The Structure of Small Beta Barrels

Philippe Youkharibache*, Stella Veretnik[1], Qingliang Li, Philip E. Bourne*[1]

National Center for Biotechnology Information, The National Library of Medicine, The National Institutes of Health, Bethesda Maryland 20894 USA.

*To whom correspondence should be addressed at philippe.youkharibache@nih.gov and pebourne@gmail.com

[1] Current address: Department of Biomedical Engineering, The University of Virginia, Charlottesville VA 22908 USA.

Abstract

The small beta barrel is a protein structural domain, highly conserved throughout evolution and hence exhibits a broad diversity of functions. Here we undertake a comprehensive review of the structural features of this domain. We begin with what characterizes the structure and the variable nomenclature that has been used to describe it. We then go on to explore the anatomy of the structure and how functional diversity is achieved, including through establishing a variety of multimeric states, which, if misformed, contribute to disease states. We conclude with work following from such a comprehensive structural study.

Introduction

Why small beta barrels are interesting?

Small beta barrels possess several intriguing features as discussed subsequently. When taken together these features make this protein structural domain an interesting study from the perspective of structure, function and evolution.

Small beta barrels occupy an extremely broad sequence space. In other words, many small barrels with similar structures and functions have little or no detectable sequence similarity, yet the folding process is robust and insensitive to the majority of sequence variations. Consequently, small barrels can be tuned for a variety of functions through variations in sequence and structural modifications to the core structural framework as described herein. Small beta barrels interact with RNA, DNA and protein; sometimes with two partners simultaneously . Small barrels are evolutionary ancient, being found in viruses, bacteria,

archaea and eukaryotes and act as fundamental components in diverse biological processes. For example, RNA biogenesis (including splicing, RNAi, sRNA) [1–4], structural organization of DNA [5] initiation of signaling cascades through DNA recognition (recombination, replication and repair, inflammation response, telomere biogenesis) [6–8] The recognition of histone tails by small barrels lies in the heart of chromatin remodeling [9], while recognition of polyproline signature makes the small barrel an ultimate adaptor domain in regulatory cascades [10]. Small beta barrels function as a single domain protein, or as part of a multi-domain protein. They also function as quaternary structures through toroid rings assembled from individual small beta barrels.

Results and Discussion

What characterizes the structure of a small beta barrel?

There are many folds named beta barrels: 53 folds in SCOPe 2.06 [11] carry a definition of barrel or pseudo-barrel and 79 X-groups appear under the architecture of beta barrel in the ECOD classification [12]. While not defined as such in the literature, here we define small beta barrels as domains, typically 60-120 residues long, with a superimposable core of approximately 35 residues, which belong to SCOP (v.2.06) folds b.34 (SH3), b.38 (SM-like), b.40 (OB), b.136 (stringent starvation protein), and b.137 (RNase P subunit p29). Given the structural and functional plasticity of small beta barrels, to provide focus, this paper concentrates on the first three folds (b.34, b.38 and b.40) which contribute the vast majority of structures and functions represented by small beta barrels.

In general, beta-barrels can be thought of as a beta sheet that twists and coils to form a closed structure in which the first strand is hydrogen bonded to the last [13,14]. This type of barrel is often described as consisting of a strongly bent antiparallel beta sheet [15–17], or as a beta sandwich [18]. Classically, barrels are defined by the *number of strands* **n** and the *shear number* **S** [14,19]. Shear number **S** determines the extend of the stagger of the beta sheet, or

the tilt of the barrel with respect to its main axis. The extent of the stagger defines the degree of twist and coil of the strands and the internal diameter of the barrel [13,14]. It is proposed that the increase in S (or tilt of the barrel) increases throughout evolution [20]. The subset of biologically relevant (n, S) pairs found in nature is rather limited, as was described by Murzin and co-workers: **n**<=**S**<=2**n**. Specifically, the combinations that are observed are **S**=8, **n**=4 to 8; **S**=10, **n**=5 to 10; **S**=12, **n**=12 [13]. Barrels can also be thought of as consisting of 2 beta sheets packed face-to-face and orthogonal to each other [21]. By this definition barrels have higher staggering and are flatter, so the two opposite sides pack together. As such small beta barrels are of this orthogonal type with low strand number, n, and high shear number S: n=4, S=8. In SCOPe version 2.06, b.34 and b.38 are defined as n=4, S=8 with an SH3 topology, while the OB fold b.40 is defined as n=5, S=10. Usually the 4th strand, as defined in SCOP for b.34 and b.38, is interrupted by a 3-10 helix, breaking it into 5 strands. In this work (as in most publications) small beta barrels are defined as containing 5 strands and represented by two orthogonally packed sheets. In the following however we will consider the highly bent second strand β2 as composed of two parts β2N and β2C as each will participate in the two orthogonally packed beta sheets of this barrel.

This work surveys a significant number of small barrel protein structures representative of different functional classes. By no means complete, we believe it to be the largest such survey to date. Table 1 summaries the specific proteins discussed in this paper.

Protein/domain name	**Ligand**	**SCOP nomenclature**	**PDB ID**
Chromo (sac7d, sso7d)	DNA, peptide	b.34.13.1	1WD1, 1KNA, 1Q3L
FMRP	RNA	b.34.9.1	4QW2
Hfq (Sm-like)	RNA	b.38.1.2	1KQ2, 3GIB, 2YLC
HIN domain	DNA, protein	b.40.16	4LNQ, 3RN5
Interdigitated Tudor (JMJD2A)	peptide	b.34.9.1	2QQR, 2GF7

Kin-17	RNA	not classified	2CKK
Mpp8	peptide	b.34.13	3QO2
OB-fold (Shiga-like toxin)	oligosaccharides	b.40.2.1	1C4Q, 1BOS
PAZ	RNA	b.34.14	4W5N,1SI3, 3O7X
Plus3	DNA, protein	b.34.21	2BZE
Retroviral integrase	dsDNA	b.34.7.1	5EJK, 4FW1
RNaseP	RNA	b.137.1	1TS9, 2KI7
SH3-like (polyPro)	peptide	b.34.2	1CKA, 1SEM, 1PSK, 2JXB
SH3-like embedded in OB-fold (eTud)	RNA, protein	b.40.4.5, b.34.5.3	3OMC, 3NTK, 4Q5Y
Sm-like/lsm	RNA	b.38.1.1	1D3B, 4WZJ, 3PGW, 4M7A, 4M75, 3S6N, 4C8Q
Spt5	RNA	b.34.5.5	4YTK
Tandem OB-SH3-like (RL2, eIF5A)	RNA	b.40.4.5, b.34.5.3	1S72, 3CPF, 4V88
Tandem Tudor (53BP1)	protein, DNA	b.34.9.1	2MWO, 1SSF, 2G3R, 3LGL
TrmB	sugars	b.38.5.1	2F5T

Table 1. Proteins used in this work. The underlined PDB ID indicates the structure used in the figures. The reference structure, Hfq, is in bold.

What nomenclature is used to describe the structure of small beta barrels?

Over decades, several independent nomenclatures for the loops within sub-groups of small barrels have emerged. The three most prominent nomenclatures are as follows (Table 2). First, are SH3-like barrels involved in signal transduction through binding to polyPro motifs (b.34.2) as well as chromatin remodeling through recognizing specific modifications on histone tails by Chromo-like (b.34.13) and Tudor-like (b.34.9). Second, are Sm-like barrels involved broadly in RNA biogenesis (b.38.1) and third are OB-fold barrels involved in cellular signaling through binding to nucleic acids and oligosaccharides (b.40). To be consistent throughout this paper and

inclusive of previous work, we cross-map the nomenclatures (Table 2) and use the nomenclature for the SH3-like fold throughout for either SH3-like (b.34) or Sm-like (b.38 or any other small barrel sharing the same topology, b.136, b.137 and b.41 for example). Given the extent of the existing literature, the OB-fold nomenclature is preserved with mapping to SH3-like when appropriate.

Strands bracketing the loops (using SH3-like numbering)	SH3-like (b.34) Loop name	SM-like (b.38) name of corresponding loop	OB-fold (b.40) name of corresponding loop
α-helix- β1 or β0-β1	N-term loop	L1	L01
β1-β2	RT	L2	-
β2-β3	n-Src	L3	L12
β3-β4	Distal	L4	L23
β4-β5	3-10 helix	L5	-

Table 2. Mapping loop names of the small barrel used in major different superfamilies sharing an SH3 topology onto the SH3-like (b.34) notation used in this work. The SH3/Sm topology, using SH3 domain nomenclature, runs (α1-β1)-(β2-β3-β4)-β5 where β2-β3-β4 is a meander. A complete description of OB loops is given in Table 3.

The reference structure used throughout this work is that of the Hfq protein with an Sm-like fold (SCOP b.38.1.2), as it represents the simplest version of a small beta barrel (Figure 1A). If one superimposes all small barrels and identifies Structurally Conserved Regions (SCRs) - Hfq appears to be the most regular structural representative containing SCRs. Therefore, for simplicity and clarity of presentation, we use Hfq as a prototypical SH3-like fold representative, even though it is not assigned as such by SCOP. Other classifications group many of the folds sharing an SH3-like topology into a single category [12]. Hfq represents a structural framework of Sm-like as well as SH3-like folds. The rationale for using the SH3-like domain nomenclature is, firstly, it is entrenched in the literature and secondly, all small barrels discussed here (with

the exception of OB), regardless of their SCOP nomenclature, have the same topology and a highly superimposable structural framework. Remarkably, OB, which has a different topology, is in some cases even more superimposable than some SH3-like folds due to similar positions of the α-helix in the OB-fold and Sm-like fold (discussed in 'More structural variations').

Only a few features are specific to the SH3-like small barrel structure. It consists of 5 beta-strands arranged in an antiparallel manner, a conserved Gly in the middle of the second β-strand (usually followed by a beta bulge) causes a strong bend in that strand (β2) dividing it into N-term (β2N) and C-term (β2C) sections. As such this then defines two orthogonal beta sheets comprising the beta barrel sandwich, thus converting it into a de facto 6-stranded barrel, with each beta sheet consisting of 3 beta-strands. A short 3-10 helix links strands β4 and β5; β4 and β5 strands straddle the barrel and belong to different beta sheets (as do β2N and β2C). This arrangement enables oligomerization of the barrels through interactions of β4-β5 of adjacent monomers - a critical feature in toroid formation (see below). Beta sheet A, also referred to as the *Meander*, is a contiguous 3-strand beta-sheet consisting of strands β2C, β3 and β4. Beta sheet B is non-contiguous and referred to as the N-C sheet since it connects the C-terminal strand to the N-terminal of the protein in an antiparallel fashion. Beta sheet B consists of strands β5, β1, β2N.

The Anatomy of Small Beta Barrels

Topological descriptions

Over the years, several different topological descriptions have arisen for describing small barrels. These are presented in Fig.1 relative to the Hfq reference structure (Fig. 1A).

Meander (Fig. 1B) has the barrel subdivided into two beta sheets: Sheet A (Meander) consisting of β2C, β3 and β4 and sheet B (N-C) consisting of β1, β2N and β5.

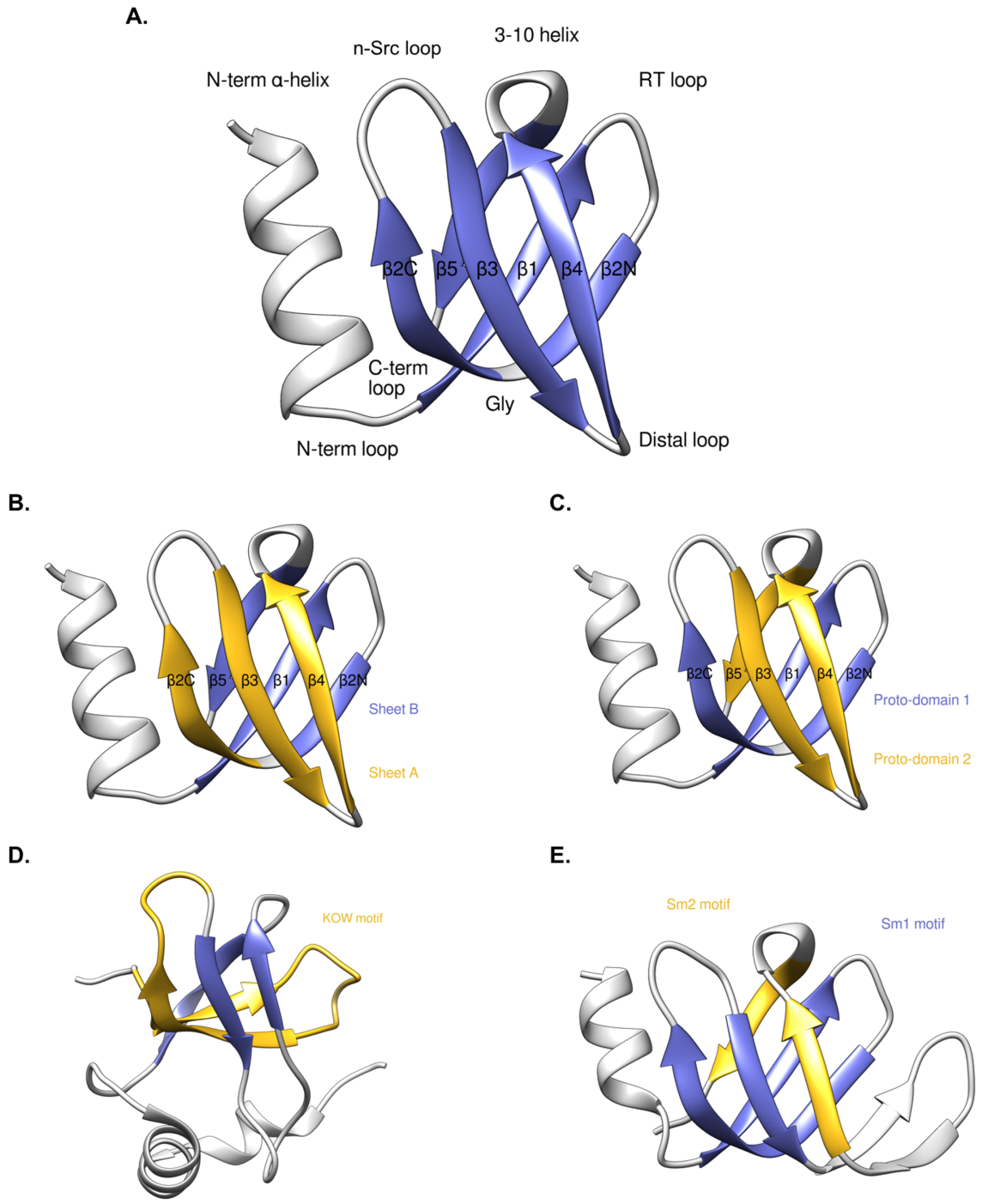


Figure 1. Structural organization of simple small barrels. The Hfq protein (PDB ID 1KQ2) from the Sm-like fold (b.38.1.2) is used as a reference structure (defining the structural framework as it represents the shortest version of the barrel where all loops are reduced to tight turns) throughout this work. The SH3 strands and loop nomenclature is used for all SH3-like barrels **A.** Small barrel: loops and strands labeled **B**. The barrel is divided into two beta sheets: *Sheet A (Meander)* consist of β2C, β3 and β4; *Sheet B (N-*

C) consists of β5, β1, β2N. **C**. The barrel is divided into two proto-domains related through C2 structural symmetry. Proto-domain 1 consist of strands β1, β2N, and β2C; Proto-domain 2 consist of strands β3, β4, and β5. **D**. The KOW motif consist of 27 residues and covers β1, β2 and the loop preceding β1. **E**. The barrels in Sm often define functional motifs Sm1 (β1- β3) and Sm2 (β4 - β5).

Proto-domains (Fig. 1C) exhibit pseudo-symmetry within each protein domain and was noticed early on, for example the C2 symmetry in serine proteases six-stranded beta barrels [19]. However, to our knowledge it has never been described in smaller barrels. The small barrel is subdivided into two proto-domains related by a C2 symmetry operation. Some domains such as serine or aspartyl proteases are believed to have arisen from ancient duplications, where the sequence signal may be lost, but structural similarity is apparent. In the case of small barrels, proto-domain 1 consists of β1, β2N and β2C; proto-domain 2 consists of β3, β4 and β5. Even if purely geometrical, the C2 symmetry of the barrel is an intrinsic feature of small barrels.

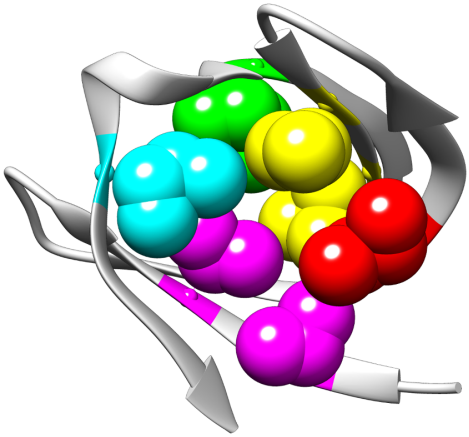
The KOW motif (Fig. 1D) [22] is found in some RNA-binding proteins (mostly small barrels in ribosomal proteins), it consist of β1, β2 and the loops preceding β1 and following β2 covering a total of 27 residues. Its hallmark is alternating hydrophilic and hydrophobic residues with an invariant Gly at position 11 [22].

Functional motifs (Fig. 1E) are described in Sm-like proteins (b.38). The Sm1 motif consists of β1- β3; the Sm2 motif consists of β4 - β5 bracketing short (4 residues) 3-10 helix [23]. The Sm2 motif with its β4 - β5 strands straddling the barrel is a very significant feature, and possibly a signature of all small barrels with an SH3-like topology. In fact superimposition of this pattern alone leads to a good structural alignment of the entire structural framework of the small barrels.

The hydrophobic core and structural framework

The hydrophobic core of the small barrel is minimalistic. It comprises the 6 elementary strands forming the conserved structural framework, (β1, β2N, β2C, β3, β4, β5) (Figure 2). These strands are short, comprising between 4 and 6 alternating inside/outside residues, unless bulges are present. Only two strands are completely saturated in terms of backbone hydrogen bonds: β1 and β3. The structural framework of β-strands is the key identifier of small barrels proteins, best represented by the Hfq barrel where all loops are reduced to tight beta turns. It is tolerant to diverse residue replacement, as long as a very small and tight hydrophobic core is preserved.

A.



B.

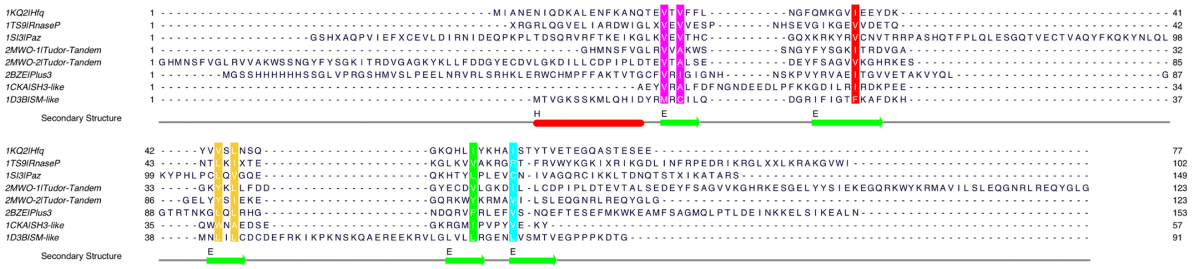


Figure 2. The hydrophobic core of the β-barrel. **A**. Seven residues (colored spheres) form the core of the β-barrel through hydrophobic interactions. Two residues have the same color if on the same strand. The β-barrel is represented as a grey ribbon. **B.** Sequence alignment of the β-barrel structures from Figure 3 (minus OB fold and Chromo). The seven core residues are highlighted with colored rectangles. Each strand contributes 1-2 residue(s) to the core. See S2 for multiple alignment.

Typically, between one or two inward-facing residues are contributed by each beta strand to the hydrophobic core. The two central strands β1 and β3 are contributing two residues each (in yellow and magenta in Figure 2) and the four lateral strands β2N, β2C, β4, and β5, usually contribute one residue each to the hydrophobic core. The β2N strand does not contribute consistently to the hydrophobic core, thus the minimum hydrophobic core consist of 7 residue (Figure 2). The hydrophobic residue in β2C follows Gly (which bends the strand) and is positioned at the beginning of the characteristic beta bulge. The hydrophobic residues in β4 and β5 are adjacent to the 3-10 helix - immediately preceding (β4) and immediately following (β5) the helix. The minimal core is what defines the stable barrel fold leaving all outwardly facing residues to interact with ligands via their side chains.

The hydrophobic core of 7 residues can be extended in a variety of ways (Table S1). Since the barrel is semi-open various decorations can add hydrophobic residues around the minimal core. For example, the N-terminal helix in Sm-like barrel extends β5-β2C of the otherwise open barrel. Similarly, the RT loop in SH3-like barrel extends the β2N,β3 side of the barrel.

The outward-facing residues on the 'edge' stands: β2C and β4 in Sheet A (Meander), β5 and β2N in Sheet B (N-C) have the potential to form hydrogen bonds with other β-strands, unless they are sterically obstructed by terminal decorations or long loops. Such strand-strand interactions can extend the β-sheet of the barrel (see Distal loop and Figure 3I) and enable formation of quaternary structures (see Oligomerization of the barrels and Figure 7).

Beyond the core: loops, decorations, extra modules

While the structural framework of the small barrel is a common denominator for different folds and the structures are easily superimposable on that framework, the other elements of the structure - loops, modules inserted within the loops, N-term and C-term extensions, are variable

(Fig. 3). These elements delineate specific structural families and define the function of the small beta barrel.

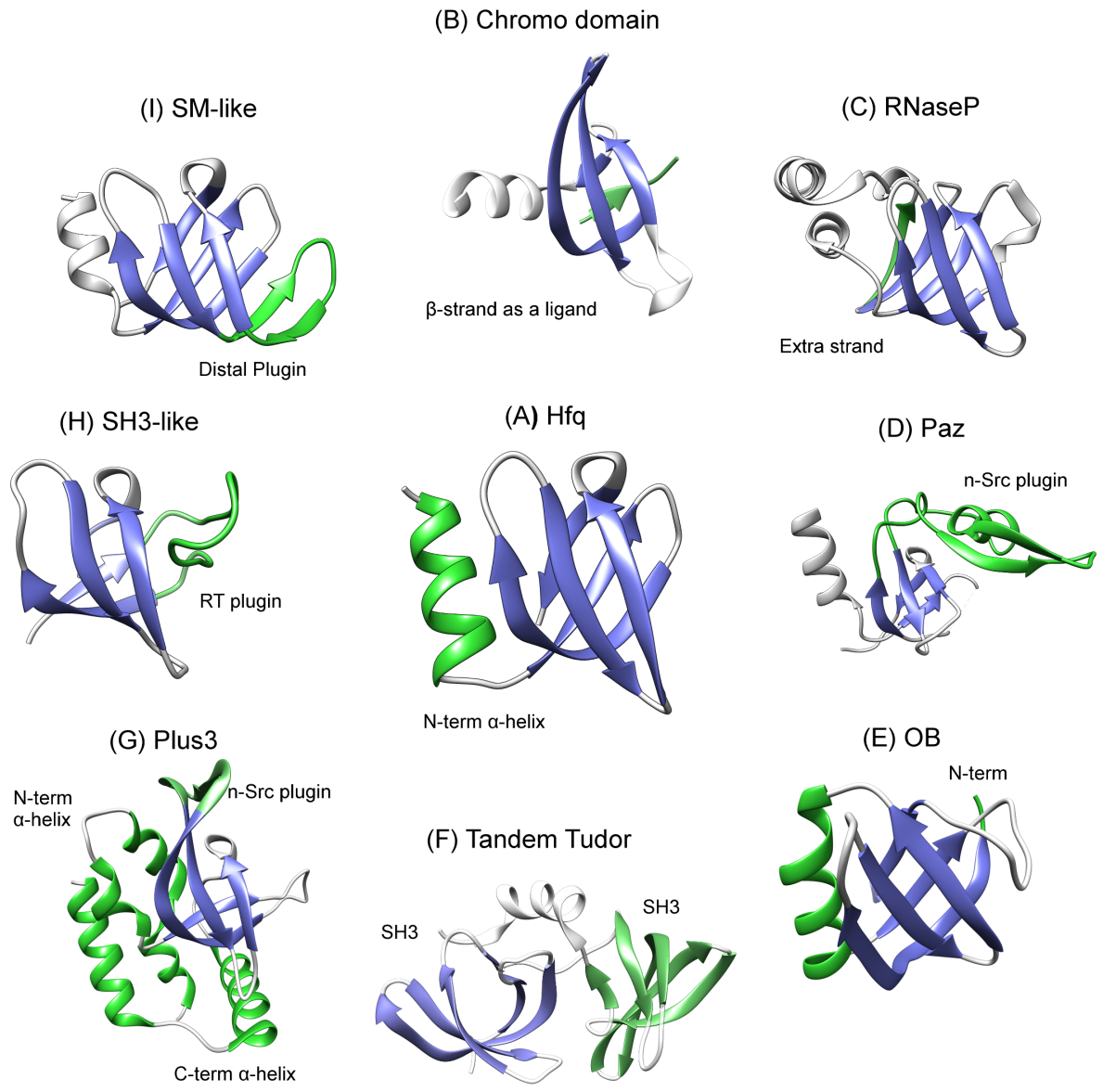


Figure 3. Key variations (green) in the small barrel structure (blue). **A.** Hfq (b.38.1.2, PDB ID:1KQ2) in the center is used as a reference structure having minimal loop lengths. **B.** Chromo domain (b.34.13.1, PDB ID:1KNA) has β5 contributed by a ligand. **C**. RNaseP subunit P29 (b.137.1, PDB ID:1TS9) has an extra strand β6. **D.** PAZ (b.34.14, PDB ID:1SI3) has a plugin into the N-src loop. **E.** OB fold (b.40, PDB ID:1C4Q) has a different topology. **F.** Tudor-Tandem SH3 (b.34.9.1, PDB ID:2MWO) has two tandemly positioned small barrels. **G.** Plus3 (b.34.21, PDB ID:2BZE) has additional helices at N- and C-termini and

an N-src plugin. **H.** SH3 (b.34.2, PDB ID:1CKA) has a plugin into the RT loop. **I.** Sm fold (b.38.1.1, PDB ID:1D3B) has a distal loop plugin.

Loop variations

Loops that connect the beta strands vary significantly in length and confer functional role(s) (see below). Overall there are 5 loops, the first precedes the first beta strand β1. A 6th loop is possible, the C-terminal loop, which connects the last beta strand to the C-terminal extension or to the 6th strand of the barrel, when present. Prior independent studies (Table 1) of each loop extension has led to independent naming schemes being used in the literature. Throughout this work we follow the annotation from the b.34 - SH3-like fold.

While there can be up to 6 loops in small beta barrels, the central four loops (RT, N-src, Distal, 3-10 helix as defined for the SH3-like fold b.34.2) are always present. Out of these 4 loops, significant changes in the length of three; RT, N-src and Distal, are observed and can be linked to specific functions. The fourth loop is almost always a short 3-10 helix (1 turn) and is, on rare occasions, a distorted version as in RPP29 [24] or replaced by a longer loop as in TrmB proteins [25]. Elongation of the loops often results in formation of additional secondary structures - as described below.

N-src loop (Fig. 3D, 3G): The elongated N-src loop is observed in two functional families. In the PAZ domain (b.34.14) of Piwi and Argonaute (RNA interference) the alpha/beta module, inserted into the N-src loop, is part of the aromatic pocket that secures the RNA molecule in place [4]. In the case of the Plus3 domain (b.34.21) of Rtf1 (elongation of transcription), the elongated N-src loop contains two tiny (3-residue long) beta strands and is involved in binding single stranded DNA [26].

RT loop (Fig. 3H): Long inserts into the RT loop, which connect strands β1 and β2, results in the classical SH3 (b.34.2) domain involved in signal transduction. The SH3 domain binds proline-rich sequences using the elongated RT loop (as well as N-src loop and 3-10 helix). The loop lies along the side of the barrel and caps one of its ends [27]; [28]. Various pairs of loops form various pockets. In the PAZ domain (b.34.14) of Piwi and Argonaute (RNA interference), aromatic residues of the elongated RT loop (Fig. 3D) are part of the aromatic pocket formed between it and the alpha/beta module (inserted into the N-src loop, see below); this pocket laterally secures the RNA substrate [4].

Distal loop (Fig. 3I): Elongation of the distal loop is observed in eukaryotic Sm proteins, which are part of the splicing machinery. An elongation in the distal loop results in elongation of the adjacent strands β3 and β4. These two long beta strands are now bent similar to that of β2 and can be seen as β3N and β3C, β4N and β4C [29]. Like β2 , they participate in the formation of two sheets simultaneously. This results in a much larger hydrogen-bonded Sheet B - now containing 5 strands, β5, β1, β2N, β3C, β4N. The original Sheet A remains the same.

3-10 helix: Connects strands β4 and β5 and is short (4 residues) and inflexible. It ultimately determines the relative positions of the β4 and β5 stands which frequently straddle the barrel. 3-10 helix is practically invariable in SH3-like folds but is absent in OB-fold for topological reasons (see below). In the cases of sac7d, sso7d and others histone-like small archaeal proteins a second 3-10 helix is found in the middle of β2 where typically a strand-bending Gly would be [5].

N- and C-term decorations, capping of the barrel, secondary structures in the loops

Alpha-helices and additional loops at the N- and C-termini are frequently observed in small beta barrels and sometimes termed ‘decorations’. Their position relative to the barrel core varies. In some cases they affect the ability of the barrel to oligomerize. The decorations, as any loop

insertion, almost always have a functionally significant role, adapted to specific situations, as demonstrated in the following selected examples.

N-term α-helix (Fig. 3A) in the Sm-like fold (b.38) is connected to the barrel via a short loop and has multiple interactions with both RNA and proteins. The α-helix stacks on top of the open barrel and lays on the proximal face when the oligomeric ring is formed [30]. In bacterial Hfq (b.38.1.2), the α-helix interacts with sRNA molecules through its three basic residues (Arg16, Arg17, and Arg19) and an acidic Gln8 [23]. In lsm proteins the same α-helix interacts with proteins Pat1C in the lsm1-7 [31] ring and with prp24 in the lsm2-8 ring [32]. In the case of Sm proteins (b.38.1.1), the same α-helix interacts with the beta sheet of the adjacent protomers during ring assembly [29]. In the case of SmD2, the long N-term results in an additional helix (h0), which interacts with U1 RNA as it leads it into the lumen of the ring [33,34].

C-term α-helices (Fig. 3C, 3G) can either augment existing binding, or interact with additional binding partners. In the case of the lsm1-7 ring (b.38.1.1), a long helix formed by the C-term tail of lsm1 lies across the central pore on the distal face of the ring, preventing the 3'-end of RNA from exiting through the distal surface [35].

N-term and C-term α-helices together (Fig. 3C, 3G) can interact to form a supporting structure/subdomain around the barrel as in the case of the Plus3 (b.34.21) domain of Rtf1 [26], where 3 N-term α-helices and a C-term α-helix form a 4-helical cluster that packs against one side of the barrel. The role of these helices is not clear, but the conservation of many residues points to an unknown functional significance.

C-term tails have the least spatial constraints among all the decorations. These can remain disordered and can vary in length significantly. Over 40 residues in SmD1 and SmD3 and over 150 residues in SmB/B' [29]. In the case of Sm proteins (b.38.1.1), the C-terminal tails of

SmB/B', SmD1 and SmD3 carry RG-rich repeats which are critical for the assembly of the barrels into the toroid ring [36–38]. Disordered C-term tails of Hfq (b.38.1.2) are proposed to extend out of the ring and be involved in interaction with various RNAs [39].

Small internal modules (Fig. 3D, 3G): consist of short secondary or super-secondary structures (α/β or purely α) inserted within the loops. These typically form a pocket against the barrel and are an integral part of barrel function. Examples include an αββ module inserted into the N-src loop of the PAZ domain (b.34.14) [4] and a β hairpin extension module appearing in the N-src loop of the Plus3 domain (b.34.21) of Rtf1 [26]. Insertions can be entire modules. For example some interdigitated Tudors can be seen as a Tudor domain inserted in the N-src loop. An eTudor is a Tudor inserted in the (SH3 equivalent) distal loop of an OB fold.

More structural variations

Extra β-strand: RNase P subunit Rpp29 (Fig. 3C)

An addition of a 6th strand to the barrel has been observed in RNase subunit P29 (b.137.1), where, after an extra beta turn β5-β6, a 6th beta strand extends the antiparallel Sheet B (N-C) to 4 antiparallel strands: (β6, β5, β1, β2N) [24,40].

Missing β-strands: chromo domain HP1 (Fig 3B)

In at least one case, that of the HP1 chromo domain (b.34.13.2), the complete barrel is formed only upon binding the peptide. HP1 exists as a 3-stranded sheet A (meander), it is the β-strand of the ligand peptide that initiates formation of the second beta sheet (N-C) around it [18].

OB fold (b.40) (Fig. 4): similar architecture, different topology.

Similar to the SH3-like barrel, the OB fold is a 5-stranded barrel, but with a somewhat different topology. One can relate the SH3-like and OB topologies through a (non-circular) permutation observed previously [41]. Our reference Hfq protein (Sm-like fold) lends itself perfectly to

comparison with the OB fold, as both topologies lead to the same structural framework (Fig. 4). To avoid confusion, we use OB strand mapping when discussing OB folds, the mapping is indicated in Table 3A.

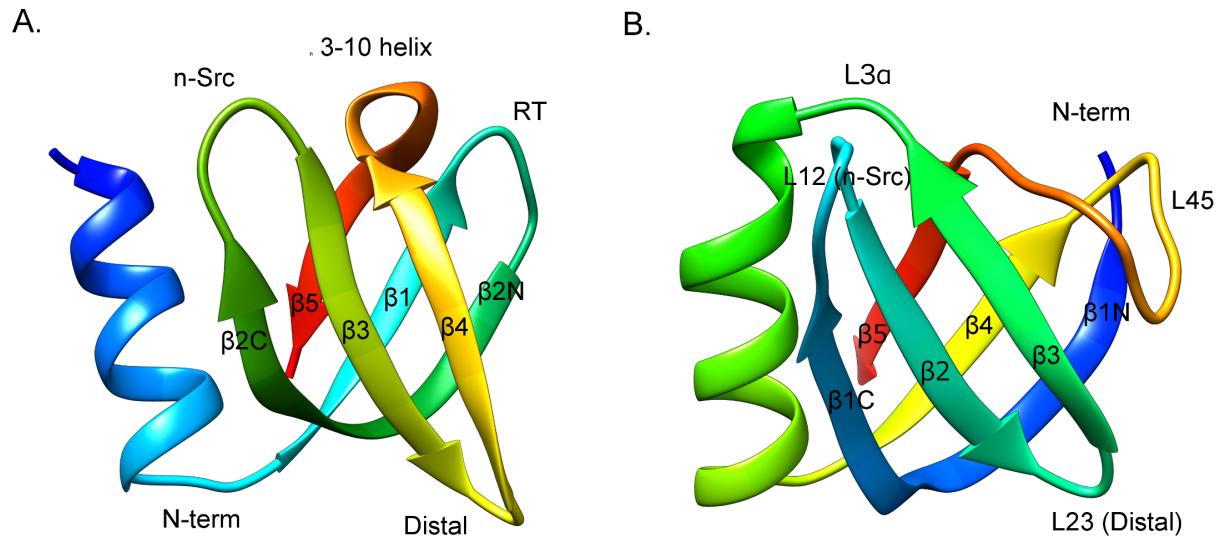


Figure 4. Comparison of OB and SH3-like folds mapping strands and loops. Coloring progresses from blue (N-terminus) to red (C-terminus). A. SH3-like fold (Hfq b.38.1.2) B. OB fold. See S3 for alignment.

The matching between the folds is particularly striking as both folds, OB-fold and Sm-like, have an N-terminal α-helix which is missing in the SH3-like fold. When starting with the Sm-like fold of Hfq, the permutation inserts the N-terminal α-helix and β1 after the Meander [β2C-β3-β4] and before β5, thus the initial topology [α-helix-β1]-[β2N-β2C-β3-β4]-[β5] results in the final topology [β2N-β2C-β3-β4]-[α-helix-β1]- [β5]. The renumbering of strands in this rearranged fold (now OB) will read [β1N-β1C-β2-β3]-[α-helix-β4]-β5]. The non-circular permutation preserves the Sheet A (Meander) in both topologies: [β2C-β3-β4] in Sm-like and [β1N-β1C-β2] in OB-fold. The structure alignment of [β2N-β2C-β3-β4-β5] in SH3 and [β1N-β1C-β2-β3]+β5 in OB results in 1.37 $Å^2$ RMS (based on the alignment of Hfq pdbid:1KQ1 and. verotoxin pdbid:1C4Q).

Of the five loops in the OB-fold (Table 3B), L12 can be clearly structurally mapped onto the N-src loop and L23 to the Distal loop of the SH3-like fold. There is no good structural correspondence between the other loops. The RT and 3-10 are unique to SH3-like topologies, while L3α, Lα4 and L45 are unique to the OB fold.

A.

OB-fold strands	SH3-like fold strands
β1	β2
β2	β3
β3	β4
β4	β1
β5	β5

B.

OB-fold strands	OB loops	Equivalent SH3 loops
β1-β2	L12	N-src
β2-β3	L23	Distal
β3-α	L3α	-
α-β4	Lα4	-
β4-β5	L45	-

Table 3. Mapping corresponding β-strands and loops between SH3-like and OB folds. The OB topology, using OB domain nomenclature, runs (β1-β2-β3)-(α-β4)-β5 where β2-β3-β4 is a meander

The hydrophobic core of the OB fold contains the 7 residues defined for the SH3-like fold, but is typically larger, by virtue of strands elongation (especially L12/N-src) and formation of possible hairpin within L45, which would then extend the Beta Sheet A by two strands.

Most notable loop variations in the OB-fold are similar to those of the SH3-like fold.

Insertion into the N-src loop: In the OB2 of BRCA2 there is an insertion of a Tower domain into the L12 loop (corresponding to the N-src loop); The Tower domain is implicated in DNA-binding. The Tower domain is a 154-residue long insert consisting of two long α-helices and a 3-helix bundle positioned between them [6,42]. In DBD-C of RPA70 a zinc-finger motif consisting of three short β-strands is inserted into the L12 loop [43].

Extension of the Distal loop: The DNA-binding domain of cdc13 contains a unique pretzel-shaped loop L23 (corresponding to the Distal loop), which significantly extends interactions of this barrel with DNA. A 30-residue long loop twists and packs across the side of the barrel and interacts with the L45 loop [44].

Change of internal α-helix and omega loop: In DBD-C of RPA70 the α-helix positioned between β3 and β4 is replaced by a helix-turn-helix, while in DBD-D of RPA32, the same α-helix is missing altogether and is replaced by a flexible loop [43].

Sequence variation and electrostatic charge

In addition to variations in structure, variations in sequence further distinguishes different barrels. Because small barrels are extremely insensitive to mutations (see discussion under "Folding of the small β-barrels"), a common evolutionary strategy is to modulate electrostatic interactions through changing the properties of the residues in loops, sheets and decorations. In some cases, it results in a switch between positively charged and negatively charged or between hydrophobic and polar/charged patches or entire sheets, resulting in different partners binding and different functions.

An interesting case is that of the HIN domains of AIM2 and p202 which are involved in the innate immune response (Figure 5) [45]. Each HIN domain consists of 2 tandem OB-fold

barrels which are shown to binds ss-, ds- and quadruplex DNA with various affinities. Well studied case of HIN domains binding dsDNA is that of innate immune response [7,46]. Even though there is 36% sequence identity between HIN domains in AIM2 and p202, the binding modes are completely different (Figure 5A,B), due to the variation in the distribution of electrostatic charges. In the case of AIM2 the binding occurs through positive charges on the convex surface of the barrel (Figure 5C,D). In the case of p202, the same surface is negatively charged and thus cannot interact with DNA. Instead the interaction occurs through the positively charged loops (of the second OB barrel) on the opposite side of the barrel (Figure 5E,F). The same loops carry hydrophobic residues in the case of AIM2 and thus cannot bind DNA [7,46] . The differences in binding surfaces result in different strength of binding which allows the two proteins to act antagonistically.

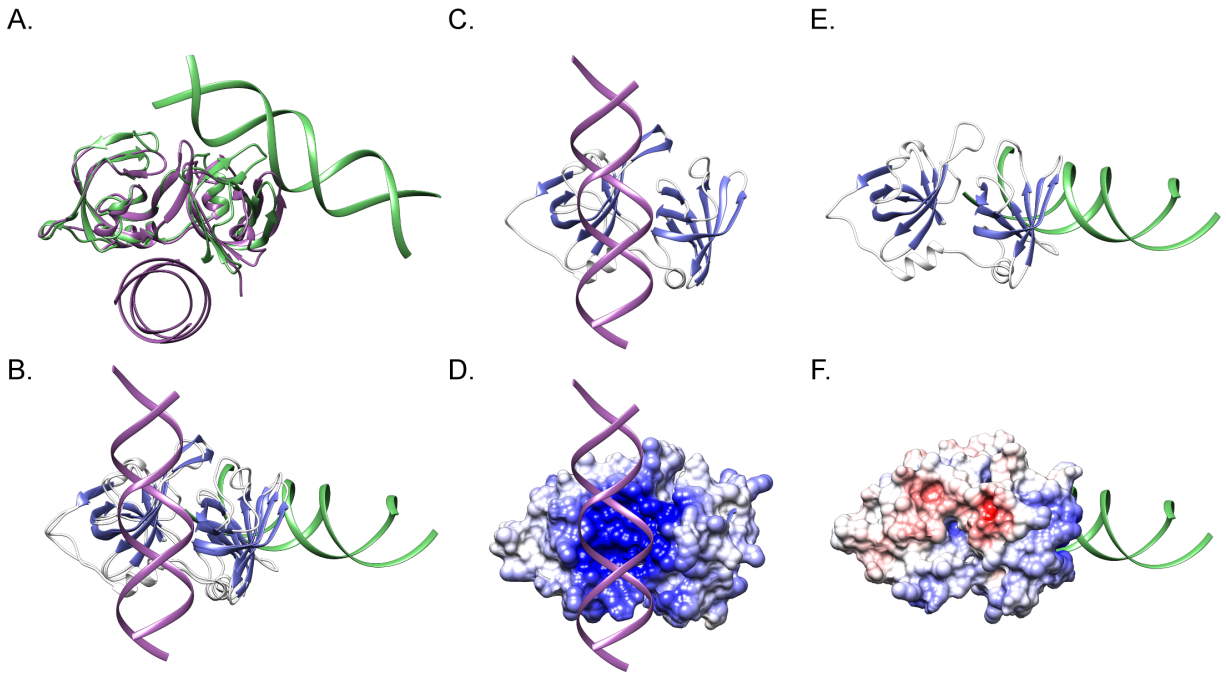


Figure 5. Changes in electrostatic charges on the surfaces of two very similar HIN domains. HIN domains consist of two tandem OB-fold barrels. **A.** Top view of superimposed HIN domains of AIM2 (purple) and of p202 (green) interacting with correspondingly colored DNA (green double helix interacts with p202, purple double helix interacts with AIM2). **B**. Front view of A, both proteins are now colored

identically. **C, D**. Interaction of HIN domain of AIM2 with DNA (PDB ID: 3RN5) through the front of domain using the first OB-fold barrel, the interacting surface is positively charged (blue). **E,F.** Interaction of the HIN domain of p202 with DNA (PDB ID: 4LNQ) occurs through the back side of the HIN domain using the second OF-fold barrel. The front side is negatively charged (red).

Another example of electrostatic variation are the five tandem small barrels containing the KOW motif which are present in transcription elongation factor Spt5 [47]. The distribution of electrostatic charge is very different in these barrels which reflects their functions. The KOW1-Linker has a very biased surface charge distribution. Its PCP (Positively Charged Patch) containing 6 basic residues can be mapped onto the KOW1 motif and is responsible for its interaction with DNA. On the other hand, the surfaces of KOW2-KOW3, which act jointly as a rigid body, have an even distribution of charge throughout and are not interacting with DNA [47].

Perhaps the most classical case illustrating the electrostatic variation in small barrels is the distribution of charge within the RT loop of polyPro-binding SH3 domains. The acidic residues in the RT loop and the basic residue within the polyproline signal are the key electrostatic interactions in addition to hydrophobic interactions involving prolines themselves. The position of the basic residue (Arg) determines the orientation of polypro peptide binding [48]. The strength of binding can be modulated by the number of acidic residues and their positioning within the RT loop [49]. In at least one case, Nck interaction with CD3epsilon, binding can be switched on or off by simply phosphorylating a key residue (Tyr) which will cause electrostatic repulsion between it and acidic residues in the RT loop [50].

Joining barrels together

Small barrels tend to work together at different structural scales. Interactions between tandem barrels within one protein are common. Individual barrels can interact to form toroidal rings that function in many aspects of RNA biogenesis in all superkingdoms of life. Finally, small barrels

can also form fibrils, which consist either of toroidal rings or individual small barrels. Each is described.

Tandem and embedded barrels

Several combinations of beta barrels positioned in tandem, or intertwined, are introduced.

SH3-SH3: tandem Tudors (Fig. 6A)

Two SH3-like barrels positioned in tandem typically form a barrel-to-barrel interface which can be constructed in various ways. In the case of 53BP1, H-bonding is formed between β2N of the first barrel and β5 of the second, thus joining individual 3-stranded β-sheets into an extended 6-stranded β-sheet. The C-terminal α-helix further strengthens the connection by interacting with multiple β-strands of both barrels [51].

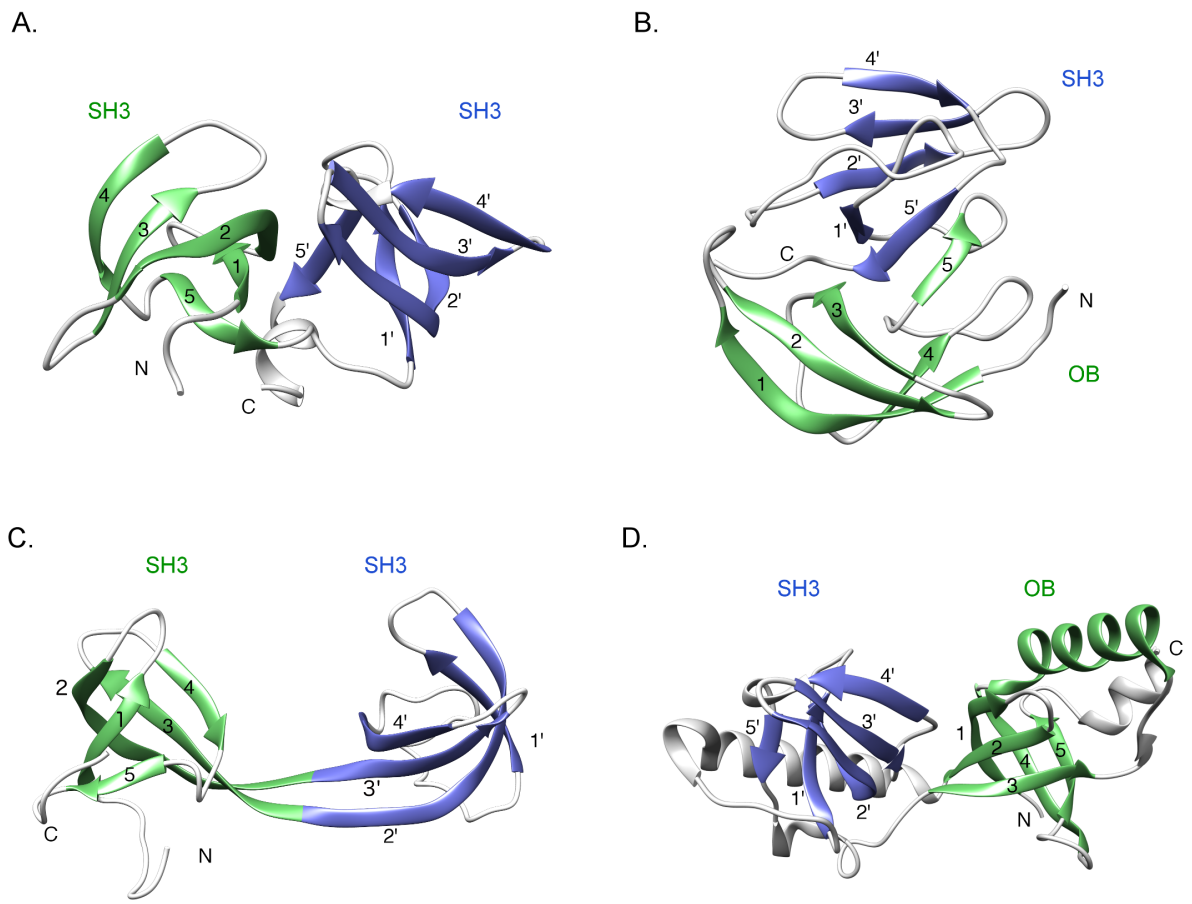


Figure 6. Small barrels can be combined in various ways. **A.** SH3-SH3 tandem barrel in 53BP1(PDB ID: 1SSF). **B.** OB-SH3 tandem barrel in ribosomal L2 (PDB ID: 1S72). **C.** SH3-SH3 interdigitated barrel in JMJD2A (PDB ID: 2QQR) . **D.** SH3 barrel embedded within OB barrel in TDRD (eTud) (PDB ID: 3OMC). The barrel closer to the N-term - whether it is a complete (A,B) or partial (C, D) barrel is colored in green, the second barrel is colored in blue. The β-strands in the SH3 and OB are labeled using nomenclature from Figure 1 and 4, respectively.

In the case of Spt5 which has 5 tandem KOW-containing Tudor domains, interactions between Tudor-2 and Tudor-3, which move as a single body, occurs through β5 of Tudor-2 and residues immediately following β5 in Tudor-3 [47]. In the case of KIN17, the interface is formed by N-terminal and C-terminal tails interacting with the linker connecting the two barrels [52].

Various linkers and sequences can lead to various tandem interfaces and different extended sheets, showing a remarkable plasticity. For example a Tandem Tudor (Agenet) would form a tandem through a b2N-b2N interface (Table 4).

OB-SH3 hybrid tandem barrels (Fig. 6B)

Combinations of OB and SH3 domains in tandem can occur in either order: in one of the oldest ribosomal proteins L2 and in translational elongation factor eIF5A [53,54]. In L2 N-terminal OB is connected to the SH3 by a 3-10 helix, which parallels the 3-10 helix between β4 and β5 in the SH3-like domain. The β5 strands from two barrels are arranged in antiparallel manner and in doing so extend the OB sheet [53] .

SH3-SH3: interdigitated Tudors (Fig. 6C)

The most involved interactions between two adjacent barrels occur in the inter-digitated Tudors JMJD2A [55] and RBBP1 [56]. This structure has been described as two barrels 'swapping' some strands, resulting in what is termed 'interdigitated barrels'. In these cases the long β2 and β3 strands are exchanged between the barrels, resulting in two structures where the first two strands belong to one 'linear' barrel and the other two strands belong to the other 'linear' barrel. An antiparallel beta sheet is formed along the entire length of β2β3-β2'β3'.

SH3 barrel embedded in OB (eTud) (Fig. 6D)

SND1 contains 5 tandem OB-fold domains with the SH3-fold inserted into the L23 (Distal loop equivalent) of the OB barrel [57]. This combination of domains is typically referred to as extended Tudor (eTudor or eTud). In Drosophila there are 11 tandem extended Tudors referred to also as maternal Tudors [58]. The extended Tudor (eTudor maternal Tudor) domain consists of the two β-strands from OB, the linker (containing an □-helix) and 5 β-strands of the SH3-like fold (Tudor) domain. Both parts of the split OB domain are essential for binding sDMA

(symmetrically dimethylated Arginines) in the protein tails. The OB-fold (SN domain) and SH-fold (Tudor domain) interact as a unit [57,59].

Oligomerization of the barrels

Single domain small barrel proteins frequently come together to form a quaternary structure. Beta strands of small barrels are typically calibrated (of equal length) and can dock sideways into each other to form backbone hydrogen bonds between lateral strands of adjacent barrels, thus lending themselves to dimerization or further oligomerization. Table 4 characterizes the combinations of beta strands H-bonding to each other that have been observed in forming dimers or oligomers (always in an antiparallel configuration).

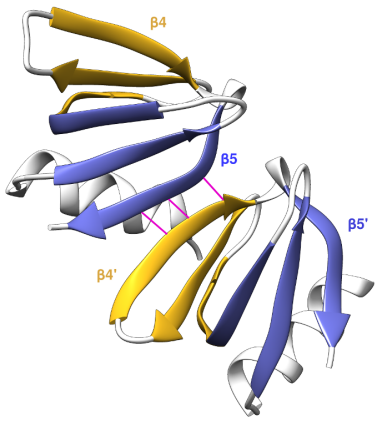
Interacting strands	**Oligomeric state**	**Protein name**	**SCOP family**	**Function**	**Reference**	**Symmetry**
β4-β5'	5-mer, 6-mer, 7-mer or 8-mer ring	Hfq, Sm, lsm	b.38.1 (Sm-like)	Splicing, RNA biogenesis	[3,29]	C5, C6,C7, C8
β2N-β5'	Dimer (covalently linked)	53BP1	b.34.9 (Royal family, Tudor)	Signal transducer in DNA repair	[51]	
β2N-β2N'	Pseudo-dimer (tandem; i.e. covalently linked)	FMRP	b.34.9 (Royal family, Agenet)	Fragile X syndrome	[60]	C2
β2C-β2C'	Dimer (Tetramers)	Mpp8	b.34.9 (Royal family, Chromo)	M phase phosphoprotein	[61]	C2
β2C-β5' (SH3 nomenclature)	5-mer ring	Verotoxin	b.40.2 (OB fold, bacterial enterotoxins)	Aids entrance of the toxin	[62]	c5
β5-β5'	Hybrid Tandem (covalently linked)	RL2	B.40+b.34 (OB + SH3)	Translation	[53]	None

Table 4. Strand-strand interactions in barrels quaternary and pseudo-quaternary arrangements.

Note that in the case of β2C-β5' SH3 strand nomenclature is used, even though this is an OB-fold and hence strand numbering differ (see Table 3A for mapping). In the case RL2, the β5-β5' antiparallel bonding is possible due to conformational changes in OB-fold domain of RL2.

The best known cases of oligomerization are the toroidal rings formed by SM-like proteins. Sm/lsm proteins are the fundamental components of the spliceosome. The bacterial counterpart, Hfq, is broadly involved in RNA biogenesis. The uniquely positioned β4-(3-10)-β5 strands, which straddle the body of the barrel, lead to interactions between the β4 strand of one monomer and the β5 strand of the adjacent monomer (Fig. 7A), ultimately connecting between 5 and 8 monomers into a doughnut-shaped ring (Fig. 7B). This process connects a 3-stranded Sheet A of one monomer with a 3-stranded Sheet B, forming a 6-stranded sheet or blades, which connects two surfaces of the toroidal ring and can be seen in a similar way as a β-propeller architecture. The two faces of the toroidal ring are formed by the two beta sheets of individual barrels: Sheet A (Meander) forms the *Distal* face, while Sheet B (N-C) forms the *Proximal* face. Lateral region of the ring [30] - also referred descriptively as outer rim [35] - consists of residues connecting two faces of the ring (Distal and Proximal) and facing outwards. In some cases (Hfq) it has a specific function as well [30,35].

A.



B.

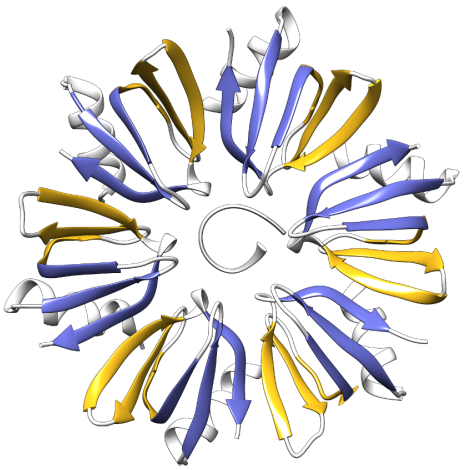


Figure 7. Oligomerization of the small barrels into a doughnut-shaped ring. **A**. hydrogen bonds are formed between β4 strand of one protomer and β5 strand of the adjacent protomer (magenta lines) **B**. A doughnut-shaped/toroidal ring is formed from 6 (Hfq) (shown), 7 (Sm/lsm) or 8 (lsm) protomers. *Meander* sheets (β2C, β3, β4) of the individual barrels create a contiguous beta sheet across the entire *Distal* face of the ring.

A further oligomerization of the toroidal rings into long tubes is observed in bacterial Hfq and Archaeal Sm-like (Sm-AP) proteins. This formation arises through stacking the faces of the rings (Proximal-to-Proximal (check), in case of SmAP) or through stacking of slabs - each slab consisting of 6 hexameric rings (for Hfq) [15].

Another case of ring formation comes from the OB-fold: a 5-mer ring is formed via hydrogen bonding β1 of one monomer with β5 of the other [62].

Dimers are frequently formed between two tandemly repeated domains - as in the case of Tudor via β2-β5 interactions (53BP1) and Agenet via β2N-β2N' interactions (FMRP, [60]).

Not all small barrels are able to oligomerize through strand-strand hydrogen bonding of the backbone. Elongation of loops, or addition of N- or C-term decorations, often prevents the strand-to-strand interaction necessary for oligomer formation. For example, the RT loop in polyPro binding SH3 domain (b.34.2) physically covers strand β4 and thus precludes β4-β5 hydrogen bonding and toroidal ring formation. Indeed oligomeric structures are completely absent from the SH3-like fold b.34. Similarly, the elongation of the N-src loop in the case of the PAZ domain and the N-term and C-term extensions in the case of Plus3 domain of Rft preclude β4-β5' formation.

Oligomerization is also possible through side chain interactions among loop residues, as in the cases of tetramer formation of HIN domains (b.40.16) [7] or through side chain interactions between strands as in dimer formation of viral integrase (b.34.7) [63,64].

Oligomerization of barrels also occurs in multi-domain proteins contributing to the formation of large structures such as cell puncturing device in bacteriophage T4 (trimer; barrel is an N-terminal domain) [65] and in MscS mechanosensitive channel in E.coli (heptamer; barrel is a central domain) [66]

Fibril formation

Beta structures are prone to polymerization and formation of fibrils, leading to amyloid formations which could be functional or disease-causing. There are several ways in which fibrils can form, either starting from individual small barrels, or starting from toroidal rings.

A common pathway to fibril formation for SH3 polyPro-binding domains begins with domain swapping between two protomers, in which any loop (RT, N-src or Distal) can function as a hinge to partially open the beta barrel and exchange beta-strands with the other protomer. Such open interacting loop regions become rigid and may contain short β-strands. These β-strands then serve as a nucleation center for amyloid formation [67]. Alternatively, the hydrophobic strand β1 may not undergo typical pairing with β5 if the latter is disordered, thereby forming non-native contacts with β1 of other protomers and forming aggregation-prone intermediates which lead to fibril formation [68].

Both of these pathways are strongly tied to the folding process. Mutations that destabilize folding are found predominantly in the open loops/hinges and unpaired strands. This ultimately leads to non-native folds with swapped domains, polymerization through beta-strands and formation of fibrils.

Fibril formation from toroidal rings proceeds through an entirely different mechanism. In E.coli Hfq rings self-assemble into slab-like layers, each layer built of 6 hexameric rings. The fibrils are then built out of such layers [15]. The C-terminal fragment of Hfq (constituting 30% of the protein) is intrinsically disordered and was shown to be critical for the assembly of fibrils into higher order cellular structures [69].

In archaea fibrils are formed by stacking hexameric rings of SmAP1 in a head-to-head manner thus forming 14-mer rings, which ultimately self-assemble into striated bundles of polar tubes [15,70].

Folding of the small β-barrels

In-depth folding studies were done on SH3 domains that bind polyPro (b.34.2) and on OB domains in Cold Shock Proteins (b.40.4.5). The folding of the beta-barrels is simple and immutable - the same fold is achieved by domains with great sequence diversity and also when the sequences are permuted. For SH3 domains the folding proceeds through two-state kinetics: Unfolded (U) --> Folded (F). The high energy transition state is characterized by having multiple conformations of partially collapsed structure and is referred to as the Transition State Assembly (TSE). It has been consistently found that the partially folded states (TSE) are highly polarized - they contain the hydrophobic nucleus which includes most of *Beta Sheet A* or *Meander*, β2-β3-β4, while *Beta Sheet B* (or *N-C*) which includes β1-β5, is disordered in the TSE [68,71–73].

In the case of OB-fold (CSPA and CSPB), an intermediate state has been recently proposed, it too consists of the 3-stranded beta sheet: β1-β2-β3, which structurally correspond to the *Meander* in SH3 [74].

The robustness of the folding process is in large part due to the cooperatively, which stresses the significance of local interactions during folding: residues that initiate the folding process are local in the sequence (referred to as structure topology) [72,75,76]. The model of the

hydrophobic zipper (HZ) [77] which begins with local interactions and eventually bring more distant residues together to form the hydrophobic core through beta-hairpin formation, supports these observations on small beta barrels. The HZ structure is formed by a group of neighboring residues (cooperativity), eliminating the reliance on specific residues for tertiary folding. Indeed the formation of a 3β-strand meander in WW proteins - a well-studied system - always initiates within one or both of the loops/turns [78–80]. The folding of CspA/CspB is also initiated within the loops - as would be expected, if interactions among the local residues drive the folding [74,81].

The significance of local interactions is also supported by circular permutation experiments of the alpha-spectrin SH3 domain [73]; [82]. In these experiments C- and N-termini are linked together and the sequence is cut open in one of the three loops, rearranging the linear order of secondary structures. The same fold is reached in all permuted structures, however, the order of folding is different, as the beta-hairpin formed by the linked ends (β1-β5) appears early in the folding process.

Perhaps the most poignant evidence of the resilience of the small beta barrel structure, as it relates to folding, is the unprecedented case of RfaH. In RfaH the C-terminal domain spontaneously switches from an alpha hairpin (when bound to N-term domain) to the small beta barrel structure (when released from interaction with N-terminal domain). Such a change in structure has far reaching functional consequences, such that RfaH has a role in both transcriptional elongation and initiation of translation [83].

Conclusions

There are at least three take home messages from this review of the structure of small beta barrels. First, biologically, while small beta barrels have a structurally immutable core, there is an absence of sequence conservation even among related protein families. From an

evolutionary perspective this implies an ancient fold, or convergent evolution, or both. It further implies a large functional repertoire derived through structural variations. With emphasis on structure, this is what is described herein. Second, pragmatically, it is impossible to cover all aspects of small beta barrels in a single paper, or series of papers. Here we highlight salient structural features and indicate the functional implications. We are in the process of completing a series of papers that delve more deeply into functions involving DNA, RNA, and protein binding, respectively. Even there Nature does not let us conveniently pigeon hole structure - function relationships. Third, systematically, there is a need for structural descriptions that go beyond the basic SCOP, CATH [84], or ECOD classifications. Historically, as in so many areas of scientific nomenclature, alternative descriptive names emerge for describing the same or closely related entities. Here, in the case of small beta barrels, we have tried to identify alternative nomenclatures and map them to each other as far as possible.

There is no doubt, given the diversity of sequence and function, that small beta barrels are a special fold which we have summarized here in a concise a way as we are able. While such a review provides details of structural immutability, it is worth looking deeper into basic structural principles that give rise to this incredible flexibility in function.

Acknowledgements

This research was supported by the Intramural Research Program of NIH National Library of Medicine.

References.

1. Wilusz CJ, Wilusz J. Lsm proteins and Hfq. RNA Biol. 2013;10: 592–601.

2. Vogel J, Luisi BF. Hfq and its constellation of RNA. Nat Rev Microbiol. 2011;9: 578–589.

3. Mura C, Randolph PS, Patterson J, Cozen AE. Archaeal and eukaryotic homologs of Hfq: A

structural and evolutionary perspective on Sm function. RNA Biol. 2013;10: 636–651.

4. Ma J-B, Ye K, Patel DJ. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature. 2004;429: 318–322.

5. Robinson H, Gao YG, McCrary BS, Edmondson SP, Shriver JW, Wang AH. The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. Nature. 1998;392: 202–205.

6. Bochkarev A, Bochkareva E. From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. Curr Opin Struct Biol. 2004;14: 36–42.

7. Yin Q, Sester DP, Tian Y, Hsiao Y-S, Lu A, Cridland JA, et al. Molecular mechanism for p202-mediated specific inhibition of AIM2 inflammasome activation. Cell Rep. 2013;4: 327–339.

8. Lewis KA, Wuttke DS. Telomerase and telomere-associated proteins: structural insights into mechanism and evolution. Structure. 2012;20: 28–39.

9. Patel DJ, Wang Z. Readout of epigenetic modifications. Annu Rev Biochem. 2013;82: 81–118.

10. McCarty JH. The Nck SH2/SH3 adaptor protein: a regulator of multiple intracellular signal transduction events. Bioessays. 1998;20: 913–921.

11. Chandonia J-M, Fox NK, Brenner SE. SCOPe: Manual Curation and Artifact Removal in the Structural Classification of Proteins - extended Database. J Mol Biol. 2017;429: 348–355.

12. Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, et al. ECOD: an evolutionary classification of protein domains. PLoS Comput Biol. 2014;10: e1003926.

13. Murzin AG, Lesk AM, Chothia C. Principles determining the structure of beta-sheet barrels in proteins. II. The observed structures. J Mol Biol. 1994;236: 1382–1400.

14. Murzin AG, Lesk AM, Chothia C. Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis. J Mol Biol. 1994;236: 1369–1381.

15. Arluison V, Mura C, Guzmán MR, Liquier J, Pellegrini O, Gingery M, et al. Three-dimensional structures of fibrillar Sm proteins: Hfq and other Sm-like proteins. J Mol Biol. 2006;356: 86–96.

16. Weber G, Trowitzsch S, Kastner B, Lührmann R, Wahl MC. Functional organization of the Sm core in the crystal structure of human U1 snRNP. EMBO J. 2010;29: 4172–4184.

17. Zhang R, So BR, Li P, Yong J, Glisovic T, Wan L, et al. Structure of a key intermediate of the SMN complex reveals Gemin2's crucial function in snRNP assembly. Cell. 2011;146: 384–395.

18. Jacobs SA, Khorasanizadeh S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science. 2002;295: 2080–2083.

19. McLachlan AD. Gene duplications in the structural evolution of chymotrypsin. J Mol Biol. 1979;128: 49–79.

20. Caetano-Anollés G, Caetano-Anollés D. An evolutionarily structured universe of protein architecture. Genome Res. 2003;13: 1563–1571.

21. Chothia C, Janin J. Orthogonal packing of beta-pleated sheets in proteins. Biochemistry. 1982;21: 3955–3965.

22. Kyrpides NC, Woese CR, Ouzounis CA. KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. Trends Biochem Sci. 1996;21: 425–426.

23. Schumacher MA, Pearson RF, Møller T, Valentin-Hansen P, Brennan RG. Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. EMBO J. 2002;21: 3546–3556.

24. Sidote DJ, Heideker J, Hoffman DW. Crystal structure of archaeal ribonuclease P protein

aRpp29 from Archaeoglobus fulgidus. Biochemistry. 2004;43: 14128–14138.

25. Krug M, Lee S-J, Diederichs K, Boos W, Welte W. Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB. J Biol Chem. 2006;281: 10976–10982.

26. de Jong RN, Truffault V, Diercks T, Ab E, Daniels MA, Kaptein R, et al. Structure and DNA binding of the human Rtf1 Plus3 domain. Structure. 2008;16: 149–159.

27. Lim WA. Reading between the lines: SH3 recognition of an intact protein. Structure. 1996;4: 657–659.

28. Yu H, Rosen MK, Shin TB, Seidel-Dugan C, Brugge JS, Schreiber SL. Solution structure of the SH3 domain of Src and identification of its ligand-binding site. Science. 1992;258: 1665–1668.

29. Kambach C, Walke S, Young R, Avis JM, de la Fortelle E, Raker VA, et al. Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell. 1999;96: 375–387.

30. Sauer E. Structure and RNA-binding properties of the bacterial LSm protein Hfq. RNA Biol. 2013;10: 610–618.

31. Wu D, Muhlrad D, Bowler MW, Jiang S, Liu Z, Parker R, et al. Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. Cell Res. 2014;24: 233–246.

32. Karaduman R, Dube P, Stark H, Fabrizio P, Kastner B, Lührmann R. Structure of yeast U6 snRNPs: arrangement of Prp24p and the LSm complex as revealed by electron microscopy. RNA. 2008;14: 2528–2537.

33. Pomeranz Krummel DA, Oubridge C, Leung AKW, Li J, Nagai K. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature. 2009;458: 475–480.

34. Li J, Leung AK, Kondo Y, Oubridge C, Nagai K. Re-refinement of the spliceosomal U4 snRNP core-domain structure. Acta Crystallogr D Struct Biol. 2016;72: 131–146.

35. Weichenrieder O. RNA binding by Hfq and ring-forming (L)Sm proteins: a trade-off between optimal sequence readout and RNA backbone conformation. RNA Biol. 2014;11: 537–549.

36. Selenko P, Sprangers R, Stier G, Bühler D, Fischer U, Sattler M. SMN tudor domain structure and its interaction with the Sm proteins. Nat Struct Biol. 2001;8: 27–31.

37. Friesen WJ, Paushkin S, Wyce A, Massenet S, Pesiridis GS, Van Duyne G, et al. The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. Mol Cell Biol. 2001;21: 8289–8300.

38. Grimm C, Chari A, Pelz J-P, Kuper J, Kisker C, Diederichs K, et al. Structural basis of assembly chaperone- mediated snRNP formation. Mol Cell. 2013;49: 692–703.

39. Beich-Frandsen M, Vecerek B, Konarev PV, Sjöblom B, Kloiber K, Hämmerle H, et al. Structural insights into the dynamics and function of the C-terminus of the E. coli RNA chaperone Hfq. Nucleic Acids Res. 2011;39: 4900–4915.

40. Numata T, Ishimatsu I, Kakuta Y, Tanaka I, Kimura M. Crystal structure of archaeal ribonuclease P protein Ph1771p from Pyrococcus horikoshii OT3: an archaeal homolog of eukaryotic ribonuclease P protein Rpp29. RNA. 2004;10: 1423–1432.

41. Agrawal V, Kishan RK. Functional evolution of two subtly different (similar) folds. BMC Struct Biol. 2001;1: 5.

42. Yang H, Jeffrey PD, Miller J, Kinnucan E, Sun Y, Thoma NH, et al. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. Science. 2002;297: 1837–1848.

43. Bochkareva E, Korolev S, Lees-Miller SP, Bochkarev A. Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. EMBO J. 2002;21:

1855–1863.

44. Mitton-Fry RM, Anderson EM, Theobald DL, Glustrom LW, Wuttke DS. Structural basis for telomeric single-stranded DNA recognition by yeast Cdc13. J Mol Biol. 2004;338: 241–255.

45. Shaw N, Liu Z-J. Role of the HIN domain in regulation of innate immune responses. Mol Cell Biol. 2014;34: 2–15.

46. Jin T, Perry A, Jiang J, Smith P, Curry JA, Unterholzner L, et al. Structures of the HIN domain:DNA complexes reveal ligand binding and activation mechanisms of the AIM2 inflammasome and IFI16 receptor. Immunity. 2012;36: 561–571.

47. Meyer PA, Li S, Zhang M, Yamada K, Takagi Y, Hartzog GA, et al. Structures and Functions of the Multiple KOW Domains of Transcription Elongation Factor Spt5. Mol Cell Biol. 2015;35: 3354–3369.

48. Lim WA, Richards FM, Fox RO. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. Nature. 1994;372: 375–379.

49. Wu X, Knudsen B, Feller SM, Zheng J, Sali A, Cowburn D, et al. Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. Structure. 1995;3: 215–226.

50. Takeuchi K, Yang H, Ng E, Park S-Y, Sun Z-YJ, Reinherz EL, et al. Structural and functional evidence that Nck interaction with CD3epsilon regulates T-cell receptor activity. J Mol Biol. 2008;380: 704–716.

51. Charier G, Couprie J, Alpha-Bazin B, Meyer V, Quéméneur E, Guérois R, et al. The Tudor tandem of 53BP1: a new structural motif involved in DNA and RG-rich peptide binding. Structure. 2004;12: 1551–1562.

52. le Maire A, Schiltz M, Stura EA, Pinon-Lataillade G, Couprie J, Moutiez M, et al. A tandem of SH3-like domains participates in RNA binding in KIN17, a human protein activated in

response to genotoxics. J Mol Biol. 2006;364: 764–776.

53. Nakagawa A, Nakashima T, Taniguchi M, Hosaka H, Kimura M, Tanaka I. The three-dimensional structure of the RNA-binding domain of ribosomal protein L2; a protein at the peptidyl transferase center of the ribosome. EMBO J. 1999;18: 1459–1467.

54. Dever TE, Gutierrez E, Shin B-S. The hypusine-containing translation factor eIF5A. Crit Rev Biochem Mol Biol. 2014;49: 413–425.

55. Huang Y, Fang J, Bedford MT, Zhang Y, Xu R-M. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science. 2006;312: 748–751.

56. Gong W, Wang J, Perrett S, Feng Y. Retinoblastoma-binding protein 1 has an interdigitated double Tudor domain with DNA binding activity. J Biol Chem. 2014;289: 4882–4895.

57. Liu K, Chen C, Guo Y, Lam R, Bian C, Xu C, et al. Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. Proc Natl Acad Sci U S A. 2010;107: 18398–18403.

58. Ren R, Liu H, Wang W, Wang M, Yang N, Dong Y-H, et al. Structure and domain organization of Drosophila Tudor. Cell Res. 2014;24: 1146–1149.

59. Friberg A, Corsini L, Mourão A, Sattler M. Structure and ligand binding of the extended Tudor domain of D. melanogaster Tudor-SN. J Mol Biol. 2009;387: 921–934.

60. Myrick LK, Hashimoto H, Cheng X, Warren ST. Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. Hum Mol Genet. 2015;24: 1733–1740.

61. Chang Y, Horton JR, Bedford MT, Zhang X, Cheng X. Structural insights for MPP8 chromodomain interaction with histone H3 lysine 9: potential effect of phosphorylation on methyl-lysine binding. J Mol Biol. 2011;408: 807–814.

62. Stein PE, Boodhoo A, Tyrrell GJ, Brunton JL, Read RJ. Crystal structure of the cell-binding

B oligomer of verotoxin-1 from E. coli. Nature. 1992;355: 748–750.

63. Lutzke RA, Plasterk RH. Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. J Virol. 1998;72: 4841–4848.

64. Yin Z, Shi K, Banerjee S, Pandey KK, Bera S, Grandgenett DP, et al. Crystal structure of the Rous sarcoma virus intasome. Nature. 2016;530: 362–366.

65. Kanamaru S, Leiman PG, Kostyuchenko VA, Chipman PR, Mesyanzhinov VV, Arisaka F, et al. Structure of the cell-puncturing device of bacteriophage T4. Nature. 2002;415: 553–557.

66. Bass RB, Strop P, Barclay M, Rees DC. Crystal structure of Escherichia coli MscS, a voltage-modulated and mechanosensitive channel. Science. 2002;298: 1582–1587.

67. Cámara-Artigas A. Crystallographic studies on protein misfolding: Domain swapping and amyloid formation in the SH3 domain. Arch Biochem Biophys. 2016;602: 116–126.

68. Neudecker P, Robustelli P, Cavalli A, Walsh P, Lundström P, Zarrine-Afsar A, et al. Structure of an intermediate state in protein folding and aggregation. Science. 2012;336: 362–366.

69. Fortas E, Piccirilli F, Malabirade A, Militello V, Trépout S, Marco S, et al. New insight into the structure and function of Hfq C-terminus. Biosci Rep. 2015;35. doi:10.1042/BSR20140128

70. Mura C, Kozhukhovsky A, Gingery M, Phillips M, Eisenberg D. The oligomerization and ligand-binding properties of Sm-like archaeal proteins (SmAPs). Protein Sci. 2003;12: 832–847.

71. Chu W-T, Zhang J-L, Zheng Q-C, Chen L, Zhang H-X. Insights into the folding and unfolding processes of wild-type and mutated SH3 domain by molecular dynamics and replica exchange molecular dynamics simulations. PLoS One. 2013;8: e64886.

72. Riddle DS, Grantcharova VP, Santiago JV, Alm E, Ruczinski I, Baker D. Experiment and theory highlight role of native state topology in SH3 folding. Nat Struct Biol. 1999;6: 1016–1024.

73. Viguera AR, Blanco FJ, Serrano L. The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. J Mol Biol. 1995;247: 670–681.

74. Huang L, Shakhnovich EI. Is there an en route folding intermediate for Cold shock proteins? Protein Sci. 2012;21: 677–685.

75. Martínez JC, Serrano L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. Nat Struct Biol. 1999;6: 1010–1016.

76. Baker D. A surprising simplicity to protein folding. Nature. 2000;405: 39–42.

77. Dill KA, Fiebig KM, Chan HS. Cooperativity in protein-folding kinetics. Proc Natl Acad Sci U S A. 1993;90: 1942–1946.

78. Davis CM, Dyer RB. The Role of Electrostatic Interactions in Folding of β-Proteins. J Am Chem Soc. 2016;138: 1456–1464.

79. Jager M, Deechongkit S, Koepf EK, Nguyen H, Gao J, Powers ET, et al. Understanding the mechanism of beta-sheet folding from a chemical and biological perspective. Biopolymers. 2008;90: 751–758.

80. Maisuradze GG, Medina J, Kachlishvili K, Krupa P, Mozolewska MA, Martin-Malpartida P, et al. Preventing fibril formation of a protein by selective mutation. Proc Natl Acad Sci U S A. 2015;112: 13549–13554.

81. Vu DM, Brewer SH, Dyer RB. Early turn formation and chain collapse drive fast folding of the major cold shock protein CspA of Escherichia coli. Biochemistry. 2012;51: 9104–9111.

82. Martínez JC, Viguera AR, Berisio R, Wilmanns M, Mateo PL, Filimonov VV, et al. Thermodynamic analysis of alpha-spectrin SH3 and two of its circular permutants with different loop lengths: discerning the reasons for rapid folding in proteins. Biochemistry. 1999;38: 549–559.

83. Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney RA, Landick R, et al. An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. Cell. 2012;150: 291–303.

84. Dawson NL, Lewis TE, Das S, Lees JG, Lee D, Ashford P, et al. CATH: an expanded resource to predict protein function through structure and sequence. Nucleic Acids Res. 2017;45: D289–D295.