Polyproline and Triple Helix Motifs in Host-Pathogen Recognition

Rita Berisio and Luigi Vitagliano*

Istituto di Biostrutture e Bioimmagini, CNR via Mezzocannone 16, I-80134 Napoli, Italy

Abstract: Secondary structure elements often mediate protein-protein interactions. Despite their low abundance in folded proteins, polyproline II (PPII) and its variant, the triple helix, are frequently involved in protein-protein interactions, likely due to their peculiar propensity to be solvent-exposed. We here review the role of PPII and triple helix in mediating host-pathogen interactions, with a particular emphasis to the structural aspects of these processes. After a brief description of the basic structural features of these elements, examples of host-pathogen interactions involving these motifs are illustrated. Literature data suggest that the role played by PPII motif in these processes is twofold. Indeed, PPII regions may directly mediate interactions between proteins of the host and the pathogen. Alternatively, PPII may act as structural spacers needed for the correct positioning of the elements needed for adhesion and infectivity.

Recent investigations have highlighted that collagen triple helix is also a common target for bacterial adhesins. Although structural data on complexes between adhesins and collagen models are rather limited, experimental and theoretical studies have unveiled some interesting clues of the recognition process. Interestingly, very recent data show that not only is the triple helix used by pathogens as a target in the host-pathogen interaction but it may also act as a bait in these processes since bacterial proteins containing triple helix regions have been shown to interact with host proteins.

As both PPII and triple helix expose several main chain non-satisfied hydrogen bond acceptors and donors, both elements are highly solvated. The preservation of the solvation state of both PPII and triple helix upon protein-protein interaction is an emerging aspect that will be here thoroughly discussed.

Keywords: Collagen, triple helix, host-pathogen interaction.

SECONDARY STRUCTURE AND PROTEIN-PROTEIN INTERACTIONS

Protein structure is usually described as a hierarchal assembly of constitutive building blocks [1]. The vast majority of folded proteins contain one or more stretches of amino acids that assume repetitive and characteristic structures in the three-dimensional space denoted as secondary structure elements. Statistical surveys of protein structure models have unveiled the existence of several secondary structure elements. The most common ones, each assumed by approximately the 30% of the residues in protein structures, are α-helix and β-structure [2-5]. Less common motifs are 3(10)-helices and π-helix, polyproline II (PPII) helices and a variant of PPII denoted as triple helix [6-15].

Secondary structure elements play a major role in the stabilization of protein folds. However, they may also be crucially involved in molecular recognition processes, such as protein-protein, protein-DNA, protein-RNA, protein-sugars interactions. The involvement of secondary structure elements in molecular recognition obviously depends on their location within the protein context. Helices, that are often solvent exposed, commonly mediate protein-protein contacts. The role of the β-sheets in these processes is more intricate being these structure elements highly reactive and typically located in the core of protein structure [1]. Nonetheless, they frequently mediate protein-protein interactions, although their reactivity, if uncontrolled, may lead to unwanted aggregation processes underlying several pathological states [16-21]. The role of α-helices and β-structures in molecular recognition has been comprehensively and effectively reviewed in previous reports [21,22]. Here, we focus our attention to the role that recognition processes mediated by PPII and triple helix with a specific emphasis in host-pathogen interactions. A particular attention will be given to the structural aspects related to these processes.

POLYPROLINE HELICES: BASIC FEATURES

Proline is unique among genetically-encoded aminoacids [23]. The specific covalent link between the Cα atom of the Pro side chain with the backbone nitrogen atom confers this residues peculiar conformational properties. As a result of its cyclic nature, Pro ψ angle is restricted to values of -65°±15° [24-26]. In addition, the particular location of the Cα atom also restricts the ψ value of the preceding residue [24-27]. As a consequence of this conformational restrictions, in proteins proline-rich regions are known to preferentially adopt an extended conformation denoted as called the poly-L-proline helix of type II (PPII). PPII helices are left-handed, all-trans extended helices characterized by (ψ, ψ) backbone dihedral angles with average values of (-75°, 145°) [6-10,28]. These extended helices have three residues per turn, with a pitch value of 9.3 Å/turn.

*Address correspondence to this author at the Istituto di Biostrutture e Bioimmagini, CNR via Mezzocannone 16, I-80134 Napoli, Italy; Tel: 00390812534506; Fax: 00390812536642; Email: luigi.vitagliano@unina.it
One of the distinctive structural properties of PPII helices, compared to the other common secondary structure elements, is the absence of local hydrogen bonding patterns involving main chain atoms. This feature leaves several non-satisfied hydrogen bond donors and acceptors free to establish intra- or inter-molecular interactions (see below). Furthermore, the absence of specific hydrogen bonding patterns makes the formation of long PPII helices rather unusual in globular proteins [6-10]. It has been proposed that the backbone conformations of unfolded proteins are not random coils but include short stretches of PPII structural motifs interspersed with turns and bends (polypeptide hypothesis for unfolded proteins) [29-32]. The absence of regular hydrogen bonds makes of PPII helices dynamic structural entities whose plasticity may be important in protein-protein interaction [33].

One open question is the nature of the determinants of PPII stability, since it does not rely on hydrogen bonding interactions. In addition to the established entropic effects favoring the formation of PPII structures by iminoacids, one appealing novel proposal is that an electronic effect of the n→p* interaction between the nonbonding n-orbital of the p* orbital of the (i-1)th peptide carbonyl oxygen atom and the anti-bonding p*orbital of the (i-1)th peptide carbonyl carbon atom provides substantial stabilization to the left-handed PPII [34]. Given the typical exposure to the solvent of these helices [35], it is also possible that solvation plays an important role in PPII stabilization.

**POLYPROLINE HELICES IN HOST-PATHOGEN RECOGNITION**

The peculiar propensity of both the side chain and the backbone atoms of the PPII helices to be solvent-exposed makes it an excellent recognition motif [35,36]. Indeed, proline-rich regions are predominantly localized in the solvent-exposed regions such as the loops, intrinsically disordered regions, or between domains that constitute the inter-molecular interface. It is, therefore, not surprising that PPII helices are frequently involved in intermolecular interactions as signal transduction, antigen recognition, cell-cell communication and cytoskeletal organization. Peptide mimics of the PPII recognition motifs have been suggested as potential antagonists of intermolecular interactions [37].

The analysis of the energetic determinants of PPII binding to their structural cognates suggests that, in general, two main forces operate in concert. The first is the stacking of proline side-chains with the hydrophobic grooves in the binding site of PPII receptors whereas the second is represented by a firm network of water-mediated hydrogen bonding interactions. In this context, it is important to note that recent investigations have highlighted the role of hydration in PPII recognition. Current data indicate that the tendency of proline-rich sequences to be highly hydrated has been exploited by nature to favor the adaptability and plasticity of the different families of protein modules for the recognition of proline-rich targets [35]. It has been suggested that interfacial water molecules may provide important clues for deciphering the specificity versus promiscuity paradox in polyproline recognition [35].

As for many other biological processes, PPII motifs also play a role in host-pathogen recognition. One of the best characterized examples in this field is represented by the ability of PPII motifs of the protein ActA of the food-borne pathogen *Listeria monocytogenes* to bind the EVH1 domain of the mammalian protein Mena, an important factor in cytoskeleton regulation.

The interaction of ActA with Mena EVH1 is the molecular event that allows the bacterium to utilize the actin cytoskeleton to move and infect neighboring cells [38-40]. Since the inhibition of this interaction may provide a strategy to hamper the motility of the pathogen and to prevent its spread in the host organism, many efforts are currently devoted to the development of peptides able to interfere with ActA/Mena interaction [41]. In this scenario many efforts are currently made to fully elucidate the structural and energetic details of this protein-protein interaction.

A direct involvement of PPII motifs in host-pathogen recognition has also been unveiled for virus infections. A prototypical example in this context is represented by the HIV-1 Nef, a multifunctional protein required for full pathogenicity of the virus [42]. Although the role of this pathogenic factor in the progression to AIDS is not mechanistically understood, it is commonly assumed that Nef necessarily functions through protein-protein interaction, as it does not present any enzymatic activity. It has been shown that the critical Nef protein interaction interface is centered on its proline-rich domain (P69VPRQVPLRP78) which contains the helical SH3 domain binding peptide motif, PXXPXR [43].

A PPII motif is also responsible for the activation of the phosphatidylinositol 3-kinase (PI3K)/Akt signaling pathway by the influenza A virus [44]. The activation of this pathway is beneficial for virus replication by inhibiting virus-induced apoptosis through phosphorylation of caspase-9. Mutational studies have demonstrated that the influenza NS1 protein is able to bind the p85 regulatory subunit of PI3K though a PXXP region (amino acids 164 to 167) [45].

In addition to the direct involvement in protein-protein interactions occurring in infectious diseases, recent investigations suggest that PPII may facilitate host-pathogen interactions through other mechanisms. Indeed, PPII motifs may act as structural spacers needed for the correct positioning of the elements needed for adhesion and infectivity. One of the most striking examples of this possibility has been highlighted by the determination of the three-dimensional structure of the A3VP1 fragment of the adhesin antigen I/II (AgI/II) localized on the cell surface of the pathogen *Streptococcus mutans*, which is the causative agent of human dental caries. The crystallographic structure of this protein has unveiled a novel and unique structural motif, a hybrid structure composed of α- and PPII helices [46]. In particular, the A3 repeat of the alanine-rich domains of the protein adopts an extended α-helix that intertwines with the P1 repeat in PPII conformation, to form a fibrillar structure which is 155 Å long (Fig. 1). Interestingly, this motif is stabilized by interactions that are commonly associated to PPII protein-protein recognition. In particular, these include: (a) hydrogen bonding of non-saturated main chain atom of PPII, (b) hydrophobic interaction of exposed proline side chains with aromatic residues, and (c) water mediated hydrogen bonding networks. It is important to emphasize that in the case of AgI/II, the PPII helix acts as a structural element that is cru-
caciole to the building of the functional architecture of an essential protein in the pathogen adhesion to the tooth surface [46, 47].

**Fig. (1).** Representation of the interaction between the α-helix (light grey) and the PPII helix (dark grey) found in AgI/II structure (PDB code 3iox). For clarity, the PPII helix was reported as a stick model.

Another recent example of a possible structural role for PPII motifs has emerged from the crystallographic characterization of the Group A streptococcal basal pilin FctB, which presents an immunoglobulin-like N-terminal domain with an extended PPII tail [48]. This tail is a proline-rich fragment (residues 123–137) with sequence PXPPXXPXXPXXPXXP. Although terminal PPII fragments are expected to be highly mobile, the PPII helix of FctB is structurally well-defined, as indicated by its surprisingly clear electron density. Although the precise role for this motif is yet to be clearly defined, the PPII helix likely acts as a "spacer" ensuring that a cell wall-anchored protein actually protrudes from the thick bacterial peptidoglycan (PGN), an "spacer" ensuring that a cell wall-anchored protein actually protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains of essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglycan (PGN), an essential bacterial cell wall polymer, formed by glycan chains protrudes from the thick bacterial peptidoglyc...
confers to this motif a structural rigidity that is missing in PPII helices. These features play a crucial role in triple helix interaction with other biological actors.

MOLECULAR RECOGNITION OF COLLAGEN TRIPLE HELIX

Collagens are the most abundant proteins in mammals. The collagen family comprises 28 members that contain at least one triple-helical domain [76]. Collagens are deposited in the extracellular matrix where most of them generate supramolecular assemblies. Traditionally, collagens have been seen as structural entities that contribute to mechanical properties, to the organization and to the shape of tissues. More recent studies have highlighted that collagens are also involved in a variety of biological processes such as survival signals, morphogenic processes, and diseases.

Despite its limited sequence variability and the simple rod-like shape, collagens are promiscuous proteins that are able to interact with several receptor families to perform their regulatory function in cell proliferation, migration, and differentiation.

Collagen-interacting proteins are expressed by both eukaryotes and prokaryotes. Indeed, a number of transmembrane proteins, collectively designed as mammalian collagen–receptors [77], are able to interact with triple helical motifs. Interestingly, recent data show that not only is the triple helix used by pathogens as a target in the host-pathogen interaction but it may also act as a bait in these processes since bacterial proteins containing triple helix regions have been shown to interact with host proteins (Fig. 2) (see below for details).

Although a full understanding of the structural bases of the collagen recognition is yet to be achieved, collagen–receptor interaction generally occurs through two distinct mechanisms [77]. In some cases, as for example integrins and discoidin domain receptors and chaperone FKBP65, the receptor recognizes a specific collagen sequence. In other cases, as glycoprotein VI, CNA, YadA, saratin, and LAIR-1, the receptor simply recognizes and binds the basic triple helix motif [77-80].

Over the years a number of complexes involving triple helix peptides and their biological cognates have been elucidated by X-ray crystallography. So far the structure of seven complexes formed by triple helical peptides are reported in the Protein Data Bank (PDB). These complexes are formed with integrin alpha 2 I domain (PDB code 1dzi) [81], CNA (2f6a) [82], SPARC (2v53) [83], DDR2 discoidin domain (2wuh) [84], MASP-1 CUB2 domain (3pob) [85], Von Willebrand factor A3 domain (4dmu) [86], matrix metalloproteinase 1 (4auo) [87] and the chaperone Hsp47/SERPINH1 (4awr) [88]. These studies have provided important clues on collagen molecular recognition, although differing views have been reported for MMP1 binding [88, 89]. With the exclusion of the collagen binding protein CNA isolated from Staphylococcus aureus, these triple helix interacting proteins are of mammalian origin. Notably, the mammalian proteins target a specific sequence region of the triple helix motif, whereas CNA targets a consensus motif (Gly-Pro-Hyp). In all cases, the interacting surface between these proteins and the triple helix is not particularly large and involves residues of at least two of the three chains of the triple helix (Fig. 3). The specific features of the complex between triple helix and the protein CNA isolated from the pathogen S. aureus will be illustrated in the following sections.

HOST COLLAGEN TRIPLE HELIX AS A TARGET FOR BACTERIAL PATHOGENS

Bacterial adherence to host tissues represents an early critical step in the infection process. Extracellular pathogens, such as enterococci, staphylococci, and streptococci, often target extracellular matrix components for attachment and

![Fig. (2). Scheme ‘collagen acts as a target and a bait’.](Image)
Fig. (3). Crystal structures of complexes between proteins and collagen triple helices. Left panel reports sequences of bound triple helices: residues of the triple helices located within 4 Å from the receptor are reported in grey boxes. Right panels report ribbon representations of their three-dimensional structures. Panels A-G refer to complexes of triple helices with integrin alpha 2 I domain (PDB code 1dzi), CNA (2f6a), SPARC (2v53), DDR2 disocidin domain (2wuh), MASP-1 CUB2 domain (3pob), Von Willebrand factor A3 domain (4dmu), matrix metalloproteinase 1 (4auo), and the chaperone Hsp47/SERPINC1 (4awr) respectively.
colony. Adherence is mediated by protein adhesins of the MSCRAMM (microbial surface components recognizing adhesive matrix molecules) family. These adhesive proteins from Gram-positive bacteria do not share sequence homology with the collagen binding I domains of integrins and do not require metal ions for collagen binding. Consistently, these proteins employ a collagen binding mechanism that is drastically different from that of the collagen binding integrins.

These MSCRAMMs contain an N-terminal signal peptide followed by a non-repetitive region called the A region, which in most cases is responsible for ligand binding (Fig. 4). The A regions are composed of two or more subdomains each adopting an immunoglobulin G-like (IgG-like) fold. Following the A region is often a segment composed of repeated sequences or motifs that is referred to as the B region [90]. In most cases, MSCRAMMs are covalently anchored to the cell wall peptidoglycan (PGN) by the sortase through its C-terminal LPXTG motif (Fig. 4).

MSCRAM proteins mediate bacterial attachment to the host through interactions with either host fibronectin or collagen triple helix. The collagen-binding protein CNA is responsible for adherence of S. aureus to collagen substrates and collagenous tissues. Consistently, antibodies against CNA inhibit bacterial binding to collagen and block bacterial adherence to cartilage [91]. CNA participates in the infectious process of pathogenic S. aureus and is shown to be a virulence factor in many different animal models of infections [92,93]. This suggests that the ability to interact with collagen provides a general advantage to the bacteria in pathogenesis.

So far, the structure of a single MSCRAMM complex (CNA from S. aureus) with its collagen ligand has been determined experimentally through X-ray crystallography [94-96]. Structure of unligated MSCRAMM s have also been reported, see for example [90,96]. Altogether, these structures have provided fundamental clues on collagen recognition by bacterial proteins mediated by specific or non-specific collagen sequence motifs. Also, they have highlighted differences compared to collagen binding integrins. Indeed, the crystal structure of the complex between the I domain of integrin α2β1 and a triple helical collagen peptide unveiled that a collagen glutamate completes the coordination sphere of the metal of the integrin [81]. A completely different mechanism, denoted as ‘collagen hug mechanism’ was proposed for collagen recognition by bacterial proteins.

The crystal structure of CNA exhibits two distinct domains (called N1 and N2) (Fig. 4). The N-terminal N1 domain exhibits an IgG-like fold, which contains two additional strands D’ and D” compared to the conventional IgG fold [94,97]. The N2 domain is similar to the previously determined crystal structure of CNA151–318 [98], which contains in addition to the D2’ and D2” strands, an extra D2”’ strand and a two-turn α-helix. The N1 domain that is completed by the C-terminal extension of the N2 domain (Fig. 4B). This organization creates a two-domain structure with a distinct hole between the two domains (Fig. 4B). The collagen triple helix is seen penetrating through the hole, of about 12 to 15Å in diameter, between the N1 and N2 domains of CNA (Fig. 4B).

Fig (4). (A) domain organisation of MSCRAMM proteins CNA and ACE: The N-terminal signal peptide (SP) is followed by A and B regions, the C-terminal cell wall sorting region W, the transmembrane region M and the cytoplasmic tail C. (B) CNA structure in complex with the collagen triple helix. The N1 and N2 domains of the A region are shown in light and dark gray, respectively. Collagen triple helix (black) is wrapped by the linker region between N1 and N2, represented in ball-and-stick.

Since the closed conformation cannot bind the collagen triple helix, the ‘Collagen Hug model’ proposes that the initial event of the collagen recognition requires that the apoprotein exists in equilibrium between an open and closed conformation, and that only the open conformation can bind collagen [95]. The following steps of the mechanism closely resemble those of the interaction mode reported for MSCRAMM recognition of fibrinogen, and denoted as the ‘Dock, Lock and Latch’ model (Fig. 5).

The binding is guided by sequential interactions of the collagen triple helix with the N2 domain with the interdomain linker and eventually with the repositioned N1 domain. The collagen peptide ligand is further secured by conformational changes of the linker region, which keeps the collagen triple helix tightly in place. Namely, MSCRAMM appears to wrap around the collagen triple helix where the N1 and N2 subdomains of CNA create a “tunnel-like” structure that “locks” the collagen ligand in between the two subdomains. In a final step, the C-terminal extension of the N2 subdomain acts as a “latch” by inserting into a trench present on the N1 subdomain by β-strand complementation [99]. Molecular dynamics simulations carried out on CNA suggest that the incomplete Ig-like fold of CNA N2 generated by the displacement of the C-terminal extension is sufficiently stable to support the Collagen Hug mechanism and also evidence a key role of collagen hydration in CNA recognition [100]. In particular, hydration maps of the CNA-collagen complex reveal the presence of several structured water molecules that mediate intermolecular interactions at the interface between the two proteins. These hydration sites feature long residence times, significant binding free energies, and a
geometrical distribution that closely resembles the hydration pattern of the isolated collagen triple helix. Therefore, CNA recognizes the collagen triple helix as a hydrated molecule [100].

Fig. (5). Collagen hug mechanism. (A), CNA open and close forms exist in equilibrium; (B) Collagen binds to the N2 domain; (C) collagen is wrapped by N1 and N2 domains and by the inter-domain linker.

Structural studies of the MSCRAMM protein ACE, the first to be identified on E. faecalis [101], showed a structural organization similar to that of the Staphylococcus aureus collagen adhesin CNA [90]. These studies corroborated the ‘Collagen Hug’ model, as they showed that point mutants which were stabilizing a closed conformation (by a disulfide bond) did not bind type I collagen. In addition, they identified key residues for collagen binding activity.

Evidence that bacterial matrix-binding proteins are virulence factors has come from studying defective mutants in adherence assays. Mutants lacking the collagen-binding protein have reduced virulence in a mouse model for septic arthritis, suggesting that bacterial colonization is ineffective. Furthermore, the isolated ligand-binding domain of the fibrinogen, fibronectin and collagen receptors strongly blocks attachment of bacterial cells to the corresponding host proteins [102,103]. The key role of collagen-binding adhesins in bacterial infections has encouraged the development of strategies to specifically block the interaction of bacteria with matrix collagen by antagonist ligands. For example, recombinant adhesion fragments and polyclonal antibodies may inhibit bacterial adhesion to host cells to varying degrees [104]. Also, specific antibodies against high-affinity binding subsegments of the MSCRAMM protein ACM were the most effective at inhibiting E. faecium adherence to host collagen [105]. Further investigations are still needed to advance our understanding of the mechanisms of bacterial adhesion and ways to hamper them.

**BACTERIAL COLLAGEN TRIPLE HELIX AS A BAIT FOR HOST-PATHOGEN INTERACTION**

Although type I is the most widespread collagen, invertebrates were shown to contain several other types of collagen genes [106]. More recently, collagen-like sequences have been identified also in prokaryotic genomes [107-111]. These collagen-like proteins have been shown to adopt a triple helix structure, with a thermal stability similar to that seen for human collagens [112-114]. However, unlike mammalian collagens, prokaryotic collagens do not contain hydroxyproline, since bacteria lack the prolyl-hydroxylase enzyme necessary for post-translational modification of Pro to Hyp.

Best characterized prokaryotic collagens are the two collagen-like proteins, Scl1 and Scl2, which have been demonstrated to be simultaneously expressed on the cell surface of *Streptococcus pyogenes* and to be able to promote bacterial adhesion to the host [107,108]. Both Scl1 and Scl2 proteins contain a signal sequence, an N-terminal variable globular domain (V), a highly charged collagen-like triple-helix domain (CL) consisting of (Gly-Xaa-Yaa)n triplet repeats and a C-terminal Gram-positive cell wall attachment domain (Fig. 6). The Scl1 and Scl2 proteins form stable triple-helical structures when expressed as recombinant proteins [115,116], and their N-terminal globular V domain adjacent to the triple-helix domain appears to be important for efficient triple-helix assembly [116, 117]. In *Streptococcus pyogenes* the CL domain can be dissected into three fragments of almost equal size with distinctive amino acid features (A, B and C). Fragment A has the highest content of polar residues, fragment B the highest Pro content, and fragment C the highest charged residue content.

**Fig. (6).** Domain organization of Scl proteins main features are: an N-terminal globular (V) domain, a collagen-like (CL) domain, and a C-terminal Gram-positive cell wall attachment domain (Gram+ anchor).

Bacterial collagen-like proteins [114] contain no Hyp, and their Pro residues are found preferentially in the Xaa positions, with the frequency of Pro residues in the X position exceeding 30%, whereas the frequency of Pro in the Y position is 5%. Despite the lack of Hyp, however, bacterial collagen-like proteins form triple-helix structures with a stability near 37 °C, close to that seen for mammalian collagens. In theory, bacteria could use Gly-Pro-Pro tripeptide units, which are highly stabilizing sequences, but Gly-Pro-Pro tripeptides are rarely found in bacterial collagen-like proteins. Interestingly, almost all bacterial collagens studied so far have melting temperatures between 36 and 38 °C, indicating that this narrow range of melting temperatures might have been evolutionarily set for host-pathogen interaction [112-114]. To explore the basis of bacterial collagen
triple-helix stability in the absence of Hyp, biophysical studies were carried out on recombinant Scl2 protein and a set of peptides modeling the Scl2 highly charged repetitive (Gly-X-Y)n sequences [116]. These studies showed that triple helix stabilization is highly pH dependent, and that the Scl2 protein uses a variety of electrostatic interactions, inter-chain hydrogen bonds, and a hydration-mediated hydrogen bonding network as an alternative to the Hyp stabilization in animal collagens [116].

Scl proteins are key to host-pathogen recognition. It has been demonstrated that SclI can bind selected human extracellular matrix components [109], cellular integrin receptors [107,118,119], and plasma components [120-122]. Importantly, human collagen receptors such as integrins α2β1 and α11β1 recognize the triple helix CL domain of Scl1, and this event results in cell signaling [107,118,119]. This finding is a strong indication that collagen-like bacterial proteins display not only structural but also functional similarities to human collagens [107-110].

FINAL CONSIDERATIONS

A survey literature data clearly indicates that PPII and triple helix motifs are able to establish interactions with a plethora of biological partners, despite their rather low sequence variability. Indeed, the over-representation of proline and proline derivatives in their sequences does not preclude to these elements the possibility to regulate diversified biological processes including those related to infections by external pathogens. It is important to note both PPII and triple helices exploit common features to establish protein-protein interactions. The first one is represented by the presence in both elements of exposed and non-satisfied main chain hydrogen bonding donor and acceptor groups. These groups may either directly bind the interacting proteins or firmly bind water molecules that, in turn, make hydrogen bonding interactions with the partner. One of the emerging ideas in the fields is that the hydration pattern of the isolated partners is, at least partially, preserved upon the formation of complexes involving PPII and triple helix. In near future, computational approaches, as those currently used for small molecules [123], may likely provide (a) interesting structural insights into these processes and (b) quantitative estimates of the energetics of water binding. The second feature of these elements is the exposure of hydrophobic side chains (especially proline residues) that are easily recognized by hydrophobic patches present on the receptor. The combination of these hydrophilic and hydrophobic patches in PPII and triple helix likely regulates the specificity of their partner recognition. It may also explain the paradox between specificity versus promiscuity of triple helix and PPII recognition.

In this scenario, the specific assembly of PPII helices in the more structurally rigid triple helix preserves the exposure of the main reactive groups of this motif. The combination of rigidity and solvent-exposure of reactive groups makes triple helix suitable for both structural and recognition roles. Finally, it should be noted that despite the emerging interest in host-pathogen recognition processes mediated by either PPII or the triple helix, the amount of structural information available is still limited. More detailed information about receptor-binding capabilities and mechanisms of different collagen-like bacterial and mammalian proteins will help us to exploit these interactions for therapeutic purposes.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

The authors thank Luca De Luca for his precious technical support. The authors also thank the UE (COST Action BM1003) and the Mizutani Foundation for glycoscience for financial support.

ABBREVIATIONS

PPII = Polyproline II
PDB = Protein Data Bank

REFERENCES

[1] Richardson, J.S. The anatomy and taxonomy of protein structure. Adv. Prot. Chem., 1981, 34, 167-339.
[2] Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers, 1983, 22, 2577-2637.
[3] Unger, R.; Sassman, J.L. The importance of short structural motifs in protein structure analysis. J. Comp.-Aid. Mol. Des., 1993, 7, 457-472.
[4] Brenner, S.E.; Chothia, C.; Hubbard, T.J. Population statistics of protein structures: Lessons from structural classifications. Curr. Opin. Struct. Biol., 1997, 7, 369-376.
[5] De Simone, A.; Berisio, R.; Zagari, A.; Vitagliano, L. Limited tendency of alpha-helical residues to form disulfide bridges: A structural explanation. J. Pept. Sci., 2006, 12, 740-747.
[6] Adzhubei, A.A.; Sternberg, M.J. Left-handed polyproline ii helices commonly occur in globular proteins. J. Mol. Biol., 1993, 229, 447-493.
[7] Stapley, B.J.; Creamer, T.P. A survey of left-handed polyproline ii helices. Prot. Sci., 1999, 8, 587-595.
[8] Cubells, M.V.; Caillé, F.; Blundell, T.L.; Lovell, S.C. Properties of polyproline ii, a secondary structure element implicated in protein-protein interactions. Proteins, 2005, 58, 880-892.
[9] Berisio, R.; Loguercio, S.; De Simone, A.; Zagari, A.; Vitagliano, L. Polyproline helices in protein structures: A statistical survey. Prot. Pept. Lett., 2006, 13, 847-854.
[10] Barlow, D.J.; Thornton, J.M. Helix geometry in proteins. J. Mol. Biol., 1988, 201, 601-619.
[11] Richardson, J.S.; Richardson, D.C. Helix lap-joints as ion-binding sites: DNA-binding motifs and ca-binding "of hands" are related by charge and sequence reversal. Proteins, 1988, 4, 229-239.
[12] Okuyama, K. Revisiting the molecular structure of collagen. Connect. Tiss. Res., 2008, 49, 299-310.
[13] Brodsky, B.; Persikov, A.V. Molecular structure of the collagen triple helix. Adv. Prot. Chem., 2005, 70, 301-339.
[14] Shoulders, M.D.; Raines, R.T. Collagen structure and stability. Annu. Rev. Biochem., 2009, 78, 929-958.
[15] Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. Recent progress on collagen triple helix structure, stability and assembly. Prot. Pept. Lett., 2002, 9, 107-116.
[16] Richardson, J.S.; Richardson, D.C. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc. Nat. Acad. Sci. USA, 2002, 99, 2754-2759.
[17] De Simone, A.; Esposito, L.; Pedone, C.; Vitagliano, L. Insights into stability and toxicity of amyloid-like oligomers by replica exchange molecular dynamics analyses. Biophys. J., 2008, 95, 1965-1973.
[18] Esposito, L.; Paladino, A.; Pedone, C.; Vitagliano, L Insights into structure, stability, and toxicity of monomeric and aggregated polyglutamine models from molecular dynamics simulations. Biophys. J., 2008, 94, 4031-4040.
De Simone, A.; Dodson, G.G.; Fraternali, F.; Zagari, A. Water molecules as structural determinants among prions of low sequence identity. *FEBS Lett.*, 2006, 580, 2488-2494.

Nagai, Y.; Popiel, H.A. Conformational changes and aggregation of expanded polyglutamine proteins as therapeutic targets of the polyglutamine diseases: Exposed beta-sheet hypothesis. *Curr. Pharmaceut. Des.*, 2008, 14, 3267-3279.

de Vega, M.J.; Martin-Martinez, M.; Gonzalez-Muniz, R. Modulation of protein-protein interactions by stabilizing/mimicking protein secondary structure elements. *Curr. Topics Med. Chem.*, 2007, 7, 33-62.

Tsai, C.J.; Lin, S.L.; Wolfson, H.J.; Nussinov, R. Protein-protein interfaces: Architectures and interactions in protein-protein interfaces and in protein cores. Their similarities and differences. *Crit. Rev. Biochem. Mol. Biol.*, 1996, 31, 127-152.

Kay, B.K.; Williamson, M.P.; Sudol, M. The importance of being proline: The interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.*, 2000, 14, 231-241.

Siligardi, G.; Drake, A.F. The importance of extended conformations and, in particular, the piii conformation for the molecular recognition of peptides. *Biopolymers*, 1995, 37, 281-292.

MacArthur, M.W.; Thornton, J.M. Influence of proline residues on protein conformation. *J. Mol. Biol.*, 1991, 218, 397-412.

Rath, A.; Davidson, A.R.; Deber, C.M. The structure of “unstructured” regions in peptides and proteins: Role of the polyproline ii helix in protein folding and recognition. *Biopolymers*, 2005, 80, 179-185.

Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A. Preferred proline puckercings in cis and trans peptide groups: Implications for collagen stability. *Prot. Sci.*, 2001, 10, 2627-2632.

Cremer, T.P.; Campbell, M.N. Determinants of the polyproline ii helix from modeling studies. *Adv. Prot. Chem.*, 2002, 62, 263-282.

Shi, Z.; Woody, R.W.; Kallenbach, N.R. Is polyproline ii a major backbone conformation in unfolded proteins? *Adv. Prot. Chem.*, 2002, 62, 163-240.

Shi, Z.; Olson, C.A.; Rose, G.D.; Baldwin, R.L.; Kallenbach, N.R. Polyproline ii structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. USA.*, 2002, 99, 19190-19195.

Eker, F.; Cao, X.; Nafie, L.; Schweitzer-Stenner, R. Tripeptides adopt stable structures in water. A combined polarized visible raman, flir, and vcd spectroscopy study. *J. Am. Chem. Soc.*, 2002, 124, 14330-14341.

Avbelj, F.; Grdadolnik, S.G.; Grdadolnik, J.; Baldwin, R.L. Intrinsic backbone preferences are fully present in blocked amino 124, 14330-14341.

Zafra-Ruano, A.; Luque, I. Interfacial water molecules in sh3 domains: Getting the full picture on polyproline recognition by protein-protein interaction domains. *FEBS lett.*, 2012, 586, 2619-2630.

Agrawal, V.; Ishan, K.V. Promiscuous binding nature of sh3 domains to their target proteins. *Proteo. Lett.*, 2002, 9, 185-193.

Srinivasan, M.; Dunker, A.K. Proline rich motifs as drug targets in immune mediated disorders. *Int. J. Pept., 2012, 12, 634769.

Niebuhr, K.; Ebel, F.; Frank, R.; Reinhard, M.; Domann, E.; Carl, U.D.; Walter, U.; Gertler, F.B.; Wehland, J.; Chakraborty, T. A novel proline-rich motif present in acta of listeria monocytogenes. Y. Sh3 binding motif 1 in influenza a virus ns1 protein is essential for pi3k/akt signaling pathway activation. *J. Virol.*, 2007, 81, 12730-12739.

Linke, C.; Young, P.G.; Kang, H.J.; Bunker, R.D.; Middleditch, M.J.; Caradoc-Davies, T.T.; Proft, T.; Baker, E.N. Crystal structure of the minor pilin fcb reveals determinants of group a streptococcal pilus anchoring. *J. Biol. Chem.*, 2010, 285, 20381-20390.

Solovyova, A.S.; Pointon, J.A.; Race, P.R.; Smith, W.D.; Kehoe, M.A.; Banfield, M.J. Solution structure of the major (spy0128) and minor (spy0125 and spy0130) pili subunits from streptococcus pyogens. *Euro. Biophys. J. : EBJ.*, 2010, 39, 469-480.

Holmgren, S.K.; Taylor, K.M.; Bretscher, L.E.; Raines, R.T. Code for collagen's stability deciphered. *Nature*, 1998, 392, 666-667.

Rath, A.; Davidson, A.R.; Deber, C.M. The structure of “unstructured” regions in peptides and proteins: Role of the polyproline ii helix in protein folding and recognition. *Biopolymers*, 2005, 80, 179-185.

Bradly, L.J.; Maddocks, S.E.; Larson, M.R.; Forsgren, N.; Persson, K.; Deivanayagam, C.C.; Jenkinson, H.F. The changing faces of molecular recognition of peptides containing 4-fluoroproline; (4(s)-fluoroproline-pro-gly)10. *J. Pept. Sci.*, 2006, 12, 5983-5988.

Ohkubo, T.; Kobayashi, Y. Characterization of collagen model triple-helical peptides. *Biopolymers*, 2001, 58, 459-464.

Berisio, R.; Granata, V.; Vitagliano, L.; Zagari, A. Imino acids and collagen triple helix stability: Characterization of collagen-like polyproline peptides containing hyp-hyp-gly sequence repeats. *J. Am. Chem. Soc.*, 2006, 128, 607-616.

Rath, A.; Davidson, A.R.; Deber, C.M. The structure of “unstructured” regions in peptides and proteins: Role of the polyproline ii helix in protein folding and recognition. *Biopolymers*, 2005, 80, 179-185.

Zafra-Ruano, A.; Luque, I. Interfacial water molecules in sh3 domains: Getting the full picture on polyproline recognition by protein-protein interaction domains. *FEBS lett.*, 2012, 586, 2619-2630.

Niebuhr, K.; Ebel, F.; Frank, R.; Reinhard, M.; Domann, E.; Carl, U.D.; Walter, U.; Gertler, F.B.; Wehland, J.; Chakraborty, T. A novel proline-rich motif present in acta of listeria monocytogenes. *J. Virol.*, 2007, 81, 12730-12739.

Okuyama, K.; Xu, X.; Iguchi, M.; Noguchi, K. Revision of collagen molecular structure. *J. Mol. Biol.*, 2007, 372, 1258-1264.

Okuyama, K.; Arnott, S.; Takayanagi, M.; Kakudo, M. Crystal and molecular structure of a collagen-like polypeptide (pro-pro-gly)10. *J. Am. Chem. Soc.*, 2003, 125, 11500-11501.
Leo, J.C.; Elovaara, H.; Brodsky, B.; Skurnik, M.; Goldman, A.; Gronwald, W.; Bomke, J.; Maurer, T.; Domogalla, B.; Huber, F.; Ishikawa, Y.; Vranka, J.; Wirz, J.; Nagata, K.; Bachinger, H.P.; Kawahara, K.; Nishi, Y.; Nakamura, S.; Uchiyama, S.; Nishiuchi, E.; Emsley, J.; Knight, C.G.; Farndale, R.W.; Barnes, M.J.; Leitinger, B.; Hohenester, E. Mammalian collagen receptors.

Improta, R.; Berisio, R.; Vitagliano, L. Contribution of dipole-dipole interactions to the stability of the collagen triple helix. Prot. Sci., 2009, 18, 2067-2076.

Perskov, A.V.; Ramshaw, J.A.; Brodsky, B. Collagen model peptides: Sequence dependence of triple-helix stability. Biopolymers, 2000, 55, 436-450.

Fallas, J.A.; Lee, M.A.; Jalan, A.A.; Hartgerink, J.D. Rational design of single-composition abe collagen heterotrimers. J. Am. Chem. Soc., 2012, 134, 1430-1433.

Ogel, J.P.; Irving, T.C.; Miller, A.; Wess, T.J. Microfibrillar structure of type i collagen in situ. Proc. Nat. Acad. Sci. USA., 2006, 103, 9001-9005.

Okuyama, K.; Bachinger, H.P.; Mizuno, K.; Boudko, S.; Engel, J.; Berisio, R.; Vitagliano, L. Re: Microfibrillar structure of type i collagen in situ. Acta crystallographica. Section D, Biol. Crystallograph., 2009, 65, 1007-1008; author reply 1009-1010.

Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. Crystal structure of a collagen-like polypeptide with repeating sequence pro-hyp-gly at 1.4 a resolution: Implications for collagen hydration. Biopolymers, 2000, 56, 8-13.

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, 75-81.

Kramer, R.Z.; Bella, J.; Mayville, P.; Brodsky, B.;berman, H.M. Sequence dependent conformational variations of collagen triple-helical structure. Nat. Struct. Biol., 1999, 6, 454-457.

Shoulders, M.D.; Satyshur, K.A.; Forest, K.T.; Raines, R.T. Structural and steric effects in side chains preorganize a protein main chain. Proc. Nat. Acad. Sci. USA., 2010, 107, 559-564.

Boudko, S.P.; Engel, J.; Okuyama, K.; Mizuno, K.; Bachinger, H.P.; Schumacher, M.A. Crystal structure of human type iii collagen gly991-gly1032 cystine knot-containing peptide shows both 7/2 and 10/3 helical symmetries. J. Biol. Chem., 2008, 283, 32580-32589.

Okuyama, K.; Miyama, K.; Morimoto, T.; Masakiyo, K.; Mizuno, K.; Bachinger, H.P. Stabilization of triple-helical structures of collagen peptides containing a hyp-thr-gly, hyp-val-gly, or hyp-ser-gly sequence. Biopolymers, 2011, 95, 628-640.

Kawahara, K.; Nishi, Y.; Nakamura, S.; Uchiyama, S.; Nishituchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y. Effect of hydration on the stability of the collagen-like triple-helical structure of [4(r)-gly sequence. Biopolymers, 2012, 913-927.

Gronwald, W.; Mayville, P.; Brodsky, B.; Herman, H.M. Crystal and molecular structure of a collagen-like peptide at 1.9 a resolution. Science, 1994, 266, 75-81.

Kramer, R.Z.; Bella, J.; Mayville, P.; Brodsky, B.; Herman, H.M. Sequence dependent conformational variations of collagen triple-helical structure. Nat. Struct. Biol., 1999, 6, 454-457.

Shoulders, M.D.; Satyshur, K.A.; Forest, K.T.; Raines, R.T. Structural and steric effects in side chains preorganize a protein main chain. Proc. Nat. Acad. Sci. USA., 2010, 107, 559-564.

Boudko, S.P.; Engel, J.; Okuyama, K.; Mizuno, K.; Bachinger, H.P.; Schumacher, M.A. Crystal structure of human type iii collagen gly991-gly1032 cystine knot-containing peptide shows both 7/2 and 10/3 helical symmetries. J. Biol. Chem., 2008, 283, 32580-32589.

Okuyama, K.; Miyama, K.; Morimoto, T.; Masakiyo, K.; Mizuno, K.; Bachinger, H.P. Stabilization of triple-helical structures of collagen peptides containing a hyp-thr-gly, hyp-val-gly, or hyp-ser-gly sequence. Biopolymers, 2011, 95, 628-640.

Kawahara, K.; Nishi, Y.; Nakamura, S.; Uchiyama, S.; Nishituchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y. Effect of hydration on the stability of the collagen-like triple-helical structure of [4(r)-gly sequence. Biopolymers, 2012, 913-927.

Gronwald, W.; Mayville, P.; Brodsky, B.; Herman, H.M. Crystal and molecular structure of a collagen-like peptide at 1.9 a resolution. Science, 1994, 266, 75-81.

Kramer, R.Z.; Bella, J.; Mayville, P.; Brodsky, B.; Herman, H.M. Sequence dependent conformational variations of collagen triple-helical structure. Nat. Struct. Biol., 1999, 6, 454-457.

Shoulders, M.D.; Satyshur, K.A.; Forest, K.T.; Raines, R.T. Structural and steric effects in side chains preorganize a protein main chain. Proc. Nat. Acad. Sci. USA., 2010, 107, 559-564.

Boudko, S.P.; Engel, J.; Okuyama, K.; Mizuno, K.; Bachinger, H.P.; Schumacher, M.A. Crystal structure of human type iii collagen gly991-gly1032 cystine knot-containing peptide shows both 7/2 and 10/3 helical symmetries. J. Biol. Chem., 2008, 283, 32580-32589.
