

# 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  $\alpha$ -helix and  $\beta$ -structure [2-5]. Less common motifs are 3(10)- and  $\pi$ -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  $\beta$ -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  $\alpha$ -helices and  $\beta$ -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  $C^{\delta}$  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  $\varphi$  angle is restricted to values of  $-65^{\circ} \pm 15^{\circ}$  [24-26]. In addition, the particular location of the  $C^{\delta}$  atom also restricts the  $\psi$  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 ( $\varphi$ ,  $\psi$ ) backbone dihedral angles with average values of ( $-75^{\circ}$ ,  $145^{\circ}$ ) [6-10,28]. These extended helices have three residues per turn, with a pitch value of  $9.3 \text{ \AA/turn}$ .

\*Address correspondence to this author at the Istituto di Biostrutture e Bioimmagini, CNR via Mezzocannone 16, I-80134 Napoli, Italy; Tel: 003990812534506; Fax: 00390812536642; Email: [luigi.vitagliano@unina.it](mailto: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 (polyproline 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  $n \rightarrow p^*$  interaction between the nonbonding  $n$ -orbital of the its 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 intermolecular 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.

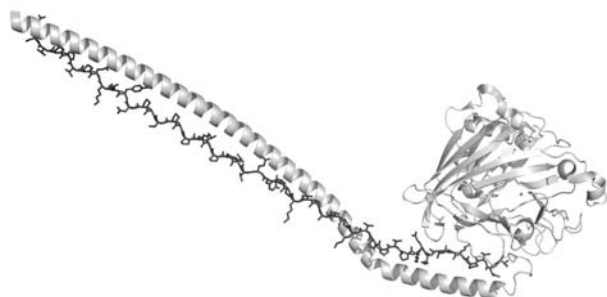
The interaction of ActA with Mena EVH1 is the molecular event that allows the bacterium to utilize the actin cytoskeleton to move and infect neighboring cells [38-40]. Since the inhibition of this interaction may provide a strategy to hamper the motility of the pathogen and to prevent its spread in the 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 prototypal 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 polyproline segment (P69VRPQVPLRP78) which contains the helical SH3 domain binding protein motif, PXXPR [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 through 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  $\alpha$ - and PPII helices [46]. In particular, the A3 repeat of the alanine-rich domains of the protein adopts an extended  $\alpha$ -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-

cial 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  $\alpha$ -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 streptococci 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 PXPPXXPPXXPPXXP. Although terminal PPII fragments are expected to be high 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 essential bacterial cell wall polymer, formed by glycan chains of  $\beta$ -(1-4)-linked-N-acetylglucosamine and N-acetylmuramic acid cross-linked by short peptide stems. Attachment to the PGN is mediated by a region called LPXTG motif, which is cleaved between the Thr and Gly residues by a transpeptidase called sortase and the carboxyl group of Thr is covalently joined to the carboxyl group of a PGN peptide stem. In this context, the PPII helix seems to be important to separate the immunoglobulin-like domain from the LPXTG sortase cleavage motif and leave space for the sortase protein to act.

The presence of a proline-rich C-terminal tail is not unique to pilin proteins and can be found in other cell wall-anchored proteins [49]. Because only one pilin protein per pilus operon appears to have a proline-rich tail, Baker and colleagues [48] have proposed that such domains preceding a sortase motif is a common structural element involved in cell wall anchoring.

### TRIPLE HELIX: BASIC STRUCTURAL FEATURES

Triple helix is a structural motif characterized by the association of three distinct polypeptide chains wrapped around a common axis [12-15]. Each of these chains assumes a PPII-like conformation. The structural features of the triple helices requires the presence of Gly residues in the inner region of the helix, every three residues of the polypeptide sequences. Therefore, sequences adopting these structural elements are characterized by the repetition of triplets

of the type Gly-Xaa-Yaa [12-15]. Although all types of amino acids may be located at positions X and Y of the triplets, they are frequently occupied by iminoacids [Pro and its post-translationally modified form 4R-hydroxy-2S-proline (4RHyp)]. The role of iminoacids in triple helix stabilization is a subject of intense research activities that has been reviewed and discussed in several literature reports [12-15,50-60].

Collagens are prototypal triple helix containing proteins. However, the triple-helical motif is also found in a variety of non-collagenous proteins, such as macrophage scavenger receptors types I and II and bacteria-binding receptor MARCO, the complement component C1q, pulmonary surfactant apoproteins A and D, acetylcholinesterase, bovine conglutinin, collectin-43, ficolins, aggrexin, ectodysplasin, mannose binding protein and in several proteins isolated by invertebrates and bacteria (see below) [61].

The triple helix is stabilized by the formation of an intermolecular hydrogen bond between the nitrogen of the Gly residue and the carbonyl group of the residue located in the X position. Further contribution to the triple helix stabilization is occasionally provided by intermolecular interactions established by specific side chains [12-15,62-66].

The non-globular shape of triple helix regions has hitherto prevented high resolution structural characterization of entire proteins carrying this structural motif [see for example ref [67,68]]. Available structural information on the triple helix have been essentially achieved through the characterization of peptide models in triple helix conformation [12-15,62,69]. High resolution studies carried out on these models have provided a detailed three dimensional picture of triple helix. Indeed, these investigations have unveiled a number of important structural details which have proven to be essential for a full understanding of triple helix stability and its possible interactions modes with solvent or other biological partners. As anticipated by initial triple helix models, in PPII side chains of residues located in positions X and Y of repetitive triplets are fully exposed to the solvent. Since these positions are often occupied by iminoacids (Pro or 4RHyp), whose side chain is partially aliphatic, one of the distinctive features of triple helix is the exposure of hydrophobic region that presents a strong propensity to participate to intermolecular interactions. High resolution studies have also shown (a) that the position X and Y of the triple helix are not conformationally equivalent and (b) triple helix geometrical parameters may locally change to generate helices with variable symmetries ranging from  $7_5$  or to  $10_7$ . The implications of these latter findings have been extensively described in previous literature papers [12-15,62,69-75]. Crystallographic studies have also provided evidence that triple helix is highly hydrated. The first hydration shell is composed by waters that interact either with non-saturated main chain hydrogen bonding donors and acceptors (nitrogens and carbonyl oxygens) or with side chain hydrophilic groups. Usually, other layers of water molecules are bound to those of the first shell to generate complex networks that are able to connect different triple helices in the crystalline state.

In conclusion, the triple helix preserves some of the features of isolated PPII helices (solvent exposure of non-satisfied hydrogen bond acceptors or donors and of hydrophobic residues). The tight association of the three chains

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).

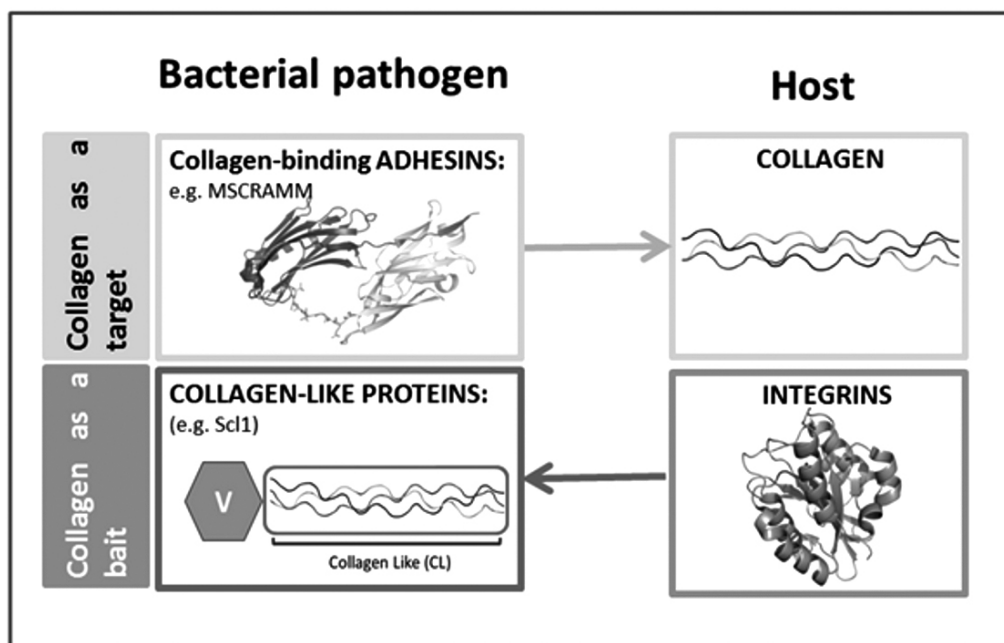
Although a full understanding of the structural bases of the collagen recognition is yet to be achieved, collagen-receptor interaction generally occurs through two distinct mechanisms [77]. In some cases, as for example integrins and discoidin domain receptors and chaperone FKBP65, the

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].

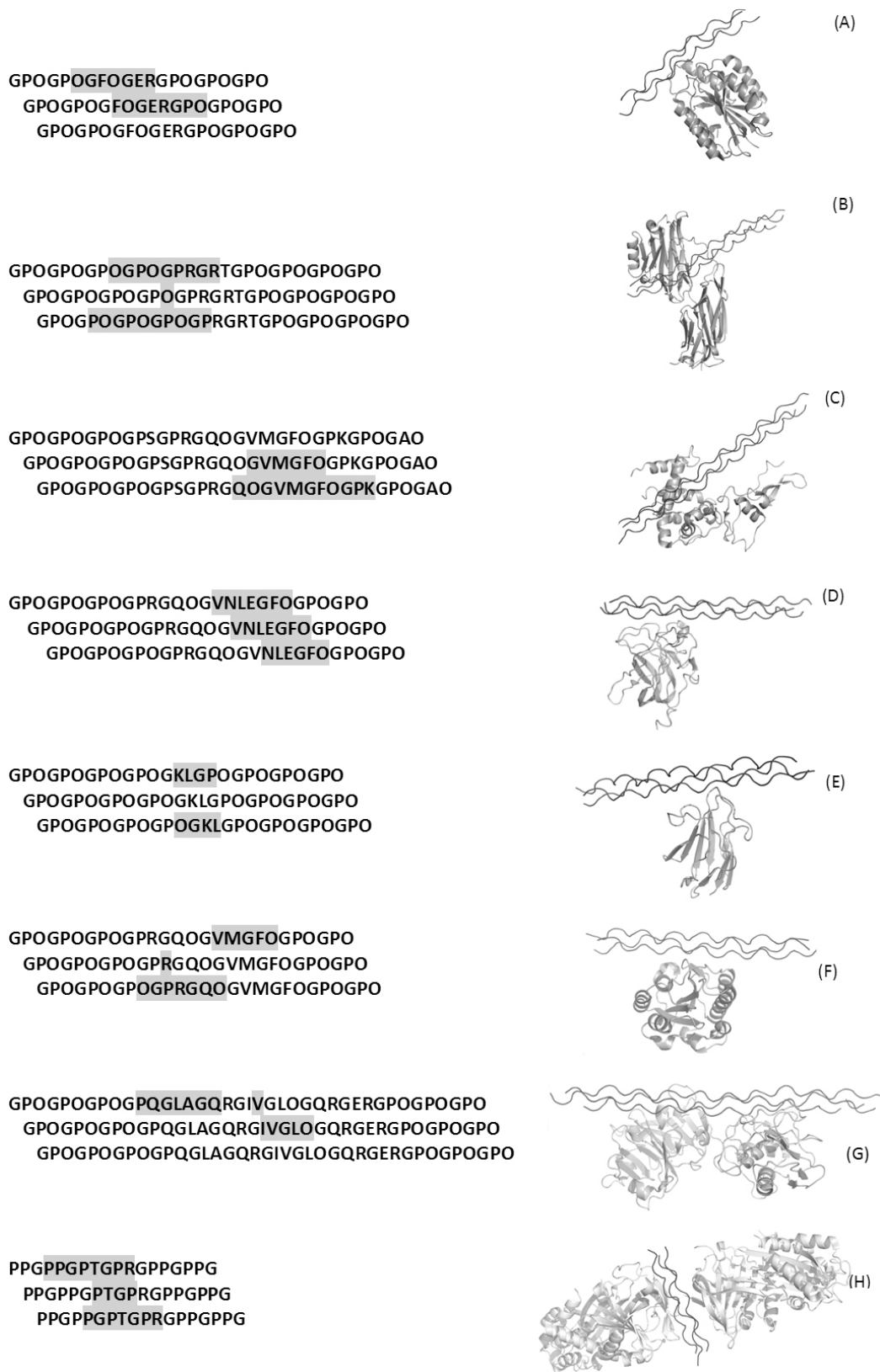
Over the years a number of complexes involving triple helix peptides and their biological cognates have been elucidated by X-ray crystallography. So far the structure of seven complexes formed by triple helical peptides are reported in the Protein Data Bank (PDB). These complexes are formed with integrin 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'.



**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 discoidin domain (2wuh), MASP-1 CUB2 domain (3pob), Von Willebrand factor A3 domain (4dmu), matrix metalloproteinase 1 (4auo), and the chaperone Hsp47/SERPINH1 (4awr) respectively.

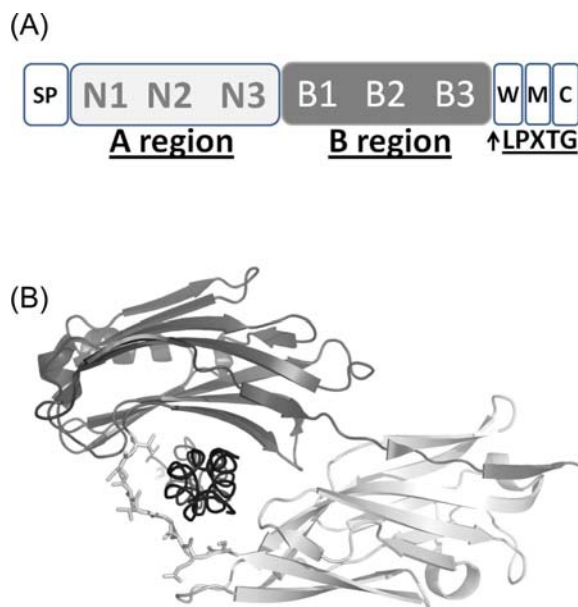
colonization. Adherence is mediated by protein adhesins of the MSCRAMM (microbial surface components recognizing adhesive matrix molecules) family. These adhesive proteins from Gram-positive bacteria do not share sequence homology with the collagen binding I domains of integrins and do not require metal 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).

MSCRAMM 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  $\alpha 2\beta 1$  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  $\alpha$ -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 can (Fig. 4B).

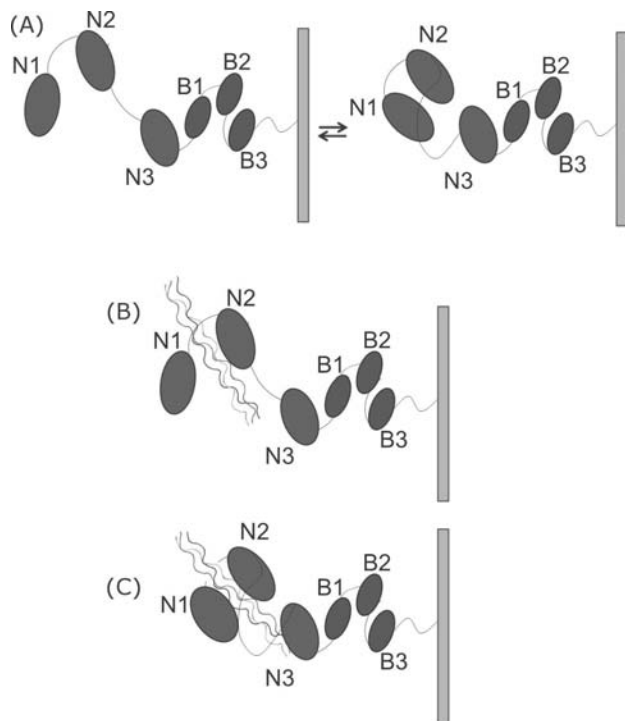


**Fig. (4).** (A) domain organisation of MSCRAMM proteins CNA and ACE: The N-terminal signal peptide (SP) is followed by A and B regions, the C-terminal cell wall sorting region W, the transmembrane region M and the cytoplasmatic tail C. (B) CNA structure in complex with the collagen triple helix. The 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 apo-form exists in equilibrium between an open and closed conformation, and that only the open conformation can bind collagen (Fig. 5) [95]. The following steps of this 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  $\beta$ -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.

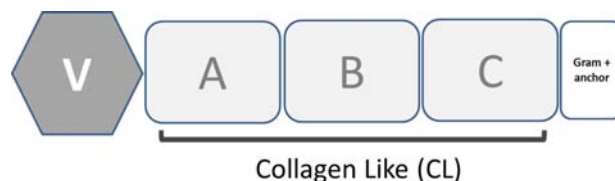
Evidence that bacterial matrix-binding proteins are virulence factors has come from studying defective mutants in adherence assays. Mutants lacking the collagen-binding protein have reduced virulence in a mouse model for septic arthritis, suggesting that bacterial colonization is ineffective. Furthermore, 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 ad-

vance 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)<sub>n</sub> 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)<sub>n</sub> 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 Scl1 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  $\alpha 2\beta 1$  and  $\alpha 11\beta 1$  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

- Richardson, J.S. The anatomy and taxonomy of protein structure. *Adv. Prot. Chem.*, **1981**, *34*, 167-339.
- Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **1983**, *22*, 2577-2637.
- Unger, R.; Sussman, J.L. The importance of short structural motifs in protein structure analysis. *J. Comp.-Aid. Mol. Des.*, **1993**, *7*, 457-472.
- 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.
- 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.
- Adzhubei, A.A.; Sternberg, M.J. Left-handed polyproline ii helices commonly occur in globular proteins. *J. Mol. Biol.*, **1993**, *229*, 472-493.
- Stapley, B.J.; Creamer, T.P. A survey of left-handed polyproline ii helices. *Prot. Sci.*, **1999**, *8*, 587-595.
- Cubellis, M.V.; Caillez, 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.
- 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.
- Barlow, D.J.; Thornton, J.M. Helix geometry in proteins. *J. Mol. Biol.*, **1988**, *201*, 601-619.
- Richardson, J.S.; Richardson, D.C. Helix lap-joints as ion-binding sites: DNA-binding motifs and ca-binding "ef hands" are related by charge and sequence reversal. *Proteins*, **1988**, *4*, 229-239.
- Okuyama, K. Revisiting the molecular structure of collagen. *Connect. Tiss. Res.*, **2008**, *49*, 299-310.
- Brodsky, B.; Persikov, A.V. Molecular structure of the collagen triple helix. *Adv. Prot. Chem.*, **2005**, *70*, 301-339.
- Shoulders, M.D.; Raines, R.T. Collagen structure and stability. *Ann. Rev. Biochem.*, **2009**, *78*, 929-958.
- 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.
- 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.
- 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.
- 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.



- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] Siligardi, G.; Drake, A.F. The importance of extended conformations and, in particular, the  $\pi$  conformation for the molecular recognition of peptides. *Biopolymers*, **1995**, *37*, 281-292.
- [25] MacArthur, M.W.; Thornton, J.M. Influence of proline residues on protein conformation. *J. Mol. Biol.*, **1991**, *218*, 397-412.
- [26] 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.
- [27] Vitagliano, L.; Berisio, R.; Mastrangelo, A.; Mazzarella, L.; Zagari, A. Preferred proline puckerings in cis and trans peptide groups: Implications for collagen stability. *Prot. Sci.*, **2001**, *10*, 2627-2632.
- [28] Creamer, T.P.; Campbell, M.N. Determinants of the polyproline ii helix from modeling studies. *Adv. Prot. Chem.*, **2002**, *62*, 263-282.
- [29] 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.
- [30] 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. Nat. Acad. Sci. USA.*, **2002**, *99*, 9190-9195.
- [31] Eker, F.; Cao, X.; Nafie, L.; Schweitzer-Stenner, R. Tripeptides adopt stable structures in water. A combined polarized visible raman, ftr, and vcd spectroscopy study. *J. Am. Chem. Soc.*, **2002**, *124*, 14330-14341.
- [32] Avbelj, F.; Grdadolnik, S.G.; Grdadolnik, J.; Baldwin, R.L. Intrinsic backbone preferences are fully present in blocked amino acids. *Proc. Nat. Acad. Sci. USA.*, **2006**, *103*, 1272-1277.
- [33] Zagrovic, B.; Lipfert, J.; Sorin, E.J.; Millett, I.S.; van Gunsteren, W.F.; Doniach, S.; Pande, V.S. Unusual compactness of a polyproline type ii structure. *Proc. Nat. Acad. Sci. USA.*, **2005**, *102*, 11698-11703.
- [34] Bartlett, G.J.; Choudhary, A.; Raines, R.T.; Woolfson, D.N.  $N-\pi$  interactions in proteins. *Nat. Chem. Biol.*, **2010**, *6*, 615-620.
- [35] Zafra-Ruano, A.; Luque, I. Interfacial water molecules in sh3 interactions: Getting the full picture on polyproline recognition by protein-protein interaction domains. *FEBS Lett.*, **2012**, *586*, 2619-2630.
- [36] Agrawal, V.; Kishan, K.V. Promiscuous binding nature of sh3 domains to their target proteins. *Prot. Pept. Lett.*, **2002**, *9*, 185-193.
- [37] Srinivasan, M.; Dunker, A.K. Proline rich motifs as drug targets in immune mediated disorders. *Int. J. Pept.*, **2012**, *2012*, 634769.
- [38] 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 and cytoskeletal proteins is the ligand for the evh1 domain, a protein module present in the ena/vasp family. *EMBO J.*, **1997**, *16*, 5433-5444.
- [39] Gertler, F.B.; Niebuhr, K.; Reinhard, M.; Wehland, J.; Soriano, P. Mena, a relative of vasp and drosophila enabled, is implicated in the control of microfilament dynamics. *Cell*, **1996**, *87*, 227-239.
- [40] Ball, L.J.; Kuhne, R.; Hoffmann, B.; Hafner, A.; Schmieder, P.; Volkmer-Engert, R.; Hof, M.; Wahl, M.; Schneider-Mergener, J.; Walter, U.; Oschkinat, H.; Jarchau, T. Dual epitope recognition by the vasp evh1 domain modulates polyproline ligand specificity and binding affinity. *EMBO J.*, **2000**, *19*, 4903-4914.
- [41] Hunke, C.; Hirsch, T.; Eichler, J. Structure-based synthetic mimicry of discontinuous protein binding sites: Inhibitors of the interaction of mena evh1 domain with proline-rich ligands. *ChemBiochem*, **2006**, *7*, 1258-1264.
- [42] Vermeire, J.; Vanbillemont, G.; Witkowski, W.; Verhasselt, B. The nef-infectivity enigma: Mechanisms of enhanced lentiviral infection. *Curr. HIV Res.*, **2011**, *9*, 474-489.
- [43] Kuo, Y.S.; Baugh, L.L.; Denial, S.J.; Watkins, R.L.; Liu, M.; Garcia, J.V.; Foster, J.L. Overlapping effector interfaces define the multiple functions of the hiv-1 nef polyproline helix. *Retrovirology*, **2012**, *9*, 47.
- [44] Hale, B.G.; Kerry, P.S.; Jackson, D.; Precious, B.L.; Gray, A.; Killip, M.J.; Randall, R.E.; Russell, R.J. Structural insights into phosphoinositide 3-kinase activation by the influenza a virus ns1 protein. *Proc. Nat. Acad. Sci. USA.*, **2010**, *107*, 1954-1959.
- [45] Shin, Y.K.; Li, Y.; Liu, Q.; Anderson, D.H.; Babiuk, L.A.; Zhou, 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.
- [46] Larson, M.R.; Rajashankar, K.R.; Patel, M.H.; Robinette, R.A.; Crowley, P.J.; Michalek, S.; Brady, L.J.; Deivanayagam, C. Elongated fibrillar structure of a streptococcal adhesin assembled by the high-affinity association of alpha- and  $\pi$ ii-helices. *Proc. Nat. Acad. Sci. USA.*, **2010**, *107*, 5983-5988.
- [47] Brady, L.J.; Maddocks, S.E.; Larson, M.R.; Forsgren, N.; Persson, K.; Deivanayagam, C.C.; Jenkinson, H.F. The changing faces of streptococcus antigen i/ii polypeptide family adhesins. *Mol. Microbiol.*, **2010**, *77*, 276-286.
- [48] 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 fctb reveals determinants of group a streptococcal pilus anchoring. *J. Biol. Chem.*, **2010**, *285*, 20381-20389.
- [49] 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 pyogenes. *Euro. Biophys. J. : EBJ.*, **2010**, *39*, 469-480.
- [50] Holmgren, S.K.; Taylor, K.M.; Bretscher, L.E.; Raines, R.T. Code for collagen's stability deciphered. *Nature*, **1998**, *392*, 666-667.
- [51] Vitagliano, L.; Berisio, R.; Mazzarella, L.; Zagari, A. Structural bases of collagen stabilization induced by proline hydroxylation. *Biopolymers*, **2001**, *58*, 459-464.
- [52] Berisio, R.; Granata, V.; Vitagliano, L.; Zagari, A. Imino acids and collagen triple helix stability: Characterization of collagen-like polypeptides containing hyp-hyp-gly sequence repeats. *J. Am. Chem. Soc.*, **2004**, *126*, 11402-11403.
- [53] Doi, M.; Nishi, Y.; Uchiyama, S.; Nishiuchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y. Characterization of collagen model peptides containing 4-fluoroproline; (4(s)-fluoroproline-pro-gly)10 forms a triple helix, but (4(r)-fluoroproline-pro-gly)10 does not. *J. Am. Chem. Soc.*, **2003**, *125*, 9922-9923.
- [54] De Simone, A.; Vitagliano, L.; Berisio, R. Role of hydration in collagen triple helix stabilization. *Biochem. Biophys. Res. Communicat.*, **2008**, *372*, 121-125.
- [55] Persikov, A.V.; Ramshaw, J.A.; Kirkpatrick, A.; Brodsky, B. Triple-helix propensity of hydroxyproline and fluoroproline: Comparison of host-guest and repeating tripeptide collagen models. *J. Am. Chem. Soc.*, **2003**, *125*, 11500-11501.
- [56] Okuyama, K.; Miyama, K.; Mizuno, K.; Bachinger, H.P. Crystal structure of (gly-pro-hyp)(9) : Implications for the collagen molecular model. *Biopolymers*, **2012**, *97*, 607-616.
- [57] Okuyama, K.; Arnott, S.; Takayanagi, M.; Kakudo, M. Crystal and molecular structure of a collagen-like polypeptide (pro-pro-gly)10. *J. Mol. Biol.*, **1981**, *152*, 427-443.
- [58] Okuyama, K.; Xu, X.; Iguchi, M.; Noguchi, K. Revision of collagen molecular structure. *Biopolymers*, **2006**, *84*, 181-191.
- [59] Shoulders, M.D.; Raines, R.T. Interstrand dipole-dipole interactions can stabilize the collagen triple helix. *J. Biol. Chem.*, **2011**, *286*, 22905-22912.
- [60] Bella, J. A new method for describing the helical conformation of collagen: Dependence of the triple helical twist on amino acid sequence. *J. Struct. Biol.*, **2010**, *170*, 377-391.
- [61] Fields, G.B. Synthesis and biological applications of collagen-model triple-helical peptides. *Org. Biomol. Chem.*, **2010**, *8*, 1237-1258.
- [62] Berisio, R.; De Simone, A.; Ruggiero, A.; Improta, R.; Vitagliano, L. Role of side chains in collagen triple helix stabilization and partner recognition. *J. Pept. Sci.*, **2009**, *15*, 131-140.

- [63] Vitagliano, L.; Nemethy, G.; Zagari, A.; Scheraga, H.A. Stabilization of the triple-helical structure of natural collagen by side-chain interactions. *Biochemistry*, **1993**, *32*, 7354-7359.
- [64] Improta, R.; Berisio, R.; Vitagliano, L. Contribution of dipole-dipole interactions to the stability of the collagen triple helix. *Prot. Sci.*, **2008**, *17*, 955-961.
- [65] Persikov, A.V.; Ramshaw, J.A.; Brodsky, B. Collagen model peptides: Sequence dependence of triple-helix stability. *Biopolymers*, **2000**, *55*, 436-450.
- [66] Fallas, J.A.; Lee, M.A.; Jalan, A.A.; Hartgerink, J.D. Rational design of single-composition abc collagen heterotrimers. *J. Am. Chem. Soc.*, **2012**, *134*, 1430-1433.
- [67] Orgel, 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.
- [68] 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.
- [69] Berisio, R.; Vitagliano, L.; Mazzarella, L.; Zagari, A. Crystal structure of a collagen-like polypeptide with repeating sequence pro-hyp-gly at 1.4 Å resolution: Implications for collagen hydration. *Biopolymers*, **2000**, *56*, 8-13.
- [70] Bella, J.; Eaton, M.; Brodsky, B.; Berman, H.M. Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science*, **1994**, *266*, 75-81.
- [71] 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.
- [72] Shoulders, M.D.; Satyshur, K.A.; Forest, K.T.; Raines, R.T. Stereoelectronic and steric effects in side chains preorganize a protein main chain. *Proc. Nat. Acad. Sci. USA.*, **2010**, *107*, 559-564.
- [73] 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 triple helical symmetries. *J. Biol. Chem.*, **2008**, *283*, 32580-32589.
- [74] 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.
- [75] Kawahara, K.; Nishi, Y.; Nakamura, S.; Uchiyama, S.; Nishiuchi, Y.; Nakazawa, T.; Ohkubo, T.; Kobayashi, Y. Effect of hydration on the stability of the collagen-like triple-helical structure of [4(r)-hydroxyprolyl-4(r)-hydroxypropylglycine]<sub>10</sub>. *Biochemistry*, **2005**, *44*, 15812-15822.
- [76] Gordon, M.K.; Hahn, R.A. Collagens. *Cell Tiss. Res.*, **2010**, *339*, 247-257.
- [77] Leitinger, B.; Hohenester, E. Mammalian collagen receptors. *Mat. Biol.*, **2007**, *26*, 146-155.
- [78] Gronwald, W.; Bomke, J.; Maurer, T.; Domogalla, B.; Huber, F.; Schumann, F.; Kremer, W.; Fink, F.; Rysiok, T.; Frech, M.; Kalbitzer, H.R. Structure of the leech protein saratin and characterization of its binding to collagen. *J. Mol. Biol.*, **2008**, *381*, 913-927.
- [79] Ishikawa, Y.; Vranka, J.; Wirz, J.; Nagata, K.; Bachinger, H.P. The rough endoplasmic reticulum-resident fk506-binding protein fkbp65 is a molecular chaperone that interacts with collagens. *J. Biol. Chem.*, **2008**, *283*, 31584-31590.
- [80] Leo, J.C.; Elovaara, H.; Brodsky, B.; Skurnik, M.; Goldman, A. The yersinia adhesin yada binds to a collagenous triple-helical conformation but without sequence specificity. *Prot. Eng. Des. Select.*, **2008**, *21*, 475-484.
- [81] Emsley, J.; Knight, C.G.; Farndale, R.W.; Barnes, M.J.; Liddington, R.C. Structural basis of collagen recognition by integrin  $\alpha 2\beta 1$ . *Cell*, **2000**, *101*, 47-56.
- [82] Zong, Y.; Xu, Y.; Liang, X.; Keene, D.R.; Hook, A.; Gurusiddappa, S.; Hook, M.; Narayana, S.V. A 'collagen hug' model for staphylococcus aureus cna binding to collagen. *EMBO J.*, **2005**, *24*, 4224-4236.
- [83] Hohenester, E.; Sasaki, T.; Giudici, C.; Farndale, R.W.; Bachinger, H.P. Structural basis of sequence-specific collagen recognition by sparc. *Proc. Nat. Acad. Sci. USA.*, **2008**, *105*, 18273-18277.
- [84] Carafoli, F.; Bihan, D.; Stathopoulos, S.; Konitsiotis, A.D.; Kvsanakul, M.; Farndale, R.W.; Leitinger, B.; Hohenester, E. Crystallographic insight into collagen recognition by discoidin domain receptor 2. *Structure*, **2009**, *17*, 1573-1581.
- [85] Gingras, A.R.; Girija, U.V.; Keeble, A.H.; Panchal, R.; Mitchell, D.A.; Moody, P.C.; Wallis, R. Structural basis of mannan-binding lectin recognition by its associated serine protease masp-1: Implications for complement activation. *Structure*, **2011**, *19*, 1635-1643.
- [86] Brondijk, T.H.; Bihan, D.; Farndale, R.W.; Huizinga, E.G. Implications for collagen I chain registry from the structure of the collagen von willebrand factor a3 domain complex. *Proc. Nat. Acad. Sci. USA.*, **2012**, *109*, 5253-5258.
- [87] Manka, S.W.; Carafoli, F.; Visse, R.; Bihan, D.; Raynal, N.; Farndale, R.W.; Murphy, G.; Enghild, J.J.; Hohenester, E.; Nagase, H. Structural insights into triple-helical collagen cleavage by matrix metalloproteinase 1. *Proc. Nat. Acad. Sci. USA.*, **2012**, *109*, 12461-12466.
- [88] Widmer, C.; Gebauer, J.M.; Brunstein, E.; Rosenbaum, S.; Zaucke, F.; Drogemuller, C.; Leeb, T.; Baumann, U. Molecular basis for the action of the collagen-specific chaperone hsp47/serpinh1 and its structure-specific client recognition. *Proc. Nat. Acad. Sci. USA.*, **2012**, *109*, 13243-13247.
- [89] Bertini, I.; Fragai, M.; Luchinat, C.; Melikian, M.; Toccafondi, M.; Lauer, J.L.; Fields, G.B. Structural basis for matrix metalloproteinase 1-catalyzed collagenolysis. *J. Am. Chem. Soc.*, **2012**, *134*, 2100-2110.
- [90] Liu, Q.; Ponnuraj, K.; Xu, Y.; Ganesh, V.K.; Sillanpaa, J.; Murray, B.E.; Narayana, S.V.; Hook, M. The enterococcus faecalis mscramm ace binds its ligand by the collagen hug model. *J. Biol. Chem.*, **2007**, *282*, 19629-19637.
- [91] Visai, L.; Xu, Y.; Casolini, F.; Rindi, S.; Hook, M.; Speziale, P. Monoclonal antibodies to cna, a collagen-binding microbial surface component recognizing adhesive matrix molecules, detach staphylococcus aureus from a collagen substrate. *J. Biol. Chem.*, **2000**, *275*, 39837-39845.
- [92] Elasri, M.O.; Thomas, J.R.; Skinner, R.A.; Blevins, J.S.; Beenken, K.E.; Nelson, C.L.; Smeltzer, M.S. Staphylococcus aureus collagen adhesin contributes to the pathogenesis of osteomyelitis. *Bone*, **2002**, *30*, 275-280.
- [93] Hienz, S.A.; Schennings, T.; Heimdahl, A.; Flock, J.I. Collagen binding of staphylococcus aureus is a virulence factor in experimental endocarditis. *J. Infect. Dis.*, **1996**, *174*, 83-88.
- [94] Ponnuraj, K.; Bowden, M.G.; Davis, S.; Gurusiddappa, S.; Moore, D.; Choe, D.; Xu, Y.; Hook, M.; Narayana, S.V. A "dock, lock, and latch" structural model for a staphylococcal adhesin binding to fibrinogen. *Cell*, **2003**, *115*, 217-228.
- [95] Zong, Y.; Xu, Y.; Liang, X.; Keene, D.R.; Hook, A.; Gurusiddappa, S.; Hook, M.; Narayana, S.V. A 'collagen hug' model for staphylococcus aureus cna binding to collagen. *EMBO J.*, **2005**, *24*, 4224-4236.
- [96] Ganesh, V.K.; Rivera, J.J.; Smeds, E.; Ko, Y.P.; Bowden, M.G.; Wann, E.R.; Gurusiddappa, S.; Fitzgerald, J.R.; Hook, M. A structural model of the staphylococcus aureus clfa-fibrinogen interaction opens new avenues for the design of anti-staphylococcal therapeutics. *PLoS Pathog.*, **2008**, *4*, e1000226.
- [97] Deivanayagam, C.C.; Wann, E.R.; Chen, W.; Carson, M.; Rajashankar, K.R.; Hook, M.; Narayana, S.V. 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.*, **2002**, *21*, 6660-6672.
- [98] Symersky, J.; Patti, J.M.; Carson, M.; House-Pompeo, K.; Teale, M.; Moore, D.; Jin, L.; Schneider, A.; DeLucas, L.J.; Hook, M.; Narayana, S.V. Structure of the collagen-binding domain from a staphylococcus aureus adhesin. *Nat. Struct. Biol.*, **1997**, *4*, 833-838.
- [99] Berisio, R.; Ciccarelli, L.; Squeglia, F.; De Simone, A.; Vitagliano, L. Structural and dynamic properties of incomplete immunoglobulin-like fold domains. *Prot. Pept. Lett.*, **2012**, *19*, 1045-53.
- [100] Vitagliano, L.; Berisio, R.; De Simone, A. Role of hydration in collagen recognition by bacterial adhesins. *Biophys. J.*, **2011**, *100*, 2253-2261.
- [101] Rich, R.L.; Kreikemeyer, B.; Owens, R.T.; LaBrenz, S.; Narayana, S.V.; Weinstock, G.M.; Murray, B.E.; Hook, M. Ace is a collagen-binding mscramm from enterococcus faecalis. *J. Biol. Chem.*, **1999**, *274*, 26939-26945.
- [102] Westerlund, B.; Korhonen, T.K. Bacterial proteins binding to the mammalian extracellular matrix. *Mol. Microbiol.*, **1993**, *9*, 687-694.

- [103] Foster, T.J.; McDevitt, D. Surface-associated proteins of staphylococcus aureus: Their possible roles in virulence. *FEMS Microbiol. Lett.*, **1994**, *118*, 199-205.
- [104] Mohamed, N.; Teeters, M.A.; Patti, J.M.; Hook, M.; Ross, J.M. Inhibition of *Staphylococcus aureus* adherence to collagen under dynamic conditions. *Infect. Immun.*, **1999**, *67*, 589-594.
- [105] Nallapareddy, S.R.; Sillanpaa, J.; Ganesh, V.K.; Hook, M.; Murray, B.E. Inhibition of enterococcus faecium adherence to collagen by antibodies against high-affinity binding subdomains of acm. *Infect. Immun.*, **2007**, *75*, 3192-3196.
- [106] Myllyharju, J.; Kivirikko, K.I. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends in genetics : TIG*, **2004**, *20*, 33-43.
- [107] Humtsoe, J.O.; Kim, J.K.; Xu, Y.; Keene, D.R.; Hook, M.; Lukomski, S.; Wary, K.K. A streptococcal collagen-like protein interacts with the alpha2beta1 integrin and induces intracellular signaling. *J. Biol. Chem.*, **2005**, *280*, 13848-13857.
- [108] Rasmussen, M.; Bjorck, L. Unique regulation of sclb - a novel collagen-like surface protein of streptococcus pyogenes. *Mol. Microbiol.*, **2001**, *40*, 1427-1438.
- [109] Caswell, C.C.; Oliver-Kozup, H.; Han, R.; Lukomska, E.; Lukomski, S. Sc11, the multifunctional adhesin of group a streptococcus, selectively binds cellular fibronectin and laminin, and mediates pathogen internalization by human cells. *FEMS Microbiol. Lett.*, **2010**, *303*, 61-68.
- [110] Chen, S.M.; Tsai, Y.S.; Wu, C.M.; Liao, S.K.; Wu, L.C.; Chang, C.S.; Liu, Y.H.; Tsai, P.J. Streptococcal collagen-like surface protein 1 promotes adhesion to the respiratory epithelial cell. *BMC Microbiol.*, **2010**, *10*, 320.
- [111] Vandersmissen, L.; De Buck, E.; Saels, V.; Coil, D.A.; Anne, J. A legionella pneumophila collagen-like protein encoded by a gene with a variable number of tandem repeats is involved in the adherence and invasion of host cells. *FEMS Microbiol. Lett.*, **2010**, *306*, 168-176.
- [112] Boydston, J.A.; Chen, P.; Steichen, C.T.; Turnbough, C.L., Jr. Orientation within the exosporium and structural stability of the collagen-like glycoprotein bcla of bacillus anthracis. *J. Bacteriol.*, **2005**, *187*, 5310-5317.
- [113] Han, R.; Zwiefka, A.; Caswell, C.C.; Xu, Y.; Keene, D.R.; Lukomska, E.; Zhao, Z.; Hook, M.; Lukomski, S. Assessment of prokaryotic collagen-like sequences derived from streptococcal scl1 and scl2 proteins as a source of recombinant gxy polymers. *Appl. Microbiol. Biotechnol.*, **2006**, *72*, 109-115.
- [114] Yu, Z.; Brodsky, B.; Inouye, M. Dissecting a bacterial collagen domain from Streptococcus pyogenes. *J. Biol. Chem.*, **2011**, *286*, 18960-18968.
- [115] Lukomski, S.; Nakashima, K.; Abdi, I.; Cipriano, V.J.; Ireland, R.M.; Reid, S.D.; Adams, G.G.; Musser, J.M. Identification and characterization of the scl gene encoding a group a streptococcus extracellular protein virulence factor with similarity to human collagen. *Infect. Immun.*, **2000**, *68*, 6542-6553.
- [116] Mohs, A.; Silva, T.; Yoshida, T.; Amin, R.; Lukomski, S.; Inouye, M.; Brodsky, B. Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J. Biol. Chem.*, **2007**, *282*, 29757-29765.
- [117] Xu, Y.; Keene, D.R.; Bujnicki, J.M.; Hook, M.; Lukomski, S. Streptococcal scl1 and scl2 proteins form collagen-like triple helices. *J. Biol. Chem.*, **2002**, *277*, 27312-27318.
- [118] Caswell, C.C.; Barczyk, M.; Keene, D.R.; Lukomska, E.; Gullberg, D.E.; Lukomski, S. Identification of the first prokaryotic collagen sequence motif that mediates binding to human collagen receptors, integrins alpha2beta1 and alpha1beta1. *J. Biol. Chem.*, **2008**, *283*, 36168-36175.
- [119] Caswell, C.C.; Lukomska, E.; Seo, N.S.; Hook, M.; Lukomski, S. Sc11-dependent internalization of group a streptococcus via direct interactions with the alpha2beta(1) integrin enhances pathogen survival and re-emergence. *Mol. Microbiol.*, **2007**, *64*, 1319-1331.
- [120] Gao, Y.; Liang, C.; Zhao, R.; Lukomski, S.; Han, R. The scl1 of m41-type group a streptococcus binds the high-density lipoprotein. *FEMS Microbiol. Lett.*, **2010**, *309*, 55-61.
- [121] Caswell, C.C.; Han, R.; Hovis, K.M.; Ciborowski, P.; Keene, D.R.; Marconi, R.T.; Lukomski, S. The scl1 protein of m6-type group a streptococcus binds the human complement regulatory protein, factor h, and inhibits the alternative pathway of complement. *Mol. Microbiol.*, **2008**, *67*, 584-596.
- [122] Reuter, M.; Caswell, C.C.; Lukomski, S.; Zipfel, P.F. Binding of the human complement regulators cfhr1 and factor h by streptococcal collagen-like protein 1 (scl1) via their conserved c termini allows control of the complement cascade at multiple levels. *J. Biol. Chem.*, **2010**, *285*, 38473-38485.
- [123] Biedermann, F.; Uzunova, V.D.; Scherman, O.A.; Nau, W.M.; De Simone, A. Release of high-energy water as an essential driving force for the high-affinity binding of cucurbit[n]urils. *J. Am. Chem. Soc.*, **2012**, *134*, 15318-15323.