CellPress

Structure
Review

The Small β-Barrel Domain: A Survey-Based Structural Analysis

Philippe Youkharibache,[1,4] Stella Veretnik,[1,3,*] Qingliang Li,[1] Kimberly A. Stanek,[2] Cameron Mura,[2,3,*] and Philip E. Bourne[1,3,*]
[1]National Center for Biotechnology Information, The National Library of Medicine, The National Institutes of Health, Bethesda, MD 20894, USA
[2]Department of Chemistry, University of Virginia, Charlottesville, VA 22904, USA
[3]Present address: Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908, USA
[4]Present address: National Cancer Institute; The National Institutes of Health, Bethesda, MD 20894, USA
*Correspondence: veretnik@sdsc.edu (S.V.), cmura@virginia.edu (C.M.), peb6a@virginia.edu (P.E.B.)
https://doi.org/10.1016/j.str.2018.09.012

The small β-barrel (SBB) is an ancient protein structural domain characterized by extremes: it features a broad range of structural varieties, a deeply intricate evolutionary history, and it is associated with a bewildering array of cellular pathways. Here, we present a thorough, survey-based analysis of the structural properties of SBBs. We first consider the defining properties of the SBB, including various systems of nomenclature used to describe it, and we introduce the unifying concept of an "*urfold*." To begin elucidating how vast functional diversity can be achieved by a relatively simple domain, we explore the anatomy of the SBB and its representative structural variants. Many SBB proteins assemble into cyclic oligomers as the biologically functional units; these oligomers often bind RNA, and typically exhibit great quaternary structural plasticity (homomeric and heteromeric rings, variable subunit stoichiometries, etc.). We conclude with three themes that emerge from the rich *structure* ↔ *function* versatility of the SBB.

Introduction

What Is the Small β-Barrel Domain? Why Study It?

The small β-barrel (SBB) is a phylogenetically pervasive and functionally diverse protein structural domain that we define and systematically analyze herein. SBBs, which comprise a subset of all β-barrel-containing protein domains, exhibit a complicated, deeply intricate evolutionary history, and they are associated with myriad physiological functions. Apart from its β-rich secondary structural composition, the SBB domain is characterized by a small size (≲100 residues over five β-strands) and a highly conserved structural framework; this "structural framework" transcends the "fold" and "superfold" levels of structural similarity classification schemes (see, e.g., (Dessailly et al., 2017), for technical descriptions), motivating us to propose the term "*urfold*" to describe the SBB concept (see below). As a reference point in terms of domain sizes, note that membrane protein β-barrels range from 8 to >20 strands, and even the smallest known ones exceed ≈150 residues, while larger ones are >700 residues (see the first tables in Fairman et al., 2011; Tamm et al., 2004); representative examples of large-, medium-, and small-sized barrels (be they membrane or soluble) are illustrated in Figure S1. Because they differ in many fundamental respects, membrane barrels appear only peripherally here; they lie beyond the main focus of this work.

A hallmark of the many known SBB-containing proteins is their unusually broad functional diversity, involving RNA, DNA, and other proteins, as detailed in the section on Variability in Cellular Pathways. The SBB structures found in these functional contexts are described in this study, and further information about biochemical functionalities can be found in the reviews cited herein. Note that the term "small β-barrel" has appeared sporadically in the literature over the past two decades. To our knowledge, no precise definition of an "SBB" has been given; also, in such cases where the term has been used in the literature, the particular 3D structures either comply with our definition (given below), correspond to a subset of our definition, or else refer to *relatively* small membrane barrels (which still significantly exceed what is defined here as a "small β-barrel"; see also Figure S1).

Although their structures superficially resemble the β-barrels found in membrane proteins (Figure S1), many of which are functionally constrained (e.g., as transporters of small molecules through a central lumen), SBBs are appreciably smaller-sized, consisting of five or six short β-strands. As detailed below, these strands are often arranged as two closely-packed, nearly orthogonal β-sheets (Figure 1). An overview of all SBBs, as well as a more specific list of those considered here, is provided in Tables S1 and S2, respectively; these particular SBBs, which we take as representative examples of what is currently known, form an admittedly limited subset of all SBBs. Viewed across the range of all known SBBs, this domain is astoundingly flexible in terms of its ligand- and substrate-binding capacities: SBB-containing proteins often act as modules that bind, either specifically or generically (depending on the particular SBB), to various RNAs, DNAs, and other proteins (e.g., rightmost column in Table S2). Despite its small size and relatively limited surface area, it appears that virtually every solvent-exposed region of an SBB can adapt to bind other biomolecules, depending on the functional context. In this regard, SBBs are reminiscent of the RNA recognition motif (RRM; Pfam clan CL0221)—a prevalent protein domain that acts as a multipurpose functional module, binding both nucleic acids and proteins via many distinct ligand-binding structural motifs and biochemically active surface patches (Cléry et al., 2008).

Unlike RRMs, many SBB-containing proteins exhibit a strong tendency to assemble into toroidal discs and other higher-order

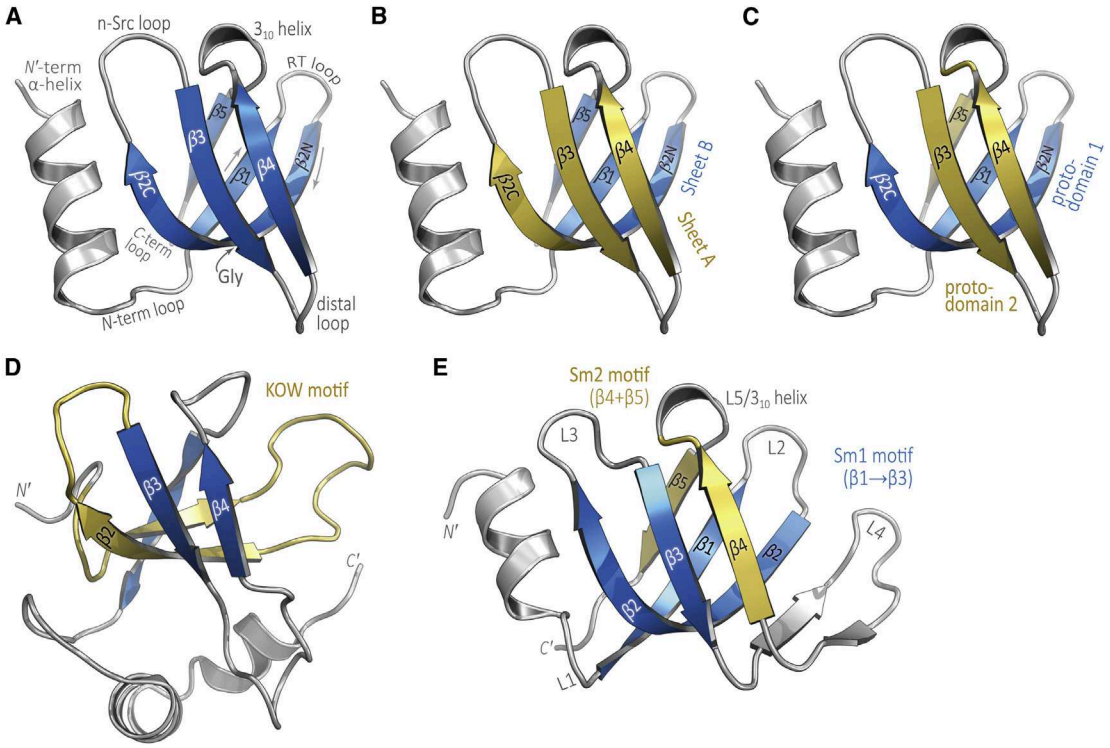

Figure 1. Structural Overview and Terminological Features of the SBB Domain

Hfq, the bacterial Sm-like protein (SCOP family b.38.1.2), is taken as a reference structure in this work. (A) The structure of *Staphylococcus aureus* Hfq, drawn from the hexamer structure (PDB: 1KQ2), with the SH3 strand numbering and loop terminology that we use for all SH3-like SBBs. The termini and other structural landmarks are labeled, including a conserved Gly near the point of greatest curvature in β2. In (B), the barrel is colored as two distinct β-sheets: *Sheet A* (*Meander*; yellow), consisting of β2C, β3, and β4, and *Sheet B* (*N-C*; blue), consisting of β5, β1, and β2N. In (C), the barrel is divided into two "proto-domains" that are related via approximate *C2* symmetry. Proto-domain 1 (blue) consists of strands β1, β2N, and β2C; proto-domain 2 (yellow) consists of strands β3, β4 and β5. The KOW motif, shown in (D), consists of 27 residues from β1, β2, and the N′-term loop preceding β1. In the Sm fold, shown in (E) for the SmD3 protein, the β-sheet is often viewed as consisting of two functional motifs: strands β1 → β3 comprise an "Sm1" signature (wherein lie many residues involved in RNA-binding), and segment β4 → β5 comprises an "Sm2" motif (which facilitates oligomerization, via these two edge strands). The loops in (E) are labeled with the Sm nomenclature (Table 1). Unless noted otherwise, all molecular illustrations were created in PyMOL and layout/post-processing was performed in Adobe Illustrator (AI).

structures that interact with various proteins or nucleic acids to serve as the biological functional unit. A notable example whereby SBB-mediated oligomerization yields an expanded range of biological functionality is the Sm/LSm class of RNA-associated proteins. The SBB domains of Sm and Sm-like (LSm) proteins, found in *Archaea*, *Bacteria*, and *Eukarya*, form a rich variety of homo- and hetero-oligomers (Mura et al., 2013). As detailed below, the tendency of some SBB-containing proteins to oligomerize enables this small module to evolve new biological functionality and elaborate versatility. Indeed, unlike RRMs (and many other protein superfamilies), SBB-containing proteins act in an astoundingly broad range of cellular pathways.

Variability Is the Theme: There Are No Golden Rules

The deep plasticity of the molecular functions of SBBs seems to stem from a unique combination of several structural and physicochemical properties of its β-rich architecture. Perhaps the most striking theme with the SBB domain is its great variability. In terms of sequence, structure, function, and evolution, it seems that "*the only golden rule is that there are no golden rules*" (GB Shaw). As outlined with some examples below, the SBB domain features extensive variability in terms of (1) *ligand-binding properties and cellular pathways* (variable molecular functions), (2) *3D structures* (variations of the fold) and *domain organization* (modularity), and (3) *oligomerization behavior and quaternary structures* (variable assembly states and supramolecular architectures).

Variability in Cellular Pathways and Molecular Activities: an Evolutionary Perspective. The evolutionarily ancient SBB domain occurs in proteins from viruses, *Bacteria*, *Archaea*, and *Eukarya*, and it appears to act as a key component in diverse biological pathways in these lineages. SBB-based cell biology ranges from generic DNA-binding by an SBB known as the oligonucleotide/oligosaccharide-binding (OB) fold (Mitton-Fry et al., 2004), to specific mRNA splicing pathways (Mura et al., 2013), small non-coding RNA (sRNA)-based regulatory circuits (Vogel and Luisi, 2011), and RNA biogenesis and decay/degradation pathways (RNAi, sRNA; Ma et al., 2004; Wilusz and Wilusz, 2013). SBB-containing proteins also function in structural scaffolding and organization of chromatin DNA in archaea (Robinson et al., 1998), as well as the maintenance of genomic integrity (Flynn and Zou, 2010); many ribosomal SBB proteins are key components of the translational machinery (Klein et al., 2004; Lomakin and Steitz, 2013; Valle et al., 2002). SBBs also act in signal transduction pathways (McCarty, 1998; Patel and Wang, 2013) and even seemingly unrelated processes such as the host immune response (Cridland et al., 2012; Jin et al., 2012; Shaw

and Liu, 2014) and mechanosensitive membrane channels (MscS; Perozo and Rees, 2003). In addition to their involvement in RNA processing, translation, and other basic (core) cellular pathways of presumably ancient/primitive origin, SBB proteins also play fundamental roles in pathways that likely evolved more recently (signaling and regulatory circuits, epigenetic modifications, etc.). As concrete examples of these, note (1) the recognition of histone tails by SBBs underlies chromatin remodeling and the regulation of gene expression (Patel and Wang, 2013); (2) recognition of a poly-proline signature motif makes the SH3 barrel a uniquely versatile adaptor/scaffold domain in regulatory cascades (McCarty, 1998); (3) eukaryotic Sm proteins form the common cores of small nuclear ribonucleoprotein complexes (snRNPs), and as such are key components of the spliceosome; and (4) the bacterial Sm homolog, known as "Hfq," is broadly involved in posttranscriptional regulatory pathways that hinge upon interactions between an sRNA and target mRNA. This broad functional repertoire distinguishes SBBs from RRM domains, which also use a variety of molecular interaction strategies, but whose primary roles are associated with posttranscriptional steps in gene expression. Considering all of the above, it is unsurprising that eliminating or compromising the functionality of many SBB-based proteins is associated with various cancers, inflammatory diseases, and other human ailments (e.g., the Sm class of SBBs was first discovered as the autoantigens in the autoimmune disease systemic lupus erythematosus; Tsokos, 2006).

Given that SBB-containing proteins span a vast range of biochemical and cellular functionalities, how much of this versatility is directly attributable to the SBB domain itself? In analyzing *structure* ↔ *function* ↔ *evolution* relationships for an SBB embedded in a larger protein, one can try to delineate the specific molecular functionality of the SBB domain itself, as distinct from (yet contributing to) the net function(s) of the whole protein. The functionality of the SBB domain can be distilled into three overarching categories, involving (1) stabilizing assemblies, (2) chaperoning interactions, and (3) relaying signals. To elaborate, SBBs can (1) stabilize macromolecular assemblies either by (a) serving as structural platforms/cores (examples are the roles of Sm and LSm oligomers in nucleating snRNP assembly, as well as verotoxin, HIN, TEBP and RPA proteins) or by (b) providing small stabilizing regions, as with individual SBB-containing ribosomal proteins, enmeshed in a network of rRNA structures. In terms of (2), chaperoning RNA···RNA or RNA···protein interactions, examples again come from the eukaryotic Sm/LSm proteins and, notably, the bacterial RNA chaperone, Hfq (Mura et al., 2013). Another example of facilitating RNA interactions is the role of the SBB-containing Argonaute in binding to small, non-coding "guide" RNAs in RNA-silencing pathways (Gorski et al., 2017). Finally, SBB proteins also (3) relay signals in biological pathways, either by (a) being part of an adaptor or scaffold protein itself, thus helping localize proteins to their target complexes (Good et al., 2011), such as for poly-Pro–binding by SH3 domains and in the recognition of modified histone tails by the Chromo/Tudor domain, or by (b) providing allosteric regulation, as in the case of ribosomal protein S12 and Spt5 (Gregory et al., 2009; Li et al., 2014; Panecka et al., 2014). Finally, there also exist two reports of an enzyme's catalytic residues lying within an SBB domain—namely, *Escherichia coli* signal peptidase and the self-cleaving transcriptional repressor LexA (Luo et al., 2001; Paetzel et al., 2002).

Intriguingly, some types of biological pathways appear to be enriched in SBB-containing proteins. In such cases, the SBB typically functions in one of a variety of specific modes, using different binding surfaces of the barrel and recognizing various binding partners (some bind to nucleic acids, others to proteins). A well-studied example is afforded by eukaryotic pre-mRNA processing pathways, where there appears to have been an evolutionary "fixation" of the SBB fold in at least five distinct (functionally unrelated) steps along the intricate *snRNP assembly → spliceosome biogenesis → intron excision* pathway; for the sake of space, this example is described in detail in the section on *Evolutionary "Fixation" of the SBB in Spliceosomal Pathways*, in the Supplemental Information.

Variability in Domain Organization. In terms of structural variability at the gross level of domain arrangement and general architecture, note that some SBBs function as single-domain, autonomous proteins, while in other cases the SBB module is part of a multi-domain protein. The former case is illustrated by several ribosomal components, such as the SH3-containing proteins L14, L21e, and L24 (Klein et al., 2004), the OB-containing S12 and S17 proteins (Brodersen et al., 2002), as well as the Melanoma Inhibitory Activity (MIA) protein (Stoll et al., 2001). Most well-characterized Sm/LSm/Hfq proteins have been found to act as single-domain SBB proteins (that, granted, oligomerize into functional rings). However, bioinformatic analyses of the domain architecture of Sm homologs suggest that C-terminal extensions—and even entire functional domains (e.g., putative methyltransferases)—can be appended to the SBB of some LSm homologs (Albrecht and Lengauer, 2004); a similar principle holds with Hfq (K.A.S. and C.M., unpublished data). Also, many SBB domains are fused or embedded within larger proteins. As two examples, note that (1) many kinases contain an SH3 domain (Morton and Campbell, 1994), and other proteins involved in signal-recognition contain SH3-like Tudor and Chromo domains (Blus et al., 2011), while (2) the RNA-binding ribosomal protein L2 consists of two fused SBB domains, with the N-terminal region adopting an OB fold and the C-terminal half being an SH3-like barrel (Diedrich et al., 2000).

No Rules, but Are There at Least Themes?—SBBs as a Unifying Structural Theme in Many Cellular Pathways, and the Concept of an "Urfold"

> Once the whole is divided, the parts need names.
> There are already enough names.
> One must know when to stop.
>
> —Tao Te Ching, Ch 32, Lao Tsu

Small barrels are quite robustly "foldable," and an unusually wide range of amino acid sequences can adopt SBB-like structures, such as the SH3 or OB folds— i.e., *a vast sequence space is compatible with the 3D architecture of this barrel*. (We avoid calling the SBB a "fold" for reasons described below.) A corollary of this principle is that a given β-barrel sequence can mutate significantly (e.g., evolutionary drift) without sacrificing the structural integrity of an SBB domain; this property, in turn, makes the β-barrel a rather extensible

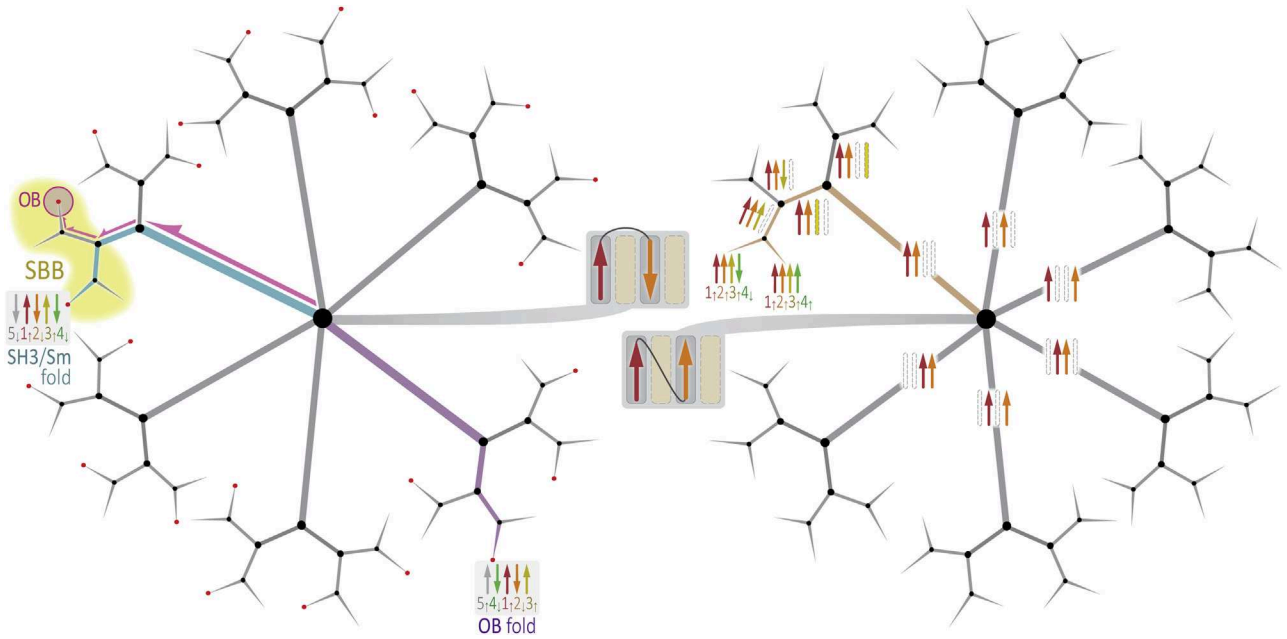

Figure 2. Exhaustive Enumeration of the 96 Unique Topologies of a Four-Stranded β-Sheet, via a Decision Tree-like Approach, Elucidates the Relationship between the SH3/Sm and OB Domains

Strand β1 is red, β2 is orange, β3 is yellow, and β4 is green; for the SH3 and OB folds, the fifth strand is also shown (gray). Branches along a sample path in this digraph are highlighted in tan (subtree at right), yielding the $1_{\uparrow}2_{\uparrow}3_{\uparrow}4_{\uparrow}$ and $1_{\uparrow}2_{\uparrow}3_{\uparrow}4_{\downarrow}$ topologies (in the nomenclature of Zhang and Kim, 2000); these two sheets, near the center of the image, are drawn simply to illustrate tree traversal. The base of the overall tree (at center) is a decision between the two possible configurations (parallel, antiparallel) for the simplest possible sheet—i.e., a tandem pair of strands (⇨∿∿⇨). Traversing the tree from this split "root" to the leaves corresponds to building-up the sheet, and the tree's branching structure elucidates the $n! \cdot 2^{n-2}$ unique topologies that are possible for a sheet of n strands; the successive branches of this unrooted k-ary tree are of degrees 2, 6, 2, 2, 2. The positions of the SH3 and OB folds are indicated by cyan and purple paths (subtree at left). Other features of β-sheets are also elucidated by this hierarchical representation, such as the fact that there are 24 unique arrangements of two sequentially adjacent β-hairpin motifs (red circles, left subtree). If the origin for strand numbering is taken as arbitrary (e.g., labeling a sequence 2→3→4 does not differ from 1→2→3), then the OB topology (pink path, left) can be seen to cluster closely with the SH3; the yellow region delimits a putative SBB "*urfold*" basin in fold-space, subsuming the SH3 and OB folds.

platform for evolving, and potentially engineering, new functionalities. This evolutionary benefit comes at a practical price (for researchers): sequence similarity levels are often so low, even among SBBs of nearly identical backbone 3D structures, that identifying small barrel structures based solely on sequence is frequently impossible (Dickey et al., 2013; Theobald and Wuttke, 2005). In general, new functional or biochemical roles can be associated with, or correlated to, specific structural features for a set of (related) proteins, but the SBB is more puzzling: with at least one type of SBB (the OB fold), function is more closely correlated with sequence phylogeny than with the structural classification (Theobald and Wuttke, 2005). Intriguingly, this phenomenon of a vast sequence and function space has been identified for some superfolds involving β-sheets, including the eight-stranded triosephosphate isomerase (TIM) barrels (Nagano et al., 2002) and the β-sandwich framework of immunoglobulins (Bork et al., 1994). Finally, note that a severe complication that arises in analyzing these and other superfolds is that a mix of both homologous (divergence from a common ancestor) and analogous (convergence to similar structures; evolutionarily unrelated) relationships may exist between various members that are grouped together at a given level (fold, superfold, etc.) under various hierarchical classification schemes, and such relationships are not readily differentiated (Alva et al., 2010; Cheng et al., 2008).

At what classification level does a protein become an "SBB"? In terms of structural hierarchies, the SBB lies above the level of *superfold*, transcending as it does the levels of homologous superfamily (in terms of similar molecular functions) and fold; in fact, it even surpasses superfold because it is not a single, coherent fold with a conserved topology (rather, it subsumes at least two superfolds: the SH3/Sm and OB). While differing in strand topology, and thus being classified as two distinct folds (Figure 2), the SH3 and OB nevertheless exhibit striking structural similarity. To address this somewhat odd duality, we propose the term "*urfold*." This term, which can be considered as describing structural similarity just above the superfold level, is meant to label the relationship between two SBBs (or, really, any distinct superfolds) that belong to what are technically different folds, yet which share (1) clear structural similarity, geometrically (β-rich architecture; Figure 3) and physicochemically (e.g., hydrophobic core in Figures 4 and S2), and (2) at least some overlap in terms of molecular functionality (e.g., nucleic acid–binding) or biophysical properties (e.g., oligomerization tendency). Note that an urfold corresponds most closely to level "A" in CATH (Class, Architecture, Topology/fold, Homologous superfamily), and its utility is that it unites neighbors in fold-space, such as the SH3/Sm and OB superfolds, into a single category. Using a term like urfold to capture the similarity of the SH3/Sm and OB architectures acknowledges the possibility

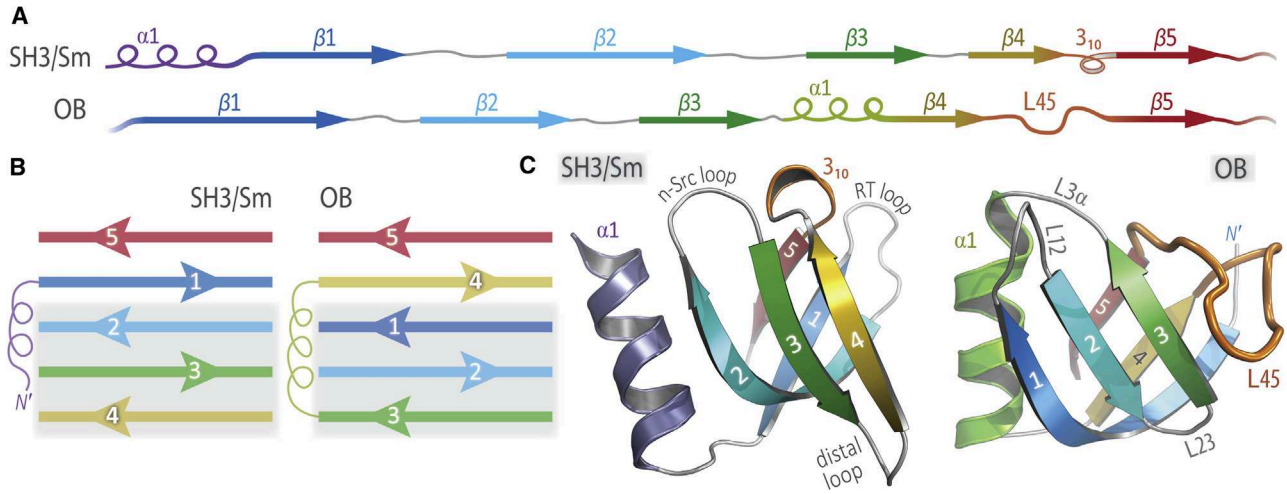

Figure 3. Comparison of the SH3 and OB Folds, with a Mapping of Secondary Structural Elements (SSEs)
These (A) one-, (B) two-, and (C) three-dimensional representations of the SH3/Sm and OB folds illustrate the similarities in their SSE content (A) and their matching architectures (C), despite differing topologies (B). Strands and other SSEs are labeled sequentially in these diagrams, with colors graded from blue (N-term) → red (C-term) and labels/numberings being matched for consistency between the panels. Representative SH3 and OB structures in (C) are Hfq (PDB: 1KQ2) and a verotoxin subunit (PDB: 1C4Q), respectively. The striking similarity between the 3D structures (C) can be elucidated in terms of sheet topologies that are equivalent under permutation (B); a light gray background in (B) accentuates the three-stranded β-meander that is architecturally and topologically identical between these folds (even neglecting permutation).

that these similarities are not purely coincidental but rather may stem from a deep evolutionarily divergence (such as in the phenomenon of homologous fold change; Grishin, 2001). Indeed, SH3-like proteins have been identified as a fold that features a "*mixture of homologous and analogous connections*" between the superfamilies comprising this fold class (Alva et al., 2010). Open questions concern how early in evolution the SBB arose, and if convergent evolution (i.e., analogy) accounts for some instances of its occurrence in extant proteins. Did the SBB arise multiple, independent times? (Is this particular β-rich architecture an "attractor" in fold-space?) Finally, we note that we chose the term urfold because the prefix "ur-" denotes "*primitive*" or "*one step higher in scope*." At present, the urfold is rather subjectively defined; development of the concept is under way (our unpublished data).

SBBs from different functional classes are found in various Structural Classification of Proteins (SCOP) superfamilies, and common themes across these superfamilies—and, indeed, even potential relationships within a superfamily—have gone largely unreported. For instance, the Sm-like superfamily (b.38.1), which is closely associated with pre-mRNA splicing and processing, also contains (1) a domain from a membrane protein channel (family b.38.1.3), (2) the hypothetical lipoprotein YgdR (b.38.1.6), and (3) the bacterial RNA chaperone, Hfq (b.38.1.2, which has been recognized as the bacterial branch of the Sm lineage). Although in the same SCOP superfamily, it is difficult to conceive, for instance, of these three sets of proteins as being homologous. A profile-based phylogenetics approach has implicated divergent evolution as being at least partly responsible for the interrelationships among various small β-barrels that are quite distant in sequence (Theobald and Wuttke, 2005). Related to this question, note that a failure to recognize potentially distant evolutionary relationships (deep homology), versus possible convergence toward the same fold (analogy), also could occur because information about SBB-containing proteins has been fragmented and disjointed in the literature. Such has been the case at least partly because of (1) the remarkable functional diversity of these proteins; (2) the small size and vast sequence space of this 3D fold, which hinders the detection of related (homologous) structures via sequence similarity beyond the twilight zone (Rost, 1999); and (3) SBBs have been assigned to 12 different SCOP folds (half of these folds populate the same "X" level in the Evolutionary Classification of Protein Domains (ECOD) database). Thus, although SBB-containing proteins have been studied for many decades, someone who is studying, say, the Sm/LSm proteins in RNA splicing may be entirely unaware that these proteins adopt the same fold as the SH3 domain that binds poly-Pro-containing sequences in signal transduction cascades, the Chromo domain involved in chromatin remodeling, the ribosomal and other proteins involved in translation, or even the membrane channels involved in a bacterial cell's response to a mechanical stress such as osmotic shock (e.g., MscS protein). Hence, classifying a newly identified small barrel structure can be a baffling and ill-posed problem, from a structural bioinformatics perspective; in fact, sometimes a new structure is reported as being similar to a specific, but not necessarily optimal, SCOP fold.

The present work largely consists of a survey-based analysis of the structural features of SBB domains, with the intent of addressing the severe "knowledge gap" alluded to above. As part of this analysis, we take a first step toward developing a systematic and coherent model of *structure* ↔ *function* relationships among SBB-containing protein superfamilies. We systematize the different nomenclatures that have arisen due to lack of cross-talk between various fields of study. We examine the anatomy of this structural fold, we introduce the term "*urfold*" to capture structure/function similarity just above the level of superfold, and we address how functional diversity can be achieved by

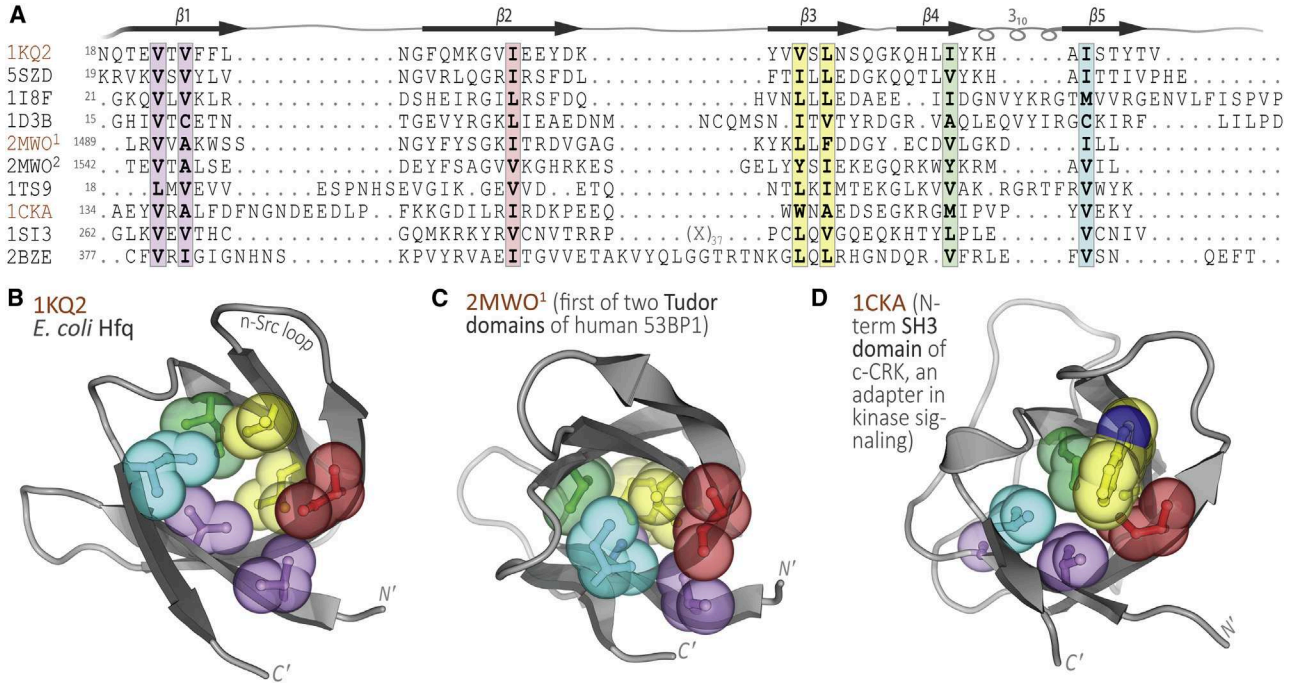

Figure 4. Sequence Analysis and Conserved Sites that Form the Core of the SBB Domain

(A) A multiple sequence alignment of β-barrel sequences reveals several highly conserved residues, highlighted here in color (Figure 5 shows structures of some of these proteins). At the sequence positions identified here, each β-strand contributes one or two conserved apolar residues to form the structural core of the barrel. For clarity, an unconserved 37-residue segment of 1SI3 is denoted "X_{37}," and superscripts distinguish the two Tudor domains that comprise the tandem repeats in 2MWO. The seven residue positions colored in this MSA are shown as space-filling spheres in the ribbon diagrams of (B) *E. coli* Hfq (1KQ2), (C) a Tudor domain (2MWO), and (D) an SH3 domain, as represented by 1CKA. In these panels, residues are color-matched to (A), two residues have the same color if on the same β-strand, and the β-barrel is drawn as a gray ribbon; a similar representation of the OB core is shown in Figure S2. The most conserved residues can be seen to form the barrel's core, chiefly via hydrophobic packing interactions (London dispersion and other van der Waals forces). This MSA was computed using the structure-guided PROMALS3D approach (Pei et al., 2008). After visualization with ESPript (Robert and Gouet, 2014), the resultant PostScript source was edited to obtain this figure; the final layout was performed in AI.

such a seemingly simple structural unit (including via a multitude of oligomeric states and polymerization into fibrils). More broadly, this work helps systematically "define" the unique SBB urfold and highlights the structural properties that enable its vast functional diversity; a detailed functional analysis of the SBB will be presented elsewhere (S.V., unpublished data).

Results and Discussion

Scope of This Work

Many protein folds are associated with the term "β-barrel." For instance, 53 folds in SCOPe 2.06 (Chandonia et al., 2017) are defined as a "barrel" or "pseudo-barrel," and 79 X-groups appear under the "β-barrel" architecture in the ECOD classification scheme (Cheng et al., 2014). The many SCOP (v.2.06) folds that belong to the SBB urfold include the following (Table S1): b.34 (SH3-like; 21 superfamilies), b.35 (GroES-like, 2 superfamilies), b.36 (PDZ domain-like, 1 superfamily), b.38 (Sm-like; 5 superfamilies), b.39 (ribosomal protein L14; 1 superfamily), b.40 (OB; 16 superfamilies), b.41 (PRC-barrel, 1 superfamily), b.55 (PH domain-like, 2 superfamilies), b.87 (LexA/signal peptidase, 1 superfamily), b.136 (stringent starvation protein B; 1 superfamily), b.137 (RNase P subunit p29; 1 superfamily), and b.138 (hydrophobin, 1 superfamily). Our selection of SCOP folds for analysis (such as those enumerated above) was done manually, by focusing on barrels of small size (as annotated in SCOP), with a strand number $n = 4$ or 5 (and occasionally $n = 6$). Many such barrels exhibit internal pseudo-symmetry (in the sense of Figure 1C; discussed below), which can be detected by structurally aligning the two halves of a barrel. Additionally, structural alignments between SBBs from different SCOP folds, using a tool such as CN3D (Hogue, 1997), as well as specific topological features (listed in "Topological Descriptions," below), were the main criteria for deciding whether a given SCOP fold is an SBB. The details of the folds which we term "SBBs" are listed in Table S1. Intriguingly, ECOD (Cheng et al., 2014)—which uses a different approach to hierarchical structural classification, versus SCOP or CATH (Dawson et al., 2017)—unites six of the above 12 SCOP folds (b.34, b.38, b.41, b.87, b.136, and b.137) into a single X-group termed "SH3."

In terms of sequence diversity, structural similarity, and functional breadth, the SBB can be termed an urfold (as defined in the Introduction); we refer to the SBB as such in this work. In many ways, the SBB is comparable to another β-rich small superfold, namely ferredoxin (d.58 in SCOP, encompassing 59 superfamilies). How both the SBB urfold and the ferredoxin superfolds achieve such immense functional diversity (e.g., whether there are parallels between them) is an intriguing question for future work. Because of the breadth of the SBB urfold (it contains 53 SCOP superfamilies), note that numerous domain names appear in our analysis ("Chromo," "Tudor," "Plus3,"

Table 1. A Map of Small β-Barrel Terminology: the Names and Labels of Key Structural Elements (Strands, Loops), and Their Correspondence between the SH3/Sm and OB Folds

Strands		Loops: SH3-Based ⇝ Sm, OB					Loops: OB-Based ⇝ SH3		
SH3/Sm ↔ OB Correspondence		Bracketing SH3 Strands	SH3 (b.34) Loop Name		Sm (b.38) Loop Name	OB (b.40) Loop Name	Bracketing OB Strands	OB Loop Name	SH3 Loop Name
β1 ↔ β4		α-helix → β1 or β0 → β1	N-term loop	↔	L1	↔ L01	β1 → β2	L12	n-Src
β2 ↔ β1	the	β1 → β2	RT	↔	L2	↔ —	β2 → β3	L23	Distal
β3 ↔ β2	Meander	β2 → β3	n-Src	↔	L3	↔ L12	β3 → α1	L3α	—
β4 ↔ β3	sub-sheet	β3 → β4	Distal	↔	L4	↔ L23	α1 → β4	Lα4	—
β5 ↔ β5		β4 → β5	3_{10} helix	↔	L5	↔ —	β4 → β5	L45	—

The SH3 topology, using the numbering of its own fold, runs (α1–β1)–(β2-β3-β4)–β5, where β2-β3-β4 is the central meander. The OB topology, in its nomenclature, runs (β1-β2-β3)–(α-β4)–β5, where β1-β2-β3 is the meander (see also Figures 1, 2; Figure 3 shows the correspondences between these elements in one-, two-, and three-dimensions). Note that only a C′-term portion of the longer strands (β1 in OB, β2 in SH3) forms part of the meander, with the N′-term region of this strand "peeling away" from the meander and hydrogen bonding with another strand (β4 in OB, β1 in SH3).

"HIN," etc.); many of these names may be unfamiliar to readers not directly involved in the corresponding (and often highly specific) biochemical field. To simplify matters, the remainder of this work focuses mainly on three SCOP folds—b.34 (SH3), b.38 (Sm), b.40 (OB)—which represent the vast majority of known SBB structures and functions (Table S1), and which offer insight into the structural and functional plasticity of small barrels. In this work, the domain names are typically the same as its SCOP superfamily name (e.g., "Chromo"); we give the SCOP classification upon first mentioning each domain. Table S2 supplies detailed information for each structural domain used in our present analysis.

General Anatomy of Small β-Barrels

Geometric and Protein Structural Characteristics of the Small β-barrels. In general, a β-barrel can be viewed as a β-sheet that twists and coils to form a closed structure wherein the first and last strands are hydrogen-bonded (Murzin et al., 1994a; 1994b). We find that all SBB-related folds are annotated as "barrels" in SCOP, although in the literature they may be variously referred to as "barrels" or "orthogonally packed β-sheets." Indeed, these small and flexible structures straddle the boundaries between these two architectures, as they have key features of both (1) antiparallel β-sheets, wherein the first and last strands hydrogen-bond (yielding a barrel), and (2) two distinct β-sheet faces that pack orthogonally (or nearly orthogonally) atop one another. An ideally cylindrical barrel would have a circular cross-section, transverse to the strand directions, whereas such a cut would be elliptical for an orthogonally-packed β-sheet; most protein structures are intermediate between these extremes.

Because they have not been formally (or precisely) defined in the literature, here we define small β-barrels as modular domains, generally ≈60 to 120 residues long, with a specific, superimposable framework of β-strands. In many cases, the spatial pattern of strands exhibits approximate two-fold rotational (pseudo-)symmetry (Figure 1C), and the side-chains that emanate from this backbone scaffold give a structural "core" of ≈35 residues. Classically, a generic barrel is defined by the number of strands, *n*, and the shear number, *S* (McLachlan, 1979; Murzin et al., 1994a). The magnitude of *S* describes the extent of stagger of the β-sheet or, equivalently, the tilt of the barrel with respect to its principal geometric axis; in turn, the magnitude of the stagger defines the degree of twist and coil of the strands, and is correlated with the internal diameter of the barrel (Murzin et al., 1994a; 1994b). Considered across all known proteins, an increase in barrel tilt (i.e., *S*) is proposed to have occurred over the course of evolution (Caetano-Anolles and Caetano-Anolles, 2003). Alternatively, many barrels also can be viewed as two β-sheets packed face-to-face, with the strands in each sheet lying roughly perpendicular to one another (Chothia and Janin, 1982). Such barrels have higher stagger values and are generally "flatter" (a more elliptical cross-section), allowing the two apposed faces to pack closely together.

SBBs are of this more orthogonal barrel type, generally with few strands (low *n*) and high shear ($S \approx 2n$). In SCOPe version 2.06, the folds b.34 (SH3) and b.38 (Sm) are defined as $n = 4$, $S = 8$, with an SH3 topology, while b.40 (OB) is defined as $n = 5$, $S = 10$ (or $S = 8$). In most cases, the fourth β-strand, as defined in SCOP for b.34 and b.38, is interrupted by a short 3_{10} helix, yielding two strands (e.g., the β4 and β5 strands of Hfq; Figure 1A). Adhering to terminological precedents found in much of the literature, here we define SBBs as containing five β-strands, arranged as two orthogonally-packed sheets. There are two distinct strand connectivities (and, thus, folds) of the small barrels treated here, namely the SH3-like and OB-like sheets. As shown in Figures 2 and 3, these two topologies are related via a (non-circular) permutation, resulting in the same 3D framework of strands. Because of their matching architectures, SH3-like and OB-like structures belong to the same "A" level in CATH. For all these reasons, we call the SBB an *urfold* (described above). To lessen confusion, in this work we default to SH3 terminology in describing the SBB urfold; the alternative topology of the OB superfold is mapped to the SH3 in Table 1 and Figure 3.

Topological Descriptions. A few structural features uniquely characterize SBBs of the SH3/Sm variety. Its architecture consists of five β-strands arranged in an antiparallel manner, as shown in Figures 1, 2, and 3. A stringently conserved glycine in the middle of strand β2 enables severe curvature of the backbone; this Gly can be followed by a β-bulge, which divides the long strand into N-terminal (β2N) and C-terminal (β2C) segments (Figures 1A–1C). As such, two orthogonal β-sheets (*Sheet A*,

Sheet B) can be defined as comprising the β-barrel; this view effectively makes the SBB a six-stranded barrel, with each β-sheet consisting of three β-strands (as is clearest in Figure 1B). A short, single-turn 3_{10} helix links strands β4 and β5; the β4 and β5 strands straddle the barrel and belong to different sheets (as do β2N and β2C). The architectural layout of these flanking strands enables barrels to associate via β4⋯β5′ interactions between monomeric subunits (a prime denotes adjacent subunits)—that, in turn, is a critical feature in forming toroidal discs and other oligomers (detailed below).

β-Sheets of the Barrel. The strands in an SBB are generally rather short, being ≈4 to 6 residues long, and are connected by loops. In some SBBs, these loops adopt known β-turn geometries (e.g., the β-turns described for *Aquifex aeolicus* Hfq; Stanek et al., 2017), while in other instances the loops are less regularly structured. *Sheet A*, also termed the *Meander* (Figure 1B), is a three-stranded β-sheet consisting of strands β2C, β3, and β4 (contiguous in sequence and in space). *Sheet B*, which consists of strands β5, β1, and β2N, and is noncontiguous (Figure 1B), closes the barrel by pairing the C-term (β5) and N-term (β1) strands of the protein in an antiparallel fashion; for that reason, this sheet has also been called the *N-C Sheet* (Figure 1B). Many other structural elements and motifs have been described over the years for different classes of SBB domains, as summarized by the various examples described below (the KOW motif, PAZ domain, etc.); these examples are provided mainly to illustrate the broad, multifaceted range of structural features that have been found in naturally occurring SBB proteins.

Proto-Domains. Proto-domains (Figure 1C) are related by pseudo-symmetry within a single domain. They were noticed relatively early in the history of protein structure; for example, the six-stranded β-barrels of serine proteases were seen to exhibit approximate *C*2 symmetry (McLachlan, 1979). Some protein domains, such as those of serine or aspartyl proteases, are thought to have arisen from ancient duplications; in such cases, the sequence signal may be lost, while structural similarity persists and is more readily apparent (e.g., as manifested by internal symmetry). To our knowledge, proto-domains have not been described in small barrels, such as analyzed here. The SBB can be viewed as two proto-domains related by *C*2 symmetry (Figure 1C). In the case of SBBs, proto-domain-1 is β1, β2N and β2C, while proto-domain-2 consists of β3, β4 and β5. Even if the existence of such proto-domains is a structural byproduct of the geometric constraints of forming a closed barrel via β-strands (with all the possible arrangements shown in Figure 2), this two-fold symmetry of the barrel nevertheless does appear to be a recurring feature of SBB domains.

The KOW Motif. The KOW motif (Figure 1D), named for Kyrpides, Ouzounis, and Woese (Kyrpides et al., 1996), is found in some RNA-binding proteins (mostly small barrels in ribosomal proteins). This motif consists of β1, β2, and the loops preceding β1 and following β2 in the SBB; together, this spans a total of ≈27 residues. A hallmark of this motif is alternating hydrophilic and hydrophobic residues, with an invariant Gly at position 11.

Functional Modules. Functional modules (Figure 1E) are exemplified by the Sm-like proteins (b.38). Here, molecular "function" is meant in a rather broad, all-inclusive sense, taking into account biochemical activities (e.g., RNA-binding), physicochemical functionalities (e.g., mediating interactions between subunits), and structural roles (e.g., scaffolding the further assembly of macromolecular complexes such as snRNPs). The "Sm1 motif" consists of β1→β3 and the "Sm2 motif" consists of β4→β5; in bacterial Sm proteins (Hfq), these two motifs are linked by a short, ≈4-residue 3_{10} helix (Schumacher et al., 2002). The Sm2 substructure, with its β4–β5 strands straddling the barrel, is a significantly conserved feature, and possibly a signature of all small barrels with the SH3-like topology. In fact, superimposing the β4–β5 pattern alone can yield high-quality structural alignments for the entire conserved structural framework (i.e., fold) of various SBB domains.

The SBB Urfold: Topological Relationship between Its OB and SH3 Categories

As a member of the SBB urfold, the OB superfold has the same 3D architecture as the SH3-like superfold, despite its different topology (Figure 2). Like the SH3 barrel, the OB domain folds as a barrel comprised of a five-stranded antiparallel β-sheet; the SH3 and OB topologies are related by a (non-circular) permutation, as noted previously (Agrawal and Kishan, 2001; Theobald and Wuttke, 2005) and illustrated in Figures 2, 3, 5A, and 5D. Our reference Sm-like fold, the Hfq protein, is well-suited to comparison with the OB fold: both topologies (Figure 2) are evidently compatible with the same 3D structural framework of strands (Figures 3B and 3C), and both feature an α-helix not found in the SH3-like fold. To mitigate terminological confusion between the OB and SH3/Sm subsets of the SBB urfold, we use the SH3↔OB strand mapping when discussing OB folds (Table 1, "Strands" column).

Permutating the strands from the Sm-like fold to the OB fold (Figures 2 and 3) places the N-terminal α-helix and β1 of Sm after its *Meander* $[\beta2C\text{-}\beta3\text{-}\beta4]_{Sm}$ and before β5. Thus, the initial SH3/Sm-like topology, [α1-β1]–[β2N-β2C-β3-β4]–[β5], maps to the final OB topology [β2N-β2C-β3-β4]–[α1-β1]–[β5], to use Sm-like numbering throughout (see Figure 3B). Renumbering the permutated strands, using OB nomenclature, then gives [β1N-β1C-β2-β3]–[α1-β4]–[β5]. The non-circular permutation preserves *Sheet A* (the *Meander*) in both topologies: [β2C-β3-β4] in SH3/Sm-like, and [β1C-β2-β3] in the OB fold. These relationships can be seen by comparing the pair of topology diagrams in Figure 3B; similarly, in three dimensions one can compare the SH3 and OB halves of Figure 3C. Structural alignment of [β2N-β2C-β3-β4-β5] in the SH3 fold and [β1N-β1C-β2-β3]+β5 in OB yields a root-mean-square deviation of 1.4 Å, using Hfq (PDB: 1KQ1) and verotoxin (PDB: 1C4Q) as SH3 and OB-fold representatives, respectively (neglecting strand β1 of the SH3 fold and the analogous strand, β4, from the OB).

Hydrophobic Core of the SBB: a Conserved Structural Framework

The hydrophobic core of the SBB is minimalistic, consisting only of the five (or six, depending on how the long β2 strand is treated) elementary strands that form the conserved structural framework: β1, β2N + β2C, β3, β4, and β5 (Figures 1 and 4). These strands are short, being composed of roughly four to six alternating (exposed/buried) residues, unless bulges are present. Only two strands, β1 and β3, are completely saturated in terms of their backbone hydrogen-bonding capacity. The three-dimensional layout of strands is the key property of SBB proteins and is

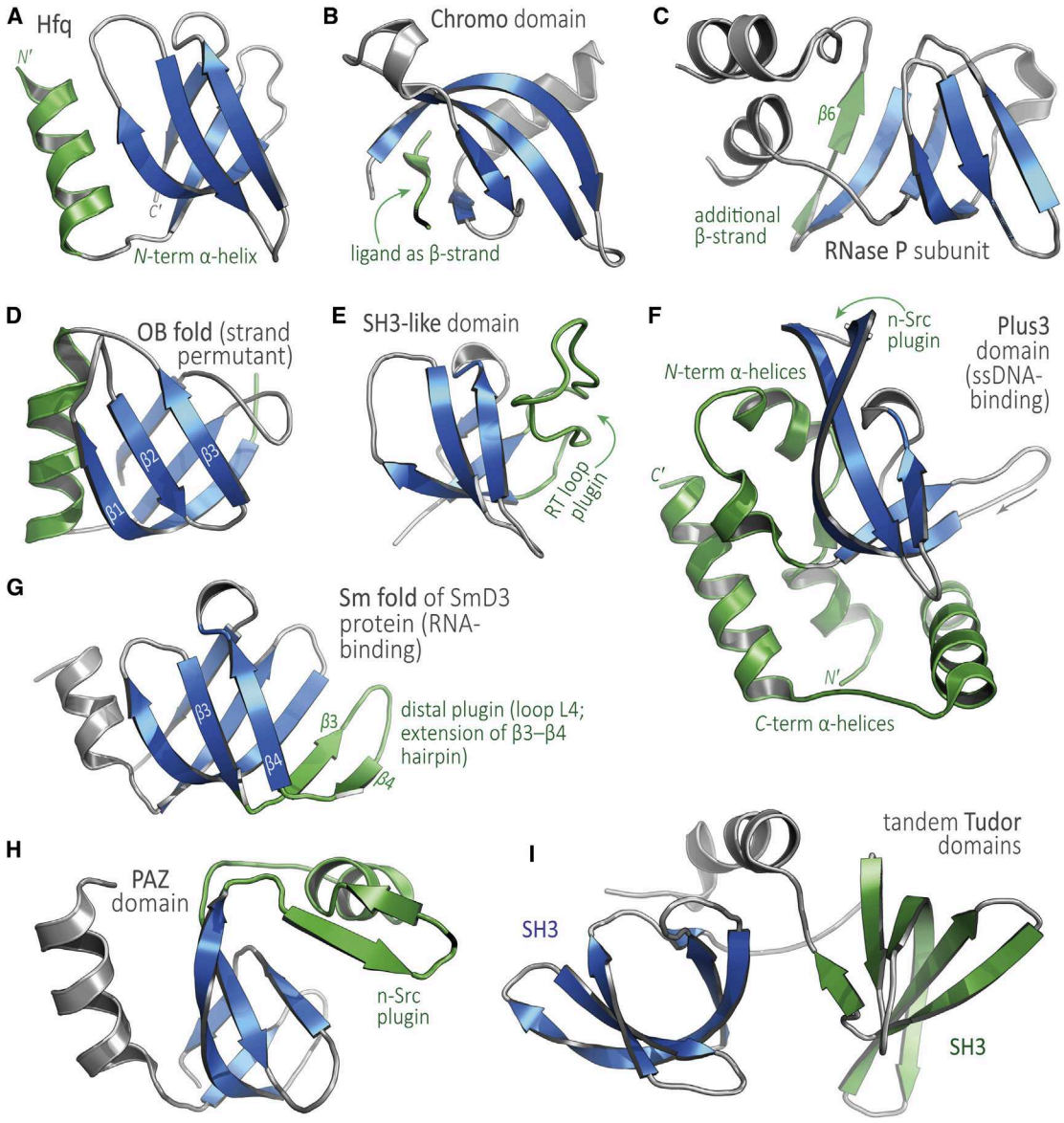

Figure 5. The SBB as a Scaffold: Insertions, Decorations, and Other Variations

In each panel the SBB core is blue and variations are green.

(A) The bacterial Sm protein Hfq (b.38.1.2) is taken as our reference structure (PDB: 1KQ2), as it features the "cleanest" and structurally simplest barrel (e.g., minimal loops).

(B) In this Chromo domain (b.34.13.1, PDB: 1KNA), the SBB's β1 strand is contributed by the cognate binding partner.

(C) The RNaseP subunit P29 (b.137.1, PDB: 1TS9) features an additional strand, β6, flanking the core SBB.

(D) The nucleic acid-binding OB fold (b.40, PDB: 1C4Q) is an SBB with a strand topology differing from the canonical SH3 sheet (see text, Figures 2 and 3).

(E) An RT loop "plugin" is found in the SH3 domain of the proto-oncogene c-Crk (b.34.2, PDB: 1CKA).

(F) In the DNA-binding Plus3 domain (b.34.21, PDB: 2BZE), additional helices extend the N- and C-termini, and an n-Src plugin also occurs.

(G) Many Sm folds feature a Distal loop plugin, as illustrated here for the canonical, RNA-binding SmD3 protein (b.38.1.1, PDB: 1D3B).

(H) The PAZ domain, shown here from a human Argonaute protein (b.34.14, PDB: 1SI3), has a plugin into the n-Src loop.

(I) The tandem-Tudor SH3 (b.34.9.1, PDB: 2MWO) has tandemly repeated SBBs joined by a flexible linker; note that the relative spatial orientation of the two barrels in this single-chain structure can vary, in contrast to the rather precise geometric positioning of adjacent SBB subunits in Sm rings (Hfq hexamers, Sm heptamers, etc. in Table 2).

well-captured by the Hfq barrel, where all loops are reduced to tight β-turns (Figure 1A). This structural framework tolerates diverse residue substitution as long as a compact, well-packed hydrophobic core is preserved—as evidenced by interdigitated barrels, barrels inserted within each other, and barrels with significant deviations from the "usual" SBB fold (Figures 5 and 7 and several sections of text below). Indeed, an antiparallel configuration of short β-strands may enjoy substantial "wiggle room" for compensatory changes, while preserving overall structural integrity and thermodynamic stability: the register of some strands may shift, strand geometries may undergo minor rearrangements or adjustments, and so on. Note that this model

is in keeping with a view of fold-space as being more continuous rather than discrete (Harrison et al., 2002).

In a typical SBB, each strand contributes one or two buried residues to the hydrophobic core (Figure 4). The two middle strands—β1 at the center of *Sheet B*, β3 at the center of *Sheet A* (Figure 1B)—contribute two apolar residues each (magenta and yellow in Figure 4), while the four lateral strands (β2N, β2C, β4, β5) typically contribute one residue each to the hydrophobic core. Not all β2N strands (in all SBB structures) contribute consistently to the hydrophobic interior, so the minimal core can be taken as consisting of the seven residue positions shown in Figure 4A. The conserved hydrophobic residue in β2C, which is typically a Val, Leu, or Ile, follows Gly (the pivot point of the highly curved β2 strand), and this residue is positioned at the beginning of the characteristic β-bulge. The apolar residues in β4 and β5 abut the 3_{10} helix, either just before (β4) or after (β5) the helix. This composite hydrophobic core defines a stable, minimal SBB fold, leaving all solvent-exposed residues to interact with ligands or other biomolecules (and that, in turn, permits the diversification of function).

The seven-residue hydrophobic core can be extended in many ways. Because the barrel is semi-open, various decorations can contribute apolar residues to the minimalistic core. For example, the N-terminal helix in the Sm-like barrel extends the β5-β2C region of the otherwise open barrel, as shown for Hfq in Figure 1A and SmD3 in Figure 1E. Similarly, the RT loop in the SH3-like barrel extends the β2N-β3 side of the barrel (Figure 5E). The hydrophobic core of the OB fold builds upon the same seven residue positions as the SH3-like fold (Figures 4B, 4C, 4D, and S2), but is typically somewhat larger by virtue of (1) strand elongation (especially those bounding L12/n-Src loop; see Figure 3, Table 1 and the next section for loop nomenclature), and (2) formation of a possible hairpin within L45 (Figure 3C), which would then extend *Sheet A* by two strands. Beyond these differences, most of the notable loop variations in the OB fold are similar to those of the SH3-like fold, as detailed elsewhere in this work (e.g., the aforementioned Tower domain in BRCA). Finally, note that a simple visual comparison of the hydrophobic cores of the SH3/Sm and a sample OB fold reveals the striking similarity of these conserved 3D architectures (Figure S2).

The solvent-facing residues on the "edge" stands—β2C and β4 in *Sheet A* (Meander), β2N and β5 in *Sheet B* (N-C)—can potentially form hydrogen bonds with other β-strands, unless they are sterically occluded by terminal decorations or long loops. Such strand···strand interactions potentially have two effects. Firstly, they enable extension of the β-sheet of the barrel in the direction of the *Sheet B* face (Figure 1B), along loop L4 (in Sm terminology; the "Distal loop," in SH3 terminology); indeed, loop L4 is known to be highly variable in length and in sequence in SBBs of eukaryotic Sm and LSm homologs (see, e.g., the L4/Distal loop of the SmD3 protein in Figure 5G). Secondly, such strand···strand interactions enable the formation of quaternary structures via hydrogen bonds and other favorable atomic contacts, stitching together β4···β5′ of adjacent subunits into oligomers (see the sections on *Possible Interfaces in Oligomeric Assemblies* and *Higher-Order Assembly of SBBs into Multimeric Rings*, and Figure 8, covering oligomerization).

Hfq as a Reference Structure for all SBBs

Hfq, a key RNA-associated bacterial protein, adopts an Sm-like fold and arguably represents the simplest example of an SBB (Figures 1A and 5A). If one superimposes all small barrels and identifies structurally conserved regions (Figure 4), then Hfq appears to be the most regular structural representative ("regular" in the sense of clean and minimalistic—short loops, strands of roughly similar length, and minimal "decorations" beyond the SBB's obligatory five-stranded core). Thus, for simplicity and clarity of presentation, we take Hfq as the archetypal representative of the SH3-like superfold, and as a useful reference point for all SBBs. Note that SH3 and Hfq are not intimately linked in SCOP: "SH3-like barrels" are defined at the fold level (b.34), whereas Hfq is classified as a family (b.38.1.2) within the Sm-like superfamily (b.38.1), under the Sm-like fold (b.38). In contrast, ECOD groups together Hfq and other Sm-like proteins at the family level, within the F (family) ⊂ T (topology) ⊂ H (homolog) ⊂ X (possible homology) levels of this hierarchical classification system. That is, Hfq falls within ECOD's SH3 category (Cheng et al., 2014). In this way, ECOD is most consistent with our model of a single, unified urfold to encompass all SBBs. In short, Hfq provides a useful structural framework for all SH3-like folds, beyond the broad Sm superfamily of which it is a key member.

Beyond the SBB Core: Loops, Decorations, Additional Modules

A Brief Overview and Note on Nomenclature. Beyond the core geometric framework of the small barrel—by which we mean the all-β structural motif found in all the various SCOP, ECOD, and CATH folds that are SBB-like—great variability is exhibited by all other structural elements too, such as the loops, modules inserted within loops, and N′- and C′-terminal extensions. A gallery of these variations is shown in Figure 5. That the SBB urfold can tolerate such variation is critical to its biological roles: these decorations and additional structural elements largely dictate the specific cellular functions of different SBB-containing proteins, irrespective of whether the structural similarity between these proteins stems from divergent evolution (homology) or, alternatively, convergent evolution (analogy).

Before analyzing the loops of SBBs, note that several different systems of nomenclature have arisen in research communities working on different (in terms of cellular functions) subsets of the universe of all small β-barrels. While the terminology for the strands is consistent (β1, β2, etc.), disparate naming schemes have emerged for the loops, thus muddying efforts at comparative analyses of different functional classes and structure ↔ function relationships. The three most prominent nomenclatures are given in Table 1. First are the SH3-like barrels that act in signal transduction, via binding to poly-Pro motifs (b.34.2), as well as chromatin remodeling via recognition of specific modifications on histone tails by Chromo-like (b.34.13) and Tudor-like (b.34.9) domains. Second are the Sm-like barrels (b.38.1), which are involved quite broadly in RNA processing and regulatory networks. Third are the OB-fold barrels (b.40), primarily involved in maintenance of genome integrity via binding to nucleic acids.

The loops in SH3 and Sm-like proteins are in an ordered, one-to-one correspondence because of the topological identity between these two folds, but the loop ordering is permutated in

the topologically distinct OB fold. Two of the OB loops have a clear counterpart in the SH3 fold: (1) L12 of OB structurally maps to the n-Src loop of SH3, and (2) L23 maps to the Distal loop of the SH3 fold (Figure 3, Table 1). The RT loop and 3_{10} helix are absent from the OB fold but present in SH3-like topologies; conversely, L3α, Lα4 and L45 are unique to the OB fold. To be consistent in this work, as well as inclusive of prior work, we cross-reference these terminological systems in Table 1, and we 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, e.g., b.136, b.137, and b.39). Given the large body of literature on OB proteins, its nomenclature is preserved here (with mapping to the SH3-like fold, when appropriate).

Specific Loop Variations, Including Insertion of Secondary Structural Elements. Loops that connect the β-strands in SBBs vary significantly in length and confer a plethora of functional roles (see below). SBBs consist of five, or sometimes six, loops. The first loop precedes the first strand, β1. The next four loops are always present: RT, n-Src, Distal, and a 3_{10} helix, as defined for the SH3-like superfamily (b.34.2) (Figure 1A). Of these four internal loops, significant variations in the lengths of three—RT, n-Src, and Distal—have been observed, and can be linked to specific biochemical functions. The fourth loop is almost always a short (single-turn) 3_{10} helix and is, infrequently, either (1) a distorted helix, as in RPP29 (Sidote et al., 2004) or (2) replaced by a longer loop, as in the TrmB proteins (Krug et al., 2006). Elongated loops often house additional secondary structures, as described below. A sixth loop is also possible (the C-terminal loop), linking the last β-strand of the SBB to a C-terminal extension. Such extensions have been found to be (1) irregularly structured (e.g., in many Hfq homologs), (2) simply a sixth strand of the barrel, as in the RNase P subunit shown in Figure 5C, or (3) an entire standalone module (e.g., a mixed α/β domain in *Pyrobaculum aerophilum* SmAP3; Mura et al., 2003b).

RT Loop, Linking β1→β2. Long insertions in the RT loop, which links strands β1 and β2 (Figure 5E), result in the classical SH3 domain (b.34.2) that is ubiquitous in signal transduction. The SH3 domain binds proline-rich peptides using the elongated RT loop (as well as the n-Src loop and 3_{10} helix). The RT loop lies alongside the barrel, capping one of its ends (Lim, 1996; Yu et al., 1992). Different pairs of loops can form various pockets. In the PAZ domain (b.34.14) of the Piwi and Argonaute proteins (involved in RNA interference), aromatic residues of the elongated RT loop (Figure 5H) are part of an aromatic pocket formed between the loop and the α/β module (inserted into the n-Src loop, see below); this pocket laterally secures the RNA substrate (Ma et al., 2004).

n-Src Loop, Linking β2→β3. An elongated n-Src loop is observed in two functional families (Figures 5F and 5H). The first case is that of the PAZ domains, described in the previous section. In the case of the Plus3 domain (b.34.21) of the transcriptional elongation factor Rtf1, the extended n-Src loop contains two short (≈3-residue) β-strands and is involved in binding single-stranded DNA (de Jong et al., 2008). In the OB2 of BRCA2 (b.40.4.3), a Tower domain is inserted in the L12 loop (corresponding to the SH3's n-Src loop). The Tower domain, which has been implicated in DNA-binding, is a 154-residue long insert that consists of two long α-helices and a three-helix bundle positioned between them (Bochkarev and Bochkareva, 2004; Yang et al., 2002). In the C-terminal DNA-binding domain (DBD-C) of RPA70 (b.40.4.3), a zinc-finger motif, consisting of three short β-strands, is inserted into the OB's L12 loop (Bochkareva et al., 2002).

Distal Loop, Linking β3→β4. Perhaps the clearest example of elongation of the Distal loop is with the eukaryotic Sm proteins (b.38.1), which are core components of the pre-mRNA splicing machinery (Will and Luhrmann, 2011). Elongation of this loop (Figure 5G) corresponds to extension of the β-hairpin formed by strands β3 and β4. In extending the Distal loop, these two long β-strands become bent, similarly to β2, and can be seen as β3N and β3C, β4N and β4C (Kambach et al., 1999; Mura et al., 2001). Like β2, they simultaneously contribute to the formation of two sheets (Figure 1E). The extended β3–β4 hairpin (Figure 5G) results in a much larger hydrogen-bonded *Sheet B*, now containing the five strands β5, β1, β2N, β3C, and β4N; the original *Sheet A* remains the same (as in Figure 1B). Much of the sequence variation among Sm proteins occurs within the Distal loop or as extensions of it (see, e.g., the SmB protein and its alternatively spliced variants; Saltzman et al., 2011). The DNA-binding OB domain of cdc13 (b.40.4.3) has a unique pretzel-shaped loop L23, corresponding to the SH3 Distal loop, that significantly extends the potential interactions of this barrel with DNA. This 30-residue–long loop twists and packs across the side of the barrel, and interacts with the L45 loop (Mitton-Fry et al., 2004).

The 3_{10} Helix, Linking β4→β5. The 3_{10} helix that links strands β4 and β5 is generally short (≈4 residues) and relatively invariant in structure. The geometry of this linker dictates the relative positions of strands β4 and β5, which, in turn, straddle the SBB's core (the "edge strands" of the SH3/Sm fold). Indeed, the dynamical flexibility and plasticity of this linker element is limited by the geometric constraint that it connect strands β4∿β5. The 3_{10} helix is present in virtually all SH3-like folds, but is absent in the OB fold for topological reasons (see below and Figure 3C; β3 does not connect to β5). This linker helix does not adhere to a strict pattern of conservation in Sm proteins—it is absent in many eukaryotic Sm structures and Sm-like archaeal protein (SmAP) homologs, but present in most bacterial (Hfq) structures. Intriguingly, sac7d, sso7d, and other histone-like small archaeal proteins feature a second 3_{10} helix in the middle of the highly bent β2 strand (Robinson et al., 1998), in place of the stereochemically forgiving glycine that is phylogenetically conserved at that position in the Sm fold (Mura et al., 2013) and in other SBB proteins.

N- and C-terminal Decorations, Capping of the Barrel, Small Internal Modules. The N- and C-terminal regions of many SBBs contain helices and additional loops/extensions, and these are sometimes termed "decorations." Their position relative to the barrel core varies and, in many cases, these structural features affect the ability of a barrel to oligomerize. As with loop insertions, barrel decorations frequently play a functionally significant role. The following select examples illustrate how SBB decorations can serve as functional adaptations.

N-terminal α-Helix. The Sm-like fold (b.38) generally features an N-terminal helix (Figure 5A) that links to the main body of the barrel via a short loop (L1 in Figure 1E). This region can engage in multiple interactions with both RNAs and proteins. The helix stacks against the open barrel and, in the context of

an intact, hexameric Hfq toroidal disc (Sauer, 2013), it lies atop the "*proximal*" face (this face of the disc, defined below, is known as the "Loop L3" face for SmAPs; Mura et al., 2013). In Hfq (b.38.1.2), the SBB's N-term α-helix mediates contacts with cognate sRNA molecules via a conserved patch of basic residues, composed of Arg16, Arg17, Arg19, and a Gln8 (Schumacher et al., 2002; Stanek et al., 2017; Sun and Wartell, 2006; residue numbers are for *E. coli* Hfq). A similar mode of RNA-binding appears to be conserved in the SmAPs (Mura et al., 2003a; Thore et al., 2003), as noted in Stanek et al. (2017). In LSm proteins, the SBB's N-term α-helix interacts with proteins Pat1C in the LSm1→7 ring (Wu et al., 2014) and with prp24 in the LSm2→8 ring (Karaduman et al., 2008). With snRNP Sm proteins (b.38.1.1), the same α-helix interacts with the β-sheet of adjacent protomers during ring assembly. For the SBB of eukaryotic paralog SmD2, a long N-terminal region harbors an additional helix (h0) that interacts with U1 snRNA as it traverses into the lumen of the hetero-heptameric Sm ring (Li et al., 2016; Pomeranz Krummel et al., 2009).

C-terminal α-Helices. C-terminal α helices (Figures 5C and 5F) can either augment existing binding interactions or mediate contacts with additional binding partners. For the LSm1→7 ring (b.38.1.1), a long helix formed by the C-term tail of the LSm1 subunit lies across the central pore on one face of the ring, preventing the 3′-end of RNA from exiting via that distal surface (Weichenrieder, 2014). Intriguingly, the novel structure of an Sm-like pentamer of putative cyanophage origin (Das et al., 2009) revealed that this homolog lacks an N-terminal helix, and instead features a well-defined C-terminal helix; any functional consequences of this structural alteration are unknown.

Both N- and C-terminal α-Helices. Both N- and C-terminal α-helices (Figures 5C and 5F), when present together, can interact to form a supporting structure/subdomain around the barrel, as in the case of the Plus3 domain (b.34.21) of Rtf1 (de Jong et al., 2008); in that system, three N-terminal α-helices and a C-terminal α-helix form a four-helix cluster that packs against one side of the barrel (Figure 5F). The role of these helices is unclear, but the conservation of many residues in that region implies some presumptive functional significance.

C-term Tails. C-term tails have, among all conceivable decorations, the least stereochemical and overall structural constraints. These regions can be disordered and can vary significantly in length—for instance, >40 residues in SmD1 and SmD3, and >150 C-term residues in SmB/B′ (Kambach et al., 1999). With spliceosomal Sm proteins (b.38.1.1), the C-term tails of SmB/B′, SmD1, and SmD3 harbor RG-rich repeats that are critical for proper assembly of these SBBs into a toroidal disc; assembly occurs via an intricately chaperoned, Arg-methylation-dependent biogenesis pathway (Friesen et al., 2001; Grimm et al., 2013; Selenko et al., 2001). In Hfq (b.38.1.2), the disordered C-term tails are proposed to extend outward from the ring and mediate contacts with RNAs (Beich-Frandsen et al., 2011), perhaps as an instance of a "fly-casting" mechanism between a disordered region and its cognate ligand (Levy et al., 2007; Shoemaker et al., 2000). Recently, the acidic C-term tails of *E. coli* Hfq were demonstrated to interact with residues of the SBB core domain. This phenomenon enables autoregulation of sRNA···mRNA annealing by assisting the release of sRNA•mRNA pairs, thus increasing specificity of sRNA binding and preventing double-stranded DNA (dsDNA) aggregation on the rings (Santiago-Frangos et al., 2017); the latter property is important, as at least some fraction of Hfq (which occurs at high intracellular concentrations) is thought to co-localize with the bacterial nucleoid. Finally, we note that in some (under-investigated) LSm homologs, lengthy regions—of up to hundreds of residues—extend the C-termini well beyond the SBB core. At least five novel groups of homologs (LSm12→16 proteins) were bioinformatically detected in eukaryotes (Albrecht and Lengauer, 2004); these extended SBBs likely act in RNA metabolic pathways (mRNA degradation, tRNA splicing, etc.), and the C-term regions in some of them have been identified as encoding putative methyltransferase activities. In these extended LSm homologs, the SBB can be viewed as a functional module that imparts a particular activity (e.g., nucleic acid binding) on the protein of which it is a part.

Small Internal Modules. Small internal modules (Figures 5E, 5F, and 5H) are short secondary or super-secondary structures (often α/β or purely α) inserted in an SBB's loops. These structural elements typically form a pocket against the barrel, and are an integral part of barrel function. Examples include an α-β-β module inserted into the n-Src loop of the PAZ domain (b.34.14; Figure 5H; Ma et al., 2004) and a β-hairpin extension module in the n-Src loop of the Plus3 domain (b.34.21; Figure 5F) of Rtf1 (de Jong et al., 2008). An "insertion" also can be an entire, domain-sized module (see below).

Further Structural Variations. The following vignettes describe further variations that have been discovered in the 3D structures of SBB-containing proteins, determined chiefly via X-ray crystallography or nuclear magnetic resonance spectroscopy.

Additional β-Strands. The structure of the RNase P subunit Rpp29 (b.137.1) illustrates that the barrel core can be extended by a sixth strand (Figure 5C). Here, an additional β-turn in the C-term region (between strands β5 and β6) is followed by the sixth β-strand, thus extending *Sheet B* (N-C) to four antiparallel strands (β6, β5, β1, β2N; Numata et al., 2004; Sidote et al., 2004).

Missing β-Strands. In at least one case, namely the Chromo domain HP1 (b.34.13.2), the intact SBB forms only by binding of the cognate peptide ligand (Figure 5B). HP1 exists as a three-stranded *Sheet A* (*Meander*), and the β-strand conformation of the peptide ligand templates the formation of the second β-sheet (i.e., the *N-C*, *Sheet B*), thereby completing the barrel (Jacobs and Khorasanizadeh, 2002).

Substituted Structural Elements. In the OB fold of the DBD-C of RPA70 (b.40.4.3), the α-helix lying between β3 and β4 (Figures 3 and 5D) is replaced by a helix-turn-helix, while in the DBD-D of RPA32 (b.40.4.3) the same α-helix is missing altogether and is replaced by a flexible Ω-loop (Bochkareva et al., 2002).

Sequence Variation, and Electrostatic Properties of SBB Surfaces

In addition to variations in structure, patterns of sequence variation can further distinguish various SBBs. Small barrels are extremely tolerant to mutations (see the section on *Folding and Stability* in the main text, as well as the Supplemental Information), and a common evolutionary strategy appears to be the modulation of electrostatic interactions by altering the residues exposed in loops, sheets, and decorations. More specifically, shifts in various physicochemical properties of substituted

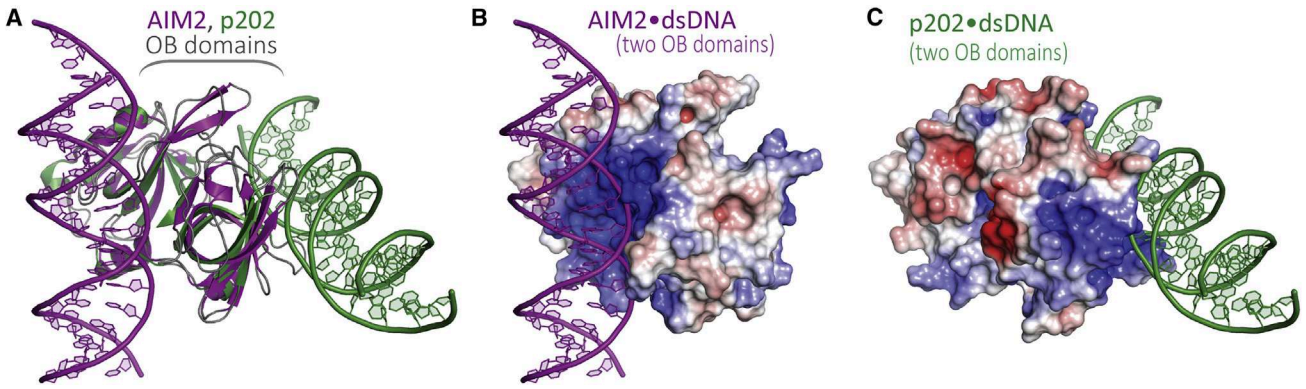

Figure 6. Functionally Relevant Electrostatic Differences between Two Otherwise Similar SBBs
The HIN domain consists of two tandem OB barrels.
(A) Superimposed structures of the HIN domains of AIM2 (purple, PDB: 3RN5) and p202 (green, PDB: 4LNQ), bound to double-stranded DNA via two distinct and separate interfaces (green duplex interacts with p202, purple duplex interacts with AIM2).
(B) is from the same perspective as (A), but shows only AIM2. The HIN domain of AIM2 is now shown as a surface representation and color-ramped by the intensity of the electrostatic potential near the molecular surface. DNA binds to the first OB barrel of the HIN domain; the interacting surface is concave and positively charged (blue).
(C) Same view as (A), but showing only p202. The HIN domain of p202 is shown as a surface representation, colored by electrostatic potential. Here, DNA binds to a positively charged patch (blue) on the *second* OB barrel of p202. All electrostatics calculations were performed with APBS (Baker et al., 2001).

residues can occur, ranging across the extremes of *acidic* ↔ *basic* (positively/negatively charged), *polar* ↔ *apolar*, *bulky* ↔ *compact*, etc. Such sequence variation adjusts the properties of protein surface patches, or even entire sheets, thus yielding drastically different ligand-binding profiles, interactions with DNA and RNA, and downstream physiological activities.

An interesting case is that of the HIN domains of AIM2 and p202, which bind double-stranded DNA and are involved in the innate immune response (Jin et al., 2012; Shaw and Liu, 2014; Yin et al., 2013). Each HIN subunit consists of tandem OB-fold barrels (Figure 6), which are known to bind single-stranded, double-stranded, and quadruplex DNA with various affinities. Despite 36% sequence identity between the HIN domains in AIM2 and p202, the duplex DNA-binding modes are entirely dissimilar (Figure 6A), largely because of differing patterns of electrostatic potential at the molecular surface. In the case of AIM2, the binding is mediated by positive charges on the convex surface of the SBB (Figure 6B). With p202, the analogous surface is negatively charged and therefore does not interact with DNA; instead, DNA is bound by the positively charged loops (of the second OB barrel), on the opposite side of the barrel (Figure 6C). The corresponding loops in AIM2 bear hydrophobic residues, and thus do not bind DNA (Jin et al., 2012; Yin et al., 2013). The different binding surfaces correspond to different DNA-binding affinities, enabling these two proteins to act in a physiologically antagonistic manner. Additional cases of electrostatic modulation via sequence variation can be found in the section on Sequence Variation, and Electrostatic Properties of SBB Surfaces in the Supplemental Information.

Joining Barrels, Covalently (In Tandem) and Non-covalently (As Oligomers)

Small barrels tend to associate with one another on different structural scales. Interactions between tandem barrels within a single polypeptide chain are common, especially in proteins that bind RNA (Cléry et al., 2008; Lunde et al., 2007) or that scaffold the binding of other proteins or nucleic acids as part of a larger biomolecular complex (Good et al., 2011). Many proteins that consist of only an SBB domain are known to assemble into multimeric rings that act in many RNA-associated pathways, across all three domains of life. Although not ubiquitous among SBBs, the property of oligomerizing into rings is rather common for small barrels involved in RNA biogenesis (e.g., pre-mRNA splicing); notably, strand-mediated oligomerization into rings also occurs with other small, β-rich proteins that are distinct from the SBB, yet which also play key roles in RNA-associated pathways (e.g., the tryptophan-activated RNA-binding attenuation protein, TRAP; reviewed in Lunde et al., 2007). Finally, small barrels (and oligomers thereof) have been found to self-associate into closed higher-order assemblies (e.g., head→head stacks of rings, often with dihedral symmetry) or, in some instances, open-ended polymeric fibrils (e.g., head→tail stacking of rings of SBBs); these polymeric and tubular assemblies of SBBs often occur in crystal lattices. Large supramolecular assemblies of SBBs are described in more detail below.

Tandem, Embedded, and Enmeshed Barrels. This section treats several combinations of β-barrels that occur either as tandem repeats, or that are intertwined within a single polypeptide chain.

SH3∩SH3 Tandem-Tudor Domains. SH3-like barrels that are repeated in tandem often form barrel···barrel interfaces, and these can be constructed in various ways. Different linkers and sequences can lead to a number of tandem interfaces with varying extended β-sheets, allowing great plasticity. For example, in p53-binding protein 1 (53BP1, Figure 7A; Tudor [b.34.9.1]), hydrogen bonding between β2N of the first barrel and β5 of the next one joins individual three-stranded β-sheets into an extended six-stranded sheet. The C-term α-helix augments the connection by interacting with multiple β-strands of both barrels (Charier et al., 2004). The tandem-Tudor-like "Agenet" domains of FMRP, the fragile X mental retardation protein (b.34.9.1; Table S2), contact each other via each SBB domain's β2N region. In the transcription elongation factor

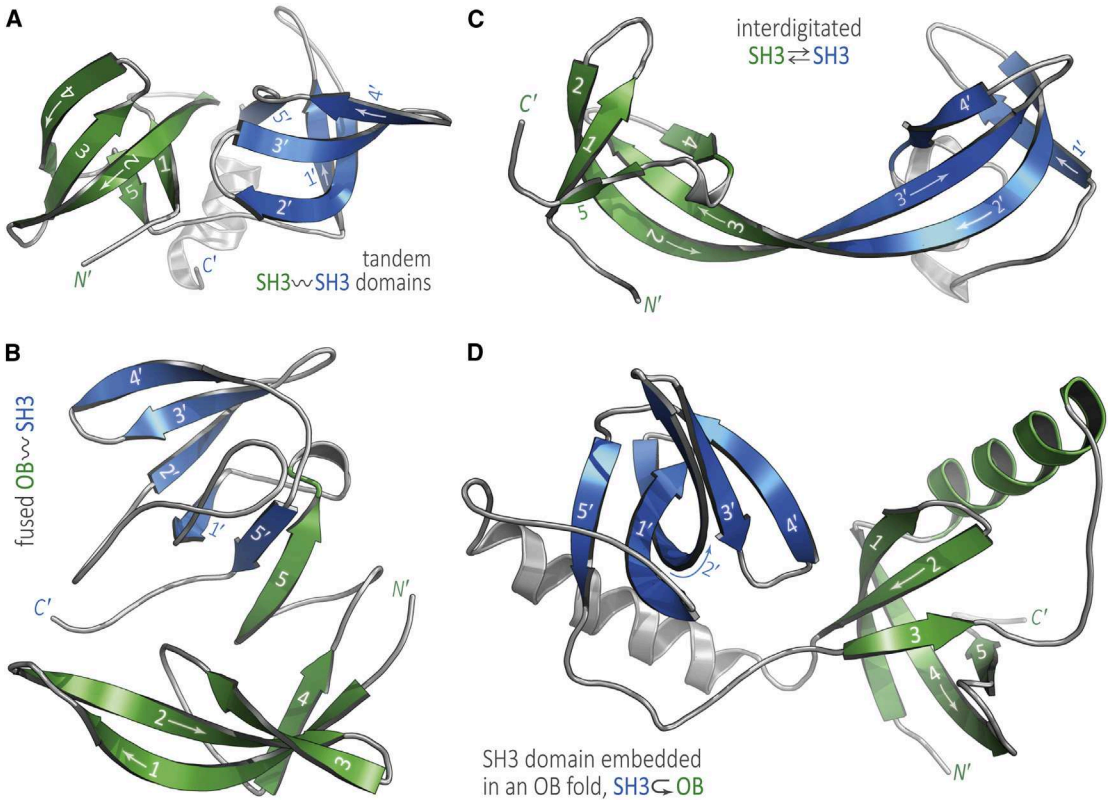

Figure 7. SBBs Can Combine in Various Ways, Including with Intricate Topologies
Here, the SBB closer to the N′-terminus—whether it is a contiguous (A and B) or noncontiguous (C and D) barrel—is colored in green and the second barrel is colored blue. The β-strands are labeled as in Figures 1 and 3. (A) An SH3∿SH3 tandem barrel occurs in 53BP1 (PDB: 1SSF), and (B) ribosomal protein L2 (PDB: 1S72) features an OB∿SH3 tandem barrel. (C) shows an SH3⇄SH3 interdigitated barrel in JMJD2A (PDB: 2QQR), and (D) illustrates an SH3 barrel embedded within the OB barrel of the TDRD (eTud) protein (PDB: 3OMC).

Spt5, which has five tandemly repeated KOW-containing Tudor domains (b.34.5.5; Figure 1D), interactions between Tudor-2 and Tudor-3, which move as a single body, occur through β5 of Tudor-2 and residues immediately following β5 in Tudor-3 (Meyer et al., 2015). In the DNA/RNA repair protein KIN17 (unclassified in SCOP) this interface is formed by N′-term and C′-term tails that interact with the linker between the two barrels (le Maire et al., 2006).

OB∿SH3 Hybrid and Fused OB∿OB Barrels. OB and SH3 domains can combine in tandem in either order, as demonstrated by the ribosomal protein L2 and eukaryotic translational elongation factor eIF5A (Dever et al., 2014; Nakagawa et al., 1999). In L2 (Figure 7B)—thought to be one of the oldest ribosomal proteins (Harish and Caetano-Anolles, 2012)—an N′-terminal OB domain (b.40.4.5) links to the SH3 domain (b.34.5.3) via a 3_{10} helix that runs parallel to the 3_{10} helix which bridges β4→β5 in the SH3-like domain. The β5 strands of the two SBBs (β5 and β5′ in Figure 7B) lie in an antiparallel orientation, extending the two β-sheets (Nakagawa et al., 1999). Conversely, in eIF5A an N′-term SH3 module (b.34.5.2) is followed by an OB domain (b.40.4.5; Dever et al., 2014). Ribosomal protein S1 offers an example of six tandem OB domains, two of which are involved in binding the 30S ribosomal subunit (Demo et al., 2017; Giraud et al., 2015); intriguingly, S1 may interact with Hfq, although this remains somewhat unclear (Hajnsdorf and Boni, 2012).

SH3⇄SH3 Interdigitated Tudors. The most intricate contacts between two adjacent barrels occur in the interdigitated Tudors, such as JMJD2A (b.34.9.1; Huang et al., 2006) and RBBP1 (not classified in SCOP; Gong et al., 2014). These structures have been described as two barrels "swapping" some strands, resulting in an "interdigitated barrel" (Figure 7C); note, however, that this phenomenon is structurally distinct from classic "domain swapping" (Liu and Eisenberg, 2002). In the architectural layout of JMJD2A and RBBP1, the long β2 and β3 strands contribute to the sheet in their parent barrel and then traverse to the adjacent barrel, yielding two compact structures (the two SBBs); the first two strands belong to one "linear" (contiguous in sequence) barrel and the other two strands belong to the other "linear" barrel. An antiparallel β-sheet forms along the length of β2-β3–β2′-β3′ (the numbering scheme/primes indicate the linear ordering in the sequence; see also the numbers and arrows in Figure 7C).

An SH3 Barrel Embedded in an OB Fold (Figure 7D). Staphylococcal nuclease domain-containing protein 1 (SND1) contains five tandem OB-fold domains (b.40.1.1), with an SH3-like Tudor domain (b.34.9.1) inserted into the L23 (Distal loop-equivalent) of the fifth OB barrel (Liu et al., 2010). Such an arrangement of OB and Tudor units is typically referred to as an extended Tudor

Table 2. Strand-Strand Interactions in Quaternary and Pseudo-quaternary Arrangements of Small β Barrels

Interacting Strands	Oligomeric States (n); [Known Symmetries]	Sample Protein	SCOP Family; Sample PDB IDs	Cellular Functions	Reference
β4···β5′	*n* = 5, 6, 7, 8 subunits, assemble as rings; [*C5*, *C6*, *C7*, *C8*]	Hfq (*Bac*), Sm/LSm (*Euk*), SmAP (*Arc*)	b.38.1 (Sm-like) PDB: 1D3B (Sm) PDB: 1KQ2 (Hfq) PDB: 4M75 (LSm)	mRNA splicing; RNA biogenesis and decay; sRNA-based regulatory pathways	Kambach et al., 1999; Mura et al., 2013
β1···β5′ (OB terminology; β2C···β5′ in SH3 labeling)	*n* = 5 ring; [*C5*]	verotoxin	b.40.2 (OB fold, bacterial enterotoxins); PDB: 1C4Q	facilitates toxin entry	Stein et al., 1992
β2N···β5′	dimer (covalently linked); [*C2*]	53BP1	b.34.9 (Royal family, Tudor) PDB: 2MWO	signal transducer in DNA repair	Charier et al., 2004
β2N···β2N′	dimer (covalently linked); [*C2*]	FMRP	b.34.9 (Royal family, Agenet) PDB: 4QW2	fragile X syndrome	Myrick et al., 2015
β2C···β2C′	dimer (tetramers); [*C2*]	Mpp8	b.34.9 (Royal family, Chromo) PDB: 3QO2	M phase phosphoprotein	Chang et al., 2011
β5···β5′	hybrid tandem (covalently linked) [no known symmetric assemblies]	RL2	b.40 + b.34 (OB + SH3) PDB: 1S72	translation	Nakagawa et al., 1999

Note that both the OB and SH3 strand nomenclatures are given for verotoxin (see the mapping in Table 1). In the case of RL2, β5···β5′ hydrogen bonding in an antiparallel orientation is possible because of conformational changes in the OB domain of this protein.

domain (“eTudor” or “eTud”). The eTudor domain consists of two β-strands from the OB, the linker (containing an α helix) and five β-strands of the Tudor domain. Both parts of this split OB domain are essential for binding the symmetrically dimethylated arginine (sDMA) residues often found in the C-terminal tails of these proteins (and other SBBs, such as the snRNP Sm proteins). In this system, the OB-fold (SN domain) and SH3-fold (Tudor domain) function as a single unit (Friberg et al., 2009; Liu et al., 2010). Notably, the *Drosophila* SND1 protein features 11 tandem extended Tudors, also known as maternal Tudors (Ren et al., 2014).

Possible Interfaces in Oligomeric Assemblies

The β-strands of an SBB, particularly those that flank the domain, are typically of roughly equal length. This simple property enables these “edge strands,” such as β4 and β5 of the SH3/Sm fold, to laterally associate via backbone hydrogen bonds and other enthalpically favorable interactions between adjacent barrels. This geometric capability, in turn, facilitates the assembly of SBB subunits into dimers, cyclic oligomers, and higher-order states, as described in the next few sections. Some of the possible strand···strand interactions between SBBs are described in Table 2 and visualized in Figure 8; a more detailed analysis of SBB interfaces can be found in the section on Possible Interfaces in Oligomeric Assemblies in the Supplemental Information.

Higher-Order Assembly of SBBs into Multimeric Rings

Many single-domain SBB proteins self-assemble into quaternary structures that are the biologically active units. A well-studied example of such oligomerization is the toroidal discs formed by Sm and Sm-like (LSm) proteins. The spatial positions of the β4–(3_{10})–β5 strands, which straddle the body of the barrel, enable interactions between the β4 strand of one monomer and β5′ of the adjacent subunit (denoted by a prime), ultimately leading to the association of between five and eight monomers into a doughnut-shaped disc (Figures 8A and 8B). This toroidal assembly can be viewed as linking a three-stranded *Sheet A* of one monomer with a three-stranded *Sheet B* (Figure 1B), giving a six-stranded composite sheet that connects the two faces of the disc (Figure 8B).

The β4–β5 substructure of the SH3/Sm-like fold is unique in its shape as well as its positioning with respect to the rest of the SBB structure. Intriguingly, this region has been identified among 40 peptides that likely originated in ancient proteins (Alva et al., 2015). Notably, we have found that structural alignment using only β4–β5 often suffices to bring the rest of an SBB structure into alignment. (This implies that whatever structural role is served by the β4/β5 pair [e.g., oligomerization, in Sm/Hfq proteins] may be a significant constraint on the evolutionary drift of those residue positions that dictate the relative geometric disposition of strands β4 and β5.) The two faces of the cyclic disc are formed by the two β-sheets of the individual barrels: *Sheet A* (*Meander*) forms the distal face, while *Sheet B* (*N-C*) forms the proximal face. The lateral periphery of the ring (Sauer, 2013), also termed the “outer rim” (Weichenrieder, 2014), houses solvent-exposed residues that form the region between the distal and proximal faces. This lateral rim appears to act as a site for auxiliary RNA interactions, at least with Hfq hexamers. In this context, transiently stable RNA^1·$(Hfq)_6$·RNA^2 complexes promote annealing of two RNA strands, yielding a host of downstream physiological effects that depend on the particular RNA pairing (RNA^1/RNA^2 are typically an sRNA/mRNA pair). The SBB residues that form the lateral site are functionally important: sRNA binding is anchored on the proximal face, while the mRNA target binds

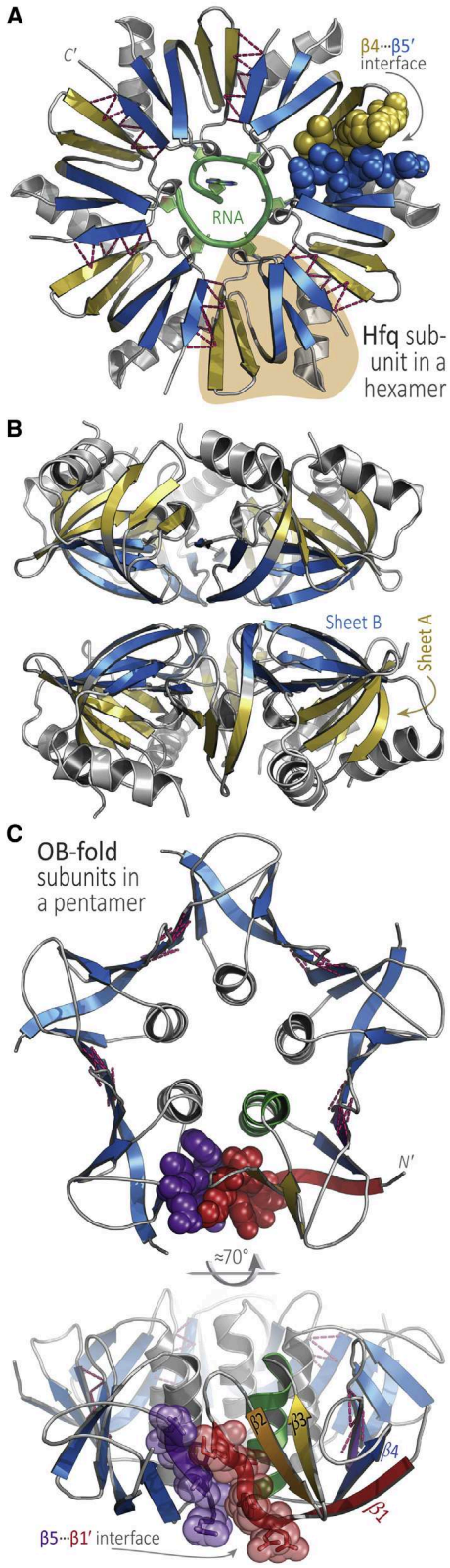

Figure 8. The SBB Is a Versatile Module for Oligomerization and Higher-Order Assembly

(A) In this ribbon diagram of an Hfq hexamer (PDB: 1KQ2), colored by the *Sheet A/B* scheme (Figure 1B), one subunit is highlighted in orange (near the 5 o'clock position) and a bound RNA is drawn in green; the C′-terminus is labeled for a subunit near 11 o'clock. The crucial β4···β5′ interface generates a toroidal disc by stitching together Hfq subunits in a head-to-tail manner; these two β-strands are rendered as space-filling spheres for one subunit, and dashed magenta lines denote hydrogen bonds and other interactions between other subunits. Together, the strands of the individual SBBs effectively create a contiguous, cyclic β-sheet that comprises the body of the ring; this cyclic sheet has 30 strands in the case of Hfq (6×5), and 35 in the case of the heptameric Sm/LSm proteins (7×5). The two faces of Hfq and other Sm rings are often found to mediate higher-order assemblies, such as in the $((Hfq)_6)_2$ assembly shown in (B). In (B), the same coloring scheme is used as in (A), revealing that the distal···distal interface of the dodecamer is built upon *Sheet B* of the SBB. To illustrate strand-mediated assembly in another SBB, (C) shows two views of a pentamer of verotoxin (PDB: 1C4Q), which adopts the OB fold. The β5···β1′ interface is shown as spheres and the strands of one subunit are labeled (lower-right). The β1 strand of the OB is structurally analogous to strand β2 of the Sm fold (see text and Figure 3). This example underscores the plasticity and versatility of SBBs as a structural scaffold for β-strand–mediated oligomerization.

mainly at the distal face. Productive RNA interactions require the two RNA strands to physically associate, and the lateral rim of Hfq's ring of SBB domains appears to facilitate that process (Santiago-Frangos and Woodson, 2018; Sauer, 2013; Stanek et al., 2017).

The SBB module also appears to enable oligomeric plasticity, including for various paralogs that form heteromeric assemblies or even for the very same protein (homomeric assemblies). As a striking example, the SBB of the *Archaeoglobus fulgidus* SmAP2 protein forms both hexamers and heptamers depending on solution-state conditions (a hexamer at low pH, without RNA, but a heptamer in the presence of U-rich RNA; Kilic et al., 2006). In terms of SBB structure, the mechanistic basis of such plasticity remains murky; recent molecular simulations suggest a role for both residue-level conformational dynamics and rigid-body structural rearrangements (McAnany and C.M, unpublished data).

Several more cases of ring formation by SBB proteins are known. Another example of oligomerization involving an Sm-like fold is the bacterial mechanosensitive channel MscS, which is a large, multi-domain protein that assembles as a homo-heptamer, encircling a membrane pore; its central domain is a small β-barrel, which forms a heptameric ring closely resembling those of the archaeal Sm homologs (Mura et al., 2003b; Steinbacher et al., 2007). In the case of the OB fold (b.40), the Shiga-like verotoxin protein forms a pentameric ring via β5···β1′ hydrogen bonding between monomeric subunits, as shown in Figure 8C (Stein et al., 1992). A final example of SBB-mediated oligomerization is provided by the cell-puncturing structure in bacteriophage T4, which contains a trimer of the gp5 protein; the N′-term domain is an SBB (specifically, an OB fold) and it forms a ring-shaped channel (Kanamaru et al., 2002).

Polymerization into Fibrils and Other Higher-Order Oligomeric States

Proteins that are β-rich are prone to aggregation, including fibril formation. The resulting polymeric species, which often resemble amyloidogenic ultrastructures (β-rich fibrils) and which can vary in their homogeneity (compositionally, and structurally), may be physiologically functional in some cases, or toxic and pathogenic in other instances. The structural unit from which a fibril forms can be an individual SBB, a toroidal disc, or a double-ring assembly (with either head→head or head→tail

stacking of rings). SBBs have been found to form fibrils via the exchange of β-strands between monomers (Camara-Artigas, 2016; Neudecker et al., 2012) as well as higher-order polymers via stacking of ring-shaped discs (Mura et al., 2003a; Schumacher et al., 2002; Stanek et al., 2017). Despite this, fibrillar aggregates of SBBs have not been extensively implicated in diseases, suggesting that SBBs might be stable or otherwise "inert" *in vivo* (robust folding pathways, well-regulated assembly states). Further details on the mechanisms of fibril formation can be found in "Polymerization into Fibrils and Other Higher-order Oligomeric States" in the Supplemental Information.

Folding and Stability

Folding of Small β Barrels. Folding of the SBB domain seems exceptionally robust—the same fold is achieved by a wide range of sequences (Figures 4 and S2), and also when the structural elements are mutated or permuted (Figures 2 and 3). The partially folded conformational states (i.e., transition state ensemble [TSE]) of SH3 and OB barrels exhibit bipartite behavior, with (1) a hydrophobic region that nucleates further folding, consisting of most of the β2–β3–β4 segment (i.e., *Sheet A*/*Meander*, β2–β3–β4), and (2) conversely, a *Sheet B* (or *N-C*), which includes β1+β5, and which is disordered in the TSE (Chu et al., 2013; Neudecker et al., 2012; Riddle et al., 1999; Viguera et al., 1995). The robustness of the folding process is largely attributable to its cooperativity, which stresses the significance of local interactions during folding: residues that initiate the folding process are local in sequence (Baker, 2000; Martínez and Serrano, 1999; Riddle et al., 1999). This finding was further confirmed by circularly permuted constructs, which assumed the same fold despite different sequential orderings of structural elements (Martínez et al., 1999; Viguera et al., 1995). Several SBB-containing structures have entire domains embedded into one of the loops of the barrel, and this apparently does not affect the folding process either (Friberg et al., 2009; Paetzel et al., 2002; Yang et al., 2002). Further details on SBB folding are described in the section on Folding of Small β Barrels in the Supplemental Information.

Structural Stability, Resistance to Thermal and Chemical Denaturation. The spatial compactness and folding robustness of the SBB has ramifications for the thermodynamic stability of SBB-containing proteins. For instance, experimental studies of Sm, LSm, and Hfq homologs have demonstrated that these SBBs resist unfolding by thermal or chemical denaturation; that property, together with their functional versatility, makes them a promising module in protein engineering and design efforts. Additional discussion and literature references are provided in the section on Structural Stability, Resistance to Thermal and Chemical Denaturation in the Supplemental Information.

Conclusions

The SBB domain pervades much of biology, including nucleic acid-related pathways (RNA metabolism, DNA maintenance, ribosome assembly, etc.) as well as other, entirely disparate, milieus (e.g., membrane channels). SBB-containing protein families and superfamilies occur across the Tree of Life, with many representatives conserved in archaeal, bacterial, and eukaryotic lineages. The ancient SBB, which we describe here as an *urfold*, may have arisen in early ribosomal proteins; it appears to have been recruited extensively, over the aeons, to serve myriad functional roles. The SBB often acts as a structural platform to scaffold the assembly of ribonucleoprotein complexes, as a protein module in signal transduction pathways, and as a chaperone of RNA···RNA interactions in bacterial sRNA-mediated regulatory circuits. The lack of a distinct, reliable sequence signature has hampered the identification of SBB proteins via sequence similarity searches; thus, the fractions of SBBs in various genomes, as well as the full breadth of their functional repertoire, remains unknown and may be underestimated. The range of biological activities described here—broad although it may seem—still represents only a subset of known SBB functionalities. Many SCOP folds and even superfolds can be classified as members of the SBB urfold; we limited the scope of this work to those select superfamilies that are highly represented in structural and bioinformatic databases (SH3, Sm/Hfq, and OB; Table S1).

The SBB transcends the classical "fold" or "superfold" concepts, insofar as it encompasses several superfolds (SH3, Sm, OB, etc.) that each contain non-homologous members. Even just one of these superfold families includes a vast swath of biochemistry and cell biology; historically, much emphasis has been on the unique properties of a particular subset of SBB proteins (e.g., Sm proteins and their roles in snRNP cores), rather than on uncovering any unifying principles. That is, any parallels between the many SBB-containing protein families have been lost in a sea of idiosyncrasies for each of the various families, so any recurring themes have gone largely unrecognized. To help identify recurrent themes and patterns, our present work has sought to systematically define and survey the SBB, chiefly in terms of structure ↔ function relationships, and their evolutionary contexts. An initial step has involved terminology: as in many areas of science, alternative (and sometimes incongruous) descriptive schemes and nomenclature systems have emerged for describing closely related entities and the basic relationships between those entities (e.g., for the Sm/Hfq and SH3 superfolds). Thus, we have identified alternative SBB nomenclatures and mapped them to one another as much as possible. More broadly, the SBB protein domain challenges us to develop systematic, formalized structure-description frameworks that can transcend hierarchical classification systems such as SCOP, CATH, and ECOD, in order to accommodate (and precisely "capture") the deep structural similarities and functional plasticity of SBB-like superfolds. We have introduced the term "*urfold*" to capture structural similarities (just above the fold level) while allowing for vast functional diversification.

A hallmark of the SBB urfold is its great variability, in terms of the 3D structures of individual domains (Figure 5), its known oligomeric states and higher-order quaternary structures (Figure 8, Table 2), and its overall functional plasticity (types of cellular pathways; RNA-, DNA-, and protein-binding capacities). Is there any set of general principles—any salient structural, physicochemical, dynamical properties—that account for such deep variation? *How does the SBB achieve such vast functional versatility, while maintaining a stable, unique structural framework that defines it as an urfold (distinct from its neighbors in fold-space)?* In elucidating SBB *sequence* ↔ *structure* ↔ *function* relationships, a key issue is that residue positions that define different regions of an SBB surface (e.g., ligand-binding patch on an Hfq ring) contribute quite differentially to functional properties (e.g., RNA-binding specificity). Variability in the functional relevance of different sites (beyond the residue variation at the sites

themselves) also complicates comparisons of one type of SBB with another (e.g., Hfq versus Tudor), and stymies efforts to decipher *structure* ↔ *function* correlations; notably, such is the case despite the substantial structural similarity between SBB domains. Indeed, the fact that the 3D structure is preserved even when strand order is permuted (i.e., SH3 versus OB superfolds) implies that this domain architecture is a resilient platform for deep sequence variation (and, thus, variation in function). In these ways, the SBB challenges our usual perspective of sequence variation yielding concomitant variation in *structure* ↔ *function* relationships. The SBB's ability to accommodate profound sequence and structural variation also raises basic questions about whether there exist well-defined boundaries of the SBB in fold-space and, assuming so, what may be its nearest structural neighbors (Figure 2). Are there other small β-barrels that are structurally distinct from the SBBs defined in this work? (Are they urfolds, too?) These are open questions.

One emerging theme from our analysis is that the SBB's core barrel is a robustly folding, compact structure for elaborating new biochemical functionality. That a wide range of sequence space can adopt the SBB urfold means that the characteristic spatial arrangement of strands in an SBB (i.e., its architecture) allows for an immense variety of residue combinations (amino acid types, rotamers) to achieve a stable hydrophobic core (Figure 4), upon which further variation can occur (i.e., solvent-exposed residues). This is a principle that bears upon both engineering and evolution. For example, evolutionary drift at relatively exposed sites will alter an SBB's electrostatic properties, thus affording a means to tune interactions with nucleic acids. The termini of the SBB often vary greatly, in terms of the presence/absence of helices or other secondary structural elements, and the loops of the SBB also vary immensely, both in sequence and in length (from a tight β-turn to the insertion of entire functional modules/domains). The extent of possible sequence and structure variation—and hence the range of potential interactions with RNA, DNA, proteins, and ligands—is astounding.

A second emerging theme is the tendency of SBB-containing proteins to oligomerize into biologically functional units. The molecular contacts that stitch together such assemblies are often mediated by the two edge strands of an SBB (as in Sm/Hfq proteins), although this is a general principle that applies beyond the flanking strands: the geometric arrangement of β-strands in the SBB predisposes it to various types of barrel···barrel interactions, including for tandemly repeated barrels, enmeshed barrels, and so on. Oligomerization confers many benefits, in terms of stability and biochemical functionality. For instance, self-assembly into homomeric complexes affords much greater surface area for binding to other biomolecules (e.g., in Hfq···RNA interactions), while the assembly of SBB domains into heteromeric complexes yields the further benefit of enabling asymmetric complexes to form (e.g., the seven Sm paralogs that nucleate the hetero-heptameric snRNP core). Indeed, the oligomeric plasticity and quaternary structural diversity of the SBB may well distinguish it among all known protein domains.

Finally, **a third emerging theme** is the severe modularity of the SBB in most SBB-containing protein families. The ≈60-residue SBB domain occurs throughout the proteome, often as part of a larger polypeptide that extends the SBB core in the N- or C-terminal directions (by just a few residues, or even entire domains). Similarly, anywhere from a few amino acids to >100 residues have been found inserted into the loops of an otherwise intact, canonical SBB domain, and some proteins contain tandem repeats of many SBBs (echoing the behavior of RRM-containing proteins). Such extensive modularity is evolutionarily adaptive. The SBB may be unique in its capacity to function robustly, and broadly, in four distinct ways: (1) as a monomer, on its own; (2) as an oligomer, on its own (or with paralogs); (3) as a domain within multi-domain proteins; and (4) as a structural component in many types of quaternary assemblies.

A detailed analysis of the biological functions of SBBs could have, as one aim, elucidation of the structural mechanisms that underpin an SBB's recognition of different classes of targets (e.g., OB···single-stranded DNA versus OB···protein binding). Notably, the RRM domain resembles the SBB, at least superficially, in terms of structure/function relationships: the RRM is a small, four-stranded antiparallel β-sheet (with helices at both termini), it binds RNA (as do many SBB proteins), it exhibits a great degree of structural variation, and it is functionally quite versatile (interacting with a variety of possible ligands, including RNA, DNA, and other proteins). Thus, intriguing directions to explore include defining other urfolds that lie near the SBB in fold-space (which are superfolds, and which can be grouped into urfolds?), as well as analyzing how the SBB urfold compares with β-rich superfolds such as ferredoxin—in terms of range of structural variation, breadth of molecular functions, and factors governing the structure ↔ function relationships within (and between) these folds and superfolds. Determining the fundamental structural and physicochemical principles that enable the deep modularity and functional plasticity of SSBs, RRMs, and other β-rich structures represents a broadly stimulating area for future work.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experiment Procedures, two figures, and two tables and can be found with this article online at https://doi.org/10.1016/j.str.2018.09.012.

ACKNOWLEDGMENTS

Portions of this work were supported by the Jeffress Memorial Trust (J-971; C.M.) and NSF Career award MCB-1350957 (C.M.).

AUTHOR CONTRIBUTIONS

P.Y. conceived of the study and performed the initial analyses; S.V. guided the project scope, conducted systematic analyses, and engaged in all stages of manuscript preparation; Q.L. generated initial molecular graphics; K.A.S. computed electrostatic maps, generated figures, and helped analyze structures and draft the manuscript; C.M. performed systematic analyses, generated figures, and helped organize and draft the manuscript; P.E.B. supervised the project and helped draft the manuscript. All authors discussed the work and contributed to the manuscript.

REFERENCES

Agrawal, V., and Kishan, R.K. (2001). Functional evolution of two subtly different (similar) folds. BMC Struct. Biol. *1*, 5.

Albrecht, M., and Lengauer, T. (2004). Novel Sm-like proteins with long C-terminal tails and associated methyltransferases. FEBS Lett. *569*, 18–26.

Alva, V., Remmert, M., Biegert, A., Lupas, A.N., and Soding, J. (2010). A galaxy of folds. Protein Sci. *19*, 124–130.

Alva, V., Soding, J., and Lupas, A.N. (2015). A vocabulary of ancient peptides at the origin of folded proteins. Elife *4*, e09410.

Baker, D. (2000). A surprising simplicity to protein folding. Nature *405*, 39–42.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U S A *98*, 10037–10041.

Beich-Frandsen, M., Vecerek, B., Konarev, P.V., Sjoblom, B., Kloiber, K., Hammerle, H., Rajkowitsch, L., Miles, A.J., Kontaxis, G., Wallace, B.A., et al. (2011). Structural insights into the dynamics and function of the C-terminus of the *E. coli* RNA chaperone Hfq. Nucleic Acids Res. *39*, 4900–4915.

Blus, B.J., Wiggins, K., and Khorasanizadeh, S. (2011). Epigenetic virtues of chromodomains. Crit. Rev. Biochem. Mol. Biol. *46*, 507–526.

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

Bochkareva, E., Korolev, S., Lees-Miller, S.P., and Bochkarev, A. (2002). Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. EMBO J. *21*, 1855–1863.

Bork, P., Holm, L., and Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. J. Mol. Biol. *242*, 309–320.

Brodersen, D.E., Clemons, W.M., Jr., Carter, A.P., Wimberly, B.T., and Ramakrishnan, V. (2002). Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: structure of the proteins and their interactions with 16 S RNA. J. Mol. Biol. *316*, 725–768.

Caetano-Anolles, G., and Caetano-Anolles, D. (2003). An evolutionarily structured universe of protein architecture. Genome Res. *13*, 1563–1571.

Camara-Artigas, A. (2016). Crystallographic studies on protein misfolding: domain swapping and amyloid formation in the SH3 domain. Arch. Biochem. Biophys. *602*, 116–126.

Chandonia, J.M., Fox, N.K., and Brenner, S.E. (2017). SCOPe: manual curation and artifact removal in the structural classification of proteins - extended database. J. Mol. Biol. *429*, 348–355.

Chang, Y., Horton, J.R., Bedford, M.T., Zhang, X., and Cheng, X. (2011). Structural insights for MPP8 chromodomain interaction with histone H3 lysine 9: potential effect of phosphorylation on methyl-lysine binding. J. Mol. Biol. *408*, 807–814.

Charier, G., Couprie, J., Alpha-Bazin, B., Meyer, V., Quemeneur, E., Guerois, R., Callebaut, I., Gilquin, B., and Zinn-Justin, S. (2004). The Tudor tandem of 53BP1: a new structural motif involved in DNA and RG-rich peptide binding. Structure *12*, 1551–1562.

Cheng, H., Kim, B.H., and Grishin, N.V. (2008). Discrimination between distant homologs and structural analogs: lessons from manually constructed, reliable data sets. J. Mol. Biol. *377*, 1265–1278.

Cheng, H., Schaeffer, R.D., Liao, Y., Kinch, L.N., Pei, J., Shi, S., Kim, B.H., and Grishin, N.V. (2014). ECOD: an evolutionary classification of protein domains. PLoS Comput. Biol. *10*, e1003926.

Chothia, C., and Janin, J. (1982). Orthogonal packing of beta-pleated sheets in proteins. Biochemistry *21*, 3955–3965.

Chu, W.T., Zhang, J.L., Zheng, Q.C., Chen, L., and Zhang, H.X. (2013). 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 *8*, e64886.

Cléry, A., Blatter, M., and Allain, F.H. (2008). RNA recognition motifs: boring? Not quite. Curr. Opin. Struct. Biol. *18*, 290–298.

Cridland, J.A., Curley, E.Z., Wykes, M.N., Schroder, K., Sweet, M.J., Roberts, T.L., Ragan, M.A., Kassahn, K.S., and Stacey, K.J. (2012). The mammalian PYHIN gene family: phylogeny, evolution and expression. BMC Evol. Biol. *12*, 140.

Das, D., Kozbial, P., Axelrod, H.L., Miller, M.D., McMullan, D., Krishna, S.S., Abdubek, P., Acosta, C., Astakhova, T., Burra, P., et al. (2009). Crystal structure of a novel Sm-like protein of putative cyanophage origin at 2.60 A resolution. Proteins *75*, 296–307.

Dawson, N.L., Lewis, T.E., Das, S., Lees, J.G., Lee, D., Ashford, P., Orengo, C.A., and Sillitoe, I. (2017). CATH: an expanded resource to predict protein function through structure and sequence. Nucleic Acids Res. *45*, D289–D295.

de Jong, R.N., Truffault, V., Diercks, T., Ab, E., Daniels, M.A., Kaptein, R., and Folkers, G.E. (2008). Structure and DNA binding of the human Rtf1 Plus3 domain. Structure *16*, 149–159.

Demo, G., Rasouly, A., Vasilyev, N., Svetlov, V., Loveland, A.B., Diaz-Avalos, R., Grigorieff, N., Nudler, E., and Korostelev, A.A. (2017). Structure of RNA polymerase bound to ribosomal 30S subunit. Elife *6*, https://doi.org/10.7554/eLife.28560.

Dessailly, B.H., Dawson, N.L., Das, S., and Orengo, C.A. (2017). Function diversity within folds and superfamilies. In From Protein Structure to Function with Bioinformatics, D.J. Rigden, ed. (Springer), pp. 295–325.

Dever, T.E., Gutierrez, E., and Shin, B.S. (2014). The hypusine-containing translation factor eIF5A. Crit. Rev. Biochem. Mol. Biol. *49*, 413–425.

Dickey, T.H., Altschuler, S.E., and Wuttke, D.S. (2013). Single-stranded DNA-binding proteins: multiple domains for multiple functions. Structure *21*, 1074–1084.

Diedrich, G., Spahn, C.M., Stelzl, U., Schafer, M.A., Wooten, T., Bochkariov, D.E., Cooperman, B.S., Traut, R.R., and Nierhaus, K.H. (2000). Ribosomal protein L2 is involved in the association of the ribosomal subunits, tRNA binding to A and P sites and peptidyl transfer. EMBO J. *19*, 5241–5250.

Fairman, J.W., Noinaj, N., and Buchanan, S.K. (2011). The structural biology of beta-barrel membrane proteins: a summary of recent reports. Curr. Opin. Struct. Biol. *21*, 523–531.

Flynn, R.L., and Zou, L. (2010). Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians. Crit. Rev. Biochem. Mol. Biol. *45*, 266–275.

Friberg, A., Corsini, L., Mourao, A., and Sattler, M. (2009). Structure and ligand binding of the extended Tudor domain of *D. melanogaster* Tudor-SN. J. Mol. Biol. *387*, 921–934.

Friesen, W.J., Paushkin, S., Wyce, A., Massenet, S., Pesiridis, G.S., Van Duyne, G., Rappsilber, J., Mann, M., and Dreyfuss, G. (2001). The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. Mol. Cell. Biol. *21*, 8289–8300.

Giraud, P., Crechet, J.B., Uzan, M., Bontems, F., and Sizun, C. (2015). Resonance assignment of the ribosome binding domain of *E. coli* ribosomal protein S1. Biomol. NMR Assign. *9*, 107–111.

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

Good, M.C., Zalatan, J.G., and Lim, W.A. (2011). Scaffold proteins: hubs for controlling the flow of cellular information. Science *332*, 680–686.

Gorski, S.A., Vogel, J., and Doudna, J.A. (2017). RNA-based recognition and targeting: sowing the seeds of specificity. Nat. Rev. Mol. Cell Biol. *18*, 215–228.

Gregory, S.T., Carr, J.F., and Dahlberg, A.E. (2009). A signal relay between ribosomal protein S12 and elongation factor EF-Tu during decoding of mRNA. RNA *15*, 208–214.

Grimm, C., Chari, A., Pelz, J.P., Kuper, J., Kisker, C., Diederichs, K., Stark, H., Schindelin, H., and Fischer, U. (2013). Structural basis of assembly chaperone-mediated snRNP formation. Mol. Cell *49*, 692–703.

Grishin, N.V. (2001). Fold change in evolution of protein structures. J. Struct. Biol. *134*, 167–185.

Hajnsdorf, E., and Boni, I.V. (2012). Multiple activities of RNA-binding proteins S1 and Hfq. Biochimie *94*, 1544–1553.

Harish, A., and Caetano-Anolles, G. (2012). Ribosomal history reveals origins of modern protein synthesis. PLoS One *7*, e32776.

Harrison, A., Pearl, F., Mott, R., Thornton, J., and Orengo, C. (2002). Quantifying the similarities within fold space. J. Mol. Biol. *323*, 909–926.

Hogue, C.W. (1997). Cn3D: a new generation of three-dimensional molecular structure viewer. Trends Biochem. Sci. *22*, 314–316.

Huang, Y., Fang, J., Bedford, M.T., Zhang, Y., and Xu, R.M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. Science *312*, 748–751.

Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080–2083.

Jin, T., Perry, A., Jiang, J., Smith, P., Curry, J.A., Unterholzner, L., Jiang, Z., Horvath, G., Rathinam, V.A., Johnstone, R.W., et al. (2012). Structures of the HIN domain:DNA complexes reveal ligand binding and activation mechanisms of the AIM2 inflammasome and IFI16 receptor. Immunity *36*, 561–571.

Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Luhrmann, R., Li, J., and Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell *96*, 375–387.

Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F., and Rossmann, M.G. (2002). Structure of the cell-puncturing device of bacteriophage T4. Nature *415*, 553–557.

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

Kilic, T., Sanglier, S., Van Dorsselaer, A., and Suck, D. (2006). Oligomerization behavior of the archaeal Sm2-type protein from *Archaeoglobus fulgidus*. Protein Sci. *15*, 2310–2317.

Klein, D.J., Moore, P.B., and Steitz, T.A. (2004). The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. J. Mol. Biol. *340*, 141–177.

Krug, M., Lee, S.J., Diederichs, K., Boos, W., and Welte, W. (2006). Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB. J. Biol. Chem. *281*, 10976–10982.

Kyrpides, N.C., Woese, C.R., and Ouzounis, C.A. (1996). KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. Trends Biochem. Sci. *21*, 425–426.

le Maire, A., Schiltz, M., Stura, E.A., Pinon-Lataillade, G., Couprie, J., Moutiez, M., Gondry, M., Angulo, J.F., and Zinn-Justin, S. (2006). A tandem of SH3-like domains participates in RNA binding in KIN17, a human protein activated in response to genotoxics. J. Mol. Biol. *364*, 764–776.

Levy, Y., Onuchic, J.N., and Wolynes, P.G. (2007). Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. J. Am. Chem. Soc. *129*, 738–739.

Li, J., Leung, A.K., Kondo, Y., Oubridge, C., and Nagai, K. (2016). Re-refinement of the spliceosomal U4 snRNP core-domain structure. Acta Crystallogr. D Struct. Biol. *72*, 131–146.

Li, W., Giles, C., and Li, S. (2014). Insights into how Spt5 functions in transcription elongation and repressing transcription coupled DNA repair. Nucleic Acids Res. *42*, 7069–7083.

Lim, W.A. (1996). Reading between the lines: SH3 recognition of an intact protein. Structure *4*, 657–659.

Liu, K., Chen, C., Guo, Y., Lam, R., Bian, C., Xu, C., Zhao, D.Y., Jin, J., MacKenzie, F., Pawson, T., and Min, J. (2010). Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. Proc. Natl. Acad. Sci. U S A *107*, 18398–18403.

Liu, Y., and Eisenberg, D. (2002). 3D domain swapping: as domains continue to swap. Protein Sci. *11*, 1285–1299.

Lomakin, I.B., and Steitz, T.A. (2013). The initiation of mammalian protein synthesis and mRNA scanning mechanism. Nature *500*, 307–311.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Luo, Y., Pfuetzner, R.A., Mosimann, S., Paetzel, M., Frey, E.A., Cherney, M., Kim, B., Little, J.W., and Strynadka, N.C. (2001). Crystal structure of LexA: a conformational switch for regulation of self-cleavage. Cell *106*, 585–594.

Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318–322.

Martínez, J.C., and Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. Nat. Struct. Biol. *6*, 1010–1016.

Martínez, J.C., Viguera, A.R., Berisio, R., Wilmanns, M., Mateo, P.L., Filimonov, V.V., and Serrano, L. (1999). 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 *38*, 549–559.

McCarty, J.H. (1998). The Nck SH2/SH3 adaptor protein: a regulator of multiple intracellular signal transduction events. Bioessays *20*, 913–921.

McLachlan, A.D. (1979). Gene duplications in the structural evolution of chymotrypsin. J. Mol. Biol. *128*, 49–79.

Meyer, P.A., Li, S., Zhang, M., Yamada, K., Takagi, Y., Hartzog, G.A., and Fu, J. (2015). Structures and functions of the multiple KOW domains of transcription elongation factor Spt5. Mol. Cell. Biol. *35*, 3354–3369.

Mitton-Fry, R.M., Anderson, E.M., Theobald, D.L., Glustrom, L.W., and Wuttke, D.S. (2004). Structural basis for telomeric single-stranded DNA recognition by yeast Cdc13. J. Mol. Biol. *338*, 241–255.

Morton, C.J., and Campbell, I.D. (1994). SH3 domains. Molecular 'Velcro'. Curr. Biol. *4*, 615–617.

Mura, C., Cascio, D., Sawaya, M.R., and Eisenberg, D.S. (2001). The crystal structure of a heptameric archaeal Sm protein: implications for the eukaryotic snRNP core. Proc. Natl. Acad. Sci. U S A *98*, 5532–5537.

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

Mura, C., Phillips, M., Kozhukhovsky, A., and Eisenberg, D. (2003b). Structure and assembly of an augmented Sm-like archaeal protein 14-mer. Proc. Natl. Acad. Sci. U S A *100*, 4539–4544.

Mura, C., Randolph, P.S., Patterson, J., and Cozen, A.E. (2013). Archaeal and eukaryotic homologs of Hfq: a structural and evolutionary perspective on Sm function. RNA Biol. *10*, 636–651.

Murzin, A.G., Lesk, A.M., and Chothia, C. (1994a). Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis. J. Mol. Biol. *236*, 1369–1381.

Murzin, A.G., Lesk, A.M., and Chothia, C. (1994b). Principles determining the structure of beta-sheet barrels in proteins. II. The observed structures. J. Mol. Biol. *236*, 1382–1400.

Myrick, L.K., Hashimoto, H., Cheng, X., and Warren, S.T. (2015). Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. Hum. Mol. Genet. *24*, 1733–1740.

Nagano, N., Orengo, C.A., and Thornton, J.M. (2002). One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J. Mol. Biol. *321*, 741–765.

Nakagawa, A., Nakashima, T., Taniguchi, M., Hosaka, H., Kimura, M., and Tanaka, I. (1999). 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. *18*, 1459–1467.

Neudecker, P., Robustelli, P., Cavalli, A., Walsh, P., Lundstrom, P., Zarrine-Afsar, A., Sharpe, S., Vendruscolo, M., and Kay, L.E. (2012). Structure of an intermediate state in protein folding and aggregation. Science *336*, 362–366.

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

Paetzel, M., Dalbey, R.E., and Strynadka, N.C. (2002). Crystal structure of a bacterial signal peptidase apoenzyme: implications for signal peptide binding and the Ser-Lys dyad mechanism. J. Biol. Chem. *277*, 9512–9519.

Panecka, J., Mura, C., and Trylska, J. (2014). Interplay of the bacterial ribosomal A-site, S12 protein mutations and paromomycin binding: a molecular dynamics study. PLoS One *9*, e111811.

Patel, D.J., and Wang, Z. (2013). Readout of epigenetic modifications. Annu. Rev. Biochem. *82*, 81–118.

Pei, J., Kim, B.H., and Grishin, N.V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. *36*, 2295–2300.

Perozo, E., and Rees, D.C. (2003). Structure and mechanism in prokaryotic mechanosensitive channels. Curr. Opin. Struct. Biol. *13*, 432–442.

Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. Nature *458*, 475–480.

Ren, R., Liu, H., Wang, W., Wang, M., Yang, N., Dong, Y.H., Gong, W., Lehmann, R., and Xu, R.M. (2014). Structure and domain organization of *Drosophila* Tudor. Cell Res. *24*, 1146–1149.

Riddle, D.S., Grantcharova, V.P., Santiago, J.V., Alm, E., Ruczinski, I., and Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. Nat. Struct. Biol. *6*, 1016–1024.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, W320–W324.

Robinson, H., Gao, Y.G., McCrary, B.S., Edmondson, S.P., Shriver, J.W., and Wang, A.H. (1998). The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. Nature *392*, 202–205.

Rost, B. (1999). Twilight zone of protein sequence alignments. Protein Eng. *12*, 85–94.

Saltzman, A.L., Pan, Q., and Blencowe, B.J. (2011). Regulation of alternative splicing by the core spliceosomal machinery. Genes Dev. *25*, 373–384.

Santiago-Frangos, A., Jeliazkov, J.R., Gray, J.J., and Woodson, S.A. (2017). Acidic C-terminal domains autoregulate the RNA chaperone Hfq. Elife *6*, https://doi.org/10.7554/eLife.27049.

Santiago-Frangos, A., and Woodson, S.A. (2018). Hfq chaperone brings speed dating to bacterial sRNA. Wiley Interdiscip. Rev. RNA *9*, e1475.

Sauer, E. (2013). Structure and RNA-binding properties of the bacterial LSm protein Hfq. RNA Biol. *10*, 610–618.

Schumacher, M.A., Pearson, R.F., Moller, T., Valentin-Hansen, P., and Brennan, R.G. (2002). Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. EMBO J. *21*, 3546–3556.

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

Shaw, N., and Liu, Z.J. (2014). Role of the HIN domain in regulation of innate immune responses. Mol. Cell. Biol. *34*, 2–15.

Shoemaker, B.A., Portman, J.J., and Wolynes, P.G. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc. Natl. Acad. Sci. U S A *97*, 8868–8873.

Sidote, D.J., Heideker, J., and Hoffman, D.W. (2004). Crystal structure of archaeal ribonuclease P protein aRpp29 from *Archaeoglobus fulgidus*. Biochemistry *43*, 14128–14138.

Stanek, K.A., Patterson-West, J., Randolph, P.S., and Mura, C. (2017). Crystal structure and RNA-binding properties of an Hfq homolog from the deep-branching Aquificae: conservation of the lateral RNA-binding mode. Acta Crystallogr. D Struct. Biol. *73*, 294–315.

Stein, P.E., Boodhoo, A., Tyrrell, G.J., Brunton, J.L., and Read, R.J. (1992). Crystal structure of the cell-binding B oligomer of verotoxin-1 from *E. coli*. Nature *355*, 748–750.

Steinbacher, S., Bass, R., Strop, P., and Rees, D.C. (2007). Structures of the prokaryotic mechanosensitive channels MscL and MscS. In Current Topics in Membranes: Mechanosensitive Ion Channels Part A, Owen P. Hamill, ed (Academic Press), pp. 1–24.

Stoll, R., Renner, C., Zweckstetter, M., Bruggert, M., Ambrosius, D., Palme, S., Engh, R.A., Golob, M., Breibach, I., Buettner, R., et al. (2001). The extracellular human melanoma inhibitory activity (MIA) protein adopts an SH3 domain-like fold. EMBO J. *20*, 340–349.

Sun, X., and Wartell, R.M. (2006). *Escherichia coli* Hfq binds A_{18} and DsrA domain II with similar 2:1 Hfq_6/RNA stoichiometry using different surface sites. Biochemistry *45*, 4875–4887.

Tamm, L.K., Hong, H., and Liang, B. (2004). Folding and assembly of beta-barrel membrane proteins. Biochim. Biophys. Acta *1666*, 250–263.

Theobald, D.L., and Wuttke, D.S. (2005). Divergent evolution within protein superfolds inferred from profile-based phylogenetics. J. Mol. Biol. *354*, 722–737.

Thore, S., Mayer, C., Sauter, C., Weeks, S., and Suck, D. (2003). Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. J. Biol. Chem. *278*, 1239–1247.

Tsokos, G.C. (2006). In the beginning was Sm. J. Immunol. *176*, 1295–1296.

Valle, M., Sengupta, J., Swami, N.K., Grassucci, R.A., Burkhardt, N., Nierhaus, K.H., Agrawal, R.K., and Frank, J. (2002). Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. EMBO J. *21*, 3557–3567.

Viguera, A.R., Blanco, F.J., and Serrano, L. (1995). The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. J. Mol. Biol. *247*, 670–681.

Vogel, J., and Luisi, B.F. (2011). Hfq and its constellation of RNA. Nat. Rev. Microbiol. *9*, 578–589.

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

Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. *3*, https://doi.org/10.1101/cshperspect.a003707.

Wilusz, C.J., and Wilusz, J. (2013). Lsm proteins and Hfq: life at the 3' end. RNA Biol. *10*, 592–601.

Wu, D., Muhlrad, D., Bowler, M.W., Jiang, S., Liu, Z., Parker, R., and Song, H. (2014). Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. Cell Res. *24*, 233–246.

Yang, H., Jeffrey, P.D., Miller, J., Kinnucan, E., Sun, Y., Thoma, N.H., Zheng, N., Chen, P.L., Lee, W.H., and Pavletich, N.P. (2002). BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. Science *297*, 1837–1848.

Yin, Q., Sester, D.P., Tian, Y., Hsiao, Y.S., Lu, A., Cridland, J.A., Sagulenko, V., Thygesen, S.J., Choubey, D., Hornung, V., et al. (2013). Molecular mechanism for p202-mediated specific inhibition of AIM2 inflammasome activation. Cell Rep. *4*, 327–339.

Yu, H., Rosen, M.K., Shin, T.B., Seidel-Dugan, C., Brugge, J.S., and Schreiber, S.L. (1992). Solution structure of the SH3 domain of Src and identification of its ligand-binding site. Science *258*, 1665–1668.

Zhang, C., and Kim, S.H. (2000). The anatomy of protein β-sheet topology. J. Mol. Biol. *299*, 1075–1089.