in the form of a grant and US for the DST INSPIRE fellowship.

Compliance with Ethical Standards

Conflict of interest Upasana Sridharan declares that she has no conflict of interest.

Karthe Ponnuraj declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Alegre-Cebollada J, Badilla CL, Fernández JM (2010) Isopeptide bonds block the mechanical extension of pili in pathogenic *Streptococcus pyogenes*. *J Biol Chem* 285:11235–11242
- Baker EN, Squire CJ, Young PG (2015) Self-generated covalent cross-links in the cell-surface adhesins of Gram-positive bacteria. *Biochem Soc Trans* 43:787–794
- Budzick JM, Poor CB, Faull KF, Whitelegge JP, He C, Schneewind O (2009) Intramolecular amide bonds stabilize pili on the surface of bacilli. *Proc Natl Acad Sci U S A* 106:19992–19997
- Cheung AL, Krishnan M, Jaffe EA, Fischetti VA (1991) Fibrinogen acts as a bridging molecule in the adherence of *Staphylococcus aureus* to cultured human endothelial cells. *J Clin Invest* 87:2236–2245
- Deivanayagam CC, Rich RL, Carson M, Owens RT, Danthuluri S, Bice T, Höök M, Narayana SV (2000) Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure* 8:67–78
- Deivanayagam CC, Wann ER, Chen W, Carson M, Rajashankar KR, Höök M, Narayana SV (2002) A novel variant of the immunoglobulin fold in surface adhesins of *Staphylococcus aureus*: crystal structure of the fibrinogen-binding MSCRAMM, clumping factor A. *EMBO J* 21:6660–6672
- Devi AS, Ogawa Y, Shimoji Y, Balakumar S, Ponnuraj K (2012) Collagen adhesin–nanoparticle interaction impairs adhesin’s ligand binding mechanism. *Biochim Biophys Acta* 1820:819–828
- Forsgren N, Lamont RJ, Persson K (2010) Two intramolecular isopeptide bonds are identified in the crystal structure of the *Streptococcus gordonii* SspB C-terminal domain. *J Mol Biol* 397:740–751
- Foster TJ, Geoghegan JA, Ganesh VK, Höök M (2014) Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. *Nat Rev Microbiol* 12:49–62
- Greenberg CS, Birckbichler PJ, Rice RH (1991) Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. *FASEB J* 5:3071–3077
- Hendrickx AP, Poor CB, Jureller JE, Budzick JM, He C, Schneewind O (2012) Isopeptide bonds of the major pilin protein BcpA influence pilus structure and bundle formation on the surface of *Bacillus cereus*. *Mol Microbiol* 85:152–163
- Kang HJ, Baker EN (2009) Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of *Streptococcus pyogenes*. *J Biol Chem* 284:20729–20737
- Kang HJ, Baker EN (2011) Intramolecular isopeptide bonds: protein crosslinks built for stress? *Trends Biochem Sci* 36:229–237
- Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN (2007) Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. *Science* 318:1625–1628

- Kang HJ, Paterson NG, Gaspar AH, Ton-That H, Baker EN (2009) The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc Natl Acad Sci U S A* 106:16967–16971
- Kline KA, Dodson KW, Caparon MG, Hultgren SJ (2010) A tale of two pili: assembly and function of pili in bacteria. *Trends Microbiol* 18:224–232
- Krishnan V (2015) Pili in gram-positive bacteria: a structural perspective. *IUBMB Life* 67:533–543
- Krishnan V, Gaspar AH, Ye N, Mandlik A, Ton-That H, Narayana SV (2007) An IgG-like domain in the minor pilin GBS52 of *Streptococcus agalactiae* mediates lung epithelial cell adhesion. *Structure* 15:893–903
- Larson MR, Rajashankar KR, Crowley PJ, Kelly C, Mitchell TJ, Brady LJ, Deivanayagam C (2011) Crystal structure of the C-terminal region of *Streptococcus mutans* antigen I/II and characterization of salivary agglutinin adherence domains. *J Biol Chem* 286:21657–21666
- Linke C, Young PG, Kang HJ, Bunker RD, Middleditch MJ, Caradoc-Davies TT, Proft T, Baker EN (2010) Crystal structure of the minor pilin FctB reveals determinants of Group A streptococcal pilus anchoring. *J Biol Chem* 285:20381–20389
- Liu Q, Ponnuraj K, Xu Y, Ganesh VK, Sillanpää J, Murray BE, Narayana SVL, Höök M (2007) The *Enterococcus faecalis* MSCRAMM ACE binds its ligand by the Collagen Hug model. *J Biol Chem* 282:19629–19637
- Mandlik A, Swierczynski A, Das A, Ton-That H (2008) Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. *Trends Microbiol* 16:33–40
- Nylander A, Forsgren N, Persson K (2011) Structure of the C-terminal domain of the surface antigen SpaP from the caries pathogen *Streptococcus mutans*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 67:23–26
- Otto M (2013) Staphylococcal Infections: mechanisms of biofilm maturation and detachment as critical determinants of pathogenicity. *Annu Rev Med* 64:175–188
- Paterson NG, Baker EN (2011) Structure of the full-length major pilin from *Streptococcus pneumoniae*: implications for isopeptide bond formation in gram-positive bacterial pili. *PLoS One* 6:e22095
- Patti JM, Allen BL, McGavin MJ, Höök M (1994) MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu Rev Microbiol* 48:585–617
- Pickart CM (2001) Mechanisms underlying ubiquitination. *Annu Rev Biochem* 70:503–533
- Ponnuraj K, Narayana SV (2007) Crystal structure of ACE19, the collagen binding subdomain of *Enterococcus faecalis* surface protein ACE. *Proteins* 69:199–203
- Ponnuraj K, Bowden MG, Davis S, Gurusiddappa S, Moore D, Choe D, Xu Y, Höök M, Narayana SV (2003) A “dock, lock, and latch” structural model for a staphylococcal adhesin binding to fibrinogen. *Cell* 115:217–228
- Seo HS, Minasov G, Seepersaud R, Doran KS, Dubrovskaya I, Shuvalova L, Anderson WF, Iverson TM, Sullam PM (2013) Characterization of fibrinogen binding by glycoproteins Srr1 and Srr2 of *Streptococcus agalactiae*. *J Biol Chem* 288:35982–35996
- Snodgrass JL, Mohamed N, Ross JM, Sau S, Lee CY, Smeltzer MS (1999) Functional analysis of the *Staphylococcus aureus* collagen adhesin B domain. *Infect Immun* 67:3952–3959
- Spraggon G, Koesema E, Scarselli M, Malito E, Biagini M, Norais N, Emolo C, Barocchi MA, Giusti F, Hillerlingmann M, Rappuoli R, Lesley S, Covacci A, Maignani V, Ferlenghi I (2010) Supramolecular organization of the repetitive backbone unit of the *Streptococcus pneumoniae* pilus. *PLoS One* 5:e10919
- Symersky J, Patti JM, Carson M, House-Pompeo K, Teale M, Moore D, Jin L, Schneider A, DeLucas LJ, Höök M, Narayana SV (1997) Structure of the collagen-binding domain from a *Staphylococcus aureus* adhesin. *Nat Struct Biol* 4:833–838
- Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G (2006) Pili in Gram-positive pathogens. *Nat Rev Microbiol* 4:509–519
- Vengadesan K, Narayana SV (2011) Structural biology of Gram-positive bacterial adhesins. *Protein Sci* 20:759–772
- Wang X, Ge J, Liu B, Hu Y, Yang M (2013a) Structures of SdrD from *Staphylococcus aureus* reveal the molecular mechanism of how the cell surface receptors recognize their ligands. *Protein Cell* 4:277–285
- Wang B, Xiao S, Edwards SA, Gräter F (2013b) Isopeptide bonds mechanically stabilize spy0128 in bacterial pili. *Biophys J* 104:2051–2057
- Xiang H, Feng Y, Wang J, Liu B, Chen Y, Liu L, Deng X, Yang M (2012) Crystal structures reveal the multi-ligand binding mechanism of *Staphylococcus aureus* ClfB. *PLoS Pathog* 8:e1002751
- Zong Y, Xu Y, Liang X, Keene DR, Höök A, Gurusiddappa S, Höök M, Narayana SV (2005) A ‘Collagen Hug’ model for *Staphylococcus aureus* CNA binding to collagen. *EMBO J* 24:4224–4236