Received: 3 April 2017 | Revised: 20 June 2017 | Accepted: 2 July 2017

DOI: 10.1002/prot.25346

RESEARCH ARTICLE

WILEY

Structural motifs in which β-strands are clipped together with the Π-like module

Alexander V. Efimov

Institute of Protein Research, Russian Academy of Sciences, Moscow Region, Russian Federation

Correspondence
Alexander V. Efimov, Institute of Protein Research, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russian Federation.
Email: efimov@protres.ru

Funding information
Russian Foundation for Basic Research (project No. 17–04-00242)

Abstract

In this study, the structural motifs that can be represented as combinations of small motifs such as β-hairpins, S-, and Z-like β-sheets and βαβ-units, and the Π-like module are described and analyzed. The Π-module consists of connected elements of the β-strand-loop-β-strand type arranged in space so that its overall fold resembles a clip or the Greek letter Π. In proteins, the Π-module itself and the structural motifs containing it exhibit unique overall folds and have specific sequence patterns of the key hydrophobic, hydrophilic and glycine residues. All this together enables us to conclude that these structural motifs can fold independently of the remaining part of the molecule and can act as nuclei and/or "ready-made" building blocks in protein folding.

KEYWORDS
α-helix, conformation, loop, overall fold, protein structure, sequence pattern

1 | INTRODUCTION

Structural motifs of globular proteins can be defined as commonly occurring folding units consisting of two or more elements of secondary structure that are adjacent along the polypeptide chain and are in close contact in three-dimensional space. While many different structural motifs have been observed to recur within proteins, only some of the motifs exhibit the definite handedness and a unique overall fold irrespective of whether they occur in homologous or nonhomologous proteins.[1–4] These structural motifs are of particular value since they can act as nuclei or "ready-made" building blocks in protein folding or can be used as starting structures in protein modeling.[4–6]

In this study, some structural motifs that can be represented as combinations of smaller modules are described and analyzed. A distinctive feature of these motifs is the module consisting of connected elements of β-strand-loop-β-strand type in which the β-strands do not form H-bonds between each other similar to that in the split β-hairpin. Its overall fold resembles a clip or the Greek letter Π, so it will be referred to here as the Π-module. In proteins, the Π-module clips together the β-strands in β-hairpins, S- and Z-like β-sheets, and βαβ-units making the obtained structural motifs more cooperative and stable. Variants of the Π-module were initially observed in the ψ-motifs,[7–9] φ-motifs,[10] and closed 3β-corners,[11] but detailed structure and amino acid sequences of their crossover loops were not considered. In this article, the Π-modules including α-helices in the crossover loops and the sequence patterns coding for them have been described and analyzed.

2 | METHODS

The structural motifs containing the Π-modules have been revealed as a result of visual inspection and protein structure comparison of the Protein Data Bank entries (http://www.rcsb.org/pdb/) as well as protein structures described in the original articles. Protein structures were visually examined using RasMol molecular graphics program.[12] Possible homologies were revealed by the Blast2Sequences program.[13] Our databases of the φ- and ψ-motifs[9,10] have also been included. In total, >300 nonhomologous proteins were selected into the database and its extension is in progress. In many proteins, the structural motifs including the Π-modules of different types occur twice and more times.

3 | RESULTS AND DISCUSSION

In proteins, there are some variants of structural motifs in which the Π-module clips together β-strands. Their structures can be represented as combinations of smaller motifs and Π-modules. Figure 1A,F show a schematic representation of the combination of the S-like β-sheet and the Π-module denoted here as the SΠ-motif and its example (region

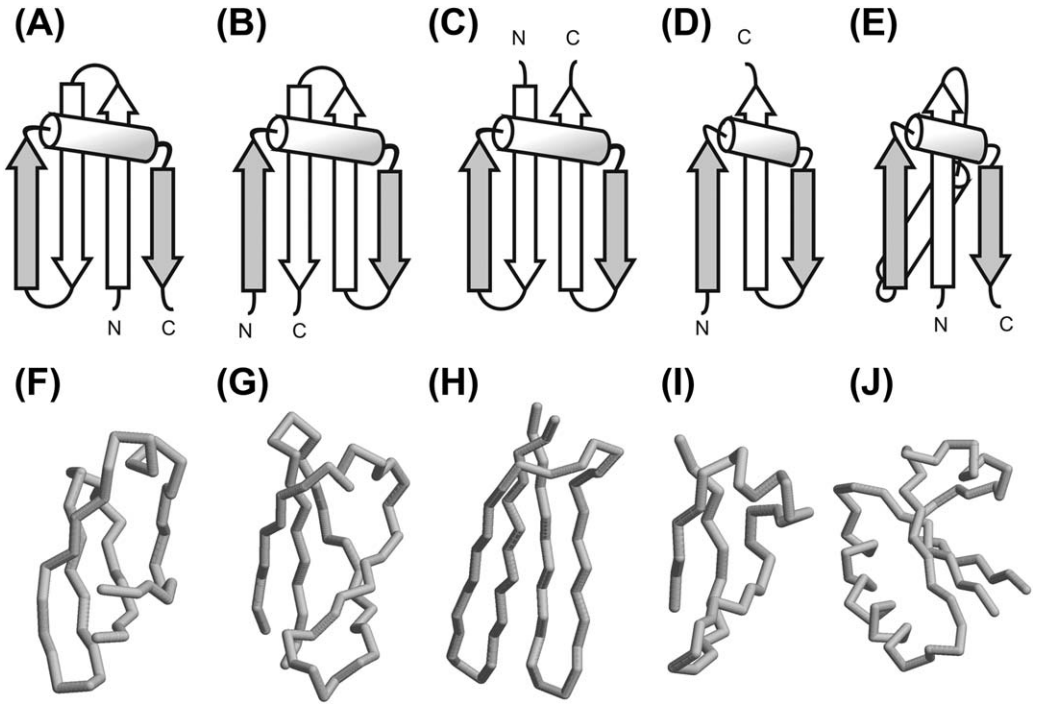

FIGURE 1 Schematic representation (A–E) and examples from known proteins (F–J) of the structural motifs in which the Π-modules clip together β-strands. β-Strands are shown with arrows directed from the N- to the C-ends and α-helices as cylinders. The Π-modules are highlighted in gray. The examples were drawn with the RasMol program.[12] See also the text

22–67 of 1JT8). Note that SΠ-motifs are included in all the OB-folds as their parts.[14] A combination of the Z-like β-sheet and the Π-module denoted here as the ΠZ-motif and its example (region 20–73 of 1EP3) are shown in Figure 1B,G. Note that in proteins Z-like β-sheets are often folded upon themselves to form the so-called 3β-corners, so the 3β-corners closed into cycles at the N-ends[11] are the ΠZ-motifs too. The φ-motif (Figure 1D) can be represented as a combination of the Π-module and the β-hairpin (its example in Figure 1I is region 770–807 of 1VLE, chain S), and the split φ-motif (Figure 1C) as a combination of the Π-module and two β-hairpins (its example in Figure 1H is region 22–72 of 3U73, chain U). Figure 1E,J show a combination of the Π-module and the right-handed βαβ-unit and its example (region 24–83 of 1GH2). This combination will be referred to here as the βαβΠ-motif (in our previous article,[9] it was called the βαβψ-combination). It should be noted that the upper row of Figure 1A–E shows the motifs schematically as relatively flattened folds although the β-sheets and β-hairpins in real structures are usually twisted and coiled (see especially examples F, G, and I).

In our previous articles devoted to the φ- and ψ-motifs,[9,10] the Π-module was represented as a split β-hairpin in which the two edge β-strands are connected by the crossover loop located above the central one or two β-strands or their extensions. In theory, there are the right- and left-turned Π-modules. When viewed from the crossover loop, the polypeptide chain runs from the N- to the C-end in the clockwise direction in the right-turned Π-modules (as in all the combinations shown in Figure 1) and in anticlockwise direction in the left-turned Π-modules. Among 50 proteins containing φ-motifs, only one protein (PDB ID 1QME) has two left-handed φ-motifs,[10] although in combinations of ψ-motifs and βαβ-units 34% of the ψ-motifs are left-handed[9] and, consequently, their Π-modules are left-turned. In contrast, all the Π-modules having α-helices in the crossover loops are right-turned in all the βαβΠ-combinations (Figure 1E,J) found in our database. Analysis of the database shows that all the Π-modules having α-helices in the crossover loops are also right-turned in all the structural motifs

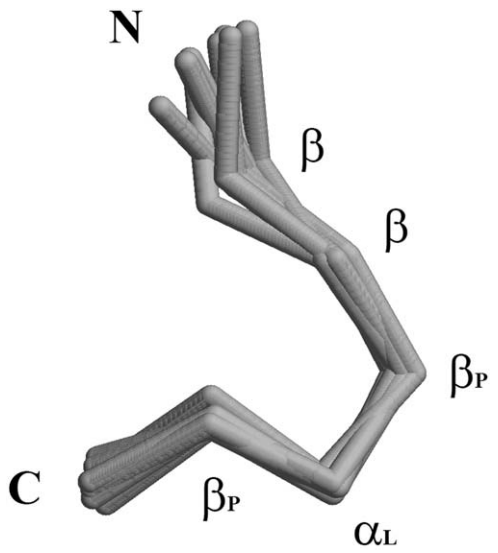

FIGURE 2 Ten superimposed examples of standard arches having $\beta\beta\beta\alpha_L\beta$-conformations found in the Π-modules of the SΠ-motifs. The structures were superimposed using the Deep View/Swiss-PDB Viewer program[15] and drawn with the RasMol program.[12] These examples were taken from the set of the modules presented in Figure 4

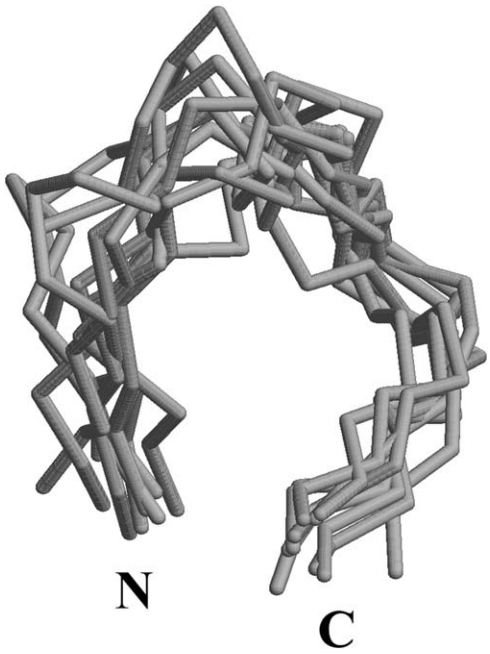

FIGURE 3 Ten Π-modules found in ΠZ-motifs superimposed on one another using Deep View/Swiss-PDB Viewer program.[15] These examples were taken from the set of modules shown in Figure 5

presented in Figure 1. The predominance of right-handed form of the Π-module having an α-helix in the crossover loop as well as of other motifs is determined by several reasons[16–18]; however, finally it is a result of the homochirality of L-amino acid residues in proteins. One of them is that the packing of the α-helix against the concave surface of the twisted and coiled β-sheet in the right-turned Π-module is more favorable than its packing against the opposite convex surface that would result in a left-turned Π-module. The other reason is that right-handed crossovers could be kinetically trapped during folding similar to that in three-helix bundles.[18]

One more structural feature of such Π-modules is that the connection region between the α-helix and the second β-strand has one or more residues in the sterically constrained α_L- or ε-conformations. Most often the connection regions contain standard arches having $\beta\beta\beta\alpha_L\beta$-conformations[19] (Figure 2), so the overall fold of such Π-modules can be described as β-strand-α-helix-arch-β-strand (Figure 3). Thus, it can be concluded that the Π-module with an α-helix in the crossover loop itself and its combinations with β-sheets and β-hairpins exhibit the definite handedness and the unique overall folds, despite their β-strands and α-helices being of different lengths, their connection regions differing in length and conformation and their sequences lacking homology.

Apparently, the isolated Π-module is not stable because its β-strands do not interact between each other and its overall fold is not compact. However, in combinations with the S- and Z-like β-sheets, β-hairpins, and βαβ-units, the Π-module fastens their β-strands like a clip that results in formation of compact structures closed into cycles. Such motifs closed into cycles are more cooperative and stable than open ones and this may be one of the main reasons of high frequencies of occurrence of the motifs in proteins.[11,20] It seems very likely that at the first step of folding the β-hairpins, S- and Z-like β-sheets, and βαβ-units are formed and then their β-strands are clipped together with the Π-module. At the next step of folding the obtained SΠ-, ΠZ-, φ-, and βαβΠ-motifs can act as nuclei and/or "ready-made" building blocks.

```
                    ββββββββααααα----------βββαLβP ββββββββββ
1.    4a75:A  pVdVfVhqskLymegfrs----LkeGe pVeFtFkK--  62-92
2.    1br9:A  fItTaPssavCgvsL-------DvgGk kEyLiAgKa-  62-90
3.    1e1O:A  rTqLyVardsLpegVyndqfk-WdlGd iIgArGtLfK  93-129
4.    1jt8:A  tRlGrIpgrLknrIw-------VreGd vViVkPw---  44-70
5.    1mjc:A  --dVfVhfsaIqndGyls----LdeGq kVsFtIe---  29-56
6.    1sro:A  -KeGlVhisqIAdkrVekvtdyLqmGq eVpVkV----  29-60
7.    1z9f:A  fFrIvTfgrLAefArty-----LtkGr lVlVeGeMr-  52-82
8.    2bh8:B  dVfVhFsagsSgaAVrgn----PqqGd rVeGkIk---  43-72
9.    2f3i:A  dLiLdVniqIYp----------VdlGd kFrLvIa---  38-61
10.   3hrz:C  tHqYiSqrkCqeaLn-------LkvNd dYlIwG----  1538-1563
11.   3i2z:A  --dVfVhfsAIqtnGfkt----LaeGq rVeFeI----  28-54
12.   3l0o:A  --dIyIspsqIrkFn-------LntGd iIsGvIr---  85-109
13.   1txy:A  qMpViVsgheNqaIths-----ItvGs rItVqGfIs-  49-79
14.   3cw2:D  --eAfLpwseVsskWvknirdvLkeNr kViViVkV--  38-68
15.   1ah9:A  vVtAhIsgkMRkknYir-----IltGd kVtVeLt---  30-57
16.   1fr3:A  eLvAaItidsVadLd-------LvpGd kVtAlVk---  33-59
17.   1ueb:A  -AvVqVplf-------------VepGe vIkVd-----  157-174

      Consensus  o•o•o•OOOO•oo---------•ooGo O•O•O•O•O
```

FIGURE 4 Structural alignment of amino acid sequences coding for the 17 representative Π-modules found in the SΠ-motifs. Each column is headed by a symbol, α, β, β_P, and α_L, showing the conformation of the residues in it; symbols β show inside β-residues pointed to the hydrophobic core. The capital letters show the residues pointed into the hydrophobic core in both α-helical and β-structural regions as well as in the arches. The bottom line represents the consensus sequence with O: preferably hydrophilics, ·: hydrophobics, and G: glycines. PDB codes of proteins are listed on the left and residue numbers of the sequence regions are shown on the right

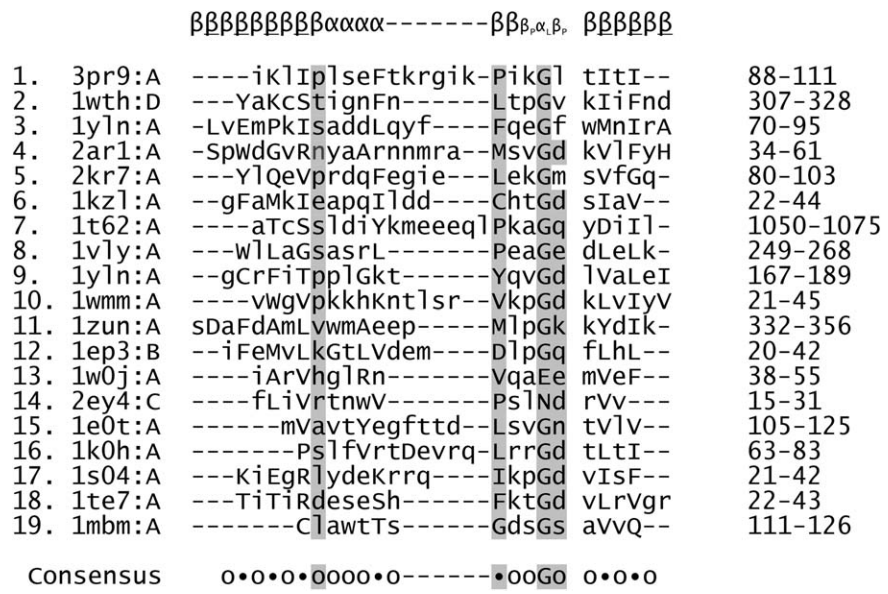

FIGURE 5 Structural alignment of amino acid sequences coding for the 19 representative Π-modules found in the ΠZ-motifs. Designations are as in Figure 4

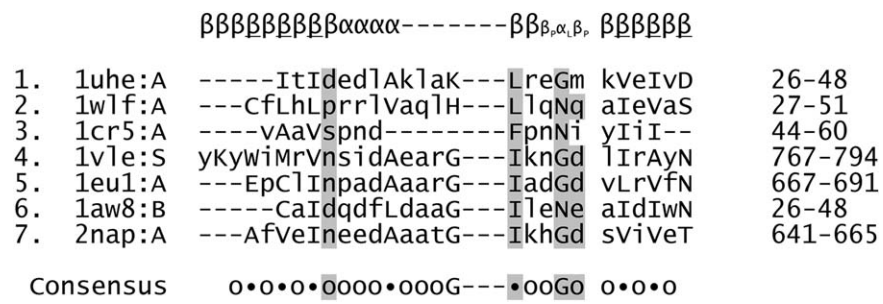

FIGURE 6 Structural alignment of amino acid sequences coding for the 7 representative Π-modules found in the φ-motifs Designations are as in Figure 4

Structural alignment of the amino acid sequences coding for a given type of Π-modules shows that they have very similar sequence patterns for the key hydrophobic, hydrophilic, glycine and proline residues irrespective of whether they occur in homologous or nonhomologous proteins. Figure 4 shows a structural alignment of the amino acid sequences coding for the Π-modules in the SΠ-motifs. In this type of the Π-modules, the crossover loop consists of an α-helix being of different length, the ββββα_Lβ-arch (see Figure 2) and a variable

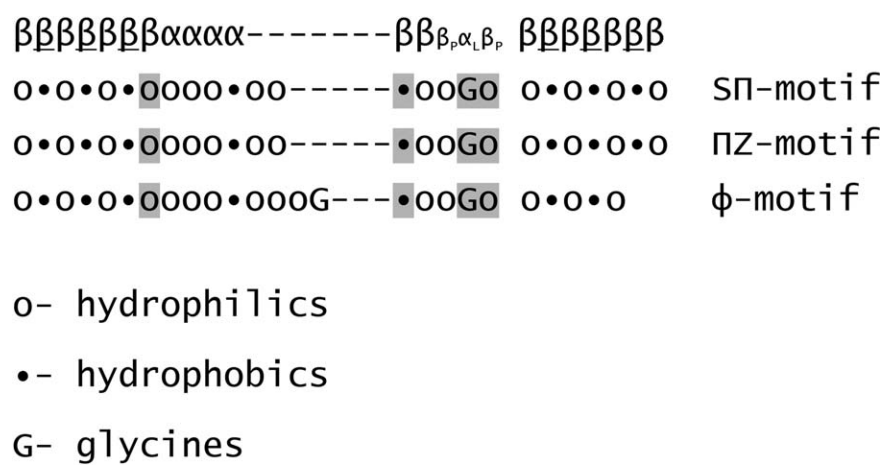

FIGURE 7 Comparison of consensus sequences coding for Π-modules in SΠ-, ΠZ-, and φ-motifs

region connecting them. The structural alignment was performed by hand. Amino acid residues occupying the equivalent positions in the Π-modules and having the same conformations are arranged column-wise. Each column is headed by the symbol showing the conformation of the residues within it. As seen, the inside positions (β) whose side chains are directed into the hydrophobic core are preferably occupied by hydrophobic residues and the outside positions (β) of the β-strands are preferably occupied by hydrophilic ones that is in a good agreement with the theory of secondary structure of proteins.[21] The last β-residue of the first β-strand which is simultaneously the residue forming the first H-bond within the α-helix has a hydrophilic, small or proline side-chain (they are highlighted in grey in Figure 4) in most examples. Note that this position is occupied by hydrophilic, small or proline residues, but not bulky hydrophobic residues in most α-helices in proteins.[3] However, the α-helices of the Π-modules have at least one hydrophobic residue per turn in the corresponding positions directed to the hydrophobic core. A distinctive feature of the $\beta\beta\beta\alpha_L\beta$-arches is that all their first β-positions are occupied by hydrophobic residues and the α_L-positions - by glycines in most cases. The sequence pattern coding for the $\beta\beta\beta\alpha_L\beta$-arch in the Π-modules is slightly different from that observed in other protein structures. Initially, the $\beta\beta\beta\alpha_L\beta$-arch has been found in the loop regions connecting two β-strands located in different layers of the two-layer β-proteins.[19] It was shown that the first β-residue of the arch (it is also the last residue of the first β-strand) is buried in the hydrophobic core and should be hydrophobic. The side-chain of the fifth arch residue (which is also the first residue of the second β-strand) is partially buried and can be both hydrophobic and hydrophilic. The fourth α_L- residue should be glycine or a residue with a flexible side-chain. Prolines in the second or third positions facilitate formation of the arch, but their presence is not necessary.[19] A detailed analysis of the $\beta\beta\beta\alpha_L\beta$-arches found in the SΠ-motifs (Figure 4) as well as in ΠZ- and φ-motifs (Figures 5 and 6) shows that the fifth β-position (it is also the first position of the second β-strand of the Π-module) is preferably occupied by hydrophilic residues of which 45% are residues of aspartic acids (Ds). The side chains of the second and third positions of the arches in the Π-modules are accessible to water molecules and therefore are preferably occupied by hydrophilic residues. All this taken together enables us to suggest a consensus sequence pattern for the Π-module of this types (see bottom lines in Figures 4–6). It should be noted that open circles represent positions preferably occupied by hydrophilic residues, closed circles—preferably by hydrophobic residues. For the positions highlighted in gray, the requirements are stronger and the residues in the first positions of the α-helices should be hydrophilic or small, in the first positions of the $\beta\beta\beta\alpha_L\beta$-arches—hydrophobic, in the α_L-positions—glycines or flexible, and in the fifth positions of the arches—hydrophilic.

Figure 7 represents the comparison of the consensus sequence patterns. As seen, the main features of the sequences coding for the Π-modules found in the SΠ-, ΠZ-, and φ-motifs are very similar, although they are not homologous and their α-helices, β-strands, and loops are of different length. It should be noted that the Π-modules found in βαβΠ-motifs (Figure 1E,J) are rather different from those described above. In our database, there are >50 βαβΠ-motifs having α-helices in the crossover loops. In most cases, their crossover loops do not contain the $\beta\beta\beta_P\alpha_L\beta_P$–arches. Therefore, their conformations and sequence patterns will be analyzed and described elsewhere.[22]

4 | CONCLUSION

This article describes five structural motifs and each of them exhibits its unique overall fold irrespective of whether it occurs in homologous or nonhomologous proteins. In spite of these structural motifs have overall folds different from one another; all of them involve very similar Π-modules. The Π-modules clip together the β-strands in the β-sheets that results in formation of more cooperative and stable structural motifs and this may be the reason of high frequencies of occurrence of the motifs in proteins. On the other hand, the structural motifs described here (ΠZ-, SΠ-, φ-, and βαβΠ-motifs) have unique folds themselves and each of them can be a core around which the remainder of the protein molecule or domain can be folded.[4,5,9–11] This hypothesis is also supported by that these motifs have very similar sequence patterns of the key hydrophobic, hydrophilic, glycine, and proline residues.

ACKNOWLEDGMENTS

The author thanks Ms. E. A. Boshkova and Dr. A.M. Kargatov for help in drawing the figures. This work was supported by the Russian Foundation for Basic Research (project No. 17–04–00242).

REFERENCES

[1] Rao ST, Rossmann MG. Comparison of super-secondary structures in proteins. *J Mol Biol.* 1973;76:241–256.

[2] Levitt M, Chothia C. Structural patterns in globular proteins. *Nature.* 1976;261:552–558.

[3] Efimov AV. Standard structures in proteins. *Prog Biophys Mol Biol.* 1993;60:201–239.

[4] Efimov AV. Favoured structural motifs in globular proteins. *Structure.* 1994;2:999–1002.

[5] Efimov AV. Structural trees for protein superfamilies. *Prot Struct Funct Bioinf.* 1997;28:241–260.

[6] Efimov AV. Super-secondary structures and modeling of protein folds. *Methods Mol Biol.* 2013;932:177–189.

[7] Suguna K, Bott RR, Padlan EA, et al. Structure and refinement at 1.8 A resolution of the aspartic proteinase from Rhizopus chinensis. *J Mol Biol.* 1987;196:877–900.

[8] Castillo RM, Mizuguchi K, Dhanaraj V, Albert A, Blundell TL, Murzin AG. A six-stranded double-psi beta barrel is shared by severalprotein superfamilies. *Structure.* 1999;7:227–236.

[9] Kargatov AM, Efimov AV. A novel structural motif and structural trees for proteins containing it. *Biochemistry (Moscow).* 2010;75:249–256.

[10] Efimov AV. Structural trees for proteins containing φ-motifs. *Biochemistry (Moscow).* 2008;73:23–28.

[11] Boshkova EA, Efimov AV. Structures closed into cycles in proteins containing 3β-corners. *Biochemistry (Moscow).* 2010;75:1258–1263.

[12] Sayle RA, Milner-White EJ. RASMOL – biomolecular graphics for all. *Trends Biochem Sci.* 1995;20:374–376.

[13] Tatusova TA, Madden TL. Blast 2 sequences – a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett*. 1999; 174:247–250.

[14] Murzin AG. OB(oligonucleotide/oligosaccharide binding)-fold: Common structural and functional solution for non-homologous sequences. *EMBO J*. 1993;12:861–867.

[15] Gluex N, Peitsch MC. SWISS-MODEL and Swiss-PDBViewer: An environment for comparative protein modeling. *Electrophoresis*. 1997;18:2714–2723.

[16] Sternberg MJE, Thornton JM. On the conformation of proteins: The handedness of the β-strand-α-helix-β-strand unit. 1976;105: 367–382.

[17] Richardson JS. Handedness of crossover connections in β-sheets. *Proc Natl Acad Sci USA*. 1976;73:2619–2623.

[18] Cole BJ, Bystroff C. Alpha-helical crossovers favor right-handed supersecondary structures by kinetic trapping: The phone cord effect in protein folding. *Protein Sci*. 2009;18:1602–1608.

[19] Efimov AV. Pseudo-homology of protein standard structures formed by two consecutive β-strands. *FEBS Lett*. 1987;224:372–376.

[20] Efimov AV. Structural motifs are closed into cycles in proteins. *Biochem Biophys Res Commun*. 2010;399:412–415.

[21] Lim VI. Structural principles of the globular organization of protein chains. A stereochemical theory of globular protein secondary structure. *J Mol Biol*. 1974;88:857–872.

[22] Kargatov AM, Efimov AV. Unique combinations of βαβ-units and Π-like modules in proteins and features of their amino acid sequences. *Mol Biol (Moscow)*. 2018, In press.

How to cite this article: Efimov AV. Structural motifs in which β-strands are clipped together with the Π-like module. *Proteins*. 2017;85:1925–1930. https://doi.org/10.1002/prot.25346